Conformational Polymorphism in Telomeric Structures: Loop Orientation and Interloop Pairing in $d(G_4T_nG_4)$

DEBASISA MOHANTY and MANJU BANSAL*

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

SYNOPSIS

Sequence repeats constituting the telomeric regions of chromosomes are known to adopt a variety of unusual structures, consisting of a G tetraplex stem and short stretches of thymines or thymines and adenines forming loops over the stem. Detailed model building and molecular mechanics studies have been carried out for these telomeric sequences to elucidate different types of loop orientations and possible conformations of thymines in the loop. The model building studies indicate that a minimum of *two* thymines have to be interspersed between guanine stretches to form folded-back structures with loops across adjacent strands in a G tetraplex (both over the small as well as large groove), while the minimum number of thymines required to build a loop across the diagonal strands in a G tetraplex is *three*. For two repeat sequences, these hairpins, resulting from different types of folding, can dimerize in three distinct ways—i.e., with loops across adjacent strands and on same side, with loops across adjacent strands and on opposite sides, and with loops across diagonal strands and on opposite sides—to form hairpin dimer structures. Energy minimization studies indicate that all possible hairpin dimers have very similar total energy values, though different structures are stabilized by different types of interactions. When the two loops are on the same side, in the hairpin dimer structures of $d(G_4T_nG_4)$, the thymines form favorably stacked tetrads in the loop region and there is interloop hydrogen bonding involving two hydrogen bonds for each thymine–thymine pair. Our molecular mechanics calculations on various folded-back as well as parallel tetraplex structures of these telomeric sequences provide a theoretical rationale for the experimentally observed feature that the presence of intervening thymine stretches stabilizes folded-back structures, while isolated stretches of guanines adopt a parallel tetraplex structure. © 1994 John Wiley & Sons, Inc.

INTRODUCTION

Sequence motifs of the type $d(T_nAG_3)_m$ or $d(T_nG_4)_m$, with n = 2–4 and m = 2 or 4, are known to occur at the ends of chromosomes.[1-4] Though the exact repeat sequence varies between species—for example T_2G_4 in *Tetrahymena*, T_4G_4 in *Oxytricha*, T_3AG_3 in *Arabidopsis*, and $TTAG_3$ in humans—the conserved feature of all these sequences is the presence of stretches of guanines interspersed by short stretches of thymines alone or with a single adenine at the 3′ end of thymine stretch. The structures adopted by such sequence motifs have been implicated to be of crucial importance for the biological function of telomeres.[3-8] Recent experimental studies[9-17] have demonstrated that these sequence motifs can adopt tetramolecular structures by association of four different strands, bimolecular structures by dimerization of two hairpins, or unimolecular structures by intramolecular folding of a single strand, and all these structures are stabilized by Hoogsteen bonded guanine tetrads.[18-22] The tetramolecular structures are essentially similar to the parallel tetraplex structures proposed for poly(G) from fiber diffraction studies,[20,21] while unimolecular and bimolecular folded-back structures consist of a four-stranded guanine stem and thymines constituting the loop regions. It is interesting to note that, while oligo-guanine strands per se prefer a four-stranded parallel tetraplex, the presence of inter-

Biopolymers, Vol. 34, 1187–1211 (1994)
© 1994 John Wiley & Sons, Inc. CCC 0006-3525/94/091187-25

* To whom correspondence should be addressed.

vening thymines stabilizes the folded-back structures over the parallel tetraplex. To date most of the information on these telomeric DNA has been obtained from studies using gel mobility, chemical probing, calorimetric measurements, or CD spectroscopy, and gives only gross structural features. The conformational details of such structures have been elucidated from the single crystal structure of $d(G_4T_4G_4)$,[16] a few high-resolution nmr experiments,[17,23-25] and model building and theoretical studies.[12,26] It is now well established that, while in four-stranded parallel structures adopted by isolated stretches of guanines all nucleotides have the *anti* conformation,[27-30] similar to the fiber models of poly(G),[20,21] in the folded-back structures the guanines adopt an alternating *syn–anti* conformation along the strand with an inverted stacking arrangement.[16,17,23] However, with the same alternating *syn–anti* conformation of guanines, the nucleotide chain can fold in a variety of ways.[31] In the crystal structure, the sequence $d(G_4T_4G_4)$ adopts a hairpin dimer structure, with the two T_4 loops being on opposite ends and joining adjacent strands across larger groove in the G tetraplex,[16] while the solution structure of the same sequence, obtained by nmr, is a hairpin dimer with diagonal loops on opposite ends of the G tetraplex.[17,23] On the other hand, hairpin dimer structures with loops on the same side and possible interloop hydrogen bonding have been proposed from theoretical studies.[12,32] Recent x-ray crystallographic[35] and nmr studies[33,34] have established that the oligonucleotide $d(G_2T_2G_2TGTG_2T_2G_2)$, which is a highly potent inhibitor for thrombin, in fact adopts an intramolecular folded-back structure with all three loops across adjacent strands and the two T_2 loops on the same side, though the strand orientations are different in crystal and solution.[34] On the other hand, as determined from nmr,[36] the human telomeric sequence $d[AG_3(TTAG_3)_3]$ adopts an intramolecular folded-back structure in solution with the central TTA stretch forming a diagonal loop on one end of the G_3 stem and the other two TTA stretches forming loops across adjacent strands on the other side of the G_3 stem. Thus, experimental studies clearly indicate that telomeric DNA can exhibit a high degree of conformational polymorphism and is capable of adopting a variety of unusual structures depending on the sequence repeat and environment, and these structures have a crucial role not only in biological processes, but their structural diversity could also be used in therapeutic applications. We have therefore carried out detailed model building and molecular mechanics studies to characterize various possible types of folding and association of these telomeric structures and to understand how the presence of thymines stabilizes folded-back structures over the parallel G tetraplex.

MODEL BUILDING OF HAIRPIN DIMER STRUCTURES

Telomeric sequences with four repeats of $d(T_nG_4)$ can adopt intramolecular folded-back structures with a Greek key or Indian key type of chain orientation (Figure 1),[31] while two repeat sequences or synthetic oligonucleotide sequences of the type $d(G_4T_nG_4)$ can adopt a variety of hairpin dimer structures (Figure 2). Since all possible intramolecular folded back structures can be obtained from different combinations of $d(G_4T_nG_4)$ hairpins, the model building studies have been carried out for various hairpins and their dimerization in different orientations have been examined in detail.

DNA hairpins, with Watson–Crick base pairs in the stem, have been extensively studied by theoretical as well as experimental methods.[37-47] As ob-

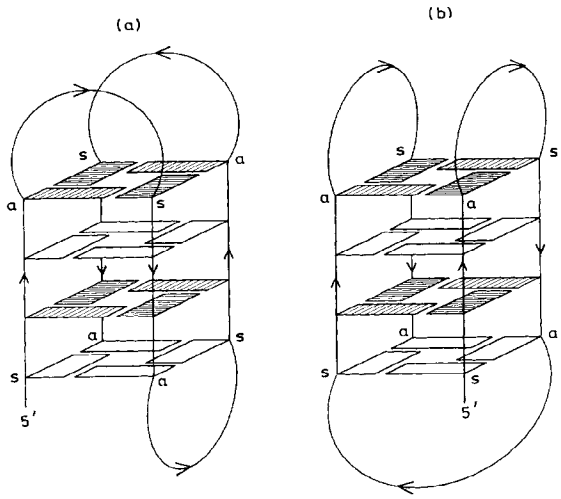

Figure 1. Schematic diagram of the monomolecular structure arising from intramolecular folding of the sequence $d[G_4(T_nG_4)_3]$. (a) Greek key type folding and (b) Indian key type folding.[31] The guanines have been represented as rectangular planes, with the longer edge indicating the donor side of guanine that contains N1 and N2, while the shorter edge indicates the acceptor side containing N7 and O6. The alternate guanines along a strand have been shaded to differentiate between the two faces of the guanines that point up due to the alternating *syn–anti* conformation about the glycosidic bonds resulting from base flipover.[26]

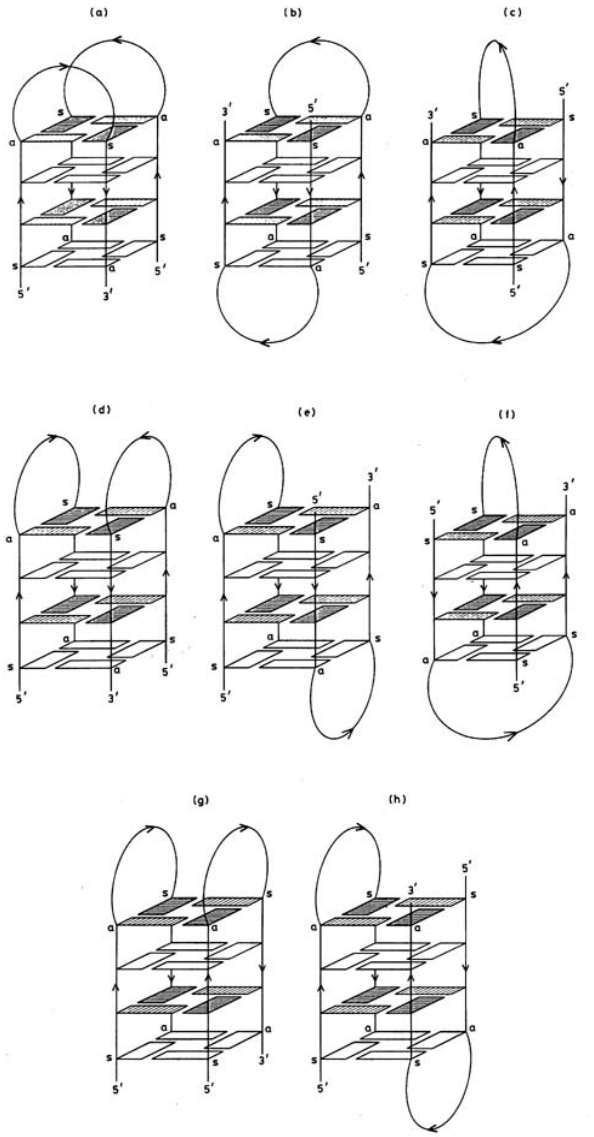

Figure 2. Schematic diagram of various possible hairpin dimer structures for the sequence $d(G_4T_nG_4)$ with alternating *syn-anti* arrangement of guanines along the nucleotide strand. Symmetric structures: (a) loops over larger groove and on same side; (b) loops over larger groove and on opposite side; (c) loops across diagonal and 5′-guanine stretches of the two hairpins form the larger groove, while the 3′-guanine stretches form the smaller groove; (d) loops over smaller groove and on same side; (e) loops over smaller groove and on opposite side; (f) loops across diagonal and 5′ guanine stretches of the two hairpins forming the smaller groove, while the 3′ guanine stretches form the larger groove. Asymmetric structures: (g) with loop on same side and one hairpin folds over the small groove, while the other hairpin folds over the large groove; (h) with loop on opposite side and one hairpin folds over the small groove, while the other hairpin folds over the large groove. In all the structures shown (a–h), the 5′ to 3′ direction of the loop joins a guanine in *anti* conformation to a guanine in *syn* conformation, but it is also possible to have structures with loops joining a guanine in *syn* conformation to a guanine in *anti* conformation.

served from these studies, the nucleotides in the loop region adopt conformations such that there could be favorable stacking interactions. An explanation based on the principle of favorable loop stacking has been put forth by Hasnoot et al.[38] for the optimum number of nucleotides required for the formation of DNA and RNA hairpins. This criteria, based on the so-called Hasnoot Plot, has been used subsequently for the model building of hairpins involved in intramolecular triplex structures.[45] The Monte Carlo technique has also been used for an exhaustive search of various possible hairpin conformations and suggests the possibility of preferred loop size being correlated to the structure of the helical stem.[46] In the Hoogsteen hydrogen-bonded *G*-tetraplex, since the structure of the helical stem is different from duplex DNA, the loops may also prefer a different conformation. In addition, when the two loops are on the same side of the tetraplex stem, as in the folded-back structure, there is a possibility of interloop hydrogen bonding.[12,32] Therefore it was decided to determine the minimum number of nucleotides required in the loop region for various types of folding and whether it is possible to build hairpins with all bases in the loop being stacked as well as being involved in intra- or interloop hydrogen bonding. In fact, in the crystal structure of $d(G_4T_4G_4)$,[16] the first and the fourth thymines in the loop are stacked over the G tetrad, while in the structure of the aptamer[34,35] two diagonal thymines form a T:T pair. In the stem region of the folded-back structures, though a variety of conformations are possible, only those with an alternating *syn-anti* arrangement of guanines along the strand have been considered, as only such structures have been observed experimentally[16,17] and have favorable intra- as well as interstrand energy unlike the *all-syn-all-anti* structures, where the *all-syn* strand is unfavorable.[26]

The $d(G_4T_nG_4)$ hairpins can have thymine stretches forming a loop across diagonal strands (Figure 2c and f) in a tetraplex or across adjacent strands (Figure 2). The hairpins with loops across adjacent strands can also be of two different types depending on the glycosidic orientation and pairing of guanines in the stem region. If the nucleotide chain folds in such a way that in each hairpin the guanines in *anti* conformation use their donor edges

to form hydrogen bonds (involving N2 and N1) with guanines in *syn* conformation containing the acceptor groups (N7 and O6), then the sugars across the G-G pairs in the stem point away from each other (Figure 4a of Ref. 26), resulting in a hairpin with a loop across the larger groove (Figure 2a and b). On the other hand, if the guanines in *anti* conformation use their acceptor edge to form hydrogen bonds with guanines in *syn* conformation providing the donor groups, then the sugars across the G-G pairs in the stem point toward each other (Figure 4a of Ref. 26), resulting in a hairpin with a loop across the smaller groove (Figure 2d and e). Each of the above three types of hairpins can have two subclasses—i.e., a loop joining a guanine in *syn* conformation to a guanine in *anti* conformation and vice versa. It is found from model building studies that the minimum number of nucleotides required to build a loop across adjacent strands of a G tetraplex is *two* and across diagonal strands is *three*. This observation can be rationalized by the fact that the minimum number of nucleotides required to build any of the above-mentioned hairpins depends on the distance between the O3′ atom of 5′ guanine preceding the loop and the C5′ atom of 3′ guanine following the loop. These distances for different antiparallel arrangements of alternating *syn–anti* strands in G-tetraplex structures that have been studied earlier are given in Table I and it can be seen that while the smallest O3′---C5′ distance across diagonal strands is $\approx$ 20Å, that across adjacent strands is 12 Å.

