Determinants of Crystal Packing by Nucleic Acid Polymers: Examples of Z-DNA and Cytosine-Tetraplex Crystal Structures

by

Curtis Alan Lockshin

S.B. Life Sciences, Massachusetts Institute of Technology (1986)

Submitted to the Department of Chemistry
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1998

Author...
Department of Chemistry
May 14, 1998

Certified by...
Alexander Rich
William Thompson Sedgwick Professor of Biophysics
Thesis Supervisor

Accepted by...
Dietmar Seyferth
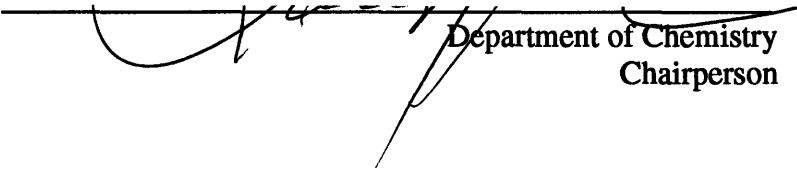Professor of Chemistry
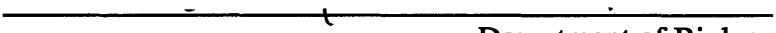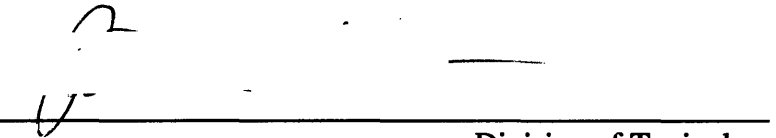Chairman, Departmental Committee on Graduate Students

Science

Thesis Committee

This doctoral thesis has been examined by a Committee of the Faculty as follows:

Professor Lawrence Stern

Department of Chemistry
Chairperson

Professor Alexander Rich

Department of Biology
Thesis Supervisor

Professor Peter Dedon

Division of Toxicology

2

DETERMINANTS OF CRYSTAL PACKING BY NUCLEIC ACID POLYMERS: EXAMPLES OF Z-DNA AND CYTOSINE-TETRAPLEX CRYSTAL STRUCTURES

by

CURTIS ALAN LOCKSHIN

ABSTRACT

Crystallization of nucleic acids requires not only that a particular sequence adopts the desired tertiary structure, but that the molecules are able to organize in a way that supports crystal growth. For nucleic acid molecules, nearest-neighbor interactions that support crystal packing include base stacking, base pairing, and other hydrogen bonding schemes. Two nucleic acid X-ray crystal structures are presented which illustrate solutions to the "crystal-packing problem".

The crystal structure of $d(TCGCGCG)_2$ in $BaCl_2$ was solved to 1.45Å resolution in space group C2. It consists of a Z-DNA $(CGCGCG)_2$ duplex with thymine "overhangs" at each end; the structure illustrates how a molecule can locally violate crystal symmetry using disordered residues to optimize packing interactions.

The crystal structure of d(CCCCACACCCC) was solved to 2.5Å resolution in space group $P4_22_12$. This oligonucleotide crystallizes as an elongated strand; three intersecting perpendicular twofold axes produce four copies of the single strand arranged as a tetraplex with both axial and lateral twofold symmetry. The tetraplex consists of two C_4 "i-motif" domains connected by an ACA knot. Residue A7 is bulged out of each strand; neighboring tetraplexes contribute to A7:A7 base pairs, which in turn stack between adjacent cytosines in the i-motif domains. This structure is of interest because it represents a fragment of the human insulin-linked polymorphic region (ILPR), which is implicated in transcriptional control of the insulin gene; it suggests that an unusual DNA structure may influence transcription as well as variable repeat number polymorphism.

Thesis Supervisor: Alexander Rich
Title: William Thompson Sedgwick Professor of Biophysics

For My Daughter, Elana Rose Lockshin,
and My Grandmother, Edna Weil.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

GENERAL INTRODUCTION

Two of the primary objectives of nucleic acid crystallography are determination of novel structures and cocrystallization of nucleic acids with ligands and proteins. Judicious selection of oligonucleotide sequence can increase the probability of successful crystallization.

Crystallization of nucleic acids requires not only that a particular sequence adopts the desired tertiary structure, but that the molecules are able to organize in a way that supports crystal growth. For nucleic acid molecules, nearest-neighbor interactions that support crystal packing include base stacking, base pairing, and other hydrogen bonding schemes. These interactions are often accomplished by residues which are externalized from the main body of the tertiary structure; these "external" residues may be a naturally occurring element of the motif, or simply residues which are unable to participate in the formation of the main body of the structure. In either case, these residues will be positioned so as to optimize the energetics of nearest-neighbor interactions. This thesis presents two nucleic acid X-ray crystal structures which illustrate solutions to the "crystal-packing problem".

The crystal structure of $d(TCGCGCG)_2$ in $BaCl_2$ was solved to 1.45Å resolution in space group C2; it consists of a Z-DNA $(CGCGCG)_2$ duplex with thymine "overhangs" at each end These thymine residues have no complementary base-pairing partners within the duplex; they must be accommodated elsewhere. One thymine residue reaches to a neighboring duplex, where it hydrogen bonds to a phosphate group; this conformation results in a pocket between the duplexes, where a barium ion coordinates to the phosphate and thymine. The second thymine residue is also turned away from the body of the duplex, where it stacks into a gap between two symmetry-related neighboring duplexes. This thymine adopts one of two conformations (50% occupancy each), in order to produce a thymine:thymine base pair with a symmetry copy of itself, thus taking

7

advantage of two hydrogen bonds as well as base stacking energetics. This structure illustrates how a molecule can locally violate crystal symmetry, using disordered residues to optimize packing interactions.

The crystal structure of d(CCCCACACCCC) was solved to 2.5Å resolution in space group $P4_22_12$. This structure is of interest because it represents a fragment of the human insulin-linked polymorphic region (ILPR), which is implicated in transcriptional control of the insulin gene. The oligonucleotide crystallizes as an elongated strand; three intersecting perpendicular twofold axes produce four copies of the single strand arranged as a tetraplex with both axial and lateral twofold symmetry. The tetraplex consists of two C_4 "i-motif" domains connected by an ACA knot. Residue A7 is bulged out of each strand; neighboring tetraplexes contribute to A7:A7 base pairs, which in turn stack between adjacent cytosines in the i-motif domains. The interaction between A7 and terminal cytosine residues constitute the only interactions which stabilize crystal packing. It may be that the A7 adenines are also an integral part of the motif's natural function.

Insights from these structures may have predictive value in the design of oligonucleotides for crystallization.

CHAPTER ONE

The 1.45 Å X-Ray Crystal Structure of d(TCGCGCG)$_2$

1.1 INTRODUCTION

A striking example of structural variability in DNA is the existence of both right-handed (B-DNA) and left-handed (Z-DNA) duplexes. While the existence of Z-DNA has been established, there has been no definitive proof of its biological role. While much work is focused on the interaction of proteins with conventional B-DNA, there is sparse knowledge of the requirements for recognition of Z-DNA by proteins. An ongoing interest in our laboratory is the cocrystallization of peptide ligands with Z-DNA.

Z-DNA was first described in structural detail in 1979[1]. Since then, a variety of high resolution crystal structures have been solved (for thorough review, see refs.2-4). Z-DNA maintains Watson-Crick base pairing while the phosphate backbone possesses a left-handed helical sense (Figure 1); in order to accommodate this, ribose pucker and phosphate torsions are changed. As a net result, the base pairs in Z-DNA are displaced from the helical axis as compared to B-DNA; the major groove in B-DNA becomes a convex surface in Z-DNA, while the B-DNA minor groove becomes narrow and deep upon transition to Z-DNA.. Z-DNA helices have an average helical rise of ~3.7Å and a width of ~20Å, as compared to a 3.4Å rise and 24Å width for B-DNA.

In contrast to B-DNA, a Z-DNA helix is built by rotation/translation of a dinucleotide repeat, with 12 base pairs constituting one full turn of the helix. For each step along the strand, helical twist (Θ) alternates between ~-8.5° and ~-51°; thus the average twist per base is ~-30°, as compared to $\Theta = 36°$ for B-DNA. In order to support Watson-Crick base pairing, residues along a strand must alternate between *anti* and *syn* glycosyl rotations. Therefore, the fundamental subunit of a Z-DNA duplex can be described as a 5'-*anti*-p-*syn*-3' dinucleotide with a twist of ~-8.5° (Figure 2a); consecutive dinucleotide subunits are joined with a 5'-*syn*-p-*anti*-3' twist of ~-51° (Figure 2b).

Of particular importance in this study is the ocurrence of a second class of inter-subunit phosphodiester geometry. While the conformation in Figure 2b, referred to as Z_I, is more common, the Z_{II} conformation (Figure 2c) is frequently observed. Z_{II} phosphodiester geometry is frequently correlated with ion binding or crystal packing interactions.

As a result of the differing inter-and intra- subunit twists, two types of base stacking are observed within the Z-DNA duplex, as illustrated in Figure 3.

The oligonucleotide d(CGCGCG) crystallizes readily as one half-turn of Z-DNA, in a tightly packed crystal lattice. While this arrangement yields fruitful high-resolution data, it presents a barrier to cocrystallization with large ligands. Data from our laboratory and others[5] indicated that peptides of sequence $(Lys\text{-}X\text{-}Lys)_n$ can influence the B-DNA/Z-DNA equilibrium of poly d(CG) in solution. We were therefore interested in cocrystallizing several such peptides (e.g. KAKAK, KGKGK) with Z-DNA. In an effort to "perturb" the Z-DNA lattice, we introduced various additions to the 5' end of synthetic d(CGCGCG) oligomers. While no cocrystals were found, the oligomer $d(TCGCGCG)_2$ crystallized alone as a Z-DNA hexamer duplex $((CGCGCG)_2)$ with "flipped out" thymine residues. One thymine residue is located near a twofold rotation axis and is disordered; it occupies one of two conformations so that symmetry mates across the twofold axis can form two hydrogen bonds. In addition, this thymine:thymine base pair is stacked between the ends of two symmetry-related duplexes. A second thymine residue is hydrogen bonded to a neighboring helix, forming a pocket which traps a barium ion. Another barium ion mediates crosslinking of adjacent duplexes via coordination to the sugar-phosphate backbone and the Z-DNA major groove.

1.2 MATERIALS AND METHODS

The sequence d(TpCpGpCpGpCpG) was produced on an automated synthesizer by the M.I.T. Biopolymers laboratory and purified by HPLC. Crystals grew from a solution containing: 5%(v/v) 2-methyl-2,4-pentanediol (MPD), 50 mM sodium cacodylate at pH 7, 5 mM $BaCl_2$, 4mm spermine, and 2 mM DNA (single-strand concentration). Hanging drops were equilibrated against a reservoir of 30% MPD; crystals grew after several weeks as elongated rods (0.2 mm x 0.2 mm x 0.6 mm).

Crystals grew in a monoclinic lattice of the space group C2. Unit cell dimensions are : a=58.844 Å, b=21. 716 Å, c=32.390 Å, α=90°, β=121.54°, γ=90° The following data sets were collected: complete low-resolution (3 Å) data on an AFC5-R rotating anode diffractometer, and a high-resolution (1.45 Å) data set (Rsym = 3%) collected on a MAR imaging plate (MAR research) at Brookhaven National Laboratory. Diffractometer data was used for molecular replacement. The structure was refined using the synchrotron data, in which a total of 5976 (of 6501 possible reflections to 1.45 Å) were observable at 2σ (F) or greater.

The structure was solved by molecular replacement using XPLOR[6], with a Z-DNA hexamer duplex d(CG)$_3$ as a search model. Rotation searches indicated a family of solutions which would place the helical axis along the unit cell diagonal, consistent with the base stacking direction evident from a Patterson map.

A favorable translation solution was found; this produced a reasonable map with strong residual density (between neighboring helices) that was assigned to a barium ion. At this point, no amount of refinement could reduce the 2.5 Å R -factor below ~43%. The molecule was then modelled as 37 rigid bodies (1 barium ion, plus each residue broken into phosphate, sugar, and base). Rigid body refinement yielded a 2.5 Å R-factor of 36%, and an excellent map into which the sugar-phosphate backbone could be fitted. Ideal geometry between the rigid bodies was restored without loss in R-factor. At this

point difference and 2Fo-Fc maps clearly showed the position of one thymine residue (residue THY11), as well as other residual density which was subsequently assigned as a partially occupied barium ion and a partially occupied (second) thymine residue (THY1). The model was adjusted accordingly and refined (simulated annealing, positional refinement, B-factor and occupancy refinement). At the late stages of refinement, it became apparent that residue THY11 occupied two distinct positions (50% occupancy each), in order to produce a THY11:THY11 "base pair" between symmetry related thymines. Furthermore, residue THY1 appeared to have some degree of disorder; a predominant location was eventually determined. The final structure also contains 39 water molecules. Final refined occupancies will be discussed in the Results section.

The final 1.45 Å R-factor is 19.3%, with bond and angle root mean square deviations (RMSD) of .011 Å and 2.2 degrees, respectively. A free R-factor based on a random 10% subset of reflections is 23.3%.

1.3 RESULTS AND DISCUSSION

Structure of the Heptamer Duplex

The Asymmetric Unit

The asymmetric unit (Figure 4) of this structure can be described as a Z-DNA hexamer duplex comprised of d(CG)$_3$, with an "overhanging" thymine residue at each end. Both thymine residues interrrupt the continuity of the Z-DNA backbone, as they are turned out from the helical axis. At one end, residue THY11 (heretofore referred to as T_{in}) is turned so as to stack over an adjacent helix; it is systematically disordered such that it forms a T_{in}:T_{in} base pair with a symmetry-related T_{in} across a twofold rotation axis. Residue THY1 (T_{out}) is released toward the solvent in order to accomodate the spatial constraints imposed by neighboring molecules; it is largely disordered, although one position (shown) seems to predominate. In addition, there are two barium ions. One of these ions, BA17, is fully occupied and mediates a hydrogen bond network between adjacent helices. The second barium ion, BA18, has partial occupancy and appears to be associated with T_{out}.

The Z-DNA Core

Several features are evident in Figure 5, which shows the central Z-DNA core of this structure as compared with the original hexamer search model. The backbone is most perturbed in the CpG steps nearest to the 5' Thymines; the CpG step proximal to T_{out}, namely C2(p)G3, is overtwisted (in the left-handed sense) by ~5 degrees. Similarly, CpG step C12(p)G13, proximal to T_{in}, is undertwisted by ~2 degrees. In general, the Z-DNA core is overtwisted toward T_{out} and undertwisted toward T_{in}. The net result is a half turn of Z-DNA duplex whose net twist is nearly that of an unperturbed d(CG)$_3$ Z-DNA

hexamer. However, The rise/residue of the central hexamer is slightly longer than average, giving rise to a slightly (~0.8 Å) longer duplex.

Of particular importance is the fact that two phosphodiesters, P4 and P14 (between G3(p)C4 and G13(p)C14, respectively), conform to Z_{II} parameters. As will be shown, P4 from one duplex and P14 from a neighboring duplex are coordinated to the barium ion BA17.

Arrangement of THY11 (T_{in})

In Figure 6, the step from THY11(T_{in}) to CYT12 is compared to an internal GpC step (Z_I conformation) from the central helix. Phosphodiester torsions around the THY11-(P12)-CYT12 linkage are modified such that T_{in} is turned about 80° from the path which would follow the propagating Z-DNA duplex.

Electron density maps indicate no disorder at the P12 phosphate group. however, residue THY11 adopts one of two positions determined by the sugar pucker; each position refines to 50% occupancy. The rationale for this arrangement is evident in Figure 7. Adjustment of pseudorotation angle (Ψ) allows two symmetry-related T_{in} residues to form two hydrogen bonds (N3-O4* and N3*-O2) across a twofold rotation axis. Because of the twofold symmetry restriction, no single thymine position could accommodate this hydrogen bonding scheme. The justification for this part of the model is derived from the density map, shown in Figure 8. Figure 8a is representative of the 1.45Å 2Fo-Fc map around residues of the central duplex; this clearly differs qualitatively from the density calculated around THY11(T_{in}), when T_{in} is modelled at a single position (Figure 8b). However, maps calculated from the superposition of two half-occupied conformations of T_{in} (Figure 8c-d) are indicative of the this model's validity.

Arrangement of THY1 (T_{out}) and Barium Ion 18 (BA18)

A simulated annealed omit map (Fo-Fc) (Figure 9) around THY1 clearly shows the path of the sugar-phosphate backbone from CYT2. However, density for the pyrimidine ring is less compelling. Furthermore, there is a region of strong residual density in the neighborhood which was modelled as a barium ion (BA18). One particular conformation for the THY1 pyrimidine seems consistent with weak density contiguous to the sugar moeity; this arrangement allows for hydrogen bonding of THY1 to BA18 and to a phosphate oxygen on a neighboring helix. Furthermore, this position accommodates a hydrogen bond between the 5' terminal-OH and and GUA6*-O6 from a neighboring duplex (see Figure 12b). This conformation for THY1 results in a backbone twist of ~170° (Figure 6). The sugar pucker of this configuration is C2'-endo.

Interestingly, THY1 and BA18 independently refine to the same occupancy (~37%), suggesting a relationship between the two. 2Fo-Fc maps in this general area are noisy and care was taken not to overinterpret the map by overaddition of solvent, or by invocation of additional conformations for THY1. Figure 10 illustrates the 2Fo-Fc electron density in the region around THY1.

The BA18 binding pocket is discussed in the next section.

Interaction of Neighboring Duplexes

Base Interactions and Stacking

The unit cell contents are shown in Figure 11a. Crystal packing results in neighboring continuous "columns" of DNA. Continuity is achieved when an end of the molecule encounters a twofold rotation axis; the column is then extended by the symmetry operation. Each end of the duplex encounters a [different] twofold axis, producing two different results. Figure 11b also details of one type of continuity.; in this case, the terminal CYT12:GUA7 base pair of the central Z-DNA hexamer is stacked

directly with its symmetry-related counterpart (CYT12*:GUA7*). This is possible because residue THY11 (T_{in}) turns away from the helical axis and into a gap in an adjacent column.

The other end of the hexamer Z-DNA core is about 3.7Å from a twofold axis (x, y+1/2) Consequently, there is an ~7.5Å gap between (symmetry related) CYT2:GUA17 base pairs into which a base pair can be stacked (Figure 12). This "base pair" is the interaction between symmetry-related T_{in} residues from two neighboring columns, as shown in Figure 12a. Combining the alternate conformations of T_{in} with the twofold rotation, two different stacking patterns are observed between T_{in} and the two symmetry-related CYT2:GUA17 base pairs (Figure 12b).

Complex with Barium Ion 17 (BA17)

BA17 (Barium ion, residue 17) occupies a pocket formed by two adjacent duplexes (Figure 13). Ba^{2+} is capable of accommodating 7 to 9 ligands; in this case, it likely coordinates to four water molecules as well as the following atoms: G15-O6 ,G15-N7 and P14-O2 (Z_{II} geometry) from one helix, and G5-O6 and P4-O2 (Z_{II} geometry) from the neighboring helix. Figures 14 & 15 show in detail the pocket around BA17.

BA17, its four coordinated waters, and two additonal water molecules participate in a belt of solvent which runs through the crystal. As shown in Figure 13, crystal symmetry results in a chain of solvent which propagates between helices (along the *b* axis).

THY1(T_{out}) forms a pocket with a neighboring duplex, which coordinates Barium Ion 18 (BA18)

Figure 13 also illustrates that T_{out} from one helix forms a pocket with the adjacent helix that is crosslinked via BA17. The hydrogen bonding scheme between T_{out}, BA18, and a neighboring phosphate is shown in Figure 16.

1.4 FIGURE LEGENDS

Figure 1. *upper:* Stereo view of d(CGCGCG)$_2$ as a half-turn of Z-DNA duplex. At right is a view down the helix, showing that the base pairs are displaced from the central axis (+) of the duplex. As a result, the (Watson-Crick) major groove becomes a convex surface, while the minor groove becomes narrow and deep. *lower:* Stereo view of d(CGCGCG)$_2$ as B-DNA. At right, axial view showing that base pairs are centrally located with respect to the helical axis (+).

Figure 2. *(a)*Stereo view of a dinucleotide subunit (in this case CpG) from which a Z-DNA helix is built. Within this subunit, nucleotide geometry is 5'-*anti*-p-*syn*-3' with a twist of ~-8.5°. *(b)* Stereo view of a 5'-*syn*-p-*anti*-3' dinucleotide (in this case GpC). This represents an inter-subunit junction; contiguous subunits along the strand are related by a twist of -51°. Phosphodiester geometry in this example is referred to as "Z$_I$". *(c)* Stereo view illustrating two possible geometries for the inter-subunit phosphodiester. The "Z$_{II}$" configuration (grey) is less common but has been observed in previous Z-DNA crystal structures, usually as a result of ion binding or crystal packing.

Figure 3.*(a)* Stereo view showing two contiguous CpG (5'-*anti*-p-*syn*-3') dinucleotide subunits, including the complementary strand. *(b)* Axial view showing intra-subunit *(upper)* and inter-subunit *(lower)* base-stacking patterns. In the former case, while there is little stacking of purine-pyrimidine ring systems, the purine is stacked with the neighboring ribose system; this presumably contributes to stabilization of the *syn* conformation.

Figure 4. The asymmetric unit of d(TCGCGCG)$_2$ consists of a Z-DNA (CG)$_3$ duplex with overhanging thymines at each end. THY11 (T$_{in}$) adopts two discrete conformations,

each with an occupancy of 50%. THY1 (T_{out}) is apparently disordered, but one position (shown) predominates with an occupancy of 37%. Barium ion BA18 is associated (via hydrogen bonding) with T_{out} and a neighboring duplex. Interestingly, the occupancy of BA18 is also 37%. Barium ion 17 (BA17) participates in a hydrogen bond network between two adjacent duplexes as well as a network of water molecules. BA17 has 100% occupancy.

