72

From Cytosine-tetraplexes to Adenine-clusters: Three Crystal Structures of DNA Telomeric Sequences

by

Li Cai

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1997

Author ....
Department of Physics
May 18, 1997

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

Accepted by ....
George F. Koster
Chairman, Departmental Committee on Graduate Students

From Cytosine-tetraplexes to Adenine-clusters: Three Crystal Structures of DNA Telomeric Sequences

by

Li Cai

Submitted to the Department of Physics
on May 18, 1997, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

Abstract

In my Ph.D. thesis, I investigate three crystal structres of sequences containing cytosine-rich strands found in telomeres, which are responsible for the stability of the chromosome and are important in both normal control of cell proliferation and abnormal growth of cancer.

The crystal structure of d(C_4), was the first X-ray structure of a novel intercalated DNA quadruplex, which consists of four strands forming two parallel-stranded duplexes with $C{\cdot}C^+$ base pairing, which intercalate to each other antiparallelly. The molecule has a small right-handed intrachain twist of 12.5°, and the base stacking distance along the tetraplex is 3.13 $\AA$.

The sequence d(AACCC) is the intermediate of both *Tetrahemena* (d(AACCCC)) and human (d(TAACCC)) telomeric sequences. The structure again shows the four stranded cytosine intercalation motif, with a significantly greater intrachain twist of 21.0°. The adenines form three different kinds of $A{\cdot}A$ base pairs, connecting cytosine tetraplexes in two orthogonal directions.

The crystal structure of d(AACCCC), telomeric cytosine-rich strand repeating sequence in *Tetrahymena*, showed two distinct cytosine tetraplexes. Each four-stranded complex is composed of two intercalated parallel-stranded duplexes pointing opposite directions, using hemiprotonated cytosine-cytosine base pairs. The outermost $C{\cdot}C^+$ base pairs, are from the 5' end of each strand in one cytosine tetraplex and from the 3' end of each strand in the other. The adenines form two different A-clusters in orthogonal directions, with their counterparts from other strands, using three base-paring modes. The A-clusters, along with the cytosine tetraplexes, form two alternating C-tetraplex-A-cluster stacking patterns, creating continuous base stacking along the x and z axes. The novel A-clusters could also be important modes of self-folding in ribozyme structures.

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

Acknowledgments

It has been a long time since I started working on these projects. First of all, I would like to thank Professor Alexander Rich for giving me the opportunity to work in his lab and to be involved in this research frontier. He introduced me to this and other exciting research areas, and his flexible management style has given me ample flexibility to explore various methodologies to solve a problem. I am grateful for his guidance through all these years. Even though I will be leaving this group soon, the tools that I have acquired will be with me forever.

The only reason that I can graduate before the turn of the century is due to two people I have been associated with through my graduate career: Professor Louis Osborne of Physics Department and Dr. Liqing Chen in our group. As my academic advisor and co-supervisor, Professor Osborne pays great attention to my research and personal progress and has been a vital source of advice whenever concerns and problems occur. Dr. Liqing Chen taught me the basics of crystallography as I joined this group and we have collaborated ever since. I always turn to him for advice whenever I encounter a difficult technical problem, and he is always there to answer. Over the years, our mentor-student relationship has grown to a longlasting friendship.

I would also like to thank the other members of the Rich lab. I have received great support from them over the years. I am especially indebted to: Curtis Lockshin, a fellow graduate student in the lab, whose overall deep knowledge of the x-ray instrumentation and insights in various other aspects of the field have been very helpful to me; Sridharan Raghavan and Niti Dube, two UROP students who have worked with me over an extented period of time. They are very intelligent and helpful, and we have had a great time working together.

Last but not least, I would like to thank my family members: parents Limin Xiong and Ti-Dao Tsai, and wife Chuan Hua Cai. It is their support, both morally and financially, that has made all this possible. My mother cooks for us whenever we go to her place, even though she does not enjoy cooking that much. Chuan pushes me to the lab every Sunday afternoon, so that she can stay away from the “annoying” football games.

Now I understand how much it takes to be a Ph.D. It takes all of the above.

Contents

List of Figures

List of Tables

Chapter 1

First Crystal Structure of An Intercalated DNA Tetraplex with $C{\cdot}C^{+}$ Base Pairs: d(C_4)

1.1 INTRODUCTION

Sequences containing C-rich strands are found in telomeres and may also occur in segments scattered throughout the genome (A more detailed discussion of the biology of telomere and the structural interest in alternative conformations of DNA is presented in Appendix A). It has long been known that nucleotide base cytosine could form three hydrogen bonds with another cytosine if they were hemiprotonated. This was first observed in the crystal structure of cytosine-5-acetic acid [1]. Although there was little doubt about this mode of base pairing($C{\cdot}C^{+}$), the number of strands involved in the structure has been controversial. A fiber x-ray diffraction pattern of polyribocytidylic acid was first interpreted as a parallel stranded duplex with such $C{\cdot}C^{+}$ base pairing [2], but later was shown to be more consistent with a single stranded model [3]. NMR study of $d(CT)_3$ indicated parallel $C{\cdot}C^{+}$ base pairing with T's extruded out [4]. More recent NMR experiments of $d(TC_5)$ and related sequences gave a rather surprisingly different structural motif: intercalated tetraplex (I-motif) in which the same $C{\cdot}C^+$ pairings were seen in two parallel stranded duplexes intercalated into each other antiparallelly [5, 6]. We have determined the first x-ray crystal structure of d(C_4) at 1.8$\AA$ resolution. Our results revealed the same I-motif structure as proposed by NMR studies, and in addition, the high resolution crystal structure allowed detailed visulization of many aspects of the four stranded intercalative system. C-tetraplex structures are likely to be found in telomeres and may be utilized in DNA self recognition or other pairing activities.

1.2 MATERIAL AND METHODS

1.2.1 DNA Synthesis and Crystallization

The oligodeoxyribonucleotide $d(C_4)$ was synthesized on a 10μM scale on an Applied Biosystems DNA synthesizer using solid phase β- cyanoethylphosphoramidite chemistry. It was then purified by reverse-phase high performance liquid chromatography (HPLC) on a C4 column (Rainin Instrument Co.), with a linear gradient of 5-40% acetonitrile in 0.1 M triethylammonium acetate buffer PH 7.0. The peak eluent containing the pure d(C_4) was collected and lyophilized overnight.

Crystals of d(C_4) were grown at room temperature using the vapor diffusion method (For interested readers, a more detailed discussion of crystal growth, derivative preparation and X-ray sources in data collection is presented in Appendix B). The best crystals were grown from a solution containing 100mM of Na-Cacodylate buffer (pH 5.0-pH 6.5) and 4mM of DNA sample (single strand concentration). It crystallized in space group I23 with unit cell dimensions a=b=c=82.3$\AA$. The crystal contains eight strands per asymmetric unit. The crystal data of the sample is summarized in Table 1.1.

Table 1.1: Crystal Data of d(C_4)

Space group	I23
a	82.3$\AA$
b	82.3$\AA$
c	82.3$\AA$
strands per unit cell	192
strands per asymmetric unit	8

1.2.2 Data Collection and Heavy Metal Soaking

All diffraction data were collected on a Rigaku Raxis II imaging plate system at room temperature and processed with the PROCESS program provided by the Molecular Structure Corporation. The native data used in phasing were collected to $2.0\AA$ resolution, 25 frames at a crystal-to-plate distance of 100mm with 2.5^o oscillations. After the structure was solved [7], another set of native data were collected to $1.8\AA$ resolution (60 frames, 3.0^o oscillation at crystal-to-plate distance of 100mm). They were used to finish the refinement.

An extensive search was carried out for heavy atom derivatives. Initial soaking with K_2PtCl_6 showed some promising results. Later systematic soaking by changing K_2PtCl_6 concentrations and/or soaking time were performed in order to optimize the heavy atom signal. The crystal for a platinum derivative were obtained by adding 1mM K_2PtCl_6 to the mother liquor and soaking the crystal for 25 days. The derivative data were collected to a resolution of $1.8\AA$ in 40 frames, with oscillation of 2.5^o and a crystal-to-plate distance of 100mm. Table 1.2 contains a summary of the data collection statistics.

1.2.3 Structure Determination by Only One Derivative

Determination of Platinum Position

The crystal used in phasing was soaked in 1mM K_2PtCl_6 for 25 days, and produced good isomorphous difference as well as anomalous scattering signals. When compared to the native data set, the heavy atom data set revealed an R factor of 23.0% at 2.0 $\AA$ resolution. The position of Pt atom was readily obtained from the isomorphous difference Patterson map and the anomalous difference map, using the Patterson search program HASSP [8]. Both the isomorphous difference Patterson and the anomalous difference Patterson maps at different resolution ranges showed the same set of strong clear peaks which could be easily interpreted as a single Platinum site, as shown in Figure 1-1. This single site is further confirmed by the Patterson search HASSP (Table 1.3). The Pt atom was on a three-fold axis in the cubic lattice. This was

Table 1.2: Summary of Data Collection Statistics

	Native	Native	$K_2P_tCl_6$
Resolution ($\AA$)	2.0	1.8	1.8
Number of observations	33025	93579	60583
Number of unique reflections	6078	8335	7797
Overall completeness (%)	95.1	95.3	89.1
Outermost shell ($\AA$)	2.25-2.0	2.0-1.8	2.0-1.8
Outermost shell completeness (%)	88.0	82.9	64.9
$\mathbf{I/\sigma(I) > 1}$ **(%)**	93.9	94.8	88.2
$\mathbf{I/\sigma(I) > 1}$ **outermost shell(%)**	86.0	82.6	64.7
$\mathbf{R_{merge}}$ **(%)**	7.1	6.3	6.8

The first native data was used in the phasing and initial refinement. The second native data was used in the final refinement. R_{merge} is the agreement R-factor between symmetry-related observations.

true only for space group I23, since the heavy atom peaks in the Patterson maps could not be interpreted with another space group $I2_13$, which has the same Laue diffraction symmetry as I23 and thus cannot be precluded as the possible space group of the d(C_4) crystal based on the diffraction symmetry alone. The largest Bijvoet differences (top 25%) and isomorphous centric reflections to a resolution of 2.0$\AA$ were then used separately to refine the position of the Pt atom by a least-squares program SHELX [9]. The position of Pt atom remains at x=0.194, y=0.194, z=0.194 after the SHELX refinement. No additional sites were observed upon inspection of a difference Fourier map calculated using SIRSAS phases from the initial Pt site.

Determination of Phases

The iterative single isomorphous replacement and single-wavelength anomalous scattering method (ISIRSAS) as described by Wang [10], in which a noise filtering pro-

Table 1.3: Summary of the Patterson Search from HASSP

List of Major Peaks in Isomorphous Single-Atom Search

Peak	X	Y	Z	Height	Probability by Chance
1	0.493	0.188	0.243	421.66	0.790
2	0.194	0.250	0.000	367.60	0.997
3	0.000	0.000	0.000	703.29	1.000
4	0.194	0.194	0.194	686.64	0.000
5	0.104	0.500	0.500	469.11	1.000

List of Major Peaks in Anomalous Single-Atom Search

Peak	X	Y	Z	Height	Probability by Chance
1	0.069	0.375	0.028	608.52	0.579
2	0.486	0.139	0.097	528.63	0.656
3	0.014	0.396	0.271	507.31	0.642
4	0.208	0.264	0.042	484.12	0.811
5	0.194	0.194	0.194	911.80	0.000
6	0.000	0.000	0.000	582.01	1.000

cedure is applied to resolve the phase ambiguity associated with SIR and SAS data to improve the initial phases, was used to generate the DNA phases. The phasing power of the SIR and SAS data (Table 1.4), are listed below. The ISIR procedure was first applied to the isomorphous data, whereas ISAS procedure was applied to the anomalous data, both to 2.0Å. The figure of merit after the SIR step is 0.28, with 5331 accepted reflection pairs, and the figure of merit after the SAS step is 0.33, with 2280 accepted reflection pairs. Subsequently, the SIR and SAS data were merged, with a figure of merit of 0.38 and 5473 paired reflections. Four cycles of iterations were then applied to the combined data, to resolve the phase ambiguity to 2.0 Å. The final cycle had an average figure of merit of 0.71, a map inversion R-factor of 0.342, and a correlation coefficient of 0.929.

The above phasing procedure also allowed the handedness of the coordinates to be determined. We have two enantiomers, "+ + +" and "- - -" handed, the second of which was created by setting the three coordinates of the Pt position to its negative. The statistics for the "- - -" enantiomer were described as above. For the "+ + +" handed enantiomer, the same filtering procedure was performed, and the final figure of merit was 0.70, the map inversion R-factor was 0.395, and the correlation coefficient was 0.891. A comparison of the final phasing statistics between the two enatiomers indicated that the "- - -" handed enatiomer was correct (Table 1.5). This choice was further verified by the interpretability of the electron density generated by the "- - -" set of phases. The electron density map generated with "- - -" set of phases clearly showed the intercalated structure (see Figure 1-2). The sugar phosphate backbones were clearly visible with intense peaks at the phosphate positions. In addition, individı al bases were visible in the map. A model was built into the density map with the program FRODO [11].

Refinement of the Structure

The model was first refined to 2.3 $\AA$. The initial R-factor was 44% for the data from 8 to 2.3 $\AA$. Refinements were then carried out using the program XPLOR [12]. Simulated annealing was used, and the R-factor fell rapidly to 23%. The original map was checked frequently during the refinement process. High resolution native data, which doubled in number of total reflections, were then used, to extend the refinement to 1.8$\AA$. Twenty cycles of restrained individual isotropic B-factor refinement followed. Well ordered water molecules were then located from the difference Fourier map$(F_o - F_c)$ and added as oxygen atoms to the model only if they had a peak height of over 3 σ in the difference density map and formed hydrogen bonds with DNA atoms or other molecules. A total of 46 water molecules were found this way. A final round of refinement completed the structural determination with an R-factor of 0.218 and r.m.s. deviation from ideal bond lengths and angles of 0.015$\AA$ and 3.7^o, respectively.

A summary of the refinement statistics is listed in Table 1.6. Figure 1-3 shows the refined model superimposed with the final 2Fo-Fc map. The atomic coordinates have been deposited in the Protein Data Bank.

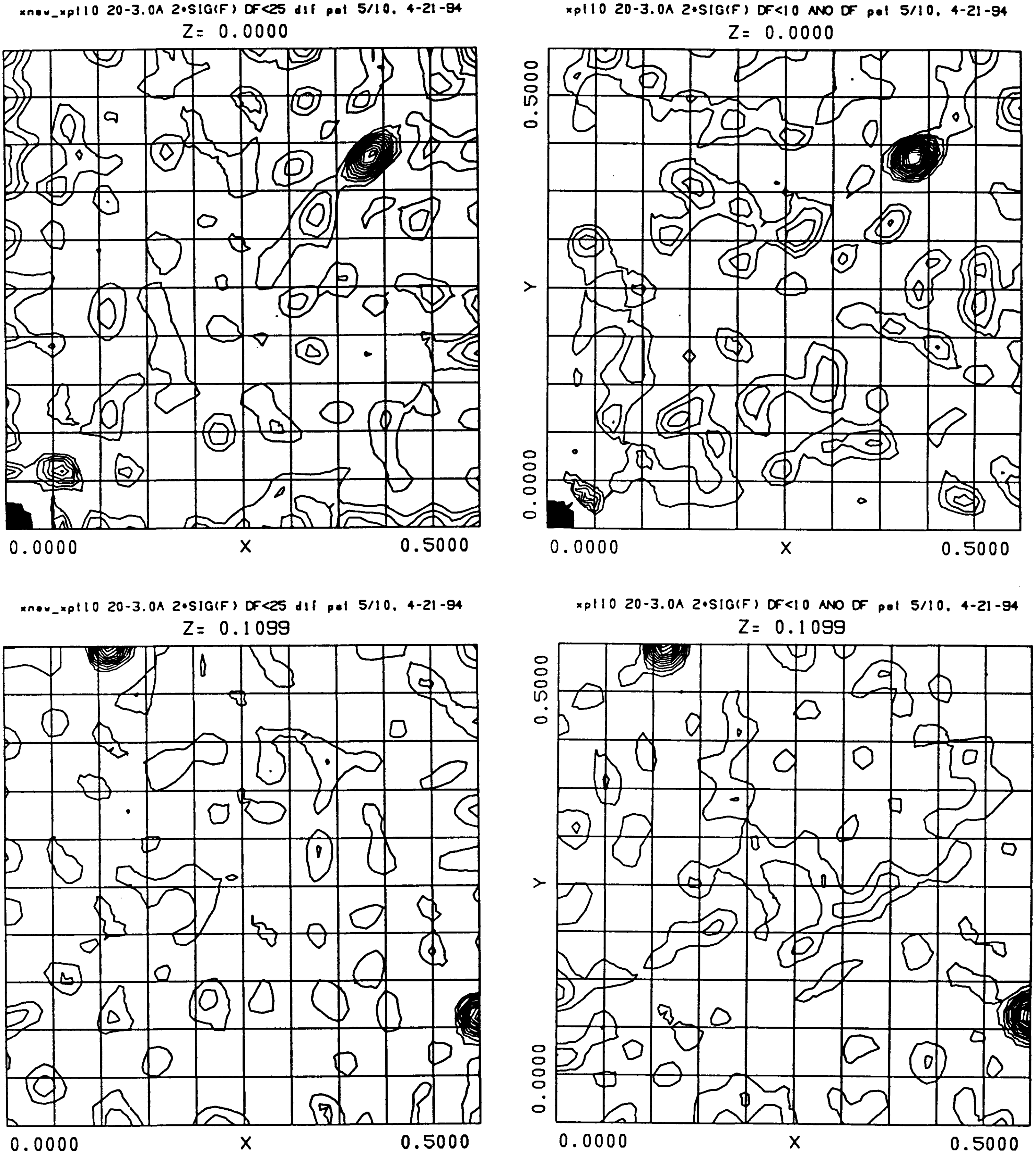

Figure 1-1: Two Sections of the Isomorphous Difference Patterson and Anomalous Difference Patterson Maps for the Crystal of d(C_4).

Table 1.4: Phasing Power of the SIR and SAS Data

SIR Data

Resolution($\mathring{A}$)	Phasing Power	Reflections
8.53	1.72	325
4.88	1.42	325
4.06	1.58	325
3.60	1.59	325
3.30	1.42	325
3.08	1.40	325
2.90	1.61	325
2.76	1.56	325
2.64	1.53	325
2.54	1.51	325
2.50	1.80	1
Total	1.54	3251

Phasing Power=$< ||F_{PH}| - |F_P|| > / < |(F_{PH}|_{obs} - |F_P|_{obs}) - (|F_{PH}|_{cal} - |F_P|_{cal})| >$

SAS Data

Resolution($\mathring{A}$)	Phasing Power	Reflections
8.00	2.38	229
4.92	2.25	229
4.15	2.12	229
3.71	1.90	229
3.40	1.87	229
3.17	1.83	229
3.00	1.91	229
2.84	1.57	229
2.69	1.63	229
2.56	1.53	229
2.50	1.56	4
Total	1.93	2294

Phasing Power=$< 2F''_H > / < |(F^+_{PH}|_{obs} - |F^-_{PH}|_{obs}) - (|F^+_{PH}|_{cal} - |F^-_{PH}|_{cal})| >$

Table 1.5: Summary of Phasing Statistics

(a) Initial phase preparation

	SIR	SAS	Merged (SIR+SAS)
Resolution limit(Å)	2.5	2.5	2.5
Reflection pairs	3110	2297	3198
Completeness (%)	94	69	96
R_{diff} (%)	17.6	5.1	
Figure of merit	0.28	0.39	0.47

(b) Handedness test

	+ hand	- hand
Initial figure of merit	0.47	0.47
After solvent flattening:		
Figure of merit	0.68	0.73
Map inversion R-factor	0.411	0.288
Correlation coefficient	0.877	0.949
Map interpretability	No	Yes

Resolution limit is the highest order of data used in phasing and is not the diffraction limit of the crystal. $R_{diff}=\Sigma \mid F_{native} - F_{derivative} \mid / \Sigma \mid F_{native} \mid$ for SIR data and $R_{diff}=\Sigma \mid F_{+} - F_{-} \mid / \Sigma \mid F \mid$ for SAS data. The coordinates for + hand are X=0.194, Y=0.194, Z=0.194; For - hand, signs of X, Y, Z of above are reversed.

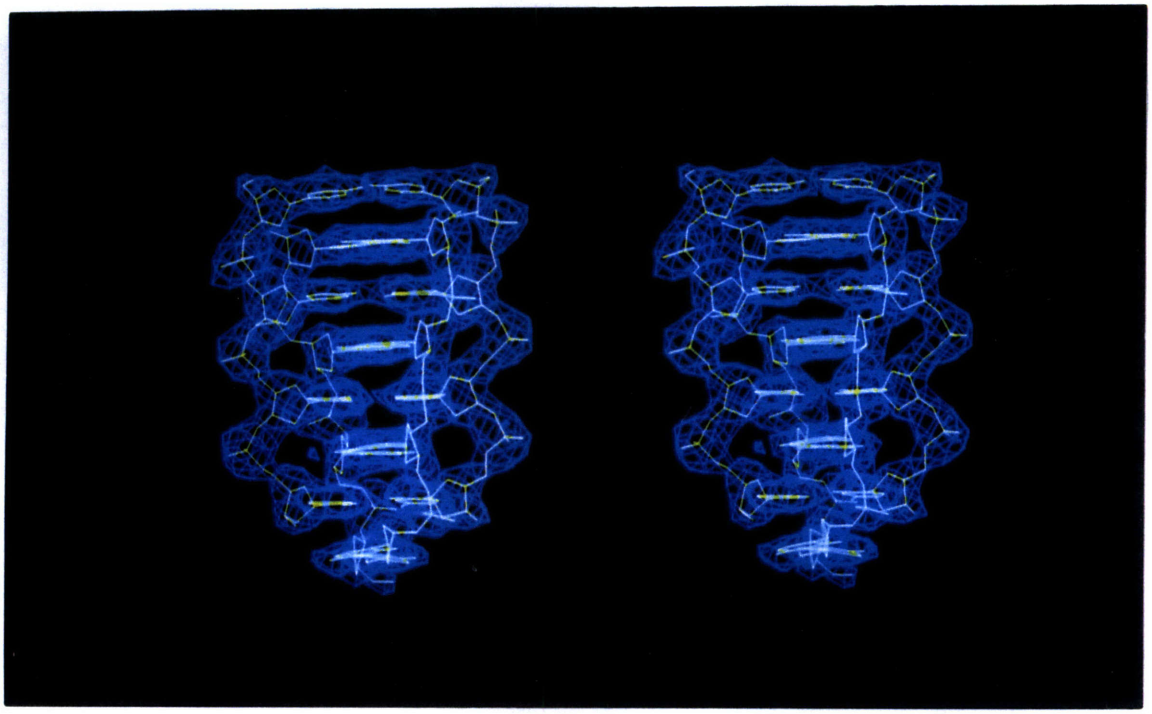

Figure 1-2: Initial (SIR+SAS) Electron Density Map (Blue) After Solvent Flattening. Inside the Density Map Is the Final Refined Model.

Table 1.6: Refinement Statistics

Resolution ($\AA$)	**8-1.8**
Number of Reflections (I>1σ(I))	8177
Completeness (%)	95.1
Number of non-hydrogen DNA atoms	584
Number of water molecules	46
r.m.s. bond length($\AA$)	0.015
r.m.s. bond angles(^{o})	3.7
R-factor	0.218

$R=\Sigma \mid F_{observed} - F_{calculated} \mid / \Sigma \mid F_{observed} \mid$

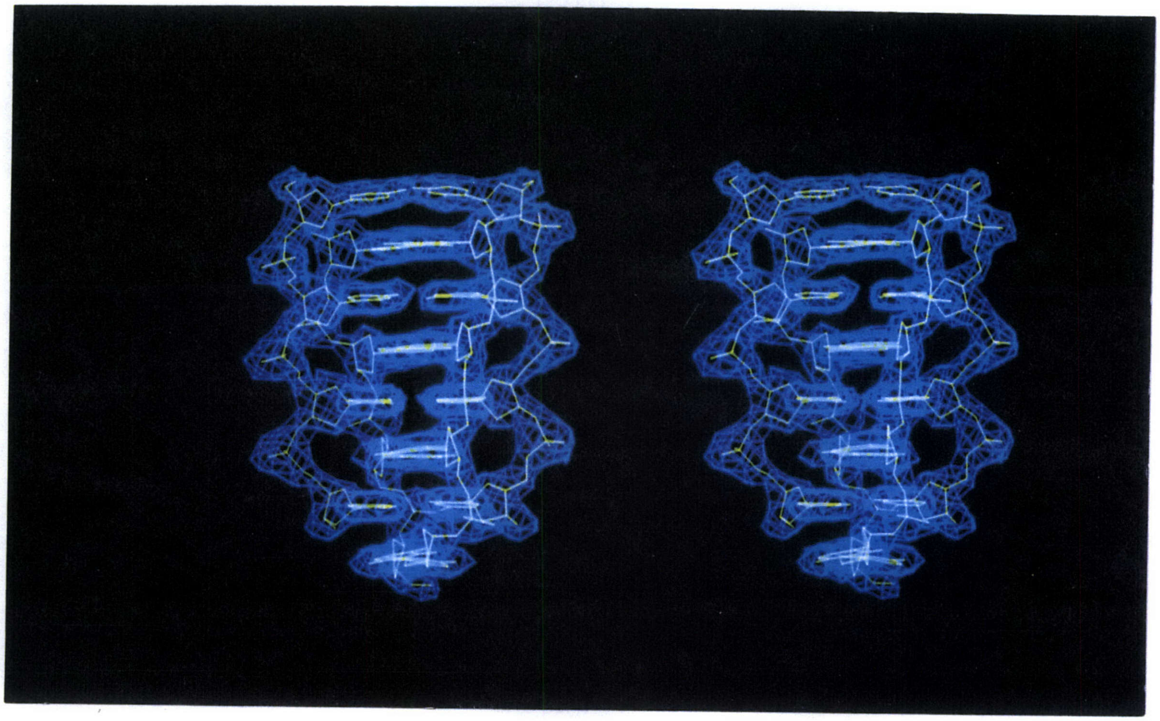

Figure 1-3: Final Refined Model Superimposed with the Calculated 2Fo-Fc Map.

1.3 RESULTS

1.3.1 Overview of the structure

Figure 1-4 shows the overall features of the d(C_4) quadruplex. The molecule consists of four d(C_4) strands which form two parallel-stranded duplexes with $C{\cdot}C^+$ base pairings. The duplexes intercalate into each other with opposite chain polarities(5'-3'). The relative positioning of the four strands is such that their sugar-phosphate backbones are grouped into two closely packed pairs. Each pair is composed of one strand from each parallel duplex. The quadruplex thus has two wide grooves and two narrow grooves and is flat shaped.

