Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Stabilization of the *i*-motif by intramolecular adenine–adenine–thymine base triple in the structure of d(ACCCT)

Jonathan Weil,[a] Tongpil Min,[a] Cheng Yang,[a] Shuren Wang,[a] Cory Sutherland,[a] Nanda Sinha[b] and ChulHee Kang[a]*

[a]Department of Biochemistry and Biophysics, Washington State University, Pullman WA 99164–4660, USA, and [b]Boston Biosystem Inc., 75A Wiggins Avenue, Bedford MA 01730, USA

Correspondence e-mail: kang@kang2.chem.wsu.edu

Received 1 July 1998
Accepted 28 September 1998

PDB Reference: d(ACCCT), 1bqj.

The crystal structure of d(ACCCT), solved by molecular replacement, shows a four-stranded *i*-motif conformation, where two parallel duplexes intercalate with one another in opposite orientations. Each duplex is stabilized by hemi-protonated C–C$^+$ base pairing between parallel strands, and a string of water molecules bridge the cytosine N4 atoms to phosphate O atoms. This structure of d(ACCCT) shows examples of reversed Hoogsteen and Watson–Crick base pairing in both intermolecular and intramolecular manners to stabilize the tetraplex. Noticeably, the four-stranded complex is further stabilized at one end by a three-base hydrogen-bonding network, in which two adenines and a thymine form four hydrogen bonds *via* a reverse Hoogsteen and an asymmetric adenine–adenine base pairing. The structure of d(ACCCT) shows a similar local structure to that found in the d(TAA) part of the crystal structure of d(TAACCC) and provides further structural evidence that these base arrangements are essential for stabilizing these novel DNA super-secondary structures.

1. Introduction

Metazoan telomeric DNA is composed of tandem repeats of d(5′-GGGTTA-3′) on one strand (G-rich strand) and the sequence d(5′-TAACCC-3′) on the complementary strand (C-rich strand). The G-rich strand at the 3′ terminus typically exists as a single-stranded overhang and is able to form a novel DNA super-secondary structure known as a guanine quartet. The degree of polymorphism has been demonstrated (Williamson, 1994; Rhodes & Giraldo, 1995; Brown & Hunter, 1997) and potential biological roles have been proposed (Kang *et al.*, 1992; Zakian, 1995). It is believed that the telomeres may be a structural regulatory element for the interaction between the telomeric DNA and the enzyme telomerase (Blackburn, 1991). Despite the high degree of polymorphism among G-quartet structures, all of these structures are mainly stabilized by the same interactions. These include a network of eight hydrogen bonds created by four guanines arranged in a plane, stacking interactions between these planes and the coordination of monovalent cations.

More recently, an NMR study of d(CCCCC) showed that it can also form a four-stranded structure with cytosine-containing parallel duplexes held together by hemiprotonated (C–C$^+$) base pairs and two duplexes intercalated with each other in opposite polarity (Fig. 1; Gehring *et al.*, 1993; Leroy *et al.*, 1994). It has been known that nucleic acid polymers with spans of cytosine residues can form parallel duplexes stabilized by hemi-protonated (C–C$^+$) base pairs (Langridge & Rich, 1963; Akinrimisi *et al.*, 1963; Inman, 1964; Hartman &

Rich, 1965). Several crystal structures of variations of C-rich metazoan telomeric DNA sequences, such as d(CCCC), d(CCCT), d(TAACCC) and d(CCCAAT), have been reported at various resolutions (Fig. 1; Chen *et al.*, 1994; Kang *et al.*, 1994, 1995; Berger *et al.*, 1995). In all these structures, the four-stranded complexes contain two parallel duplexes intercalated within one another in an antiparallel orientation. These individual parallel duplexes are stabilized by hemiprotonated ($C–C^+$) base pairs, even though the crystals were prepared near neutral pH. At neutral pH, cytosine should not be protonated, since the pK_a of N3 of cytosine is 4.7. It has been assumed that the pK_a for cytosine in this environment is raised at least to 7 in order to stabilize the structure with three hydrogen bonds (Inman, 1964; Kang *et al.*, 1995). Raman spectra indicate that the quadruples of the d(CCCT) crystal structure are conserved in aqueous solution (Benevides *et al.*, 1996). The C-rich strand is also additionally stabilized by hydrogen-bonding networks of Watson–Crick base pairing as well as non-Watson–Crick base pairing of adenine–adenine, thymine–thymine and adenine–thymine base pairs, which occur between the flanking non-cytosine bases of the telomeric DNA sequences (Kang *et al.*, 1994, 1995; Berger *et al.*, 1995). The crystal structure of d(TAACCC) has revealed that it utilizes the sequence TAA at the ends to form a novel loop in which the 5′ end of one molecule is in close proximity to the 3′ end of another molecule (Kang *et al.*, 1995). This loop at the end of the C-quartet is stabilized by three interactions: (i) the first T forms a hydrogen bond to the third A in a Hoogsteen base pair, (ii) the loop is stabilized by stacking interactions and (iii) in addition, the phosphate group between the second A and the third A is stabilized by a hydrogen-bonded interaction with the N4 amino group of cytosine. All of these interactions combine to form a very tight loop involving a TAA sequence at the 5′ end of the molecule. This C-strand telomere structure strongly suggests that the loop may provide a mechanism for stabilizing the interaction of two pairs of C-strand repeats. The presence of a stable loop associated with the four-stranded cytosine structure reinforces the possibility that conformations of this type may be found in the telomere.