Figure 5. The central core of d(TCGCGCG)$_2$ (black) compared to the original search model, a Z-DNA d(CGCGCG)$_2$ duplex (light grey). Overall, the central (CG)$_3$ core adopts an unremarkable Z-DNA conformation. The most important distinction is in the phosphodiester geometries at GUA13(p)CYT14 and GUA3(p)CYT4. These two linkages possess Z_{II} geometry; as will be shown (Figs 10-11), these two phosphate groups contribute to the coordination of a barium ion (BA17).

Figure 6. Propagation of the Z-DNA helix is interrupted at the thymine residues. *Central figure*: 5' Elongation of each strand via standard Z-DNA helix parameters is shown in light grey. *Left*: Detail of the THY11(p)CYT12 step compared to a Z_I GpC step (light grey). The net result is an ~80° turn in the path of the backbone, away from the helix' axis. *Right*: Detail of the THY1(p)CYT2 step compared to a Z_{II} GpC step (light grey); the result is an ~170° turn in the path of the backbone.

Figure 7. Base-pairing rationale for the systematic disorder of THY11 (T_{in}). Alternate conformations T_{in} and $T_{in}*$ (solid model) meet at a twofold rotation axis (↑). Application of strict symmetry to either conformation (thin lines) would result in either van der Waals collision (in the case of T_{in}:T_{in}) or a separation too great for hydrogen bonding ($T_{in}*$:$T_{in}*$). Adjustment of pseudorotation angle Ψ permits the formation of two hydrogen bonds (N3-O2* & N3*-O2) between two symmetry-related, alternately

conformed THY11 residues. Ψ values are 183° (C3'-endo-C2'-exo)and 215° (C3'-endo-C4'-endo), for T_{in} and T_{in} *, respectively. One would expect that, by necessity, each conformation would be 50% occupied; refinement proved this to be the case. A water molecule (grey sphere) is located on the twofold axis ($\uparrow$). By definition, this water has 50% occupancy; it is hydrogen bonded to the 50% occupied T_{in}*.

Figure 8. Density map evidence for the dual conformation of THY11. *(a)* 1.45 Å 2Fo-Fc density around an internal cytosine (CYT4), contoured at 1σ. This is an indication of the map quality around all non-thymine residues. *(b)* 1.45 Å 2Fo-Fc (contoured at 1σ) density around THY11, when it is modeled as a single conformation. Given the potential for detail evident in *(a)*, this part of the map is unsatisfying in terms of map quality and goodness-of-fit of the model itself. *(c)* 1.45 Å 2Fo-Fc (contoured at 1σ) density around THY11, when it is modeled as two alternate conformations, 50% occupied each. *(d)* 1.45 Å 2Fo-Fc (contoured at 0.7σ) map around the "two-conformation" model for THY11. Overall, this model allows for a better fit of the sugar moeties to the map; furthermore, the density around the pyrimidine ring can now be rationalized as the superposition of these two states.

Figure 9. Evidence for the position for THY1 Fo-Fc (1.5σ) omit maps around residue THY1. The map indicates only one compelling alternative for the path of the sugar. However, omit density around the ring is less clear; lack of sufficient coverage and clarity suggests disorder of T1, leading to < 100% occupancy at this position.

Figure 10. 1.45 Å 2Fo-Fc density around residues THY1, CYT2, and BA18. *(a)* 1σ contour (grey), showing the density calculated around THY1, as compared to density for CYT2. *(b)* 0.7σ contour (grey). In each figure a 5σ contour is shown in black. The only

atoms that appear at 5σ are BA18 and CYT2-P; it is largely for this reason that the peak was assigned as a barium ion.

Figure 11. The unit cell containing four asymmetric units. *(a)* view of the *ac* plane. Helices are aligned with the shorter *ac* diagonal; crystal packing results in "infinite" adjacent columns of DNA (see inset left, view along *ac* diagonal). Duplexes Z and Z' are related by the twofold rotation axis (x+1/2,z+1/2) (◆) .Symmetry-related sets of CYT12:GUA7 base pairs stack directly upon each other, continuing the column; T_{in} is turned to stack over an adjacent duplex (grey) near the twofold rotation axis (x+1/2,z) (see Fig. 11). *(b)* Detail of the twofold rotation axis (x+1/2,z+1/2).

Figure 12. Environment around the twofold rotation axis (x+1/2,z). *(a)* Duplexes Z and Z" (black) are related by the twofold axis (◆); this axis also gives rise to the T_{in}:T_{in}* base pair from two symmetry-related neighbors (grey). The stacking distance between T_{in} and Z (or Z") is ~3.6Å. *(b)* *(upper)* stacking of duplex Z (black) over a T_{in}:T_{in}* base pair (grey). *(lower)* stacking of the T_{in}:T_{in}* base pair (black) over duplex Z" (grey).

Figure 13. Stereo view of neighboring duplexes. Duplex Z stacks its T_{in} residue over duplex Z'; T_{out} from Z' is turned toward the viewer. At the lower half of the figure, residue T_{out} from duplex Z turns to form a pocket (↑) with duplex Z'; this pocket is able to coordinate a barium ion, BA18 (see Fig 15). Barium ions BA17 (large spheres) crosslink Z and Z'; in addition, each BA17 coordinates four water molecules (small spheres), contributing to a continuous "belt" of solvent which runs between helices in the crystal.

Figure 14. Details of Barium ion coordination. *(upper)* stereo view showing the environment around BA17 (large sphere). BA17 coordinates to four water molecules

(small spheres) and to five atoms (dark grey) on two neighboring strands Z and Z',
namely, CYT4-O2P, GUA5-O6, CYT14'-O2P,GUA15'-O6,and GUA15'-N7. *(lower)*
stereo view showing the hydrogen bonding scheme between THY1 (strand Z), P13
(strand Z'), and barium ion BA18. Paricipating atoms are shown in dark grey (see Figure
15 for detail). Electron density in this area is of insufficent quality to determine solvent
coordination by BA18.

Figure 15. Hydrogen-bond distances between barium ion BA17 and its 9 potential
ligands. *(upper)* distances between BA17 and DNA heteroatoms. *(lower)* distances
between BA17 and its four coordinated water molecules.

Figure 16. Hydrogen bonding scheme which stabilizes the arrangement of THY1 (T_{out})
with barium ion BA18 and a neighboring helix (Z').

1.5 REFERENCES

1. Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., Crawford, J. L. ,van Boom, J. H., van der Marel, G. , Rich, A. (1979) *Nature* 282:680-686

2. Rich, A., Nordheim, A., Wang, A. H.-J. (1984) *Ann. Rev. Biochem.* 53:791-846

3. Jovin, T. M., Soumpasis, D. M. (1987) *Ann. Rev. Phys. Chem.* 38:521-60

4. Moeers, B.H.M., and Ho, P.S. (1997) *Biopolymers:Nucleic Acid Sciences* **44**, 1-2:65

5. Takeuchi, H., Hanamura, N., Hayasaka, H., Harada, I. (1991) *FEBS Letters* **279**, 2:253

6. Brunger, A.T. (1992) *XPLOR 3.1: A System for X-ray Crystallography and NMR Yale University Press*
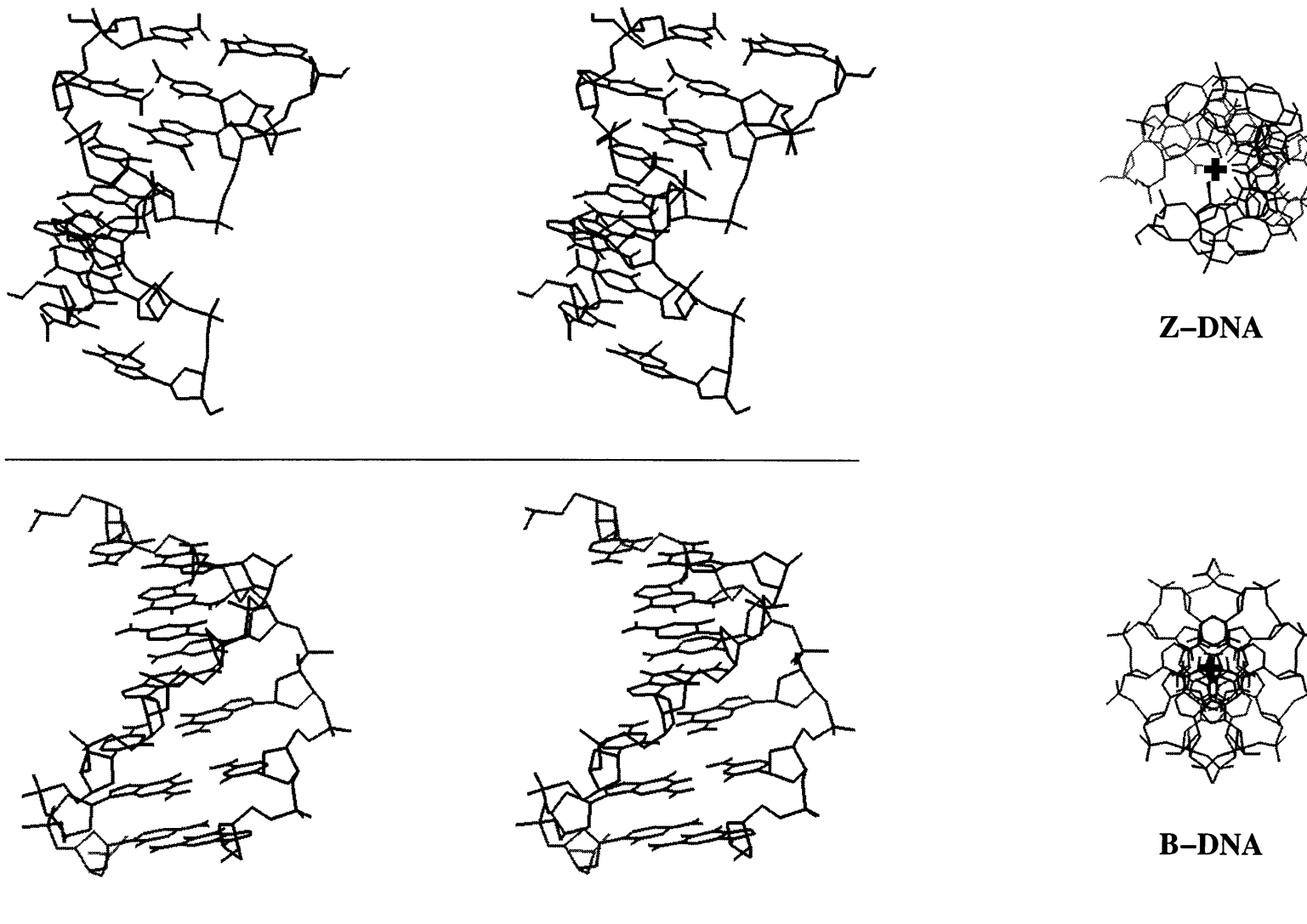
Z–DNA

B–DNA

Figure 1. *(upper)* Stereo view of d(CGCGCG)$_2$ as ½–turn of Z–DNA.
(lower) Stereo view of d(CGCGCG)$_2$ as B–DNA. Axial views at right.
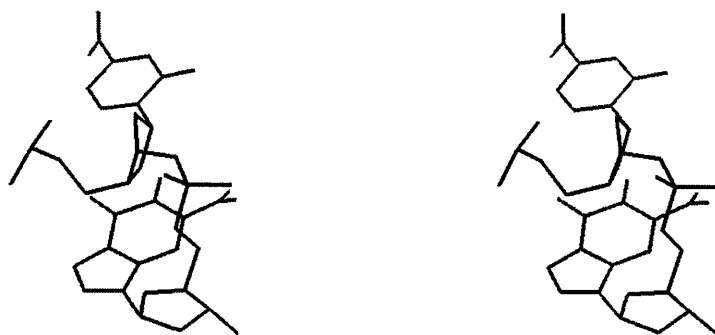
(a) intra–subunit **(CpG)**
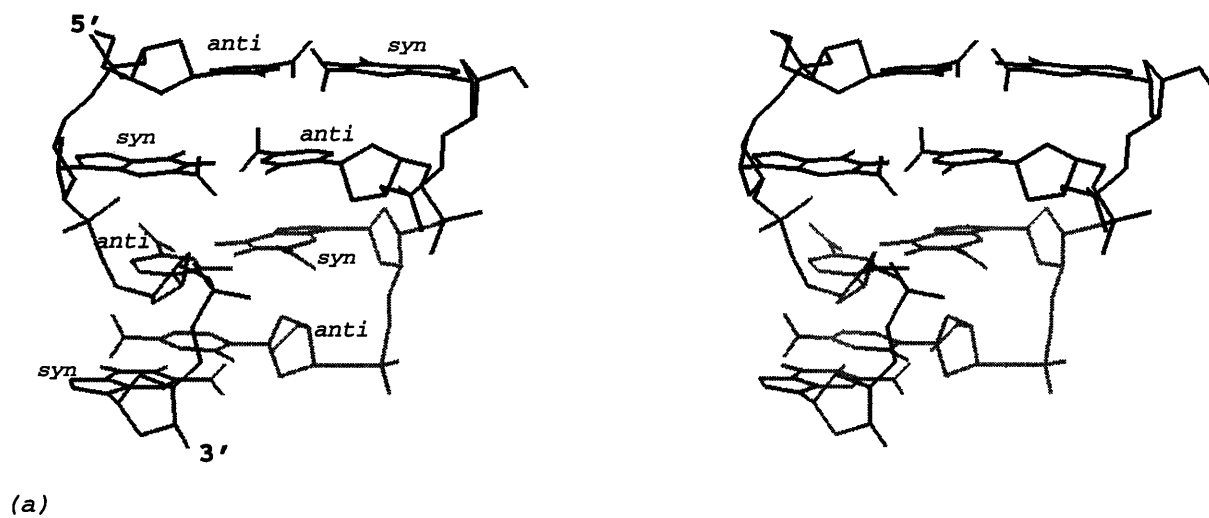

Z_I
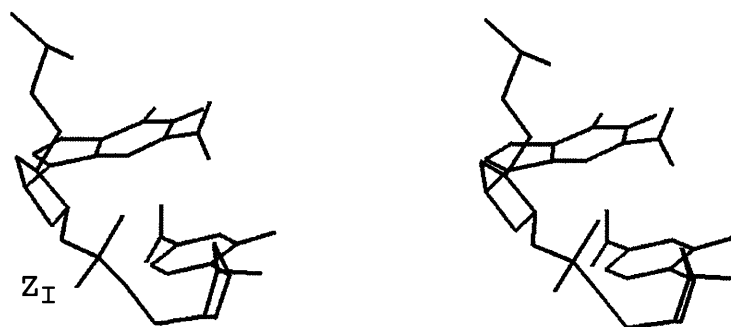
(b) inter–subunit **(GpC)**

Z_{II}
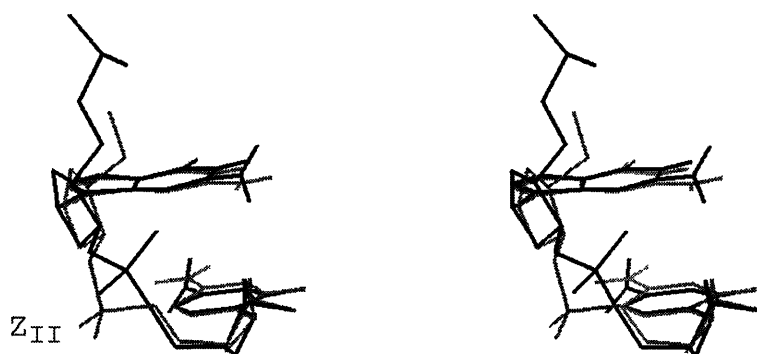
(c) inter–subunit **(GpC)**

Figure 2. *(a)* Stereo view of a CpG dinucleotide repeat subunit . *(b)* Stereo view of a GpC intra–subunit junction.

(a)

5'-anti-p-syn-3'
(CpG)

5'-syn-p-anti-3'
(GpC)
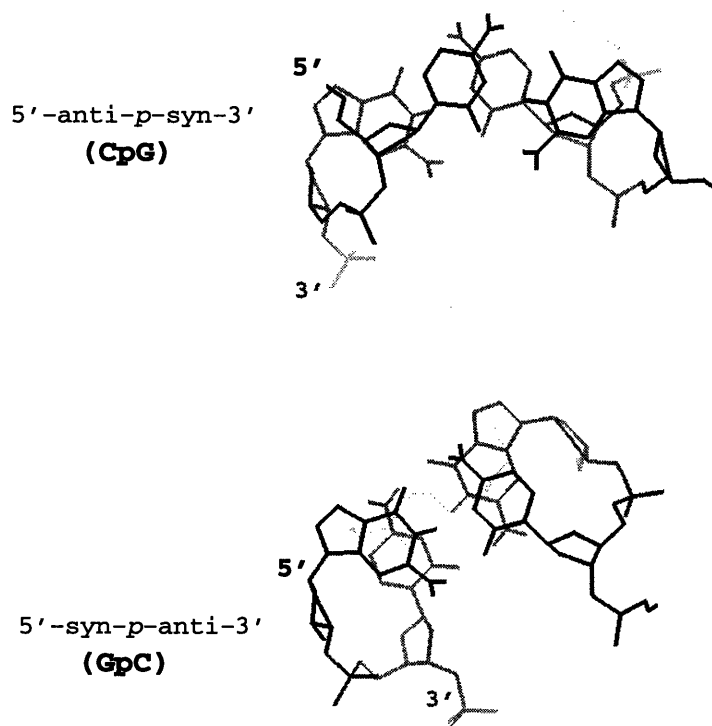
(b)

Figure 3. *(a)* Stereo view of two contiguous CpG dinucleotide repeat subunits *(b)* intra– and inter–subunit stacking patterns .
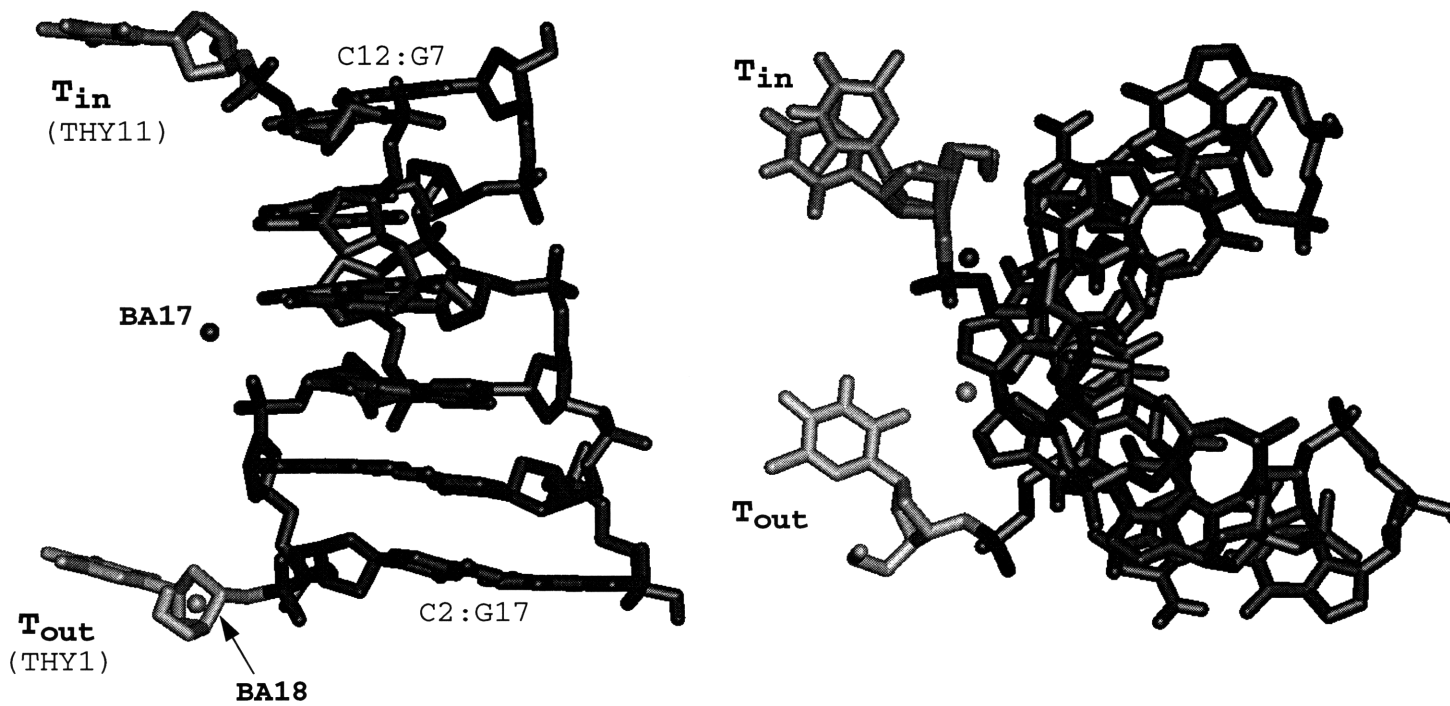
Figure 4. The asymmetric unit of d(TCGCGCG)$_2$, including two Barium ions.

