Biophysical Journal Volume 69 September 1995 1046–1067

Chain Folding and A:T Pairing in Human Telomeric DNA: A Model-Building and Molecular Dynamics Study

Debasisa Mohanty and Manju Bansal
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

ABSTRACT The various types of chain folding and possible intraloop as well as interloop base pairing in human telomeric DNA containing d(TTAG$_3$) repeats have been investigated by model-building, molecular mechanics, and molecular dynamics techniques. Model-building and molecular mechanics studies indicate that it is possible to build a variety of energetically favorable folded-back structures with the two TTA loops on same side and the 5′ end thymines in the two loops forming TATA tetrads involving a number of different intraloop as well as interloop A:T pairing schemes. In these folded-back structures, although both intraloop and interloop Watson–Crick pairing is feasible, no structure is possible with interloop Hoogsteen pairing. MD studies of representative structures indicate that the guanine-tetraplex stem is very rigid and, while the loop regions are relatively much more flexible, most of the hydrogen bonds remain intact throughout the 350-ps in vacuo simulation. The various possible TTA loop structures, although they are energetically similar, have characteristic inter proton distances, which could give rise to unique cross-peaks in two-dimensional nuclear Overhauser effect spectroscopy (NOESY) experiments. These folded-back structures with A:T pairings in the loop region help in rationalizing the data from chemical probing and other biochemical studies on human telomeric DNA.

INTRODUCTION

The ends of eukaryotic chromosomes, called telomeres, are known to play an important role in the stability of chromosomes and complete replication of chromosomal ends (Blackburn and Szostak, 1984; Henderson and Blackburn, 1989; Blackburn, 1991, 1992). The telomeric DNA contains a double-stranded region with tandem repeats of short guanine-rich nucleotides at the 3′ end, and the guanine-rich strand of the duplex terminates with a short single-stranded overhang of this repeat sequence. The telomeric repeat in general consists of runs of four contiguous guanines interspersed with stretches of two to four thymines (Blackburn and Szostak, 1984; Henderson and Blackburn, 1989; Blackburn, 1991, 1992) or of an adenine (A) and three contiguous guanines (Gs) interspersed with stretches of two to three thymines (Ts) (Moyzis et al., 1988). The exact sequence repeat varies among species, for example, while sequence motifs of the type d(T_2G_4) and d(T_4G_4) are found in the telomeres of *Tetrahymena* and *Oxytricha*, respectively, in the human telomeric DNA d(TTAG$_3$) repeats are observed, whereas the telomeric repeat for *Arabidopsis* is d(T_3AG_3) (Blackburn and Szostak, 1984; Henderson and Blackburn, 1989; Blackburn, 1991, 1992; Moyzis et al., 1988). The presence of guanine stretches gives these sequence motifs the ability to adopt novel four-stranded structures (Sen and Gilbert, 1988, 1990; Sundquist and Klug, 1989; Henderson et al., 1987; Williamson et al., 1989) because of the formation of Hoogsteen bonded guanine tetrads (Sasisekharan et al., 1975; Pinnavaia et al., 1975; Guschlbauer et al., 1990). Apart from telomeric DNA, the G-tetraplex structures adopted by guanine-rich oligonucleotides have in general received wide attention, because a number of proteins have been identified that bind specifically to these structures (Walsh and Gualberto, 1992; Pearson et al., 1993). Formation of G-tetraplex structure has also been postulated to be crucial for dimerization of HIV RNA (Sundquist and Heaphy, 1993). Guanine-rich oligonucleotides are also known to be inhibitors for fibrinogen action in thrombin (Wang et al., 1993; Macaya et al., 1993; Schultze et al., 1994, Padmanabhan et al., 1993) and HIV viral mediated cell fusion (Wyatt et al., 1994).

The conformational features of these G-tetraplex structures have been investigated by both theoretical (Panyutin et al., 1990; Balagurumoorthy et al., 1992; Mohanty and Bansal, 1993, 1994; Ross and Hardin, 1994) and experimental (Arnott et al., 1974; Zimmermann et al., 1975; Jin et al., 1990; Balagurumoorthy et al., 1992; Guo et al., 1992a; Scaria et al., 1992; Wang and Patel, 1992; Aboul-ela et al., 1992; Kang et al., 1992; Smith and Feigon, 1992, 1993) methods. A recent review (Williamson, 1994) gives a detailed account of the various studies of G-tetraplex structures that have been reported. These studies indicate that sequences with isolated stretches of guanines adopt tetramolecular structures by association of four guanine strands in a parallel orientation, and all guanines remain in anti conformation (Wang and Patel, 1992; Cheong and Moore, 1992; Aboul-ela et al., 1994; Gupta et al., 1993). This observation has been further supported by the recently reported high-resolution crystal structure of d(TG_4T) (Laughlan et al., 1994). On the other hand, sequences with stretches of guanines interspersed by thymines adopt folded-back structures with a G-tetraplex stem and thymines forming loops con-

Received for publication 27 February 1995 and in final form 25 May 1995.

Address reprint requests to Dr. Manju Bansal, Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India. Tel.: 91–080-3092534; Fax: 91–080-3341683; E-mail: mb@mbu.iisc.ernet.in. Dr. Mohanty's current address: Department of Physical Chemistry, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.

0006-3495/95/09/1046/22 $2.00

necting the adjacent or diagonal strands of the G tetraplex (Kang et al., 1992; Smith and Feigon, 1992, 1993). Depending on the number of guanine stretches present in a strand, such folded-back structures could arise from dimerization of two hairpins (Kang et al., 1992; Smith and Feigon, 1992, 1993) or intramolecular folding of a single strand (Wang et al., 1993; Macaya et al., 1993; Schultze et al.; 1994, Padmanabhan et al., 1993; Wang and Patel, 1993), and such structures necessarily have adjacent or diagonal strands in an antiparallel orientation. The guanines in these antiparallel G-tetraplex structures generally adopt an alternating *syn–anti* conformation along the strand. The x-ray crystallographic structure for $d(G_4T_4G_4)$ (Kang et al., 1992) has a hairpin dimer arrangement with T_4 loops over adjacent strands of a G tetraplex occurring on opposite ends of the tetraplex stem. However, the structure of the same sequence in solution when studied by nuclear magnetic resonance (NMR) has been found to have the T_4 loops connecting the diagonal strands (Smith and Feigon, 1992, 1993). In addition, a number of hairpin dimer structures with two thymine loops on the same side and interloop T:T pairing have also been proposed from theoretical studies (Balagurumoorthy et al., 1992; Bansal and Mohanty, 1993; Mohanty and Bansal, 1994) for the sequence motifs $d(G_4T_nG_4)$, with $n = 2–4$. Thus the various folded-back and hairpin dimer structures for telomeric DNA differ not only in the folding pattern but also in the conformation and orientation of thymines in the loop region. The possibility of occurrence of thymine pairs in the loop region of telomeric DNA had been first proposed from model-building studies (Balagurumoorthy et al., 1992; Bansal and Mohanty, 1993; Mohanty and Bansal, 1994). In the molecular dynamics studies (Ross and Hardin, 1994) of the hairpin dimer structure of $d(G_4T_2G_4)$ two interloop T:T pairs were observed in the final structure, even though the simulation was carried out with several starting structures without T:T pairs. On the other hand, no T:T pair was found in the NMR and crystal structures for the telomeric DNA. However, subsequently a Greek-key type of structure with two T_2 loops on the same side and a single T:T pair between the two T_2 loops was observed in the NMR (Macaya et al., 1993; Schultze et al., 1994) and x-ray crystallographic (Padmanabhan et al., 1993) studies of $d(G_2T_2G_2TGTG_2T_2G_2)$, even though the folding patterns in crystal and solution were topologically different. Similarly, a recent NMR study (Wang et al., 1994) of the oligonucleotide $d(G_4UTUTG_4TUTUG_4UUTTG_4)$ has indicated that this sequence folds into a Greek-key motif with T:T and U:U pairs being present in the loop region, whereas earlier NMR studies of a very similar sequence $d(T_4G_4)_4$ (Smith and Feigon, 1992, 1993) had indicated that it folds into an Indian-key motif with no T:T pairs. The variety of studies of telomeric DNA reported indicate that the structure is quite polymorphic, and the major source of the conformational variability is the conformation and orientation of the loop nucleotides, whereas the guanine stem has relatively less conformational variability.

In view of the greater conformational variability of the loop regions in telomeric DNA, it is interesting to investigate the possible structures adopted by the human telomeric sequence $d(TTAG_3)_m$ or its variant $d(T_3AG_3)_m$ found in *Arabidopsis*, because the presence of adenine in these sequences gives rise to the possibility of the occurrence of A:G or A:T pairs in the structure. However, relatively few studies (Balagurumoorthy and Brahmachari, 1994; Guo et al., 1992b) of telomeric DNA of humans or *Arabidopsis* have been reported in contrast to the large number of experimental studies of sequences of the type $d(T_nG_4)_m$. The single NMR study (Wang and Patel, 1993) of $d(AG_3(TTAG_3)_3)$ indicates that the molecule folds into an Indian-key motif with a G_3 stem and TTA loops, but no base pairing is observed in the loops. On the other hand, chemical probing studies (Balagurumoorthy and Brahmachari, 1994) of $d(G_3TTAG_3)$ and $d(TTAG_3)$ repeats of different lengths indicate that N7 of adenine is inaccessible, thus suggesting the possibility of the participation of adenine in hydrogen bonding interactions. Similarly, the higher stability of $d(G_4TTAG_4)$ compared with $d(G_4T_3G_4)$ has been explained on the basis of the occurrence of A:T pairs in the loop regions (Guo et al., 1992b). Therefore, in the present theoretical study, model-building, molecular mechanics, and molecular dynamics techniques have been used to elucidate various possible structures for $d(T_nAG_3)$ repeats.

The specific questions addressed in this study are the following: 1. Will these sequences fold, with a T_n loop and AG_3 stem consisting of two G tetrads and two AGAG tetrads or with a T_nA loop and G_3 stem? 2. Is it possible to have A:T pairs in the loop regions of folded-back structures with G_3 stems and T_nA loops? If so, what are the various types of A:T pairs compatible with different types of chain folding? 3. Are the hydrogen bonds in the loop region really stable?

MODEL BUILDING

A sequence of the type $d(T_nAG_3)_4$ or $d(G_3(T_nAG_3)_3)$ can adopt a monomolecular structure by Greek-key- or Indian-key-type folding (Frank-Kamenetskii, 1992), and these structures can have a G-tetraplex stem with three guanine tetrads (Fig. 1 *a* and *b*) and loops containing adenine and thymine residues; alternatively, adenine could occur in the stem, resulting in structures with the tetraplex stem consisting of two G tetrads and two AGAG tetrads (Fig. 1 *c* and *d*) and only thymines constituting the loops. The structures with AGAG tetrads in the stem and T_n loops would be similar to the structures of $d(T_nG_4)_4$ (Mohanty and Bansal, 1994), as the AGAG tetrad is isostructural with the guanine tetrad. Structures with G_3 stems and T_nA loops are expected to be different from their thymine counterparts from the point of view of stacking and hydrogen bonding in the loop region, because the first thymine and the adenine could in principle form intraloop as well as interloop A:T pairing involving Watson-Crick and Hoogsteen types of hydrogen bonding. Moreover, sequences of the type $d(G_3T_nAG_3)$ are expected to adopt only hairpin dimer structures with G_3 stems and T_nA loops. Inasmuch as all possible intramolecular folded-back structures with G_3 stems and T_nA loops can in principle be obtained from a combination of hairpins, model building was carried out for different possible hairpin dimer structures for $d(G_3T_nAG_3)$ with G_3 stems and T_nA loops.

The sequences of the type $d(G_3T_nAG_3)$ can form three types of hairpin dimer with the two T_nA loops over adjacent strands and occurring on the

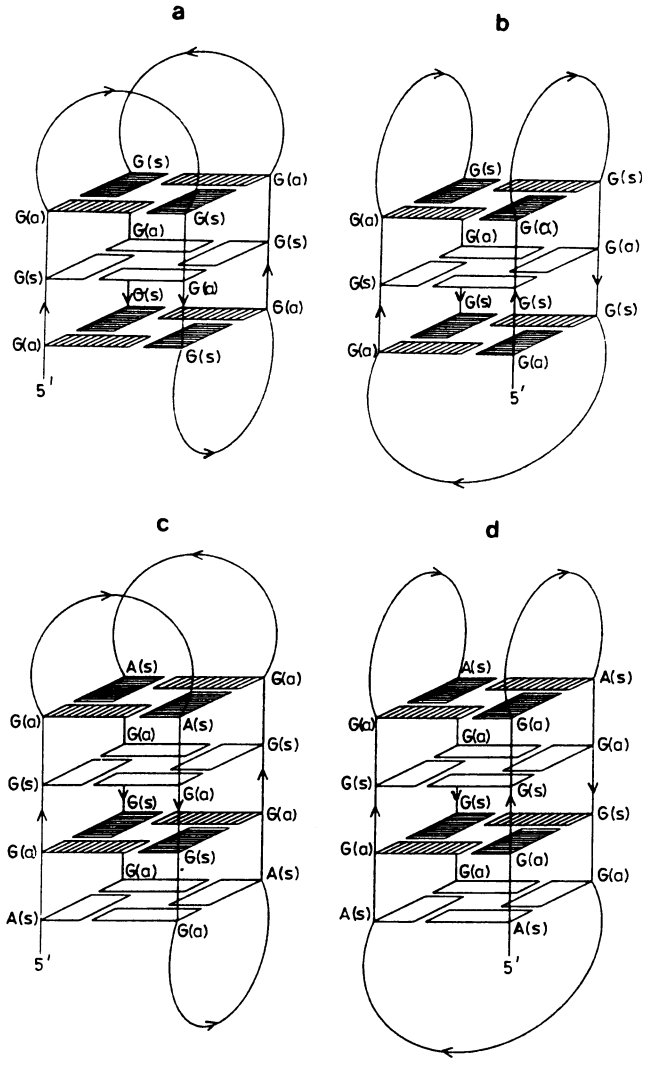

FIGURE 1 Schematic diagrams of the monomolecular structures arising from intramolecular folding of the sequence $d(T_nAG_3)_4$, n = 2, 3. (*a*) Greek-key-type folding with G_3 in the stem and T_nA in the loop. (*b*) Indian-key-type folding with G_3 in the stem and T_nA in the loop. (*c*) Greek-key-type folding with AG_3 in the stem and T_2 or T_3 in the loop. (*d*) Indian-key-type folding with AG_3 in the stem and T_3 in the loop. This type of structure is not possible for $d(T_2AG_3)_4$, as a minimum of three nucleotides are required for building a loop across the diagonal strands in a G tetraplex. The guanines and adenine in the tetraplex stem have been represented as rectangular planes, with the longer edge indicating the side of guanine that contains N1 and N2 or the side of adenine that contains N1 and C2 and the shorter edge indicating the side containing N7 of guanine or adenine. The alternate guanines along a strand, which have their α faces visible from the top, have been shaded to show alternating syn–anti arrangement of guanines along the strand resulting from base flipover (Mohanty and Bansal, 1993).

same side of the tetraplex stem (Fig. 2 *a*) or on opposite ends (Fig. 2 *b*), or with the loops over diagonal strands and at opposite ends of the G tetraplex (Fig. 2 *c*). Even though the guanines in the antiparallel G-tetraplex stem of these hairpin dimers can have all-syn–all-anti or alternating syn–anti conformation (Mohanty and Bansal, 1993, 1994; Balagurumoorthy et al., 1992), only the ones with an alternating syn–anti arrangement of guanines along the strand have been observed in x-ray and NMR experiments. Such structures have favorable intra-strand as well as interstrand interaction, unlike the all-syn–all-anti structures, in which the all-syn strand is energetically unfavorable (Mohanty and Bansal, 1993). It is important to note that, once we assume an alternating syn–anti arrangement of guanines in the stem region, for the hairpin dimers with three guanines in the stem the structures with loops at opposite ends (Fig. 2 *b* and *c*) have to be asymmetric, with the T_nA loop in one hairpin joining a guanine in syn conformation to a guanine in anti conformation while the other joins a guanine in anti conformation to a guanine in syn conformation (Fig. 2 *b* and *c*). However, hairpin dimers with two loops on same side of the G tetraplex can be symmetric (Fig. 2 *a*) as well as asymmetric. Thus the hairpin dimers with G_3 stems are different from the hairpin dimers with G_4 stems, where all three types of dimer can be symmetric because of the even number of guanines in the stem (Mohanty and Bansal, 1994; Smith and Feigon, 1993). In the symmetric hairpin dimers with two loops on the same side the thymines and adenines can form a TATA tetrad (Chernyi et al., 1990), which is nearly isostructural to the G tetrad and can accommodate maximum number of hydrogen bonds between the adenines and thymines that form the tetrad. Such symmetric dimer structures with TATA tetrads can be obtained by dimerization of hairpins with the loop across the large groove (i.e., the groove in which donor atoms N2 and N1 of the guanine in the anti conformation are hydrogen bonded to acceptors N7 and O6 of the guanine in the syn conformation (Fig. 3 *a*)) or from hairpins with the loop across the small groove (i.e., the groove in which donor atoms N2 and N1 of the guanine in the syn conformation are hydrogen bonded to acceptors N7 and O6 of the guanine in the anti conformation (Fig. 3 *a*)). Thus the objective of the model-building study was to build hairpins with appropriate orientation of adenine and thymines in the loop so that, in the symmetric dimer structure obtained by applying a twofold rotation about the helical axis of the stem, both intraloop and interloop hydrogen bonds can be formed.