The studies on proteins *in vivo* that bind specifically to telomeres in yeast and humans suggest these telomeric DNA super-secondary structures may represent some structural motif for the recognition of enzyme or protein factors (Rhodes & Giraldo, 1995; Blackburn, 1994). It has also been

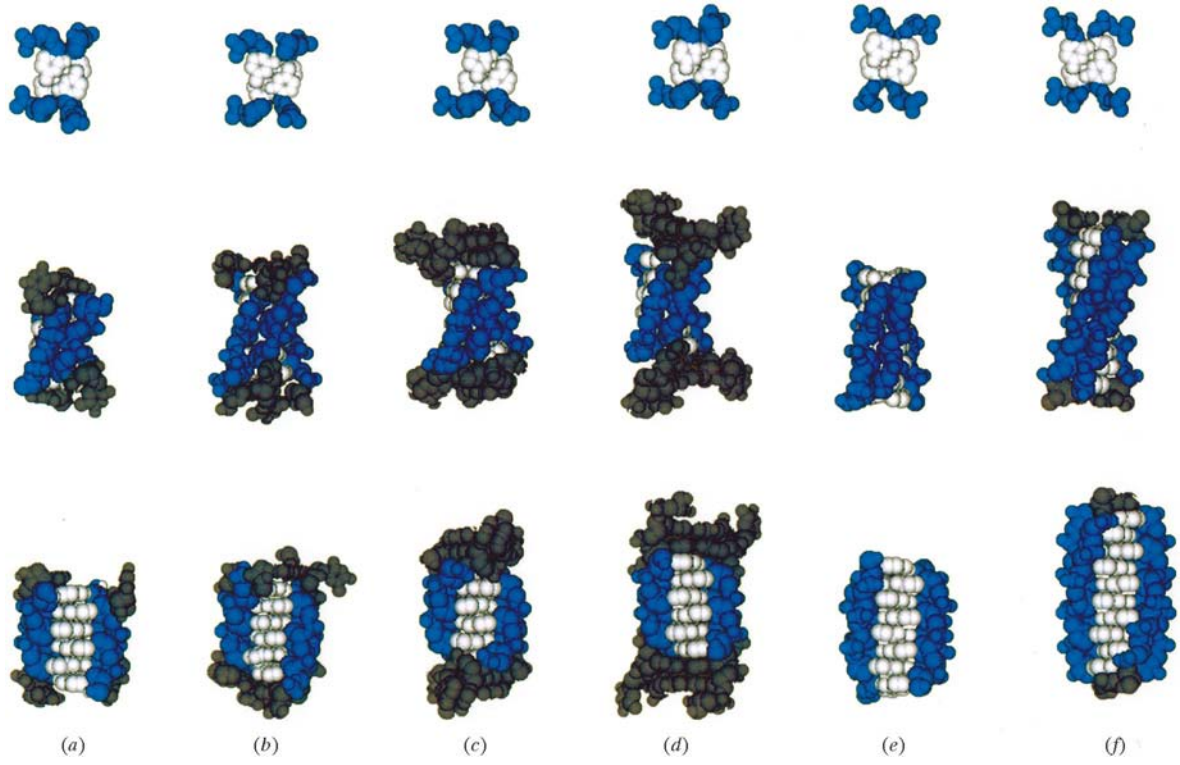


Figure 1
Space-filling representations of the four-stranded C-rich DNAs. (*a*) d(CCCT) (1.4 Å, $C222_1$, a = 28.281, b = 44.341, c = 50.47 Å; Kang *et al.*, 1994), (*b*) d(ACCCT) (2.2 Å, $I222$, $a = b = c$ = 93.81 Å; Kang *et al.*, 1994), (*c*) d(TAACCC) (1.85 Å, $F222$, a = 59.94, b = 81.33, c = 26.86 Å; Kang *et al.*, 1995), (*d*) d(CCCAAT) (2.0 Å, $P62$, $a = b$ = 32.18, c = 52.49 Å; Berger *et al.*, 1995), (*e*) d(CCCC) (1.8 Å, $I23$, $a = b = c$ = 82.3 Å; Chen *et al.*, 1994), (*f*) d(TCCCCC) (NMR model; Gehring *et al.*, 1993). The backbone of the C-rich part is coloured blue, the base is coloured white and the non-C part is coloured gray. First row: top views of the molecules showing two adjacent $C–C^+$ base pairs. In the left grooves of (*a*), (*b*), (*c*) and (*d*), the phosphates are bent over and rotated towards each other. On the other hand, the left and right grooves of (*e*) and (*f*) are more or less the same size. Second row: side views of the molecules showing the manner in which the two sugar phosphate chains at the narrow end of the molecules come close together in their antiparallel orientation. Note that in case of (*a*), (*b*), (*c*) and (*d*), the sugar–phosphate chains at the left side of the molecules are all fairly straight, whereas the chains on the right side of molecules have the phosphates rotated sharply away from the center of the molecules. In contrast, in molecules (*e*) and (*f*) both strands shows similar conformations. Third row: side views of the molecules showing the two sugar–phosphate chains in their antiparallel orientation at the wider groove of the molecules.

shown that the composition of the telomeric DNA itself plays a critical role in telomere regulation (McEachern & Blackburn, 1995).