T$_{in}$

ZII geometry
at G13(p)C14

T$_{out}$

T$_{in}$

ZII geometry
at G3(p)C4

T$_{out}$
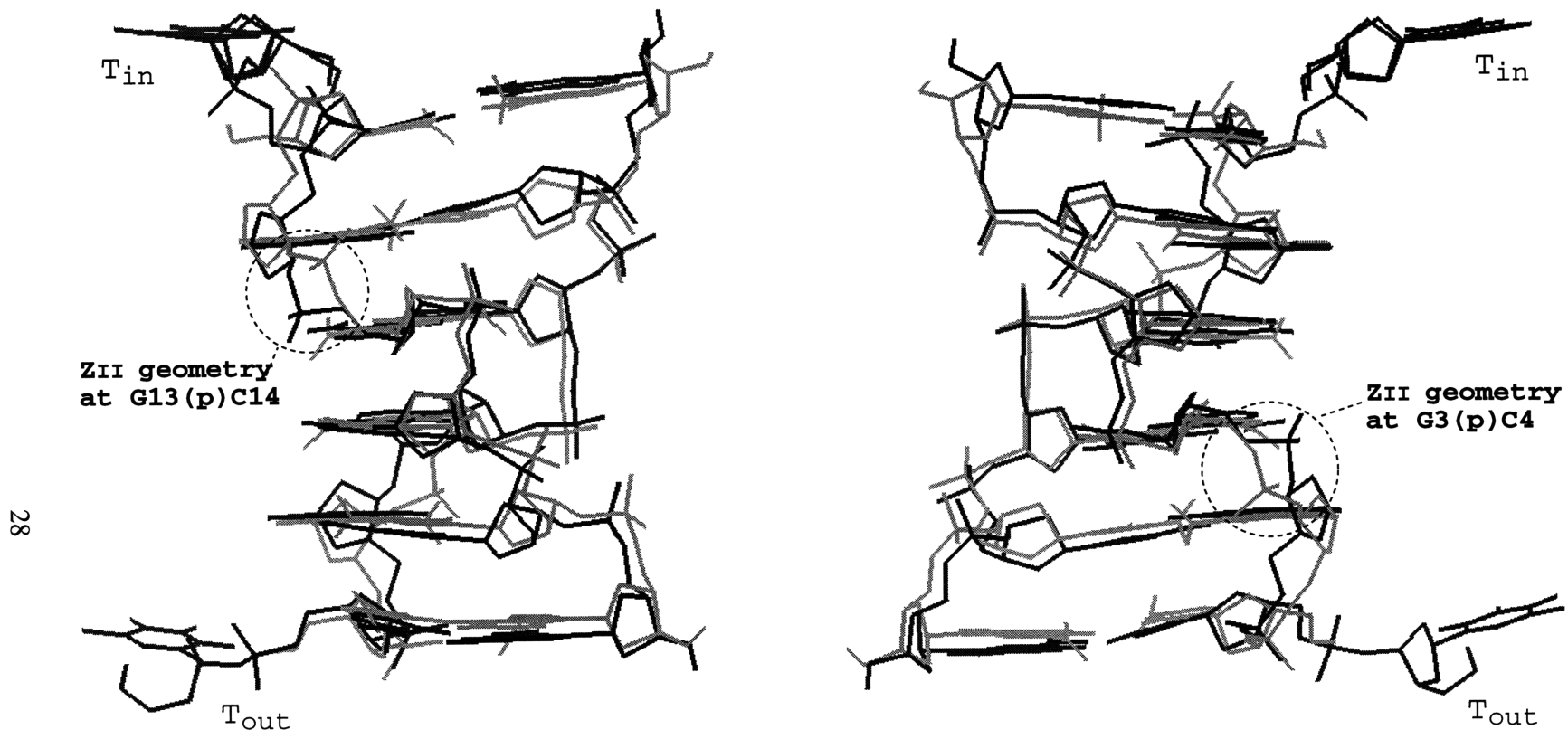
Figure 5. The Z–DNA hexamer core of d(TCGCGCG)$_2$ compared to the Z–DNA structure of d(CGCGCG)$_2$.

Internal GpC
(Z$_{\text{I}}$ Phosphate
Torsions)

THY11

C12:G7

G13:C6

G3:C12

C2:G17

THY1

Internal GpC
(Z$_{\text{II}}$ Phosphate
Torsions)
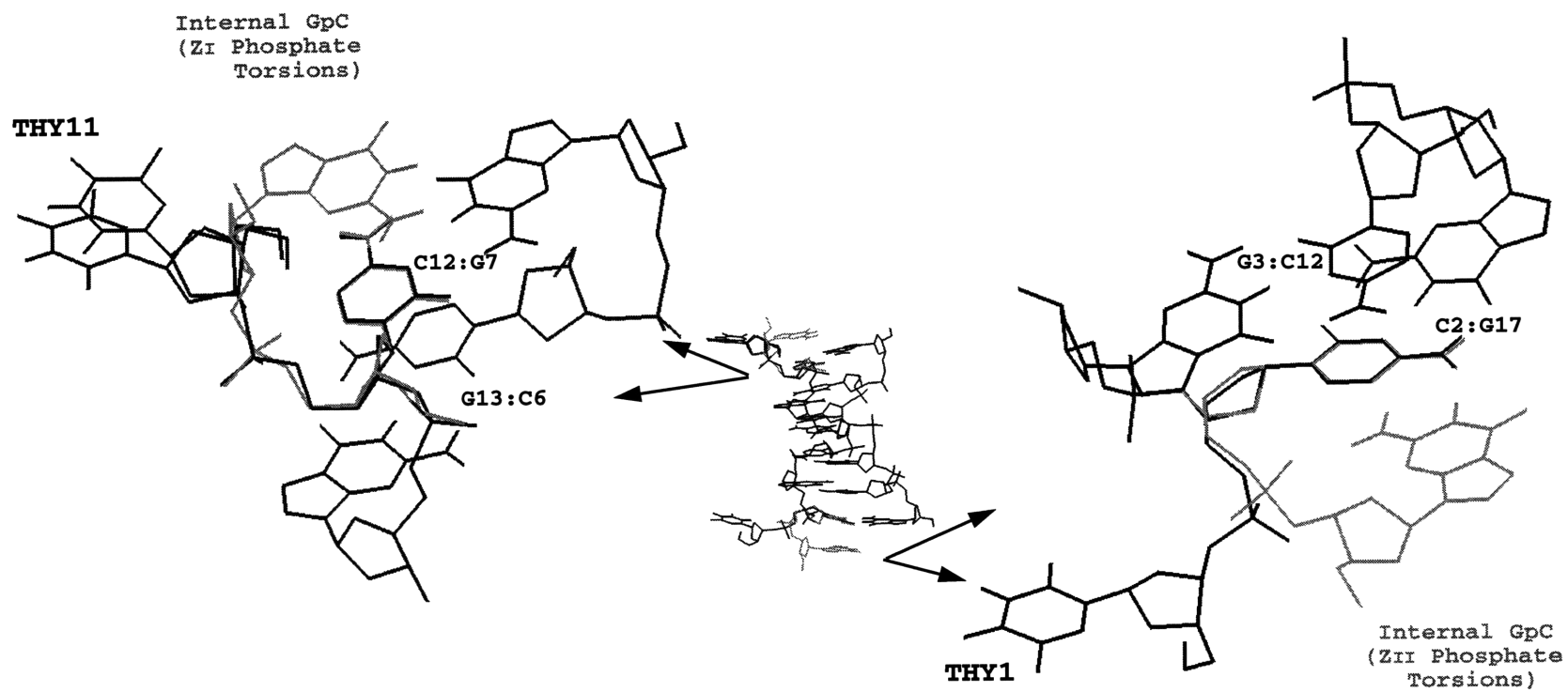
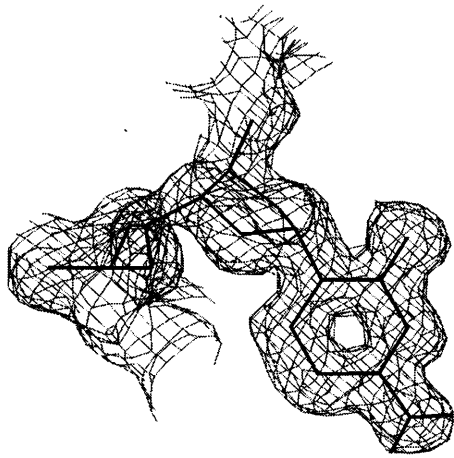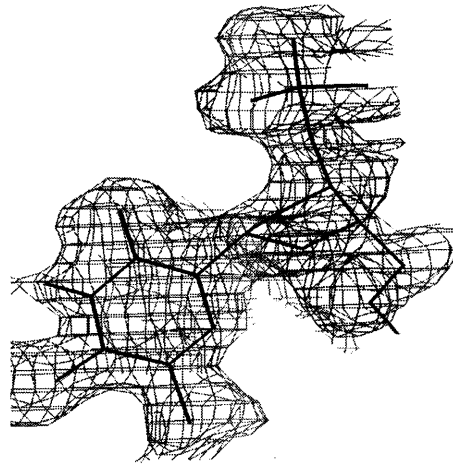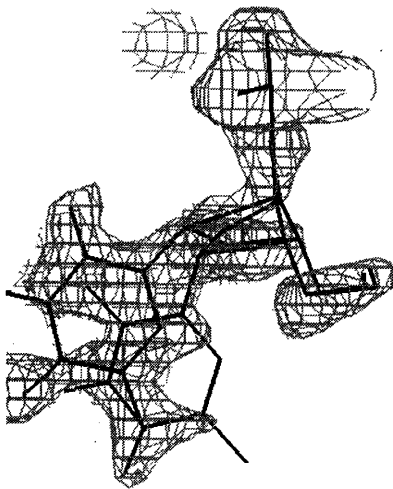**Figure 6.** The regularity of the Z–DNA backbone is interruped at the terminal thymine steps.
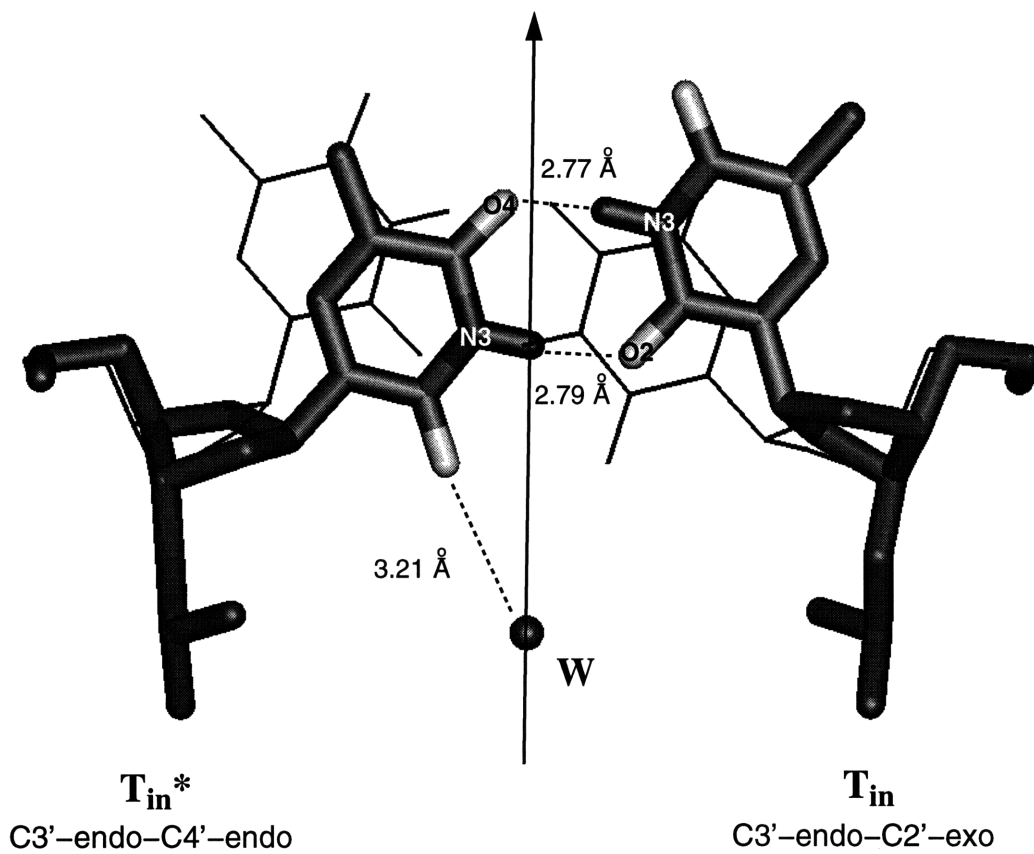
Figure 7. Systematic disorder localized at a twofold rotation axis permits the formation of two hydrogen bonds between two symmetry–related THY11 (T_{in}) residues.
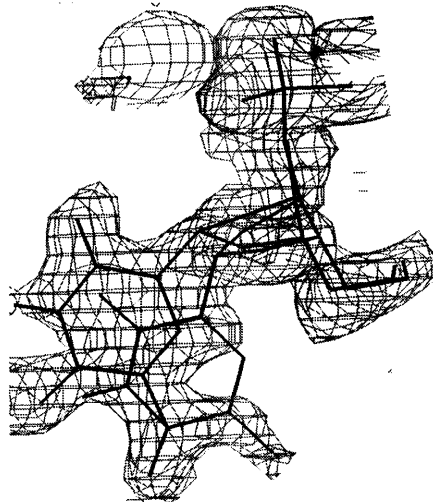
*(a)***CYT4** 2Fo-Fc, 1σ contour

*(b)***THY11 single conformation**
2Fo-Fc, 1σ contour

*(c)***THY11 two conformations**
2Fo-Fc, 1σ contour

*(d)***THY11 two conformations**
2Fo-Fc, 0.7σ contour

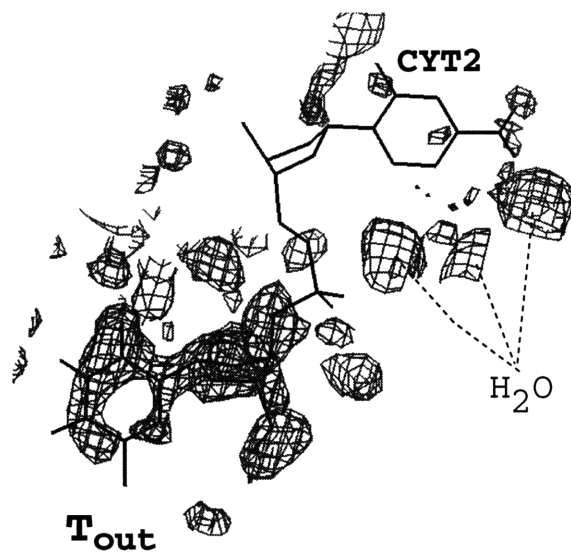**Figure 8.** Density maps suggest a dual conformation for residue THY11 (T_{in}).

31
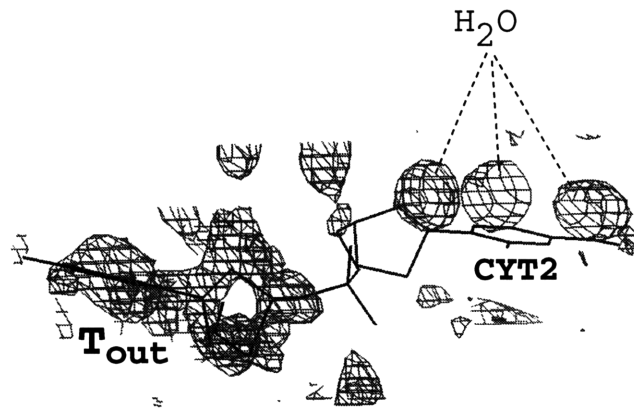
Figure 9. Simulated annealed omit (Fo−Fc) map around residue THY1 (T$_{out}$).
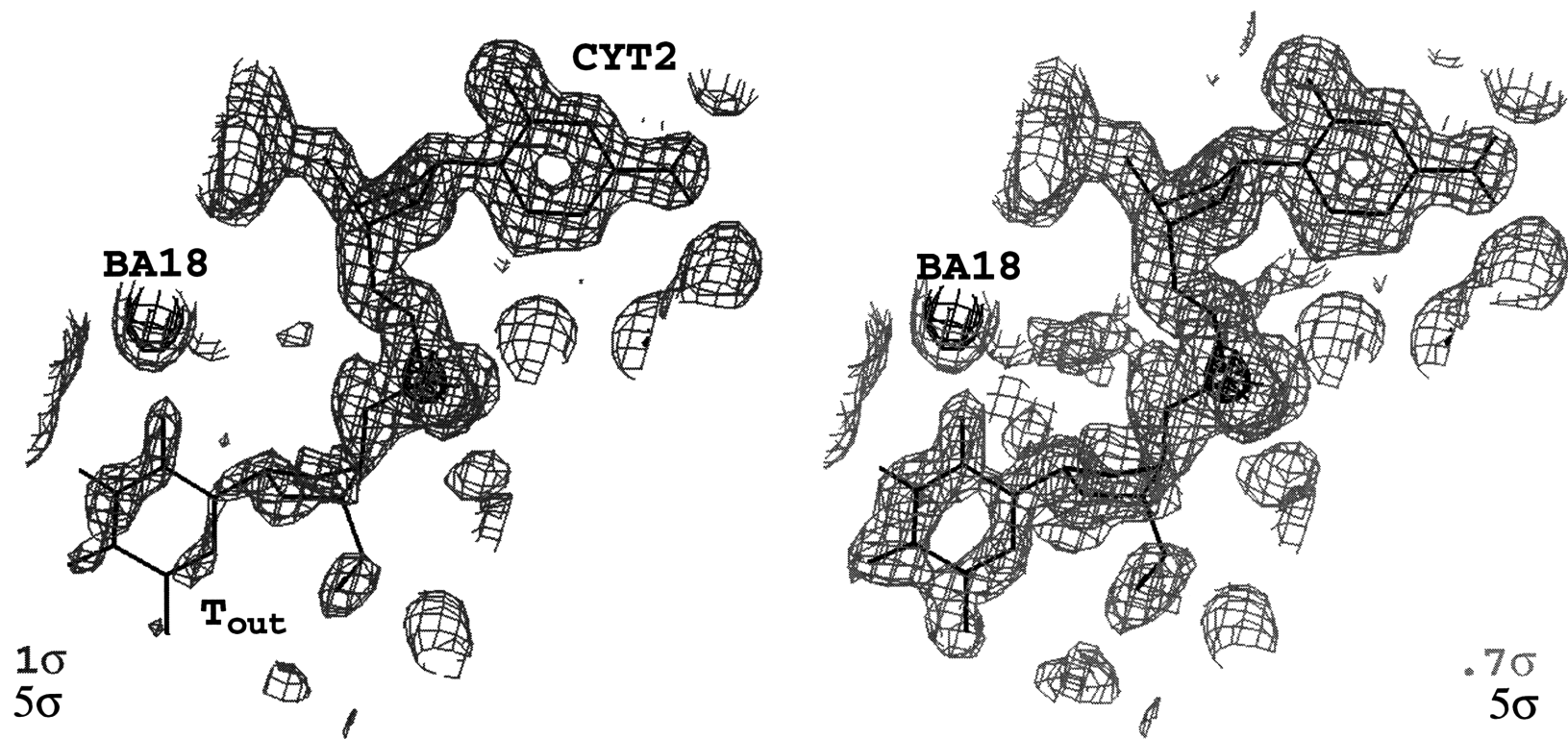
Figure 10. (2Fo−Fc) maps around residues THY1 (T_{out}), CYT2, and barium ion BA18.

(a)

(b)

G7:C12
C12:G7

Z

T_in

T_in

Z'

T_out

T_in

Z

T_in

Z'