The various types of TATA tetrad (Chernyi et al., 1990) that are expected from symmetry considerations are the following: Two Watson–Crick (Fig. 3 *b*) (or Hoogsteen (Fig. 3 *c*)) AT pairs are related by a twofold axis so that the A and the T of one Watson–Crick (or Hoogsteen) pair form one hydrogen bond each with the T and the A of the other Watson–Crick (or Hoogsteen) pair. There are several ways in which this type of base pairing can be accommodated in the hairpin dimer.

WC1 or HG1

An A:T Watson–Crick (Fig. 4 *a*) (or a Hoogsteen (Fig. 4 *c*)) pair is in the large groove and either of the following conditions exists: a) The loop is also across the large groove of the tetraplex. Thus in the dimer one loop connects arms 1 and 2 and the other loop connects arms 3 and 4 in Fig. 4 *a* and *c*). These types of structure will be referred to as WC1a or HG1a. b) The loop is across the small groove of the tetraplex. Thus in the dimer one loop connects arms 1 and 4 and the other loop connects arms 2 and 3 in Fig. 4 *a* and *c*. These types of structure will be referred as WC1b or HG1b.

WC2 or HG2

An A:T Watson–Crick (Fig. 4 *b*) (or a Hoogsteen (Fig. 4 *d*)) pair is in the small groove and either of the following conditions exists: a) The loop is also across the small groove, i.e., one loop connects arms 1 and 4 and the symmetry-related loop connects arms 2 and 3 in Fig. 4 *b* and *d*. These types of structure will be referred as WC2a or HG2a. b) The loop is across the large groove, i.e., one loop connects arms 1 and 2 and the symmetry-related loop connects arms 3 and 4 in Fig. 4 *b* and *d*. These types of structure will be referred as WC2b or HG2b.

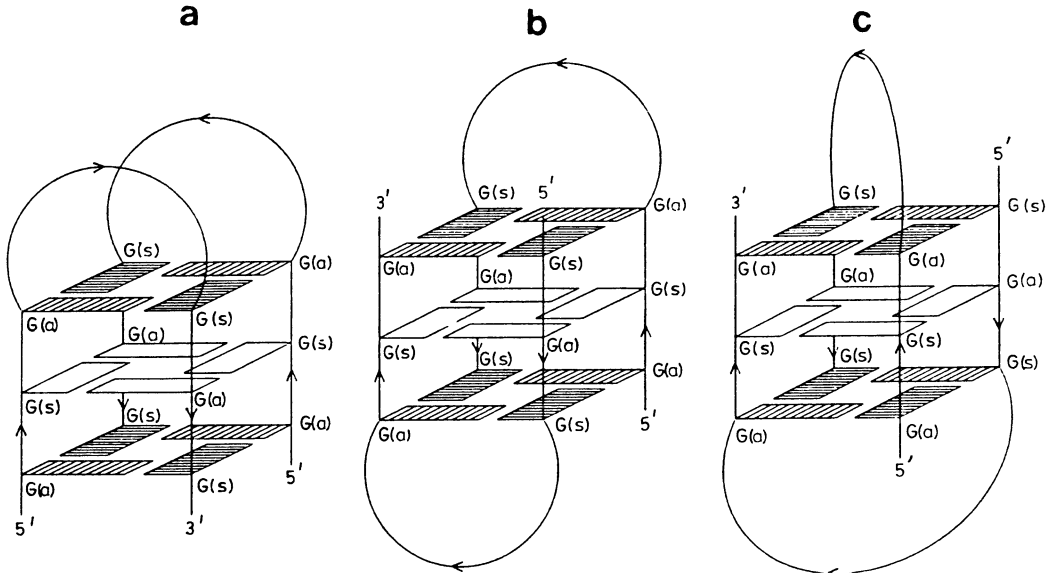

FIGURE 2 Schematic diagrams of three types of hairpin dimer structures for the sequence $d(G_3T_nAG_3)$ with G_3 in the stem and T_nA in the loop and alternating syn–anti arrangement of guanines along the nucleotide strand. (*a*) Loops on same side connecting adjacent strands. (*b*) Loops on opposite sides connecting adjacent strands. (*c*) Loops across diagonal strands and on opposite sides. The guanines in the stem have been represented as rectangular planes similar these in to Fig. 1. As can be seen, while in (*a*) the hairpin dimer is symmetric, in (*b*) and (*c*) the dimers are asymmetric, i.e., one loop joins a guanine in a syn conformation to a guanine in an anti conformation, and the other loop joins a guanine in an anti conformation to a guanine in a syn conformation.

The TATA tetrad was fixed at a z height of 3.4 Å above the G tetrad, with its center on the helix axis of the G stem. The helical twist of the TATA tetrad and the glycosidic torsion angles (χ_T and χ_A) of T and A were varied in the range 0° to 360° to yield the values of χ_T and χ_A that can give rise to a stereochemically satisfactory (Bansal and Sasisekharan, 1986) backbone link for a TTA or T_3A loop. It is interesting to note that when the Watson–Crick or Hoogsteen pair is in the large groove (WC1a, WC1b, HG1a, and HG1b) the glycosidic torsion angle of the thymine always remains in the anti range (180° to 300°), but when the Watson–Crick or Hoogsteen pair is in the small groove (WC2a, WC2b, HG2a, and HG2b) the glycosidic torsion of the thymine is restricted to the range between 0° and 150°. This can be rationalized from a simple analysis of the relative orientation of the backbone chains resulting from various combination of glycosidic orientations in G as well as TATA tetrads (shown in Fig. 3). As can be seen from Fig. 3 and 4 *a*, when a Watson–Crick-type TATA tetrad, with the sugar rings of both A and T in anti conformation, is stacked over the G tetrad with the β faces of A and T (i.e. the faces corresponding to anticlockwise numbering of at-

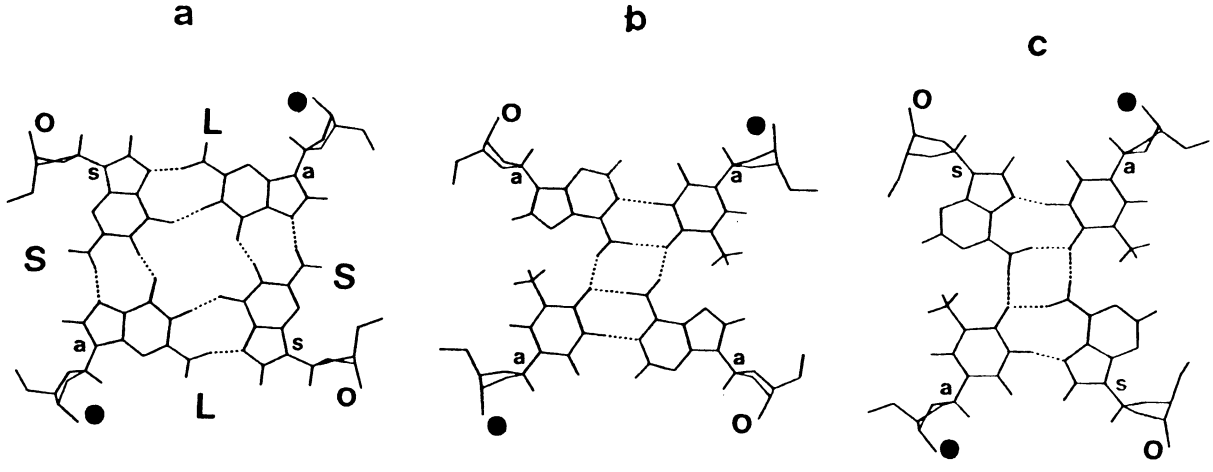

FIGURE 3 Orientation of backbone chains in (*a*) a G tetrad with a syn–anti–syn–anti arrangement of glycosidic torsion angles in the tetrad. The symbols L and S indicate the large and the small grooves, respectively, resulting from the glycosidic orientations in the tetrad. (*b*) Watson–Crick-type TATA tetrad with both thymine and adenine in an anti conformation about the glycosidic bond. (c) Hoogsteen-type TATA tetrad with thymine in an anti and adenine in a syn conformation. (●) The 5′ → 3′ chains pointing up. (○) The 5′ → 3′ chains pointing down.

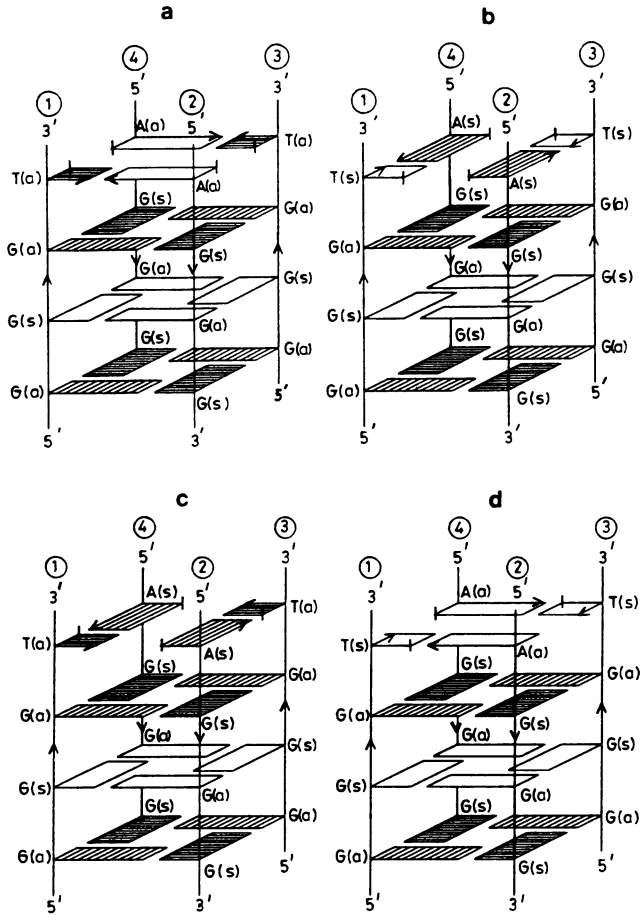

FIGURE 4 Schematic diagrams showing the TATA tetrad stacked over the G tetrad in a G_3 tetraplex. (*a*) Watson–Crick-type TATA tetrad with an A:T Watson–Crick pair in the large groove of the G tetraplex (WC1-type structures). (*b*) Watson–Crick-type TATA tetrad with an A:T Watson–Crick pair in the small groove of the G tetraplex (WC2-type structures). (*c*) Hoogsteen-type TATA tetrad with an A:T Hoogsteen pair in the large groove of the G tetraplex (HG1-type structures). (*d*) Hoogsteen-type TATA tetrad with an A:T Hoogsteen pair in the small groove of the G tetraplex (HG2-type structures). The guanines in the stem have been represented as rectangular planes similar to those in Fig. 1. The adenine is represented as rectangular plane with an arrow indicating the side that contains N1 and C2 and a bar indicating the side containing N7. Similarly the thymine is represented as a square with an arrow indicating the side that contains N3 and O2 and a bar indicating the side containing the methyl group. The α faces of guanine and adenine, and the β-faces of thymine, when visible from the top, have been shaded. Strands 1 and 2 (and the symmetry-related pair 3 and 4) form the large groove of the tetraplex, while strands 1 and 4 (and symmetry-related pair 3 and 2) form the small groove.

oms (Rose et al., 1980)) visible from the top and the A:T Watson–Crick pair in the large groove of the G tetraplex, O3′ atoms of two diagonal guanines will point toward O5′ atoms of the thymines and O5′ atoms of the other two guanines will point toward O3′ atoms of the adenines, thus leading to stereochemically satisfactory backbone linkage. Hence the A:T Watson–Crick pair can be in the large groove (as in WC1a and WC1b) with both T and A being anti, but, as can be seen from Figs. 3 and 4 *b*, for the A:T Watson–Crick pair to be in the small groove all the bases of the TATA tetrad have to be flipped about their glycosidic bonds, thus leading to syn orientation for both T and A. Similarly, when the A:T Hoogsteen pair is in the large groove (HG1a and HG1b) the glycosidic conformations for T and A have to be anti and syn, respectively (Figs. 3 and 4 *c*), but if the A:T Hoogsteen pair is in the small groove (HG2a and HG2b) T is likely to be in the syn conformation and A in the anti conformation (Figs. 3 and 4 *d*). In the case of $d(G_3TTAG_3)$ no hairpin structure of the HG1b or HG2b type is possible, as in these models the loop has to be built over an A:T pair with the C1′–C1′ distance being 12.63 Å (see Fig. 3 *c*); it was impossible to bridge this distance with a single thymine i.e., the middle thymine of the TTA loop, but such structures are possible for $d(G_3T_3AG_3)$ hairpins, which have two thymines connecting the A:T pair.

As indicated from grid search, because structures with A:T base pairs in the small groove have glycosidic torsion for thymine in syn or low anti conformation, which is usually not observed, only two representative structures of WC2 and HG2 type were considered, whereas for other types all possible structures were considered. In addition to Watson–Crick and Hoogsteen ATAT tetrads, the other possible symmetric ATAT tetrads are the ones with reverse Watson–Crick (RWC) or reverse Hoogsteen pairing between the two loops, and in such tetrads adenines can also form a diagonal pair with two hydrogen bonds (Chernyi et al., 1990). Thus as a representative case, a structure of the RWC type with the loop going from anti to syn, was also built. Therefore for $d(G_3TTAG_3)$ the following hairpins were considered for detailed energy minimization study: WC1a (both possibilities, i.e., loop going from syn to anti and from anti to syn), WC1b (both the possibilities, i.e., loop going from syn to anti and from anti to syn), WC2a (loop going from anti to syn), HG1a (both possibilities, i.e., loop going from syn to anti and from anti to syn), HG2a (loop going from anti to syn), and RWC (loop going from anti to syn).