2. Experimental

Crystals of d(ACCCT) were grown by the hanging-drop vapor-diffusion method from a solution containing 5 m*M* d(ACCCT), 5%(*v*/*v*) 2-propanol, 10%(*v*/*v*) ethanol, 0.5 *M* $(NH_4)_2SO_4$, 150 m*M* NaCl, 5 m*M* $MgCl_2$ and 2 m*M* spermine. Hexagonal crystals grew slowly over a period of one month and reached a size of 0.3 × 0.3 × 0.1 mm. X-ray diffraction analysis revealed that the molecule had crystallized in the cubic space group *I*23 with unit-cell parameters $a = b = c = 93.81$ Å. The data set was collected on a Rigaku R-AXIS IIC imaging plate by the flash-freezing method (103 K; crystal-to-detector distance of 100 mm). The final R_{sym} value was 5%, generating 3679 unique reflections above the 1σ level and 2643 reflections above the 2σ level in the resolution range 15.0–2.4 Å. The molecular-replacement method *X-PLOR* (Brünger, 1992) was used to attempt to solve the structure of d(ACCCT), using a part of the crystal structure of d(CCCT) (Kang *et al.*, 1994) as a starting model. Initially, a cross-rotational search with 15–3.0 Å data produced more than 89 unique orientations, whose rotation-function (RF) values ranged from 0.27 to 0.23. 16 independent solution sets were left after a subsequent Patterson correlation stage using data in the 12.0–2.4 Å resolution range. After translational searches with these various orientations using 2175 reflections from 15.0 to 3.0 Å resolution, a unique position for four-stranded d(CCC) which contained no symmetry overlap was generated. The molecular position of this solution in the lattice showed that the helical axis was not oriented along any axis of the unit cell, which was consistent with the self-rotation search. Initial rigid-body refinement was carried out using 2636 reflections from data in the resolution range 15.0–2.5 Å and produced an *R* value of 41%. After several cycles of positional refinement, temperature-factor refinement and simulated-annealed omit-map calculation, we were able to fit the remaining seven residues to the electron density. The residue T5 is disordered in our crystal structure and the corresponding electron density was not visible from the early stage of refinement. The higher resolution data set was collected on a Rigaku R-AXIS IIC imaging plate. The final R_{sym} value of this new data set was 4.5%, generating 7136 unique reflections above the 2σ level in the 15.0–2.2 Å resolution range. The *R* factor for the final model containing 380 non-H atoms is 19.2% (free *R* factor 23.5%) for 6914 reflections above the 2σ level between 8.0 and 2.2 Å resolution. The r.m.s. deviation is 0.009 Å for bonds and 1.691° for angles. 39 water molecules are included in the structure. The coordinates and structure factors have been deposited in the Protein Data Bank.

3. Results and discussion

In the crystal structure of d(ACCCT), representing a truncated analog of the C-rich telomeric single-repeat sequence in metazoans, the four strands make up an intercalation motif, in which two parallel duplexes form a tetraplex by orienting themselves in antiparallel directions (Fig. 2). Each parallel duplex exhibits a slight right-handed twist of about 14.5°. Inspection of the central portion of the molecule reveals a flat lath-like structure with two broad and two narrow grooves,