T_out
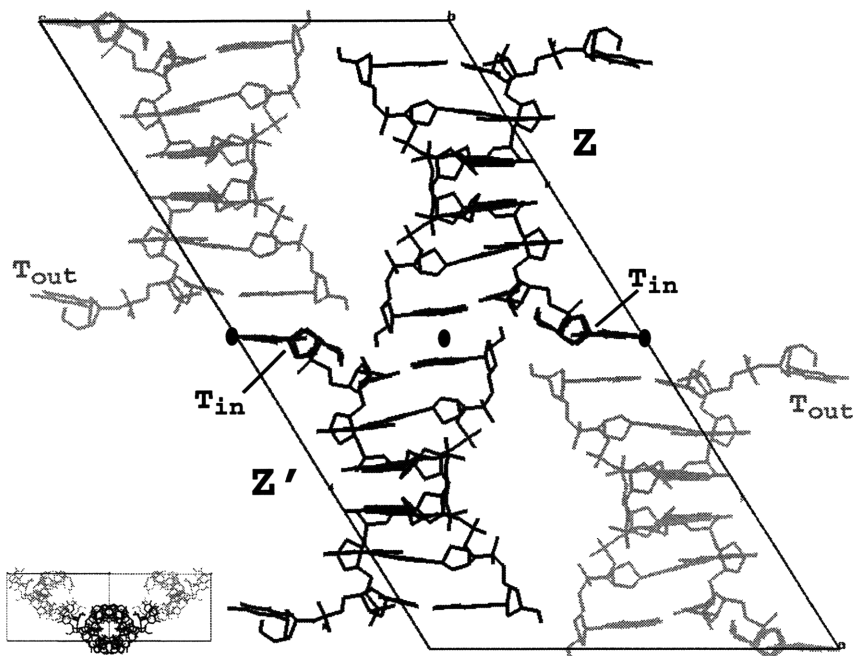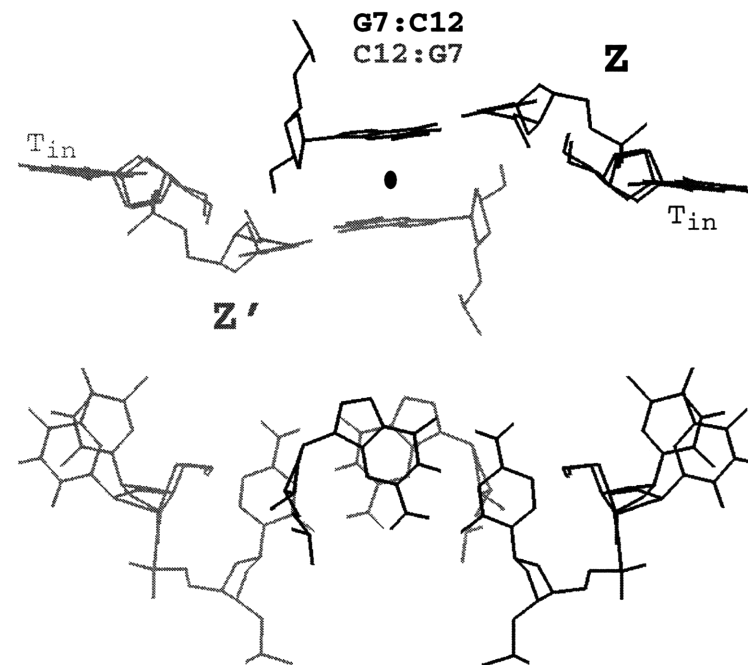
Figure 11. *(a)* View of the unit cell containing four duplexes. *(b)* stacking of symmetry−related duplexes at the twofold axis (x+1/2,z+1/2) (●).
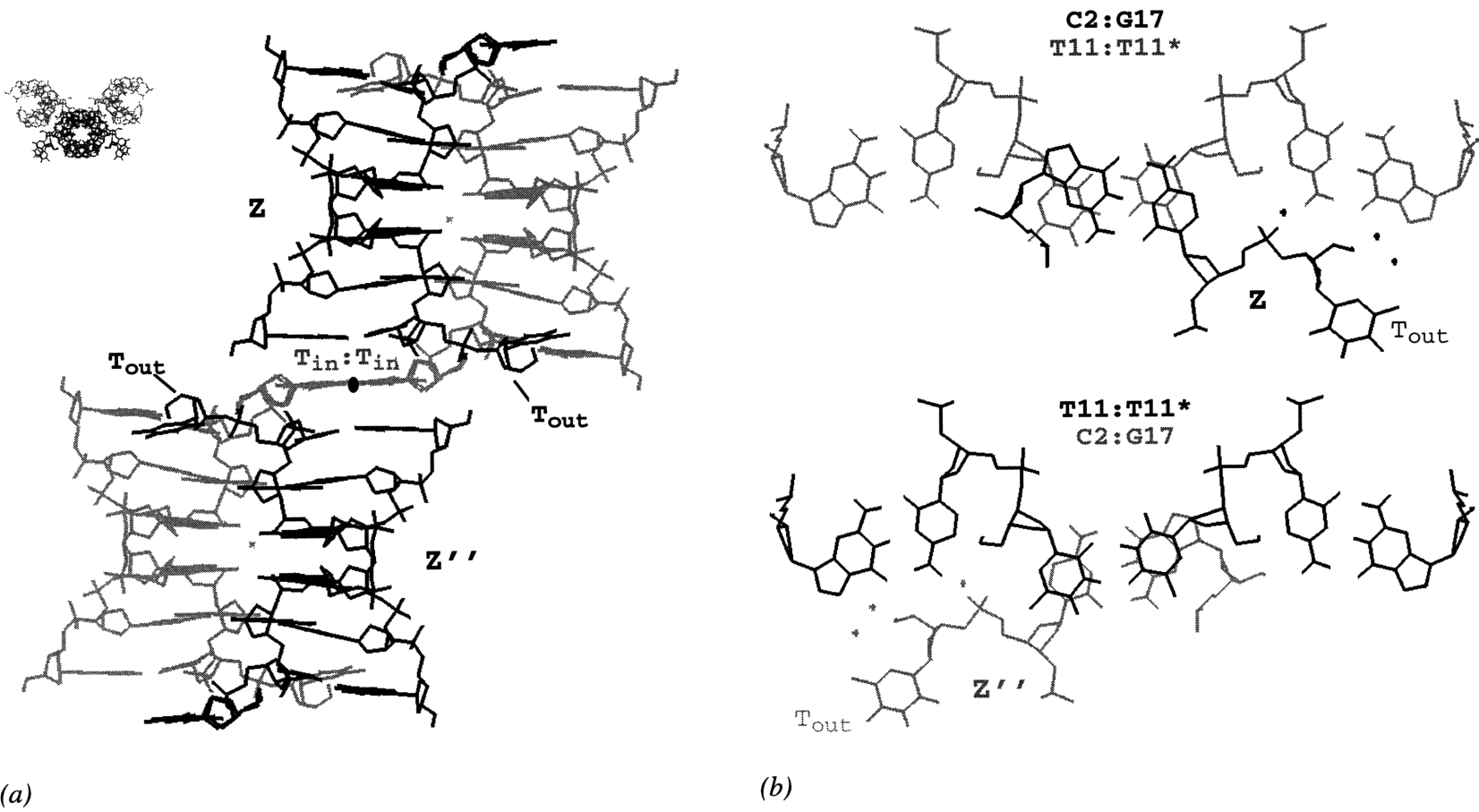
Figure 12. *(a)* Environment at the twofold rotation axis (x+1/2,z) (●). *(b)* alternate stacking arrangements of $T_{in}:T_{in}*$ base pair over CYT2:GUA17.
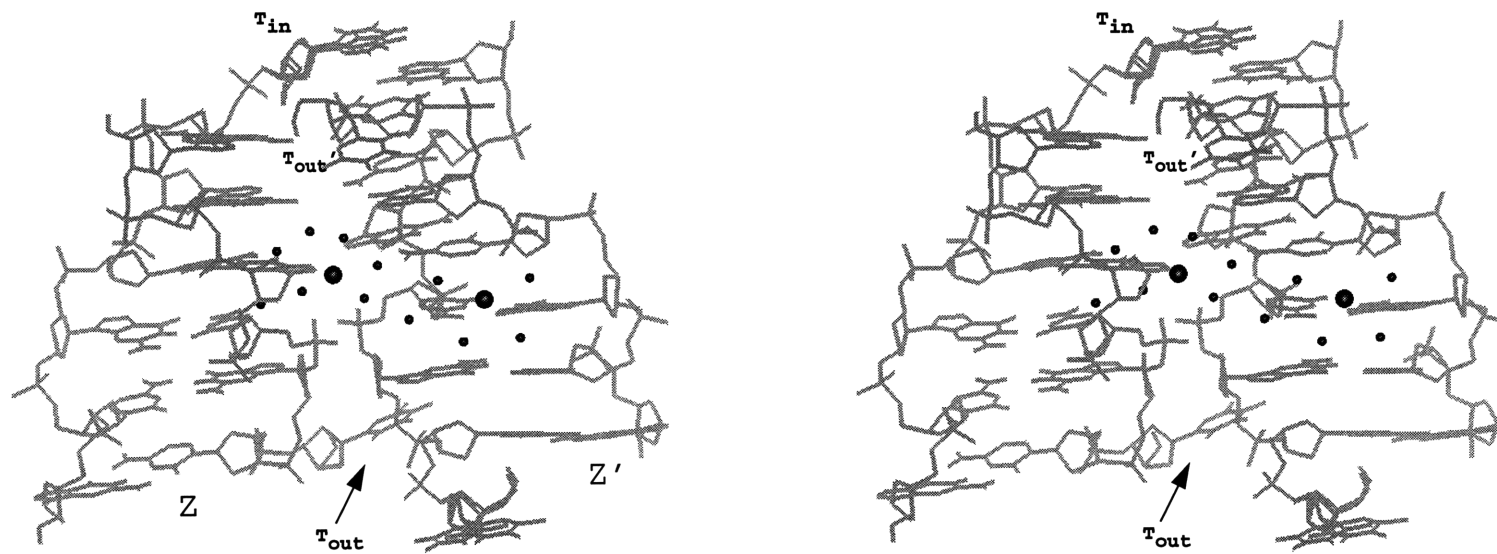
Figure 13. Stereo view of neighboring duplexes (Z and Z') and their coordinated barium–solvent belt.
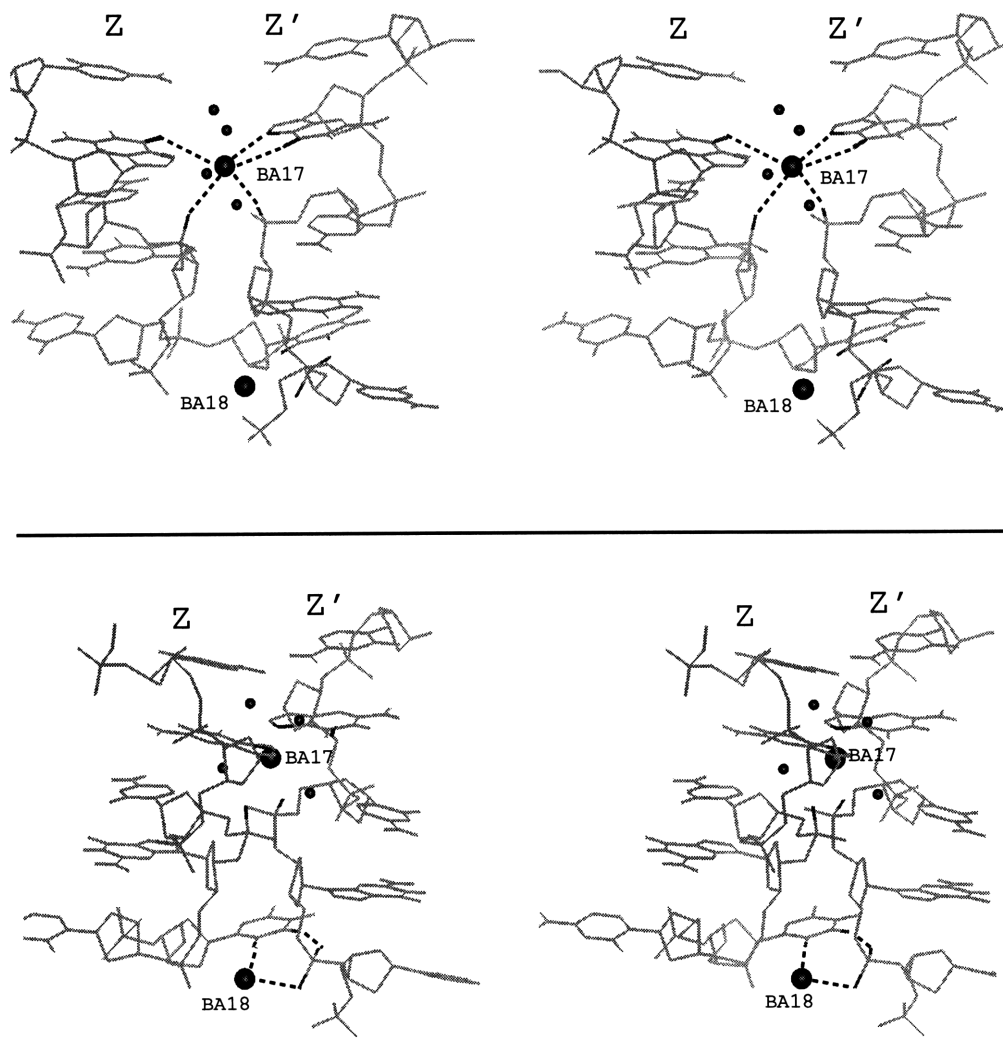
Figure 14. Details of Barium ion coordination. *upper:* stereo view showing the coodination if BA17 to strands from neighboring duplexes. *lower:* stereo view showing coordination of BA18 to THY1 and a phosphate group from the neighboring duplex.
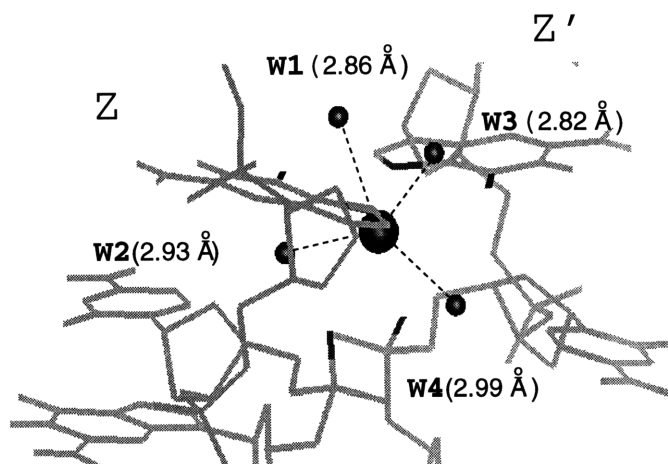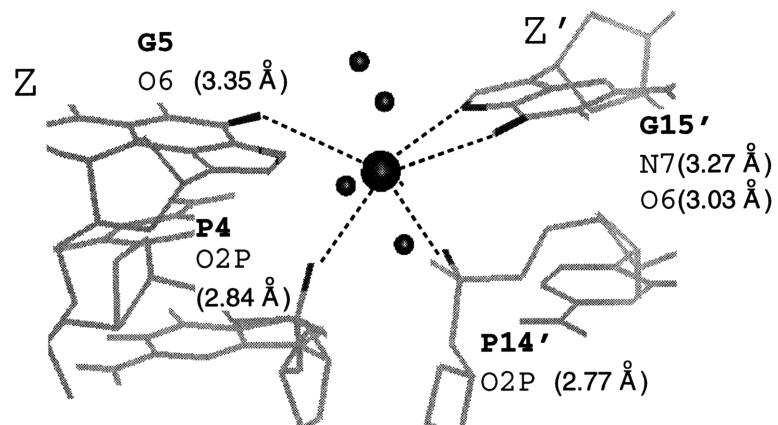
Figure 15. Details of Barium Ion BA17 geometry.

Figure 16. Details of Barium Ion BA18 geometry.

CHAPTER TWO

The 2.5 Å X-Ray Crystal Structure of
d(CCCCACACCCC)

2.1 INTRODUCTION

Sequences of DNA containing long repeats of G-rich and C-rich strands are found in telomeres[1] as well as other regions of the genome. G-rich sequences are capable of forming four-stranded structures stabilized by cyclic hydrogen bonding of four coplanar guanines[2-4] (the "G-quartet"). Biological contexts for G-quartets have been accumulating with the identification of G-quartet binding proteins.[5-8].

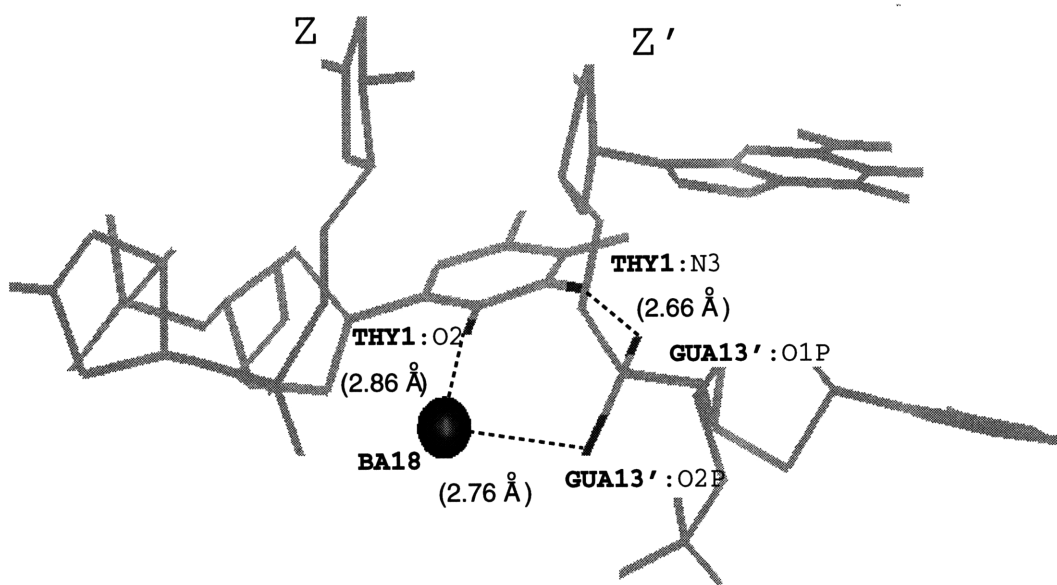A possible structural disposition for the complementary C-rich strand was first identified by Gueron and his associates[9], analyzing $d(TC_5)$ by NMR. They proposed a four-stranded structure (termed the "i-motif") in which two parallel-stranded duplexes intercalate to each other with opposite polarity; the parallel duplexes are held together by hemiprotonated cytosine-cytosine[+] base pairs (Figure 1). The rise/residue within the parallel duplex is ~6.4 Å, leading to stacking distance of ~3.2 Å within the intercalated tetraplex. The canonical i-motif tetraplex has gradual right-handed inter-and intra-strand twists.

A number of C-rich i-motif structures have been determined by X-ray crystallography and NMR[10-13,23]. The first i-motif crystal structure solved[10], $d(C_4)_4$, provided examples of tetraplexes with either 3' or 5' overhangs. In another crystal structure of $d(CCCAAT)$[13], it was demonstrated that the i-motif can incorporate Adenine:Adenine base pairing within the C-rich tetraplex. Base stacking in the i-motif is limited to exocyclic groups over adjacent π-systems, rather than π-system/π-system overlap; however, inspection of the various i-motif structures suggests that the four-stranded molecule is stabilized by inter-strand C-H$\cdots$O hydrogen bonds between (antiparallel) riboses[14].

Another example of long tandem repeats of G-rich and C-rich strands is the insulin-linked polymorphic region (ILPR). The ILPR is 363 base pairs upstream from the human insulin gene; it has been linked to insulin-dependent diabetes mellitus[15]. The

locus of the ILPR (or "insulin minisatellite") and its consensus sequence are shown in Figure 2.

Various types of polymorphism within the ILPR have been identified, including variation in the consensus sequence and in the number of tandem repeats[18-21]. Polymorphic classes which are variable in repeat number are associated with different levels of transcription of the insulin gene[21,22].

Biochemical[16] and structural[17] evidence indicate that the G-rich strand probably adopts a G-quartet conformation. Furthermore, a recent NMR structure[23] for part of the C-rich repeat ($d(C_4TGTC_4)$) indicates that this sequence is capable of forming an i-motif tetraplex from two hairpin duplexes with TGT loops.

Here we report the X-ray crystal structure of the $d(C_4ACAC_4)$ subset of the ILPR consensus sequence. The structure is a tetraplex consisting of two $(C_4)_4$ domains in an i-motif conformation, connected by an $(ACA)_4$ "knot". The knot includes adenine residues which are flipped out from the body of the tetraplex. These adenine residues hydrogen bond with, and stack between, neighboring tetraplexes; they provide the principle interactions which stabilize packing of molecules in the crystal.

This structure provides further evidence for possible unusual DNA structures within the ILPR, which may be responsible for transcriptional regulation as well as variable repeat number polymorphisms.

2.2 MATERIALS AND METHODS

The sequence $d(C_4ACAC_4)$ was produced on an automated synthesizer by the M.I.T. Biopolymers laboratory and purified by HPLC. Crystals grew from a solution containing: 5%(v/v) 2-methyl-2,4-pentanediol (MPD), 50 mM sodium cacodylate at pH 5.5, 5 mM KCl, 2mm spermine, and 2 mM DNA (single-strand concentration). Hanging drops were equilibrated against a reservoir of 30% MPD.

Crystals grew in a tetragonal lattice of the space group $P4_22_12$. Unit cell dimensions are : a=b= 49.97Å, c= 28.06Å, $\alpha=\beta=\gamma=$ 90°. Data was collected on an RAXIS IIc (Molecular Structure Corp.) The structure was refined using 1357 (of 1447 possible reflections to 2.5 Å) observable at 2σ (F) or greater.

The structure was solved by molecular replacement using AmoRe[26], with both 3'- and 5'- overhang i-motif models of $d(C_4)$[10]. One 3'-overhang model consistenly yielded good rotation-translation solutions (Correlation Coefficient >60%, R-factor <50 %) Subsequent refinement with XPLOR[25] allowed for the identification of strand ends and proper assignment of residues. Omit and (2Fo-Fc) maps were sufficient to trace the missing parts of the chain. Additionally, omit maps indicated that two phosphate groups, P2 (between CYT1 and CYT2) and P9 (between CYT8 and CYT9) were disordered. In order to allow for alternative phosphodiester geometries at these locations, the O5'-OPO-atoms were assigned two conformations (50% occupancy each) a the P2 and P9 phosphodiesters. This permitted variation in α, β, and γ phosphodiester angles without requiring alternate sugar puckers, particularly on the 3'-adjacent riboses.

After all residues were assigned, and the chain properly traced, simulated annealing and B-factor refinement was able to reduce the R-factor to about 28%. This was attributed to poor R-factor performance in the highest resolution shells (2.8- 2.5 Å), which in turn was the result of significant anisotropy in the dataset. In an effort correct for this, the program MAXSCALE (generously provided by Dr. Mark Rould, M.I.T.)

was used to estimate local scale factors. MAXSCALE, in this application, derives a scale factor based on the distribution of F_{calc}s from the model.

Rescaling of the data, followed by further refinement (simulated annealing, positional refinement, B-factor refinement) yielded improved density maps and improved R-factor performance. Very few solvent were discernible in the maps; only 10 waters are included.

The final 2.5 Å R-factor is 24.1%, with bond and angle root mean square deviations (RMSD) of .016 Å and 2.4 degrees, respectively. A free R-factor based on a random 10% subset of reflections is 29.3%.

2.3 RESULTS AND DISCUSSION

The asymmetric unit of d(CCCCACACCCC) is an elongated single strand, as shown in Figure 3. The strand lies along the ab diagonal in the unit cell (Figure 4a,b); this places its long axis along a twofold rotation axis, one of three (mutually perpendicular) twofold axes which meet at the center of the unit cell. Application of the "axial" twofold generates a parallel symmetry copy; subsequent application of either "lateral" twofold generates an antiparallel copy of the two parallel strands. The resulting tetraplex thus consists of two parallel duplexes which intercalate to each other in an antiparallel way (Figure 4c,d). The tetraplex can be described as two($C_4)_4$ i-motif domains (comprised of residues C1-C4 and C8-C11) connected by a knot consisting of four copies of residues A5-C6-A7 ($(ACA)_4$). Except for residue A7, all base pairing is between members of a parallel duplex. A7 residues base pair with symmetry related A7s from a neighboring tetraplex, forming two layers of A7:A7 base pairs (Figure 4c,d, Figures 14 & 15).

The parallel duplex produced by the axial twofold symmetry operation is shown in Figure 5. Ten of the eleven residues are base-paired to their identical symmetry copies (e.g C1:C1*) across the axial twofold. Hydrogen bond lengths for C:C base pairs in the parallel duplex are as follows: N3-N3* =2.7Å ± .18 Å, O2-N4* =2.6 ± 0.15 Å. The intra-duplex A5:A5* base pair has an N6-N7* hydrogen bond length of 2.73 Å. (the inter-tetraplex A7:A7* hydrogen bond (between N6-N7*) is 2.89 Å.

Figure 6 illustrates the tetraplex (two antiparallel duplexes) produced by the lateral twofold operation. Four copies of ($^{5'}$C1-C4$^{3'}$) and ($^{5'}$C8-C11$^{3'}$) comprise a ($C_4)_4$ i-motif domain at either end of the tetraplex. These domains have 5' overhangs at C1 and C8, and 3' underhangs at C4 and C11. Each i-motif domain presents its 5'-overhanging C8 residues toward the central (A5-C6-A7)$_4$ knot, leaving 5'-C1 overhangs at the distal ends. A5 residues confer 3'-overhangs (the 3'-A5 caps) above the 5'-C8 layers of the i-

motif domains. Each strand then proceeds through the C6-A7 "loop", where they connect with the 5'-C8 overhangs of the next i-motif domain. Figure 7 shows the relationship between the $(C_4)_4$ i-motif domains, the 3'-A5 caps, and the C6-A7 loops.