Since the aim was to build hairpin dimers with optimal base stacking and hydrogen-bond formation, an attempt was made to build hairpins with appropriately stacked thymines such that, when two such hairpins form symmetric dimers with loops on the same side (Figure 2a and d), thymines from two hairpins are linked through a pair of hydrogen bonds. Thus for the T_2 hairpin loop, two thymines were kept stacked over the GG pair at a height of 3.4 Å in appropriate orientation, so that after applying a twofold rotation about the helix axis of the stem, the 5′ thymine in each loop forms two hydrogen bonds (N3---O4) with the 3′ thymine of the other hairpin. Thus the thymines from two hairpins form a tetrad stacked over the G tetrad. Starting from such an arrangement of bases, backbone atoms in the loop consisting of three phosphates and two sugar moieties were built following stereochemical guidelines.[48] It was possible to build stereochemically satisfactory structures with both thymines in *anti* conformation and sugar rings having C2′-*endo* pucker. Similarly, a structure could be built for the T_3 loop, wherein the first and the third thymines were stacked over the G tetrad at a height of 3.4 Å, while the middle thymine stacks over the plane of first and third thymines. The dyad-related thymines of the other hairpin form hydrogen bonds with these three thymines, involving N3 as donor and O4 as acceptor. The backbone of this T_3 loop also had all the three glycosidic torsions in *anti* conformation and sugar puckers were C2′-*endo*. A similar structure was also built for the T_4 loop, where the first and the fourth thymines in the loop stacked over the GG pair, while the second and third thymines stacked over the first and fourth thymines, and all glycosidic torsions in the loop were *anti*. In the dimer structure obtained by applying a twofold rotation about the helix axis, while the second and third thymine formed tetrads, identical to tetrads in two and three base loops, with hydrogen bonds involving N3 and O4, the first and fourth thymine formed a tetrad with a different hydrogen-bonding scheme involving N3 to O4 and N3 to O2. It is important to note that

Table I The C5′---O3′ Distances (in Å) Across Antiparallel Strands in G-Tetraplex Stuctures Analyzed Earlier[26,a]

	Diagonal		Adjacent			
			Small Groove		Large Groove	
	syn to *anti*	*anti* to *syn*	*syn* to *anti*	*anti* to *syn*	*syn* to *anti*	*anti* to *syn*
S	21.7	22.4	13.3	12.3	17.4	16.7
N	22.0	19.9	14.6	13.2	16.8	15.3
R	21.9	19.7	13.5	12.4	17.4	15.5

[a] S represents starting structure for energy minimization, while N and R represent structures minimized with normal and reduced charge on phosphate oxygens.

such hairpins with stacking and hydrogen bonding in the loop are possible for all types of hairpins across adjacent strands, i.e., over large as well as small grooves and joining guanine in *syn* conformation to guanine in *anti* conformation and vice versa. In addition, structures with diagonal loops similar to the ones observed in nmr studies[17,23] were built for T_3 and T_4 hairpins without any hydrogen bonding. In each of the above-mentioned hairpins, the guanine stem was generated with alternating *syn–anti* conformation along the chain, following the model building procedure described earlier.[26]

The various $d(G_4T_nG_4)$ hairpins described above were used to generate the following different types of symmetric hairpin dimers:

1. Loops across adjacent strands and on the same side of the tetraplex stem obtained by a twofold rotation about the helix axis of the stem (Figure 2a and d). The resulting hairpin dimer has adjacent strands in antiparallel orientation and there are two large and two small grooves in the G tetraplex stem.
2. Loops across adjacent strands and on opposite sides of the tetraplex stem obtained by a twofold rotation about an axis perpendicular to the helix axis and passing between the second and third G tetrad (Figure 2b and e). The resulting hairpin dimer has a G tetraplex stem exactly identical to the previous one and also has two large and two small grooves.
3. Loops across diagonal strands and on opposite sides of the tetraplex stem obtained by a twofold rotation about an axis perpendicular to the helix axis and passing between the second and third G tetrad (Figure 2c and f). The resulting hairpin dimer has diagonal strands antiparallel; thus adjacent strands have both parallel and antiparallel orientation as shown in Figure 2c and f. Thus in the G tetraplex there are two medium grooves resulting from *syn–syn* or *anti–anti* G-G pairing across the parallel orientation of guanine strands and one large and one small groove arising from the antiparallel orientation of guanine strands.[26] It is important to note that though there is only one type of diagonal loop hairpin, two types of dimers can be formed (Figure 2c and f) depending on whether the 5′-terminal guanine stretches of the two hairpins form the large groove or the small groove.

Asymmetric hairpin dimers, which can be obtained by dimerizing a large groove hairpin with a small groove hairpin (Figure 2g and h), have not been considered for energy minimization studies. However, such an arrangement of hairpin loops has been observed in the intramolecular folded-back structure of $d[AG_3(TTAG_3)_3]$.[36] It is important to note that hairpin dimers of the type 2 and 3 (above) were symmetric because of the even number of guanines present in the tetraplex stem, whereas for an odd number of guanines such dimers become asymmetric, with the loop in one hairpin connecting a guanine in *syn* conformation to a guanine in *anti* conformation while in the other hairpin the G at the 5′ end of the loop is in *anti* conformation and the G at the 3′ end of the loop is in *syn* conformation.

Parallel tetraplex structures were also built for $d(G_4T_nG_4)$, n = 2–4, starting from Arnott's fiber model[20] and keeping the thymines in the stem in exactly the same conformation as guanines, in order to compare the energies of parallel vs folded-back structures.

ENERGY MINIMIZATION

The various parallel tetraplex and hairpin dimer structures obtained from model building and also the crystal structure of $d(G_4T_4G_4)$ (Ref. 16; coordinates obtained from the Nucleic Acid Database, see Ref. 49), were energy minimized to an rms gradient of 0.09 kcal/mole/Å using the AMBER 3.0[50,51] all-atom force field. In the minimization no constraints were applied on the hydrogen bonds or any dihedral angles. Since a large number of structures with different types of folding and different numbers of nucleotides in the loop were to be considered for minimization, to reduce the computational requirement solvent atoms were not included explicitly in the calculations. Thus solvent effects were mimicked implicitly by using a distance-dependent dielectric function $\epsilon_{ij} = R_{ij}$,[52,53] and no cutoff distance was used in the calculation of nonbonded pairs. In addition to the calculations carried out with normal AMBER electrostatic charges, the structures were also minimized with a reduced charge on anionic phosphate oxygens (by an amount of 0.7 of electronic charge, so that total charge on the molecule becomes −0.3 per phosphate group), to mimic the screening of phosphate–phosphate repulsion. Such an approach has been used earlier in the simulations of DNA duplex and triplex structures,[54–56] in our earlier molecular mechanics study on G tetraplexes,[26] and also in the rMD refinement of triplex[57] and tetraplex[36] structures obtained from nmr experiments.

In addition, to study the effect of the K^+ ion, for the $d(G_4T_4G_4)$ parallel tetraplex structure and three representative hairpin dimers, with loops across adjacent strands on same side, opposite side, and across diagonal strands, were also energy minimized, keeping a K^+ ion between adjacent G tetrads.

Similarly, to estimate the contribution of solvent molecules to the relative stability of parallel tetraplex and different hairpin dimer structures, minimization was carried out for four representative structures of $d(G_4T_4G_4)$ (three types of hairpin dimers and parallel tetraplex) with the inclusion of explicit water molecules and counterions, using AMBER 4.0.[58] The structures obtained from model building were first energy minimized for 100 cycles without solvent to remove bad contacts if any. The counterions (Na^+) were placed at a distance of 5 Å from the phosphorus atom along the bisector of OA-P-OB angle, using the EDIT module of AMBER 4.0.[58] Thus while the hairpin dimers had 22 Na^+ counterions, the parallel tetraplex had 44 counterions associated with it. The DNA with its associated counterions were then placed in a cube of TIP3P water and water molecules that had oxygens closer than 3.0 Å or hydrogens closer than 2.0 Å from any atom of DNA or counterion were removed from the box. Similarly, water molecules that had oxygen or hydrogen atoms farther than 9.0 Å from any counterion or DNA atom were also removed. The minimization was carried out to an rms gradient of 0.09 kcal/mol/Å using the periodic boundary condition, a nonbonded cutoff distance of 9.0 Å, and a constant dielectric of $\epsilon = 1$. No other constraint was imposed on the system during the minimization.

RESULTS AND DISCUSSION

Analysis of Structural Features

Hydrogen-Bond Patterns. The energy-minimized structures for $d(G_4T_4G_4)$ in the parallel tetraplex and hairpin dimer conformation are shown in Figures 3 and 4(a and b) as representative examples of the parallel and hairpin dimer structures. In the minimized structures of parallel tetraplexes the thymines form hydrogen-bonded tetrads (with the N3 of each thymine being hydrogen bonded to O4 in the adjacent strand with an H---O distance of 1.9 Å; (Figure 5), even though they were not hydrogen bonded in the starting structures. Such hydrogen-bonded tetrads have been observed for uracils in the nmr structures of $r(UG_4U)$.[28] Though such hydrogen bonds have not been unambiguously identified

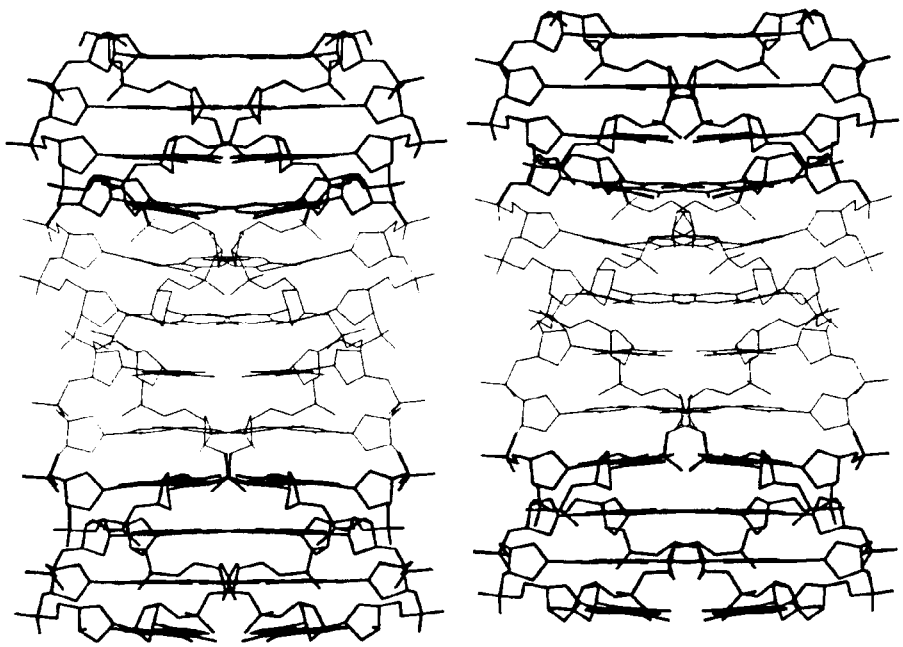

Figure 3. Stereo diagram of the parallel tetraplex structure of $d(G_4T_4G_4)$ minimized with reduced charge. The guanine stretches have been drawn in thick lines while the T_4 stretch is shown in thin lines.