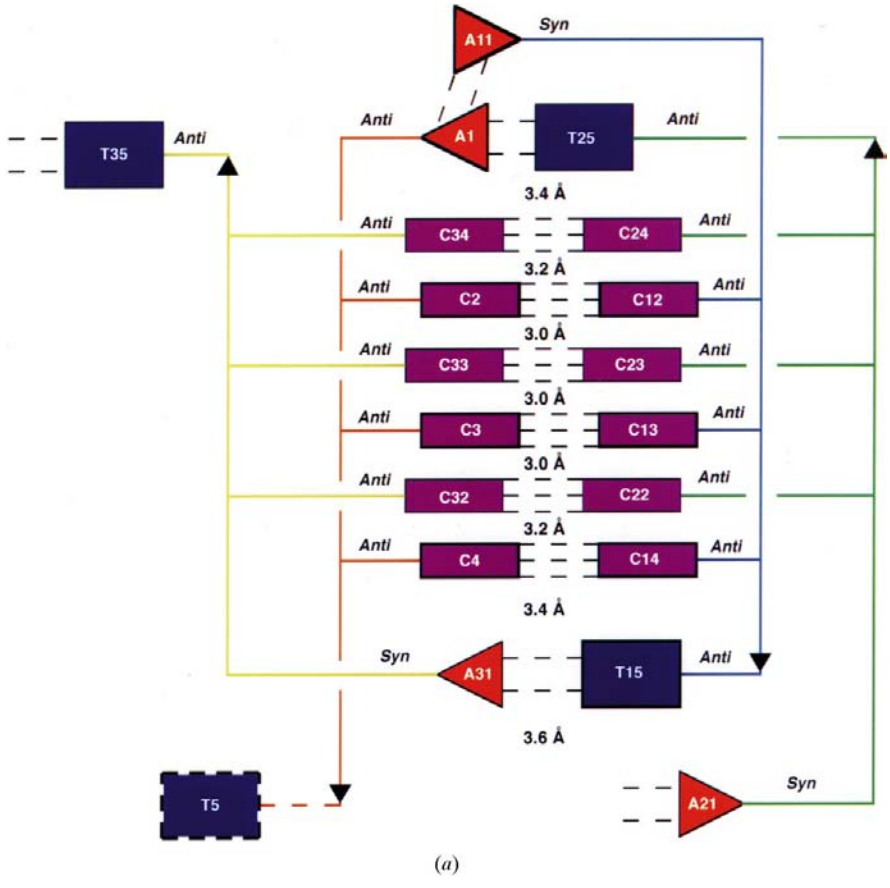
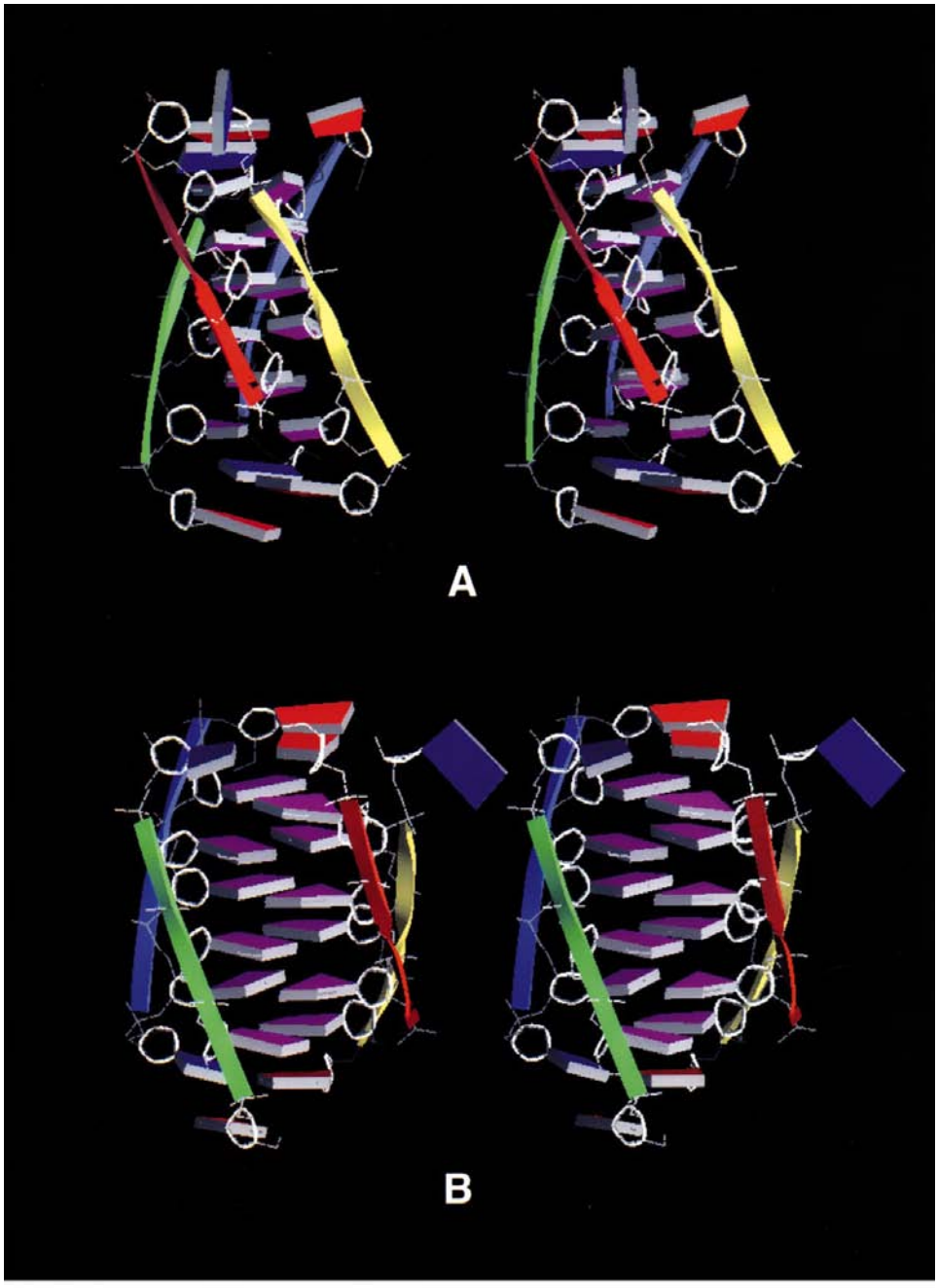


Figure 2
(*a*) Schematic diagram illustrating the numbering systems and interactions of the four strands of d(ACCCT). The molecule is composed of two sets of parallel-stranded duplexes held together by C–C^+ hydrogen bonds. The two duplexes intercalate and have opposite orientations. At the bottom and top, a Watson–Crick base pair and a base triple are shown, respectively. A disordered residue, T5, is drawn with bold dashed lines. Hydrogen bonds are drawn with dashed lines. *Syn* and *anti* indicate conformations around the glycosyl bond. The base-stacking distances (Å) have been indicated between the corresponding paired bases. Adenine, thymine and cytosine bases are depicted by red, blue and magenta colours, respectively.

which is similar to the *i*-motif parts of other C-rich strand crystal structures, with the exceptions of d(CCCC) (Chen *et al.*, 1994) and d(TCCCCC) (Gehring *et al.*, 1993; Fig. 1). The mean distance between phosphates in the narrow grooves is 8.1 (1.5) Å and is 15.0 (1.7) Å in the broad grooves (Table 1). The distance between phosphate groups in the two narrow grooves and one of the broad grooves varies in an alternating pattern. This asymmetry is seen most clearly in Figs. 1(*a*)–1(*d*). In the first row of Fig. 1, two continuous layers of C–C^+ base pairs in five different crystal structures and one NMR structure are shown. In molecules (*a*), (*b*), (*c*) and (*d*), the two strands at the right-hand side of the molecules are relatively flat, while the strands at the left-hand side have curved conformations owing to the fact that the phosphate groups are rotated away from the flatter strands. This asymmetric structure of sugar–phosphate backbones is not visible in the structures of d(CCCC) and d(TCCCCC).

The core part of this $[d(ACCCT)]_4$ molecule is much more compact than its B-DNA counterpart, and the central four layers of cytosines (C22–C32, C13–C3, C23–C33, C2–C12) are very planar, the average distance between consecutive planes of cytosine base pairs averaging 3.0 Å (Fig. 2*a*). However, the two outer layers of cytosine base pairs (C4–C14, C24–C34) are propeller-twisted with respect to each other by the rotation of the cytosine base pairs around the C1′—N1′ bond. There is an average twist angle of 11° seen in these two layers and these terminal cytosine base pairs (C4–C14, C24–C34) have adjacent cytosine bases (C22–C32, C2–C12, respectively) at a distance of 3.2 Å and adjacent non-cytosine bases at a distance of 3.4 Å (Fig. 2*a*).