Figure 8 gives details of duplex and tetraplex geometry. With respect to rise/residue strand twist, and base stacking distances, the $(C_4)_4$ i-motif domain conforms to previously observed values[10-13]. Both Adenine residues have C1'-exo sugar puckers; seven of nine cytosine residues, including the loop C6, have C4'-exo pucker, the most commonly observed sugar conformation in the cited crystal structures.

Figure 9 shows one of the i-motif domains with a 3'-A5 cap. The cap can be considered part of the i-motif in that it continues with an ~6.4 Å rise; however, contrary to the right handed strand twist of the $(C_4)_4$ i-motif (~ 10° to 20°), the 3'-A5 cap is attached with a strand twist of ~-18°. It may be that the choice of Adenine at this position in the sequence is for its ability to form a base pair in this left-handed turn. A view down the helical axis illustrates the narrow and wide grooves characteristic of i-motif tetraplexes [10-13]. Figure 10 shows the base stacking between each adjacent layer in the $(C_4)_4$ i-motif domain and 3'-A5 cap. With the exception of the A5/C8 overlap, there is no π-system overlap between the pyrimidine rings; base stacking is limited to exocyclic groups. This is also typical of previously observed i-motif structures.

The C6-A7 loop in a parallel duplex is shown in Figure 11. The strands make an ~ 3.5 Å rise, with +52° twist, between A5 and C6; the C6-A7 loops then propagate to C8, with a net rise of ~10.5Å, and net twist of ~50°. Figure 12 shows all four strands of the C6-A7 loop connected to the capped i-motif domains. A5:A5 and C6:C6 base pairs have high propeller twists (See Figure 8); as a consequence, adjacent C8:C8 base pairs have high propeller twists as well. Electron density maps around A5:A5* and C6:C6* base pairs are shown in Figure 13.

A7 residues from the C6-A7 loop are pointed away from the helical axis of the tetraplex; four A7 residues from two neighboring tetraplexes interact to form two stacked

layers of A7:A7 base pairs (Figure 14); The N6:N7' hydrogen bond is 2.89 Å; the stacking distance between A7:A7 base pairs is ~3.5 Å. Each A7 also makes a long hydrogen bond (~3.4 Å) via N7 with a phosphate oxygen of the A7' residue to which it is base paired (Figure 14). The two A7:A7' layers are stacked between adjacent perpendicular tetraplexes; each A7:A7' base pair thus stacks over a neighboring C1:C1* base pair, at a stacking distance of ~3.4 Å (Figure 15). The interactions between neighboring A7 and C1 residues are, almost exclusively, the only forces which stabilize the organization of tetraplexes in the crystal.

The $(ACA)_4$ knot perturbs the narrow and wide grooves of the tetraplex. Using interstrand phosphate distances as a measure of groove with, the narrow groove is widened dramatically (up to ~4 Å) across the C6-A7 loop (Figures 16-17).Conversely, the wide groove (Figures 18-19) narrows only slightly across the $(ACA)_4$ knot.

This structure may be relevant to replicative and transcriptional processes. The formation of G-quartet and i-motif structures in the ILPR may present a challenge to the replicative machinery in its attempt to reproduce the two complementary strands. Furthermore, transition from double stranded DNA to four stranded structures should effect the supercoiling in the region proximal to the insulin gene. This, in turn, should effect the dynamics of the transcriptional apparatus, which itself is producing supercoils as it processes through the gene. [24]

It is possible to reconcile this structure of CCCCACACCCC with the results from Catasti et al[24], who observed by NMR that the ILPR consensus sub-fragment CCCCTGTCCCC formed an i-motif quadruplex using antiparallel hairpins with TGT loops. Figure 20 illustrates how four contiguous repeats of the full ILPR consensus could form a tetraplex using the $(ACA)_4$ knot presented here, and (TGT) loops to connect 5'-overhangs to 3'-underhangs.

47

2.4 REFERENCES

1. Blackburn, E.H. (1991) *Nature* **350**, 569

2. Kang, C.H., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. (1992) *Nature* **356**, 126

3. Smith, F.W. & Feigon, J. (1992) *Nature* **356,** 164

4. Wang, Y. & Patel, D.J. (1995) *J. Mol. Bio*, **251**(1), 76

5. Fang, G. & Cech, T.R. (1993) *Cell* **74**, 875

6. Weisman-Schomer, P., & Fry, M. (1993) *J. Biol. Chem.* **268**, 3306

7. Liu, Z., Franz, J.D., Gilbert, W., & Tye, B.-K.(1993) *Proc. Natl. Acad. Sci. USA* **90**, 3157

8. Schierer, T., Henderson, E. (1994) *Biochemistry* **33**, 240

9. Gehring, K., Leroy, J.-L., & Gueron, M. (1993) *Nature* **363**, 561

10. Chen, L., Cai, L., Zhang, X., & Rich, A. (1994) *Biochemistry*, **33**, 13540

11. Kang, C.H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. & Rich, A. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 11636

12. Kang, C.H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. & Rich, A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 3874

13. Berger, I., Kang, C.H., Fredian, A., Ratliff, R., Moyzis, R. & Rich, A. (1995) *Nat. Struct Biol* **2**, 416

14. Berger, I., Egli, M. & Rich, A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 12116

15. Wicker, L.S., Todd, J.A., & Peterson, L.B. (1995) *Annu. Rev. Immunol.* **13**, 179

16. Hammond-Kosack, M.C., Dobrinski, B., Lurz, R., Docherty, K, Kilpatrick, M.W. (1992) *Nucl. Acids Res.*, **20**(2), 231

17. Catasti, P., Chen, X., Moyzis, R., Bradbury, E.M., & Gupta, G. (1996) *J. Mol. Biol.*, **264**, 534

18. Bell, G.I., Karam, J.H., & Rutter, W.J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5759

19. Bell, G.I., Selby, M.J., & Rutter, W.J. (1982) *Nature* **295**, 31

20. Lucassen, A.K., Screaton, G.R., Julier, C., Elliot, T.J., Lathrop, M., & Bell, J.I.
    (1995) *Hum. Molec. Genet.* **4**, 501

21. Kennedy, G.C., German, M.S., & Rutter, W.J. (1995) *Nature Genetics*, **9**(3), 293

22. Bennett, S.T., Lucassen, A.K., Gough, S.C.L., Powell, E.E., Undlien, D.E., Pritchard,
    L.E.R, Merriman, M.E., Kawaguchi, Y. Dromsfiled, M.J., Pociot, F., Nerup, J.,
    Bouzekri, N., Cambon-Thomsen, A., Ronningen, K.S., Barnett, A.H., Bain, S.C.,
    & Todd, J.A. (1995) *Nature Genet.* **9**, 284

23. Catasti, P., Chen, X., Deavan, L.L., Moyzis, R., Bradbury, E.M., & Gupta, G. (1997)
    J. Mol. Biol., **272**, 369

24. Liu, L.F., & Wang, J.C. (1987) *Proc. Natl. Acad. Sci. USA* **84**,7024

25. Brunger, A.T. (1992) *XPLOR 3.1: A System for X-ray Crystallography and NMR*
    Yale University Press

26. Navaza, J. (1994) AMoRe:An Automated Package for Molecular Replacement.
    Acta Cryst. **50** 157
    Collaborative Computational Project, number 4 (1994)
    The CCP4 Suite: Programs for Protein Crystallography
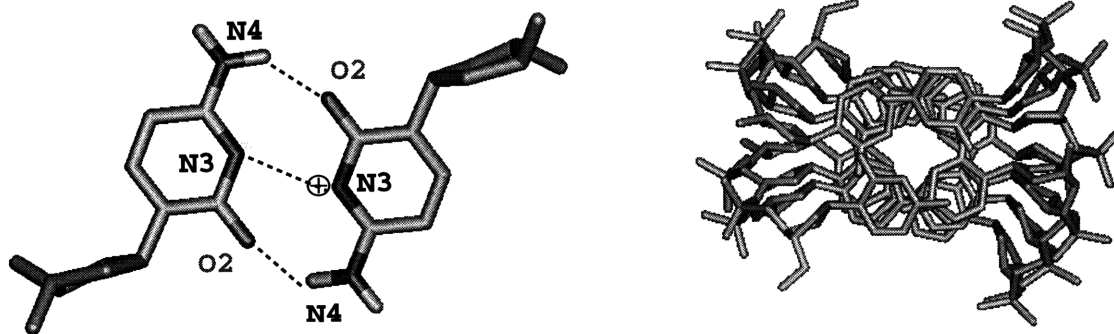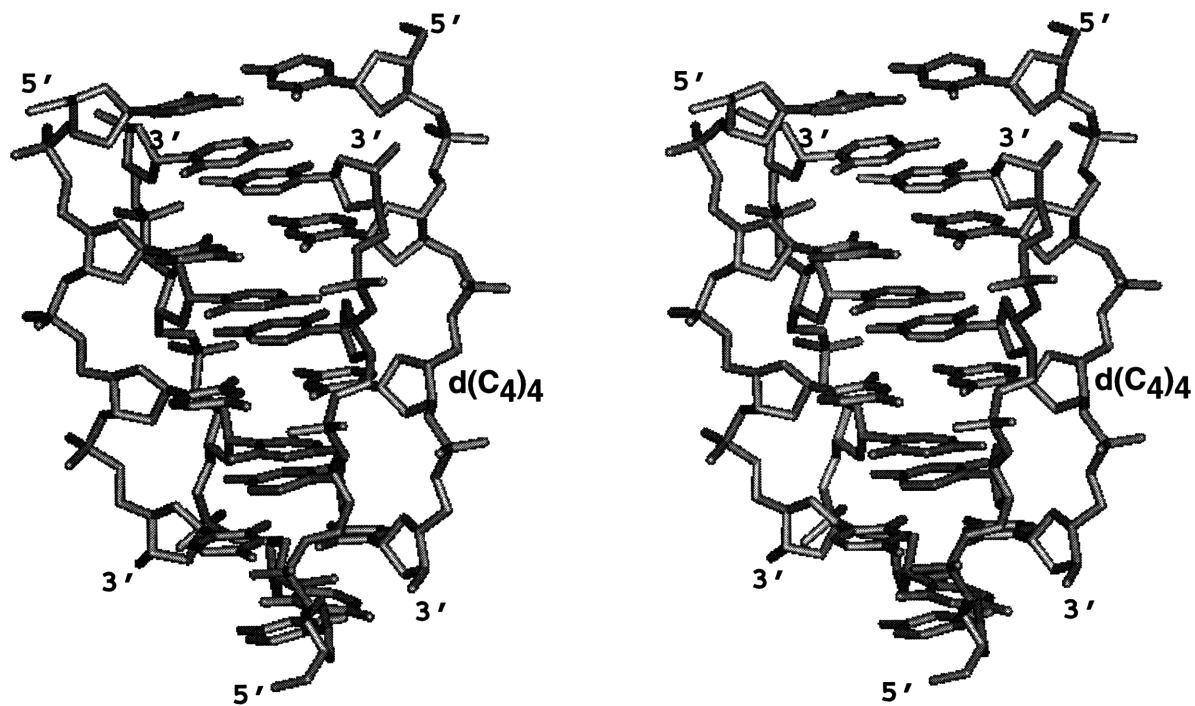   *Acta Cryst.* **D50** 760

Figure 1. An i–motif Tetraplex and its Constituent Hemiprotonated C:C+   Base Pair
(upper) Stereo view of a $(C_4)_4$ i–motif consisting of two antiparallel sets of parallel
duplexes. The parallel duplexes are held together Cytosine–Cytosine base pairs *(lower left);*
if only one N3 in the pair is protonated ($\oplus$), three hydrogen bonds can be formed. It is
important that this type of C:C+ base pair has a perpendicular twofold rotation axis.
(lower right) View down the tetraplex' axis, showing that it has wide and narrow grooves.

~363 bp

ILPR   INSULIN GENE

enhancer   promoter

ILPR
Consensus Sequence:
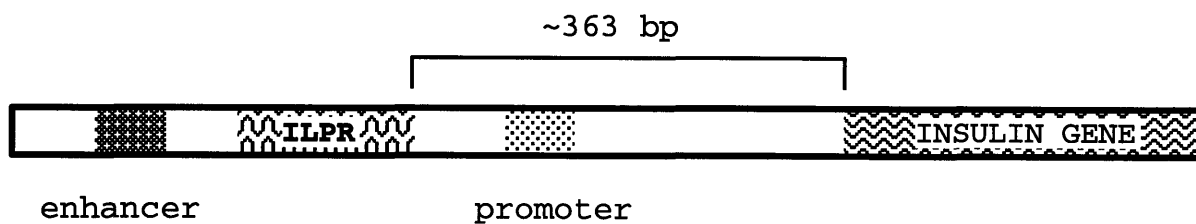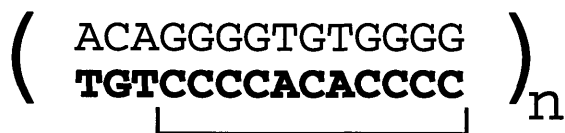
$$\left(\begin{array}{c} \text{ACAGGGGTGTGGGG} \\ \textbf{TGTCCCCACACCCC} \end{array} \right)_n \qquad n = \sim 45\text{–}160 \text{ in humans}[21]$$

↑ n => ↑ transcription

Figure 2. Location of the Insulin–Linked Poplymorphic Region (ILPR)
The ILPR, or "insulin minisatellite", consists of a variable number of repeats of the consensus
sequence. An increase in transcription rate is observed with increasing repeat length (n).
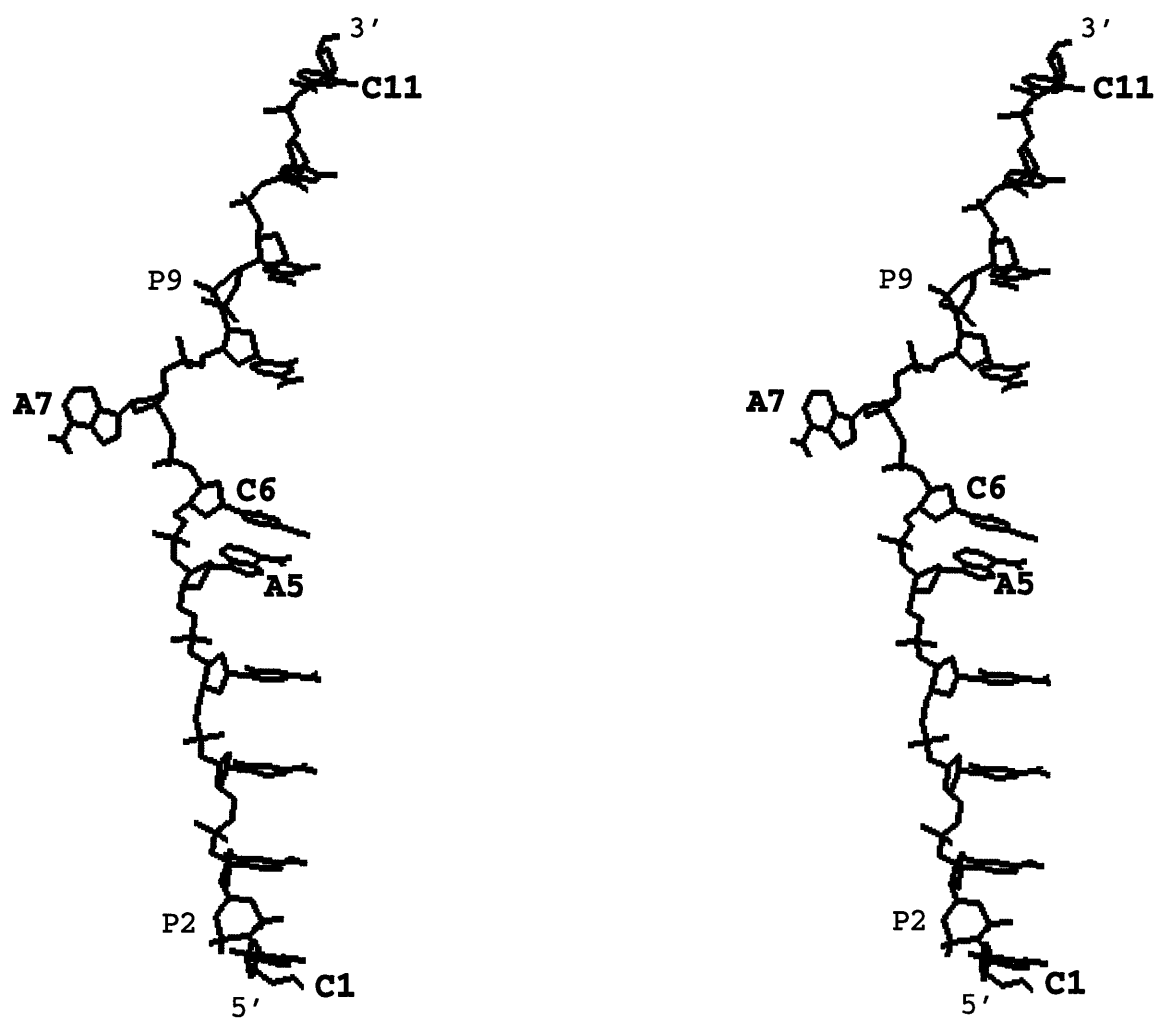The subject of this study is a subset of the C–rich strand,namely CCCCACACCCC.

51

Figure 3. Stereo view of the assymetric unit of d(CCCCACACCCC)
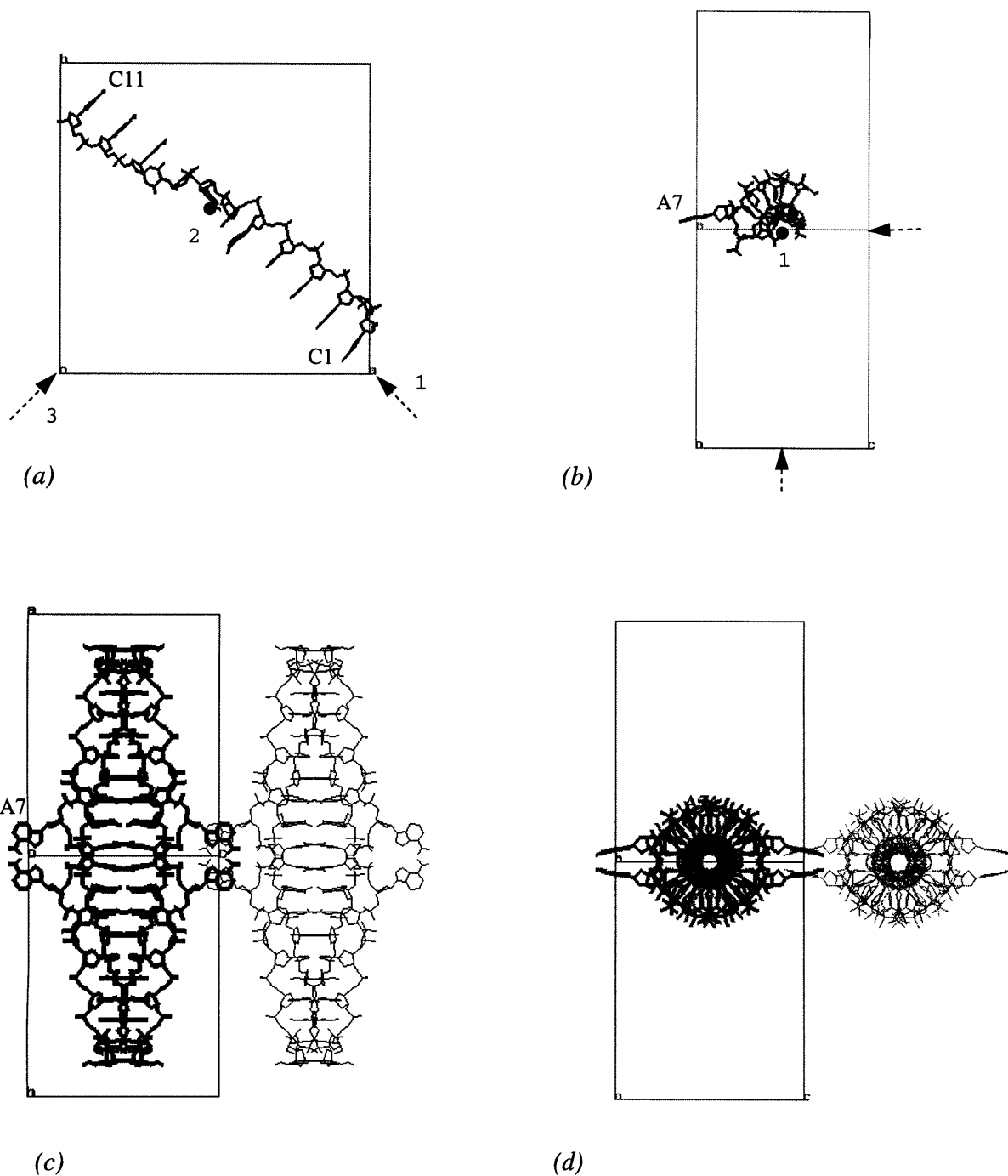Phosphate groups P2 and P9 have multiple conformations as shown