In all the structures for $d(G_3TTAG_3)$ the middle T of the TTA loop remained stacked over the ATAT tetrad while in the anti conformation and could also form interloop hydrogen bonds, with two loops on same side, as in the case of the hairpin dimer structures for $d(G_4T_3G_4)$ (Mohanty and Bansal, 1994). However, hairpin dimers with loops on opposite ends of the tetraplex can be obtained only by an asymmetric combination of syn to anti and anti to syn hairpins because of the occurrence of three guanines in the stem. Two such hairpin dimer structures were considered for energy minimization, one with intraloop Watson–Crick pairing and the other with intraloop Hoogsteen pairing. No diagonal hairpin structures were considered for $d(G_3TTAG_3)$, as they are expected to be similar to the T_3 diagonal structure (Mohanty and Bansal, 1994) in the absence of any obvious base pairing scheme in the loop region. However, an asymmetric dimer (WC1a/WC2a) consisting of large and small groove hairpins with loops on same side and Watson–Crick pairing in the loop was considered for energy minimization. This structure has diagonal strands in antiparallel orientation and

is a representative case of the hairpin dimers involved in Indian-key type of chain folding (Frank-Kamenetskii, 1992). The As and Ts in the tetrad in this structure were arranged as TTAA, unlike the TATA tetrads in other structures.

The hairpin dimer structures for $d(G_3T_3AG_3)$ with loops on same side are essentially similar to the structures for $d(G_3TTAG_3)$ except for the fact that T and A in the loop are bridged by two thymines instead of one thymine as in the case of $d(G_3TTAG_3)$. These two thymines can stack over the TATA tetrad and can form interloop T:T pairing, as in the case of $d(G_4T_2G_4)$ (Mohanty and Bansal, 1994; Balagurumoorthy et al., 1992). Only two models were built for T_3A loop structures, as these structures had certain features different from TTA loop structures. One was of the type WC1a and in this case both T and A had an anti conformation, unlike the TTA loop structure in which A was in the syn conformation. The other was of the type HG1b, which was not possible for $d(G_3TTAG_3)$.

In addition to the hairpin dimer structures, four strands of $d(G_3TTAG_3)$ can also adopt a parallel tetraplex structure (Arnott et al.; 1974, Wang and Patel, 1992; Laughlan et al., 1994) with six G tetrads, two T tetrads, and one A tetrad or an antiparallel tetraplex structure with six G tetrads, two TATA tetrads, and one T tetrad. Therefore a parallel tetraplex structure was built for $d(G_4TTAG_4)$ starting from Arnott's fiber model (Arnott et al., 1974) and keeping the thymines and adenines in the stem in exactly the same conformation as for the guanines. The antiparallel tetraplex structure was built starting from an antiparallel G-tetraplex structure (Mohanty and Bansal, 1993) of nine nucleotide lengths and replacing the middle three guanines by TTA with appropriate glycosidic torsion angles so that thymines and adenines are in the anti conformation.

Intramolecular folded-back structures were also built for $d(TTAG_3)_4$, with hairpins with TTA or TT loops and the 5′ TTA or TT kept in a single-stranded helical conformation.

ENERGY MINIMIZATION AND MOLECULAR DYNAMICS

The various parallel tetraplex, hairpin dimer, and intramolecular folded-back structures obtained from model building were energy minimized to an rms gradient of 0.09 kcal/mole Å by use of an AMBER (1987) all-atom force field (Weiner et al., 1986). Solvent effects were mimicked implicitly by use of a distance-dependent dielectric function $\epsilon_{ij} = R_{ij}$ (Weiner et al., 1984, 1986; Bansal and Pattabiraman, 1989), and no cutoff was used for nonbonded pairs. In addition to the calculations carried out with normal AMBER electrostatic charges, the structures were also minimized with a reduced charge (by an amount 0.7 of electronic charge, so that total charge on the molecule (became -0.3 per phosphate group) on anionic phosphate oxygens, to mimic the effects of counter ions. Such an approach was used earlier in the simulation of DNA duplex (Tidor et al., 1983; Ferentz et al., 1993) and triplex structures (Cheng and Pettitt, 1992); and, as observed in our earlier molecular mechanics study of G tetraplexes (Mohanty and Bansal, 1993, 1994), it leads to groove widths in better agreement with the crystal structure. In the minimizations no constraints were used on any of the hydrogen bonds or any other parameters.

Molecular dynamics (MD) calculations at constant temperature were carried out with AMBER 4.0 for two representative intramolecularly folded-back structures with sequence $d(G_3(TTAG_3)_3)$. The MD studies were carried out with a reduced charge on phosphate oxygens and a distant dependent dielectric function $\epsilon_{ij} = R_{ij}$ similarly to the minimization studies, but two K^+ ions were placed at the two cation binding sites between the three G tetrads. A cutoff of 100 Å was used to calculate the nonbonded interactions, and nonbonded pairs were updated every 20 steps. The structures were slowly heated in six steps from 0 to 300 K in the first 12 ps of the MD at a rate of 50 K for every 2 ps of MD run. Then the system was coupled to a heat bath at 300 K and MD was continued for a further period of 338 ps, resulting in a total simulation time of 350 ps. SHAKE, a logarithm, was used to constrain all bonds involving hydrogens, and a time step of 1 fs was used in the MD integrations. The structures were saved after every 100 steps of MD run, i.e., every 0.1 ps.

RESULTS AND DISCUSSION

Does adenine occur in the loop or in the stem?

A comparison of energy values in energy-minimized folded-back structures reveals that when adenine remains in the stem the structure has better van der Waal's interaction, owing to more favorable stacking of two AGAG tetrads on either side of the two GGGG tetrads. As can be seen from Fig. 5 the AGAG tetrad has six hydrogen bonds occurring between the O6 and N1 atoms of guanine and the N6 and N7 of adenine. In addition N7 of guanine and N1 of adenine

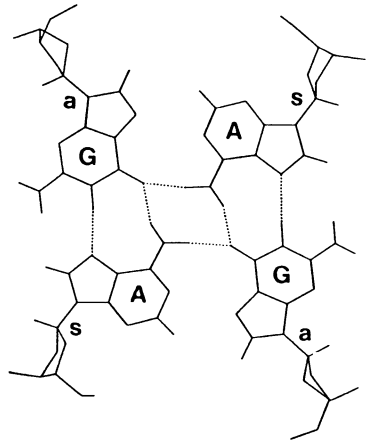

FIGURE 5 Hydrogen bonding pattern in the AGAG tetrad in the energy-minimized structure of $d(TTAG_3)_4$ with adenine in the stems and T_2 loops connecting adjacent strands.

are within 3 Å in the G:A pairs, which have only one hydrogen bond; hence at low pH, if N1 of adenine is protonated, the AGAG tetrad could be stabilized by eight hydrogen bonds. The better van der Waals stacking in these structures is compensated for by less favorable electrostatic interaction compared with those with TTA loops and three G tetrads. In fact TTA loop structures have slightly better total energy than structures in which adenine remains in the stem. The structure with adenine in the stem and AGAG tetrads also lacks the quartet of oxygens required for formation of the cation binding site (Sundquist and Klug, 1989) in the tetraplex stem. Thus in presence of cations these structures are likely to be less stable than TTA loop structures with three G tetrads. The NMR study of $d(AG_3(TTAG_3)_3)$ indicates that the sequence folds into an Indian-key motif with a G_3 stem and a TTA loop. Chemical probing studies (Balagurumoorthy and Brahmachari, 1994) on $d(TTAG_3)_4$ and corresponding control sequence $d(G_3TTAG_3)$ also indicate that $d(TTAG_3)_4$ adopts a structure with adenine in the loop region. Therefore only those structures with G_3 stems and TTA loops are discussed in detail.

Energy component analysis for various hairpin dimer structures

Energy-minimization studies indicate that the various hairpin dimer structures described in the section on model building are energetically almost equally favorable (Table 1). Among the models with Watson–Crick and Hoogsteen base pairs, WC1a (a–s) with the loop going from anti to syn is the most favorable (−773 kcal/mol), HG1a (a–s) is the least favorable (−734 kcal/mol), and all other models have intermediate energy values. The small difference in energy arises mostly from bond, angle, and dihedral distortion energy components, whereas electrostatic and nonbonded energy components are very similar for all the Watson–Crick and Hoogsteen types of structure. However, the RWC model, i.e., with interloop reverse Watson–Crick pairing and diagonal A:A pairing, has almost 10–15 kcal/mol better electrostatic energy than other models, mostly because of loop stem interaction, but this structure has a larger distortion energy component; thus there is no overall gain in total energy. The WC1a/WC2a structure, which is an asymmetric dimer of two different Watson–Crick hairpins, also has energy values similar to those of the symmetric dimers, even though the arrangement of bases in the top tetrad is different. The models with TTA loops on opposite ends of the G_3 tetraplex have energy values similar to those with loops on same side. In fact the Watson–Crick type of structure WC–opp, which consists of two WC1a types of hairpin, one syn to anti and the other anti to syn, has a total energy value that is close to the average of the energy of WC1a (a–s) and WC1a (s–a), as seen from Table 1. This is also true for Hoogsteen-type structure HG–opp, with loops on opposite ends of the G_3 tetraplex. If we analyze the

TABLE 1 Energy contributions (kcal/mol) from G_3 and TTA stretches of various hairpin dimer, parallel tetramer, and antiparallel tetramer structures of $d(G_3TTAG_3)$, minimized with reduced charge on phosphate oxygens.

Structure	Intra G_3 Stretch (a)	Intra TTA Stretch (b)	Inter G_3 TTA (c)	Total (a + b + c)
WC1a(a–s)	−542	−160	−71	−773
d	176	87	8	271
v	−146	−31	−70	−247
e	−572	−216	−9	−797
WC1a(s–a)	−541	−143	−63	−747
d	177	103	13	293
v	−139	−26	−72	−237
e	−579	−220	−4	−803
WC1b(a–s)	−541	−156	−61	−758
d	184	90	9	283
v	−140	−27	−70	−237
e	−585	−219	0	−804
WC1b(s–a)	−542	−147	−74	−763
d	181	97	7	285
v	−137	−29	−76	−242
e	−586	−215	−5	−806
WC2a(a–s)	−547	−131	−68	−746
d	176	105	8	289
v	−145	−26	−67	−238
e	−578	−210	−9	−797
HG1a(a–s)	−545	−119	−70	−734
d	174	114	11	299
v	−145	−21	−70	−236
e	−574	−212	−11	−797
HG1a(s–a)	−534	−139	−78	−751
d	181	104	8	293
v	−139	−27	−76	−242
e	−576	−216	−10	−802
HG2a(a–s)	−549	−155	−64	−768
d	173	95	11	279
v	−143	−31	−69	−243
e	−579	−219	−6	−804
RWC (a–s)	−537	−119	−62	−718
d	184	119	14	317
v	−144	−17	−61	−222
e	−577	−221	−15	−813
WC1a/WC2a (a–s)	−544	−129	−75	−748
d	177	98	11	286
v	−141	−24	−70	−235
e	−580	−203	−16	−799
WC–opp	−548	−121	−94	−763
d	172	99	10	281
v	−141	−20	−82	−243
e	−579	−200	−22	−801
HG–opp	−550	−113	−84	−747
d	173	106	10	289
v	−141	−21	−83	−245
e	−582	−198	−11	−791
Tet–par	−1159	−355	−142	−1656
d	314	162	10	486
v	−288	−91	−160	−539
e	−1185	−426	8	−1603
Tet–apr	−1090	−337	−137	−1564
d	351	186	12	549
v	−278	−93	−153	−524
e	−1163	−430	4	−1589

The symbols d, v, and e correspond to the distortion (sum of bond length, angle, and dihedral components), van der Waals (sum of 6–12 and 10–12 terms), and electrostatic energy, respectively. The first row for each structure gives the total energy.

energy components in detail (Table 1), we can see that the G_3 stem has very similar energy in all the models (near −540 kcal/mol), whereas the energy of the loop varies within a range of 40 kcal/mol because of different distortion energy values for different loops. The loop–stem interaction is more or less similar for various models and varies within a range of 10 kcal/mol. Compared with structures with loops on the same side, the structures with loops on opposite ends have similar stem energy, less favorable intraloop energy, but better loop–stem interaction energy, a trend similar to that observed for hairpin dimer structures with T_2, T_3, and T_4 loops (Mohanty and Bansal, 1994). The structures when minimized by normal charge on phosphate oxygens show energy component trends very similar to those for reduced charge, and therefore corresponding energy tables have not been shown.

Parallel tetraplex versus hairpin dimer structures

The parallel tetraplex structure (Tet–par) for $d(G_3TTAG_3)$, with all bases in the anti conformation is more favorable than the antiparallel tetraplex structure (Tet–apr) by almost 90 kcal/mol (Table 1) when minimized with reduced charge. In fact an almost 60-kcal/mol difference arises from unfavorable distortion energy in the antiparallel structure, whereas the differences arising from van der Waals and electrostatic components are relatively smaller. This feature has also been observed for an all-G tetraplex (Mohanty and Bansal, 1993). The TTA stretch in the anti structure has unfavorable distortion energy compared with the TTA stretch in the parallel structure, probably because the starting antiparallel structure had an alternating backbone conformation and only χ was changed to the anti conformation. If we compare the conformational energy (reduced charge) of the tetraplexes with that of the hairpin dimer WC1a(a–s) (after scaling it by a factor of 2 to account for the difference in the number of strands), then parallel tetraplex is more favorable, whereas the antiparallel tetraplex is almost equal in energy. However, for structures minimized with a normal charge on phosphate oxygens the hairpin dimers are more favorable because of large phosphate–phosphate repulsion in tetramolecular structures. Thus tetramolecular structures are likely to be observed only if phosphate charges are screened by counter ions at high salt concentrations. Our earlier molecular mechanics calculations of $d(G_4T_nG_4)$ had indicated a similar result. In fact it was shown recently from Raman spectroscopic studies that, although $d(T_4G_4)_4$ adopts a folded-back structure at low salt concentrations, it adopts a parallel tetraplex structure at high salt (Miura and Thomas, 1994).

It is important to note that, in addition being stabilized by an ambient counter-ion atmosphere, the effect of which was modeled by reduction of the charge on phosphate groups, the G tetraplexes are also known to be stabilized by binding of K^+ ions to the central cavity between two adjacent G tetrads (Kang et al., 1992; Gupta et al., 1993). However, our earlier molecular mechanics study of $d(G_4T_4G_4)$ had indicated that the structural features of G tetraplexes minimized with K^+ ions in the cavities were almost identical to those minimized without K^+ ions. Hence in the present study the various types of parallel and folded-back structure have been energy minimized, without K^+ ions being put between adjacent G tetrads, and the minimization was carried out without the imposition of any constraints. In spite of that, the K^+ binding site is preserved in all the models. Our earlier energy-minimization study of $d(G_4T_4G_4)$, which included K^+ ions in the cavity, also indicated that the binding of K^+ ions in the cavities of the G-tetraplex stem contributes a significant stabilization energy to the structure, in agreement with the experimental observations. However, the energy contribution from K^+ binding to the G_4 stem is nearly identical for all types of tetraplex, i.e., parallel, Greek-key-type folded-back, and Indian-key-type folded-back structures (Mohanty and Bansal, 1994). This can be attributed to the fact that in all three types of G tetraplexes the K^+—O distance remains the same, regardless of the twist between the G tetrads or the types of stacking arrangement i.e., normal or inverted stacking. Hence binding of K^+ to the central cavities is unlikely to stabilize preferentially any particular type of G tetraplex compared with the others, but can stabilize G tetraplexes of all types, as reported already for antiparallel and parallel structures (Kang et al., 1992; Gupta et al., 1993).