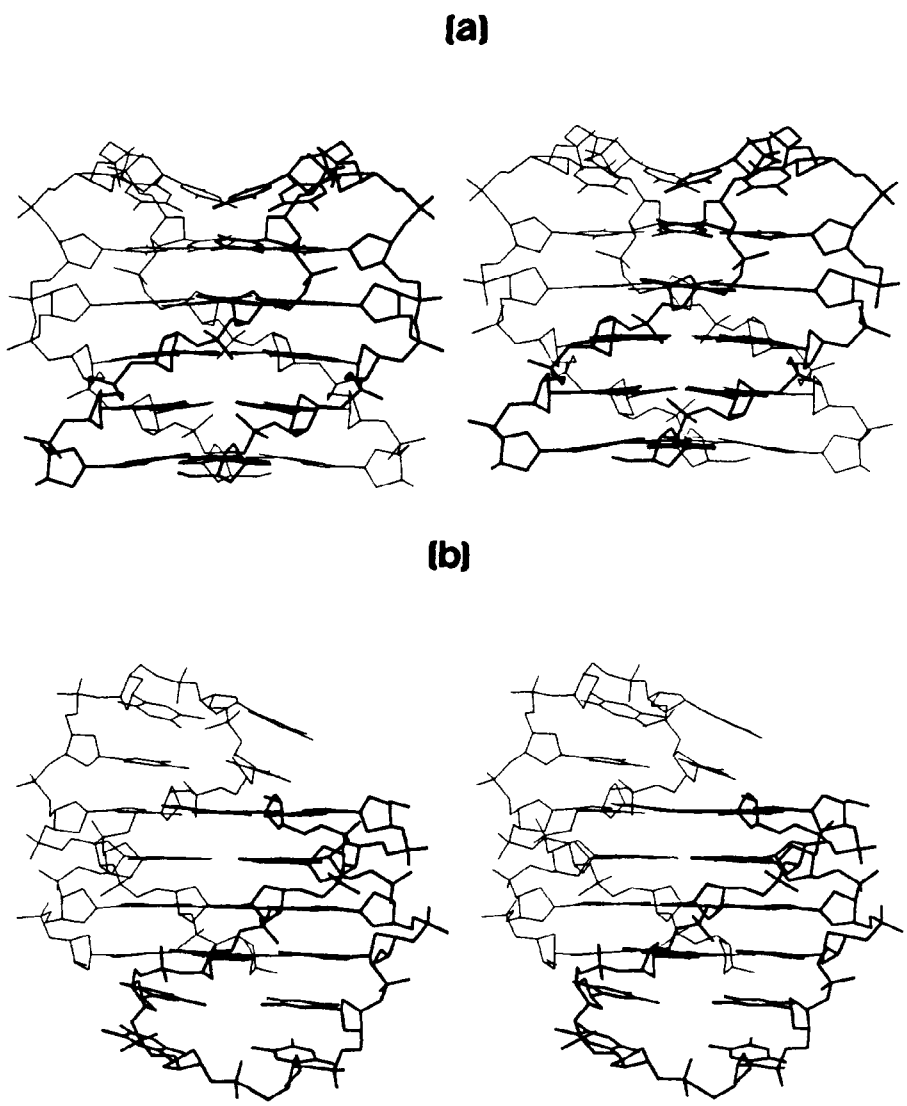

Figure 4. Stereo diagram of the hairpin dimer structures of d($G_4T_4G_4$) with the 5′ to 3′ direction of the T_4 loop joining a guanine in *anti* conformation to a guanine in *syn* conformation across the smaller groove and minimized with reduced charge. (a) Loops on same side; (b) loops on opposite sides. In this figure one hairpin has been drawn in thick lines, while the other has been drawn in thin lines.

for thymines, stacking of thymines over the G tetrads have been reported from nmr studies.[30,59,60] In the thymine tetrads, the methyl carbon atom of each thymine is at a distance of 3.0 Å from O2 of the thymine in the adjacent strand, thus giving rise to the additional possibility of C—H---O hydrogen bonds.[61] The parallel tetraplex structures minimized with both normal and reduced charge have very similar structural features.

In the energy-minimized structures of hairpin dimers (Figure 4a and b), the stem regions are very similar to the minimized structures of antiparallel G tetraplexes without the loops.[26] The G tetrads in the stem regions of the energy-minimized hairpin dimers remain intact and the cation binding site is preserved, even though the structures have been energy minimized without any constraints. The structures, when minimized with reduced charge on phosphate oxygens, have large and small grooves similar to the starting structures, but when mini-

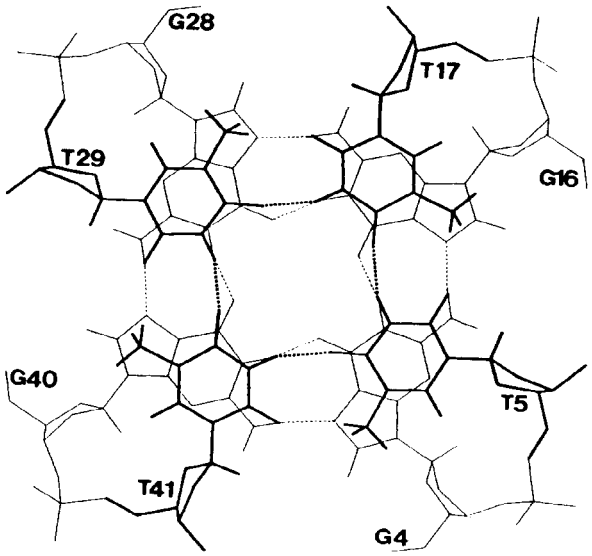

Figure 5. Line diagram showing thymine tetrad (thick lines), formed by hydrogen bonding of O4 of each thymine with N3 of adjacent thymine, stacked over the guanine tetrad (thin lines) in the GT step of the energy-minimized parallel tetraplex structure for d($G_4T_4G_4$). The hydrogen bonds are shown as dotted lines. The nucleotides in the structure are numbered sequentially in the 5′ to 3′ direction as 1–12, 13–24, 25–36 and 37–48 in the first, second, third, and fourth strand, respectively.

mized with normal charge the small groove widens, while the width of the large groove remains similar to that in the starting structure. In contrast in the parallel tetraplex structure, which has four identical grooves, the groove widths remain the same in both sets of minimizations. This feature is also observed for the tetraplexes involving guanine alone.[26]

The stacking and hydrogen bonding in various hairpin dimer structures with loops on the same side, over the small groove and minimized with reduced charge, are shown in Figure 6(a–d). Figure 7(a and b) show the stacking of the first and the fourth thymines over the G tetrad, in the A and B chains (coordinates were obtained from the Nucleic Acid Database[49]) of the energy-minimized (with reduced charge) crystal structure of d($G_4T_4G_4$) in which the loops are on the opposite ends of the G stem. The structures minimized with normal charge on phosphate oxygens also have very similar stacking and hydrogen-bonding features. In all the T_n loop structures the thymines are also well stacked, even though they have larger buckle- and propeller-type distortions compared to G tetrads (Figure 6). In the structures with loops on opposite sides, the thymines remain stacked even in the absence of any hydrogen bonding in the loop. The H–Y distance and X—H---Y angle values for various hydrogen bonds in the loop regions of the energy-minimized structures (with reduced charge) are listed in Table II. The residues in a hairpin dimer with N nucleotides in each hairpin have been number sequentially in the 5′ to 3′ direction as 1 to N in one hairpin and $N + 1$ to $2N$ in the other hairpin. For example, the residues in the d($G_4T_4G_4$) dimer are identified as G1-G2-G3-G4-T5-T6-T7-T8-G9-G10-G11-G12 in the first hairpin and G13-G14-G15-G16-T17-T18-T19-T20-G21-G22-G23-G24 in the second hairpin of the dimer. As can be seen from Table II, for the T_n loops all the hydrogen bonds present in the starting model structures are retained in the energy-minimized structures, except for the two hydrogen bonds in the T_4 loop structure involving N3(T20)---O2(T5) and N3(T8)---O2(T17), which have H---Y distance of 3.5 Å. However, the N3 of thymines T20 and T8 form new hydrogen bonds with the O4 of T6 and T18 respectively as acceptors (also listed in Table II), which were not present in the starting structure. Thus the O4 of thymines T6 and T18 are now involved in bifurcated hydrogen bonding. It is interesting to note that, in the symmetric hairpin dimers with n nucleotides in the loop and both loops on same side (Figure 6a–d), since the 5′ to 3′ strand directions in the two loops are antiparallel, the nth residue in one hairpin is hydrogen bonded to the first residue in the loop of the other hairpin.

The various possible thymine–thymine pairing considered here are representative ones consistent with the dyad rotation between the two hairpins and the *anti* conformation of the thymines in the hairpin loops. It is interesting to note that the structure of DNA aptamer d($G_2T_2G_2TGTG_2T_2G_2$) has an interloop pairing[34,35] essentially similar to the T-T pairing proposed by us earlier for the hairpin dimer d($G_4T_2G_4$).[12] Thus other possible hydrogen bonds discussed above could also be present in the structures of the T_3 and T_4 loops, but this can only be confirmed by x-ray and nmr studies on these sequences. It may also be possible to have other types of intra- or interloop T-T pairing in the asymmetric or folded-back type of structures. Similarly, in the human telomeric sequence d($TTAG_3$)$_4$, which is expected to adopt folded-back structures with a G_3 stem and TTA loops, a very large number of structures are possible with different intra- and interloop pairing,[62] since adenine can be involved in Hoogsteen-type base pairing using its N7 edge, in addition to the Watson–Crick pairing using N1 as an acceptor.

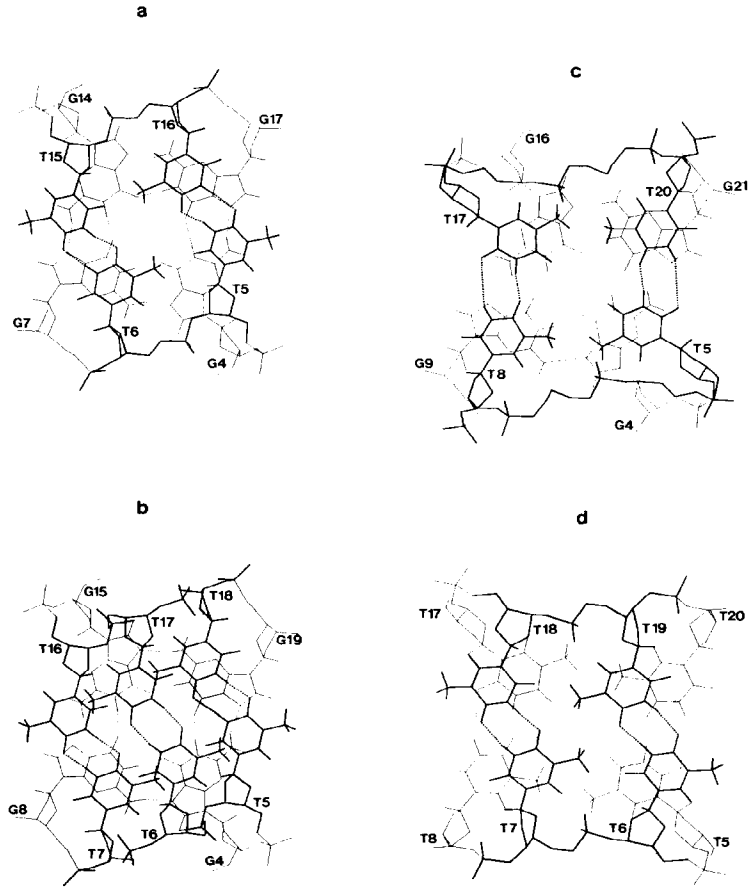

Figure 6. Stacking and hydrogen-bonding patterns in the hairpin dimer structures with two loops on same side and joining a guanine in *anti* conformation to a guanine in *syn* conformation across the smaller groove. In the hairpin dimers with N nucleotides forming the hairpin, the nucleotides in one chain are numbered 1 to N in the 5′ to 3′ direction, while they are numbered as $N + 1$ to $2N$ in the other chain. (a) $d(G_4T_2G_4)$: The thymine tetrad (in thick lines), formed by T5 and T6 in one hairpin forming a hydrogen bond with T16 and T15 respectively in the other hairpin, is stacked over the guanine tetrad (shown in thin lines). (b) $d(G_4T_3G_4)$: The tetrad formed by T5, T7, T16, and T18 has hydrogen bonds as listed in Table II and is stacked over the guanine tetrad (shown in thin lines), while the TT pair involving T6 and T17 is stacked over the thymine tetrad. (c) $d(G_4T_4G_4)$: The stacking diagram of the thymine tetrad formed by T5, T8, T17, and T20 (thick lines) over the guanine tetrad (thin lines). In this thymine tetrad T5 forms a pair of hydrogen bonds with T20, while T8 forms another pair of hydrogen bonds with T17. In the figure only backbone atoms have been shown for T6, T7, T18, and T19, which form the outermost thymine tetrad. (d) $d(G_4T_4G_4)$: Stacking pattern of the tetrad formed by T6, T7, T18, and T19 (drawn in thick lines) over the first tetrad (drawn in thin lines). In the outermost thymine tetrad T6 forms two hydrogen bonds with T19, while T7 forms two hydrogen bonds with T18.