52

Figure 4. The Asymmetric Unit in Relation to Symmetry Operators.
(a) view of ab plane. The asymmetric unit lies along an ab diagonal, one of three twofold rotation axes (1,2,3)that meet in the center of the unit cell. *(b) view along axis 1.*Axis 1 is coincident with the long axis of the strand. This "axial" twofold operation yields a parallel duplex (see Fig 5). Subsequent application of either "lateral" twofold results in a four–stranded molecule of antiparallel duplexes. *(c)* View down lateral axis 3 of the tetraplex. Neighboring tetraplexes interact via their respective A7 residues. *(d)* Axial view of the tetraplex.
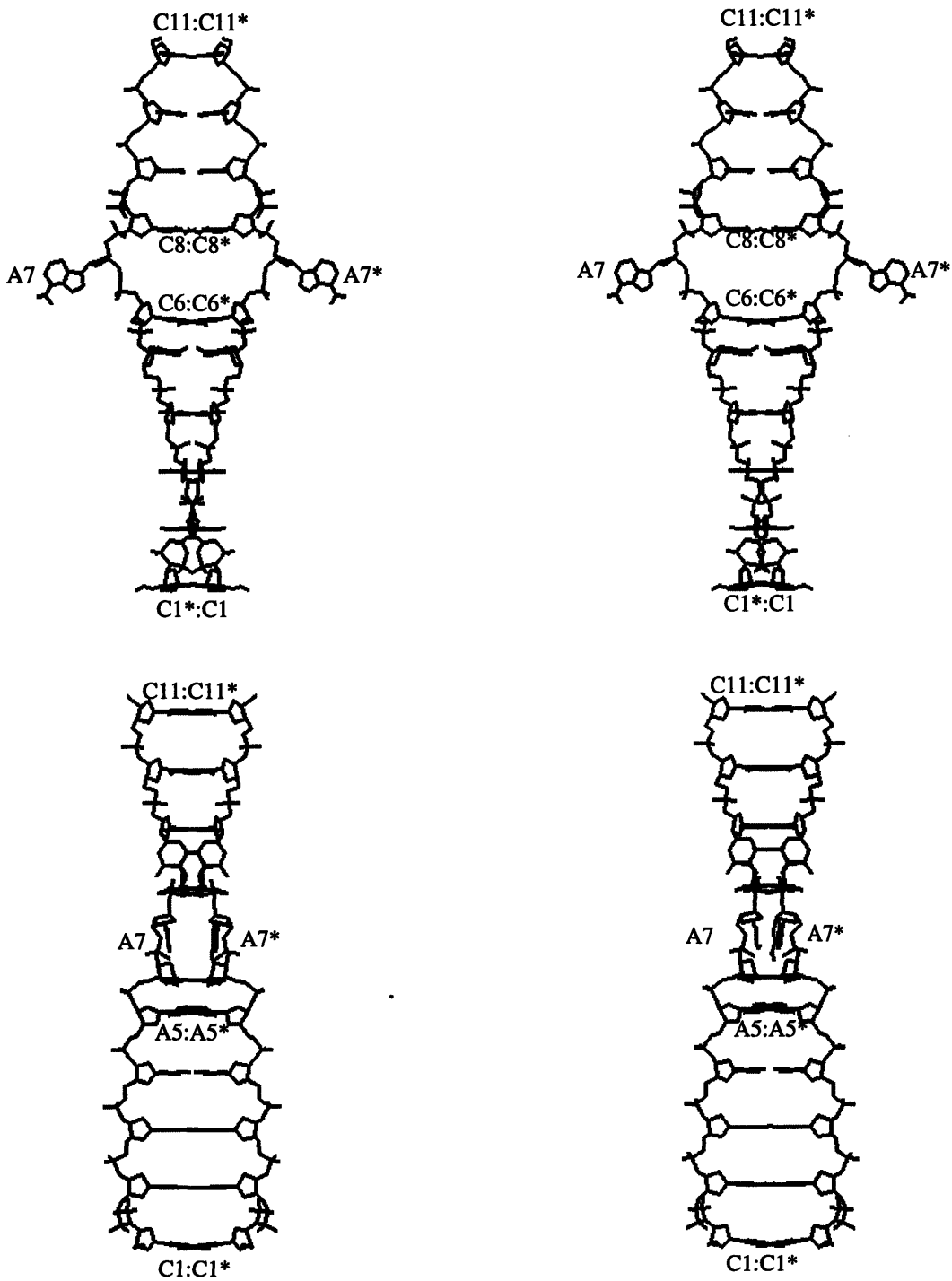
Figure 5. Stereo Views of the Parallel Duplex Produced by the Axial Twofold Symmetry. All bases except A7 participate in intra-duplex base pairing. Every residue pairs with a twofold symmetry copy of itself (e.g. C1:C1*). The A7:A7 base pairs are inter-tetraplex with a neighboring molecule. (See Figs. 4, 14, 15).
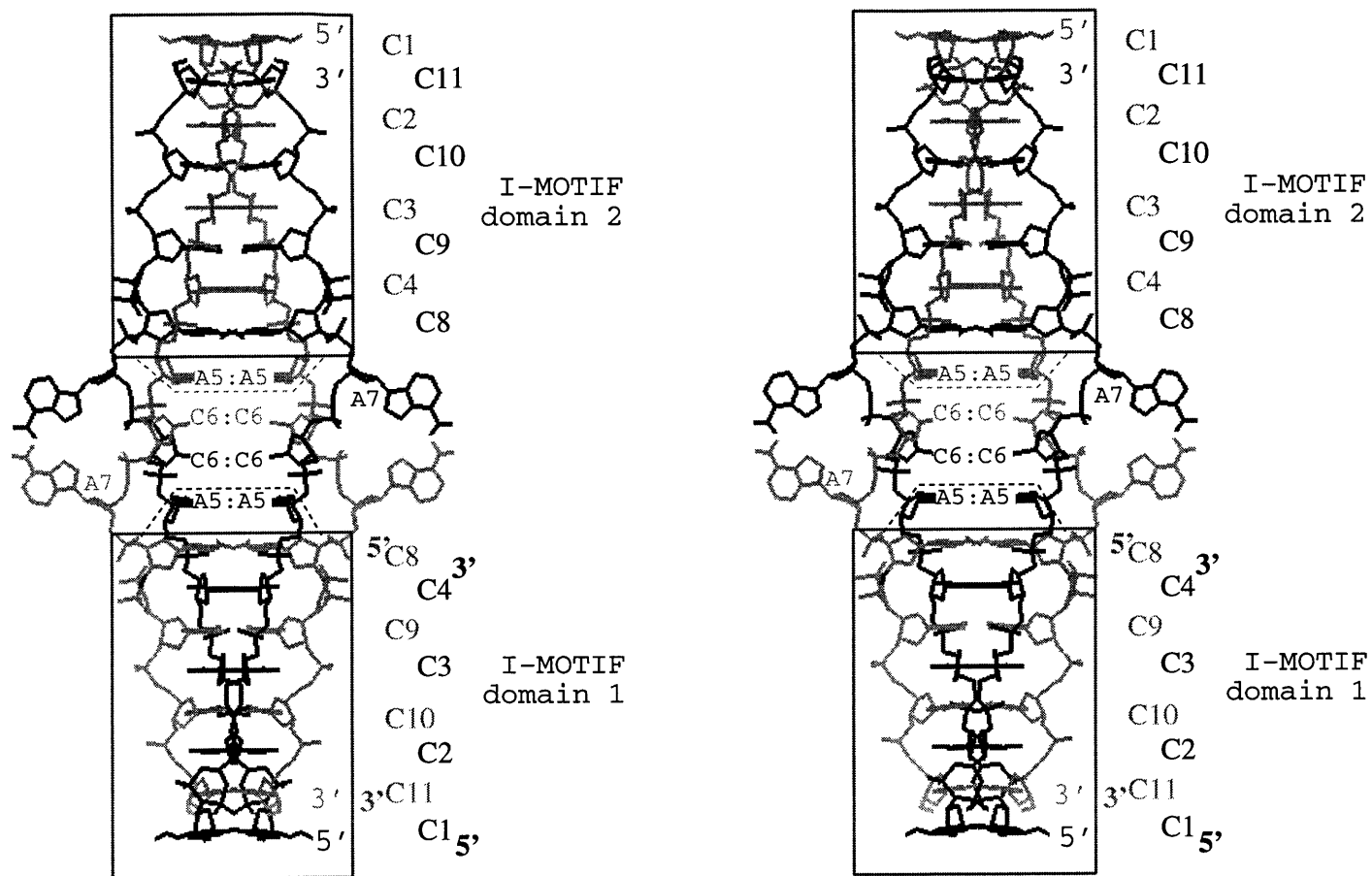
Figure 6. Lateral Twofold Symmetry Produces a Tetraplex From A Parallel Duplex.
The molecule consists of two i–motif domains with 5'overhang at C1 and C8 (boxed) connected by the A5–C6–A7 knot. A5:A5 base pairs place a 3' cap over the 5'–C8 overhangs. The remainder of the knot consists of the C6–A7 loops, which connect A5 residues to the 5'–C8 overhang on the next i–motif domain.
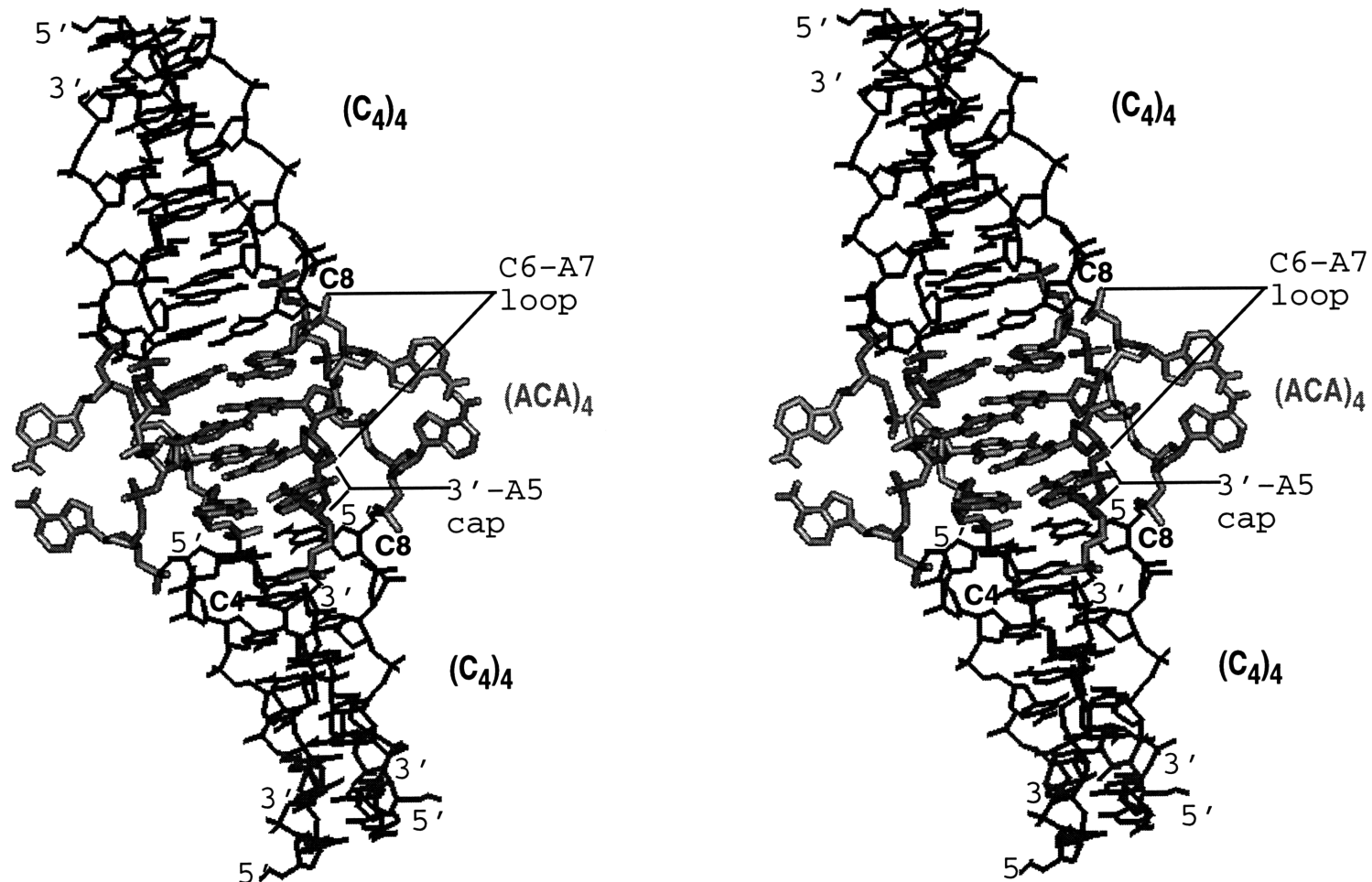
Figure 7. Stereo view of (ACA)₄ Knot Connecting Two (C₄)₄ i–motif domains

Four copies of the single strand produce two $(C_4)_4$ i–motif domains with 5'–overhangs (C1:C1*, C8:C8*)(black lines). The $(ACA)_4$ knot (grey solid) is connected to each strand and connects the 3' end (C4) of one i–motif domain with the 5' (C8) end of the next. Residue A5 may be considered part of the i–motif in that it persists with an ~6.4 Angstrom rise; however, it is unusual in that it has a left–handed strand twist, contrary to the right–handed twist within the $(C_4)_4$ i–motif. It is essentially a 3' "cap" to the 5'–overhang $(C_4)_4$ i–motif domain; the remainder of the knot is the C6–A7 loop.

rise(Å) strand twist(θ_T) base pair propeller twist

-5.0°

C11

6.3   13.2°

3.6°

C10

6.3   9.1°

6.2°

C9

6.5   21.4°

12.5°

C8

3.5   171.9°

13.4°

A7

7.0   -122°

8.5°

C6

3.5   51.8°

13.9°

A5

6.4   -18.8°

-5.0°

C4

6.4   17.6°

6.3°

C3

6.2   13.9°

4.1°

C2

6.2   18.6°

17.4°

C1

base stacking rise(Å)   sugar pucker

(A7 C1'-exo)

C6' 3.6
C6   C4'-exo
3.5
A5   C1'-exo
3.3   C8'   C4'-exo
3.2
C4   C4'-exo
3.2   C9'   C4'-exo
3.2
C3   C4'-endo
3.1   C10' O1'-endo
3.2
C2   C4'-exo
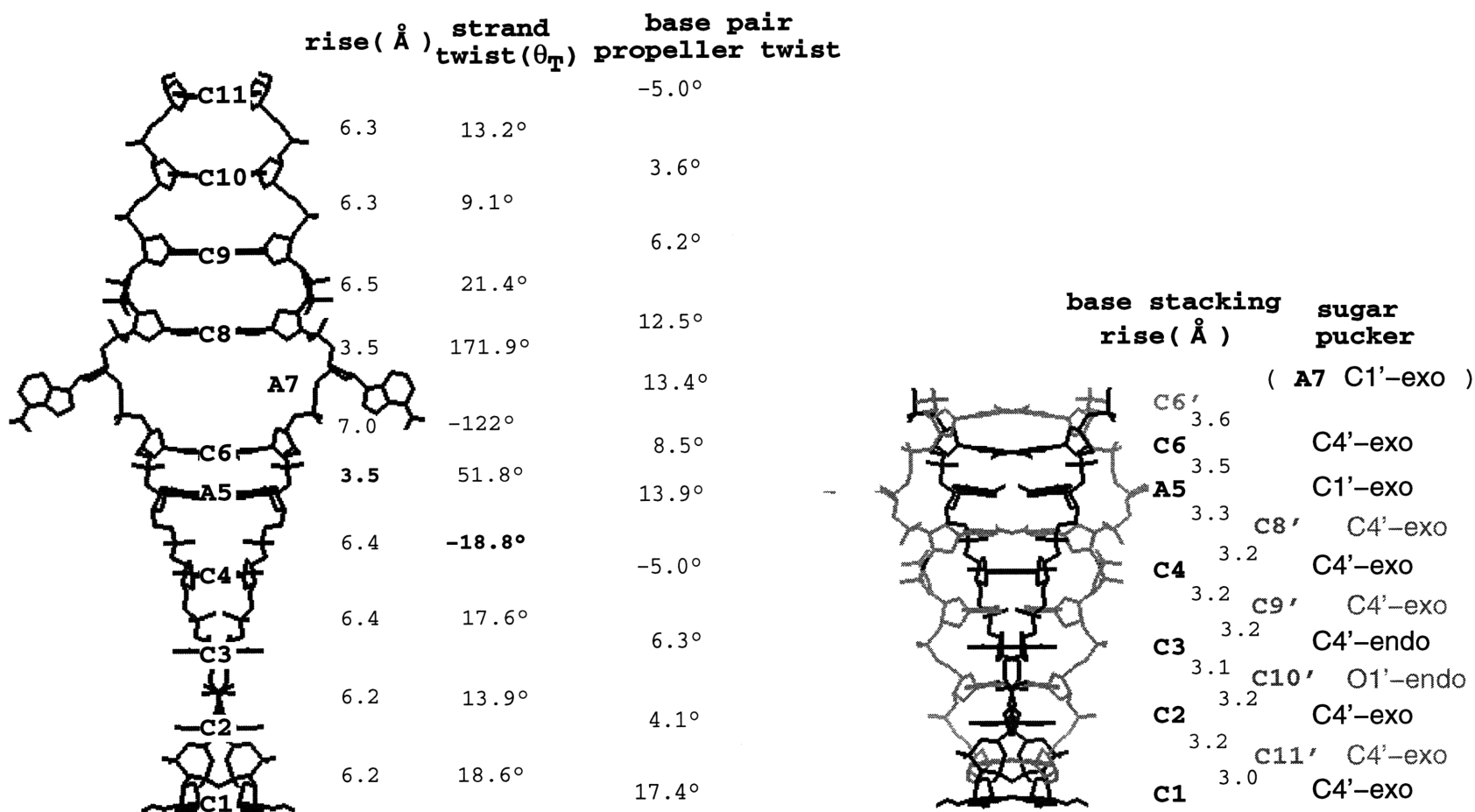3.2   C11' C4'-exo
3.0
C1   C4'-exo

Figure 8. Strand Parameters of the Parallel Duplex; Base Stacking and Sugar Pucker Within the Tetraplex
(left) Helical rise and twist parameters for the parallel duplex.Cytosine residues C1–C4 and C8–C11 have strand twists and rise/residue values typical of observed i–motif structures. Base pair propeller twist is highest in the knot residues (A5–C6–A7) and in neighboring cytosine base pairs. The inter–tetraplex A7:A7 base pair also has a high propeller twist; this impacts the propeller twist of the C1:C1* base pair, which is stacked over A7:A7 (see Figs 14, 15.). *(right)* Stacking distances between adjacent base pairs in the i–motif and the knot. Sugar pucker is also indicated.
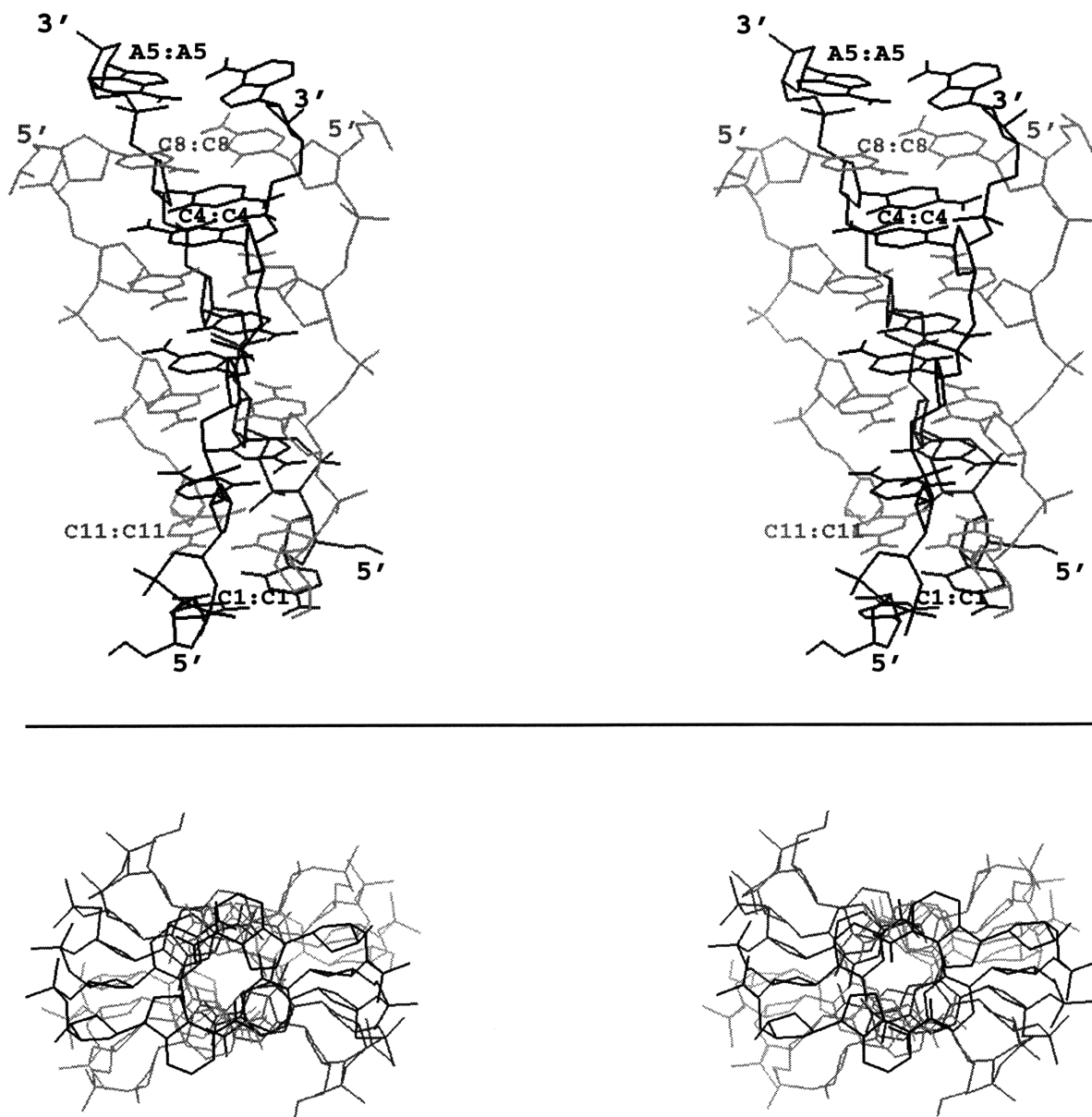
Figure 9. 5'–overhang (C₄)₄ Tetraplex with a 3'–A5 Cap.