Analysis of structural features in the energy-minimized structures

Stacking and hydrogen bonding in hairpin dimer structures

Stacking and hydrogen bonding in various energy-minimized hairpin dimer structures are shown in Fig. 6 *a–h*. It is interesting to note that the various hydrogen bonds in the loop region are retained in the energy-minimized structures, even though no constraints were applied to the hydrogen bonds during minimization. Table 2 gives a list of various hydrogen bonds in the loop region, the H–Y distances, and X—H–Y angle values. In models WC1a, WC2a, HG1a, and HG2a, because the TTA loop can have perfect AT base pairs with Watson–Crick or Hoogsteen pairing, these structures are effectively hairpins with a single thymine in the loop. It is important to note that a similar single base loop is not possible over a GG pair (Balagurumoorthy et al., 1992) if the guanines are constrained to have conformations identical to those of guanines in the tetraplex stem. In structures with Watson–Crick and Hoogsteen type hydrogen bonded TATA tetrads (Fig. 6 *a–e*) the methyl hydrogens of thymines remain close (≈3.0 Å) to N7 or N1 of adenine, thus giving rise to the additional possibility of C—H–N hydrogen bond formation, similar to C—H–O hydrogen bonds (Taylor and Kennard, 1982; Bruskov et al., 1989; Zhurkin et al., 1994). In all the structures shown in Fig. 6 *a–f* the N7 of adenine is either involved in hydrogen bonding or is close to a methyl group hydrogen, thus making it

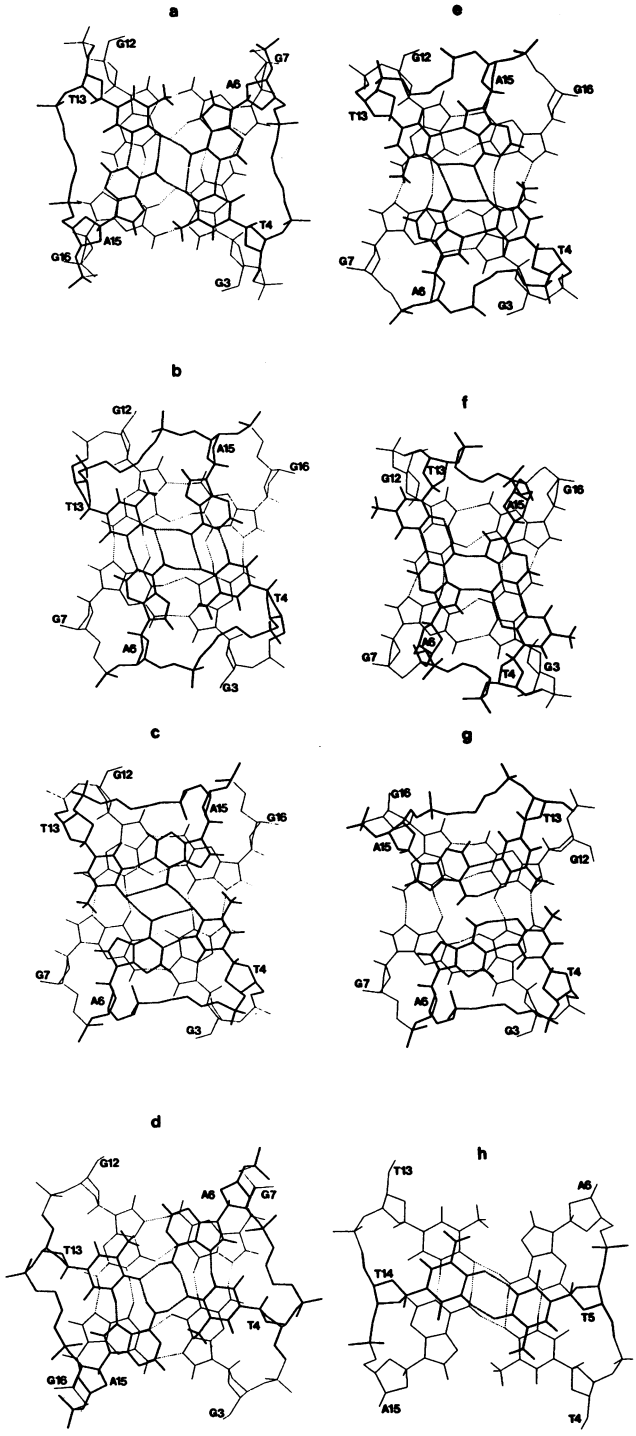

FIGURE 6 Hydrogen bonding pattern in a TATA or TTAA tetrad and stacking of this tetrad (thick lines) over the guanine tetrads in various hairpin dimer models of $d(G_3TTAG_3)$ with two loops on same side. The residues are numbered G1-G2-G3-T4-T5-A6-G7-G8-G9 in one hairpin and G10-G11-G12-T13-T14-A15-G16-G17-G18 in the other hairpin of the dimer. (*a*) WC1a(a–s), (*b*) WC1b(a–s), (*c*) WC2a(a–s), (*d*) HG1a(a–s), (*e*) HG2a(a–s), (*f*) RWC(a–s) and (*g*) an asymmetric hairpin dimer of WC1a(a–s) and WC2a(a–s), with the tetraplex stem in an Indian-key arrangement (i.e., diagonal strands are antiparallel). In *a–g*, only backbone atoms have been shown for the middle thymines T(5) and T(14). (*h*) Representative example of the stacking of the TT pair (thick lines) formed by T(5) and T(14) over the TATA tetrad (thin lines) in the structure WC1a(a–s).

less accessible to chemical probes; whereas in the TTAA tetrad shown in Fig. 6 *g* the N7 of both the adenines are more accessible. Chemical probing experiments can therefore distinguish among various types of models discussed here, and the results of such studies (Balagurumoorthy and Brahmachari, 1994) in fact suggest the possibility of A:T pairing in the loop region.

In TTA loop structures with loops on same side the TATA as well as the G tetrads have very little buckle and propeller-type distortion, and the top TT pair is also stacked over the TATA tetrad (shown in Fig. 6 *h*).

The various models with TTA loops, although they are similar in terms of energy (Table 1), have distinct structural features that can give rise to unique cross-peaks in nuclear Overhauser effect spectroscopy (NOESY) experiments. Table 3 gives, for five representative models, a list of hydrogen pairs that are separated by less than 4 Å in some models, although the separation is much larger in other models. As can be seen from Table 3 and the stacking diagrams (Fig. 6), H8 of A6 will come close (3.3 Å) to H8 of G3 only in model HG2a(a–s) (Fig. 6 *e*); in models (Fig. 6 *a–d*) that have A:T pairs in the large grooves or Watson–Crick A:T pairs in the small grooves this distance is larger than 8 Å. The model WC2a(a–s) (Fig. 6 *c*) has H8 of G3 at a distance of 3.3 Å from H2 of A6; the same proton pair is separated by more than 7 Å in all other models (Fig. 6 *a, b, d, e*), which have A:T pairs in the large grooves or Hoogsteen-type A:T pairs in the small grooves. Similarly, NOE will be observed between HN2A/HN2B protons of G3 and H2 of A6 only in model WC1a(a–s) (Fig. 6 *a*), which has a WC A:T pair in the large groove. Table 3 also indicates that NOEs should generally be observed between H3 of T4 and either H2 or H8 of A6, corresponding to Watson–Crick or Hoogsteen A:T pairs, respectively. Thus these NOEs could be used as diagnostic features for these structures.

Backbone and glycosidic torsion angles

In all the hairpin dimer structures for $d(G_3TTAG_3)$ the guanines that are one nucleotide away from the loop all have torsion angles identical to those in the antiparallel G tetraplex without the loops; i.e., α, β, γ, δ, ϵ, and ζ for guanine in the syn conformation are in the range g^-, g^-, g^-, $t(^2E)$, t, and g^-, and for guanines in the anti conformation they are in the g^-, t, g^+, $t(_1E)$, t, and g^- regions. Even in the guanines, immediately flanking the loop at the 5′ and the 3′ ends, only a few torsion angles have changed from their values in the tetraplex stem. The backbone torsion angles for the nucleotides in the loop take up a wide range of values in various models, but the glycosidic torsion is the one that is directly correlated to the chain folding and the type of base pairing in the loop, as discussed in the Model Building section. The glycosidic torsion angles (Table 4) for thymines and adenines in various models remain close to their values in the starting structures. In models WC1a(a–s) and WC1a(s–a), which have loops across the large grooves and

TABLE 2 List of hydrogen bonds involving the residues in the loop region of $d(G_3TTAG_3)$ in different energy-minimized (reduced-charge) hairpin dimer structures

$d(G_3TTAG_3)$		WC1a(a–s)		WC1a(s–a)		WC2a(a–s)	
Donor (X)	Acceptor (Y)	Dist (Å) (H–Y)	Angle (°) (X–H—Y)	Dist (Å) (H–Y)	Angle (°) (X–H—Y)	Dist (Å) (H–Y)	Angle (°) (X–H—Y)
N6-A(6)	O4-T(4)	2.0	158.2	2.1	162.3	2.6	147.7
N3-T(4)	N1-A(6)	1.8	166.8	1.8	161.8	1.8	163.3
N6-A(15)	O4-T(4)	1.9	164.7	1.9	154.9	1.8	169.3
N3-T(14)	O4-T(5)	1.8	156.6	1.8	163.0	1.8	157.2
N3-T(5)	O4-T(14)	1.8	156.6	1.8	163.0	1.8	157.4
N6-A(6)	O4-T(13)	1.9	164.7	1.9	154.9	1.8	168.4
N6-A(15)	O4-T(13)	2.0	158.2	2.1	162.3	2.6	142.0
N3-T(13)	N1-A(15)	1.8	166.8	1.8	161.8	1.8	163.0

		WC1b(a–s)		WC1b(s–a)	
		Dist (Å) (H–Y)	Angle (°) (X–H—Y)	Dist (Å) (H–Y)	Angle (°) (X–H—Y)
N6-A(6)	O4-T(4)	2.2	131.5	2.0	149.4
N6-A(15)	O4-T(4)	1.8	170.2	1.8	170.2
N3-T(4)	N1-A(15)	1.8	165.1	1.8	163.8
N3-T(14)	O4-T(5)	1.8	149.6	1.8	147.9
N3-T(5)	O4-T(14)	1.8	149.6	1.8	147.9
N6-A(6)	O4-T(13)	1.8	170.2	1.8	170.1
N3-T(13)	N1-A(6)	1.8	165.1	1.8	163.8
N6-A(15)	O4-T(13)	2.2	131.5	2.0	149.4

		HG1a(a–s)		HG1a(s–a)		HG2a(a–s)	
		Dist (Å) (H–Y)	Angle (°) (X–H—Y)	Dist (Å) (H–Y)	Angle (°) (X–H—Y)	Dist (Å) (H–Y)	Angle (°) (X–H—Y)
N6-A(6)	O4-T(4)	1.9	158.5	1.9	155.7	2.0	147.9
N3-T(4)	N7-A(6)	1.8	160.2	1.8	167.3	1.8	172.1
N6-A(15)	O4-T(4)	1.9	147.5	1.9	164.4	1.9	177.5
N3-T(14)	O4-T(5)	1.8	156.7	1.8	157.6	1.8	151.5
N3-T(5)	O4-T(14)	1.8	156.7	1.8	157.6	1.8	151.5
N6-A(6)	O4-T(13)	1.9	147.5	1.9	164.4	1.9	177.5
N6-A(15)	O4-T(13)	1.9	158.4	1.9	155.7	2.0	148.0
N3-T(13)	N7-A(15)	1.8	160.2	1.8	167.3	1.8	172.1

		RWC	
		Dist (Å) (H–Y)	Angle (°) (X–H—Y)
N3-T(4)	N1-A(15)	1.8	163.6
N6-A(15)	O2-T(4)	1.9	162.7
N3-T(14)	O4-T(5)	1.8	172.5
N3-T(5)	O4-T(14)	1.8	172.5
N6-A(15)	N7-A(6)	1.7	166.7
N6-A(6)	N7-A(15)	1.7	166.7
N6-A(6)	O2-T(13)	1.9	162.7
N3-T(13)	N1-A(6)	1.8	163.6

		WC1a/WC2a	
		Dist (Å)	Angle (°)
N6-A(6)	O4-T(4)	2.5	165.5
N3-T(4)	N1-A(6)	1.8	165.1
N6-A(15)	O4-T(13)	3.0	115.5
N3-T(13)	N1-A(15)	1.8	165.7
N3-T(14)	O4-T(5)	1.8	156.3
N6-A(15)	O6-G(7)	1.9	152.6
N6-A(15)	O6-G(12)	1.9	161.0

In a hairpin dimer containing two strands of N nucleotides each, nucleotides in one chain are numbered 1 to N in the 5′ to 3′ direction, and the nucleotides in the other chain are numbered $N + 1$ to $2N$.

TABLE 3 Comparison of interproton distances in five representative models of d(G_3TTAG_3)

Proton 1		Proton 2		WC1a (a–s)	WC2a (a–s)	HG1a (a–s)	HG2a (a–s)	RWC (a–s)
H1′	G(3)	H8	G(3)	3.9	3.9	3.9	3.9	3.9
H1′	G(3)*	H6	T(4)*	5.5	6.1	3.7	5.8	5.7
H1	G(3)	H3	T(4)	3.3	3.6	3.7	4.2	3.7
H8	G(3)*	H8	A(6)*	12.9	8.1	11.4	3.3	7.0
H8	G(3)*	H2	A(6)*	8.0	3.3	13.6	8.8	10.7
H1	G(3)	HN6A	A(6)	3.8	4.9	3.9	3.4	5.2
H1	G(3)	HN6B	A(6)	2.9	3.3	4.8	4.6	6.1
HN2A	G(3)	H3	T(4)	3.6	4.8	3.5	5.4	4.1
HN2A	G(3)*	H2	A(6)*	3.0	6.8	8.4	10.5	10.3
HN2B	G(3)	H6	T(4)	7.3	3.8	3.9	4.1	7.3
HN2B	G(3)	H3	T(4)	3.9	5.1	4.6	5.6	4.1
HN2B	G(3)*	H2	A(6)*	2.9	6.6	10.1	11.4	11.6
HN2B	G(3)*	H1′	T(4)*	5.4	5.8	3.5	6.4	8.4
H1′	T(4)	H6	T(4)	3.7	2.7	3.7	2.7	3.5
H3	T(4)‡	H8	A(6)‡	7.1	7.1	3.1	2.8	5.4
H3	T(4)	HN6A	A(6)	3.9	3.9	2.4	2.9	6.3
H3	T(4)	HN6B	A(6)	2.4	2.5	4.0	4.5	7.8
H3	T(4)‡	H2	A(6)‡	2.8	2.8	6.4	6.6	9.9
H1′	A(6)	H8	A(6)	3.2	2.5	3.4	3.9	3.9
H1′	A(6)‡	H1′	G(7)‡	3.7	5.8	4.2	6.4	5.9
H1′	A(6)*	H8	G(7)*	3.9	6.8	5.4	7.4	5.3
H8	A(6)	HN2A	G(7)	4.3	5.8	6.9	3.2	5.4
H8	A(6)	HN2B	G(7)	3.9	5.2	6.5	3.8	6.1
HN6A	A(6)	H1	G(7)	3.4	3.4	3.6	3.2	5.8
HN6B	A(6)	H1	G(7)	4.3	3.2	3.3	3.9	6.5
H2	A(6)‡	HN2A	G(7)‡	7.9	3.3	3.0	6.2	7.9
H2	A(6)‡	HN2B	G(7)‡	8.2	4.3	3.4	6.1	8.1
H1′	G(7)	H8	G(7)	2.5	2.5	2.6	2.6	2.5

The pairs that have a distance of less than 4 Å in any one of the models and can give rise to NOE signals in NMR spectra have been listed. The distances that are unique for any given model, i.e., proton pairs that have a distance of less than 4 Å in only one of these models while being greater than 5 Å in other models, are indicated by *. Proton pairs that would give NOE signals in any two models are indicated by ‡.