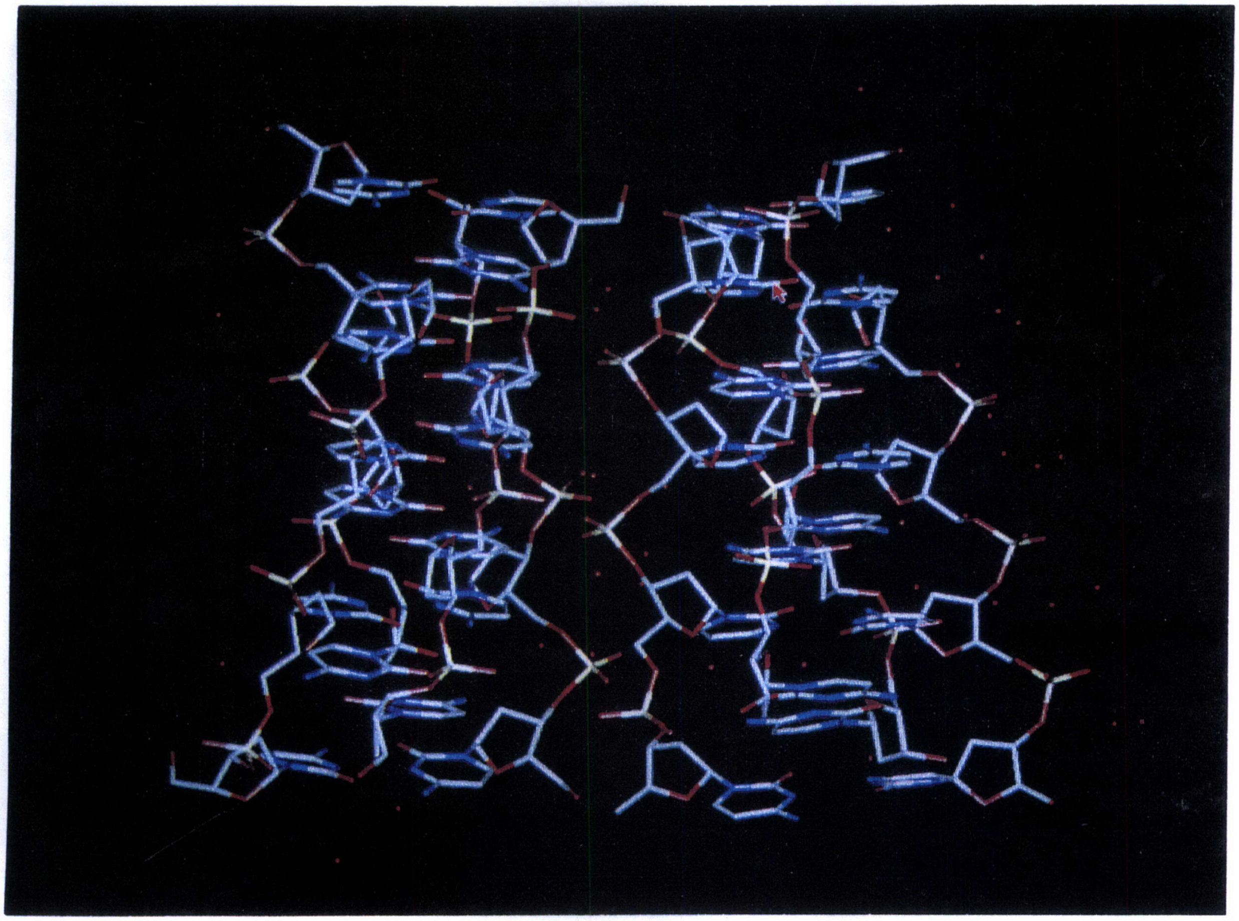

Figure 1-4: Overviews of the d(C_4) Quadruplex.

1.3.2 Comparison of tetraplexes A and B

The unit cell of the d(C_4) crystal is very big and contains a total of 192 strands, of which 8 strands are crystallographically independent. Thus the asymmetric unit consists of two tetraplexes labelled as A and B. Each tetraplex has 16 nucleotides which are numbered from 1 to 16 for tetraplex A and 21 to 36 for tetraplex B, respectively. The two quadruplexes are very similar to each other, and have an r.m.s. positional difference of 0.51$\AA$ when they are superimposed (Figure 1-5). The smallest difference is between fourth strands, residue C13 to C16 of A versus C33 to C36 of B, which have r.m.s. difference of just 0.19$\AA$, and the two strands are almost identical. The third strands, C9-C12 of A versus C29-C32 of B, show the biggest r.m.s. deviation of 0.63$\AA$.

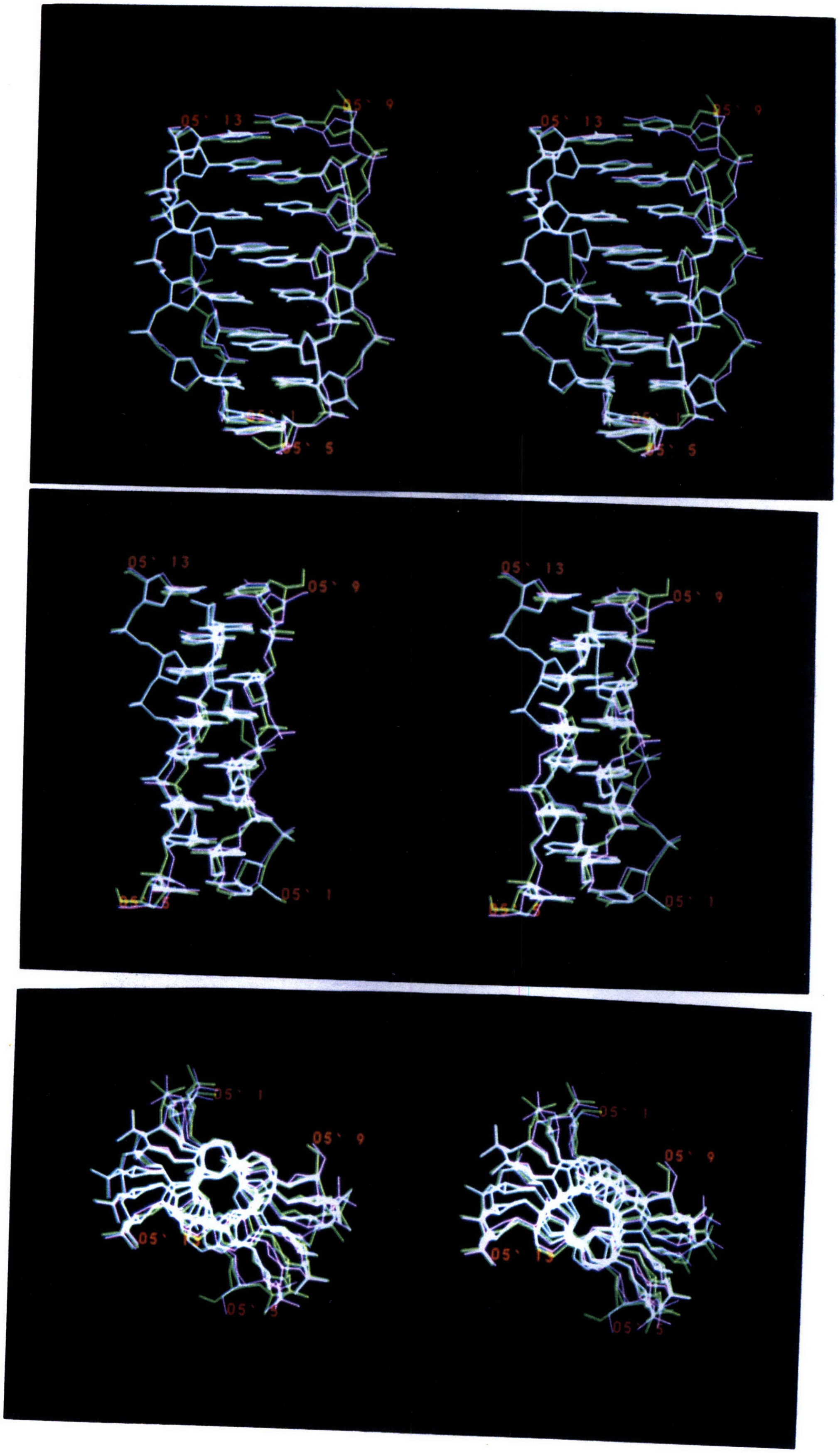

Figure 1-5: Superposition of Tetraplexes A and B in Stereo.

It should be pointed out that the major differences occur along the sugar-phosphate backbones, especially at phosphate positions, such as those at phosphates 3, 6, 7, 11 and 12. Residue C9 exhibits large difference from residue C29 due to the different crystal packing environment of tetramers A and B (see crystal packing section). Cytosine bases show much better agreement between A and B with an r.m.s. difference of $0.25\AA$ which reflects the less flexible nature of $C{\cdot}C^+$ base pairing due to three strong planar hydrogen bonds. In contrast, the sugar-phosphate backbones are intrinsically more flexible partly due to their lack of torsional restraints and partly due to strong electrostatic repulsion between phosphate groups in the narrow grooves.

1.3.3 $C{\cdot}C^+$ base pairing

In order for cytosine to form three interbase hydrogen bonds, it must be hemiprotonated at imino position (N3). This can be achieved by lower PH to acidic range [13]. The d(C_4) crystals could be grown at pH range 4.0 to 6.0 and thus were expected to contain such hemiprotonated *cytosine·cytosine* base pairs. Our results clearly showed that eight such $C{\cdot}C^+$ base pairs were found in quadruplex. The imino-imino($N3{\cdots}N3$) distances vary from $2.65\AA$ to $2.92\AA$ in tetraplexes A and B (Table 1.7), with the average to be 2.76±0.07 $\AA$. The two flanking, carbonyl-amino hydrogen bonds($O2{\cdots}N4$) have a distance range from 2.62 $\AA$ to $2.93\AA$, with the average being 2.75±0.08 $\AA$. These hydrogen bonds are somewhat shorter than those reported in the crystal structure of cytosine-5-acetic acid [1] and 1-methylcytosine hemi-hydroiodide hemihydrate [14]. In the case of cytosine-5-acetic acid, $N3{\cdots}N3$ and $O2{\cdots}N4$ distances are $2.82\AA$ and $2.79\AA$, respectively, while the corresponding bonds in 1-methlcytosine are $2.83\AA$ and $2.76\AA$. We think that the shorter hydrogen bonds in the d(C_4) quadruplex are the results of stronger interbase interactions caused by enhanced cooperativity among so many intercalation base pairs.

Table 1.7: Hydrogen Bonds of $C{\cdot}C^+$ Base Pairs

Tetraplex	Base pair	$O2\cdots N4'$ ($\AA$)	$N3\cdots N3'$ ($\AA$)	$N4\cdots O2'$ ($\AA$)
A	C1-C5	2.77	2.79	2.87
	C2-C6	2.82	2.87	2.81
	C3-C7	2.80	2.80	2.68
	C4-C8	2.78	2.70	2.64
	C9-C13	2.76	2.81	2.93
	C10-C14	2.69	2.78	2.77
	C11-C15	2.76	2.78	2.66
	C12-C16	2.68	2.74	2.70
B	C21-C25	2.62	2.67	2.73
	C22-C26	2.69	2.74	2.76
	C23-C27	2.79	2.71	2.67
	C24-C28	2.88	2.92	2.84
	C29-C33	2.78	2.79	2.78
	C30-C34	2.64	2.71	2.77
	C31-C35	2.77	2.77	2.81
	C32-C36	2.70	2.65	2.65
	Average	2.75	2.76	2.75
	Standard Deviation	0.07	0.07	0.08

1.3.4 Base Stacking Mode

One of the remarkable features about the quadruplex structures is its unique mode of base-base stacking(Figure 1-4). In contrast to the conventional intra-duplex base stacking seen in the antiparallel paired B- and Z-DNA structures, the $C{\cdot}C^+$ base pairs form one parallel stranded duplex stacking on those from another parallel duplex, which runs in the opposite direction, thus creating an antiparallelly intercalated structure. It is interesting to note the two slightly different ways these adjacent base pairs stack on each other (Figure 1-6). There are seven such base steps in the eight-layered molecule, each step has a base stacking distance (rise) of 3.13$\AA$ on average (Table 1.8). Only the exocyclic carbonyl(O2) and amino(N4) groups are significantly involved in stacking and are overlapping with their dipole, pointing to approximately opposite directions. More overlapping of amino groups than carbonyl ones are seen in four of the seven base steps, while the reverse is true for the rest three. The alternation in overlap difference is due to the fact that there exist two different types of base step twist (Table 1.8), which are alternating along helical axis. One is right-handed twist of 37.4^o on average and the other is left-handed twist of -24.4^o. Thus we see an alternation of slightly wider and narrower minor grooves, which are consistent with the extent carbonyl and amino groups overlap. Since the rise and twist described here are correlating base pairs from different duplexes, we would like to call this pseudo-rise and pseudo-twist, respectively, in order to distinguish them from those intra-duplex properties usually associated with covalently linked bases. The introduction of the pseudo parameters is necessary to fully describe the intercalated structure. The parallel paired duplexes have on average a helical rise of 6.26 $\AA$ and a small right-handed twist of 12.5^o (Table1.9). The rises are quite constant through all the four duplexes, but the twists vary a lot ranging from 8.5^o to 16.2^o. In fact tetraplex B has a larger twist (13^o) than tetraplex A (12.1^o). The relatively large standard deviation (2.6^o) of twist angles reflects the flexibility of the structure.

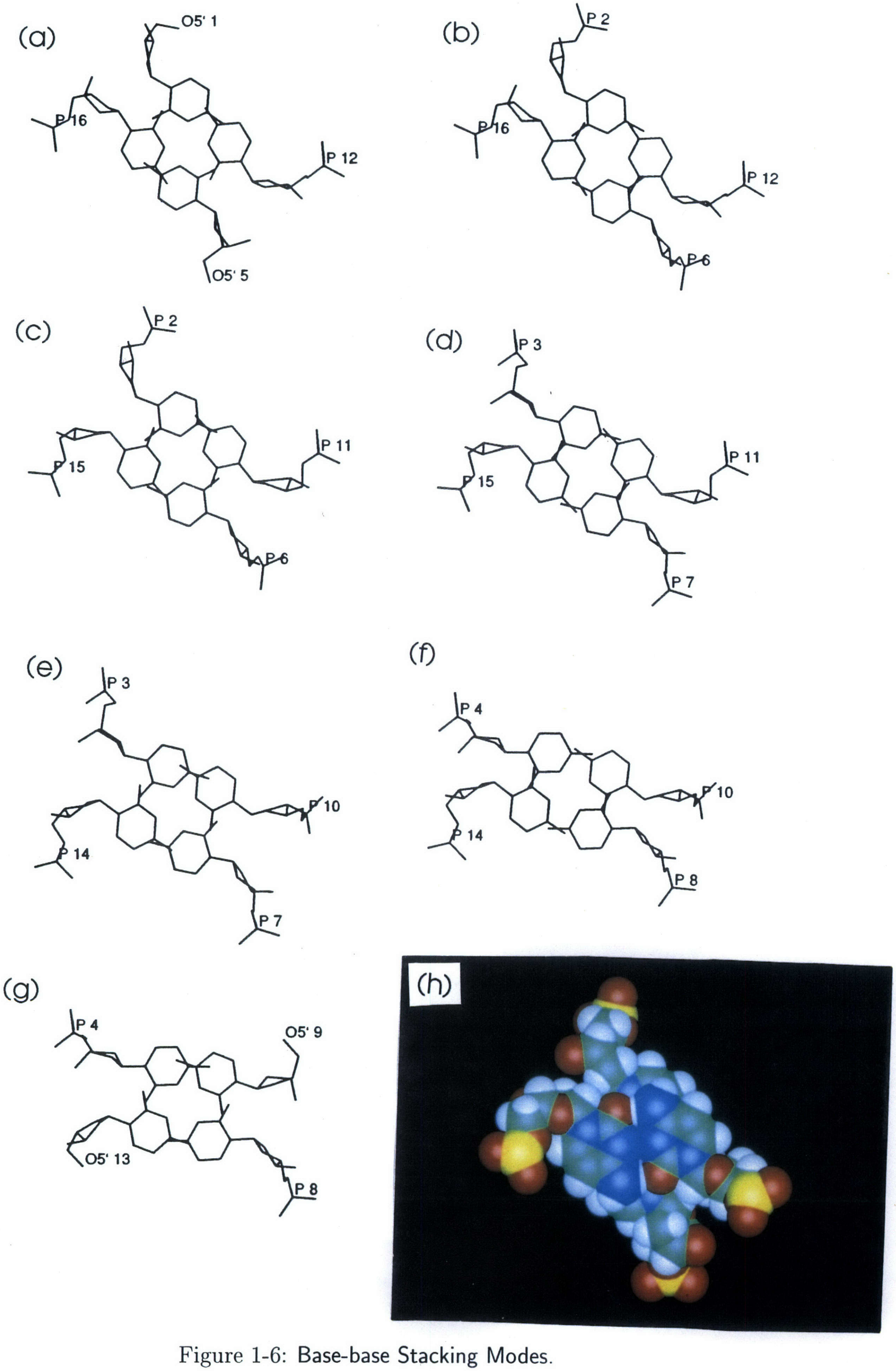

Figure 1-6: Base-base Stacking Modes.

Table 1.8: Quadruplex Parameters

Quadruplex	Base step	Pseudo-Rise(Å)	Pseudo-Twist(°)
A	C1·C5-C16·C12	3.17	37.1
	C16·C12-C2·C6	3.13	-25.8
	C2·C6-C15·C11	3.12	34.1
	C15·C11-C3·C7	3.10	-25.4
	C3·C7-C14·C10	3.12	38.5
	C14·C10-C4·C8	3.13	-23.3
	C4·C8-C13·C9	3.11	39.3
B	C21·C25-C36·C32	3.12	40.7
	C36·C32-C22·C26	3.10	-25.8
	C22·C26-C35·C31	3.17	36.1
	C35·C31-C23·C27	3.11	-23.7
	C23·C27-C34·C30	3.19	34.8
	C34·C30-C24·C28	3.11	-22.4
	C24·C28-C33·C29	3.12	38.5
	Overall average	3.13	37.4/-24.4
	Overall standard deviation	0.03	2.3/1.5

Right-handed twist is positive, left-handed twist is negative.

Table 1.9: Duplex Parameters

Parallel duplex	Base step	Rise($\mathring{A}$)	Twist(^{o})
A1	C1·C5-C2·C6	6.30	11.3
	C2·C6-C3·C7	6.22	8.7
	C3·C7-C4·C8	6.26	15.2
A2	C9·C13-C10·C14	6.25	16.0
	C10·C14-C11·C15	6.23	13.1
	C11·C15-C12·C16	6.25	8.5
B1	C21·C25-C22·C26	6.22	15.0
	C22·C26-C23·C27	6.28	12.5
	C23·C27-C24·C28	6.31	12.4
B2	C29·C33-C30·C34	6.23	16.2
	C30·C34-C31·C35	6.30	11.2
	C31·C35-C32·C36	6.27	10.5
	A Average	6.25	12.1
	A Standard Deviation	0.03	3.2
	B Average	6.27	13.0
	B Standard Deviation	0.04	2.2
	Overall average	6.26	12.5
	Overall standard deviation	0.03	2.6

1.3.5 Wide and Narrow Grooves

The intercalation of two parallel duplexes yields a quadruplex with two very wide and two very narrow grooves which are basically symmetrical about the helical axis. The narrow groove is made of two closely packed strands running antiparallelly (Figure 1-7). The two backbone chains fit to each other remarkably well in a zig-zag way. They are so close to each other that some inter-chain P-P distances are even shorter than intra-chain ones. The shortest inter-chain P-P distance is 5.54$\AA$ between P20 and P31, while the shortest intra-chain P-P distance is 6.03$\AA$(P35-P36) (Table 1.10). The shortest P-P distance across a wide groove is 14.5$\AA$, between P2 and P12. The average intra-chain P-P distance is 6.5±0.4 $\AA$, and the corresponding inter-chain P-P distance across a narrow groove is 7.9±1.4 $\AA$, and 15.8±0.7 $\AA$ across the wide groove. So the P-P separation in wide groove is twice as much as in the minor groove. If we take the Van der Waals radius of phosphate(PO_4) as 2.9$\AA$(roughly), and deduct twice that from the P-P distances, we obtain the widths of the wide and narrow grooves to be approximately 10$\AA$ and 2.1$\AA$ respectively. Thus the molecule has a pair of major grooves, which are so wide that they can accommodate many foreign molecules and a pair of minor grooves, which are too narrow to trap anything (see hydration section).

Table 1.10: Phosphate-Phosphate Distances

(a) Tetraplex A

P atoms	P2	P3	P4	P6	P7	P8	P10	P11	P12	P14	P15	P16
P2	–	6.22	12.98	16.27	17.82	20.07	21.02	16.12	14.51	19.54	14.34	10.24
P3	–	–	6.92	18.25	17.49	18.26	18.36	15.39	15.87	14.43	9.81	7.79
P4	–	–	–	20.78	17.89	17.16	16.86	16.45	18.56	8.59	6.07	8.75
P6	–	–	–	–	6.98	13.13	18.95	13.15	7.86	20.72	17.69	16.12
P7	–	–	–	–	–	6.48	13.18	9.27	7.40	16.00	14.91	15.85
P8	–	–	–	–	–	–	7.31	6.88	9.68	14.58	15.61	18.35
P10	–	–	–	–	–	–	–	7.04	12.99	16.00	17.82	21.04
P11	–	–	–	–	–	–	–	–	6.17	17.39	16.78	18.01
P12	–	–	–	–	–	–	–	–	–	19.83	17.65	17.12
P14	–	–	–	–	–	–	–	–	–	–	6.17	12.20
P15	–	–	–	–	–	–	–	–	–	–	–	6.04
P16	–	–	–	–	–	–	–	–	–	–	–	–

(b) Tetraplex B

P atoms	P22	P23	P24	P26	P27	P28	P30	P31	P32	P34	P35	P36
P22	–	6.38	13.28	16.79	17.90	20.55	21.78	17.60	15.05	19.74	14.51	10.37
P23	–	–	7.04	18.36	17.41	18.41	18.71	16.34	16.12	14.39	9.78	7.86
P24	–	–	–	20.63	17.89	17.13	16.89	16.80	18.67	8.37	6.02	8.92
P26	–	–	–	–	6.74	13.20	19.49	13.32	7.85	20.43	17.45	15.93
P27	–	–	–	–	–	6.49	13.23	7.96	6.24	16.52	15.36	16.15
P28	–	–	–	–	–	–	7.48	5.54	9.69	14.85	15.89	18.60
P30	–	–	–	–	–	–	–	6.91	13.71	16.21	18.15	21.45
P31	–	–	–	–	–	–	–	–	6.82	17.25	17.00	18.54
P32	–	–	–	–	–	–	–	–	–	19.91	17.71	17.15
P34	–	–	–	–	–	–	–	–	–	–	6.17	12.19
P35	–	–	–	–	–	–	–	–	–	–	–	6.03
P36	–	–	–	–	–	–	–	–	–	–	–	–

1.3.6 Backbone Conformation

Unusually small twist(12.5°) and large rise(6.26Å) render a quite extended sugar-phosphate backbone (Figure 1-7,Table 1.11). The high standard deviation for most of the backbone torsion angles also reflect highly flexible backbone conformation. However, the glycosidic angle χ which defines the orientation of the base relative to the sugar varies only a little in the range of 229.6° to 250.6°, with an average of 236.0° and a standard deviation of only 4.8°. Thus all the bases are in anti orientation. The variability in sugar puckering is clearly seen: among 32 sugar rings in the crystal asymmetric unit, 15 of them have C4'-exo conformation, 5 C3'-endo, 5 C1'-exo, 4 C2'-endo and 3 O4'-endo. A-DNA and B-DNA normally adopt C3'-endo and C2'-endo puckers, respectively, while Z-DNA has C2'-endo at cytosine and C3'-endo at guanines. The intercalated DNA quadruplex (I-DNA) reported here certainly prefers a different puckering mode from those of A-, B- and Z-DNA, since only a minority of sugar rings in I-DNA have C2'-endo and C3'-endo puckers. The four C2'-endo puckers are found at the 5'-end which suggests that this puckering mode may be associated with the end effect. The fact that about half of the 32 rings adopt C4'-exo pucker implies that I-DNA prefers this puckering mode.

1.3.7 Hydration

Another interesting feature of the I-DNA structure is its mode of interaction with water molecules (Figure 1-8,Table 1.12). At current 1.8 Å resolution, 46 water molecules were identified. Tetraplex A is surrounded by 38 water molecules (Figure 1-8) and tetraplex B by 34. Many of the water molecules are shared between tetraplexes A and B (see below). The two wide grooves are coated with a layer of water molecules, most of which form hydrogen bonds to the hydrogen of amino group N4 which is not included in the $C{\cdot}C^{+}$ pairing. This kind of hydrogen bonding is consistent throughout both tetraplexes, 13 out of 16 N4 atoms in tetraplex A and 11 out of 16 in tetraplex B participate in interaction(Table 1.12). Most of the 8 N4 atoms not involved in such hydrogen bonding are from end bases(C8, C9, C13, C21, C25 and C29) and thus may be attributed to the end effect. In fact, residues C9, C13 and C29 are involved in inter-tetraplex base stacking(see crystal packing section). The N4 atoms from the two non-end bases C30, C31 participate in the wide groove-wide grove interaction between tetraplexes A and B(also see crystal packing below). In addition to the above water-base interaction in the wide grooves, backbone phosphate oxygens also attract a lot of water molecules through $O \cdots H - O$ hydrogen bonding. It is interesting to note that most of the water-phosphate interaction occur at the wide groove side of the backbones, some outside the narrow grooves, but none inside the narrow grooves which are too small for water molecules to fit in (see wide and narrow grooves section). It should also be pointed out that many of water molecules described above are participating in inter-tetraplex interaction, i.e. they are bridging between tetraplexes A and B or their symmetry-related ones. They are doing so by forming a hydrogen bond with base amino N4 at one end and forming another one with backbone phosphate oxygen O1P or O2P at the other end. This type of water bridging may play an important role in stabilizing the I-DNA structure in crystal lattice and/or in solution. 19 out of 46 water molecules in the crystal structure have been identified as such bridging agents.