Backbone and Glycosidic Torsion Angles. Backbone and glycosidic torsion angles for nucleotides in the loop and the two guanines flanking the loop are shown in Figure 8 for structures minimized using reduced charge on phosphate oxygens. As observed in our earlier energy minimization studies on antiparallel G tetraplexes without the loop, the backbone torsion angles α, β, γ, δ, ϵ, and ζ for the guanine in *syn* conformation are in the range g^-, g^-, g^-, t (^2E), t, and g^-, while for the guanines in *anti* conformation

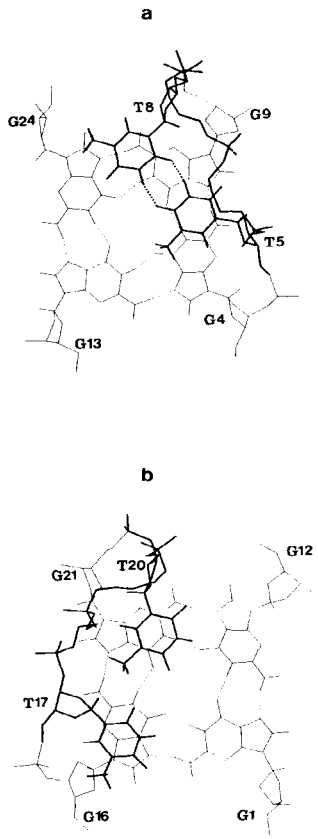

Figure 7. Stacking of first and fourth thymines of the loop (thick lines) over the G tetrads (thin lines) in the A and B chains of the crystal structure of d($G_4T_4G_4$) (Ref. 48) after energy minimization. In this structure the loops connect a guanine in *anti* conformation to a guanine in *syn* conformation across the large groove, and only backbone atoms are shown for the thymines T6, T7, T18, and T19. (a) T5 and T8 in the A chain stacked over the G tetrad formed by G4, G9, G13, and G24. (b) T17 and T20 of B chain stacked over the G tetrad formed by G16, G21, G1, and G12. Since the two loops are on the opposite side of the G-tetraplex, while loop T5 to T8 is above the G tetrad in (a), the loop T17 to T20 in (b) is below the G tetrad.

they are in g^-, t, g^+, t ($_1E$), t, and g^- conformation. In all the structures with T_n loops, the guanines that are one nucleotide away from the loop have all the torsion angles (Figure 8a and h) identical to those in the antiparallel G tetraplex without the loops. Even in the guanines, immediately flanking the loop at the 5′ and the 3′ end, only a few torsion angles have changed from their values in the tetraplex stem. In the various hairpin structures, the changes in torsion angles (Figure 8) involve α, β, γ, ϵ, and ζ of the guanine at the 5′ end of the loop, and α, β, and γ of the guanine at the 3′ end of the loop. However, for any given structure there is a change in only a few of these torsion angles. As can be seen from Figure 8b and g, while for T_2 loop structures with guanines in *syn* conformation connected to guanines in *anti* conformation, only α and β have changed in the guanine at the 5′ end of the loop, for *anti* to *syn* structure of the T_2 loop, changes have occurred in the α and β of the guanine at the 3′ end of the loop. However for T_3 and T_4 loop structures, the torsion angles have changed both in the guanines at the 5′ as well as 3′ end of the loop. In various structures of T_n loops, all the thymines have *anti* conformation about the glycosidic bond, while the sugar puckers show considerable variability, as seen from the range in the torsion angle δ (Figure 8).

Relation Between Glycosidic Torsion of Thymines and Their Stacking and Hydrogen-Bonding Patterns. As can be seen from Figure 6a, which has a T_2 loop over the smaller groove, the two thymines are stacked with their α face (numbering of atoms in the ring appears clockwise)[63] being visible from the top and the β face (numbering of atoms in the ring appears anticlockwise)[63] pointing toward the guanines. The two sugar rings are oriented with their oxygens (O4′) pointing almost in the same direction and have *anti* conformation about the glycosidic bond for both the thymines. Similarly, for the T_3 loop (Figure 6b) the thymines are stacked with their α face pointing up and the oxygens (O4′) of all the sugars are oriented in the same direction with *anti* conformation about the glycosidic bond. Due to the typical positioning of thymines in the loop, it is possible to have chain connectivity with sugars in G4 (or G15) and T5 (or T16) pointing away from each other. In the T_4 loop structure (Figure 6c and d), the thymines T6, T7, T18, and T19 in the top tetrad have stacking with the α face visible from the top, but in the lower tetrad, while T8 and T20 are stacked with the α face visible from the top, T5 and T17 have to be oriented with the β face pointing up, in order to have an *anti* conformation about the glycosidic bond. Since the positioning of the top tetrad restricts the possible sugar orientations of thymines in the lower tetrad, an orientation of T5 and T17 with the α face pointing up would result in a *syn* conformation. The orientation of T5/T17 with the β face on the top results in a hydrogen-bonding scheme in the lower tetrad of the T_4 loop, which is different from that in the top tetrad or the tetrads in T_2 and T_3 loop structures. In the top tetrad of the T_4 loop (Figure 6d) and also in the T_2 and T_3

Table II The List of Hydrogen Bonds Involving the Residues in the Loop Region of Different Energy-Minimized (Reduced-Charge) Hairpin Dimer Structures of $d(G_4T_nG_4)$, n = 2–4, with Loops on the Same Side and Across the Small Groove.[a]

$d(G_4T_2G_4)$		*anti* to *syn*		*syn* to *anti*	
Donor X	Acceptor Y	Distance H---Y	Angle X–H---Y	Distance H---Y	Angle X–H---Y
N3-T(16)	O4-T(5)	1.8	169.1	1.9	158.6
N3-T(5)	O4-T(16)	1.8	147.0	1.8	149.9
N3-T(15)	O4-T(6)	1.8	147.0	1.8	149.9
N3-T(6)	O4-T(15)	1.8	169.1	1.9	158.6
$d(G_4T_3G_4)$		*anti* to *syn*		*syn* to *anti*	
Donor X	Acceptor Y	Distance H---Y	Angle X–H---Y	Distance H---Y	Angle X–H---Y
N3-T(18)	O4-T(5)	1.8	174.3	1.8	154.2
N3-T(5)	O4-T(18)	1.8	170.7	1.8	164.8
N3-T(17)	O4-T(6)	1.8	169.5	1.8	160.1
N3-T(6)	O4-T(17)	1.8	169.5	1.8	160.1
N3-T(16)	O4-T(7)	1.8	170.7	1.8	164.8
N3-T(7)	O4-T(16)	1.8	174.4	1.8	154.2
$d(G_4T_4G_4)$		*anti* to *syn*		*syn* to *anti*	
Donor X	Acceptor Y	Distance H---Y	Angle X–H---Y	Distance H---Y	Angle X–H---Y
N3-T(5)	O4-T(20)	1.8	167.6	1.8	161.8
N3-T(20)	O2-T(5)	3.5	138.0	1.8	164.0
N3-T(19)	O4-T(6)	2.0	169.2	1.8	162.0
N3-T(6)	O4-T(19)	1.8	165.7	1.8	144.9
N3-T(18)	O4-T(7)	1.8	165.7	1.8	144.9
N3-T(7)	O4-T(18)	2.0	169.3	1.9	162.1
N3-T(17)	O4-T(8)	1.8	167.6	1.8	161.8
N3-T(8)	O2-T(17)	3.5	138.0	1.8	164.0
N3-T(8)	O4-T(18)	2.0	130.8		
N3-T(20)	O4-T(6)	2.0	130.8		

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

loop (Figure 6 a and b), the thymines are stacked with their methyl groups pointing toward the 5′ arm of the hairpin, while in the lower tetrad of the T_4 loop (Figure 6c), the two thymines in each hairpin are oriented with their methyl groups pointing toward each other. It is interesting to note that the stereochemical requirement of *anti* conformation for thymines would result in structures having thymines stacked with their β face pointing up for T_2 and T_3 loop across the larger groove. Similarly, for the T_4 structure with loop across the larger groove, the thymines in the top tetrad will have their β face visible from the top, and in the lower tetrad T5 and T17 will have the β face pointing up while in T8 and T20 it will point down, thus resulting in an intraloop T-T pairing in the lower tetrad unlike the interloop T-T pairing of the T_4 loop across the smaller groove (Figure 6c). In fact in the crystal structure of $d(G_4T_4G_4)$, which has loops across the larger groove and at opposite ends of the G tetraplex, in one hair-

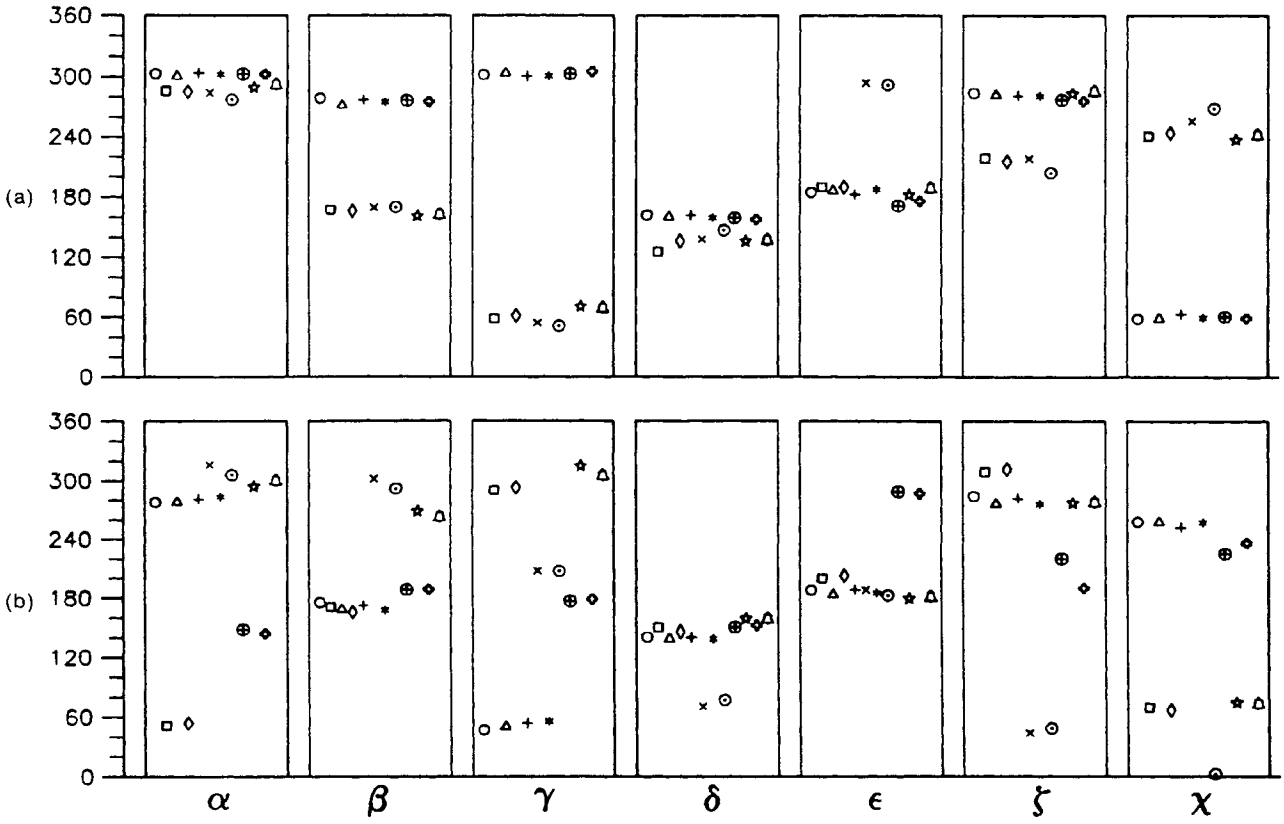

Figure 8. Backbone and glycosidic torsion angles in hairpin dimer structures of $d(G_4T_nG_4)$, n = 2–4, with loops across the smaller groove and minimized with reduced charge on phosphate oxygens. The torsion angles in hairpin dimers with the 5′ to 3′ chain direction of the loop joining a guanine in *anti* to a guanine in *syn* conformation and with two loops on same and opposite side of the G stem are represented by symbols; ○ and △ for $d(G_4T_2G_4)$, + and * for $d(G_4T_3G_4)$, ⊕ and ⊞ for $d(G_4T_4G_4)$, respectively. The *syn* to *anti* loops for $d(G_4T_2G_4)$, $d(G_4T_3G_4)$, and $d(G_4T_4G_4)$ are represented by □, ×, and ☆, respectively, when the loops are on same side, and by ◇, ⊙, and ⏃ when the loops are on opposite sides. Torsion angles in the two hairpins constituting each of the dimers are almost identical, and the average value over the two hairpins in the dimer have been plotted for the thymines in the loop region and two guanines flanking the loop in the 5′ as well as 3′ directions. (a) G3 of $d(G_4T_2G_4)$, $d(G_4T_3G_4)$, and $d(G_4T_4G_4)$. (b) G4 of $d(G_4T_2G_4)$, $d(G_4T_3G_4)$, and $d(G_4T_4G_4)$. (c) T5 of $d(G_4T_2G_4)$, $d(G_4T_3G_4)$, and $d(G_4T_4G_4)$. (d) T6 of $d(G_4T_3T_4)$ and $d(G_4T_4G_4)$. (e) T7 of d $(G_4T_4G_4)$. (f) T6 of $d(G_4T_2G_4)$, T7 of $d(G_4T_3G_4)$, and T8 of $d(G_4T_4G_4)$. (g) G7 of $d(G_4T_2G_4)$, G8 of $d(G_4T_3G_4)$, and G9 of $d(G_4T_4G_4)$. (h) G8 of $d(G_4T_2G_4)$, G9 of $d(G_4T_3G_4)$, and G10 of $d(G_4T_4G_4)$.

pin the first thymine in the loop is stacked with its β face pointing up while the fourth thymine is stacked with its α face pointing up. It is found that though there is no hydrogen bonding between these two thymines in the crystal structure, energy minimization of the structure results in an intraloop T-T pair involving N3—H---O4 and N3—H---O2 [Figure 7(a)]. However, in the other hairpin the fourth thymine (T20) is stacked with its α face pointing toward the guanine tetrad [Figure 7(b)], thus resulting in a *syn* conformation, but both the thymines have orientations essentially similar to that in the thymine tetrad of the T_3 loop structures discussed earlier.