(*b*)

Figure 2 (continued)
(*b*) Schematic diagram of the four-stranded d(ACCCT) molecule in two different orientations, showing a narrow groove (A) and broad groove (B). This figure was produced using *GRASP* (Nicholls *et al.*, 1991). The same colours are used as in Fig. 2(*a*).

The hydration pattern of d(ACCCT) resembles those of d(CCCT) and d(CCCAAT), but is more evenly hydrated. In this d(ACCCT) crystal structure, all cytosines have their N4 amino group hydrogen bonded to a water molecule, which bridges to a neighboring phosphate's oxygen in turn directly or through a second water molecule (Fig. 3).

All the cytosine residues of d(ACCCT) have glycosidic bonds in the *anti* conformation (Fig. 2) and most of the cytosine sugars exhibit C3′-*endo* puckering. Raman spectra of d(CCCT) indicate that the C3′-*endo* deoxynucleotides are conserved in aqueous solution (Benevides *et al.*, 1996), despite the fact that the C3′-*endo* conformation is rarely encountered for aqueous DNA molecules. In the case of previously examined crystal structures, C3′-*endo* nucleoside conformers are invariably converted in solution to the more energetically favorable C2′-*endo* conformation (Benevides *et al.*, 1996). However, the adenines

Table 1
Interstrand and intrastrand phosphorus–phosphorus distances (Å).

Intrastrand

C2 (P)–C3 (P)	6.0	C22 (P)–C23 (P)	6.0
C3 (P)–C4 (P)	5.9	C23 (P)–C24 (P)	6.3
C4 (P)–T5 (P)	6.2	C24 (P)–T25 (P)	6.3
C12 (P)–C13 (P)	6.8	C32 (P)–C33 (P)	5.7
C13 (P)–C14 (P)	6.7	C33 (P)–C34 (P)	6.8
C14 (P)–T15 (P)	6.4	C34 (P)–T35 (P)	6.7

Interstrand

Across the narrow groove		Across the broad groove	
C2 (P)–T35 (P)	7.9	C2 (P)–T25 (P)	14.6
T35 (P)–C3 (P)	5.7	T25 (P)–C3 (P)	15.3
C3 (P)–C34 (P)	9.2	C3 (P)–C24 (P)	12.4
C34 (P)–C4 (P)	6.8	C24 (P)–C4 (P)	14.2
C4 (P)–C33 (P)	9.3	C4 (P)–C23 (P)	12.1
C33 (P)–T5 (P)	6.7	C23 (P)–T5 (P)	13.7
T5 (P)–C32 (P)	9.6	T5 (P)–C22 (P)	12.1
C12 (P)–T25 (P)	10.0	C12 (P)–T35 (P)	16.1
T25 (P)–C13 (P)	6.2	T35 (P)–C13 (P)	16.6
C13 (P)–C24 (P)	9.0	C13 (P)–C34 (P)	16.0
C24 (P)–C14 (P)	5.9	C34 (P)–C14 (P)	16.3
C14 (P)–C23 (P)	8.2	C14 (P)–C33 (P)	16.8
C23 (P)–T15 (P)	8.1	C33 (P)–T15 (P)	17.1
T15 (P)–C22 (P)	10.6	T15 (P)–C32 (P)	16.1

and thymidines that are located at the periphery of the molecule exhibit variability in sugar puckering and base conformation (Fig. 2). This degree of variation might be a consequence of their involvement in intramolecular and intermolecular stabilization.

The main body of the molecule is extended beyond the C–C$^+$ base pairs by adding adenine and thymidine residues. At the bottom and top, the four chains organized by the three cytosines in each strand are stacked by either a base pair or a base triple, respectively (Fig. 2). Adenine residue A31 and thymidine residue T15 form a Watson–Crick base pair, which is stacked upon the cytosine C–C$^+$ base pair (C4–C22). The A21 base is stacked upon this A–T base pair. This particular local structure of two adenines and one thymine has been observed in the crystal structure of d(TAACCC) (Kang *et al.*, 1995). Based upon the formation of a very tight loop by two adenines and one thymine in the crystal structure of d(TAACCC), it has been proposed that the C-rich telomeres have the potential to stably interact and thus to recognize another C-strand telomere sequence (Kang *et al.*, 1995; Ahmed *et al.*, 1994). The observation of the same type of interactions between two adenines and one thymine in the crystal structure of d(ACCCT), despite a different crystal system and sequence, provides further evidence for the *in vivo* existence of this type of interaction between two pairs of C-strand.

Adenine residue A21 also plays an essential role in building the lattice, together with thymidine 35 (Fig. 2*a*). This A21 and a symmetry-related T35 curve back to the molecule where they form a reverse Hoogsteen base pair, which is stacked on top of the Watson–Crick A21–T15 base pair forming a continuous helical axis in the crystal lattice.