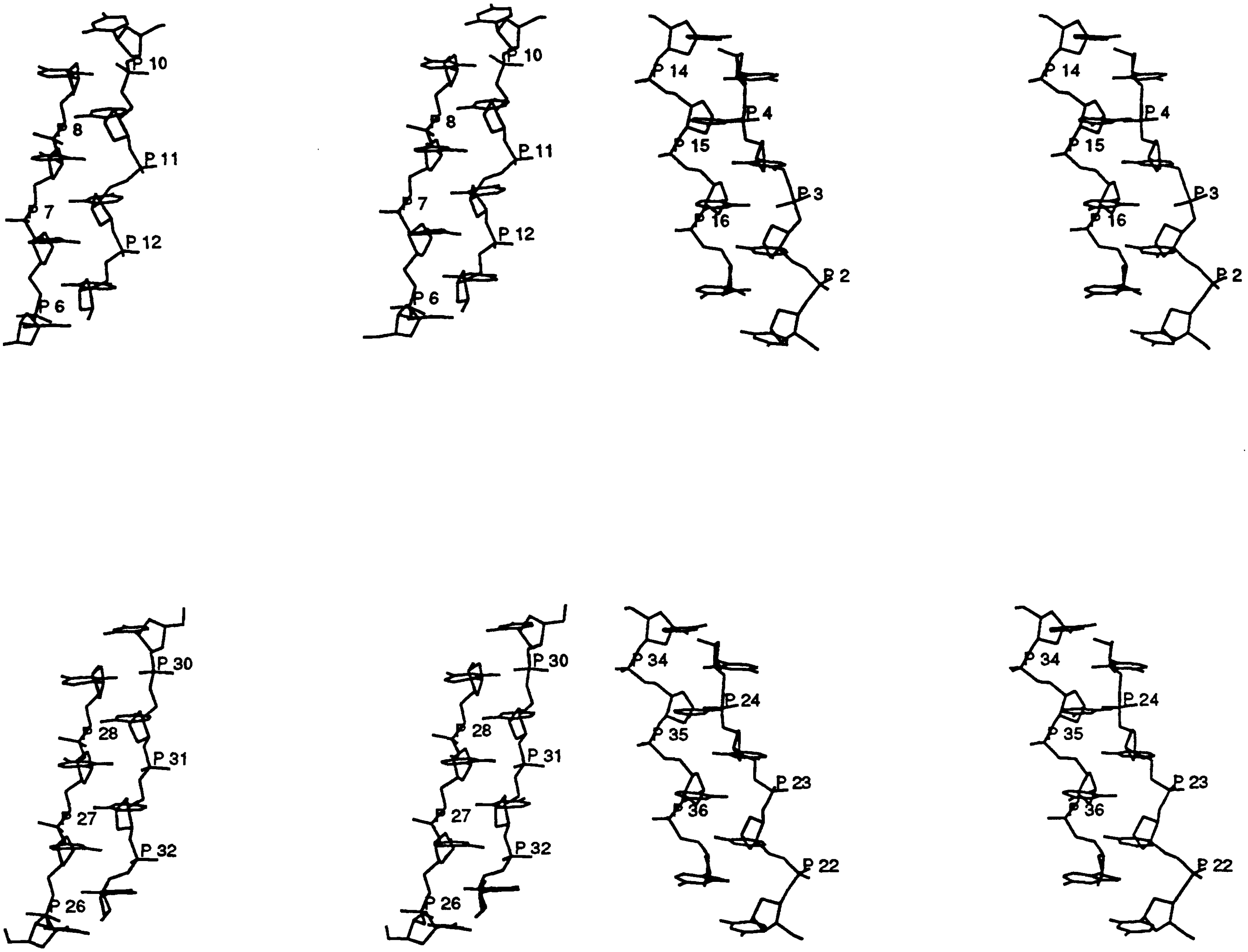

Figure 1-7: Stereo View of Closely Packed Strands.

Table 1.11: Glycosidic Torsion Angles and Sugar Puckers

Strand	**Residue**	α	β	γ	δ	ϵ	ζ	χ	**Pseud**	**Pucker**
a1	C1	n.a.	n.a.	59.5	80.1	205.2	273.3	233.1	15.3	C3'-endo
	C2	305.2	173.3	54.1	61.7	82.3	77.3	238.5	34.6	C3'-endo
	C3	181.3	117.5	159.4	128.4	215.8	291.1	235.2	134.1	C1'-exo
	C4	171.6	153.4	179.5	105.3	n.a.	n.a.	232.4	109.9	O4'-endo
a2	C5	n.a.	n.a.	291.2	157.9	219.3	286.4	244.0	184.3	C2'-endo
	C6	109.9	182.3	186.3	77.9	189.2	289.8	243.4	58.3	C4'-exo
	C7	191.4	160.2	167.8	95.4	205.1	278.8	233.3	68.6	C3'-exo
	C8	157.5	176.9	174.8	87.3	n.a.	n.a.	231.0	48.5	C4'-exo
a3	C9	n.a.	n.a.	65.3	152.0	261.8	246.7	238.3	168.2	C2'-endo
	C10	89.0	211.8	184.1	74.7	189.4	277.8	238.3	57.0	C4'-exo
	C11	289.0	184.6	58.2	74.0	174.7	280.1	247.0	45.9	C4'-exo
	C12	171.5	183.1	177.6	87.4	n.a.	n.a.	232.6	37.4	C4'-exo
a4	C13	n.a.	n.a.	59.2	77.6	190.6	278.4	232.5	17.3	C3'-endo
	C14	293.6	176.5	66.1	80.0	197.3	280.4	233.6	27.5	C3'-endo
	C15	291.7	181.7	65.4	80.8	191.7	280.3	236.0	41.6	C3'-endo
	C16	305.2	173.9	80.8	129.7	n.a.	n.a.	250.6	129.6	C1'-exo
b1	C21	n.a.	n.a.	60.5	82.4	203.9	272.9	237.6	38.9	C3'-endo
	C22	305.0	170.0	64.5	82.2	191.1	269.1	235.1	58.7	C4'-exo
	C23	178.9	195.9	181.0	114.9	211.8	293.0	232.4	114.4	C1'-exo
	C24	171.9	145.8	177.2	106.3	n.a.	n.a.	233.6	101.0	O4'-endo
b2	C25	n.a.	n.a.	60.0	149.4	223.0	289.2	240.7	178.8	C2'-endo
	C26	152.5	164.3	173.4	86.0	203.5	271.7	235.9	80.2	O4'-endo
	C27	155.2	176.3	181.9	91.7	201.7	278.9	231.7	62.2	C4'-exo
	C28	163.5	178.5	169.8	87.1	n.a.	n.a.	231.6	40.1	C4'-exo
b3	C29	n.a.	n.a.	148.7	158.1	211.6	296.3	236.3	181.0	C2'-endo
	C30	160.3	152.0	166.4	86.1	190.1	292.3	234.4	72.4	O4'-endo
	C31	170.5	165.1	171.4	75.5	198.3	293.1	232.0	42.8	C4'-exo
	C32	280.8	176.5	74.9	76.7	n.a.	n.a.	229.6	69.3	C4'-exo
b4	C33	n.a.	n.a.	60.1	81.7	202.7	268.5	233.5	37.1	C3'-endo
	C34	295.9	173.9	64.4	74.9	189.7	284.6	233.0	34.2	C4'-exo
	C35	298.5	179.2	58.0	81.3	196.0	275.3	236.4	47.0	C3'-endo
	C36	301.1	163.5	92.2	110.5	n.a.	n.a.	237.1	115.2	O4'-endo

Backbone torsion angles for the bonds in the backbone P-O5'-C5'-C4'-C3'-O3'-P are α, β, γ, δ, ϵ and ζ, respectively, and the glycosidic angle is χ. (a) is calculated by the program NEWHEL93.

Table 1.12: Hydrogen Bonds Between d(C_4) and Water Molecules

DNA atom	Water molecule	distance (Å)	DNA atom	Water molecule	distance (Å)
1N4	74O*	3.37	2N4	41O	2.93
3O2P	80O	2.91	3N4	68O	2.88
4O1P	61O	2.69	4O5'	61O	3.34
4N4	56O	2.92	4O3'	79O	3.03
5O5'	79O	3.24	5N4	76O	3.05
6O1P	66O	2.72	6N4	42O	2.82
7N4	49O	2.87	10O1P	59O	3.19
10O1P	70O	3.39	10N4	65O	2.91
11O1P	50O	2.65	11O1P	69O	3.02
11O2P	54O	2.72	11O5'	69O	3.39
11N4	63O	2.85	12O2P	55O	2.89
12N4	77O	2.88	13O4'	79O*	3.27
14O1P	48O	2.83	14O1P	78O	3.28
14O1P	75O	3.31	14O2P	58O	2.68
14O2P	46O	3.15	14N4	53O	2.82
15O1P	45O	2.73	15O1P	46O	3.30
15O2P	43O	3.13	15N4	67O*	2.84
16O1P	44O	2.77	16O1P	47O	2.94
16O1P	43O	3.19	16O2P	71O*	3.27
16N4	52O	2.83	22O1P	60O	2.74
22O1P	56O	2.93	22O2P	73O	2.73
22O1P	65O	3.27	22N4	70O	2.99
22O3'	72O	3.07	23O1P	69O	2.57
23O1P	68O	2.82	23O2P	63O	3.27
23O5'	72O	2.89	23N4	54O	2.88
24N4	55O	2.82	25O2	83O	3.01
26N4	47O*	3.04	27N4	45O*	2.87
27N4	45O*	2.87	28N4	48O*	2.83
30O2P	82O	3.18	32N4	59O	3.00
33N4	75O*	3.09	34O1P	49O*	2.88
34O1P	53O*	2.92	34O1P	62O*	2.97
34O2P	86O*	2.96	34O2P	67O	3.34
34N4	46O*	2.75	35O1P	85O*	2.86
35O1P	42O*	2.93	35O1P	67O	3.03
35O2P	52O*	3.26	35N4	43O*	2.82
36O1P	51O*	2.66	36O1P	76O*	3.02
36O1P	52O*	3.23	36N4	71O	2.90

1.3.8 Crystal Packing

The analysis of a DNA crystal structure is not complete without discussing its crystal packing, since conformation of DNA, especially in the end residues, is greatly influenced by packing environment in the crystal lattice. The crystal has very high symmetry(cubic space group I23) with 24 asymmetric units in the unit cell, each containing two tetraplexes (A and B). Within the asymmetric unit, tetraplex A and tetraplex B contact laterally with each other through their wide grooves(Figure 1-9), and there is a pseudo 2-fold symmetry axis between them. The two molecules are not packed parallel to each other, but with a tilt of about 30^o between their helical axes. There are only two direct hydrogen bonds between molecules A and B: N4 of C30 in B is hydrogen bonded to O2P of C12 in A (distance $3.05\mathring{A}$) and N4 of C31 in B forms bifurcated hydrogen bonds to two phosphate oxygenr O1P and O2P of C11 in A(distances $3.09\mathring{A}$ and $3.24\mathring{A}$). The interaction between A and B is mainly through bridging water molecules (see hydration section above). Similar wide groove to wide groove packing is observed between two tetraplexes, each from different asymmetric units, but there are no direct hydrogen bonds and the tilt is about 35^o instead of 30^o. In addition to lateral wide groove packing, base stacking is also seen for both molecules A and B. Molecule A has base stacking at both ends with its symmetry-related ones. It involves base C5 of one molecule(A) and base C9 of another molecule(A'), which stack directly on each other (Figure 1-10). The two tetraplexes do not stack well but with a tilt of about 25^o between their helical axes. Thus bases C5 and C9 are tilted to accommodate that. This is true at both ends of molecule A. Molecule B is different from molecule A as it has base stacking at one end only, which is around a crystal 2-fold axis(Figure 1-11). The base pair $C29 \cdot C33$ from one molecule B stack on the same pair from another molecule B related by the 2-fold axis. They stack on each other so well that the two molecules could form a single continuous quadruplex if the missing phosphates were added to the 5'-ends of C29 and C33. The other end of tetraplex B is wide open and is about $6.3\mathring{A}$ away from a crystal 3-fold axis which is by symmetry surrounded by three B quadruplexes(Figure 1-12). Thus there is a

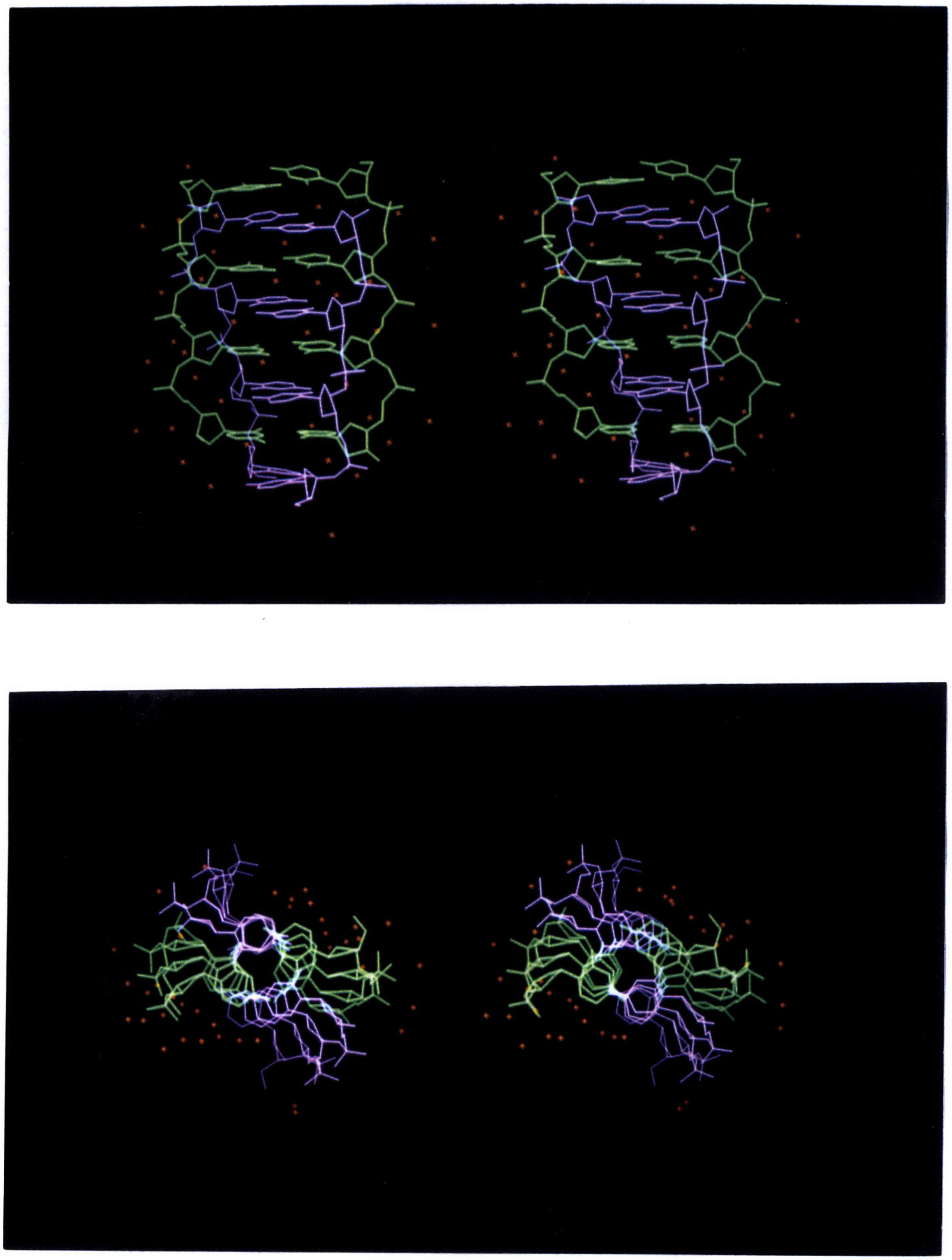

Figure 1-8: Stereo Views of Hydration in Tetraplex A.

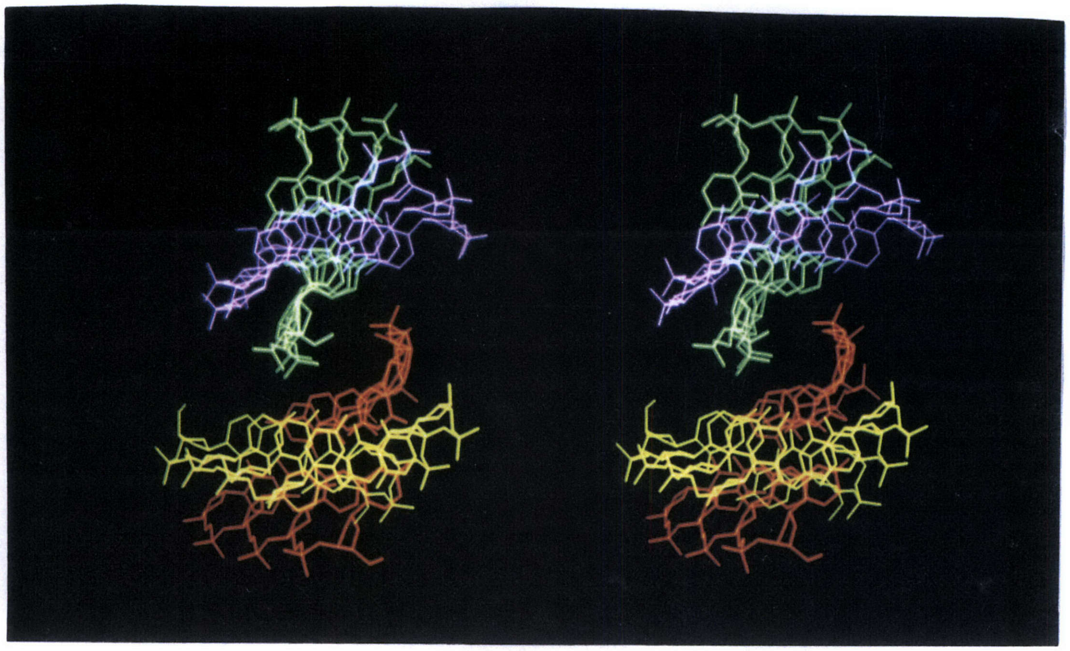

Figure 1-9: Stereo View of Packing Between Tetraplexes A and B.

cavity of about 6.3$\AA$ in diameter there which is big enough for the platinum to be soaked in and sit nicely at the center of the cavity. It is also interesting to point out that next to the cavity there exists a huge empty hole about 44$\AA$ in diameter (data not shown). There are two such big holes per unit cell, which account for about 16% of the unit cell volume.

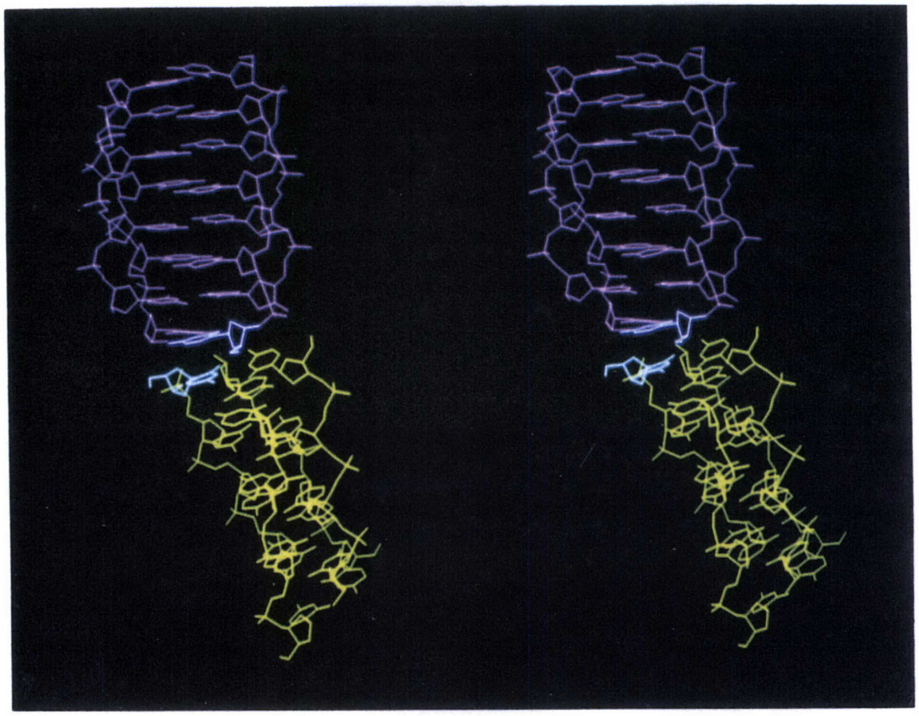

Figure 1-10: Stereo View of Base Stacking Between Tetraplex A and Its Symmetry Related One (A').

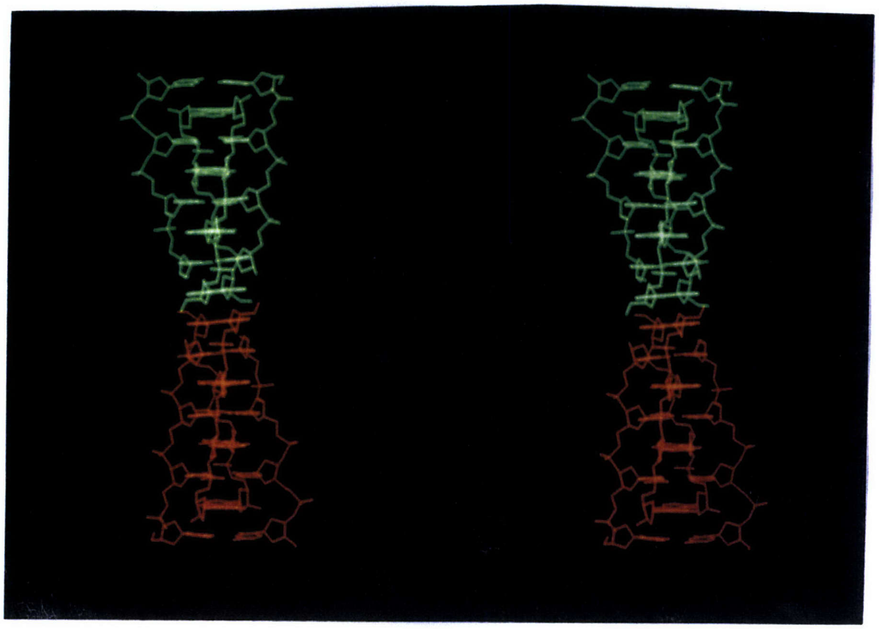

Figure 1-11: Stereo View of Base Stacking Between Tetraplex B and Its Symmetry-related One (B').

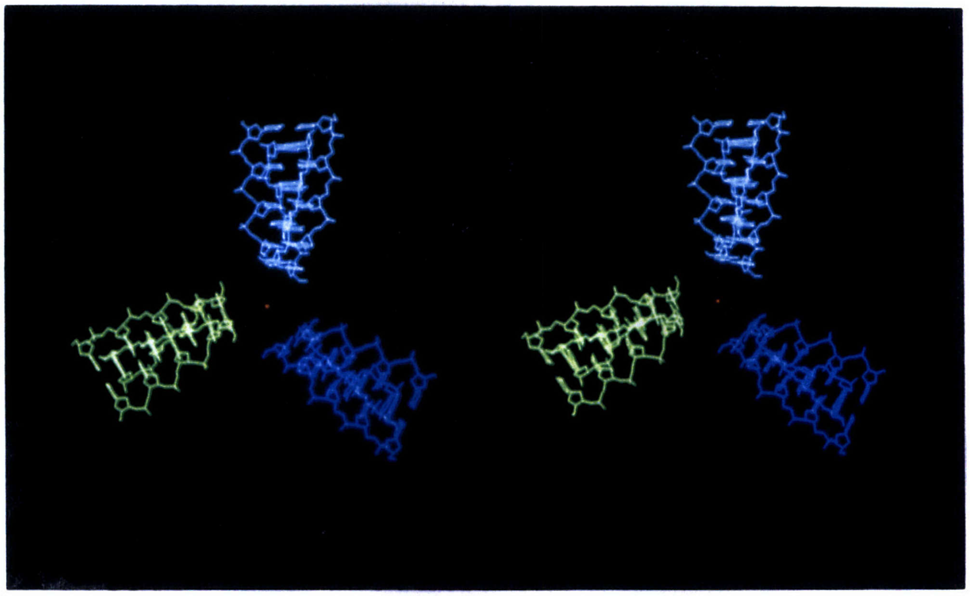

Figure 1-12: Stereo View of Three Tetraplex B's Related by a Crystal 3-fold Axis.

1.4 DISCUSSION

1.4.1 Comparison with the NMR structure of $d(TC_5)$

Comparison of our d(C_4) crystal structure with the d(TC_5) NMR structure has been made previously [7]. This was the first X-ray structure of I-motif DNA which confirmed the overall features proposed by the NMR studies [5]. In addition, the crystal structure revealed many details which could not be derived by NMR experiments. We would like to briefly discuss the inter-residue sugar-sugar H1'-H1' interaction whose unusually strong NMR nuclear Overhauser effects (NOEs) are the fingerprint for such I-motif structures. There are 28 such close H1'-H1' pairs in the d(C_4) crystal structure which have an average distance of 3.13±0.20 $\AA$ (Table 1.13). This is in good agreement with 3.0$\AA$ in the d(TC_5) NMR structure.

1.4.2 Comparison With Other DNA Structures

One of the remarkable features about DNA is its structural polymorphism. DNA can adopt a variety of conformations depending on internally its sequences (base composition), and externally pH, counter ions, salt concentration, drugs, binding proteins, etc. We have already known A- and B-DNA which are the right-handed antiparallel double helix; Z-DNA, a left-handed antiparallel duplex; triple helices (not well known), where a third strand binds in the major groove of a standard duplex; and guanine quadruple helix which involves four guanine bases in a sequential planar arrangement. Now that we have another well-characterized type of quadruplex, I-DNA, it would be very interesting to compare it with other DNA structures. The unique features of I-DNA structure are the parallel $C{\cdot}C^+$ base pairing, antiparallel intercalation of two parallel duplexes, small right-handed intra-strand twist and closely packed backbones. It is those features that distinguish it from other DNA structures and make it a very interesting DNA structural motif which deserve a further and more detailed analysis.

1.4.3 Biological implications

Tracts of cytosine-rich DNA have been found in telomeres, which cap the ends of linear embaryotic chromosomes and consist of simple tandem repeats that are G-rich on one strand and C-rich on the other [15, 16]. We now know that both the G-rich and C-rich strands can form quadruplexes in vitro, although G-quartet is in a square-planar arrangement [17, 18] and C-quartet is in an intercalation mode [5, 7]. Do these structures occur in vivo? The fact that C and G are complementary in a DNA double helix suggests that they may have some biological implications. Any sequence, such as telomeric DNA, that forms a C-quadruplex on one strand has the ability to form a G-quadruplex on the other strand. The two structures could act in concert and/or one could promote the formation of the other (Figure 1-13). The self pairing nature of both structures implies that they may play an important role in DNA self recognition, which is essentıal in many biological activities, such as meotic chromosome pairing and recombination.