Energy Component Analysis

Orientation of Thymine Loops in the Dimers of $d(G_4T_nG_4)$, n = 2–4. A comparison of the energy values of the hairpin dimers indicates that the energy of the stem region remains very similar in the various

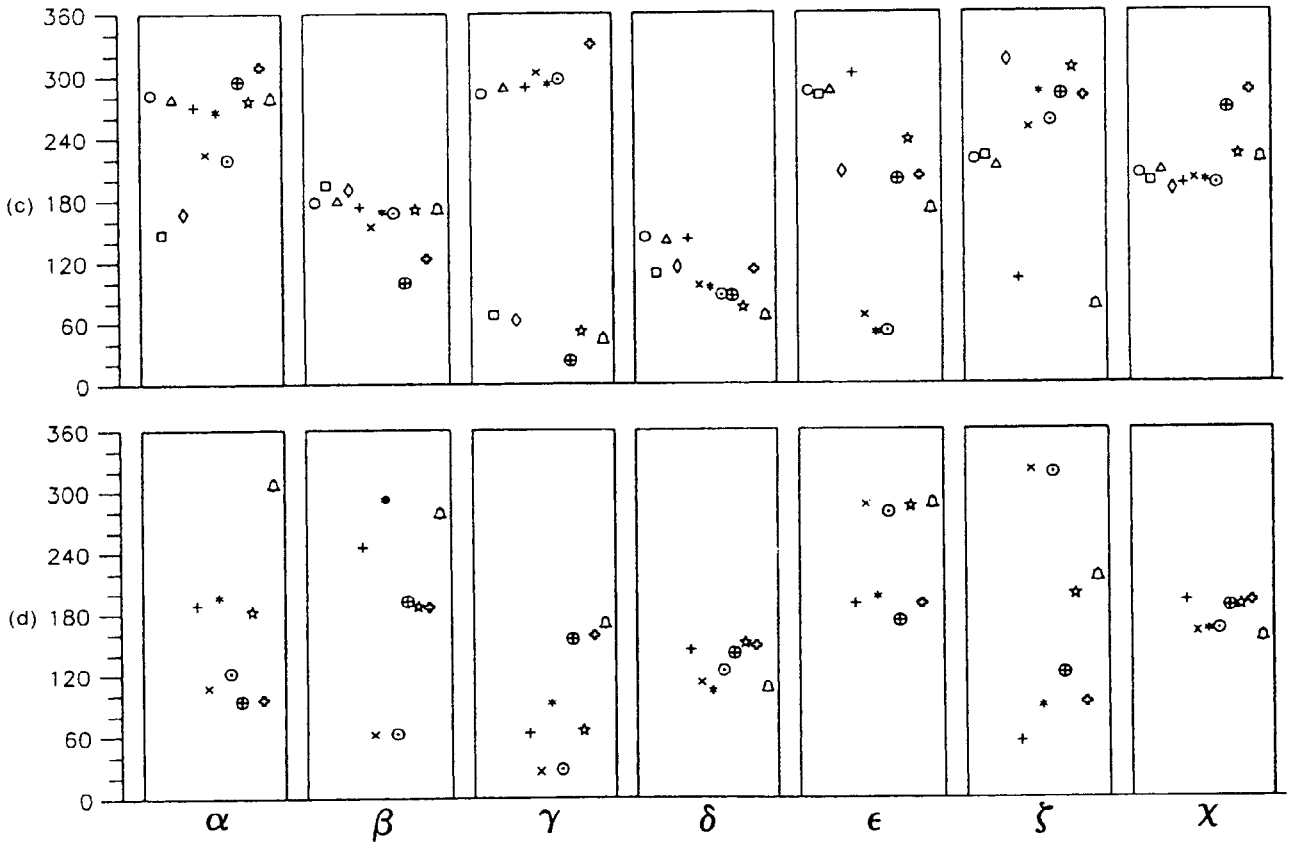

Figure 8. (*Continued from the previous page.*)

structures with different possible orientations of the loop, i.e., same side, opposite side, and diagonal. As can be seen from Table IIIa and b, the stem energy is between −770 and −780 kcal/mole for structures with an *anti* to *syn* loop and between −740 and −750 kcal/mole for those with an *syn* to *anti* loop. This difference in stem energy is expected since in the G tetraplex, the more the number of *syn-anti* steps the better is the energy, a feature observed in our previous calculations.[26] When the loops are on the same side of the G tetraplex and form interloop hydrogen bonds, the energy contribution from the loops is better because of favorable electrostatic component. However, the loop stem interaction energy is more favorable when the loops are on opposite ends. Thus for the T_n hairpin dimers built with the constraint of stacking and hydrogen bonding, the structures with loops on the same side (Table IIIa) have energy values similar to the corresponding structures with loops on opposite side (Table IIIb) irrespective of whether the loop is from *syn* to *anti* or *anti* to *syn*. This result is also true for structures minimized with normal charge on phosphate oxygens, even though the various energy components differ in magnitude and the loop stem interaction even differs in sign (Table IVa and b). Among the diagonal loop structures (nmr type) considered for minimization, while the T_3 loop structure has energy very similar to the corresponding adjacent loop structures, because of favorable van der Waal's interaction between loop and stem, the T_4 diagonal loop has less favorable loop–stem interaction and hence a slightly unfavorable total energy. However, since a large number of loop conformations can be generated, it may be possible to have T_4 diagonal loop structures with energy values comparable to adjacent loop structures. As can be seen from Tables IIIb and IVb, the energy-minimized crystal structure of the T_4 loop also has energy values similar to the structures obtained from our model building.

The detailed molecular mechanics energy calculations carried out by us indicate that for hairpin dimers the three different modes of dimerization lead to structures with very similar energy. In fact, the energy contribution from the G-tetraplex stem, which is the major stabilizing factor in these struc-

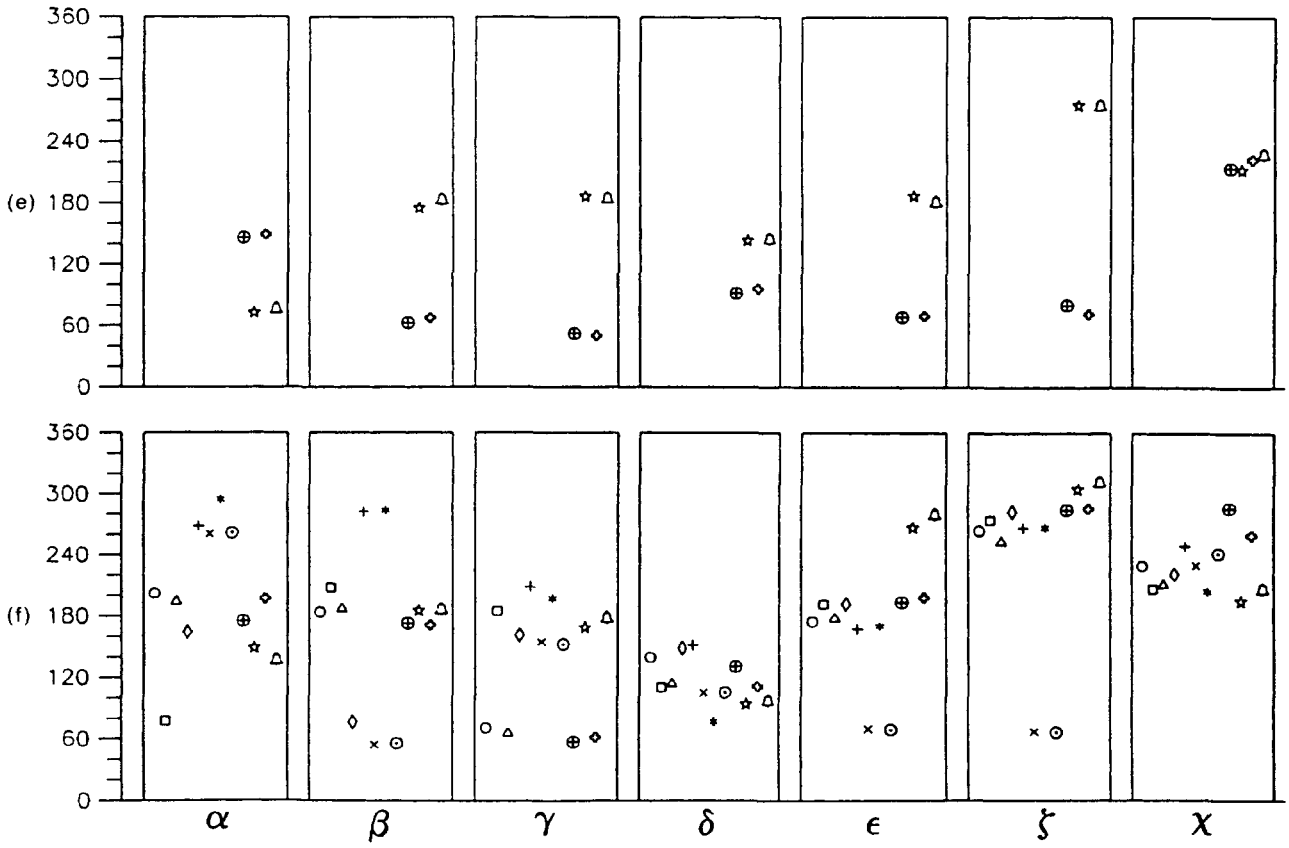

Figure 8. (*Continued from the previous page.*)

tures, is almost identical in all these structures. The various structures differ only in the contribution from the loop region, which in any case is very small as compared to the contribution from the stem region. While structures with loops on same side have favorable interloop interaction, structures with loops on opposite end can have an equally favorable contribution from loop–stem interaction. Therefore, though only hairpin dimer structures with loops on opposite ends have been observed in x-ray and nmr studies,[16,17] structures with loops on the same side could also be a possible class of telomeric structures. Such structures have the possibility of interloop hydrogen bonding, and for folded-back structures of four repeat sequences, two loops necessarily have to be on the same side.

Effect of Number of Thymines on the Loop Energy.

The total energy for the hairpin dimer structures, minimized with reduced charge, shows an improvement (Table IIIa and b) as the number of thymines in the loop increases from 2 to 4, the contribution arising mainly from energy of the loop. This trend is valid both for structures with loops on the same as well as opposite sides and for loops going from *syn* to *anti* and *anti* to *syn*. In fact, the total energy improves by 30–40 kcal with each additional thymine.

However, when the structures are minimized with normal charge, the loop–stem interaction energy is positive, and it becomes unfavorable with an increasing number of thymines in the loop (Table IVa and b), even though the energy of the loop becomes more favorable with an increasing number of thymines. Thus in the overall contribution from the loop, though the T_4 loop has better energy than the T_3 and T_2 loops, the T_3 loop is unfavorable compared to T_2. In addition, there is a significantly larger fluctuation in the energy of the tetraplex stem in these structures. A single point energy calculation for these structures using a reduced charge on phosphate groups also gives favorable energy of $\approx$ 40 kcal/mole for each additional thymine in the loop. Thus the energy contribution from the loop depends not only on the structure of the loop but also on the charge assigned to the phosphate groups.