Of particular interest is the base triple formed between adenine 1 (A1), adenine 11 (A11) and thymine 25 (T25) at the

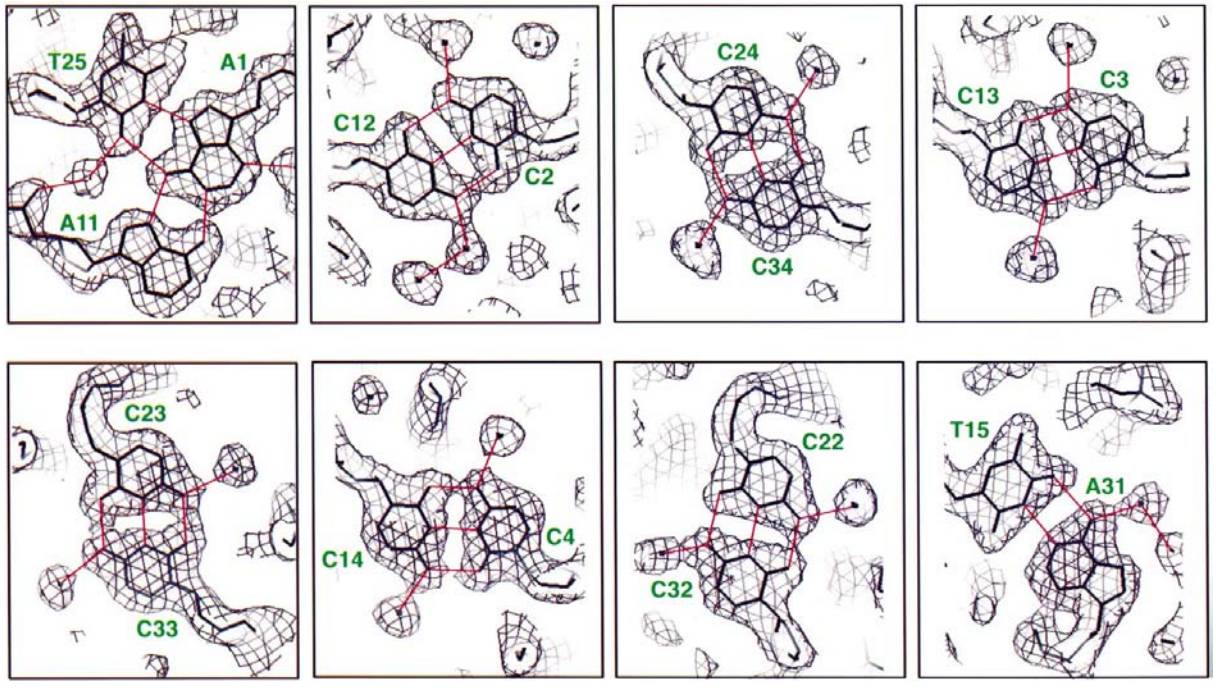
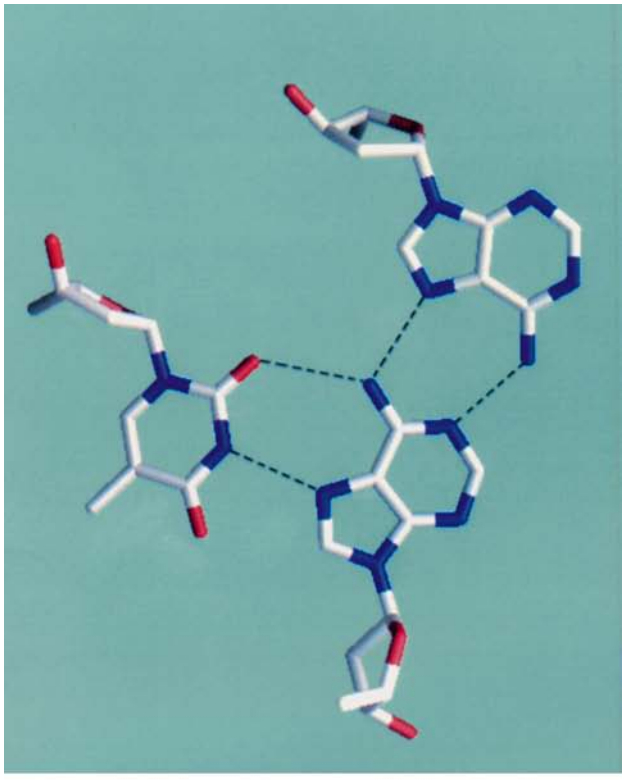


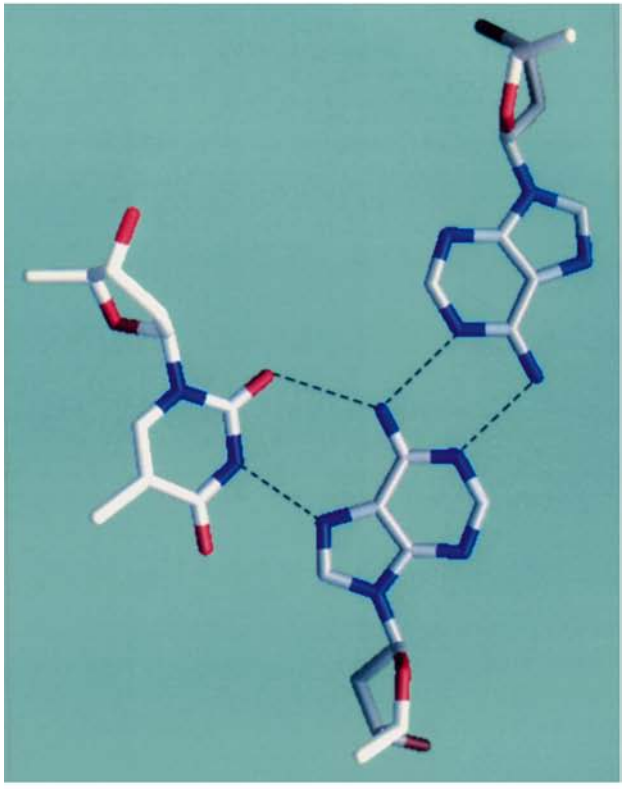
Figure 3
Hydrogen bonding with associated $2F_o - F_c$ electron-density maps (1σ level). A hydrogen bond is formed between A1 and T25 in a reverse Hoogsteen base pairing and another hydrogen bond is formed between A1 and A11 in an asymmetrical A–A pattern. Three hydrogen bonds are shown between cytosine residues of the central C–C$^+$ base pair. All the cytosine bases have their N4 amino group hydrogen bonded to a water molecule. Hydrogen bonds are formed between T15 and A31 in a Watson–Crick pairing. This figure was produced using *O* (Jones *et al.*, 1991).