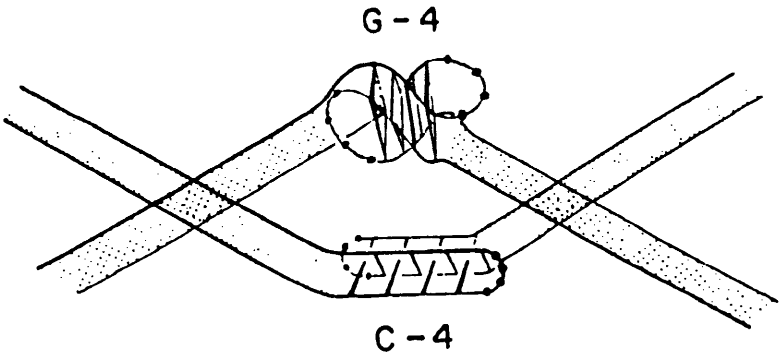

Figure 1-13: Model That C-tetraplex and G-quadruplex Could Interact.

Table 1.13: Inter-residue Sugar-sugar H1'-H1' Distances

Tetraplex A	**H1’-H1’ distance(Å)**	**Tetraplex B**	**H1’-H1’ distances(Å)**
C1-C16	2.92	C21-C36	3.29
C16-C2	3.40	C36-C22	2.95
C2-C15	3.11	C22-C35	3.33
C15-C3	3.01	C35-C23	3.02
C3-C14	3.36	C23-C34	3.29
C14-C4	2.95	C34-C24	3.24
C4-C13	3.07	C24-C33	3.07
C5-C12	3.04	C25-C32	3.24
C12-C6	2.96	C32-C26	3.07
C6-C11	2.78	C26-C31	3.62
C11-C7	3.24	C31-C27	2.84
C7-C10	3.29	C27-C30	3.38
C10-C8	3.13	C30-C28	3.02
C8-C9	2.77	C28-C29	3.31
Average (A and B)	3.13Å		
Standard deviation	0.20Å		

Figure Legends of Chapter 1

Figure 1-1. Two sections of the isomorphous difference Patterson and anomalous difference Patterson maps for the crystal of d(C_4). (Top) Isomorphous and anomalous difference Patterson maps (from left to right) at Z=0 Harker section with data from 20.0$\AA$ to 3.0$\AA$. Two maps clearly show the peak at the same position (0.398, 0.398). (Bottom) Isomorphous and anomalous difference Patterson maps (from left to right) at Z=0.1099 with data from 20.0$\AA$ to 3.0$\AA$. We were able to determine the Pt position from the these maps, and the position was further confirmed by HASSP.

Figure 1-2. Superposition of the initial (SIR+SAS) density map after solvent flattenirg with the final refined structural model. The initial electron density map clearly shows the intercalated structure. The phosphates are identifiable as intense peaks and the $C{\cdot}C^+$ base pairing within each layer are clearly visible. The map is contoured at 1σ density level.

Figure 1-3. Superposition of the final 2Fo-Fc map after the refinement with the final refined structural model. The model fits in the map nicely, and shows a very stable intercalated structure. Compared with Figure 1-2, the final map is clearer, even though the model fits both maps very well. This map is contoured at 1σ density level.

Figure 1-4. Overview of the d(C_4) quadruplex. (a) A stereo sideview of tetraplex A mainly through its wide groove. The molecule is made of two parallel-stranded duplexes, one is colored purple (Cl-C8) and the other green (C9-C16). Each strand is labelled at its 5' end (atom O5'). All the atoms including hydrogens are shown. Strand Cl-C4 is paired to strand C5-C8, running upward (5'-3'). Strand C9-C12 is parallel to C13-C16, but running downward. Thus the two duplexes intercalate into each other antiparallelly. (b) to (d) Space filling models of molecule A with different

views: (b) Sideview from the back of (a) showing the wide groove; (c) Sideview from the left of (a) showing the narrow groove; (d) End view from the top of (a). Color codes: red-oxygen, yellow-phosphorous, blue-nitrogen, green-carbon and white-hydrogen.

Figure 1-5. Superposition of tetraplexes A and B in stereo. Molecule A is colored purple, molecule B green. Only 5'-ends of moleculeA is labelled. Superimposing is such that C1 of A is on C21 of B, C2 on C22, and so on. Hydrogen atoms are not shown for clarity. (a) Sideview to the wide groove. (b) Sideview to the narrow groove. (c) Endview from the top.

Figure 1-6. Base-base stacking modes. (a) to (g) Endviews of seven two-layer base-base stacking for the eight-layer molecule A. The two $C{\cdot}C^+$ base pairs stack on each other with an average stacking distance of 3.13$\AA$. Only exocyclic atoms O2 and N4 are significantly involved in stacking. (h) A space filling model of (e). Color codes are the same as those in Figure 1-4.

Figure 1-7. Stereo views of closely packed strands. Phosphorous atoms are labelled. (a) and (b) are from molecule A, and (c) and (d) are from molecule B.

Figure 1-8. Stereo views of hydration in tetraplex A. The color codes are the same as those in Figure 1-4 for DNA, and the read crosses represent water molecules. (a) Sideview, (b) Endview.

Figure 1-9. Stereo view of packing between A and B. The view is down the pseudo 2-fold axis relating A (top) and B (bottom). Color codes are purple for C1-C8, green for C9-C16, red for C21-C28, and yellow C29-C36.

Figure 1-10. Stereo view of base stacking between tetraplex A (purple) and its symmetry-related one (A', yellow). The bases C5 (light blue) of A stacks on the base C9 (light blue) of A'. A similar stacking occurs at the top of A (not shown).

Figure 1-11. Stereo view of base stacking between tetraplex B (green) and its symmetry-related one (B', red). A crystal 2-fold axis is between B and B'.

Figure 1-12. A stereo view of three tetraplex B's related by a crystal 3-fold axis. The red dot represents the platinum atom in the derivative crystal.

Figure 1-13. A possible model that a cytosine-tetraplex and a guanine-quartet could interact in DNA recognition.

Chapter 2

Crystal Structure of d(AACCC): A Three Demensional Network Formed By Two Cytosine-tetraplexes and Two Adenine-clusters

2.1 INTRODUCTION

Telomere DNA at chromosome ends determines the stability of the chromosome and thus is important in in both the normal control of cell proliferation and the abnormal growth of cancer [19, 20, 21]. The newer roles being assigned to telomeres include aiding in gene regulation and possibly serving as a "mitotic clock" for the cells of higher animals. The first telomere was isolated from ciliate *Tetrahymena thermophila* in the early 1970s [15]. Its G-rich strand consists of a short repeating sequences of d(GGGGTT). and the complementary C-rich strand consists of short repeats of d(AACCCC). Confirmation of sequence conservation of telomeres came in 1988, when Robert Moysis and his colleagues at Los Alamos National Laboratory isolated the first human telomeres and show that they also consists of repeating sequences of d(TAACCC) and its complement [22]. Its evolutionary conservation among vertebrate species was later demonstrated [23, 24]. Earlier we have solved the first crystal structure of C-rich DNA segment d(C_4), revealing a novel four-stranded intercalated motif [7]. In an attempt to elucidate the three-dimensional structure of the telomere sequences, given its biological and structural significance, we crystallized and solved d(AACCC), the common sequence of both *Tetrahymena* and human telomeres, to 2.0$\AA$ resolution. It confirms the novel cytosine intercalated motif, shows a great deal of structural variation of the motif, and in the mean time, reveals two kinds of novel adenine clusters that are used to build the three demensional crystal lattice by elaborate interactions of symmetry related adenine residues.

2.2 MATERIAL AND METHODS

2.2.1 DNA Synthesis and Crystallization

The oligodeoxyribonucleotide d(AACCC) was synthsized on a 10μM scale on an Applied Biosystems DNA synthesizer using solid phase β- cyanoethylphosphoramidite chemistry. It was then purified by reverse-phase high performance liquid chromatography (HPLC) on a C4 column(Rainin Instrument Co.), with a linear gradient of 5-40% acetonitrile in 0.1 M triethylammonium acetate buffer pH 7.0. The peak eluent containg the pure d(AACCC) was collected and lyophilized overnight.

Crystals of d(AACCC) were grown at room temperature using the sitting drop vapor diffusion method. The best crystals of diffraction quality were grown from a solution containing 200mM of Na-Cacodylate buffer of PH 6.0 and 3mM of DNA sample (single strand concentration). It crystallized in space group $P22_12_1$, with unit cell dimensions of a=29.7$\AA$, b=52.4$\AA$, and c=64.5$\AA$. The crystal contains eight strands per asymmetric unit. The crystal data of the sample is summarized in Table 2.1.

Table 2.1: Crystal Data of d(AACCC).

Space group	$P22_12_1$
a	29.7$\AA$
b	52.4$\AA$
c	64.5$\AA$
strands per unit cell	32
strands per asymmetric unit	8

2.2.2 Data Collection and Heavy Metal Soaking

All diffraction data were collected on a Rigaku Raxis II imaging plate system at room temperature and processed with the PROCESS program provided by the Molecular Structure Corporation. The native data used in phasing were collected to 2.0$\AA$ resolution, 35 frames at a crystal-to-plate distance of 85 mm with 3.0^o oscillations. This data set was used to carry out the refinement.

The presence of eight strands in an asymmetric unit made building initial models exceedingly difficult, and we again opted for the heavy atom derivatives soaking method. After an extensive search, we found soaking with $HgCl_2$ showed promising results. Systematic soaking was later carried out by changing $HgCl_2$ concentrations and/or soaking time in order to optimize the heavy atom signal. The best derivative was obtained by soaking the crystals in a mother liquor with 10mM $HgCl_2$ for 20 hours. The derivative data were collected to a resolution of 2.0$\AA$ in 50 frames, with oscillation of 3.0^o and a crystal-to-plate distance of 85mm. Table 2.2 contains a summary of the data collection statistics.

2.2.3 Structure Determination by Only One Heavy Atom Derivative

Determination of Mercury Position

The crystal used in phasing was soaked in 10mM $HgCl_2$ for 20 hours, and produced good isomorphous difference as well as anomalous scattering signals. When compared to the native data set, the heavy atom data set revealed an R factor of 15.99% at 2.5 $\AA$ resolution. The position of Hg atom was readily obtained from the isomorphous difference Patterson map and the anomalous difference map, using the Patterson search program HASSP [8]. Both the isomorphous difference Patterson and the anomalous difference Patterson maps with data to 3.0 $\AA$ resolution showed the same set of strong clear peaks which could be easily interpreted as a single Hg site, as shown in Figure (2-1). The Patterson search by HASSP confirmed this single site (Table 2.3). The largest Bijvoet differences (top 25%) to a resolution of 3.0$\AA$ and isomorphous

Table 2.2: Summary of Data Collection Statistics

	Native	$HgCl_2$
Resolution ($\AA$)	2.0	2.0
Number of observations	61483	66316
Number of unique reflections	6602	6825
Overall completeness (%)	94.0	95.1
Outermost shell ($\AA$)	2.25-2.0	2.25-2.0
Outermost shell completeness (%)	84.0	90.2
$I/\sigma(I) > 1$ (%)	90.9	94.3
$I/\sigma(I) > 1$ outermost shell(%)	83.8	88.0
R_{merge} (%)	8.2	6.5

R_{merge} is the agreement R-factor between symmetry-related observations.

centric data to resolutions of 3.0$\AA$ and 2.5$\AA$ were then used separately to refine the position of the Hg atom by a least-squares program SHELX [9]. SHELX refines the Hg position to x=0.013, y=0.487, and z=0.727. No additional sites were observed upon inspection of a difference Fourier map calculated using SIRSAS phases from the initial Hg site.

Determination of Phases

The iterative single isomorphous replacement and single-wavelength anomalous scattering method (ISIRSAS) [10], in which a noise filtering procedure is applied to resolve the phase ambiguity associated with SIR and SAS data to improve the initial phases, was used to generate the DNA phases. The phasing power of the SIR and SAS data (Table 2.4), are listed below. The ISIR procedure was first applied to the isomorphous data, whereas ISAS procedure was applied to the anomalous data, both to 3.0$\AA$. The figure of merit after the SIR step is 0.28, with 1918 accepted reflection pairs, and the figure of merit after the SAS step is 0.36, with 852 accepted reflection pairs. Subsequently, the SIR and SAS data were merged, with a figure of merit of 0.39 and 1934 paired reflections. Four cycles of iterations were then applied to the combined data to resolve the phase ambiguity to 3.0 Å, followed by six cycles of phase extension to 2.5 Å. The final cycle had an average figure of merit of 0.64, a map inversion R-factor of 0.290, and a correlation coefficient of 0.956 for the "+ + +" handed enatiomer.

The above phasing procedure also allowed the handedness of the coordinates to be determined. We have two enantiomers, "+ + +" and "- - -" handed, the second of which was created by setting the three coordinates of the Hg position to its negative. The statistics for the "+ + +" enatiomer were described as above. For the "- - -" handed enantiomer, the same filtering procedure was performed, and the final figure of merit was 0.64, the map inversion R-factor was 0.302, and the correlation coefficient was 0.952. A comparison of the final phasing statistics between the two enatiomers indicated that the "+ + +" handed enatiomer was correct (Table 2.5). This choice was further verified by the interpretability of the electron density generated by the "+ + +" set of phases. The electron density map generated with "+ + +" set of phases clearly showed the intercalated structure. Figure 2-2 is a portion of the ISIRSAS map showing a part of the intercalated C-tetraplex. The sugar phosphate backbones were clearly visible with intense peaks at the phosphate positions. In addition, individual bases were visible in the map. A model was built into the density map with the program FRODO [11].

Refinement of the Structure

The model generated by the ISIRSAS map was then refined to 2.3 Å. The initial R-factor was 50% for the data from 8 to 2.0 Å. Refinements were then carried out using the program XPLOR [12]. Simulated annealing was used, and the R-factor fell rapidly to 25%. The original map was checked frequently during the refinement process. Twenty cycles of restrained individual isotropic B-factor refinement followed. Well ordered water molecules were then located from the difference Fourier map$(F_o - F_c)$ and added as oxygen atoms to the model only if they had a peak height of over

3 σ in the difference density map and formed hydrogen bonds with DNA atoms or other molecules. A total of 57 water molecules were found this way. A final round of refinement completed the structural determination with an R-factor of 0.199 and r.m.s. deviation from ideal bond lengths and angles of 0.016$\AA$ and 3.5, respectively. A summary of the refinement statistics is listed in Table 2.6.

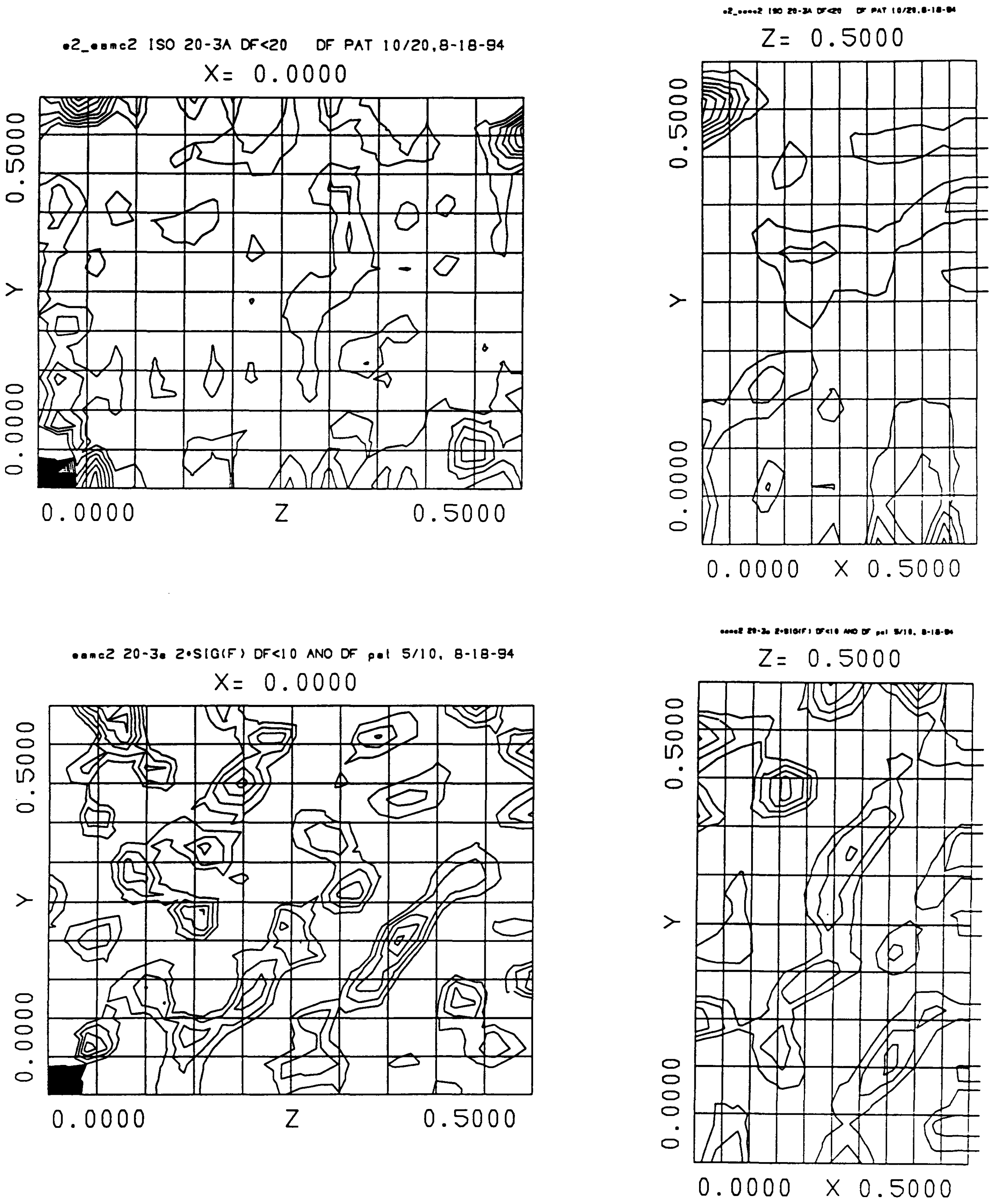

Figure 2-1: Harker Sections of the Isomophous Difference Patterson and Anomolous Difference Patterson Maps for the Crystal of d(AACCC).

Table 2.3: Summary of the Patterson Search from HASSP

List of Major Peaks in Isomorphous Single-Atom Search

Peak	X	Y	Z	Height	Probability by Chance
1	0.016	0.477	0.727	4750.4	0.000
2	0.156	0.047	0.219	1165.0	0.988
3	0.000	0.023	0.094	1152.7	0.923
4	0.188	0.047	0.016	1092.2	0.996
5	0.000	0.227	0.172	1031.4	0.830
6	0.000	0.234	0.250	753.15	0.780
7	0.000	0.219	0.125	576.90	1.000

List of Major Peaks in Anomolous Single-Atom Search

Peak	X	Y	Z	Height	Probability by Chance
1	0.063	0.055	0.063	1642.9	0.540
2	0.016	0.469	0.727	1626.5	0.561
3	0.125	0.117	0.078	1579.9	0.620
4	0.188	0.141	0.219	1530.1	0.682
5	0.031	0.234	0.109	1311.6	0.904
6	0.031	0.234	0.164	1311.6	0.904
7	0.000	0.031	0.023	1166.7	0.537

Table 2.4: Phasing Power of the SIR and SAS Data

SIR Data

Resolution($\AA$)	Phasing Power	Reflections
6.72	2.03	666
3.97	1.75	666
3.31	1.56	666
2.93	1.69	666
2.68	1.69	666
2.49	1.62	666
2.35	1.60	666
2.23	1.55	666
2.13	1.52	666
2.04	1.56	666
2.00	6.46	1
Total	1.70	6661

Phasing Power=$< ||F_{PH}| - |F_P|| > / < |(F_{PH}|_{obs} - |F_P|_{obs}) - (|F_{PH}|_{cal} - |F_P|_{cal})| >$

SAS Data

Resolution($\AA$)	Phasing Power	Reflections
6.74	2.13	311
4.29	2.02	311
3.61	1.91	311
3.22	1.91	311
2.95	1.84	311
2.72	1.75	311
2.55	1.52	311
2.39	1.53	311
2.25	1.47	311
2.10	1.32	311
2.00	1.14	8
Total	1.79	3118

Phasing Power=$< 2F''_H > / < |(F^+_{PH}|_{obs} - |F^-_{PH}|_{obs}) - (|F^+_{PH}|_{cal} - |F^-_{PH}|_{cal})| >$

Table 2.5: Summary of Phasing Statistics

(a) Initial phase preparation

	SIR	SAS	Merged (SIR+SAS)
Resolution limit(Å)	3.0	3.0	3.0
Reflection pairs	1918	852	1934
Completeness (%)	93	72	94
R_{diff} (%)	16.0	5.5	
Figure of merit	0.28	0.36	0.39

(b) Handedness test

	+ hand	- hand
Initial figure of merit	0.39	0.39
After solvent flattening:		
Figure of merit	0.64	0.64
Map inversion R-factor	0.290	0.302
Correlation coefficient	0.956	0.952
Map interpretability	Yes	No

Resolution limit is the highest order of data used in phasing and is not the diffraction limit of the crystal. R_{diff}=$\Sigma \mid F_{native} - F_{derivative} \mid / \Sigma \mid F_{native} \mid$ for SIR data and R_{diff}=$\Sigma \mid F_{+} - F_{-} \mid / \Sigma \mid F \mid$ for SAS data. The coordinates for + hand are X=0.0129, Y=0.4867, Z=0.7271; For - hand, signs of X, Y, Z of above are reversed.

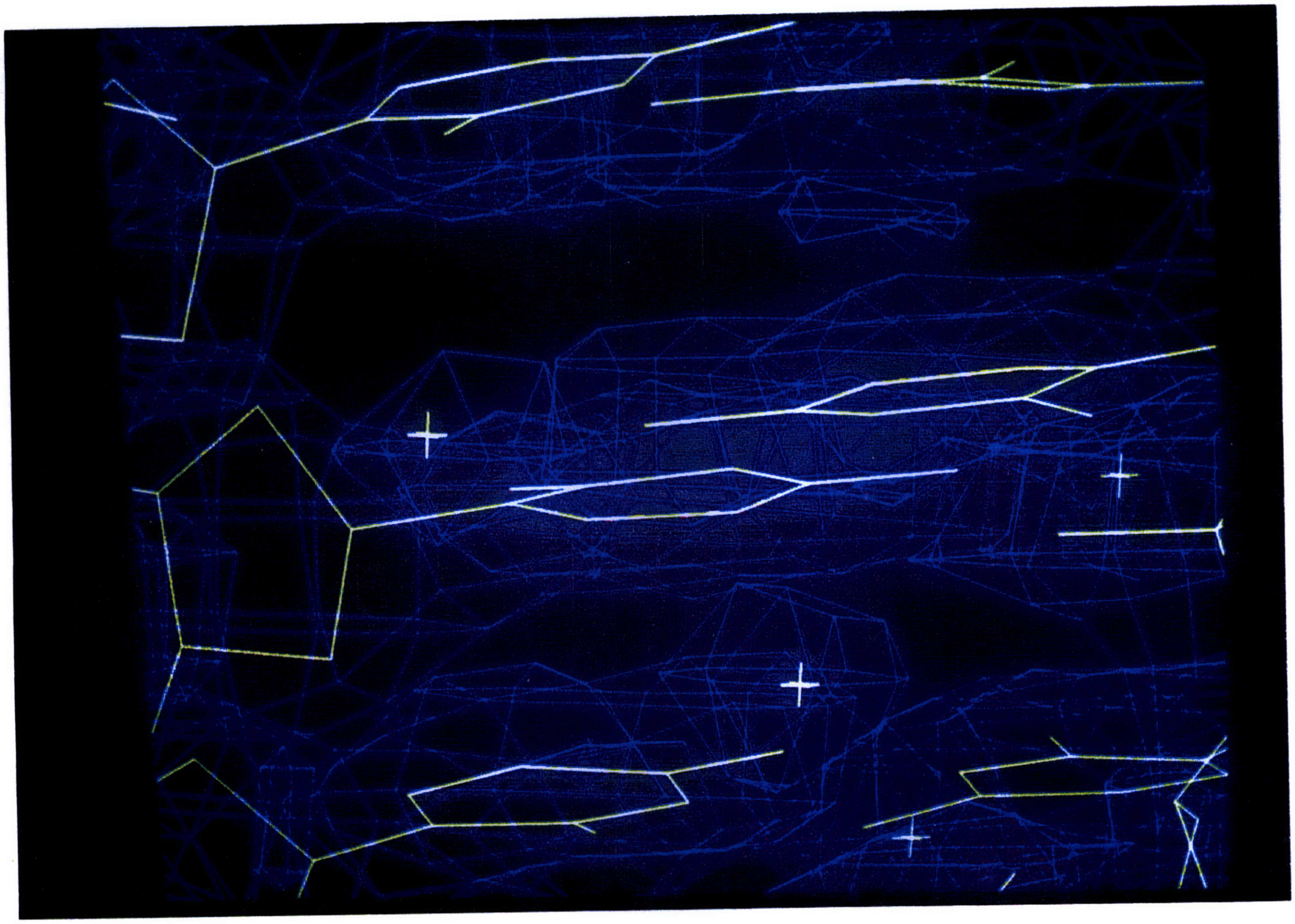

Figure 2-2: A Portion of the Initial (SIR+SAS) Electron Density Map (Blue) After Solvent Flattening. Inside the Density Map Is the Final Refined Model.

Table 2.6: Refinement Statistics

Resolution ($\AA$)	**8-2.0**
Number of Reflections (I>1σ(I)	6602
Completeness (%)	94.0
Number of non-hydrogen DNA atoms	768
Number of water molecules	57
r.m.s. bond length($\AA$)	0.016
r.m.s. bond angles(^{o})	3.5
R-factor	0.199

$$R=\Sigma \mid F_{observed} - F_{calculated} \mid / \Sigma \mid F_{observed} \mid$$

2.3 RESULTS

2.3.1 Overview of the structure

The oligonucleotide d(AACCC) crystallizes in the orthorhormbic space group $P22_12_1$ with eight strands in an asymmetric unit. These eight strands in turn form two cytosine tetraplexes, labelled as tetraplex 1 and tetraplex 2, in Figure 2-3 and Figure 2-4, respectively, along with adenines. At the center of each figure shown are the four cytosine segments of four different chains organized into an intercalation motif which, similar to that of d(C_4), is composed of a four stranded molecule in which two parallel duplexes intercalate with opposite polarity, using *cytosine·cytosine* base pairs. The overall orientations of the two tetraplexes are perpendicular to each other, with C-tetraplex 1 stacking along x-axis and C-tetraplex 2 stacking along z-axis. Similar to d(C_4), d(AACCC) has two cytosine tetraplexes in an asymmetric unit. In contrast to d(C_4), however, the two cytosine tetraplexes in d(AACCC) are connected by two novel adenine clusters in orthogonal directions, whereas the interaction between tetraplex A and B in d(C_4) is through bridging water molecules. This new feature, not only gives us a glimpse of what the telomere structure might be, but also shows us a new folding motif that large nucleic acid molecules might adopt when they are in close contact.