TABLE 4 Glycosidic torsion angle χ (in degrees) for the nucleotides in the loop region of d(G_3TTAG_3) hairpin dimer structures listed in Table 1

Structure	T4/T13	T5/T14	A6/A15
WC1a(a–s)	209	205	−33
WC1a(s–a)	236	236	5
WC1b(a–s)	231	217	252
WC1b(s–a)	231	185	245
WC2a(a–s)	132	175	77
HG1a(a–s)	247	245	172
HG1a(s–a)	231	234	165
HG2a(a–s)	132	211	233
RWC (a–s)	190	194	260

The residue numbering is same as in Fig. 6 and the torsion angles are averages of T4 and T13, T5 and T14, and A6 and A15.

intraloop A:T Watson–Crick pairing, the thymines (T4/T13 and T5/T14) are in anti conformation and the adenine (A6/A15) is in syn conformation (Table 4), because, as discussed above, in this case the loop cannot be closed and thus the adenine remains in an anti conformation. However, in the WC1a(a–s) structure for d($G_3T_3AG_3$) the adenine can also have an anti conformation. For d(G_3TTAG_3), in models with Hoogsteen-type A:T pairing in large-groove (HG1a) and interloop Watson–Crick pairing (WC1b(a–s) and WC1b(s–a)), the adenine as well as the thymines are in an anti conformation (Table 4), but in the HG1b structure for d($G_3T_3AG_3$) the adenine will have a syn conformation because of the typical orientation of thymines in the top thymine tetrad. However, in models WC2a(a–s) and HG2a(a–s), which have loops across the smaller grooves and intraloop A:T Watson–Crick or Hoogsteen pairing, T4/T13 has glycosidic torsions in the low anti region (≈130°. Table 4), which is not usually observed. However, as discussed above, one can build such models only by keeping this particular thymine either in a syn (which is stereochemically unfavorable) or in a low anti glycosidic orientation (≈150°). Similarly, the adenine in Watson–Crick type models with intraloop A:T pairing in the small groove (WC2a) is also constrained to have a syn glycosidic orientation (Table 4). Therefore, as can be seen from Table 4, for d(G_3TTAG_3) if all nucleotides in the loop are to have anti conformation and Watson–Crick A:T pairing, then the hairpin dimer structure will have loop across the small groove and interloop Watson–Crick pairing (WC1b type), which would necessitate the loops being on the same side. On the other hand, if all nucleotides in the loop are to have anti conformation and Hoogsteen A:T pairing, then the hairpin dimer structure can have loops across the large groove and intraloop Hoogsteen pairing (HG1a type), and the hairpin dimer can have loops on same side or on opposite sides of the G_3 tetraplex stem.

Models for intramolecularly folded human telomeric sequence with A:T pairs in the loop region

The four repeat human telomeric DNA d$(TTAG_3)_4$ can adopt structures with G_3 tetraplex stems and TTA loops by Greek-key (Fig. 1 *a*) or Indian-key (Fig. 1 *b*) types of intramolecular folding, but the structure with the maximum number of A:T pairs will be the Greek-key structure. The Greek-key structure can be obtained by two topologically different ways of nucleotide chain folding. In one case two loops on the same side will connect guanine strands across the small groove and the single loop on the other side will be across the large groove, while in the other case two loops on the same side will connect guanine strands across the large groove and the single loop on the other side will connect strands over the small groove. However, from the correlation mentioned above between the glycosidic torsion angle of loop nucleotides and the occurrence of base pairs in the loop, we can conclude that only in the first case can the chain fold

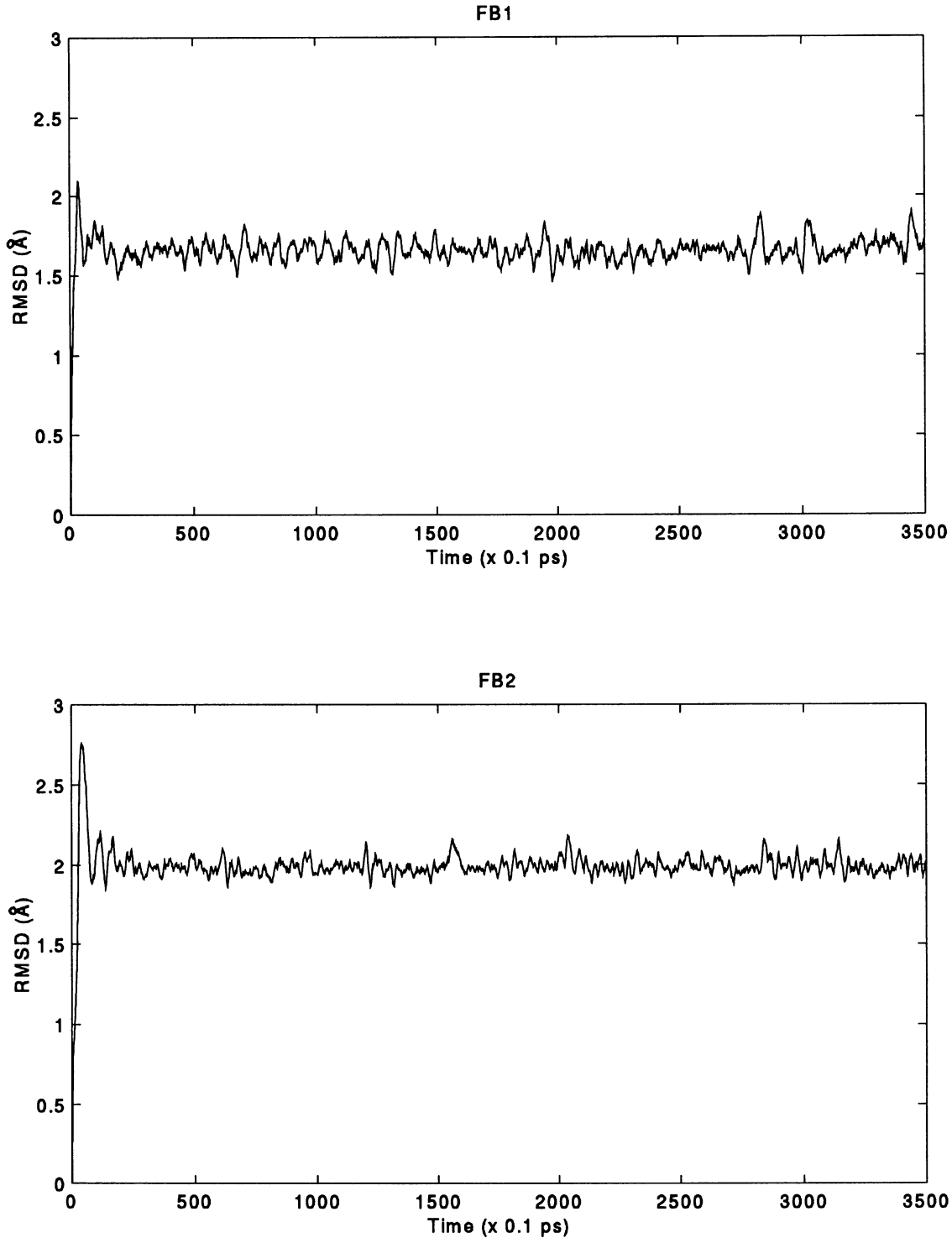

FIGURE 7 Plots showing the rms deviations, during the 350-ps MD run, of structures FB1 and FB2 from their respective starting model structures.

with all nucleotides in the loop in an anti conformation and have interloop Watson–Crick pairing between two small groove loops on one end of the tetraplex and intraloop Hoogsteen pairing in the single large groove loop on the other end.

Stability of the folded-back structures with A:T pairs in the loop region, as revealed by MD simulations

Because the MD technique is known to be more efficient than molecular mechanics in scanning conformational

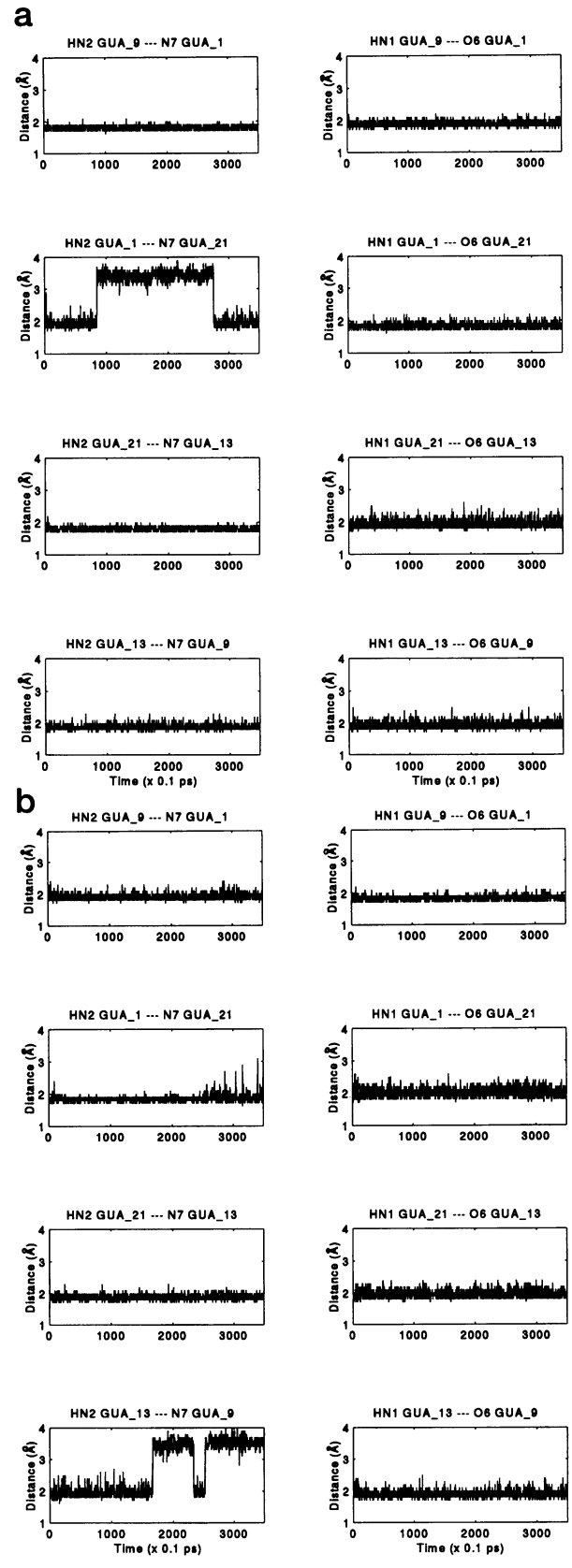

FIGURE 8 Plots of hydrogen-to-acceptor atom distance versus time for all the eight hydrogen bonds in the G tetrad, consisting of guanines G1, G9, G13, and G21 in the intramolecular folded-back structures: (*a*) FB1, (*b*) FB2.

space, we carried out MD simulations for two representative folded-back structures for $d(G_3(TTAG_3)_3)$ to find out whether these structures really represent energetically stable minima in the conformational space. Both of the structures considered had Greek-key-type folding, because these structures have the maximum number of A:T pairs, but the folding topology and base parings in the loop region were different in the two structures. In one, the two loops on the same side were over the small groove joining a guanine in an anti conformation to a guanine in a syn conformation and with interloop A:T Watson–Crick pairs, and the single loop on the other end joining a guanine in a syn conformation to a guanine in an anti had intraloop Hoogsteen A:T pairs. Thus the folded-back structure had two hairpin loops of the WC1b(a–s) type on one end (as shown in Fig. 6 *b*) and a single hairpin of the HG1a(s–a) type on the other end. This structure will be called FB1. The other folded-back structure (FB2) had hairpin loops of the HG1a(s–a) type on one end and a single hairpin of the type WC2a(a–s) on the other end.

As mentioned in the methods the MD calculations were carried out with distance-dependent dielectric function and reduced charge on the phosphate groups, as for the molecular mechanics studies. However, because the structures were expected to undergo relatively larger structural changes during the MD simulations than those observed during energy minimization, there was a possibility of the K^+ binding site's being distorted in the absence of explicit K^+ ions. Therefore in the MD calculations two K^+ ions were included between the three G tetrads of the folded-back structures FB1 and FB2. A 350- ps MD simulation was carried out for each of the folded-back structures, at a constant temperature of 300 K (by coupling to a heat bath). For all the structures saved along the trajectory, rms deviations from the starting structure was calculated, and the plot of rms deviation versus time is shown in Fig. 7. As can be seen from that figure, after the first few picoseconds of the MD run, while the rms deviation from the starting structure stabilizes in the range between 1.6 and 1.8 Å for structure FB1, the rms deviation for FB2 stabilizes at a value close to 2.0 Å. This stable value of the rms deviation indicates that the structures are well equilibrated. The rms deviations are reasonably small (<2.0 Å) for all the 595 atoms, including hydrogens. Thus the MD simulation clearly shows that the structures obtained from our model building and energy minimization are energetically quite stable.

The stability of the folded-back structures is further evident from an analysis of the various hydrogen bonds in the stem as well as the loop regions. In the stem region of the folded-back structures almost all the hydrogen bonds remain intact throughout the MD run. Fig. 8 *a* and *b* shows the trajectories for the all eight hydrogen bonds in the first G tetrads of FB1 and FB2 as representative examples. It can be seen from Fig. 8 that the hydrogen-to-acceptor distance fluctuates around 1.9 Å for all hydrogen bonds. In the case of FB1 only the hydrogen bond HN2 G1–N7 G21 breaks (i.e. the HN2–N7 distance increases to 3.5 Å) during the

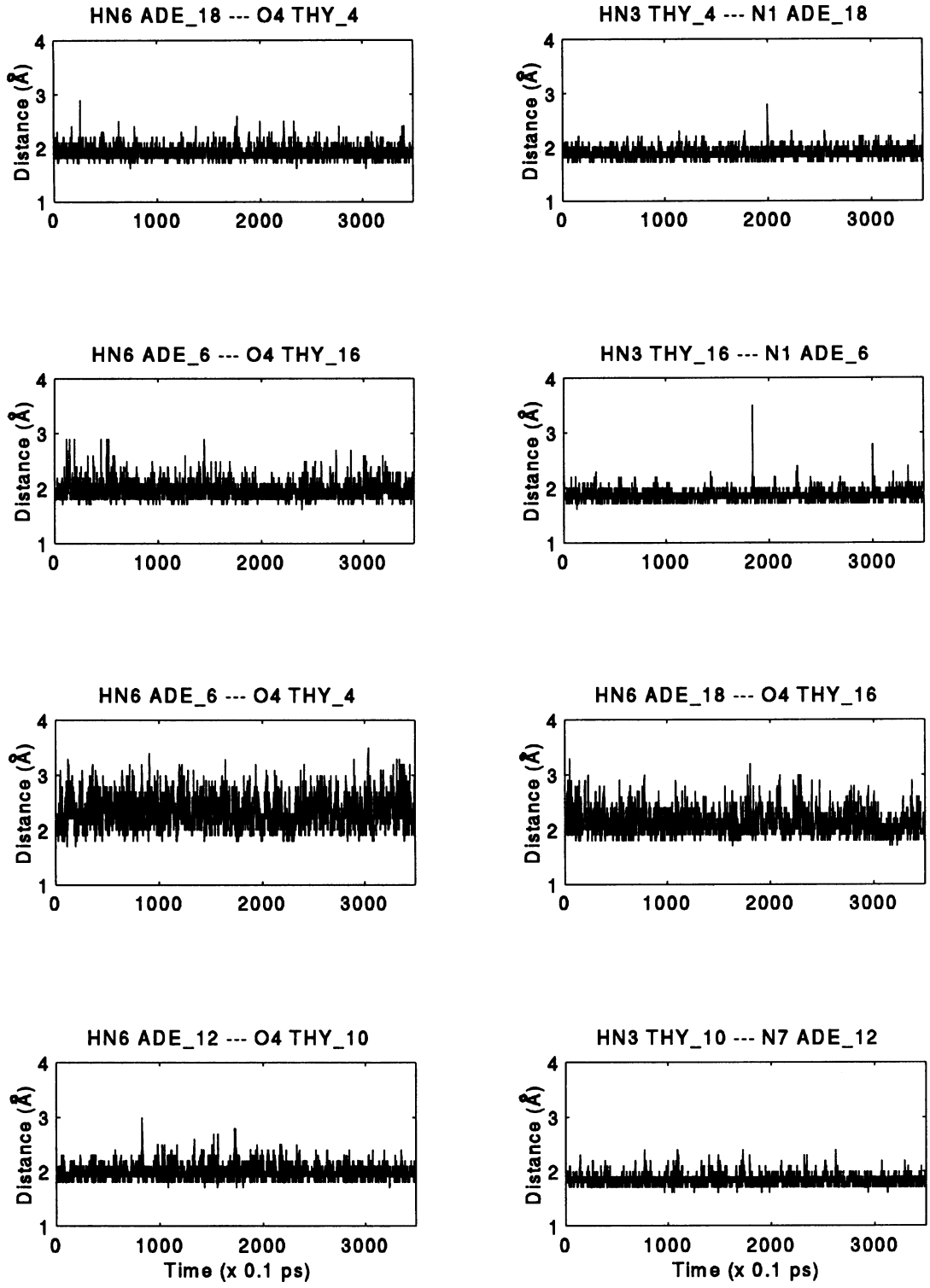

FIGURE 9 Plots of hydrogen-to-acceptor atom distance versus time for the eight hydrogen bonds in the various A:T pairs in the loop regions of the intramolecular folded-back structure FB1. The interloop Watson–Crick A:T pairs T4:A18 and T16:A6 form the TATA tetrad stacked over the G tetrad. T10 and A12 form an intraloop Hoogsteen A:T pair in the single TTA loop over the large groove at the other end of the G-tetraplex stem.