(upper) Stereo view of one i–motif domain capped by a 3'–A5 base pair. Strands from residue C1 to C4 have a right–handed twist (see Fig. 8) until residue A5 is encountered; the step from C4 to A5 has a twist of ~–18°. C8 to C11 have all right–handed twists. *(lower)* Stereo view from A5 down the i–motif axis, toward the 5'–C1 overhang, showing that the domain has wide and narrow grooves.
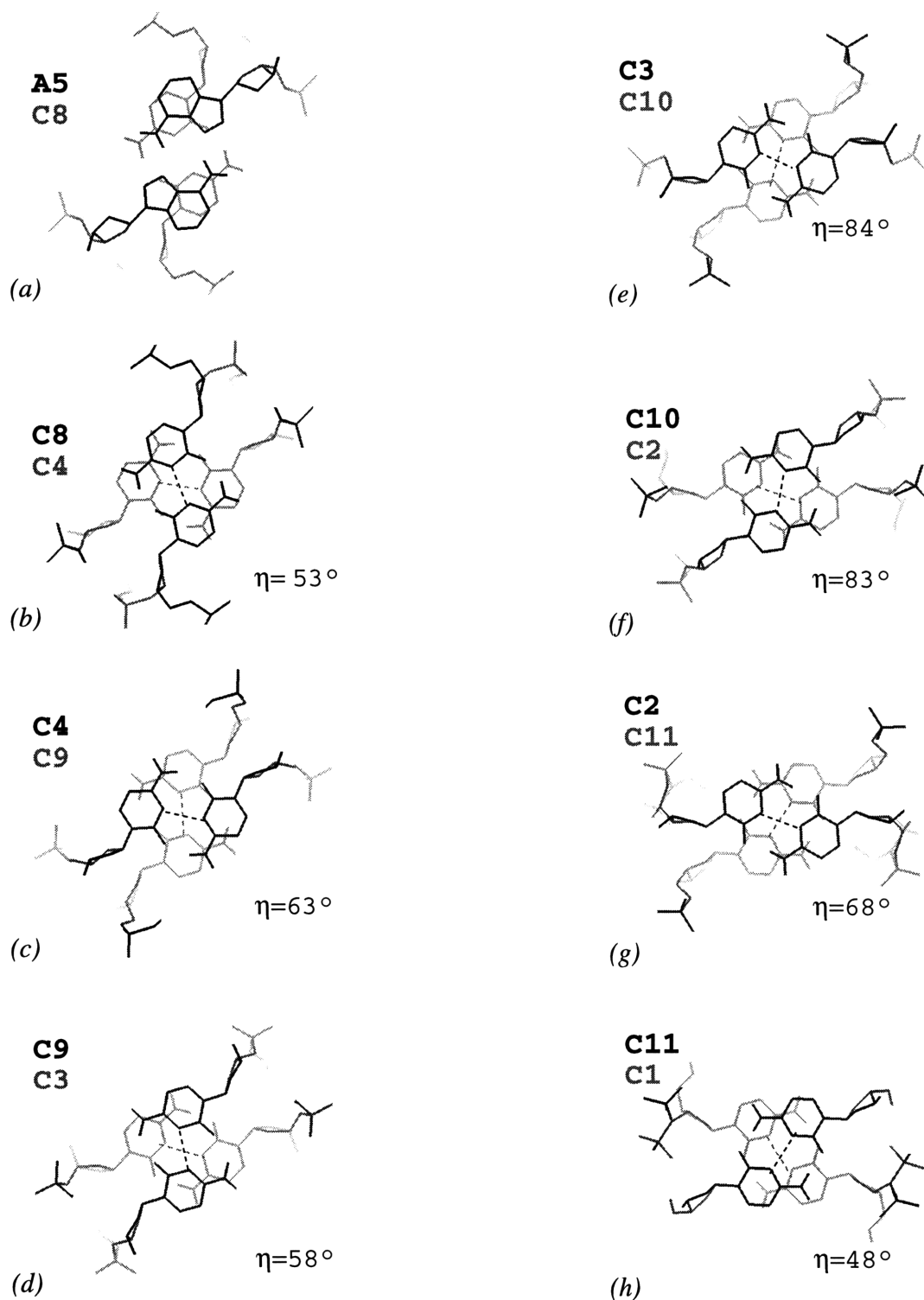
58

Figure 10. Base Stacking Within the $(C_4)_4$ i–motif and 3'–Cap.

(a–h) Base stacking between antiparallel duplexes from the 3'–A5 cap to the 5'–terminal base pair (C1:C1). Twist angle η, the acute angle between N3–N3 vectors, an indicator of inter–duplex twist, is shown. Stacking is primarily by exocyclic groups, not ring systems.
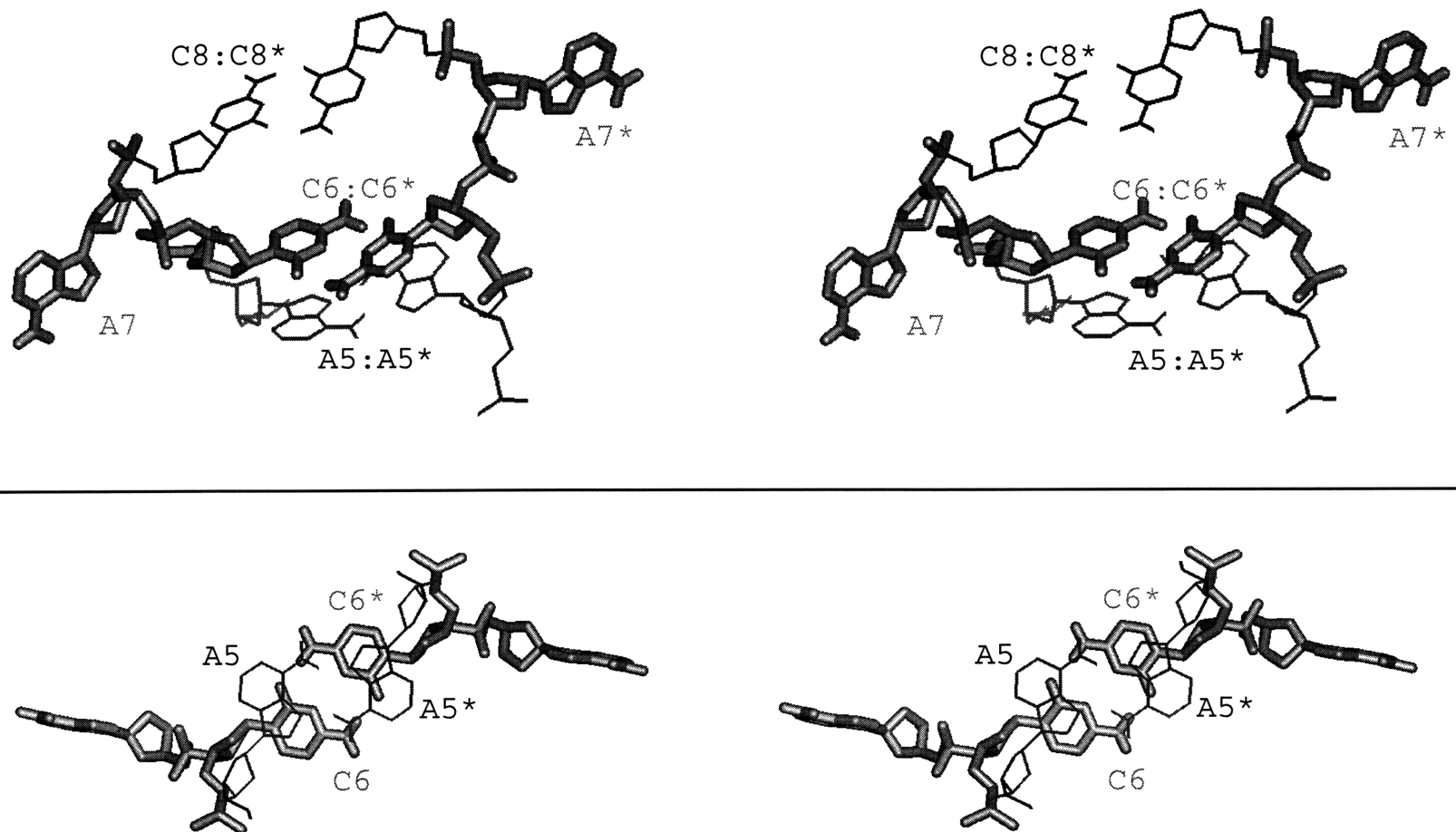
Figure 11. The C6–A7 Loop Connects the 3'A5–cap to the 5'–Overhang (C8) of the Next i–motif Domain
(upper) Stereo view of C6–A7 loops in a parallel duplex. The net rise from C6 to C8 is ~10.5 Angstroms, with a net twist of ~50°. *(lower)* Stereo view showing the stacking of A5:A5* over C6:C6*

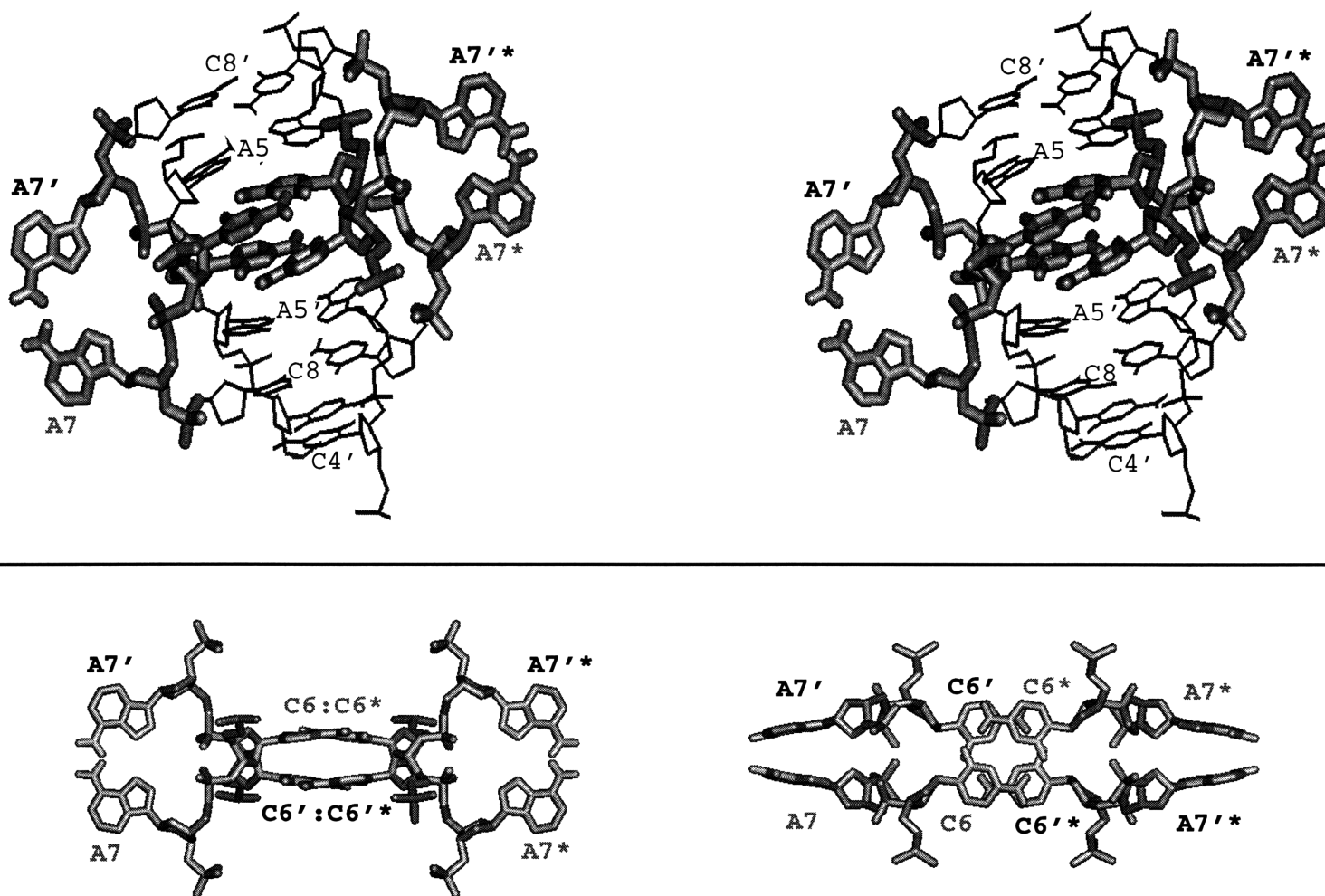**Figure 12. The (ACA)$_4$ Knot Connects two (C$_4$)$_4$ i–motif domains Between C4 and C8.**

(upper) Stero view: Four copies of the single strand produce two (C$_4$)$_4$ i–motif domains with 5'–overhangs (C8). The (ACA)$_4$ knot is connected to each strand with a left–handed twist from residue C4 to A5. Strands related by "*" are members of the same parallel duplex. *(lower)* Side and top views the four–stranded C6–A7 loop.