2.3.2 Cytosine Tetraplexes

Two cytosine-tetraplexes, along with two adenine residues at the 5' end of each strand, are shown in Figure 2-3 and Figure 2-4, respectively. Looking at each strand carefully in both figures, we note that the adenine residues are rather loosely connected to the cytosine residues of the strand, which is part of the rigid tetraplex. In each case, the adenine residues point away from the C-tetraplexes, with the orientation of the bases perpendicular to those of the cytosine bases. In Figure 2-3, the $C{\cdot}C^+$ base pairs of C-tetraplex 1 stack along x-axis, with the adenine residues pointing in the z-direction. In Figure 2-4, the $C{\cdot}C^+$ base pairs of C-tetraplex 2 stack along z-axis, with the adenine residues pointing the x-direction.

While the two tetraplexes appear to be rather similar, careful inspection of Figure 2-3 and Figure 2-4 shows that there are striking differences. The relative positions of the adenine residues are different. More importantly, the configurations of the C-tetraplexes differ in a subtle way. In C-tetraplex 2, the outmost layers of the cytosine residues come from the 5' ends of the strands, a configuration similar to that observed in the two tetraplexes in d(C_4). In C-tetraplex 1, however, the outmost layers of the cytosine residues come from the 3' ends of the strands, adding an interesting variation to the intercalating motif.

Comparison With Previously Solved C-rich Structures

The cytosine tetraplexes observed in this structure bear striking resemblence to those observed in d(C_4) and d(TAACCC), the sequence repeat found in human telomere. They also differ in many subtle ways. The stacking of $C{\cdot}C^+$ base pairs is again limited to the exocyclic amino and carbonyl groups, with an average stacking distance of $3.1\AA$. Each tetraplex features two very wide major grooves and two very narrow grooves, like what we observed in d(C_4) (see Figure 2-5 and Figure 2-6). The narrowness of the minor grooves is demonstrated by the short P-P distances across the minor groove, with average distance close to the intra-chain P-P distance of $6.3\AA$. Table 2.7 summarizes the phosphate-phosphate distances of C-tetraplex 1 and C-tetraplex 2. In Figure 2-5 and Figure 2-6, we can see that the narrow grooves of both tetraplexes are so compact that there is no room to trap any foreign molecules, such as water molecules, while in the wide major grooves, there is a lot of room to accomodate water molecules.

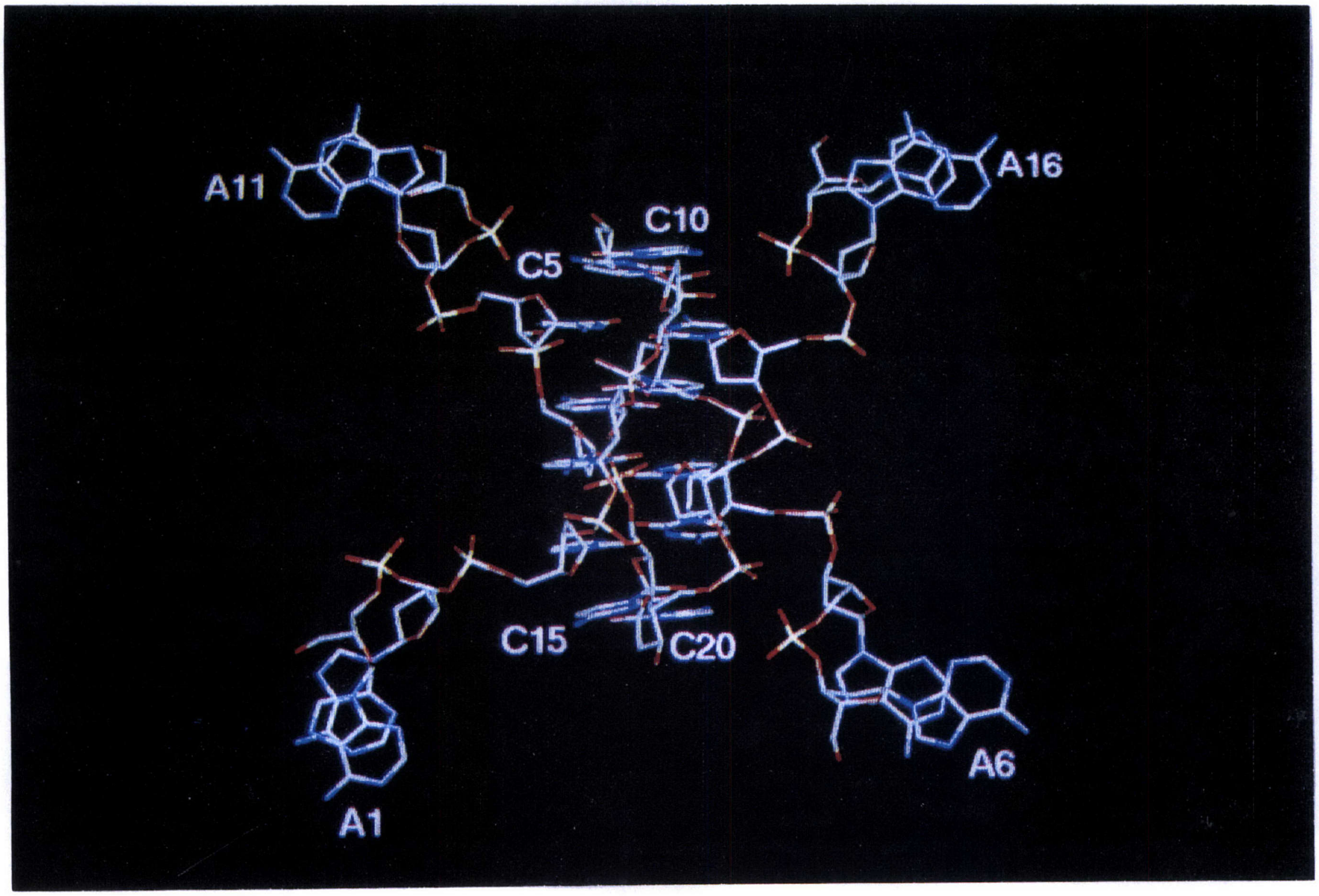

Figure 2-3: Cytosine Tetraplex 1 of d(AACCC).

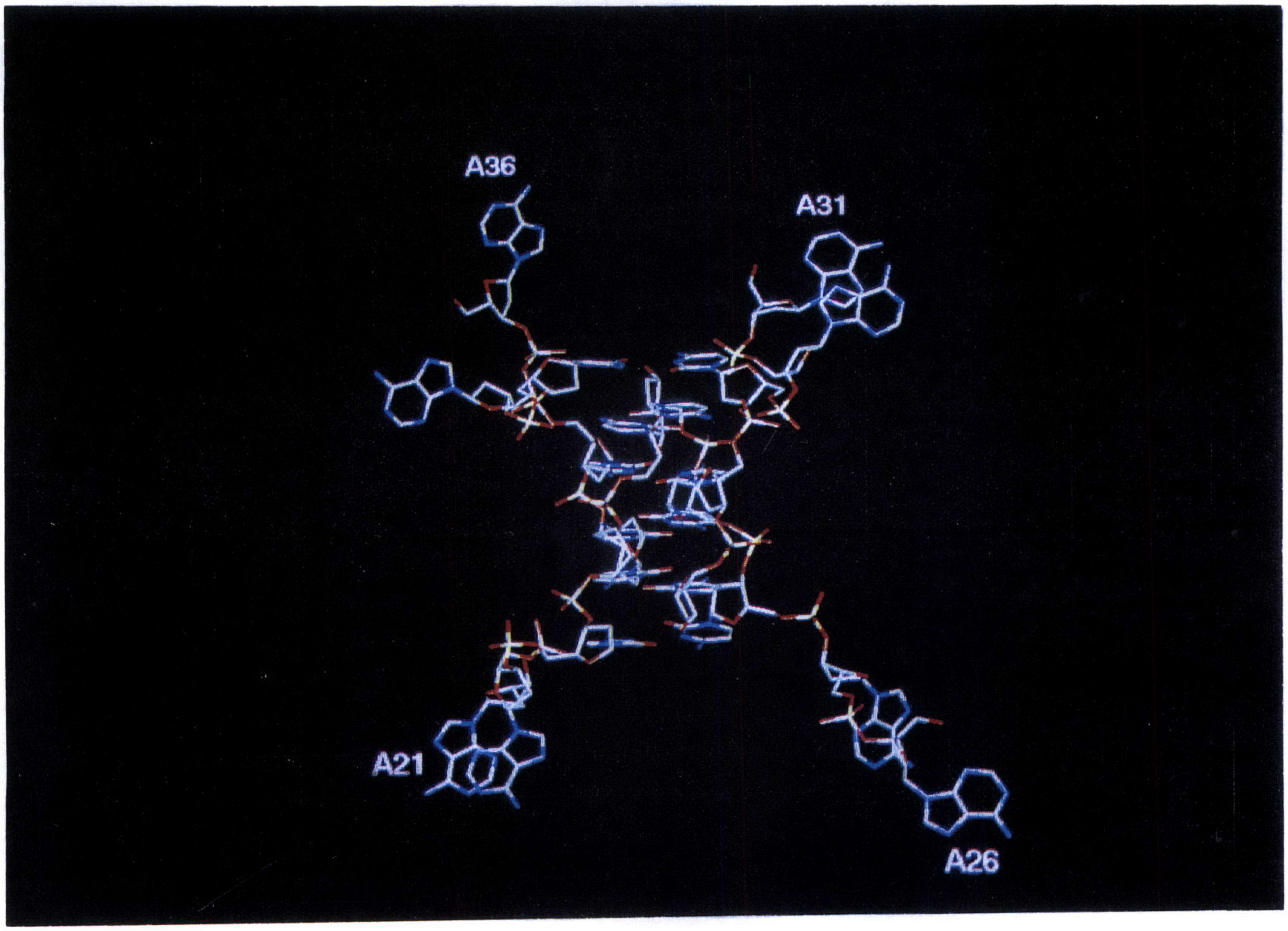

Figure 2-4: Cytosine Tetraplex 2 of d(AACCC).

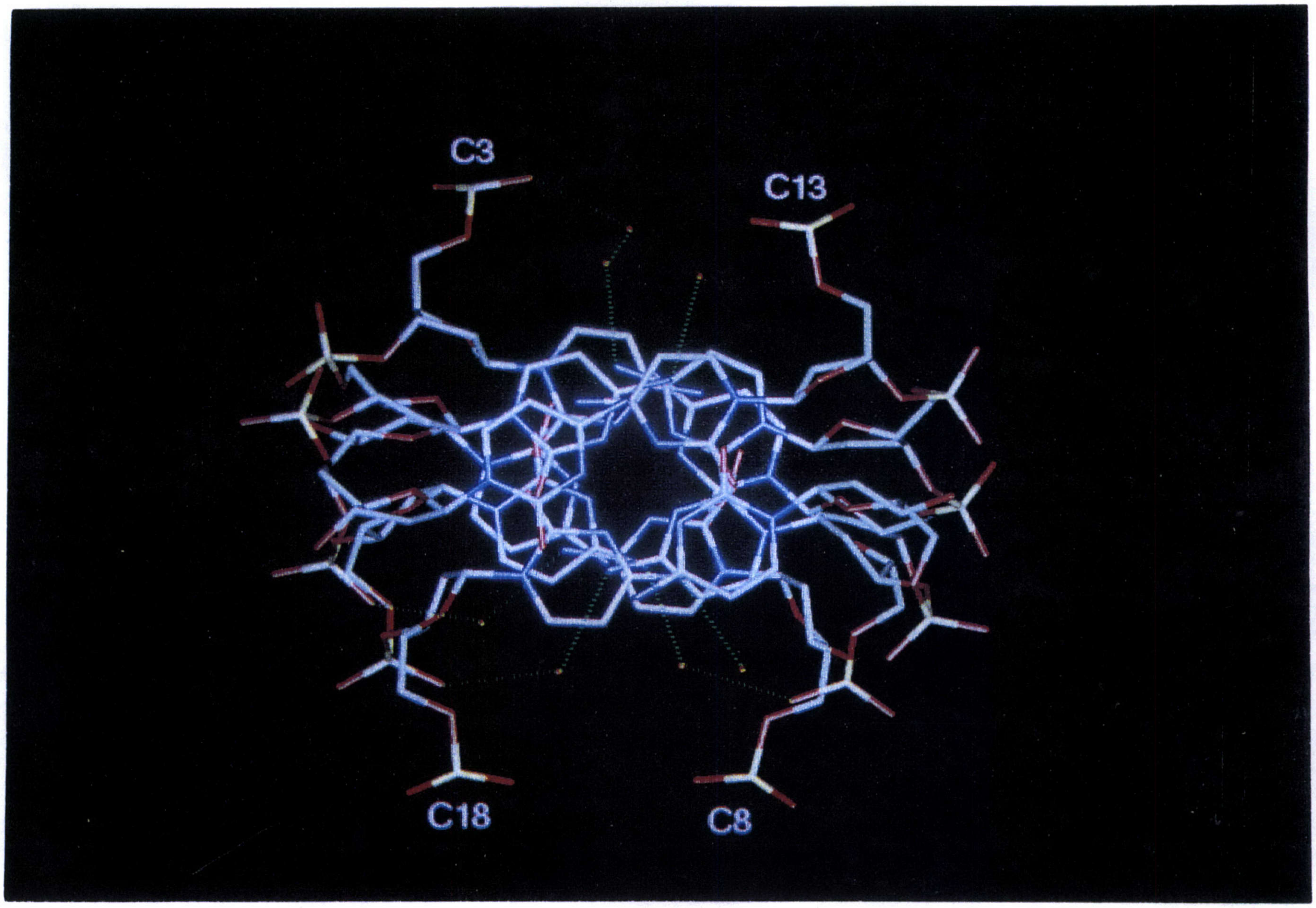

Figure 2-5: View of C-tetraplex 1 From Its Helical Axis.

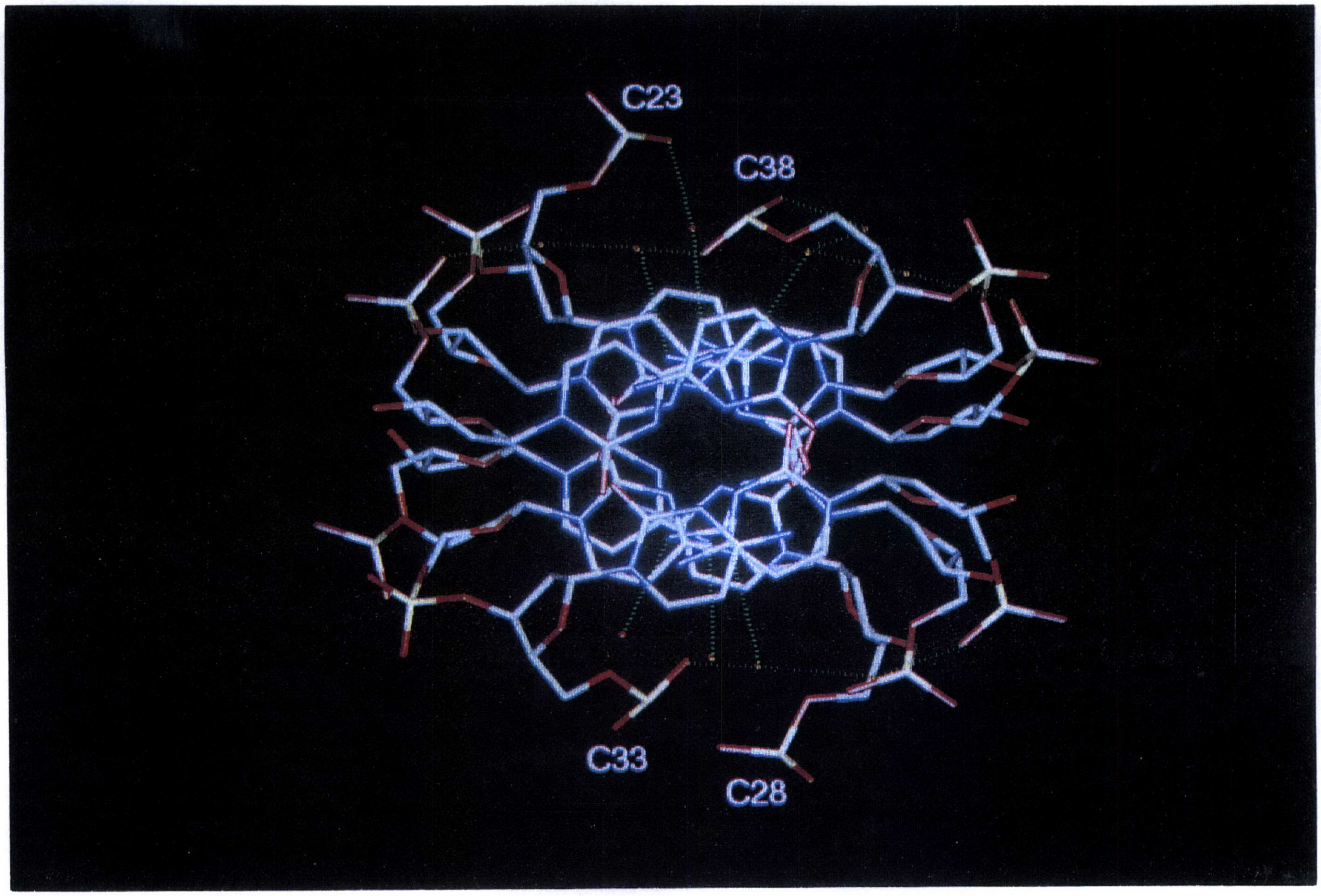

Figure 2-6: View of C-tetraplex 2 From Its Helical Axis.

Table 2.7: Phosphate-Phosphate Distances

(a) Tetramer A

P	P3	P4	P5	P8	P9	P10	P13	P14	P15	P18	P19	P20
P3	–	6.15	11.23	17.62	16.79	18.55	12.67	14.79	15.11	17.96	13.76	10.88
P4	–	–	6.72	15.90	15.82	17.99	14.66	16.31	16.49	13.40	8.76	6.22
P5	–	–	–	17.75	16.41	16.74	13.76	16.37	18.59	10.46	7.99	10.06
P8	–	–	–	–	5.45	11.44	18.40	13.84	9.18	12.00	11.15	12.06
P9	–	–	–	–	–	6.08	14.03	8.96	6.25	11.37	11.59	13.91
P10	–	–	–	–	–	–	11.20	6.45	8.78	12.56	14.18	17.75
P13	–	–	–	–	–	–	–	6.49	12.49	17.30	16.53	18.23
P14	–	–	–	–	–	–	–	–	6.92	16.15	15.90	17.76
P15	–	–	–	–	–	–	–	–	–	16.63	15.51	15.94
P18	–	–	–	–	–	–	–	–	–	–	5.20	11.23
P19	–	–	–	–	–	–	–	–	–	–	–	6.10
P20	–	–	–	–	–	–	–	–	–	–	–	–

(b) Tetramer B

P	P23	P24	P25	P28	P29	P30	P33	P34	P35	P38	P39	P40
P23	–	5.51	11.22	17.15	16.33	17.80	19.24	18.60	13.87	12.55	15.29	13.12
P24	–	–	6.24	16.32	15.85	17.26	15.81	13.85	9.02	11.23	15.48	14.46
P25	–	–	–	18.08	16.90	16.91	12.60	8.93	7.11	9.31	14.75	15.82
P28	–	–	–	–	5.44	11.39	13.24	16.93	13.41	19.36	18.88	14.19
P29	–	–	–	–	–	6.05	10.93	15.77	13.95	15.68	14.08	9.58
P30	–	–	–	–	–	–	9.80	15.32	15.88	12.79	9.71	7.22
P33	–	–	–	–	–	–	–	6.50	10.00	13.14	14.22	14.52
P34	–	–	–	–	–	–	–	–	7.12	13.23	16.81	18.06
P34	–	–	–	–	–	–	–	–	–	14.11	17.89	17.08
P38	–	–	–	–	–	–	–	–	–	–	6.14	9.99
P39	–	–	–	–	–	–	–	–	–	–	–	6.37
P40	–	–	–	–	–	–	–	–	–	–	–	–

The symmetric feature of the major grooves gives the tetraplex a different look from d(TAACCC). The broad grooves in d(TAACCC) are very different, with the phosphate groups from one side projecting away from the center of the molecule, while the other side is essentially flat. In d(C_4), such a rotation of phosphate grooves is nonexistent, as seen in Figure 2-7. In Figure 2-5 and Figure 2-6, the phosphate groups on both sides are rotated away from the center and bent over toward each other. This feature results shorter phosphate-phosphate distances across both major grooves (see Table 2.7).

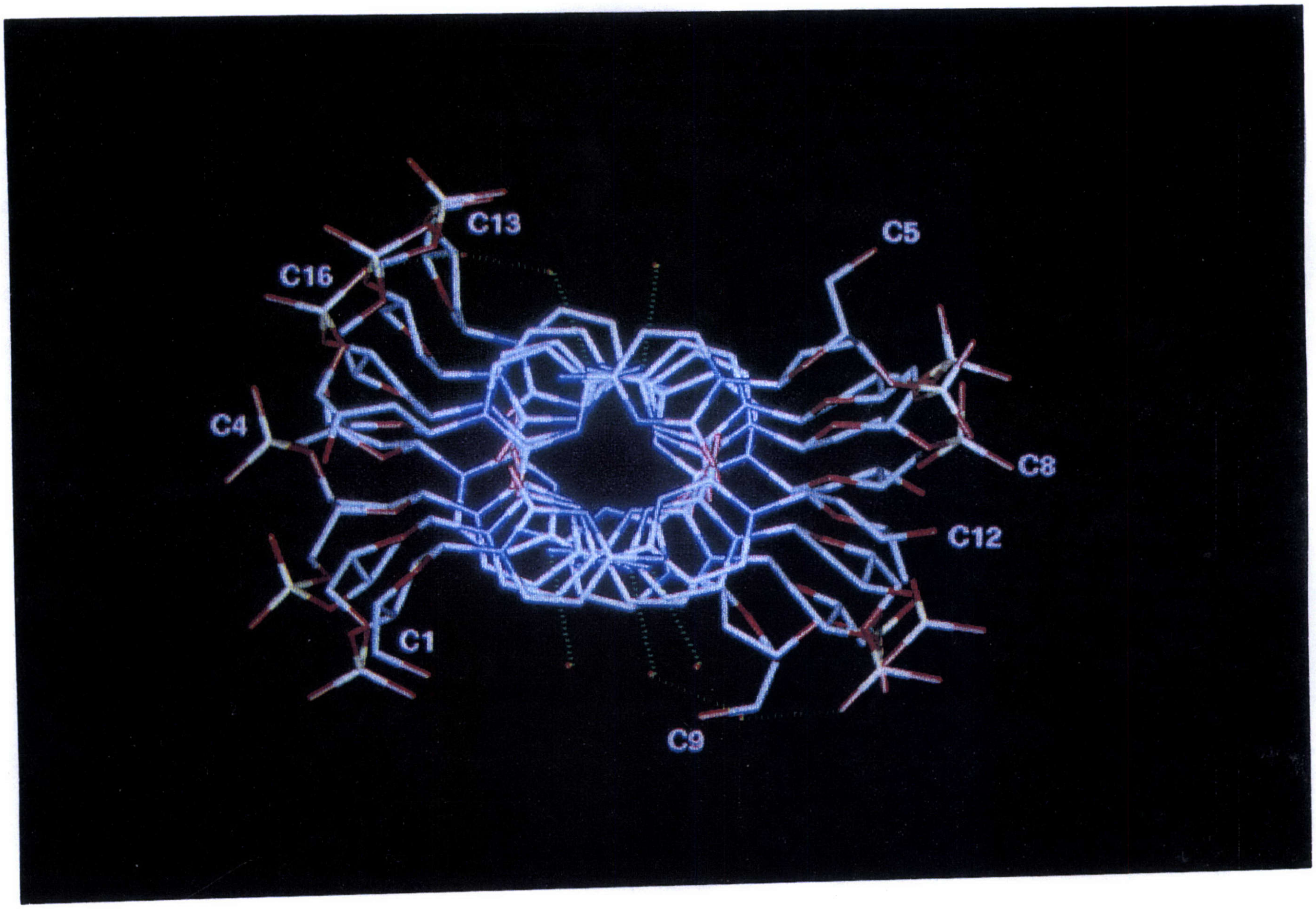

Figure 2-7: View of Tetraplex A of d(C_4) From Its Helical Axis.

In both C-tetraplex 1 and C-tetraplex 2, the molecules twist in a right-handed manner. The twist angles between two covalently linked cytosine residues vary greatly, ranging from 12.9^o to 25.5^o, with a standard deviation of 4.6^o. The average twist angles, however, are fairly similar for both tetraplexes, at about 21^o. This value is significantly greater than that of d(C_4), whose average twist angle is only 12.5^o. We can reason that the much larger twist angle for d(AACCC) is due to the presence of adenine residues and crystal packing. As we see in Figure 2-3 and Figure 2-4, the adenine residues point away from the C-tetraplexes, with an orientation perpendicular to that of the cytosine bases. The cytosine portion of each strand thus needs to twist by a certain angle to accomodate this configuration. This reasoning is supported by the fact that despite the unusually large standard deviation (4.6^o) for the twist angles, the average twist angle values for C-tetraplex 1 (21.47^o) and C-tetraplex 2 (20.22^o) are very close, as both have to turn about the same amount. One can further reason that with more (less) cytosine bases, the twist angle will become smaller (larger). This hypothesis is supported by our observation in the structure of d(AACCCC), the *Tetrahymena* telomere sequence, as presented[19, 20] in Chapter 3 of this thesis. The average twist in d(AACCCC) is about 16^o, smaller than the average value for d(AACCC), as we expected. Table 2.8 and Table 2.9 summarize the values of duplex and quadruplex parameters, respectively.