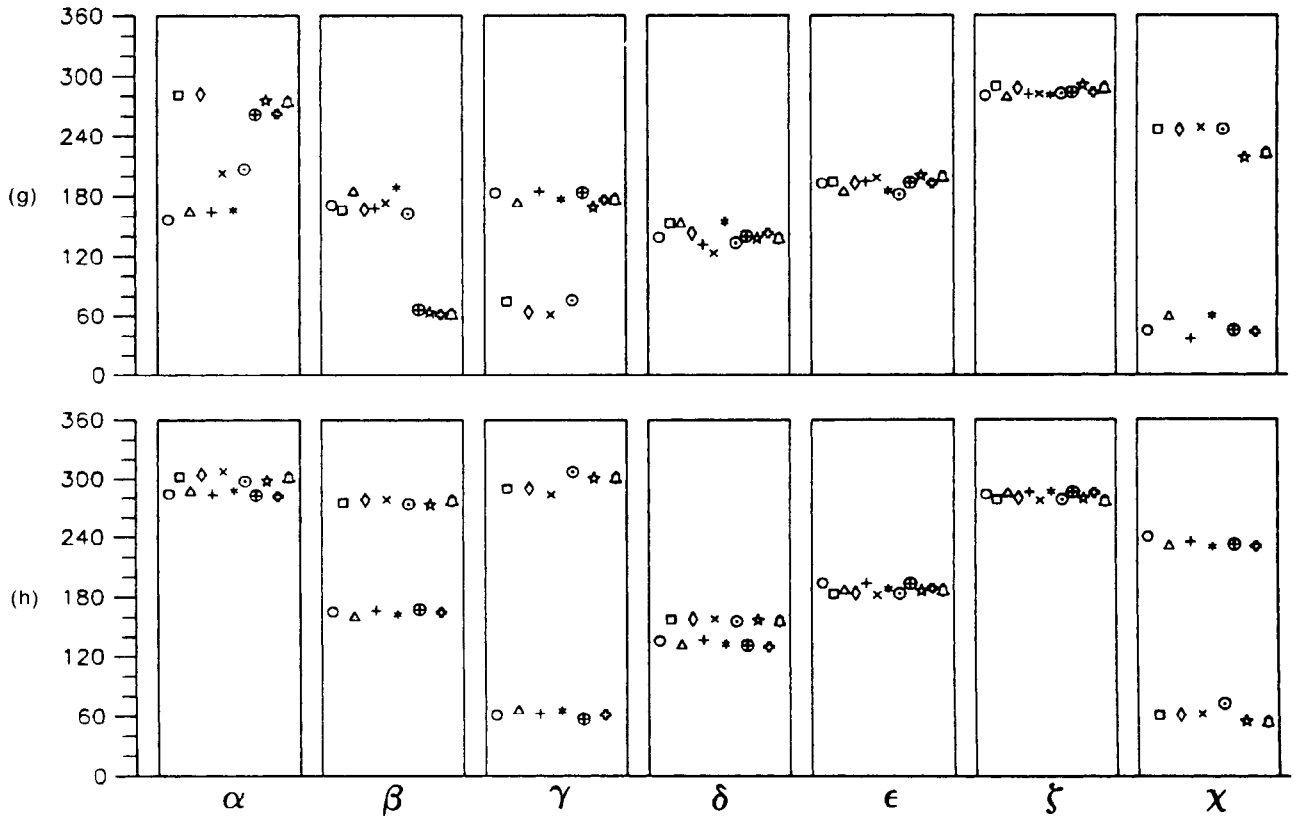

Figure 8. (*Continued from the previous page.*)

The dependence of the stability of various hairpin dimer structures on loop size has been examined in several experimental studies.[12,14] While for sequences of the type $d(G_4T_nG_4)$ the stability increases as the number of thymines increases[12] from 2 to 4, the trend is reversed for sequences of the type $d(T_nG_4T_nG_4)$. Our calculations, carried out on both types of sequences, indicate a higher stability for sequences with larger number of nucleotides in the loop. However, the energy values obtained from molecular mechanics calculations only give the enthalpy contribution, and the observed difference between the two types of structures could probably arise from entropic contribution.

Stability of Folded-Back Structures Compared to Parallel Tetraplexes.

A comparison of the conformational energy of the parallel tetraplex with corresponding hairpin dimers (Table Va) indicates that the equivalent energy of the hairpin dimer structures (corresponding to twice the energy of the hairpin dimer, which contains only two chains, while the parallel tetraplex has four) is more favorable by at least 100 kcal/mole for all the three sequences. Such an observation is really surprising considering the fact that for an all-G stretch a parallel alignment of strands is more favorable compared to an *anti* parallel arrangement. However, an analysis of the detailed breakup of different energy components indicates that though the parallel tetraplex has less distortion and more favorable van der Waal's energy components compared to hairpin dimers, the electrostatic component makes the hairpin dimers more favorable. The electrostatic contribution from base and sugar atoms are very similar in both types of structures (Table Va), but the contribution from phosphate groups are highly unfavorable for the parallel tetraplex. This unfavorable electrostatic interaction in fact arises from the large repulsion between phosphates in the thymine and guanine stacks in the parallel tetraplex, as can be seen from Table VIa, where contributions from guanine and thymine stacks have been listed separately. In $d(G_4T_2G_4)$, while the T_2 stack has an intrastack energy of −160 kcal/mole, its inter-

Table III Energy Contributions (kcal/mole) from Loop and Stem Regions of the Hairpin Dimer Stuctures for d($G_4T_nG_4$), n = 2–4, Minimized with Reduced Charge on Anionic Phosphate Oxygens.[a]

(a) Loops on Same Side

	Intrastem (a)	Intraloop (b)	Interloop-stem (c)	Total from Loop ($b + c$)	Total ($a + b + c$)
anti to *syn*			Loop Across Smaller Groove with T-T Pairing		
T_2	**−773**	**−72**	**−60**	**−132**	**−905**
d	227	62	7	69	296
v	−222	−2	−58	−60	−282
e	−778	−132	−9	−141	−919
T_3	**−771**	**−110**	**−62**	**−172**	**−943**
d	228	99	7	106	334
v	−222	−14	−63	−77	−299
e	−777	−195	−6	−201	−978
T_4	**−772**	**−147**	**−66**	**−213**	**−985**
d	227	140	8	148	375
v	−222	−37	−69	−106	−328
e	−777	−250	−5	−255	−1032
syn to *anti*			Loop Across Smaller Groove with T-T Pairing		
T_2	**−752**	**−76**	**−61**	**−137**	**−889**
d	235	60	8	68	303
v	−220	−3	−61	−64	−284
e	−767	−133	−8	−141	−908
T_3	**−738**	**−103**	**−60**	**−163**	**−901**
d	245	115	9	124	369
v	−220	−19	−65	−84	−304
e	−763	−199	−4	−203	−966
T_4	**−755**	**−198**	**−68**	**−266**	**−1021**
d	238	119	5	124	362
v	−223	−44	−67	−111	−334
e	−770	−273	−6	−279	−1049

(b) Loops on Opposite Side

	Intrastem (a)	Intraloop (b)	Interloop-stem (c)	Total from Loop ($b + c$)	Total ($a + b + c$)
anti to *syn*			Loop Across Smaller Groove		
T_2	**−777**	**−52**	**−88**	**−140**	**−917**
d	228	60	7	67	295
v	−225	4	−70	−66	−291
e	−780	−116	−25	−141	−921
T_3	**−775**	**−82**	**−85**	**−167**	**−942**
d	225	98	7	105	330
v	−223	−11	−73	−84	−307
e	−777	−169	−19	−188	−965
T_4	**−768**	**−114**	**−85**	**−199**	**−967**
d	228	135	8	143	371
v	−224	−22	−74	−96	−320
e	−772	−227	−19	−246	−1018

Table III (*Continued*)

	(b) Loops on Opposite Side				
	Intrastem (*a*)	Intraloop (*b*)	Interloop-stem (*c*)	Total from Loop (*b* + *c*)	Total (*a* + *b* + *c*)
syn to *anti*			Loop Across Smaller Groove		
T_2	**−751**	**−51**	**−65**	**−116**	**−867**
d	236	59	9	68	304
v	−217	5	−63	−58	−275
e	−770	−115	−11	−126	−896
T_3	**−741**	**−75**	**−76**	**−151**	**−892**
d	248	116	10	126	374
v	−215	−13	−82	−95	−310
e	−774	−178	−4	−182	−956
T_4	**−758**	**−131**	**−63**	**−194**	**−952**
d	234	124	5	129	363
v	−221	−24	−61	−85	−306
e	−771	−231	−7	−238	−1009
anti to *syn*			Loop Across Diagonal Strands (nmr Type)		
T_3	**−780**	**−94**	**−85**	**−179**	**−959**
d	224	83	8	91	315
v	−218	6	−73	−67	−285
e	−786	−183	−20	−203	−989
T_4	**−780**	**−116**	**−45**	**−161**	**−941**
d	228	131	8	139	367
v	−221	−7	−46	−53	−274
e	−787	−240	−7	−247	−1034
anti to *syn*			Energy Minimized Crystal Structure of $d(G_4T_4G_4)$		
T_4	**−738**	**−153**	**−78**	**−231**	**−969**
d	258	124	7	131	389
v	−222	−34	−77	−111	−333
e	−774	−243	−8	−251	−1025

[a] The intraloop component includes contribution from both the loops together. The rows indicated by *d*, *v*, and *e* give distortion (sum of bond length, angle and dihedral), van der Waals (sum of 6–12 and 10–12), and electrostatic energy, respectively, while the first row for each sequence gives the total energy.

action energy with the guanine stacks flanking it is +159 kcal/mole, which is mostly due to the large electrostatic repulsion between the phosphates of guanine and thymine stacks. Addition of thymines in T_3 and T_4 structures results in increased unfavorable contributions (Table VIa). In contrast, when these sequences adopt hairpin dimer structures, the interaction energy (Table IVa and b) of the thymine loop with guanine stem, while still positive, is much less unfavorable when compared to the parallel tetraplex, and the thymine stretches make a net contribution of −30 to −60 kcal/mole in the folded-back hairpin structures. Thus the presence of thymine stretches destabilizes the parallel tetraplex structures while they seem to have a stabilizing effect in the hairpin dimer structures. It is important to note that such phosphate–phosphate repulsion is also present in an all-G parallel tetraplex, but the favorable contributions arising from the stacking and base pairing in the guanine tetrads compensates for it, while the contribution from the thymine tetrads is unable to do so. It can be seen from Table VIa that, while a G_4 stack has a contribution of about −650 kcal/mole in a parallel tetraplex, the contribution from T_4 stack is only about −280 kcal/mole. In addition, there is an unfavorable interaction energy component of +290 kcal/mole between the G_4 and T_4 stretch. On the other hand,

Table IV Energy Contributions (kcal/mole) from Loop and Stem Regions of the Hairpin Dimer Stuctures for d($G_4T_nG_4$), n = 2–4, Minimized with Normal Charge on Anionic Phosphate Oxygens.[a]

(a) Loops on Same Side

	Intrastem (a)	Intraloop (b)	Interloop-stem (c)	Total from Loop ($b + c$)	Total ($a + b + c$)
anti to *syn*		Loop Across Smaller Groove with T-T Pairing			
T_2	**−679**	**−63**	**17**	**−46**	**−725**
d	230	65	8	73	303
v	−214	2	−57	−55	−269
e	−695	−130	66	−64	−759
T_3	**−670**	**−82**	**50**	**−32**	**−702**
d	234	112	7	119	353
v	−212	−16	−54	−70	−282
e	−692	−178	97	−81	−773
T_4	**−635**	**−126**	**64**	**−62**	**−697**
d	242	139	9	148	390
v	−215	−43	−63	−106	−321
e	−662	−222	118	−104	−766
syn to *anti*		Loop Across Smaller Groove with T-T Pairing			
T_2	**−630**	**−72**	**8**	**−64**	**−694**
d	241	57	7	64	305
v	−206	0	−57	−57	−263
e	−665	−129	58	−71	−736
T_3	**−611**	**−90**	**51**	**−39**	**−650**
d	255	112	9	121	376
v	−214	−22	−58	−80	−294
e	−652	−180	100	−80	−732
T_4	**−635**	**−158**	**57**	**−101**	**−736**
d	238	124	5	129	367
v	−216	−39	−63	−102	−318
e	−657	−243	115	−128	−785

(b) Loops on Opposite Side

	Intrastem (a)	Intraloop (b)	Interloop-stem (c)	Total from Loop ($b + c$)	Total ($a + b + c$)
anti to *syn*		Loop Across Smaller Groove			
T_2	**−678**	**−52**	**8**	**−44**	**−722**
d	233	61	9	70	303
v	−213	4	−69	−65	−278
e	−698	−117	68	−49	−747
T_3	**−686**	**−71**	**29**	**−42**	**−728**
d	227	102	8	110	337
v	−213	−4	−63	−67	−280
e	−700	−169	84	−85	−785
T_4	**−670**	**−108**	**42**	**−66**	**−736**
d	234	131	10	141	375
v	−211	−24	−76	−100	−311
e	−693	−215	108	−107	−800

Table IV (*Continued*)

	Intrastem (*a*)	Intraloop (*b*)	Interloop-stem (*c*)	Total from Loop (*b* + *c*)	Total (*a* + *b* + *c*)
			(b) Loops on Opposite Side		
syn to *anti*			Loop Across Smaller Groove		
T_2	**−638**	**−38**	**18**	**−20**	**−658**
d	244	71	6	77	321
v	−214	8	−58	−50	−264
e	−668	−117	70	−47	−715
T_3	**−623**	**−52**	**44**	**−8**	**−631**
d	259	114	7	121	380
v	−207	−4	−65	−69	−276
e	−675	−162	102	−60	−735
T_4	**−628**	**−120**	**52**	**−68**	**−696**
d	243	118	4	122	365
v	−217	−21	−68	−89	−306
e	−654	−217	116	−101	−755
anti to *syn*			Loop Across Diagonal Strands (nmr Type)		
T_3	**−678**	**−78**	**60**	**−18**	**−696**
d	225	91	8	99	324
v	−217	9	−72	−63	−280
e	−686	−178	124	−54	−740
T_4	**−679**	**−112**	**93**	**−19**	**−698**
d	230	129	8	137	367
v	−212	−13	−43	−56	−268
e	−697	−228	128	−100	−797
anti to *syn*			Energy Minimized Crystal Structure of $d(G_4T_4G_4)$		
T_4	**−638**	**−129**	**55**	**−74**	**−712**
d	261	129	9	138	399
v	−216	−28	−68	−96	−312
e	−683	−230	114	−116	−799

[a] The intraloop component includes contribution from both the loops together. The symbol convention is the same as in Table III.

when two chains of the same sequence $d(G_4T_4G_4)$ adopt a hairpin dimer structure, the thymine stretch forming the loop has an energy of −126 kcal/mole and an interaction of +64 kcal/mole with the guanine stretch, and thus the presence of T_4 in a loop conformation contributes a net energy of −62 kcal/mole to the structure and this explains the observed propensity of $d(G_4T_nG_4)$ sequences for folded-back structures.[12,14] A sequence of the type $d(T_nG_4)_4$ or $d[G_4(T_nG_4)_3]$, with three or four stretches of thymines, if it adopts a parallel tetraplex will have much larger phosphate–phosphate repulsion than in $d(G_4T_nG_4)$ with a single thymine stretch. Thus a longer repeat sequence would have a much higher propensity for folded-back structure.