time period between 90 and 280 ps, but the bond is subsequently re-formed. However, the other hydrogen bond between the same G:G pair remains intact throughout, indicating a propeller-twistlike movement between the bases. Similarly, as shown in Fig. 8 *b*, in the structure FB2 of the two hydrogen bonds between G9 and G13, the one involving HN1 G13–O6 G9 remains intact throughout the MD run but the other hydrogen bond (HN2 G13–N7 G9) shows sharp transitions between 2.0 and 3.5 Å and finally remains near 3.5 Å from 260 ps until the end of the run. It is

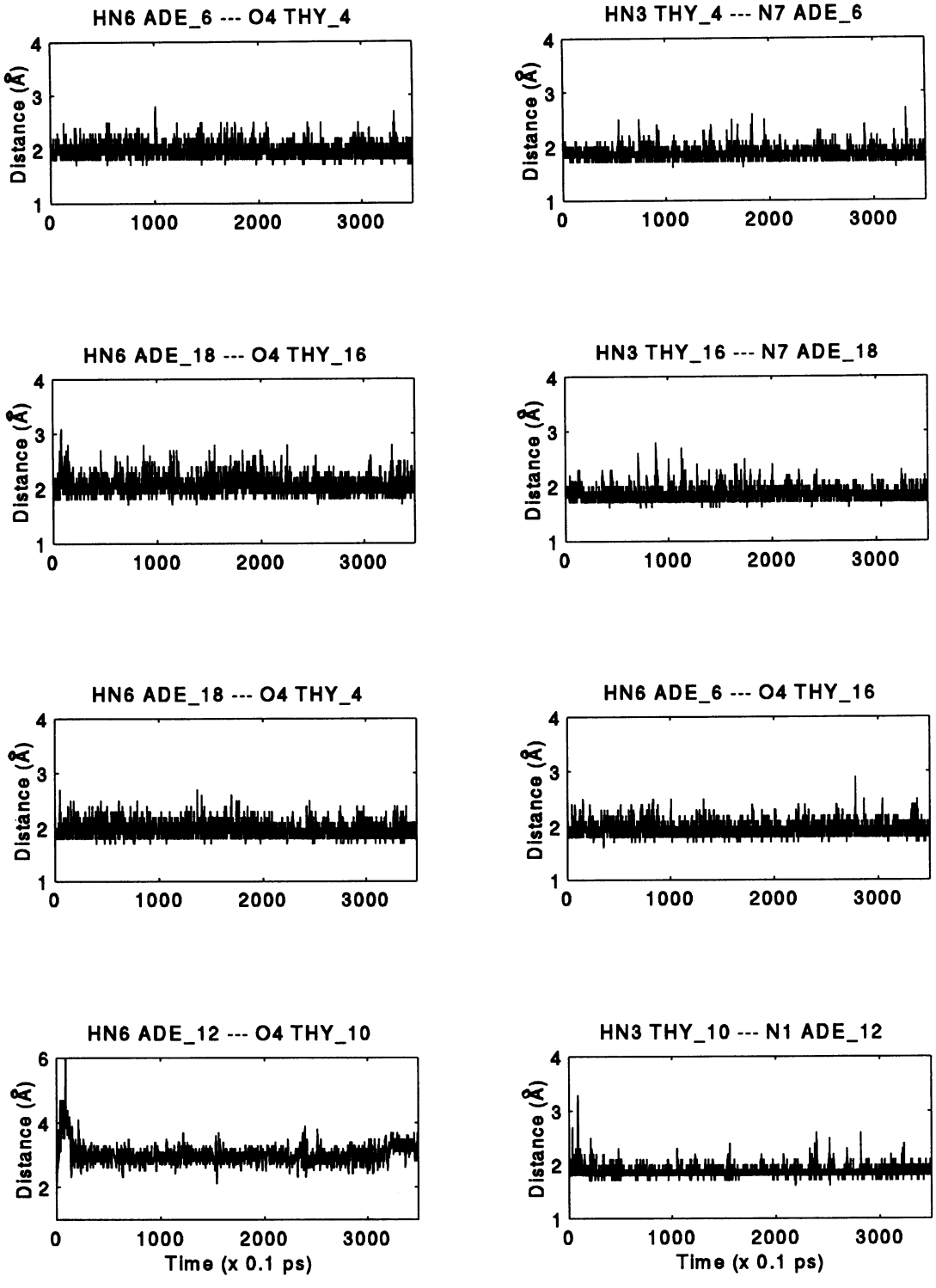

FIGURE 10 Plots of hydrogen-to-acceptor atom distance versus time for the eight hydrogen bonds forming the various A:T pairs in the loop regions of the intramolecular folded-back structure FB2. The intraloop Hoogsteen A:T pairs T4:A6 and T16:A18 form the TATA tetrad stacked over the G tetrad. T10 and A12 form an intraloop Watson–Crick A:T pair in the single TTA loop over the small groove at the other end of the G-tetraplex stem.

interesting to note that in structure FB2 this particular G:G pair precedes the A12:T10 pair in the loop, which also shows relatively large-scale structural fluctuations, as discussed in the next paragraph. Thus the sharp transitions observed in this G:G pair can be related to structural fluctuations in the A12:T10 pair. On the whole, the fluctuations observed for the hydrogen bonds in the stem region are smaller than the fluctuations observed for the hydrogen bonds in the loop region. The higher rigidity of the stem region compared with the loop can at least partly be attrib-

uted to the presence of two K^+ ions between the G tetrads of the G_3 stem.

Most of the hydrogen bonds in the loop regions of both of the folded-back structures also remain intact during the entire duration of the MD simulation, with only three hydrogen bonds being broken, i.e., the hydrogen-to-acceptor distance becomes more than 2.5 Å. Figs. 9 and 10 show the plots for all the hydrogen bonds that occur in the TATA tetrad and also in the isolated A:T pair in the single loop in the folded-back structures FB1 and FB2. As can be seen from Fig. 9, in structure FB1 all the hydrogen bonds in the two interloop Watson–Crick pairs (i.e., A18:T4 and A6:T16) and in the intraloop Hoogsteen pair (A12:T10) remain intact and show a mean fluctuation of only 0.2 Å about the mean hydrogen-to-acceptor distance of 2.0 Å, with a few excursions to slightly longer distances. However, the intraloop hydrogen bonds in the TATA tetrad of FB1 (i.e., HN6 A6—O4 T4 and HN6 A18—O4 T16) show larger fluctuations, in the range between 1.9 and 3.0 Å, indicating that these hydrogen bonds break and form during the course of the simulation.

In structure FB2 all the intraloop as well as the interloop hydrogen bonds in the TATA tetrad remain intact and fluctuate about the mean hydrogen-to-acceptor distance of 2.0 Å (Fig. 10). On the other hand, the two hydrogen bonds in the intraloop WC A12:T10 pair are initially broken. As can be seen from Fig. 10, at the very beginning of the simulation, while the HN6 A12–O4 T10 distance increases to 6.0 Å, the HN3 T10–N1 A12 distance approaches 3.4 Å, indicating a complete rupture of this A:T pair. However, after ~30 ps of the MD run the hydrogen bond HN3 T10—N1 A12 is again formed and remains close to 2.0 Å throughout the rest of the MD run, but the potential hydrogen bond between HN6 A12—O4 T10 is not formed, although the distance does reduce to ~3.0 Å after 30 ps and fluctuates about that value for the rest of the 350-ps MD run. It may be noted that this particular A:T pair is a WC pair in the small groove, and, as discussed in the model building section, it is essential for the thymine to be in syn or low anti glycosidic orientation to retain both the hydrogen bonds in this A:T pair. As discussed in the next subsection, the breakage of one of the hydrogen bonds in the A12:T10 pair is correlated to a transition in the glycosidic orientation of T10. The hydrogen bonds linking the T:T pairs in both of the folded-back structures remain intact throughout the 350-ps run.

Torsion angle fluctuations in the MD structures

The backbone and glycosidic torsion angles for the two folded-back structures, averaged over the last 300 ps of the MD simulation, are listed in Table 5, along with the standard deviations and also the corresponding values in the starting structure. As can be seen from Table 5, many of the guanines in the stem region of FB1 and FB2 do not undergo any dihedral transition but throughout the simulation show small fluctuations about their starting values. The dihedral transitions in some of the guanines take place in the first few picoseconds, and the dihedral angle values subsequently stabilize. In structure FB1, while the guanines in the anti conformation have their backbone torsion angles close to their starting values (Table 5), most of the guanines in the syn conformation have undergone dihedral transitions involving a change in (β, γ) from (g^-, g^-) to (t, g^+) within the first 10 ps. This dihedral transition for the syn guanines is shown in Fig. 11 *a* and *b* for the nucleotide G14 in the middle tetrad of FB1, as a representative case. On the otherhand, in structure FB2, α, β, and γ of some of the guanines in the syn conformation changed from (g^-, g^-, g^-) to (g^+, t, t), whereas guanines in the anti conformation showed a transition in ζ from g^- to t (Table 5). Only the guanines G7 and G8 of FB2 show sharp dihedral transitions after 200 ps of MD run, with a (g^-, g^-) to (t, g^+) transition in (β, γ) of G8, as seen in Fig. 11 *c* and *d*, accompanying a g^- to t transition in ζ of G7 (Fig. 11*e*). Hence, in these cases, averages of torsion angles over the period 50–350 ps have very large values of standard deviations, and the average torsion angle values do not give a true indication of the dihedral angles. In the energy-minimized structures the six backbone torsion angles for guanines, located one nucleotide away from the loop, were g^-, t, g^+, $t(_1E)$, t, and g^- with an anti glycosidic conformation and g^-, g^-, g^-, $t(^2E)$, t, and g^- for the guanines with syn glycosidic orientation. As mentioned above, after the MD simulation the six backbone torsion angles in many of the syn guanines changed to a g^-, t, g^+, $t(_1E/^2E)$, t, g^- combination, which is known to be stereochemically more favorable (Bansal and Sasisekharan, 1986). However, as can be seen from Table 5, in the structures obtained from MD a few other unusual combinations of torsion angles are found, as also observed for the antiparallel G-tetraplex structures obtained from crystallography and NMR.

In contrast to the guanines in the stem, the backbone torsion angles of some loop nucleotides show large fluctuations with significant dihedral transitions throughout the 350-ps MD run. This is reflected in the larger values of standard deviations listed in Table 5. The transitions generally occur in a pairwise correlated manner for the backbone torsion angles (α, γ) and (ε, ζ), between the three staggered conformations. The torsion angle trajectories for α and γ in T4 of FB1, ε and ζ in T17, and α and β in A18 of FB2 are shown in Fig. 12 as representative examples. Similar transitions are observed for several other loop residues constituting the TATA tetrads and TT pairs. It is interesting to note that the loop nucleotides can undergo such large torsion angle fluctuations while retaining all the intraloop and interloop hydrogen bonds, as shown in Fig. 9 and 10.

The glycosidic torsion angles of almost all the nucleotides in the two folded-back structures remain very close to the values obtained from the minimization studies, except for T10 in the structure FB2. As discussed in the Model Building section, this angle was constrained to have a value in the syn or low anti range because of the requirements for

TABLE 5 Mean backbone and glycosidic torsion angles in the two folded-back structures averaged over the time interval 50 to 350 ps