Table 2.8: Duplex Parameters

Parallel duplex	Base step	Rise(Å)	Twist(°)
A1	C3·C8-C4·C9	6.34	23.32
	C4·C9-C5·C10	6.21	18.41
A2	C13·C18-C14·C19	6.20	23.03
	C14·C19-C15·C20	6.24	21.11
B1	C23*·C28-C24*·C29	6.44	25.48
	C24*·C29-C25*·C30	6.21	12.90
B2	C40*·C35*-C39*·C34*	6.25	16.52
	C39*·C34*-C38*·C33*	6.38	25.98
	A Average	6.25	21.47
	A Standard Deviation	0.06	2.26
	B Average	6.32	20.22
	B Standard Deviation	0.11	6.58
	Overall average	6.28	20.84
	Overall standard deviation	0.09	4.58

Table 2.9: Quadruplex Parameters

Quadruplex	Base step	Pseudo-Rise(Å)	Pseudo-Twist(^{o})
A	C20·C15-C3·C8	3.23	23.71
	C3·C8-C19·C14	3.11	-44.43
	C19·C14-C4·C9	3.08	21.11
	C4·C9-C18·C13	3.13	-44.14
	C18·C13-C5·C10	3.11	25.92
B	C23*·C28-C40*·C35*	3.31	45.55
	C40*·C35*-C24*·C29	3.12	-20.14
	C24*·C29-C39*·C34*	3.13	36.65
	C39*·C35*-C25*·C30	3.08	-23.75
	C25*·C30-C38*·C33*	3.29	49.69
	Overall average	3.16	33.77/-33.11
	Overall standard deviation	0.08	12.03/12.98

Right-handed twist is positive, left-handed twist is negative.

The three strong hydrogen bonds in a $C{\cdot}C^{+}$ base pair and the stacking interaction between the base pair layers give the I-motif a very rigid, stable structural appearance despite the variation of strand polarity observed in this structure. The sugar-phosphate backbone, however, varies considerably due to the electrostatic repulsion as the two anti-parallel backbone chains fit so closely in the minor groove. When we superimpose cytosine-tetraplex 2 of d(AACCC) with tetraplex A of d(C_4), the r.m.s. differences are quite considerable, especially between the sugar-phosphate backbones. The overall tetraplexes show an r.m.s. difference of 1.4$\AA$, and the backbones have an r.m.s. of 1.85$\AA$. On the contrary, the bases only have an r.m.s. of 0.54$\AA$. This observation leads us to believe that most of the structural variations among the cytosine tetraplexes will rise from the sugar-phosphate backbone.

Hydration in the Major Groove

Another interesting feature of d(AACCC) is its hydration pattern in the major grooves. There are 57 water molecules identified in an asymmetric unit, most of which form hydrogen bonds to the hydrogens of amino group N4, which is not included in the $C{\cdot}C^{+}$ base pairing, or form hydrogen bonds to the backbone phosphate oxygens. We have observed similar features in d(C_4). However, as described earlier in this chapter, the phosphate groups of major grooves are rotated away from the center. This configuration is stabilized by the bridging water molecules. The phosphate oxygens are connected to water molecules, which in turn form hydrogen bonds either directly or through another bridging water molecule to the N4 of cytosines. This hydration pattern is observed in both tetraplexes (Figure 2-5 and Figure 2-6), and is nearly absent in d(C_4) (Figure 2-7).

Continuous Stacking Between $C{\cdot}C^{+}$ And $A{\cdot}A$ Base Pairs

Figure 2-8 shows an asymmetric unit which consists of eight independent strands. This particular selection of asymmetric unit shown is such that the center shows four strands forming C-tetraplex 1, which stacks directly on top of three $A{\cdot}A$ base pairs. There are two other adenine residues stacked on top of each other at an angle about

38^o from the rest of $A{\cdot}A$ base pairs. We term this arrangement of eight adenine residues A-cluster 2, which will be discussed shortly. The stacking of C-tetraplex 1 and A-cluster 2 is along the x-axis. Thus, the cytosines of the four strands in tetraplex 1 and the adenines of the other four strands form a stacking column of nine layers along the x-axis. Similar stacking also occurs along the z-axis. If one chose the asymmetric unit in another way, four strands would form C-tetraplex 2, which stacks on four layers of $A{\cdot}A$ base pairs, formed by the adenines from the other four strands. This arrangement of the other eight adenine residues is termed A-cluster 1. Therefore, a ten-layer stacking pattern occurs along the z-axis. As we shall see in the crystal packing section, the nine-layer column stacks continuously on itself, forming an infinite stacking pattern along the x-axis. Similarly, repeats of the ten-layer column stack continuously along the z-axis.

One interesting note: while the cytosine base stacking distance continues to be $3.2\AA$ (Table 2.9), a feature which is in agreement with the stacking distances in d(C_4), the stacking distances of adenine pairs are $3.5\AA$. The increase in stacking distance is due to the involvement of aromatic rings in stacking, which is similar to that of B-DNA. The sugar puckering modes of the adenines are also close to those of B-DNA. Out of the total of 16 adenine residues, 9 are C2'-endo, 3 are C3'-exo, and 4 are C1'-exo. For the cytosines, the most common one is still C4'-exo, which is the same as what we saw in d(C_4). This, together with O4'-endo and C3'-endo, amount for the majority of the sugar puckering modes for the cytosines. Table 2.10 lists all the torsion angles for the residues.

2.3.3 Two Adenine Clusters

In Figure 2-3 and Figure 2-4, we have seen the adenine residues project away from the central C-tetraplexes and face the directions that are perpendicular to the helical axes of the respective C-tetraplexes. In Figure 2-8, eight adenine residues from four different strands form a cluster under cytosine tetraplex 1. The eight adenines from the top four strands of the figure form a different A-cluster with the symmetry-related residues. The two A-clusters thus serve as the building blocks of the entire lattice. Figure 2-9 shows A-cluster 1. The eight adenine residues form four $A{\cdot}A$ base pairs, which stack on each other in the z-direction. As described above, A-cluster 1 is stacked on C-tetraplex 2, and this stacking alternates continuously along the z-direction. A-cluster 2 is shown in Figure 2-10. Note there are only three $A{\cdot}A$ base pairs present in this case. The other two, A27 and A36, stack on top of each other and do not form base pairs. The three base pairs stack along x-axis. Similar to A-cluster 1, A-cluster 2 stacks on C-tetraplex 1, and this stacking pattern continues along the x-axis. Figure 2-11 shows the stacking of A-cluster 1 and C-tetraplex 2 along the edge of the unit cell in z-direction. This combines with the stacking along x-axis, forming a three-dimensional network (Figure 2-12). Figure 2-13 gives another view of A-cluster 2, with four connecting cytosine strands. Note the white sphere in the middle. That is the mercury atom sitting between two layers of $A{\cdot}A$ base pairs. A more detailed description of A-clusters and $A{\cdot}A$ base pairing modes will be given in chapter 3, when the structure of *Tetrahymena* telomere sequence d(AACCCC) is presented.

2.4 DISCUSSION

Structure of d(AACCC) is very important as it gives us a first look of what the telomere structures might be with the addition of adenine residues. It again confirms the intercalating cytosine motif and gives a few unexpected variations of the motif. It is interesting to note that its overall structure is quite different from that of d(TAACCC), the human telomere repeat, with a single difference of thymine residue. The adenine residues play a very important role in building the lattice, and we expect the non-cytosine residues would be more flexible in adopting different structural motifs under different conditions.

The ability of adenine residues to interact in different ways by forming base pairs and stacking gives a new dimension to the topic of C-rich telomere structures. It could play an important role in stabilizing long stretches of telomere sequences by adopting some of the features seen in this structure. Moreover, stretches of adenines also exist in other segments of the genome, and these novel features could be useful not only in telomeres, but also in other large nucleic acid assembly to facilitate folding and long-range tertiary interactions.

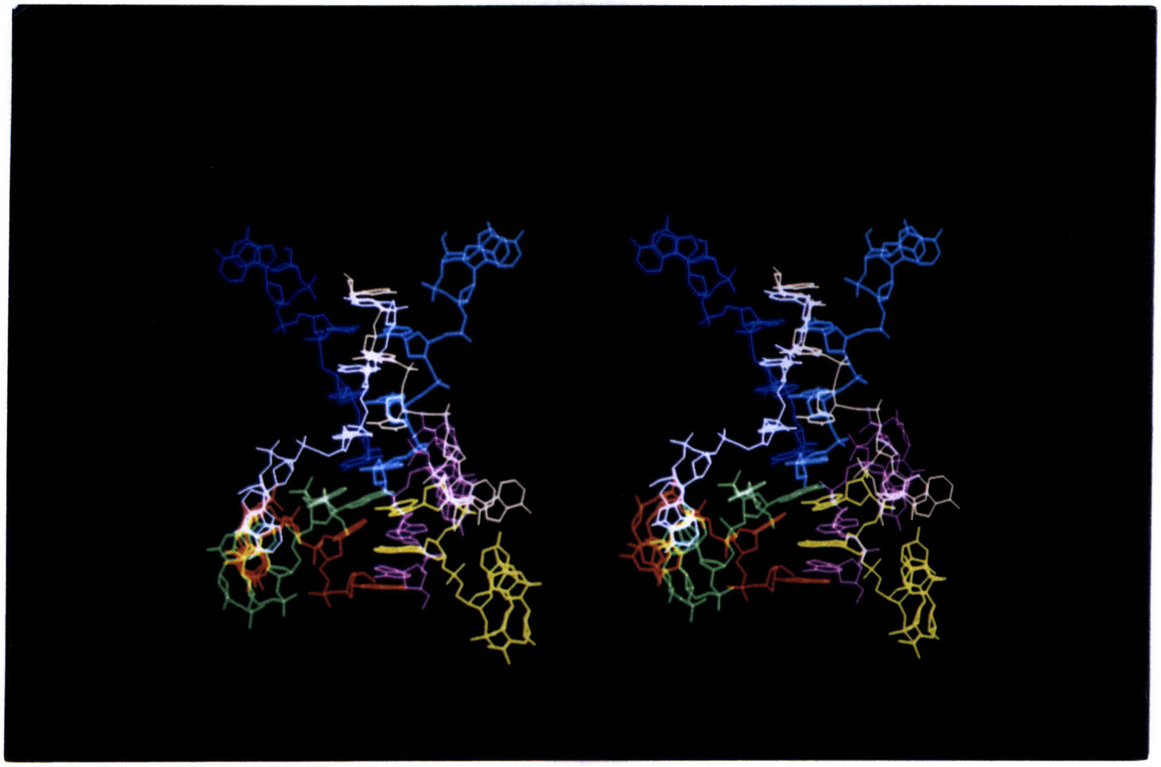

Figure 2-8: Continuous Stacking Between $C{\cdot}C^+$ and $A{\cdot}A$ Base Pairs.

Table 2.10: Glycosidic Torsion Angles and Sugar Puckers of d(AACCC)

Strand	Residue	α	β	γ	δ	ϵ	ζ	χ	Pseud	Pucker
a1	A1	n.a.	n.a.	52.8	137.1	190.4	273.1	258.3	205.1	C3'-exo
	A2	296.5	165.1	50.5	132.2	248.8	141.9	282.6	139.0	C1'-exo
	C3	286.6	228.0	72.0	139.8	246.3	249.7	235.4	143.1	C1'-exo
	C4	109.4	197.7	198.4	84.1	194.2	264.8	233.8	74.7	O4'-endo
	C5	173.3	197.1	173.0	90.0	n.a.	n.a.	240.7	63.7	C4'-exo
a2	A6	n.a.	n.a.	56.8	144.8	191.6	267.1	257.4	171.3	C2'-endo
	A7	282.0	192.2	55.0	141.4	273.4	295.5	298.4	170.7	C2'-endo
	C8	289.0	191.5	55.4	78.6	205.2	272.6	222.5	20.0	C3'-endo
	C9	296.3	178.9	62.2	85.6	198.1	275.9	233.4	35.2	C3'-endo
	C10	317.3	154.5	69.0	75.3	n.a.	n.a.	224.6	41.5	C4'-exo
a3	A11	n.a.	n.a.	174.9	158.3	183.4	270.1	265.4	197.6	C3'-exo
	A12	290.8	164.9	54.2	132.2	271.0	290.8	276.7	148.2	C2'-endo
	C13	298.6	200.5	64.2	143.8	229.2	291.2	239.4	150.8	C2'-endo
	C14	198.5	124.6	168.3	107.3	196.9	298.9	233.7	104.8	O4'-endo
	C15	161.6	170.6	182.7	92.4	n.a.	n.a.	247.9	96.4	O4'-endo
a4	A16	n.a.	n.a.	182.7	153.0	192.4	279.9	249.2	177.3	C2'-endo
	A17	281.7	166.9	53.2	122.0	269.9	294.1	269.5	123.8	C3'-exo
	C18	280.4	198.5	54.2	80.0	213.7	263.4	215.1	32.1	C3'-endo
	C19	300.5	175.9	55.1	84.4	193.5	274.8	227.2	21.5	C3'-endo
	C20	312.6	164.5	62.5	82.3	n.a.	n.a.	225.6	28.3	C3'-endo
b1	A21	n.a.	n.a.	176.0	158.8	196.3	276.0	80.1	176.5	C2'-endo
	A22	288.4	169.6	55.2	134.7	244.2	291.1	288.7	140.9	C1'-exo
	C23	277.6	178.3	52.3	102.3	245.4	245.3	217.1	43.7	C4'-exo
	C24	301.0	160.7	58.2	85.5	194.6	272.7	236.4	49.7	C4'-exo
	C25	298.7	166.9	68.0	70.5	n.a.	n.a.	242.7	44.4	C4'-exo
b2	A26	n.a.	n.a.	58.4	131.6	261.7	61.5	53.2	153.3	C2'-endo
	A27	99.0	172.1	183.4	154.6	269.5	292.1	62.7	172.6	C2'-endo
	C28	292.5	202.1	66.2	93.5	213.4	264.2	213.4	48.5	C4'-exo
	C29	296.8	175.7	62.3	85.1	190.7	280.6	237.4	27.9	C3'-endo
	C30	174.7	188.7	177.1	91.8	n.a.	n.a.	234.4	45.5	C4'-exo
b3	A31	n.a.	n.a.	56.3	140.9	200.2	272.2	66.2	151.4	C2'-endo
	A32	286.4	158.2	51.3	115.4	243.8	283.3	280.3	115.2	C1'-exo
	C33	296.3	190.1	55.8	150.0	240.5	293.3	199.6	184.4	C3'-exo
	C34	126.9	185.2	177.2	104.3	212.7	266.0	239.1	83.1	O4'-endo
	C35	284.7	173.3	74.5	91.0	n.a.	n.a.	233.5	80.0	O4'-endo
b4	A36	n.a.	n.a.	183.0	100.8	206.2	96.8	66.9	108.7	C1'-exo
	A37	168.4	167.7	54.3	146.2	263.6	296.8	292.1	166.8	C2'-endo
	C38	296.1	207.0	57.7	145.6	260.6	280.2	193.6	195.3	C3'-exo
	C39	293.9	151.1	57.2	80.3	184.2	272.1	238.0	76.5	O4'-endo
	C40	172.2	183.6	169.3	81.1	n.a.	n.a.	237.7	44.2	C4'-exo

Backbone torsion angles for the bonds in the backbone P-O5'-C5'-C4'-C3'-O3'-P are α, β, γ, δ, ϵ and ζ, respectively, and the glycosidic angle is χ. (a) is calculated by the program NEWHEL93.

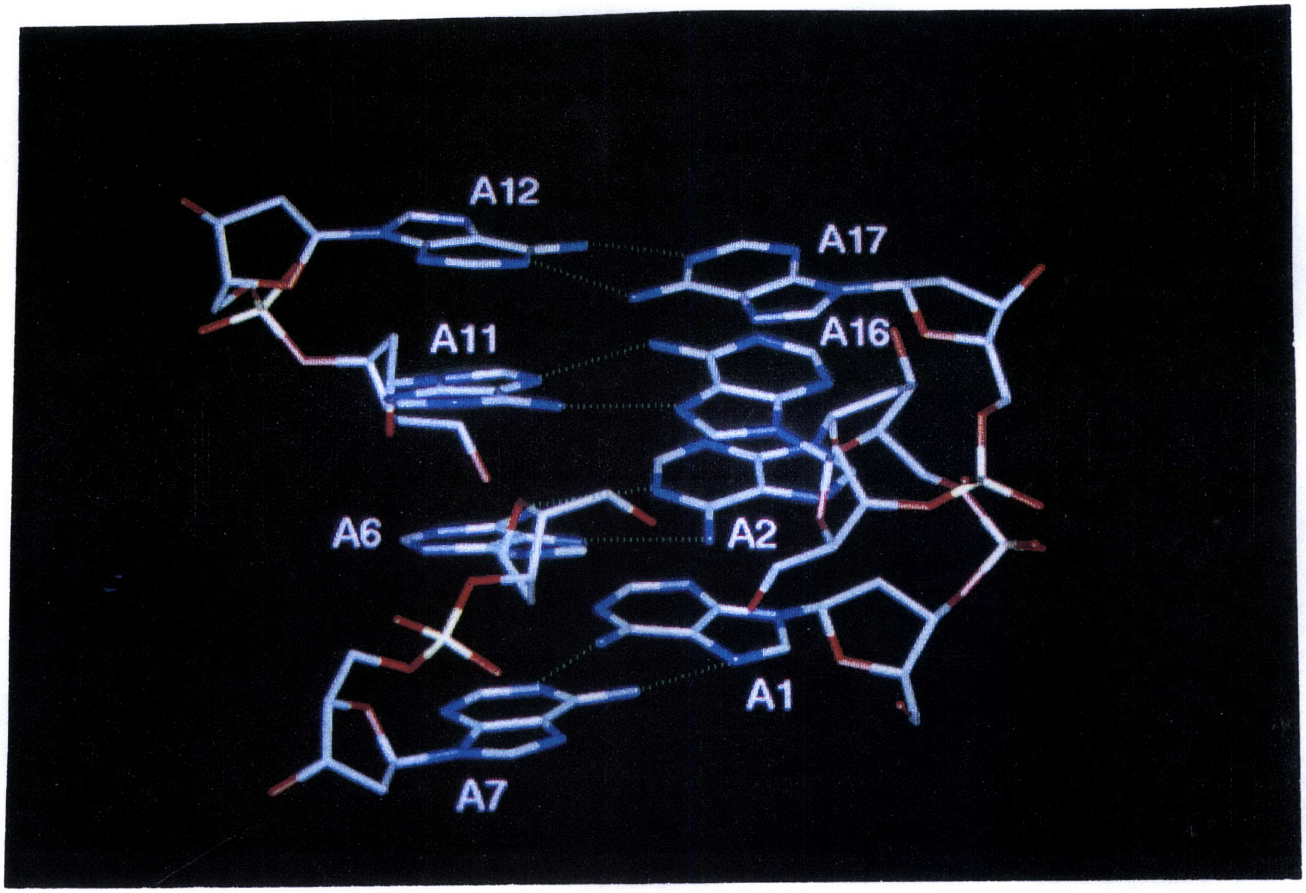

Figure 2-9: A-cluster 1 of d(AACCC).

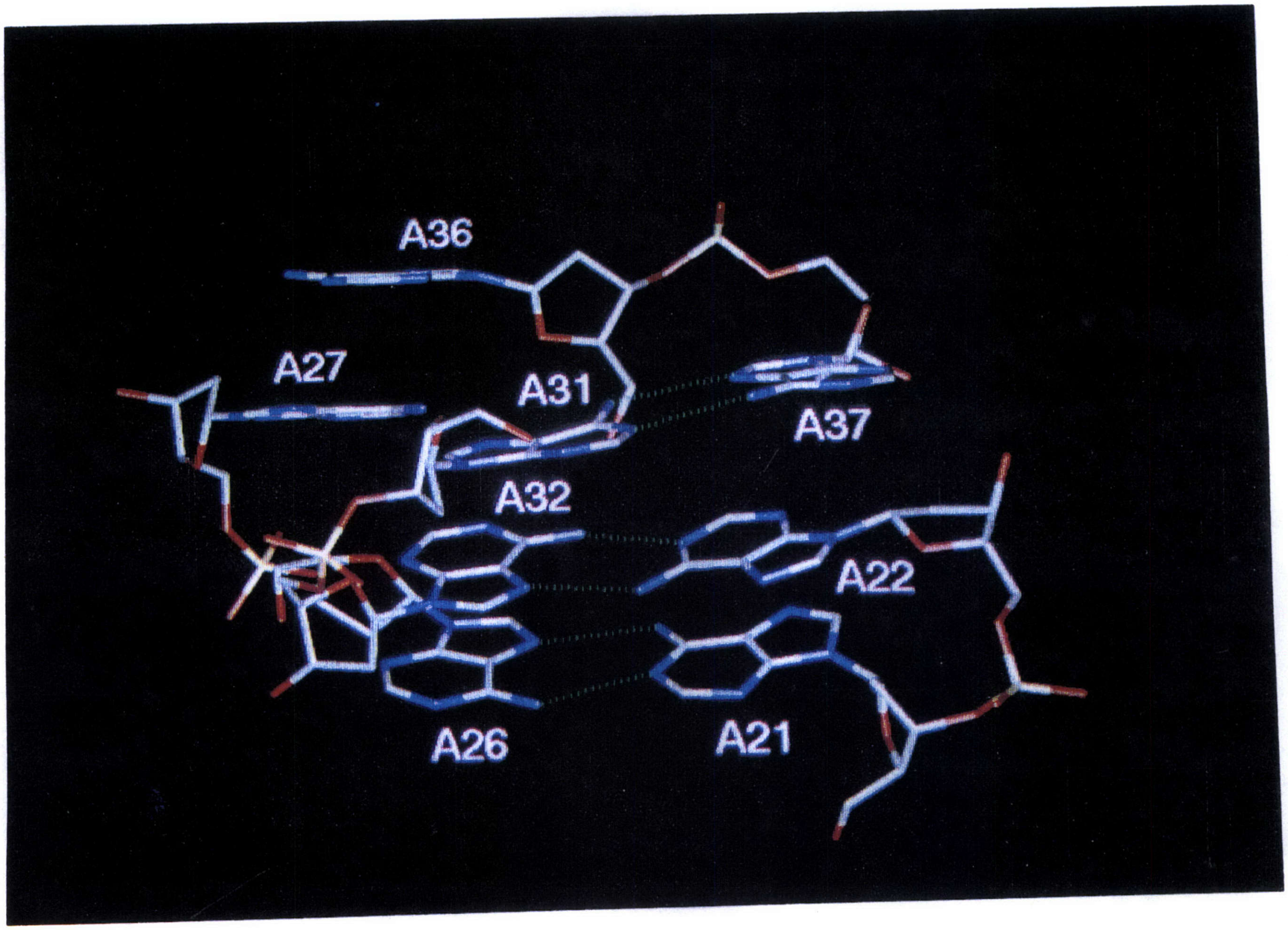

Figure 2-10: A-cluster 2 of d(AACCC).

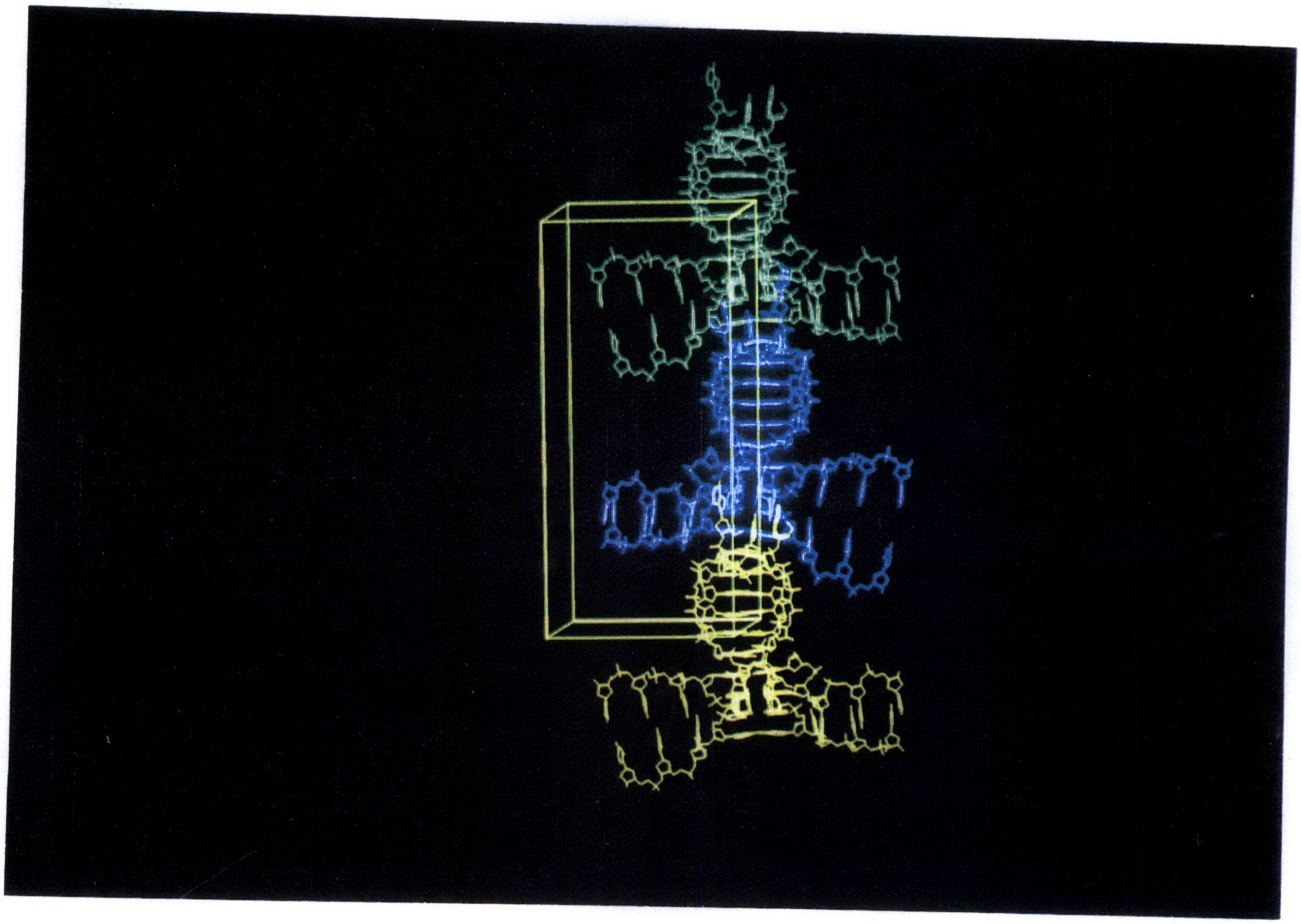

Figure 2-11: Stacking of A-cluster 1 and C-tetraplex 2 Along Z-axis.