When the energy minimizations are carried out with a reduced charge on anionic phosphate oxygens, the total electrostatic energy becomes very similar in parallel as well as folded-back structures, as can be seen from Table Vb. The better nonbonded interaction and less distortion energy now make the total energy of the parallel tetraplex 100 to 200 kcal/mole more favorable when compared to the hairpin dimers. However, the electrostatic interaction between the G stretch and T stretch remains better in the hairpin dimers (Table IIIa and b) compared to parallel tetraplexes (Table VIb), even though total electrostatic energy is similar for both the structures.

Thus the results from the two sets of minimizations indicate the two different ways in which thymine stretches can be stabilized in telomeric se-

Table V Energy Components[a] (kcal/mole) for Hairpin Dimer and Parallel Tetraplex Structures of $d(G_4T_nG_4)$, $n = 2–4$[b]

(a) Normal Charge

	Adjacent Loop		Diagonal Loop,	
	Same Side	Opposite Side	Opposite Side	Parallel Tetraplex
$d(G_4T_2G_4)$				
Bond + angle + dihedral	298	298		560
Nonbonded van der Waals	−259	−270		−590
Electrostatic	−813	−802		−1388
Base	−776	−767		−1544
Phos	−37	−35		156
Hydrogen bond 10–12	−9	−8		−20
Total energy	**−783**	**−782**		**−1438**
$d(G_4T_3G_4)$				
Bond + angle + dihedral	350	338	320	620
Nonbonded van der Waals	−273	−273	−271	−634
Electrostatic	−832	−811	−808	−1406
Base	−825	−796	−796	−1630
Phos	−7	−15	−12	224
Hydrogen bond 10–12	−9	−8	−8	−24
Total energy	**−764**	**−754**	**−767**	**−1444**
$d(G_4T_4G_4)$				
Bond + angle + dihedral	386	370	362	680
Nonbonded van der Waals	−310	−300	−258	−668
Electrostatic	−855	−860	−863	−1444
Base	−843	−843	−831	−1730
Phos	−12	−17	−32	286
Hydrogen bond 10–12	−11	−9	−8	−24
Total energy	**−790**	**−799**	**−767**	**−1456**

(b) Reduced Charge

	Adjacent Loop		Diagonal Loop,	
	Same Side	Opposite Side	Opposite Side	Parallel Tetraplex
$d(G_4T_2G_4)$				
Bond + angle + dihedral	290	290		532
Nonbonded van der Waals	−272	−279		−621
Electrostatic	−934	−938		−1876
Base	−764	−760		−1518
Phos	−170	−178		−358
Hydrogen bond 10–12	−10	−10		−18
Total energy	**−926**	**−937**		**−1983**
$d(G_4T_3G_4)$				
Bond + angle + dihedral	330	324	310	596
Nonbonded van der Waals	−289	−296	−275	−630
Electrostatic	−994	−981	−1005	−1990

Table V (*Continued*)

(b) Reduced Charge

	Adjacent Loop		Diagonal Loop,	
	Same Side	Opposite Side	Opposite Side	Parallel Tetraplex
$d(G_4T_4G_4)$				
Bond + angle + dihedral	370	366	362	640
Nonbonded van der Waals	−316	−308	−266	−720
Electrostatic	−1050	−1036	−1053	−2120
Base	−850	−830	−844	−1696
Phos	−200	−206	−209	−424
Hydrogen bond 10–12	−11	−11	−8	−22
Total energy	**−1007**	**−989**	**−965**	**−2222**

[a] The terminal 5'OH and 3'OH have not been included in this analysis and hence the energy values differ marginally from those given in Tables III, IV, and VI.

[b] In the hairpin dimer structures the 5' to 3' direction of the loop joins a guanine in *anti* conformation to a guanine in *syn* conformation across the smaller groove or across diagonal. The component of electrostatic interaction indicated by "Base" is the contribution from base as well as sugar atoms, but without 05' and 03' atoms, while the component indicated as "Phos" gives the contribution from phoshate groups (phosphate atom, anionic oxygens, O3' and 05'), as well as the interaction between base and phosphates.

quences. While an all-G helix is always more stable in the parallel tetraplex conformation, when guanines are interspersed by stretches of thymines, they can either adopt a folded-back structure or be stabilized in a parallel tetraplex arrangement if there is significant screening of the phosphate groups. In fact, sequences $d(TG_3T)$ and $d(TG_3T_2G_3T)$ have been reported to adopt a parallel tetraplex structure and exhibit an unusually smaller dependence of T_m on salt concentration,[64] indicating a significantly reduced charge density on the DNA.

The stability of folded-back structures compared to the parallel tetraplex has also been investigated by a number of experimental techniques. For all-G sequences it has been unequivocally established by design of 5'-5' linked hairpins[65] that parallel arrangement of guanines is favored over an antiparallel arrangement. This agrees very well with our molecular mechanics calculations reported earlier for parallel and antiparallel G tetraplexes.[26] For sequence repeats of the type $d(T_nG_4)$, it has been suggested that, though a parallel tetraplex should be thermodynamically most stable, folded-back structures are kinetically trapped intermediates due to the relative ease of hairpin formation compared to parallel tetraplex.[7] However, a parallel tetraplex with thymines need not always be more stable than folded-back structures. Our molecular mechanics calculations on parallel tetraplex and hairpin dimer structures indicate that the parallel tetraplexes containing thymines are energetically unfavorable when compared to hairpin dimers, due to weak stacking energy and interchain phosphate repulsion between G and T stretches. Therefore, when thymines are interspersed between stretches of guanines, the telomeric sequences would prefer to adopt folded-back structures. However, if phosphate–phosphate repulsion is reduced due to shielding of phosphate groups by ionic environment (as indicated by the calculations with reduced charge on phosphate groups), they can be stabilized in a parallel tetraplex arrangement. It has been reported by Sen and Gilbert[7,66] that parallel tetraplex structures can be formed from oligomers containing multiple stretches of contiguous guanines by incubating with 1.9 *M* NaCl and 0.1 *M* KCl. In has also been observed that though the sequences $d(G_4T_nG_4)$ and $d(T_nG_4T_nG_4)$ adopt hairpin dimers in the presence of Na^+, in the presence of K^+ alone they too adopt predominantly parallel tetraplex structures.[12,14]

Energy Contribution from Cation Binding. It is important to note that both Greek key (all adjacent strands being antiparallel) as well as Indian key (adjacent strands being both parallel and antiparallel) types of structures have identical inverted stacking arrangement of bases, and only the glycosidic conformations in the four guanines of the tetrad differ in the two cases. In the former it is *syn-anti-syn-anti* while in the later it is *syn-syn-anti-anti*. Thus the two structures only differ in the backbone conformation, while the guanine stacks are identical

Table VI Energy Contribution (kcal/mole) from Guanine and Thymine Stretches of the Parallel Tetraplex Structure for $d(G_4T_nG_4)$, $n = 2–4$[a]

(a) Normal Charge

	Intra-G-Stem (a)	Intra-T-Stem (b)	Inter-G&T-Stem (c)	Total from T-Stem ($b + c$)	Total ($a + b + c$)
T_2	**−1275**	**−160**	**159**	**−1**	**−1276**
d	445	108	12	120	565
v	−429	−41	−140	−181	−610
e	−1291	−227	287	60	−1231
T_3	**−1289**	**−228**	**238**	**10**	**−1279**
d	450	163	12	175	625
v	−426	−88	−145	−233	−659
e	−1313	−303	371	68	−1245
T_4	**−1296**	**−279**	**290**	**11**	**−1285**
d	460	214	11	225	685
v	−426	−126	−142	−268	−694
e	−1330	−367	421	54	−1276

(b) Reduced Charge

	Intra-G-Stem (a)	Intra-T-Stem (b)	Inter-G&T-Stem (c)	Total from T-Stem ($b + c$)	Total ($a + b + c$)
T_2	**−1587**	**−200**	**−141**	**−341**	**−1928**
d	420	103	13	116	536
v	−450	−37	−154	−191	−641
e	−1557	−266	0	−266	−1823
T_3	**−1577**	**−307**	**−111**	**−418**	**−1995**
d	429	159	13	172	601
v	−433	−80	−138	−218	−651
e	−1573	−386	14	−372	−1945
T_4	**−1594**	**−453**	**−120**	**−573**	**−2167**
d	423	207	13	220	643
v	−448	−133	−161	−294	−742
e	−1569	−527	28	−499	−2068

[a] The symbol convention is same as in Table III.

and hence the cation binding site is unaffected by the type of folding. Even if we assume that there is a change in the helical twist due to the change in backbone orientation, the distance between the cation and the oxygens will remain unchanged. As discussed by Klug et al.,[6] if the K^+/Na^+ is at the center of the cube and oxygens (O6 of guanine) form the corners of the cube, the K^+/Na^+–O distances remain unchanged, irrespective of the twist between the faces. Therefore, as expected, the results of our calculations for $d(G_4T_4G_4)_2$ in all three models (i.e., loops on same side, loops across the edge at opposite ends of the stem, and also diagonal loops at opposite ends) with a K^+ ion between adjacent G tetrads, indicate that the additional stabilization energy gained due to the binding of cation is similar in all three structures. In fact, three K^+ ions bound to the G_4 stem have an electrostatic interaction of −240 kcal/mole with the stem and −10 to −12 kcal/mole with the loop, in all three models. Similarly, in the parallel structure for $d(G_4T_4G_4)$, the three K^+ ions bound to the G_4 stem have an electrostatic interaction of −250 kcal/mole with the G_4 stretch and −14 kcal/mole with the T_4 stretch. Thus the various types of G-tetraplex structures gain almost equal energy due to cation binding.

Analysis of Structures Minimized with Explicit Solvent and Counterions

The structures minimized with explicit solvent and counterions have structural features essentially

similar to the structures minimized without explicit solvent. For example, the parallel tetraplex structure for d($G_4T_4G_4$), which does not have T:T hydrogen bonds in the starting structure, has hydrogen-bonded thymine tetrads in the energy-minimized structures with explicit solvent. As mentioned earlier, this feature has also been observed in both normal and reduced charge minimization with $\epsilon = R_{ij}$ without solvent. The parallel tetraplex minimized with explicit solvent and counterion also has four equal grooves similar to that in the structures minimized with normal and reduced charge. The hairpin dimer structures minimized with explicit solvent and counterion have large and small grooves similar to the structures minimized with reduced charge and $\epsilon = R_{ij}$. However, as mentioned earlier when the minimization is carried out with $\epsilon = R_{ij}$ and normal charge, the small groove widens. In the hairpin dimer structure with two loops on the same side, out of eight hydrogen bonds between the thymines in the starting structure, the hydrogen bonds N3-T(20)---O2-T(5), N3-T(8)---O2-T(17), and N3-T(19)---O4-T(6) have a H---Y distance of 2.7–3.0 Å. In fact in the structures minimized without explicit solvent the former two hydrogen bonds also had a distance of 3.5 Å as shown in Table II. In all the minimizations with explicit solvent, all hydrogen bonds between guanines remain intact even in the absence of any constraints and the counterions remain close to the phosphates.