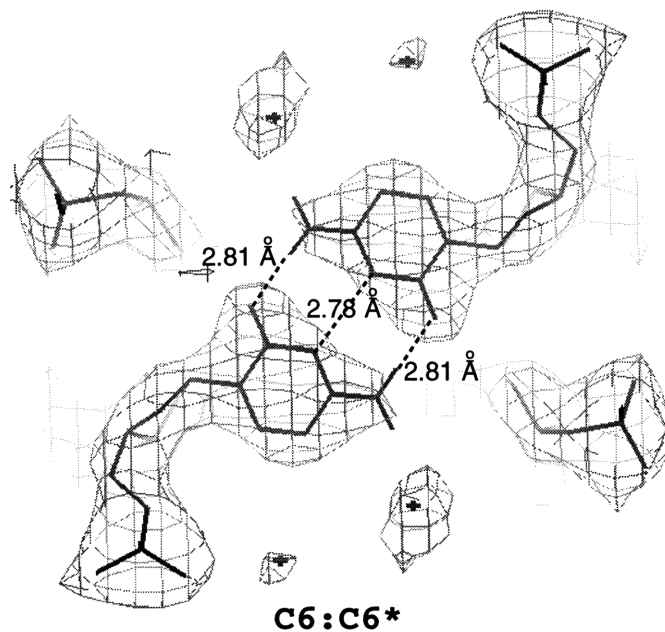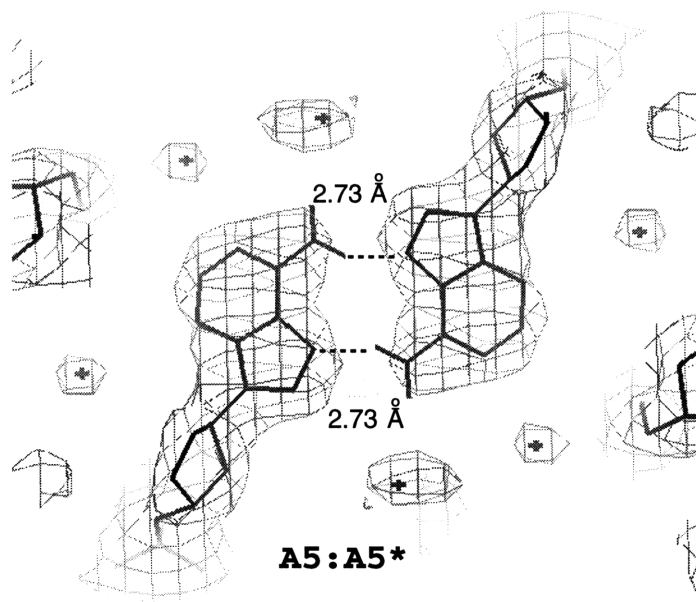**Figure 13. 2Fo–Fc Density Maps Around C6:C6* and A5:A5* Base Pairs**
(upper) A5:A5* base pair, indicating N6–N7 hydrogen bond distance. *(lower)* C6:C6*
base pair, indicating N3–N3* and O2–N4*(O4*–N4) hydrogen bond lengths.
Maps contoured at 1σ.
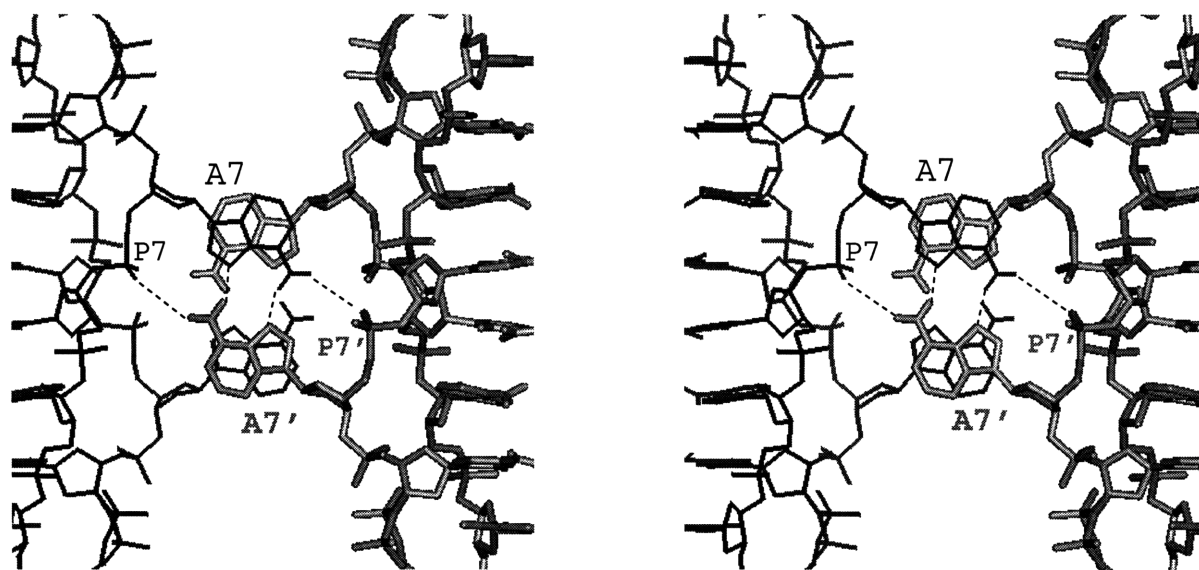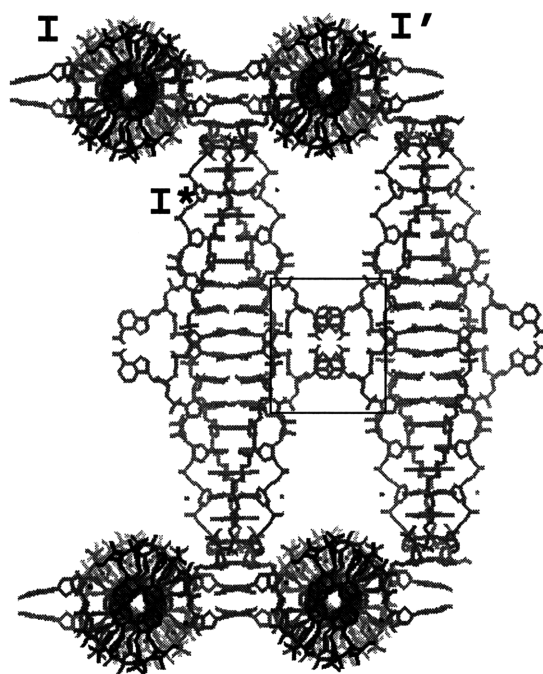
Figure 14. Two Neighboring Tetraplexes Are Connected by A7:A7 Base Pairs.
(upper) Neighboring tetraplexes I and I' contribute two A7 residues each to two layers of base pairs, which stack between the 5' overhanging base pairs (C1:C1*) of perpendicular helices. (I*). *(lower)* Stereo view of two layers of A7:A7' base pairs, showing the stacking and hydrogen bonding scheme. The A7:A7' interaction consists of hydrogen bonds between N6–N7 (2.89 $\overset{\circ}{A}$) and N7–O1P (3.4 $\overset{\circ}{A}$) from the same base pair partner.
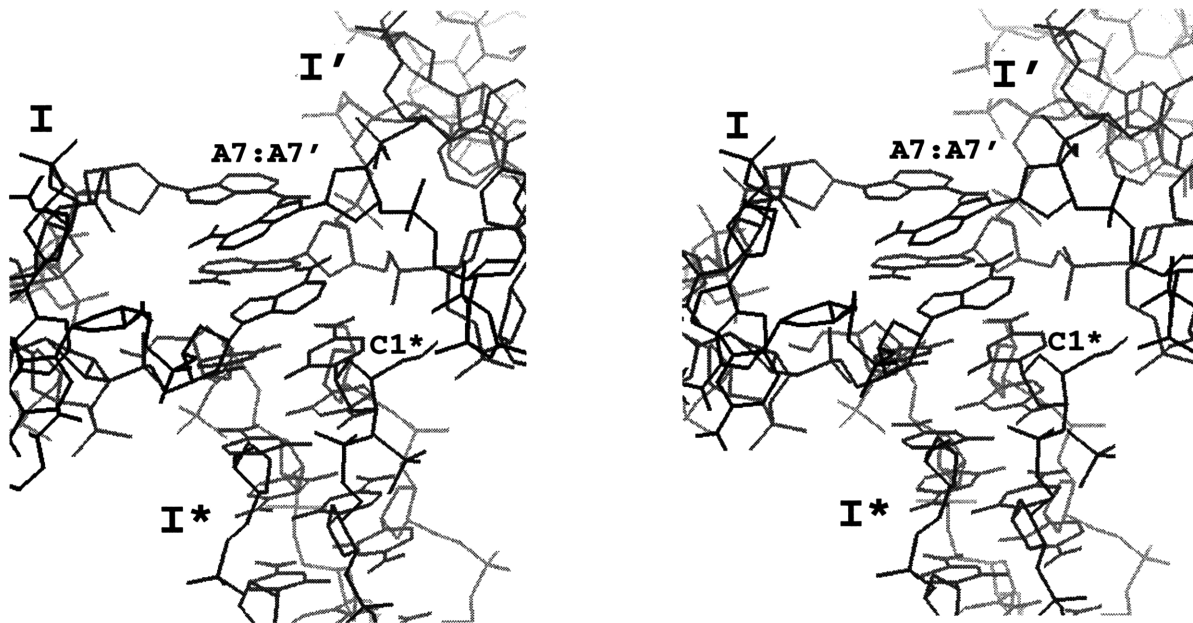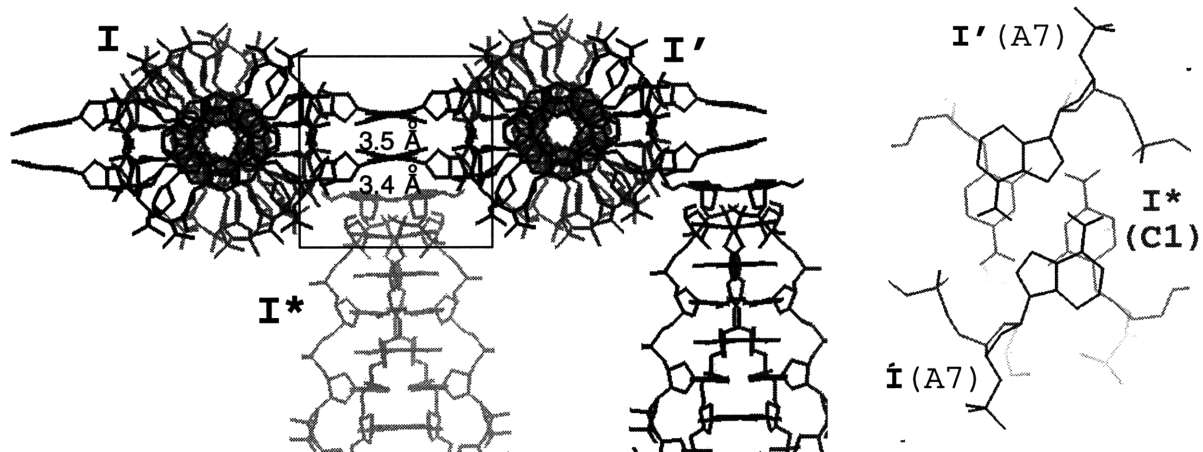
Figure 15. A7:A7 Base Stacking over NeighboringTetraplex.
(upper) Detail showing stacking of A7 :A7' over C1:C1*, with stacking distances.
(lower) Stereo view illustrating propeller twisting of A7 base pairs, which appears to influence propeller twist on C1.
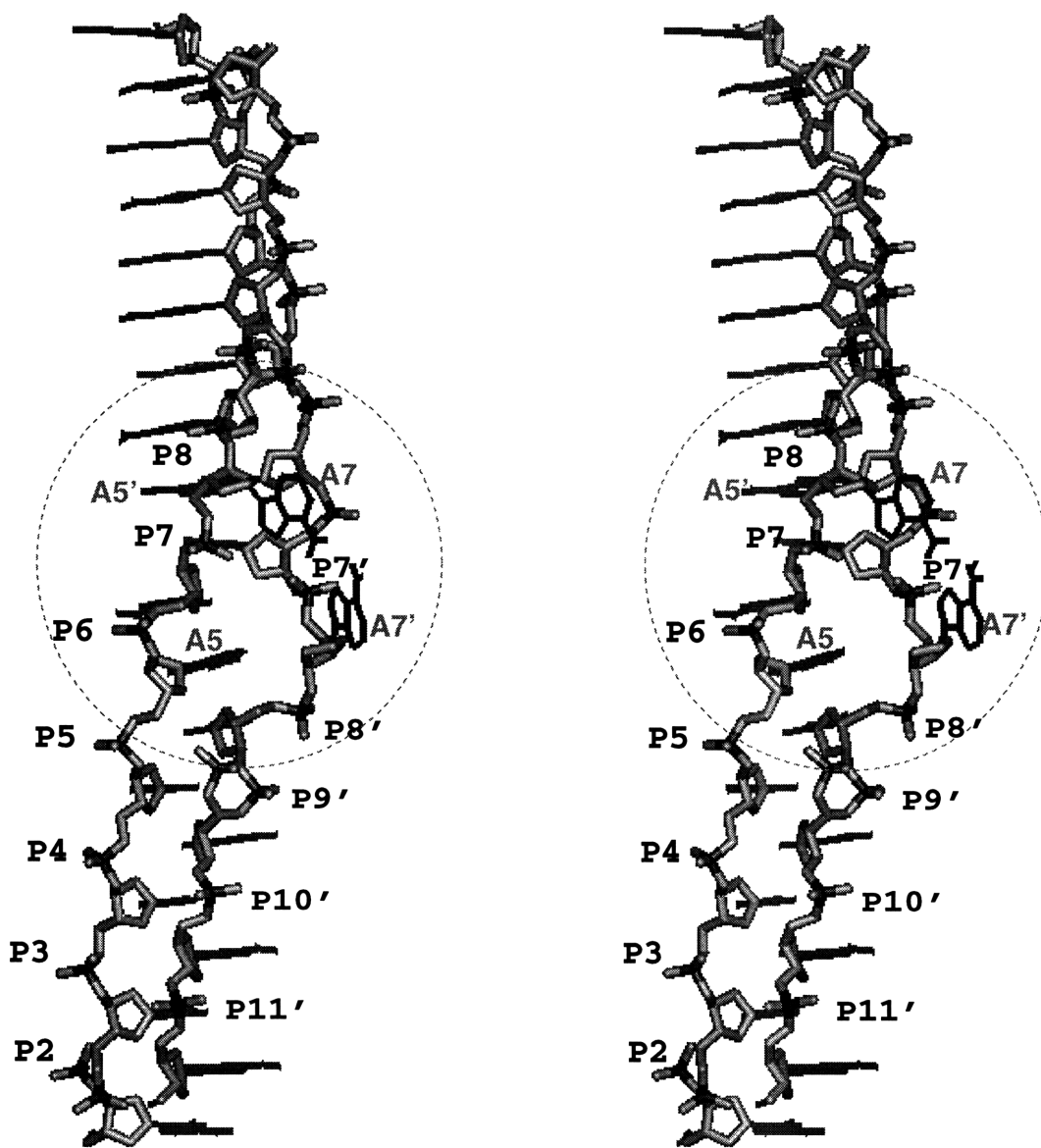
Figure 16. Stereo View of The Narrow Groove
Dotted circle denotes the area of the A5−C6−A7 knot.
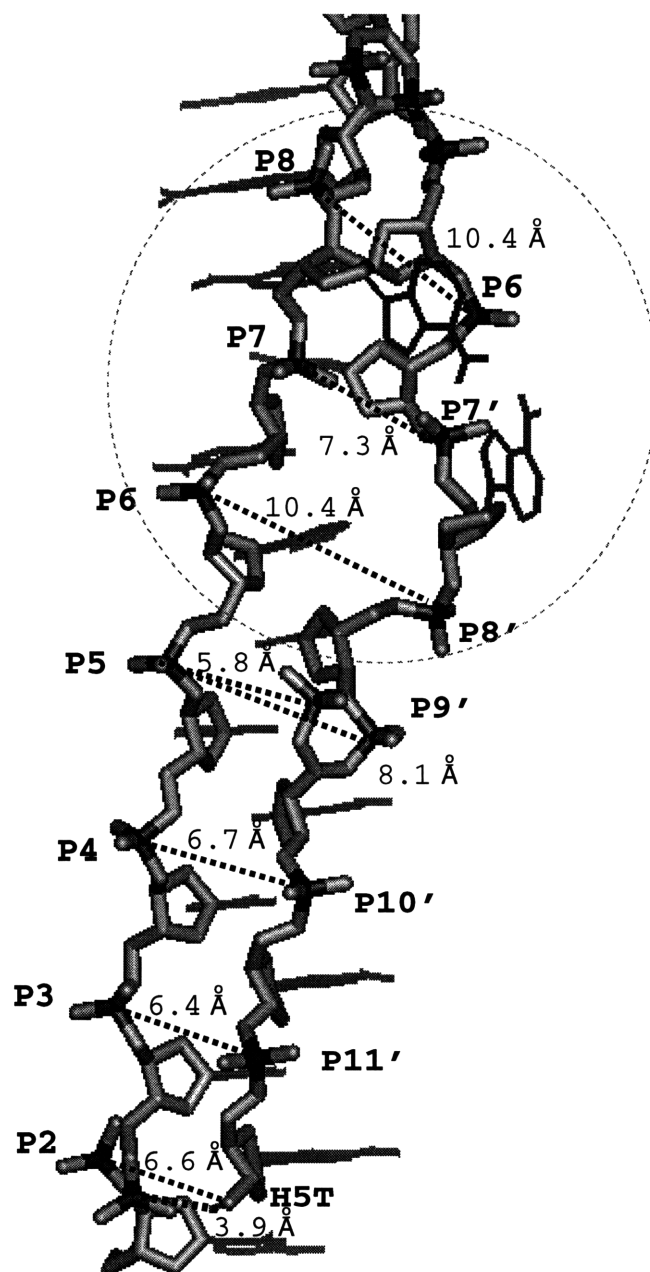
Figure 17. Interphoshate Distances Across The Narrow Groove.
Two antiparallel strands produce the narrow groove, which widens dramatically in the A5–C6–A7 knot (dotted circle). The 5' terminal hydrogen (H5T) on C1 is within hydrogen bonding distance of a P2 oxygen, in one of the P2 conformations.
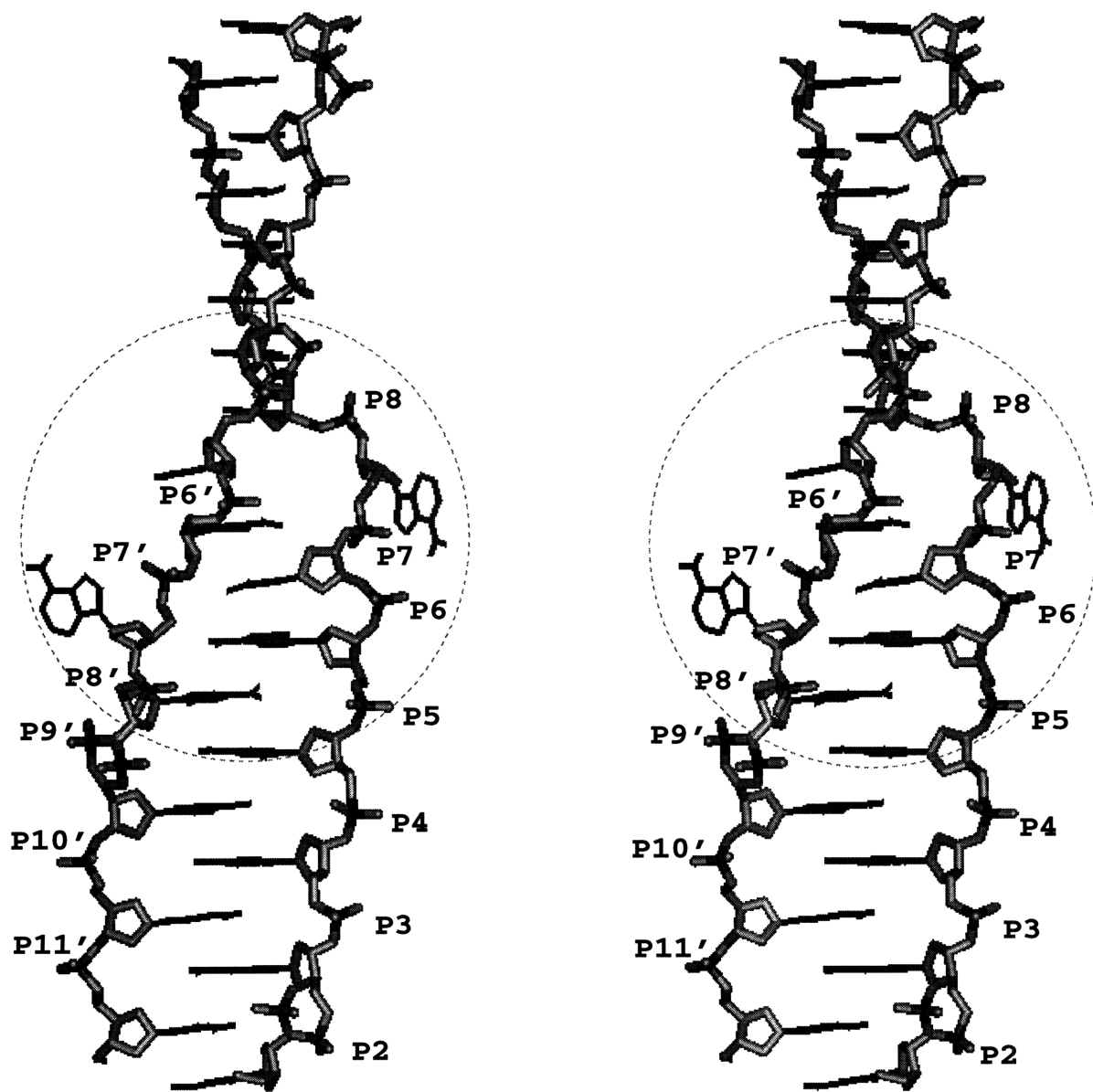
Figure 18. Stereo View of The Wide Groove
Dotted circle denotes the area of the A5–C6–A7 knot.
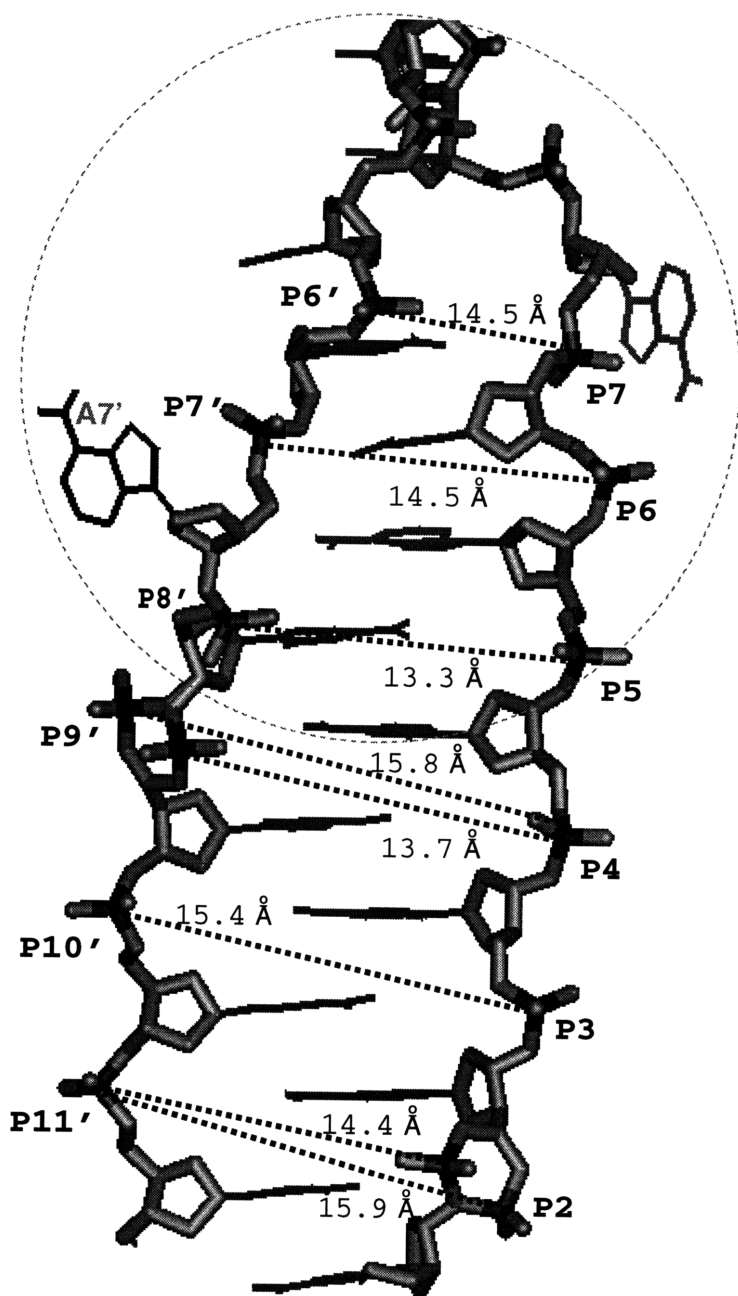
Figure 19. Interphoshate Distances Across The Wide Groove.
There is only a slight narrowing of the groove, where A7' exits the loop to C8'
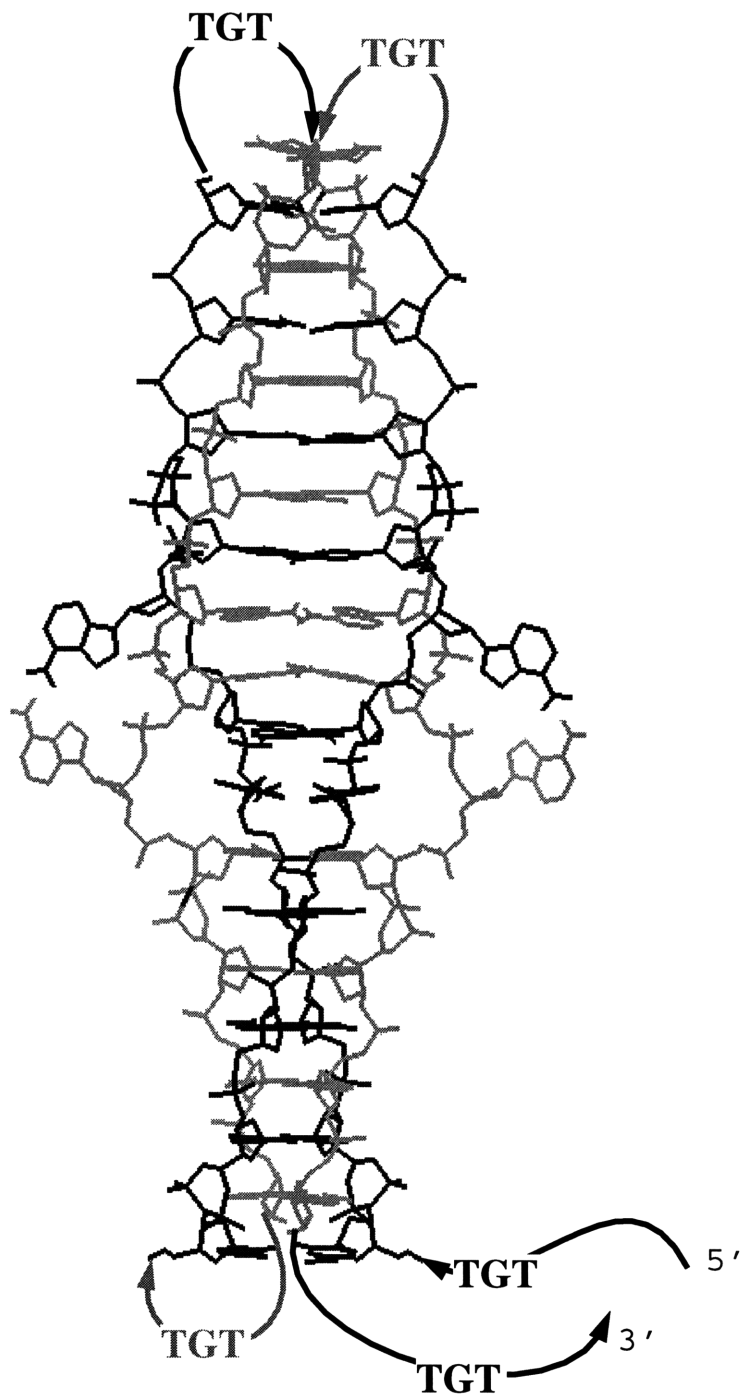
Figure 20. Proposed Organization for the Full ILPR Consensus Sequence.
Four contiguous repeats of (5'–TGTCCCCACACCCC–3') could use ACA as a knot, while using TGT as a hairpin loop to connect 5' to 3' ends. A TGT loop was observed in the NMR strucure of (5'–CCCCTGTCCCC).[23]

BIOGRAPHICAL NOTE

Curtis Alan Lockshin

Education   **Massachusetts Institute of Technology**   Cambridge, MA
Ph.D. Degree in Chemistry, June 1998.

Massachusetts Institute of Technology   Cambridge, MA
S.B. Degree in Life Sciences, June 1986.

Patents   No. 5,670,483 Iss: 9/23/97. "Stable macroscopic membranes formed by self-assembly of amphiphilic peptides and uses therefore". Inventors: Shuguang Zhang, Curtis Lockshin, Alexander Rich, Todd Holmes.

Publications

Schrier, R.D., Gnann, J.W. Jr., Landes, R., Lockshin, C., Richman, D., McCutchan, A., Kennedy, C., Oldstone, M.B.A., Nelson, J.A. (1989). T-cell recognition of HIV synthetic peptides in a natural infection. *J. Immunology* , **142**(4), 1166-1176.

Lubon, H., Ghazal, P., Hennighausen, L., Reynolds-Kohler, C., Lockshin, C., Nelson, J.A. (1989). Cell-specific activity of the modulator region in the human cytomegalovirus major immediate early gene. *Mol. Cell. Biol.*, **9**(3), 1342-1345.

Zhang, S., Lockshin, C., Herbert, A., Winter, E., Rich, A. (1992). Zuotin, a putative Z-DNA binding protein in Saccharomyces cerevisiae. *EMBO J.*, **11**(10), 3787-3796.

Zhang, S., Holmes, T., Lockshin, C., Rich, A. (1993). Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proc. Natl. Acad. Sci. USA*, **90**(8), 3334-3338.

Zhang, S., Lockshin, C., Cook, R., Rich, A. (1994). Unusually stable ß-sheet formation in an ionic self-complementary oligopeptide. *Biopolymers*, **34**(5), 663-672.

Kang, C.H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R., Rich, A. (1994). Crystal structure of intercalated four-stranded d(C3T) at 1.4 Å resolution. *Proc. Natl. Acad. Sci. USA*, **91**(24), 11636-11640.

Kang, C.H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R., Rich, A. (1995). Stable loop in the crystal structure of the intercalated four-stranded cytosine-rich metazoan telomere. *Proc. Natl. Acad. Sci. USA*, **92**(9), 3874, 3878.

Kang, C.H., Chan, R., Berger, I., Lockshin, C., Green, L., Gold, L., Rich, A. (1995). Crystal structure of the T4 regA translational regulator protein at 1.9 Å resolution. *Science*, **268**(5214), 1170-1173.