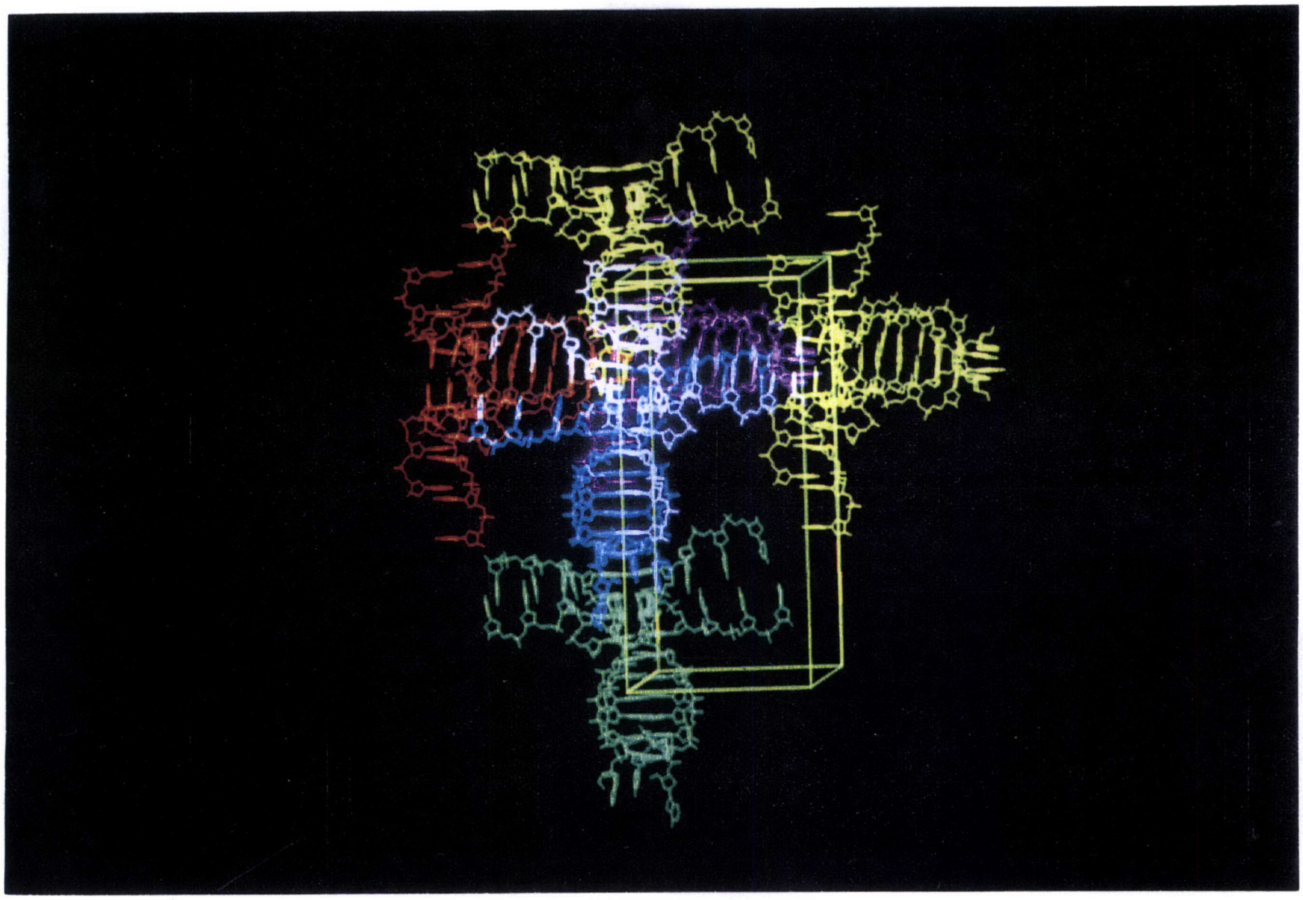

Figure 2-12: 3-D Network Formed by Continuous Stacking along X- and Z-axes.

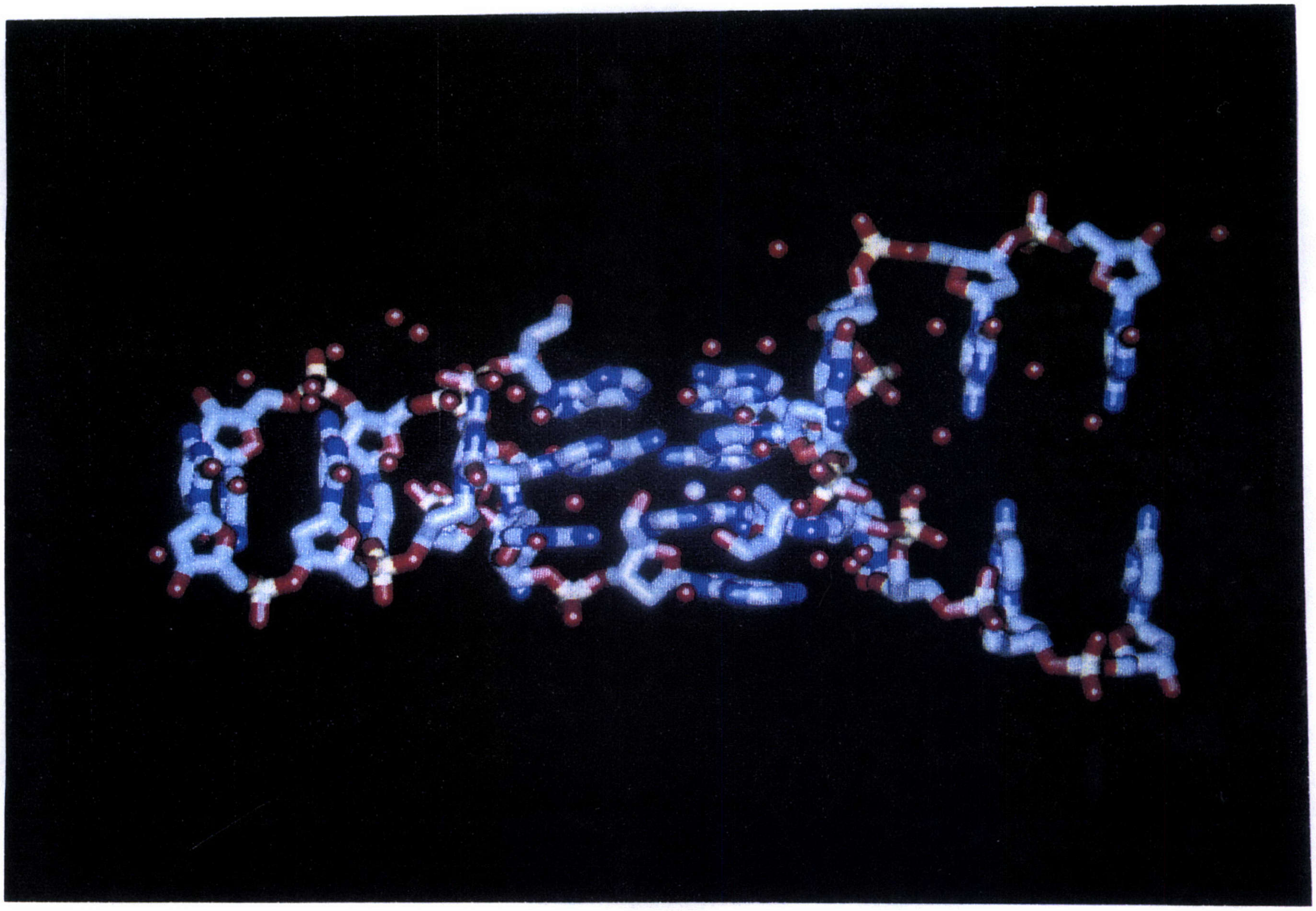

Figure 2-13: Position of Hg atom within A-cluster 2.

Figure Legends

Figure 2-1. Harker sections of the isomorphous difference Patterson and anomalous difference Patterson maps for the crystal of d(AACCC). (Top) Isomorphous difference Patterson map with data from 20.0Å to 3.0Å. The Harker section at X=0 and Z=0.5 are shown here. (Bottom) Anomalous difference Patterson map with data from 8.0Å to 3.0Å. The Harker section at X=0 and Z=0.5 show the Hg-Hg peak are at the same positions. It confirms the Hg position found in the isomorphous map.

Figure 2-2. Superposition of the part of the initial (SIR+SAS) density map after solvent flattening with the final refined structural model. The phosphates are identifiable as intense peaks (not shown here) and the $C{\cdot}C^+$ and $A{\cdot}A$ base pairs within each layer are clearly visible. The map is contoured at 1σ density level.

Figure 2-3. Cytosine tetraplex 1 of the structure d(AACCC). The center of the molecule is composed of intercalating cytosine residues held together by $C{\cdot}C^+$ base pairs. Note that there are two adenine residues at the 5' end of each strand, and they are projecting away from the center of the molecule. The outermost $C{\cdot}C^+$ base pairs of the tetraplex are from the 3' end of each strand.

Figure 2-4. cytosine tetraplex 2 of the structure d(AACCC). (Left) The intercalating motive here is very similar to that of tetraplex 1. However, the outermost $C{\cdot}C^+$ base pairs of the tetraplex are from the 5' end of each strand.

Figure 2-5. Six adjacent layers of $C{\cdot}C^+$ base pairs from tetraplex 1 along with water molecules in the major groove. Several water bridges are present, where the water molecule forms hydrogen bonds with both phosphate oxygen and N4 of cytosine, thus stabilizing the rotated phosphate groups. This kind of water bridge was rare

in d(C_4), and the symmetric nature of the broad grooves makes it different from d(TAACCC).

Figure 2-6. Six adjacent layers of $C{\cdot}C^+$ base pairs from tetraplex 2 along with water molecules in the major groove. Again, several water bridges are present, where the water molecule forms hydrogen bonds with both phosphate oxygen and N4 of cytosine, thus stabilizing the rotated phosphate groups.

Figure 2-7. Eight adjacent layers of $C{\cdot}C^+$ base pairs from tetraplex A of d(C_4) along with water molecules in the major groove. The broad grooves are very flat and symmetric. Most water molecules form hydrogen bonds with just the N4 of cytosine.

Figure 2-8. Eight strands are present in an asymmetric unit. The cytosines of the top four strands form C-tetraplex 1, while the adenines of the bottom four strands form A-cluster2, which consists of three $A{\cdot}A$ base pairs and two stacking adenine residues. The C-tetraplex 1 in turn stacks on the A-cluster 2, forming nine layers of continuous stacking.

Figure 2-9. A-cluster 1 of d(AACCC). (a) A schematic diagram illustrating the formation of A-cluster 1 and its relation with the rest of the cytosine residues of the strands. There are two parallel $A{\cdot}A$ base pairs, namely, A11-A16 and A12-A17. The other two $A{\cdot}A$ base pairs, A2-A6 and A1-A7, are antiparallel. Every cytosine portion of four strands combines with three other symmetry related cytosines strands (not shown) to form tetraplex 1. Therefore, there are four cytosine tetraplexes 1 connected by A-cluster 1. It stacks on two cytosine tetraplexes 2 (not shown), from top and bottom, forming continuous stacking along z.

Figure 2-10. A-cluster 2 of d(AACCC). Note there are only three stacking base pairs. The other two bases form stacking of their own, which is tilted about 38^o with the other base pairs. Like A-cluster 1, A-cluster 2 connects four cytosine tetraplexes 2. Of the three $A{\cdot}A$ base pairs, A31-A37 and A21-A26 are parallel. A22-A32 is antiparallel. A-cluster 2 connects four cytosine strands which belong to four different cytosine tetraplexes. It stacks on two cytosine tetraplexes 1 (not shown), from top and bottom, forming continuous stacking along x.

Figure 2-11. Continuous stacking z axis formed by C-tetraplex 1 and A-cluster 2. The box shown is the unit cell of the crystal. Similar stacking exits along the x-axis, formed by C-tetraplex 2 and A-cluster 1.

Figure 2-12. View of the three dimensinal network formed by continuous stacking along x and z axes. The box shown is the unit cell of the crystal.

Figure 2-13. View of the position of the mercury atom that is used in heavy atom soaking. It is the white sphere sitting between two base pair layers in A-cluster 2.

Chapter 3

Intercalated Cytosine Motif and Novel Adenine Clusters in the Crystal Structure of Tetrahymena Telomere

3.1 INTRODUCTION

In Chapter 2, we have stated several known biological importances of telomere DNA at chromosome ends. The first telomere isolated was from ciliate *Tetrahymena thermophila* in the early 1970s, with its C-rich strand being multiple repeats of d(AACCCC), and its complementary G-rich strand being multiple repeats of d(GGGGTT) [15]. Structural studies in recent years showed that guanine-rich repeats can form tetraplexes by using four guanines hydrogen-boned in a cyclic fashion [17, 18, 25].

Recent NMR experiments of d(TC_5) and the first crystal structure of cytosine rich DNA sequence d(C_4), gave a rather novel structural motif: intercalated tetraplex (I-motif), in which the same $C{\cdot}C^+$ pairings were seen in two parallel stranded duplexes intercalated into each other antiparallelly [5, 6, 7]. The crystal structure, in addition, revealed more detailed structural information. Subsequently, several more crystal studies also confirmed the I-motif and showed structural variation among different sequences [26, 27, 28]. We previously solved d(AACCC) (see chapter 2), a sequence that closely resembles the *Tetrahymena* telomere sequence, giving us a first look of what the *Tetrahymena* telomere structure might be, and how the adenines relate structurally to the cytosine stretches.

The crystal structure of d(AACCCC), telomeric cytosine-rich repeating sequence in *Tetrahymena*, was recently solved to 2.5 $\AA$ resolution. The adenines form two different novel A-clusters in orthogonal directions, with their counterparts from other strands, using three base-paring modes. Two distinct cytosine tetraplexes are found. Each four-stranded complex is composed of two intercalated parallel-stranded duplexes pointing in opposite directions, using hemiprotonated cytosine-cytosine base pairs. The outermost $C{\cdot}C^+$ base pairs, are from the 5' end of each strand in one cytosine tetraplex and from the 3' end of each strand in the other. The A-clusters, along with the cytosine tetraplexes, form two alternating A-cluster-C-tetraplex stacking patterns, creating continuous base stacking along the x and z axes.

3.2 MATERIAL AND METHODS

3.2.1 DNA Synthesis and Crystallization

The oligodeoxyribonucleotide d(AACCCC) was synthesized on a 10μM scale on an Applied Biosystems DNA synthesizer using solid phase β- cyanoethylphosphoramidite chemistry. It was then purified by reverse-phase high performance liquid chromatography (HPLC) on a C4 column (Rainin Instrument Co.), with a linear gradient of 5-40% acetonitrile in 0.1 M triethylammonium acetate buffer PH 7.0. The peak eluent containing the pure d(AACCCC) was collected and lyophilized overnight.

Crystals of d(AACCCC) were grown at room temperature using the vapor diffusion method from solutions containing 2.0 mM d(AACCCC) and 100 mM sodium cacodylate buffer with various pHs equilibrated with a reservoir of 70% ammonium sulfate. The best crystal, with a size of 0.3mm x 0.2mm x 0.1mm, came out of the buffer with pH 7.5. The crystal diffracted to 2.5 $\AA$ resolution. It crystallized in space group $P22_12_1$ with unit cell dimensions a=35.93 $\AA$, b=52.33 $\AA$, c=76.94 $\AA$. The crystal contains eight strands per asymmetric unit. The crystal data of the sample is summarized in Table 3.1.

Table 3.1: Crystal Data of d(AACCCC)

Space group	$P22_12_1$
a	35.93$\AA$
b	52.33$\AA$
c	76.94$\AA$
strands per unit cell	32
strands per asymmetric unit	8

3.2.2 Data Collection

All diffraction data were collected on a Rigaku Raxis II imaging plate system at 4^oC and processed with the PROCESS program provided by the Molecular Structure Corporation. The data set was collected to 2.5$\AA$ resolution, with 64 frames at a crystal-to-plate distance of 120mm with 4^o oscillations. There are 4628 independent reflections above the 1σ (I) level from 20-2.5$\AA$. 75% of the reflections in the resolution shell between 2.75-2.5$\AA$ were observed. Overall completeness from 20-2.5$\AA$ is 86.5%. Data collection statistics are summarized in Table 3.2.

Table 3.2: Summary of Data Collection Statistics

	Native
Resolution ($\AA$)	2.5
Number of observations	33551
Number of unique reflections	4628
Overall completeness (%)	86.5
Outermost shell ($\AA$)	2.75-2.5
Outermost shell completeness (%)	75
R_{merge} (%)	9.63

3.2.3 Refinement of the Structure

We solved the crystal structure d(AACCC) by single isomorphous replacement and single anomalous scattering method (see Chapter 2). We used the two different cytosine tetraplexes of that structure as the starting models in an attempt to solve d(AACCCC) by molecular replacement using XPLOR [12]. Rotation and translation searches with these two models at various resolution ranges of the d(AACCCC) diffraction data always led to the same orientations (i.e. along x and z axes) of the molecule in the lattice, clearly showing the asymmetric unit contained two cytosine tetraplexes, which means there are eight independent strands of d(AACCCC) in the asymmetric unit, as we expected from the volume calculations. Four initial models were then built, using all the combinations of the two initial tetraplexes. The position of the molecule showed the orientation of helical axis of one tetraplex parallel to x-axis and the helical axis of the other parallel to the z-axis. This stacking pattern is in complete agreement with the native Patterson map of the molecule. After several cycles of rigid body refinement using 10-2.5$\AA$ data, the difference map of the correct model allowed us to identify the missing adenines and the extra cytosines. We then carried out simulated annealing refinement, leading to an R factor of 25.2%. Twenty cycles of restrained individual isotropic B-factor refinement followed. Well ordered water molecules were then located from the difference Fourier map ($F_o - F_c$) and added as oxygen atoms to the model only if they had a peak height of over 3σ in the difference density map. A total of 60 water molecules were found in this way. A final round of refinement completed the structural determination with an R-factor of 0.213 and r.m.s deviations from ideal bond lengths and angels of 0.016$\AA$ and 3.7^o, respectively. The free R-factor [29] based on a random subset of 10% of the reflections is 29%. A summary of the refinement statistics is listed in Table 3.3. The atomic coordinates have been deposited in the Protein Data Bank.

Table 3.3: Refinement Statistics

Resolution ($\AA$)	**8-2.5**
Number of Reflections (I>1σ(I))	4628
Completeness (%)	86.5
Number of non-hydrogen DNA atoms	836
Number of water molecules	60
r.m.s. bond length($\AA$)	0.016
r.m.s. bond angles(^{o})	3.7
R-factor	0.213

$$R=\Sigma \mid F_{observed} - F_{calculated} \mid / \Sigma \mid F_{observed} \mid$$

3.3 RESULTS

3.3.1 Overview

The cytosines of the eight strands in an asymmetric unit, organized into two intercalated tetraplexes, closely resemble those in the structure of d(AACCC). However, with the addition of one cytosine residue per strand, the twist of the strands is significantly less severe than that of d(AACCC), and unlike d(AACCC), where the phosphate groups in both broad groves rotate away from the center and bend over towards each other, the strands are straight and flat in this case. The non-cytosine bases in various telomeric structures have shown a greater degree of variability. In metazoan telomeric sequence d(TAACCC), a stabilized loop was formed by TAA. In *Tetrahymena* telomeric sequence d(AACCCC), however, the structure displays an interesting structural motif first observed in d(AACCC): the adenine cluster. The adenines, sitting at the 5' end of each strand, form two different types of A-clusters, with three stacking $A{\cdot}A$ base pairs in one and four stacking $A{\cdot}A$ base pairs in the other. There are three different base-pairing modes present. The stacking $A{\cdot}A$ base pairs in each A-cluster also stack upon the two different modes of cytosine tetraplexes in two orthogonal directions to form alternating A-cluster-C-tetraplex base stacking continuously along x and z axes. These novel features were in agreement with our previously solved structure d(AACCC) (L. Chen, L. Cai, Q. Gao, and A. Rich, unpublished data), the common sequence of both tetrahymena and human telomere sequences. Notable differences occur in the geometry of cytosine tetraplexes.

3.3.2 Two Different Cytosine Tetraplexes

The oligonucleotide d(AACCCC) crystallizes in the orthorhombic space group $P22_12_1$. There are eight strands in the asymmetric unit, enough to form two cytosine tetraplexes. Figure 3-1 and Figure 3-2 show tetraplex 1 and tetraplex 2, respectively, along with the adenines. At the center of each figure shown are the four cytosines from four different chains organized into an intercalation motif. In Figure 3-1, the cytosine bases stack along the x-axis, and in Figure 3-2, the cytosine bases stack along the z-axis, with an average stacking distance of 3.2 $\AA$. The stacking distance of 3.2 $\AA$ is in agreement with those of previously solved C-tetraplex structures and occurs when the stacking is limited to the exocyclic amino and carbonyl groups. The adenines from each strand project out at the top and bottom, with the planes containing the A bases being nearly perpendicular to those containing the cytosine bases (with the exception of A52 and A71). Thus the adenines in Figure 3-1 are perpendicular to the z-axis and the adenines in Figure 3-2 are perpendicular to the x-axis.

Careful inspection of the two cytosine tetraplexes clearly shows that the configuration of the two tetraplexes differ in a subtle way. In Figure 3-1, tetraplex 1, which runs along the x-axis, has the utmost $C{\cdot}C^{+}$ layer coming from the 3' end of the strands. In Figure 3-2, tetraplex 2, which runs along the z-axis, has the utmost $C{\cdot}C^{+}$ layer coming from the 5' end of the strands. This motif is in agreement with that of d(AACCC), and this formation in Figure 3-1 was not observed previously in other published structures and is the major variation among the C-tetraplexes.

In each tetraplex, the interaction of two parallel duplexes yields a quadruplex with two very wide and two very narrow grooves which, like d(C_4) and d(AACCC), are basically symmetrical about the helical axis. The narrow groove is made of two closely packed strands running antiparallelly. The two backbone chains fit to each other remarkably well in a zig-zag way. They are so close to each other that some inter-chain P-P distances are even shorter than intra-chain ones. In tetraplex 1, the average intra-chain P-P distance is 6.33 $\AA$. The average inter-chain P-P distance across the minor groove is 6.36 $\AA$, with the shortest being 5.62 $\AA$. The average inter-chain P-P distance across the minor groove for tetraplex 2 is very comparable, at 6.81 $\AA$. The minor groove is so narrow that if we take away the van der Waals radius of phosphate, there is very little room left to trap anything. Indeed, we find no water molecules inside the minor groove.

On the contrary, the major grooves are very wide. The average inter-chain P-P distances across the major grooves of tetramer 1 and 2 are 16.09 $\AA$ and 15.19 $\AA$, respectively. Table 3.4 summarizes the phosphate-phosphate distances of cytosine-tetraplex1 and cytosine-tetraplex2.

Table 3.4: Phosphate-Phosphate Distances

(a) Tetramer A

P	P3	P4	P5	P6	P13	P14	P15	P16	P22	P23	P24	P25	P32	P33	P34	P35
P3	–	6.10	11.53	17.75	17.44	16.88	19.02	22.37	18.51	18.90	15.57	14.13	22.57	18.17	13.49	10.90
P4	–	–	6.73	13.37	16.36	16.27	18.32	21.42	18.40	19.16	16.68	16.39	18.20	13.12	8.01	6.18
P5	–	–	–	6.65	18.37	17.11	17.11	18.49	14.87	16.80	16.57	18.80	13.45	7.62	6.45	9.65
P6	–	–	–	–	21.95	19.89	18.02	17.35	13.67	16.61	18.63	22.63	10.69	5.62	10.28	15.31
P13	–	–	–	–	–	5.12	11.27	17.24	22.77	18.62	13.43	10.50	17.41	18.03	14.04	13.33
P14	–	–	–	–	–	–	6.32	12.42	18.39	13.72	8.71	7.73	15.59	16.68	14.19	14.90
P15	–	–	–	–	–	–	–	6.14	13.86	8.50	6.11	10.08	13.62	15.75	15.83	18.33
P16	–	–	–	–	–	–	–	–	10.94	6.06	8.98	15.06	13.43	16.26	18.64	22.33
P22	–	–	–	–	–	–	–	–	–	6.37	11.53	17.70	17.27	16.21	18.73	22.12
P23	–	–	–	–	–	–	–	–	–	–	6.53	13.32	16.50	17.10	18.67	21.65
P24	–	–	–	–	–	–	–	–	–	–	–	6.82	17.66	17.84	17.00	18.44
P25	–	–	–	–	–	–	–	–	–	–	–	–	21.46	21.07	17.95	17.26
P32	–	–	–	–	–	–	–	–	–	–	–	–	–	6.45	11.31	16.84
P33	–	–	–	–	–	–	–	–	–	–	–	–	–	–	6.88	12.72
P34	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	5.86
P35	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

(b) Tetramer B

P	P43*	P44*	P45*	P46*	P53*	P54*	P55*	P56*	P63	P64	P65	P66	P73*	P74*	P75*	P76*
P43*	–	6.52	10.73	17.04	17.14	16.48	17.85	20.93	23.31	23.73	17.74	13.12	18.95	20.69	16.26	13.50
P44*	–	–	6.87	13.23	15.81	16.06	17.94	21.10	20.67	19.85	13.21	7.62	18.74	21.82	18.06	16.09
P45*	–	–	–	6.63	18.42	17.02	16.81	18.09	16.03	14.18	9.00	6.83	13.01	17.38	15.53	16.15
P46*	–	–	–	–	21.20	19.03	17.41	16.93	11.96	8.35	6.50	9.49	10.53	15.89	16.17	18.74
P53*	–	–	–	–	–	5.55	11.40	17.31	19.02	22.21	16.32	13.69	24.57	24.14	18.26	13.67
P54*	–	–	–	–	–	–	5.92	11.94	15.41	19.65	14.79	13.75	20.33	19.06	13.16	9.13
P55*	–	–	–	–	–	–	–	6.07	11.96	17.23	14.26	15.31	16.13	13.81	8.49	7.14
P56*	–	–	–	–	–	–	–	–	9.82	15.62	15.22	18.14	12.79	9.29	6.77	10.00
P63	–	–	–	–	–	–	–	–	–	6.65	9.21	14.88	13.06	13.99	14.12	17.22
P64	–	–	–	–	–	–	–	–	–	–	7.32	13.69	13.40	17.01	18.11	21.32
P65	–	–	–	–	–	–	–	–	–	–	–	6.46	14.38	17.83	16.36	17.50
P66	–	–	–	–	–	–	–	–	–	–	–	–	17.40	20.68	17.58	16.69
P73*	–	–	–	–	–	–	–	–	–	–	–	–	–	6.50	10.35	15.84
P74*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	6.80	13.25
P75*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	6.47
P76*	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

The symmetric feature of two broad grooves is very different from that of the metazoan telomeric structure d(TAACCC), where one broad groove is very flat and the phosphate groups in the other broad groove are rotated away from the center and bend over towards each other, stabilized by the bridging water molecules between phosphate oxygens and cytosine N4 groups. Both major grooves in d(AACCCC) are very flat, differing very much from d(AACCC). In d(AACCC), both broad grooves are heavily hydrated, and the phosphate groups bend significantly in the same fashion as the one exhibited in the bending broad groove in d(TAACCC). Figure 3-3 shows flat nature of the broad groves with two stacked $C{\cdot}C^{+}$ base pairs from tetraplex 2 of d(AACCCC), together with a couple of water molecules that are within 3.3 $\AA$ from base pairs. It clearly shows two very wide and very flat grooves, and the heavy hydration which forms bridging water molecules between phosphate oxygens and cytosine N4 groups is clearly absent here.