An analysis of different energy contributions, i.e., energy of the DNA alone, interaction of DNA with water, interaction of DNA with counterions, counterion water interaction, and ion–ion repulsion (all calculated with $\epsilon = 1$ and nonbonded cutoff of 9.0 Å) reveals that the three hairpin dimers have similar energy values irrespective of the orientation of the loops. Similarly, if we consider all the above-mentioned energy components for the parallel tetraplex and compare them with the equivalent energy (twice) of the hairpin dimer, the hairpin dimer is found to be energetically more favorable. However for a more realistic comparison, the equilibrium distribution of water and ion around the structure have to be determined from molecular dynamics simulations.

Relevance of the Structures Obtained from Model Building *vis-à-vis* the Crystal and NMR Structures

The approach of the present study has been to search for possible thymine loop structures with optimum stacking and hydrogen bonding, keeping all thymines in the favorable *anti* conformation. A comparison of such structures with structures without T : T pairing, as observed in nmr and crystallographic studies, indicates that these hydrogen-bonded structures are comparable in energy and hence equally possible. This assumption of stacking and hydrogen bonding in the loop seems reasonable, considering that the first and fourth thymines in the loop are also stacked over the G tetrad in the crystal structure of d($G_4T_4G_4$) and point in appropriate orientations to form potential hydrogen bonds. However, they do not form direct hydrogen bonds but are bridged by water molecules. In fact, as shown in Figure 7(a), upon energy minimization without explicit solvent, they result in a T : T pair. As discussed earlier, our models with loops over the large groove also have a similar intraloop T : T pair and when the loop is over the small groove the T : T pair will be of the interloop type to keep thymines in *anti* conformation. Similarly in the x-ray[35] as well as nmr[34] structure of the aptamer sequence d($G_2T_2G_2TGTG_2T_2G_2$), which folds with two T_2 loops on the same side of the stem, a T : T pair does occur between two diagonal thymines. However, the folding topology is different in crystal and solution.[33–35] In the crystal the two T_2 loops are over the large groove and the TGT loop is over the small groove, while in the solution structures obtained by nmr the two T_2 loops are over the small groove and the TGT loop is over the large large groove. The two nmr studies on the same aptamer give different results, while in one case[34] T : T pairing is observed; in the other case[33] these two thymines are found to be only stacked over the G stem and no T : T pairing is reported. In the structures obtained after energy minimization, though the thymines remain stacked and hydrogen bonded, the thymine tetrads show larger buckle and propeller type distortions compared to guanine tetrads, indicating that they can undergo fluctuations to take up other nonsymmetric arrangements as observed experimentally. Rich and co-workers[16] also mention that, within the crystal itself, the loops show variable conformation and between the solution and crystal structure there is a large variation in conformation and orientation of the loop. Thus the telomeric DNA can exhibit a high degree of structural polymorphism and the present model building elucidates a class of such structures with potential to adopt interloop T : T pairing.

CONCLUSION

The present study indicates that, while a parallel tetraplex is favorable for an all-G sequence,[26] the repeat sequences of the type d(T_nG_4) would favour a folded-back structure. Our molecular mechanics

calculations qualitatively explain the observed propensity for telomeric sequences to adopt folded-back structures and also indicate how it might be possible to stabilize parallel tetraplexes for these sequences under very high ionic strengths. This result provides a theoretical rationale for the experimental observations of a variety of structures, for G containing oligonucleotides. For any given sequence a large number of folded-back structures are possible, which are structurally quite different, but have very similar energy. Our model building study indicates that it is possible to have folded-back structures with loops joining adjacent strands over the large as well as small groove of the tetraplex, with a minimum of two nucleotides. A minimum of three nucleotides are required for a loop joining diagonal strands. It is possible to have energetically favorable folded-back structures with stacking and hydrogen bonding in the loop region, when two loops are on the same side of the tetraplex. Though in the crystal and nmr structures of $d(G_4T_4G_4)$ the thymine loops are on opposite ends,[16,17] for four repeat sequences two thymine loops will necessarily be on same side. In fact, such pairing has been observed in the folded-back structure of an aptamer sequence,[34,35] and as pointed out for this structure,[34] the interloop T-T pairing might have a role in the biological function of these sequences. Similar interloop pairing may also be expected to occur in other telomeric sequences as postulated from our theoretical model building and molecular mechanics studies. The comparison of structures minimized without explicit solvent with those obtained from minimization with explicit solvent and counterion indicates that they have very similar structural features.

Financial assistance to MB from DBT (India) is gratefully acknowledged. DM thanks CSIR (India) for a fellowship.

REFERENCES

1. Blackburn, E. H. & Szostak, J. W. (1984) *Ann. Rev. Biochem.* **53,** 163–194.
2. Henderson, E. R. & Blackburn, E. H. (1989) *Mol. Cell. Biol.* **9,** 345–348.
3. Blackburn, E. H. (1990) *J. Biol. Chem.* **265,** 5919–5921.
4. Blackburn, E. H. (1991) *Nature* **350,** 569–573.
5. Sen, D. & Gilbert, W. (1988) *Nature* **334,** 364–366.
6. Sundquist, W. I. & Klug, A. (1989) *Nature* **342,** 825–829.
7. Sen, D. & Gilbert, W. (1990) *Nature* **344,** 410–414.
8. Zhaler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. (1991) *Nature* **350,** 718–720.
9. Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., Jr. and Blackburn, E. H. (1987) *Cell* **51,** 899–908.
10. Williamson, J. R., Raghuraman, M. K. & Cech, T. R. (1989) *Cell* **59,** 871–880.
11. Jin, R., Breslauer, K. J., Jones, R. A. & Gaffney, B. L. (1990) *Science* **250,** 543–546.
12. Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M. & Sasisekharan, V. (1992) *Nucleic Acids Res.* **20,** 4061–4067.
13. Guo, Q., Lu, M. & Kallenbach, N. R. (1992) *J. Biol. Chem.* **267,** 15293–15300.
14. Guo, Q., Lu, M. & Kallenbach, N. R. (1992) *Biochemistry* **32,** 3596–3603.
15. Scaria, P. V., Shire, S. J. & Shafer, R. H. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10336–10340.
16. Kang, C., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. (1992) *Nature* **356,** 126–131.
17. Smith, F. W. & Feigon, J. (1992) *Nature* **356,** 164–168.
18. Sasisekharan, V., Zimmermann, S. B. & Davies, D. R. (1975) *J. Mol. Biol.* **92,** 171–179.
19. Pinnavaia, T. J., Miles, H. T. & Becker, E. D. (1975) *J. Am. Chem. Soc.* **97,** 7198–7200.
20. Arnott, S., Chandrasekharan, R. & Martilla, C. M. (1974) *Biochem. J.* **141,** 537–543.
21. Zimmermann, S. B., Cohen, G. H. & Davies, D. R. (1975) *J. Mol. Biol.* **92,** 181–192.
22. Guschlbauer, W., Chantot, J. F. & Thiele, D. (1990) *J. Biomol. Struct. Dynam.* **8,** 491–511.
23. Smith, F. W. & Feigon, J. (1993) *Biochemistry* **32,** 8682–8692.
24. Wang, Y., Jin, R., Gaffney, B., Jones, R. A. & Breslauer, K. J. (1991) *Nucleic Acids Res.* **19,** 4619–4622.
25. Wang, Y., de los Santos, C., Gao, X., Greene, K., Live, D. & Patel, D. J. (1991) *J. Mol. Biol.* **222,** 819–832.
26. Mohanty, D. & Bansal, M. (1993) *Nucleic Acids Res.* **21,** 1767–1774.
27. Wang, Y. & Patel, D. J. (1992) *Biochemistry* **31,** 8112–8119.
28. Cheong, C. & Moore, P. B. (1992) *Biochemistry* **31,** 8406–8414.
29. Aboul-ela, F., Murchie, A. I. H., Lilly, D. M. (1992) *Nature* **360,** 280–282.
30. Gupta, G., Garcia, A. E., Guo, Q., Lu, M. & Kallenbach, N. R. (1993) *Biochemistry* **32,** 7098–7103.
31. Frank-Kamenetskii, M. (1992) *Nature* **356,** 105–105.
32. Bansal, M. & Mohanty, D. (1993) *J. Biomol. Struct. Dynam.* **10,** a010.
33. Wang, K. Y., McCurdy, S., Shea, R. G., Swaminathanan, S. & Bolton, P. H. (1993) *Biochemistry* **32,** 1899–1904.
34. Schultze, P., Macaya, R. F. & Feigon, J. (1994) *J. Mol. Biol.* **235,** 1532–1547.
35. Padmanabhan, K., Padmanabhan, K. P., Ferrara, J. D., Sadler, J. E. & Tulinsky, A. (1993) *J. Biol. Chem.* **268,** 17651–17654.
36. Wang, Y. & Patel, D. J. (1993) *Structure* **1,** 263–282.
37. Hasnoot, C. A. G., de Bruin, S. H., Berendsen, R. G., Janssen, H. G. J. M., Binnendijk, T. J. J., Hilbers,

C. W., van der Marel, G. A. & van Boom, J. H. (1983) *J. Biomol. Struct. Dynam.* **1,** 115–129.
38. Hasnoot, C. A. G., Hilbers, C. W., van der Marel, J. H., Singh, U. C., Pattabiraman, N. & Kollman, P. A. (1986) *J. Biomol. Struct. Dynam.* **3,** 843–847.
39. Hilbers, C. W., Hasnoot, C. A. G., de Bruin, S. H., Joorden, J. J., van der Marel, G. A. & van Boom, J. H. (1985) *Biochimie* **67,** 685–695.
40. Hare, D. R. & Reid, B. R. (1986) *Biochemistry* **25,** 5341–5350.
41. Chattopadhyaya, R., Grzeskowiak, K. & Dickerson, R. E. (1988) *Nature* **334,** 175–179.
42. Xodo, L. E., Manzini, G., Quadriofoglio, F., van der Marel, G. A. & van Boom, J. H. (1991) *Nucleic Acids Res.* **19,** 1505–1511.
43. Blommers, M. J. J., van de Ven, F. J. M., van der Marel, G. A., van Boom, J. H. & Hilbers, C. W. (1991) *Eur. J. Biochem.* **201,** 33–51.
44. Raghunathan, G., Jernigan, R. L., Miles, H. T. & Sasisekharan, V. (1991) *Biochemistry* **30,** 782–788.
45. Harvey, S. C., Luo, J. & Lavery, R. (1988) *Nucleic Acids Res.* **16,** 11795–11809.
46. Erie, D. A., Suri, A. K., Breslauer, K. J., Jones, R. A. & Olson, W. K. (1993) *Biochemistry* **32,** 436–454.
47. Boulard, Y., Gabarro-Arpa, J., Cognet, J. A. H., Le Bret, M., Buy, A., Teoule, R., Guschlbauer, W. & Fazakerley, G. V. (1991) *Nucleic Acids Res.* **19,** 5159–5167.
48. Bansal, M. & Sasisekharan, V. (1986) in *Theoretical Chemistry of Biological Systems,* Naray-Szabo, G., Ed., Elsevier, Amsterdam, New York, pp. 127–288.
49. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992) *Biophys. J.* **63,** 751–759.
50. AMBER Ver. 3.0 (1987) Molecular Mechanics Program from University of California, San Francisco.
51. Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986) *J. Comp. Chem.* **7,** 230–252.
52. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. Jr. & Weiner, P. (1984) *J. Am. Chem. Soc.* **106,** 765–784.
53. Bansal, M. & Pattabiraman, N. (1989) *Biopolymers* **28,** 531–548.
54. Tidor, B., Irikura, K. K., Brooks, B. R. & Karplus, M. (1983) *J. Biomol. Struct. Dynam.* **1,** 231–252.
55. Chuprina, V. P., Heinemann, U., Nurislamov, A. A., Zielenkiewicz, P., Dickerson, R. E. & Saenger, W. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 593–597.
56. Cheng, Y. K. & Pettitt, B. M. (1992) *J. Am. Chem. Soc.* **114,** 4465–4474.
57. Radhakrishnan, I. & Patel, D. J. (1993) *Structure* **1,** 135–152.
58. Pearlman, D. A., Case, D. A., Caldwell, J. C., Seibel, G. L., Singh, U. C., Weiner, P. & Kollman, P. A. (1991) AMBER 4.0, University of California, San Francisco.
59. Sarma, M. H., Luo, J., Umemoto, K., Yuan, R. D. & Sarma, R. H. (1992) *J. Biomol. Struct. Dynam.* **9,** 8112–8119.
60. Xu, Q., Deng, H. & Braunlin, W. H. (1993) *Biochemistry* **32,** 13130–13137.
61. Taylor, R. & Kennard, O. (1982) *J. Am. Chem. Soc.* **104,** 5063–5070.
62. Mohanty, D. & Bansal, M., manuscript in preparation.
63. Rose, I. A., Hanson, K. R., Wilkinson, K. D. & Wimmer, M. J. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 2439–2441.
64. Jin, R., Gaffney, R. L., Wang, C., Jones, R. A. & Breslauer, K. J. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8832–8836.
65. Lu, M., Guo, Q. & Kallenbach, N. R. (1993) *Biochemistry* **32,** 589–601.
66. Sen, D. & Gilbert, W. (1992) *Methods Enzymol.* **211,** 191–199.

Received November 29, 1993
Accepted April 5, 1994