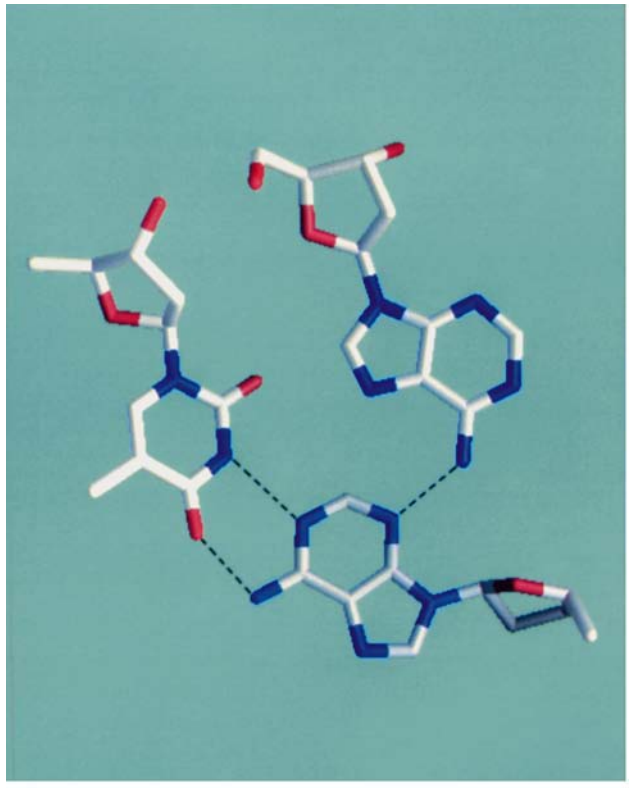
(*a*)

top of the molecule (Fig. 2, Fig. 3). A1 and A11 are contributed from a parallel-stranded duplex and T25 belongs to the oppositely oriented strand. Two hydrogen bonds are formed between A1 and T25 in a reverse Hoogsteen base pairing and another two hydrogen bonds are formed between A1 and A11 in an asymmetric homo-purine pattern (Fig. 3, Fig. 4*a*). This adenine–adenine–thymine (A–A–T) base triple has previously been reported in the crystal structure of the hexamer d(CCCAAT), in which the respective A–A–T base triples are formed between symmetry-related molecules, in contrast to the intramolecular A–A–T base triple of this d(ACCCT) structure. Another type of intermolecular (A–A–T) base triple has been found in the crystal packing of d(CCCAAT), where a *trans* adenine–adenine base pair and a reverse Hoogsteen base pair are involved in a base triple (Fig. 4*b*). A different type of (A–A–T) base triple has also been found in the four-stranded structure of the *Bombyx mori* single-repeat telomeric DNA sequence, d(TAGG) (Kettani *et al.*, 1997) (Fig. 4*c*). This d(TAGG) quadruplex contains two adjacent G–G–G–G tetrads sandwiched between two (T–A–A) base triples.

The structure, function and application of base triples have been reported in a wide variety of contexts including a potential antisense therapeutic application (Radhakrishnan & Patel, 1994; Frank-Kamenetskii & Mirkin, 1995). However, there is very little information available regarding the



(*b*)



(*c*)

Figure 4
The hydrogen-bonding alignments and relative base arrangements found in base triples of (*a*) d(ACCCT) (Kang *et al.*, 1994); (*b*) d(TAGG) (Kettani *et al.*, 1997) and (*c*) d(CCCAAT) (Berger *et al.*, 1995).

geometry of base triples, particularly the purine–purine–pyrimidine type involving adenines and thymines. Fiber diffraction (Liu *et al.*, 1994) and spectroscopic studies (Radhakrishnan & Patel, 1994; Sklen'ar & Feigon, 1990) have shown three different base triples, A–A–T, G–C–G and T–A–T, which have been reported to be formed by single-stranded oligonucleotides intercalating into the major groove of the duplex DNA.

Recently, a model for antiparallel double-stranded DNA has been proposed where the individual steps are base triples rather than base pairs (Kettani *et al.*, 1997; Kuryavyi & Jovin, 1994). The model was put forward to explain the nucleic acid architecture postulated to exist for the expanded trinucleotide repeat sequences which result in the onset of human diseases such as fragile X syndrome (CGG repeats) and myotonic dystrophy (CAG repeats) (Sinden & Wells, 1992).

The crystal structure of a peptide nucleic acid (PNA; Nielsen *et al.*, 1991) was also determined in complex with a DNA molecule (Betts *et al.*, 1995), where two base triples, C–G–C and T–A–T, have been shown. In each case, two bases form Watson–Crick base pairs (G–C and A–T) and, subsequently, the Hoogsteen base pairs are found between the remaining bases (C–C and A–T). In the crystal structure of d(ACCCT), the triple base pairing between A–A–T contains no Watson–Crick base-pairing scheme between the adenine and thymine as is seen in the PNA structure. In the PNA structures, the N3 atom of the cytosine in the GCG base triple is protonated and hydrogen bonded to the N7 position of the corresponding guanine. This phenomenon is analogous to the state of the N3 atom of the cytosine base pairs which stabilize the central six layers of the d(ACCCT) structure. The d(ACCCT) and PNA molecules were crystallized at the pH values of 7.0 and 8.5, respectively, while the pK_a of the N3 position is 4.7. These are good examples that cytosine pK_a can be shifted by the formation of stable structures.