	Angle						
Nucleotide	α	β	γ	δ	ε	ζ	χ
Structure FB1							
G1			60 ± 7	141 ± 6	187 ± 6	266 ± 7	210 ± 8
			100	135	204	277	240
G2	292 ± 7	197 ± 7	50 ± 7	128 ± 7	187 ± 6	273 ± 7	45 ± 8
	325	276	287	155	166	277	55
G3	128 ± 10	186 ± 7	173 ± 7	70 ± 7	78 ± 8	76 ± 7	202 ± 7
	196	203	127	110	44	115	225
T4	105 ± 57	179 ± 8	125 ± 54	79 ± 7	241 ± 28	268 ± 95	215 ± 12
	270	159	299	139	180	3	253
T5	171 ± 31	184 ± 12	64 ± 10	77 ± 7	186 ± 6	285 ± 7	190 ± 8
	284	235	229	162	207	250	197
A6	284 ± 9	177 ± 7	61 ± 7	76 ± 5	183 ± 6	281 ± 6	213 ± 7
	231	222	72	153	188	231	266
G7	172 ± 8	188 ± 7	179 ± 7	151 ± 5	192 ± 5	284 ± 5	45 ± 6
	74	211	250	145	173	277	45
G8	279 ± 7	171 ± 5	56 ± 6	123 ± 8	187 ± 6	250 ± 9	237 ± 8
	212	208	102	128	209	269	238
G9	282 ± 8	190 ± 6	48 ± 7	134 ± 5	169 ± 5	278 ± 6	68 ± 7
	332	260	284	167	187	248	63
T10	155 ± 9	189 ± 6	185 ± 7	122 ± 11	207 ± 7	302 ± 8	208 ± 10
	65	174	263	168	177	255	239
T11	236 ± 9	79 ± 7	177 ± 5	145 ± 8	184 ± 6	248 ± 10	187 ± 7
	73	178	269	133	94	273	234
A12	276 ± 8	166 ± 6	57 ± 6	99 ± 7	60 ± 6	290 ± 5	158 ± 7
	222	211	110	121	91	298	173
G13	294 ± 7	165 ± 5	86 ± 6	154 ± 5	171 ± 5	262 ± 6	262 ± 6
	288	158	105	139	204	276	243
G14	294 ± 6	184 ± 6	56 ± 6	117 ± 7	188 ± 6	273 ± 7	47 ± 7
	325	277	287	155	167	277	54
G15	125 ± 10	186 ± 6	175 ± 7	71 ± 5	79 ± 7	78 ± 7	201 ± 6
	196	203	127	110	44	115	225
T16	64 ± 8	187 ± 8	167 ± 8	84 ± 6	260 ± 10	309 ± 11	209 ± 9
	270	159	299	139	180	3	253
T17	200 ± 45	188 ± 12	101 ± 94	83 ± 18	190 ± 8	278 ± 12	197 ± 10
	284	235	229	162	207	250	197
A18	275 ± 12	176 ± 7	59 ± 8	77 ± 5	179 ± 6	284 ± 6	211 ± 8
	231	222	72	153	188	231	266
G19	172 ± 9	194 ± 7	178 ± 7	147 ± 6	192 ± 5	282 ± 5	45 ± 7
	74	211	250	144	173	277	45
G20	283 ± 7	171 ± 5	59 ± 6	137 ± 5	193 ± 5	288 ± 5	232 ± 7
	211	209	101	127	206	276	238
G21	300 ± 5	274 ± 7	302 ± 8	155 ± 5			59 ± 7
	327	268	288	145			51
Structure FB2							
G1			64 ± 5	142 ± 4	185 ± 5	274 ± 5	48 ± 7
			290	156	171	277	50
G2	289 ± 6	167 ± 5	65 ± 6	132 ± 6	190 ± 5	225 ± 7	239 ± 6
	211	205	104	131	207	271	239
G3	56 ± 7	177 ± 7	288 ± 8	152 ± 5	176 ± 6	274 ± 7	61 ± 7
	331	260	283	169	187	246	64
T4	149 ± 11	184 ± 8	177 ± 8	139 ± 8	186 ± 9	258 ± 20	225 ± 10
	66	173	261	169	175	257	235
T5	105 ± 47	176 ± 11	236 ± 53	146 ± 16	298 ± 17	117 ± 37	229 ± 22
	77	176	267	135	96	271	235
A6	201 ± 39	177 ± 41	62 ± 10	108 ± 12	60 ± 7	285 ± 10	174 ± 9
	226	207	106	125	92	299	175
G7	159 ± 10	247 ± 17	178 ± 7	146 ± 8	190 ± 8	247 ± 40	240 ± 8
	289	158	105	139	204	275	243
G8	307 ± 8	232 ± 48	189 ± 121	148 ± 12	191 ± 6	279 ± 5	52 ± 9
	326	276	285	159	167	275	55
G9	280 ± 6	175 ± 6	54 ± 5	121 ± 5	62 ± 6	77 ± 6	260 ± 7
	207	204	117	121	46	56	237
T10	95 ± 8	187 ± 5	61 ± 7	142 ± 5	199 ± 7	228 ± 7	180 ± 7
	125	275	354	161	193	255	154
T11	93 ± 11	178 ± 7	186 ± 8	156 ± 6	194 ± 5	66 ± 6	180 ± 6
	81	152	200	144	130	65	188
A12	178 ± 6	98 ± 6	51 ± 5	134 ± 5	212 ± 24	158 ± 13	139 ± 6
	162	171	52	147	189	225	106
G13	82 ± 7	251 ± 19	189 ± 5	149 ± 5	185 ± 5	276 ± 6	58 ± 9
	65	229	253	152	173	277	49
G14	285 ± 7	171 ± 5	57 ± 7	144 ± 5	282 ± 7	160 ± 12	255 ± 7
	212	207	101	129	208	270	238
G15	73 ± 7	194 ± 9	197 ± 7	127 ± 16	187 ± 7	273 ± 8	47 ± 24
	332	260	284	169	187	246	64
T16	145 ± 14	185 ± 9	175 ± 8	126 ± 28	186 ± 9	285 ± 13	219 ± 15
	67	173	261	169	175	257	235
T17	153 ± 17	185 ± 9	182 ± 9	135 ± 10	252 ± 44	145 ± 33	210 ± 10
	77	176	267	135	96	271	235
A18	258 ± 38	150 ± 25	60 ± 9	131 ± 11	190 ± 7	260 ± 11	149 ± 16
	226	207	106	125	92	299	175
G19	276 ± 8	69 ± 8	173 ± 7	140 ± 16	193 ± 10	252 ± 12	228 ± 9
	289	158	105	139	204	275	243
G20	124 ± 15	168 ± 9	186 ± 8	79 ± 4	199 ± 6	287 ± 5	27 ± 10
	325	276	286	160	170	277	55
G21	291 ± 6	175 ± 5	71 ± 8	131 ± 30			254 ± 13
	210	205	109	117			228

The numbers in the second row for each nucleotide represent the torsion angle values in the starting structure.

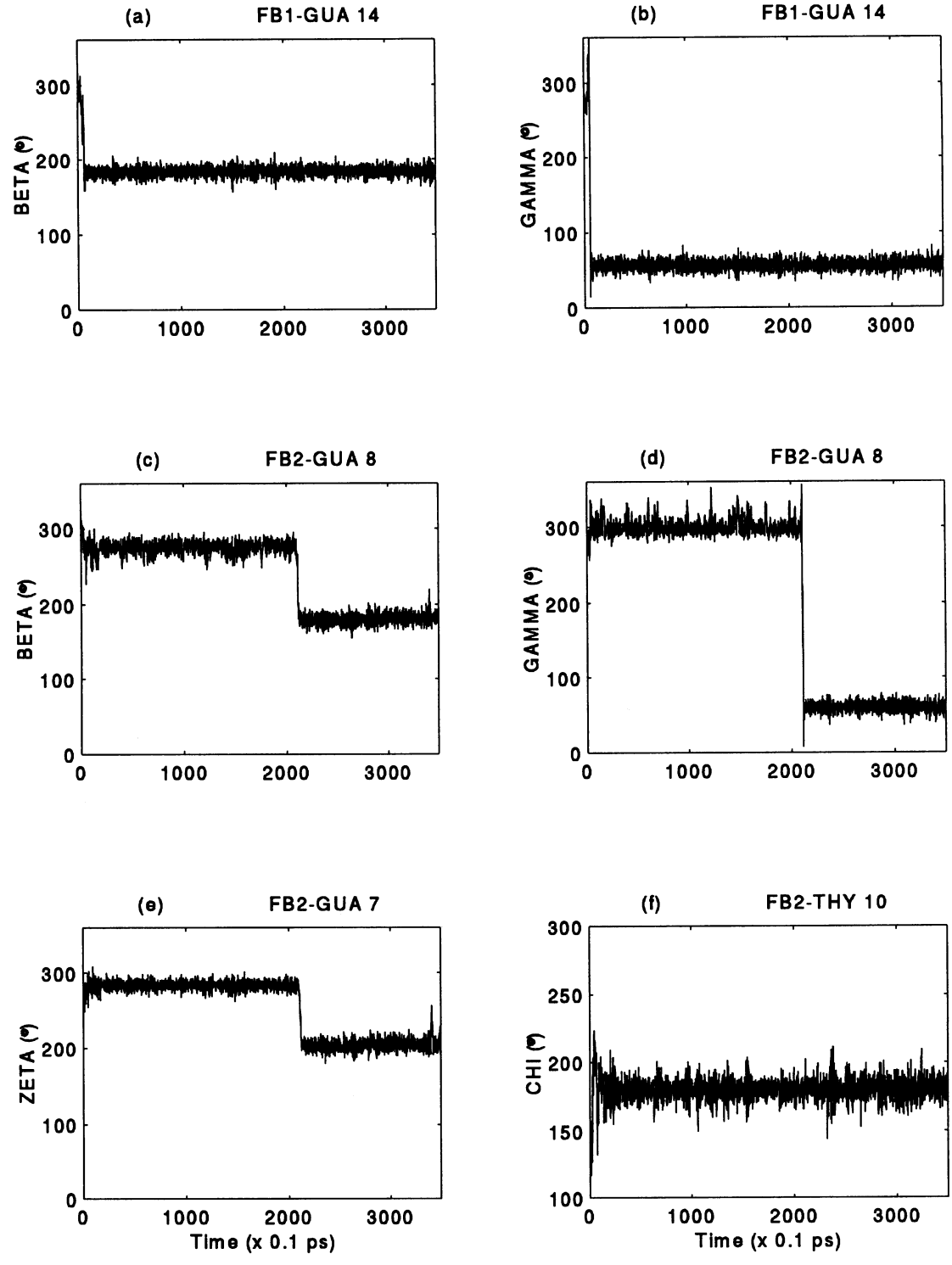

FIGURE 11 Time trajectories showing the fluctuations in some representative torsion angles in the guanine residues of the G-tetraplex stem regions: (*a*) β of G14 in FB1, (*b*) γ of G14 in FB1, (*c*) β of G8 in FB2, (*d*) γ of G8 in FB2, (*e*) ζ of G7 in FB2, (*f*) the glycosidic torsion χ for the residue T10 in the loop region of FB2.

the formation of a Watson–Crick A:T pair in the small groove. However, one of the hydrogen bonds in the A12:T10 pair breaks during the MD simulation, with the hydrogen-to-acceptor distance increasing to more than 3.0 Å (Fig. 10) and the glycosidic torsion for T10 changing from 154° to 180° (as shown in Fig. 11 *f*). In this structure (FB2) the intraloop A12:T10 pair is not a part of the TATA tetrad, unlike the WC2a(a–s) type of hairpin dimer (Fig. 6 *c*) studied by energy minimization in which the two small-groove intraloop A:T pairs form a tetrad. In the energy-

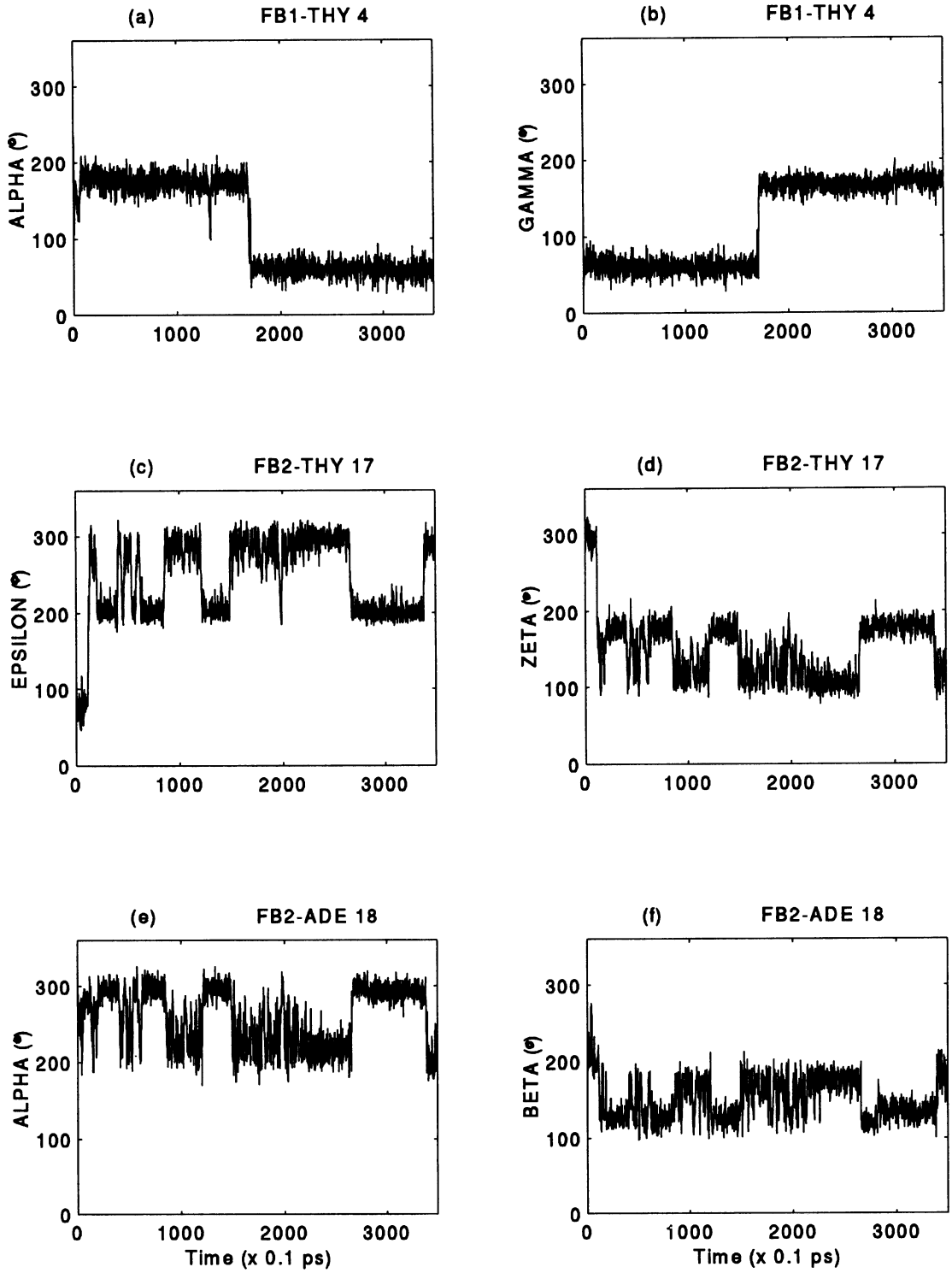

FIGURE 12 Time trajectories showing the relatively large fluctuations seen in some representative torsion angles in the loop regions of the folded-back structures FB1 and FB2 during the 350-ps MD simulation. (*a*) α of T4 in FB1, (*b*) γ of T4 in FB1, (*c*) ε of T17 in FB2, (*d*) ζ of T17 in FB2, (*e*) α of A18 in FB2, (*f*) β of A18 in FB2.

minimized structure of the WC2a(a–s) hairpin dimer all the hydrogen bonds in the TATA tetrad were retained, and the glycosidic torsion of the first thymine changed from 150° to 132° (Table 4). Thus this result from MD calculations supports the results obtained from simple model building with respect to correlations between glycosidic orientations and A:T pairing.

The FB1 and FB2 structures obtained at the end of 350 ps of a MD run were energy minimized, and the stereo diagrams of the energy minimized final structures are

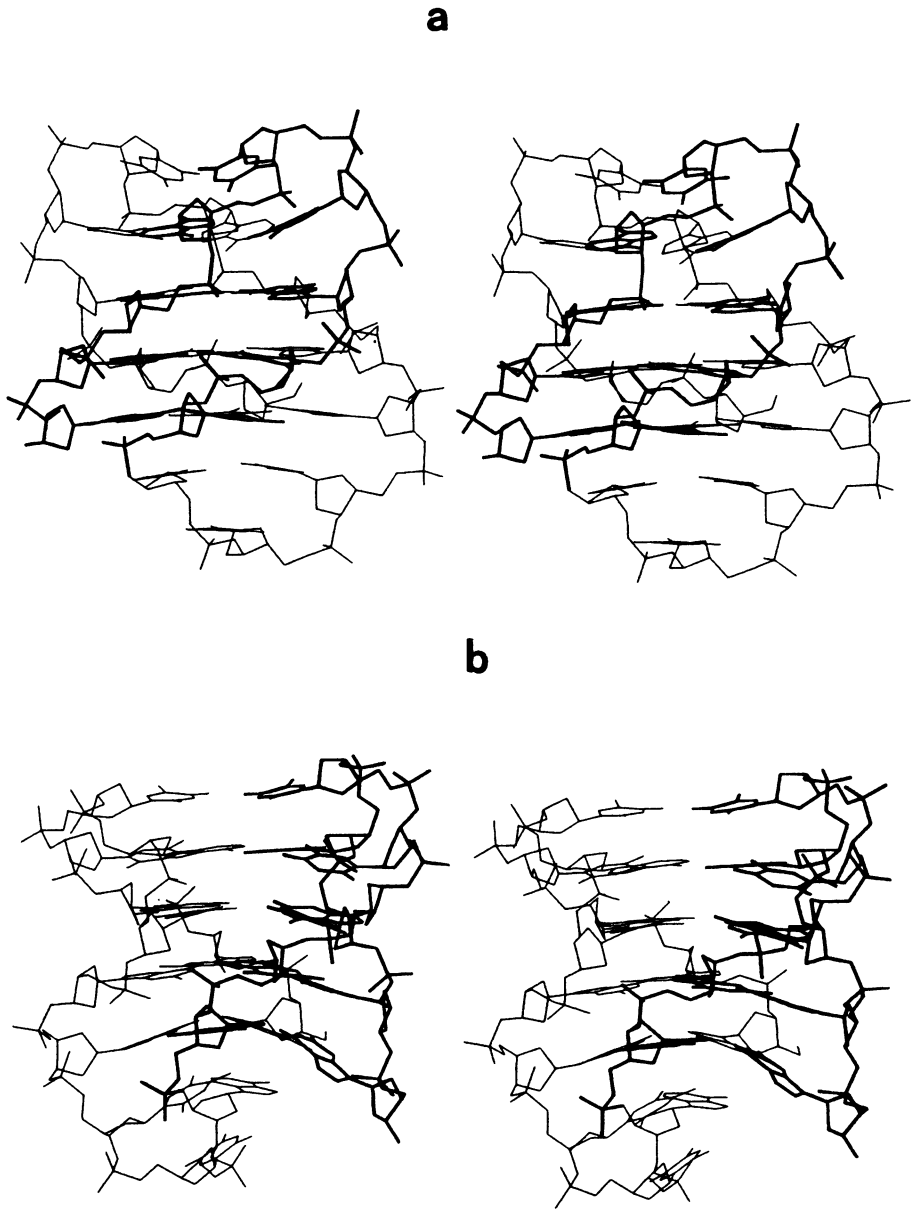

FIGURE 13 Stereo diagrams of the folded-back structures obtained by energy minimizing the final MD structure at the end of 350 ps. The nucleotides G1 to A12 have been shown in thin lines; the nucleotides G13 to G21 have been shown in thick lines. (*a*) FB1: The two loops over the small grooves are on one side of the G_3 stem and form a TATA tetrad with interloop A:T Watson–Crick pairing, and the single loop over the large groove at the other end of the G_3 stem has an intraloop Hoogsteen A:T pair. (*b*) FB2: The two loops over the large grooves are on one side of the G_3 stem and form a TATA tetrad with intraloop A:T Hoogsteen pairing, and the single loop over the small groove at the other end of the G_3 stem has an intraloop Watson–Crick A:T pair.

shown in Fig. 13 *a* and *b*. It can be seen from the figure that both of the folded-back structures maintain two small and two large grooves. In FB1 all the G tetrads are almost planar, with very small values of buckle and propeller types or out-of-plane distortions, while in FB2 the outermost G tetrad (G1-G9-G13-G21) shows relatively larger out-of-plane distortions.

In summary, the MD simulations indicate that the various model structures of human telomeric DNA, with different types of A:T pairing in the loop, represent different stable conformations. The structures during the MD run remain close to the starting models in terms of rms deviations, and most of the hydrogen bonds in the stem as well as in the loop regions remain intact. However, because our calculation did not include explicit solvent molecules, the possibility that some hydrogen bonds in the loop may break to form new hydrogen bonds with water cannot be ruled out. Such a possibility is unlikely for the stem, as it is held tightly by the K^+ ions in the cavities in addition to the eight hydrogen bonds in each tetrad. The MD simulations also illustrate some interesting features about the dynamics of the folded-back G tetraplexes in general. As mentioned above, the stem regions in the folded-back structures are extremely rigid, in terms of the fluctuations observed in the hydrogen bonds as well as of torsion angles. On the other hand, the loop region can have significant structural fluctuations in the torsion angles, while retaining a number of intraloop and interloop hydrogen bonds.

CONCLUSION

The results of the present study indicate that the human telomeric sequence repeat d($TTAG_3$) and the related se-

quence d(T_3AG_3) will prefer to adopt folded-back structures with G_3 stems and TTA or T_3A loops rather than structures with AGAG tetrads and T_2 or T_3 loops. The various G_3 tetraplex structures, although energetically they are very similar, have a wide range of structural variability because the presence of adenine along with thymines in the loop gives these sequence motifs the ability to adopt several different types of intraloop as well as interloop A:T pairs, leading to a number of different TATA tetrads when two loops are on the same side of the G_3-tetraplex stem. There have been relatively few studies of structural characterization of human telomeric DNA so far, and the various possible model structures presented here can help in rationalizing the results of chemical probing studies of these sequence motifs. The calculation of interproton distances indicates that each of the different possible structures could give rise to unique NOE signals in NMR experiments. It is also observed from model building that, although A:T pairs with Watson–Crick or Hoogsteen hydrogen bonding schemes can be accommodated in an antiparallel G tetraplex, the thymine can have an anti conformation only when the A:T pair occurs in the large groove of the tetraplex, whereas an A:T pair in the small groove of the tetraplex forces the thymine to adopt a syn conformation. Thus by taking into consideration the known propensity of thymines to adopt an anti conformation, it might be possible to design guanine-rich oligonucleotides with selective positioning of A and T, which would favor their folding into a particular type of structure.

The MD simulations on two of the folded-back model structures indicate that these structures, with various possible A:T pairs in the loop region, represent energetically stable conformations for human telomeric DNA. The MD simulations also indicate that the antiparallel G-tetraplex stem in these folded-back structures is very robust, and the network of G:G hydrogen bonds, as well as the presence of K^+ ions between the G tetrads, makes it an extremely rigid and stable structure. Considerably higher flexibility is exhibited by the backbone of the loop nucleotides, even though all nucleotides in the loop are also involved in hydrogen bonding and most of these are retained during the MD simulations. The fact that folded-back telomeric structures have a very rigid G-tetraplex stem and relatively flexible loop regions has also been indicated by NMR studies (Wang and Patel, 1992, Smith and Feigon, 1992, 1993).

The authors are grateful to SERC, IISc, for computational facilities. We also thank Mr. Ravi Kiran for help in preparation of the manuscript. D.M. was supported by a fellowship from the Department of Biotechnology, India.

REFERENCES

Aboul-ela, F., A. I. H. Murchie, and D. M. Lilly. 1992. NMR study of parallel-stranded DNA tetraplex formation by d(TG_4T). *Nature (London).* 360:280–282.

Aboul-ela, F., A. I. H. Murchie, D. G. Norman, and D. M. J. Lilley. 1994. Solution structure of a parallel-stranded tetraplex formed by d(TG_4T) in presence of sodium ions by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* 243:458–471.

AMBER, V. 3.0. 1987. Molecular mechanics program from the University of California, San Francisco.

Arnott, S., R. Chandrasekharan, and C. M. Martilla. 1974. Structure of polyinosinic acid and polyguanylic acid. *Biochem. J.* 141:537–543.

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

Balagurumoorthy, P., S. K. Brahmachari, D. Mohanty, M. Bansal, and V. Sasisekharan. 1992. Hairpin and parallel quartet structure for telomeric sequences. *Nucl. Acids Res.* 20:4061–4067.

Bansal, M., and D. Mohanty. 1993. Molecular mechanics studies of telomeric DNA. *J. Biomol. Struct. Dyn.* 10:a010.

Bansal, M., and N. Pattabiraman. 1989. Molecular mechanics studies on poly(purine).poly(pyrimidine) sequences in DNA: polymorphism and local variability. *Biopolymers.* 28:531–548.

Bansal, M., and V. Sasisekharan. 1986. Molecular model-building of DNA: constraints and restraints. *In* Theoretical Chemistry of Biological Systems. G. N. Szabo, editor. Elsevier, Amsterdam. 127–288.

Blackburn, E. H. 1991. Structure and function of telomeres. *Nature (London).* 350:569–573.

Blackburn, E. H. 1992. Telomerases. *Ann. Rev. Biochem.* 61:113–129.

Blackburn, E. H., and J. W. Szostak. 1984. The molecular structure of centromeres and telomeres. *Ann. Rev. Biochem.* 53:163–194.

Bruskov, V. I., V. N. Bushuev, M. S. Okon, N. V. Shulyupina, and V. I. Poltev. 1989. C—H . . . O hydrogen bonding in solutions of methylated nucleic acid base analogs as revealed by NMR. *Biopolymers.* 28: 589–604.

Cheng, Y. K., and B. M. Pettitt. 1992. Hoogsteen versus reversed-Hoogsteen base pairing: DNA triple helices. *J. Am. Chem. Soc.* 114: 4465–4474.

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

Chernyi, A. A., Yu P. Lysov, I. A. Il'ychova, A. S. Zibrov, A. K. Shchyolkina, O. F. Borisova, O. K. Mamaeva, and V. L. Florentiev. 1990. Four-stranded DNA helices: conformational analysis of regular poly(dT).poly(dA).poly(dA).poly(dT) helices with various types of base binding. *J. Biomol. Struct. Dyn.* 8:513–527.

Ferentz, A. E., J. Wiorkiewicz-Kuczera, M. Karplus, and G. L. Verdine. 1993. Molecular dynamics simulation of disulfide cross-linked DNA decamers. *J. Am. Chem. Soc.* 115:7569–7583.

Frank-Kamenetskii, M. 1992. The quadruplex and the vase. *Nature (London).* 356:105–105.

Guo, Q., M. Lu, and N. R. Kallenbach. 1992a. Effect of thymine tract length on the structure and stability of model telomeric sequences. *Biochemistry.* 32:3596–3603.

Guo, Q., M. Lu, and N. R. Kallenbach. 1992b. Adenine affects the structure and stability of telomeric sequences. *J. Biol. Chem.* 267:15293–15300.

Gupta, G., A. E. Garcia, Q. Guo, M. Lu, and N. R. Kallenbach. 1993. Structure of a parallel-stranded tetramer of the *Oxytricha* telomere repeat oligonucleotide d(T_4G_4). *Biochemistry.* 32:7098–7103.

Guschlbauer, W., J. F. Chantot, and D. Thiele. 1990. Four-stranded nucleic acid structures 25 years later: from guanosine gels to telomer DNA. *J. Biomol. Struct. Dyn.* 8:491–511.

Henderson, E., C. C. Hardin, S. K. Walk, I. Tinoco, Jr., and E. H. Blackburn. 1987. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine–guanine base pairs. *Cell.* 51: 899–908.

Henderson, E. R., and E. H. Blackburn. 1989. An overhanging 3′ terminus is a conserved feature of telomeres. *Mol. Cell. Biol.* 9:345–348.

Jin, R., K. J. Breslauer, R. A. Jones, and B. L. Gaffney. 1990. Tetraplex formation of a guanine-containing nonameric DNA fragment. *Science.* 250:543–546.

Kang, C., X. Zhang, R. Ratliff, R. Moyzis, and A. Rich. 1992. Crystal structure of four-stranded *Oxytricha* telomeric DNA. *Nature (London).* 356:126–131.

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

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

Miura, T., and G. J. Thomas, Jr. 1994. Structural polymorphism of telomeric DNA: interquadruplex and duplex–quadruplex conversions probed by Raman spectroscopy. *Biochemistry.* 33:7848–7856.

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

Mohanty, D., and M. Bansal. 1994. Conformational polymorphism in telomeric structures: loop orientation and interloop pairing in $d(G_4T_nG_4)$. *Biopolymers.* 34:1187–1211.

Moyzis, R. K., J. M. Buckingham, L. S. Cram, M. Dani, L. L. Deaven, M. D. Jones, J. Meyne, R. L. Ratliff, and J.-R. Wu. 1988. A highly conserved repetitive DNA sequence, $(TTAGGG)_n$, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. USA.* 85: 6622–6626.

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

Panyutin, I. G., O. I. Kovalsky, E. I. Budowsky, R. E. Dickerson, M. E. Rikhirev, and A. A. Lipanov. 1990. G-DNA: A twice folded DNA structure adopted by single-stranded oligo(dG) and its implications for telomeres. *Proc. Natl. Acad. Sci. USA.* 87:867–870.

Pearson, A. M., A. Rich, and M. Krieger. 1993. Polynucleotide binding to macrophage scavanger receptors depends on the formation of base-quartet-stabilized four stranded helixes. *J. Biol. Chem.* 268:3546–3554.

Pinnavaia, T. J., H. T. Miles, and E. D. Becker. 1975. Self-assembled 5′-guanosine monophosphate. Nuclear magnetic resonance evidence for a regular, ordered structure and slow chemical exchange. *J. Am. Chem. Soc.* 97:7198–7200.

Rose, I. A., K. R. Hanson, K. D. Wilkinson, and M. J. Wimmer. 1980. A suggestion for naming faces of ring compounds. *Proc. Natl. Acad. Sci. USA.* 77:2439–2441.

Ross, W. S., and C. C. Hardin. 1994. Ion-induced stabilization of the G-DNA quadruplex: free energy perturbation studies. *J. Am. Chem. Soc.* 116:6070–6080.

Sasisekharan, V., S. B. Zimmermann, and D. R. Davies. 1975. The structure of helical 5′-guanosine monophosphate. *J. Mol. Biol.* 92:171–179.

Scaria, P. V., S. J. Shire, and R. H. Shafer. 1992. Quadruplex structure of $d(G_3T_4G_3)$ stabilized by K^+ or Na^+ is an asymmetric hairpin dimer. *Proc. Natl. Acad. Sci. USA.* 89:10336–10340.

Schultze, P., R. F. Macaya, and J. Feigon. 1994. Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). *J. Mol. Biol.* 235:1532–1547.

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

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

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

Smith, F. W., and J. Feigon. 1993. Strand orientation in the DNA quadruplex formed from the *Oxytricha* telomere repeat oligonucleotide $d(G_4T_4G_4)$ in solution. *Biochemistry.* 32:8682–8692.

Sundquist, W. I., and S. Heaphy. 1993. Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus I genomic RNA. *Proc. Natl. Acad. Sci. USA.* 90:3393–3397.

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

Taylor, R., and O. Kennard. 1982. Crystallographic evidence for the existence of C—H...O, C—H...N and C—H...Cl hydrogen bonds. *J. Am. Chem. Soc.* 104:5063–5070.

Tidor, B., K. K. Irikura, B. R. Brooks, and M. Karplus. 1983. Dynamics of DNA oligomers. *J. Biomol. Struct. Dyn.* 1:231–252.

Walsh, K., and A. Gualberto. 1992. MyoD binds to the guanine tetrad nucleic acid structure. *J. Biol. Chem.* 267:13714–13718.

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

Wang, K. Y., S. Swaminathan, and P. H. Bolton. 1994. Tertiary structure motif of *Oxytricha* telomere DNA. *Biochemistry.* 33:7517–7527.

Wang, Y., and D. J. Patel. 1992. Guanine residues in $d(T_2AG_3)$ and $d(T_2G_4)$ form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry.* 31:8112–8119.

Wang, Y., and D. J. Patel. 1993. Solution structure of human telomeric repeat $d[AG_3(T_2AG_3)_3]$ G-tetraplex. *Structure.* 1:263–282.

Weiner, S. J., P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. Weiner. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106:765–784.

Weiner, S. J., P. A. Kollman, D. T. Nguyen, and D. A. Case. 1986. An all atom force field for simulation of proteins and nucleic acids. *J. Comp. Chem.* 7:230–252.

Williamson, J. R. 1994. G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.* 23:703–730.

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

Wyatt, J. R., T. A. Vickers, J. L. Robertson, R. W. J. Buckheit, T. Klimkait, E. DeBaets, P. W. Davis, B. Rayner, J. L. Imbach, and D. J. Ecker. 1994. Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immuno deficiency virus envelope-mediated cell fusion. *Proc. Natl. Acad. Sci. USA.* 91:1356–1360.

Zhurkin, V. B., G. Raghunathan, N. B. Ulyanov, R. D. Camerini-Otero, and R. L. Jernigan. 1994. A parallel DNA triplex as a model for homologous recombination. *J. Mol. Biol.* 239:181–200.

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