The crystal structure of the decamer d(GGCCAATTGG)$_2$ also showed a unique packing arrangement where the terminal guanines of one duplex form Hoogsteen base pairs with the G–C Watson–Crick base pairs of a second duplex in either a parallel or antiparallel fashion (Vlieghe *et al.*, 1996). This G–C–G base triple again demonstrates the bonding motif in triple helices where a Hoogsteen base pair exists. Unlike the purine–purine–pyrimidine base triple of d(ACCCT), the base triple in the d(GGCCAATTGG)$_2$ structure shows that all of the bases involved in triple-base interactions are in the *anti* conformation relative to their sugars.

Even though all the base triples observed are somewhat different from one another, observations of base triples in various DNA structures lend encouragement to the proposal that a triple DNA structure could form for defined sequences and could play an important role not only in the formation of triple-stranded complexes but also in the stabilization of various nucleic acid tertiary structures (Berger *et al.*, 1995).

In the three crystal structures d(TAACCC), d(CCCT) and d(ACCCT), which represent various sequence repeats or truncated variations of the cytosine-rich strand of metazoan telomeres tandem repeats, the hemiprotonated cytosine base pairs which make up the intercalated portion of these three molecules are very similar. However, the variation occurs at the ends of the molecules, where the various adenine and thymine compositions contribute to different structural conformations. In spite of sequence and crystal-system differences, the adenines and thymines located between the consecutive cytosine repeats serve a vital function in contributing to a super-secondary strand polymorphism. Structural similarity of these non-cytosine residues is also observed in the crystal structures of d(TAACCC) and d(ACCCT). Presently, there is no direct biological evidence that the C-rich strand can exist by itself as can the single-stranded G-rich overhang. However, it is tempting to speculate that the higher order structures observed in these G-rich and/or C-rich DNA, whether intrastrand or interstrand, or a combination of the two in the telomere or in the middle of chromosome, may facilitate many biological phenomena by providing stability in the form of a structural intermediate.

This work was supported by ACS Grant PRF #30735-G4. The authors acknowledge the technical assistance from Stephen Fry of Boston Biosystems, Inc.

References

Ahmed, S., Kintana, A. & Henderson, E. (1994). *Nature Struct. Biol.* **1**, 83–88.
Akinrimisi, E. O., Sander, C. & Ts'o, P. O. P. (1963). *Biochemistry*, **2**, 340–344.
Benevides, J. M., Kang, C. H. & Thomas, G. J. (1996). *Biochemistry*, **35**, 5747–5755.
Berger, I., Kang, C. H., Fredian, A., Ratliff, R., Moyzis, R. & Rich, A. (1995). *Nature Struct. Biol.* **2**, 416–425.
Betts, L., Josey, J., Veal, J. & Jordan, S. (1995). *Science*, **270**, 1838–1841.
Blackburn, E. H. (1991). *Nature (London)*, **350**, 569–573.
Blackburn, E. H. (1994). *Cell*, **77**, 621–623.
Brown, T. & Hunter, W. N. (1997). *Biopolymers*, **44**, 91–103.
Brünger, A. T. (1992). *X-PLOR, Version 3.0. A System for Crystallography and NMR.* Yale University, New Haven, CT, USA.
Chen, L., Cai, L., Zhang, X. & Rich, A. (1994). *Biochemistry*, **33**, 13540–13546.
Frank-Kamenetskii, M. D. & Mirkin, S. M. (1995). *Annu. Rev. Biochem.* **64**, 65–95.
Gehring, K., Leroy, J. L. & Guéron, M. (1993). *Nature (London)*, **363**, 561–565.
Hartman, K. A. Jr & Rich, A. (1965). *J. Am. Chem. Soc.* **87**, 2033–2039.
Inman, R. B. (1964). *J. Mol. Biol.* **9**, 624–637.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kang, C. H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. & Rich, A. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 11636–11640.
Kang, C. H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. & Rich, A. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 3874–3878.
Kang, C. H., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. (1992). *Nature (London)*, **356**, 126–131.
Kettani, A., Bouaziz, S., Wang, W., Jones, R. A. & Patel, D. J. (1997). *Nature Struct. Biol.* **4**, 382–389.
Kuryavyi, V. V. & Jovin, T. M. (1994). *Nature Genet.* **9**, 339–341.
Langridge, R. & Rich, A. (1963). *Nature (London)*, **198**, 725–728.
Leroy, J. L., Guéron, M., Mergny, J. L. & Helne, C. (1994). *Nucleic Acids Res.* **22**, 1600–1606.

Liu, K., Miles, H. T., Parris, K. D. & Sasisekharan, V. (1994). *Nature Struct. Biol.* **1**, 11–12.
McEachern, M. & Blackburn, E. H. (1995). *Nature (London)*, **376**, 403–409.
Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281–296.
Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. (1991). *Science*, **254**, 1497–1500.
Radhakrishnan, I. & Patel, D. J. (1994). *Biochemistry*, **33**, 11405–11416.
Rhodes, D. & Giraldo, R. (1995). *Curr. Biol.* **5**, 311–322.
Sinden, R. R. & Wells, R. D. (1992). *Curr. Opin. Biotechnol.* **3**, 612–622.
Sklen'ar, V. & Feigon, J. (1990). *Nature (London)*, **345**, 836–838.
Vlieghe, D., Meervelt, L. V., Dautant, A., Gallois, B., Precigoux, G. & Kennard, O. (1996). *Science*, **273**, 1702–1705.
Williamson, J. R. (1994). *Annu. Rev. Biophys. Biophys. Chem.* **23**, 703–730.
Zakian, V. A. (1995). *Science*, **270**, 1601–1607.