In both tetraplexes 1 and 2, the molecules twist slowly in a right-handed manner. The average twist in both tetraplexes is 16.6^{o}, with a standard deviation of 3.4^{o}. Thus, one cytosine base pair is on average twisted 16.6^{o} relative to its covalent neighbor. This is somewhat larger than the twist value in d(C_4), which is 12.4^{o}, but smaller than that of d(AACCC), which is 20.8^{o}. Table 3.5 and Table 3.6 summarize the duplex and quadruplex parameters, respectively.

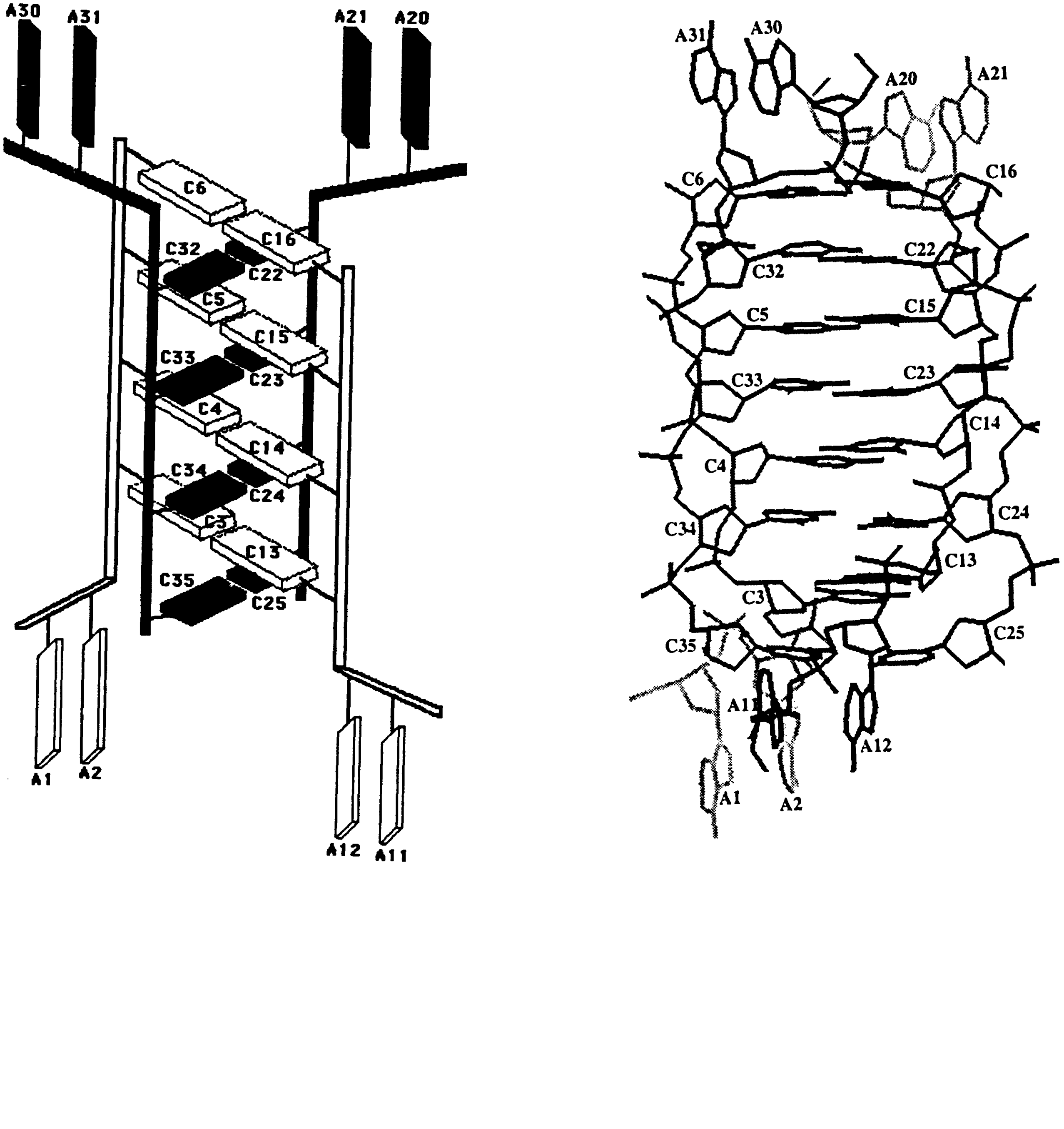

Figure 3-1: Four Strands That Form Cytosine-tetraplex 1 in the Asymmetric Unit of d(AACCCC).

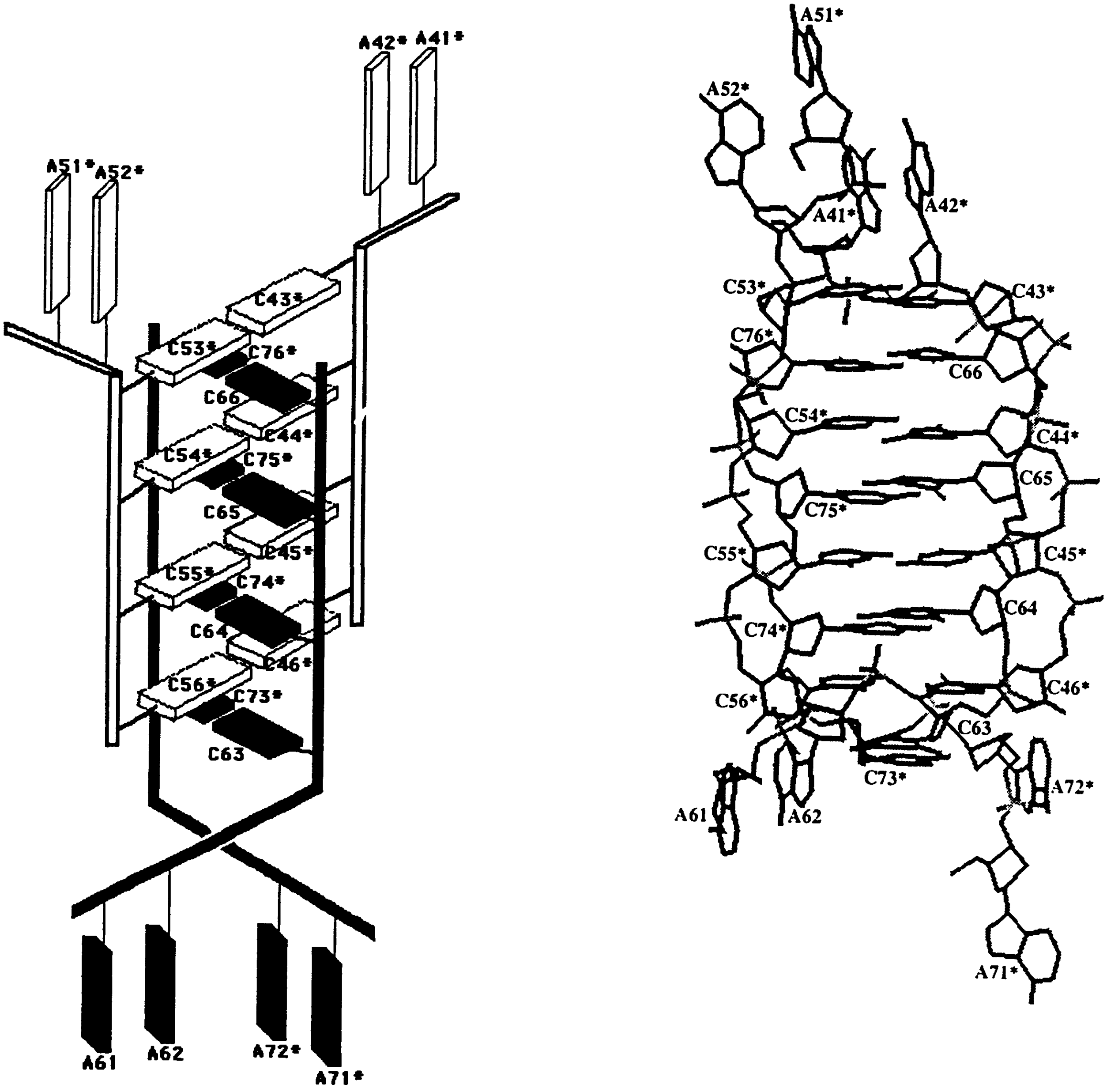

Figure 3-2: Four Strands That Form Cytosine-tetraplex 2 in the Asymmetric Unit of d(AACCCC).

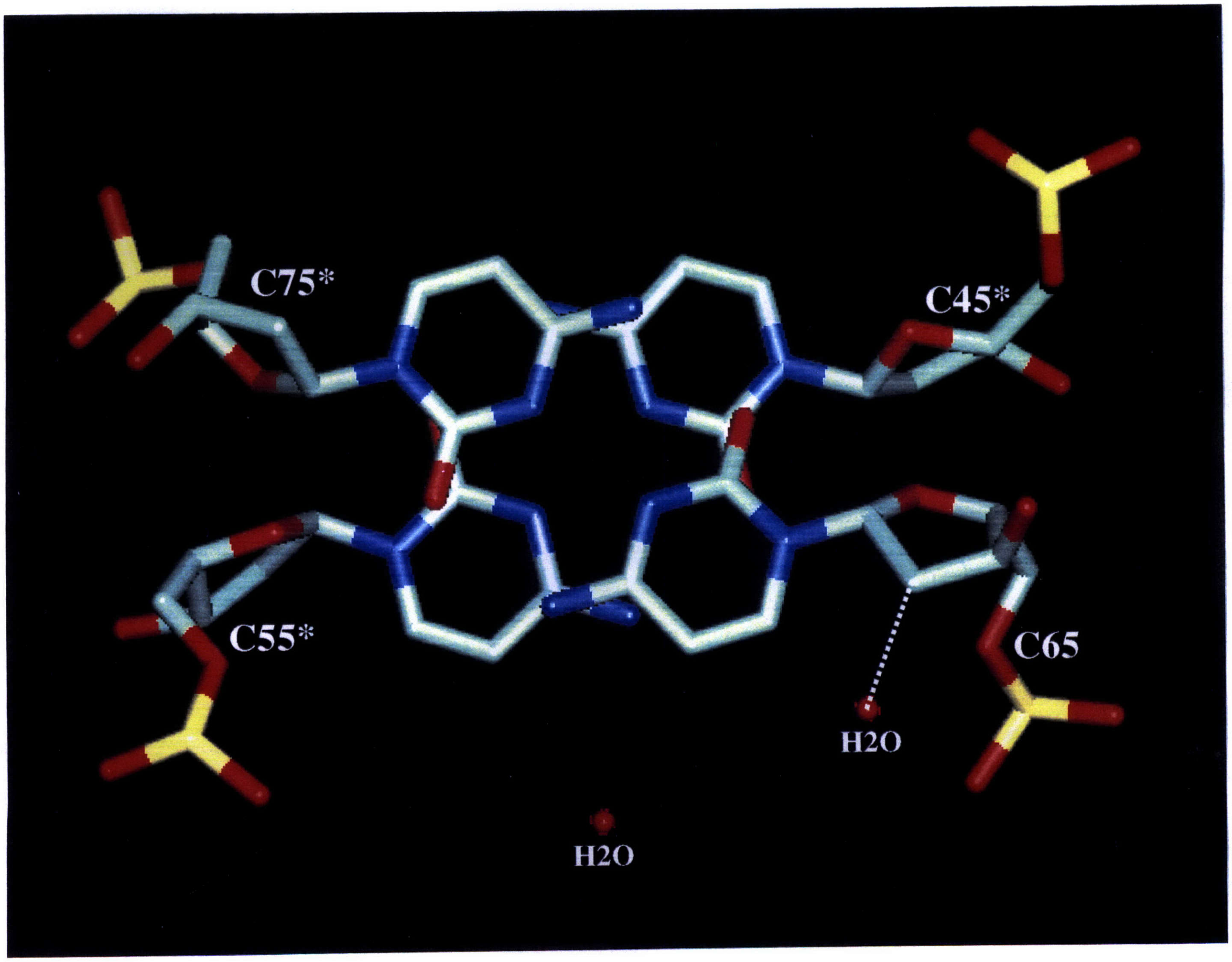

Figure 3-3: Absence of the Water Bridges.

Table 3.5: Duplex Parameters

Parallel duplex	Base step	Rise($\AA$)	Twist(^{o})
A1	C3·C13-C4·C14	6.23	20.32
	C4·C14-C5·C15	6.21	16.20
	C5·C15-C6·C16	6.43	13.49
A2	C22·C32-C23·C33	6.33	17.25
	C23·C33-C24·C34	6.13	20.50
	C24·C34-C25·C35	6.53	11.18
B1	C43*·C53*-C44*·C54*	6.47	20.15
	C44*·C54*-C45*·C55*	6.31	12.03
	C45*·C55*-C46*·C56*	6.17	13.02
B2	C66·C76*-C65·C75*	6.31	19.25
	C65·C75*-C64·C74*	6.36	9.79
	C64·C74*-C63·C73*	6.38	26.10
	A Average	6.31	16.49
	A Standard Deviation	0.03	3.40
	B Average	6.33	16.72
	B Standard Deviation	0.04	5.60
	Overall average	6.32	16.61
	Overall standard deviation	0.03	3.40

Table 3.6: Quadruplex Parameters

Quadruplex	Base step	Pseudo-Rise(Å)	Pseudo-Twist(^{o})
A	C6·C16-C22·C32	3.27	-28.86
	C22·C32-C5·C15	3.16	42.30
	C5·C15-C23·C33	3.17	-26.38
	C23·C33-C4·C14	3.04	42.34
	C4·C14-C24·C34	3.09	-21.98
	C24·C34-C3·C13	3.14	41.96
	C3·C13-C25·C35	3.39	-31.60
B	C43*·C53*-C66·C76*	3.23	38.91
	C66·C76*-C44*·C54*	3.24	-18.77
	C44*·C54*-C65·C75*	3.07	37.98
	C65·C75*-C45*·C55*	3.24	-26.56
	C45*·C55*-C64·C74*	3.12	36.32
	C64·C74*-C46*·C56*	3.05	-24.34
	C46*·C56*-C63·C73*	3.33	50.37
	Overall average	3.18	41.45/-25.50
	Overall standard deviation	0.03	4.20/4.00

Right-handed twist is positive, left-handed twist is negative.

3.3.3 Two Adenine Clusters

The adenine bases, as shown in Figure 3-1 and Figure 3-2, project away from the direction of the cytosine-tetraplex which is formed by the strands that the adenine residues are on. The A bases form base pairs with A-bases of neighboring strands (some of them are symmetry-related), creating two different kinds of A clusters in two orthogonal directions. In Figure 3-4, the A-cluster 1 is running in the z direction. There are total of four $A{\cdot}A$ base pairs in the cluster, stacking on top of each other with an average stacking distance of 3.5 $\AA$. The increase in stacking distance from 3.2 $\AA$ in cytosine-cytosine base-pair-stacking, to 3.5 $\AA$ when $A{\cdot}A$ base pairs are involved, is due to the involvement of aromatic rings in stacking. When stacking interactions involve the aromatic rings, such as the case in B-DNA, the stacking distance is generally 3.4-3.5 $\AA$. These four base pairs are in turn sandwiched between two symmetry-related cytosine tetraplexes running in the z-direction from the top and the bottom, effectively forming a continuous stacking along z. Another adenine cluster, A-cluster 2, also made up of eight adenine residues, is running along the x direction (Figure 3-5). There are only three $A{\cdot}A$ base pairs, though, stacking along the x-direction. The other two adenine bases, A52 and A71, stack upon each other. These two bases form a 38^o angle with the rest of the paired adenines in the cluster. Similarly, the three stacking $A{\cdot}A$ base pairs in this cluster are also sandwiched between two symmetry-related cytosine tetraplexes in the x-direction, and this alternating C-tetraplex-A-cluster stacking pattern creates continuous stacking along x-axis. Thus, a rather interesting three dimensional network is formed (see Figure 3-6), connected by the bridging adenine residues.

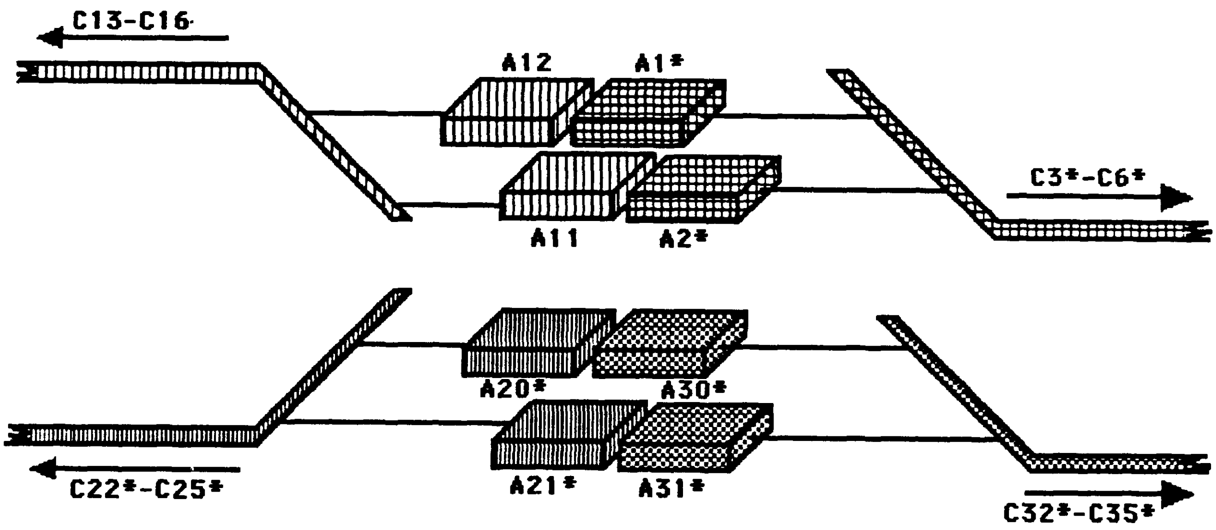

Figure 3-4(a): the schematic drawing of A-cluster 1.

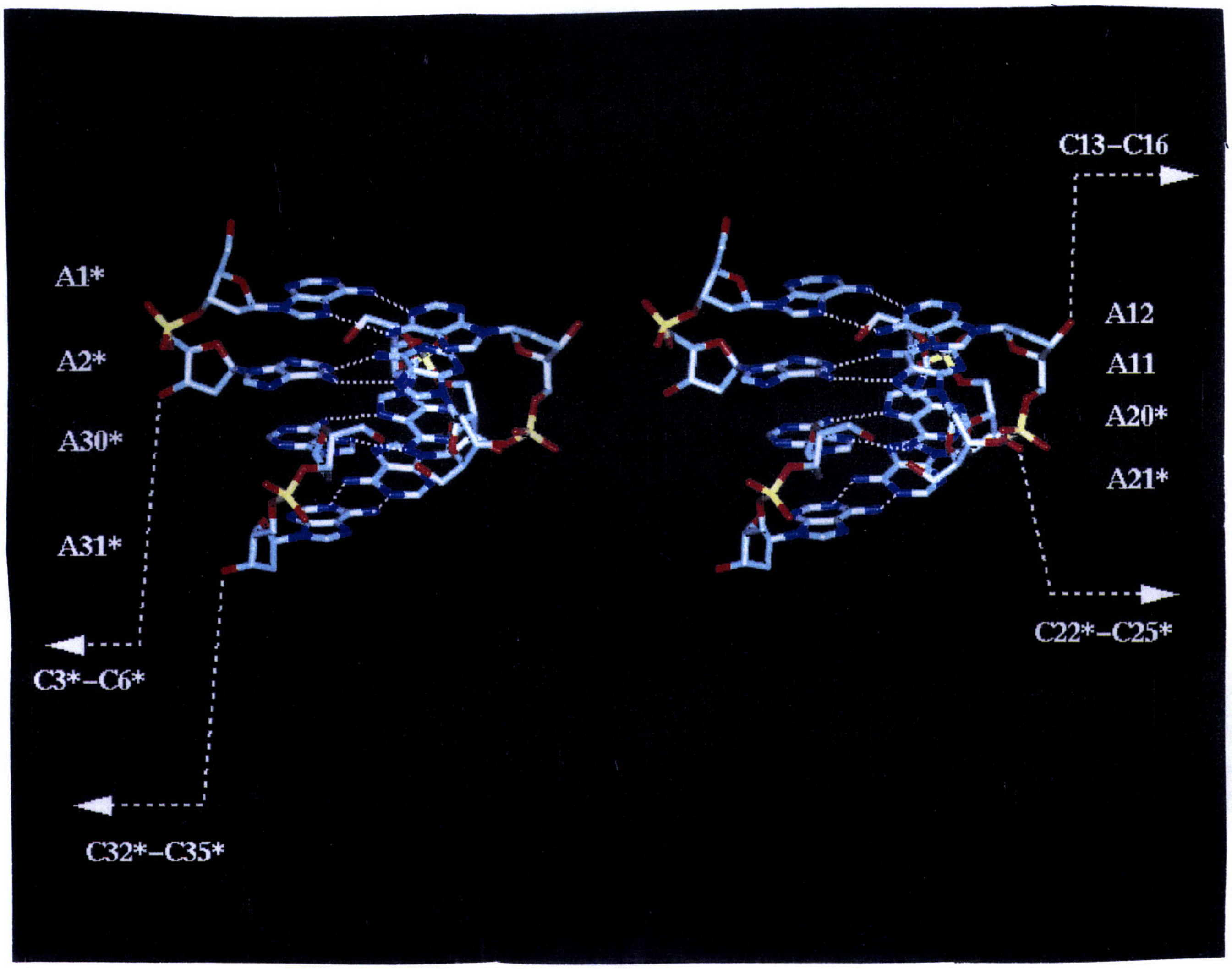

Figure 3-4(b): the stereo view of A-cluster 1.

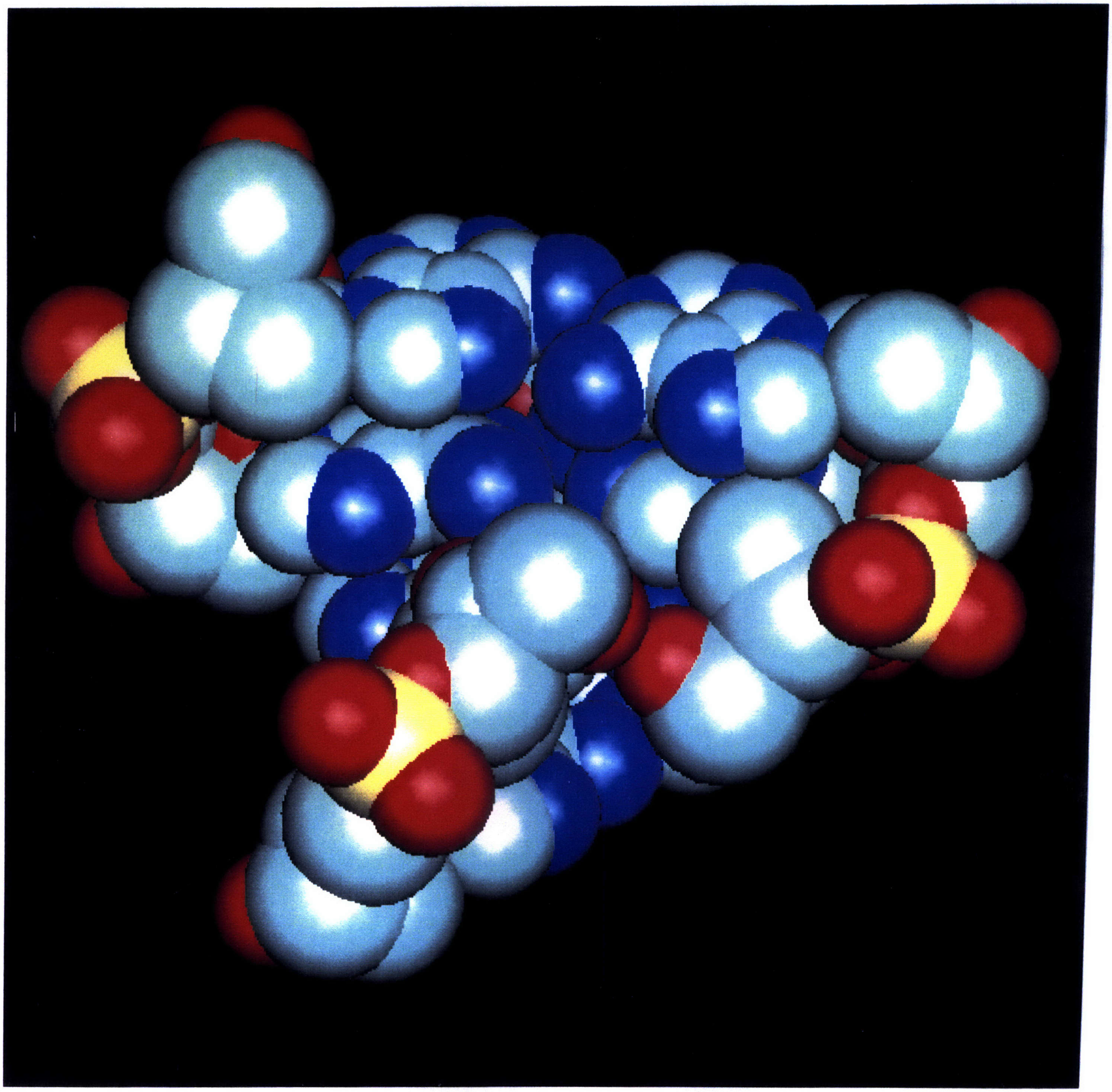

Figure 3-4(c): the Van der Waals representation of A-cluster1.

Figure 3-4: Adenine-cluster 1.

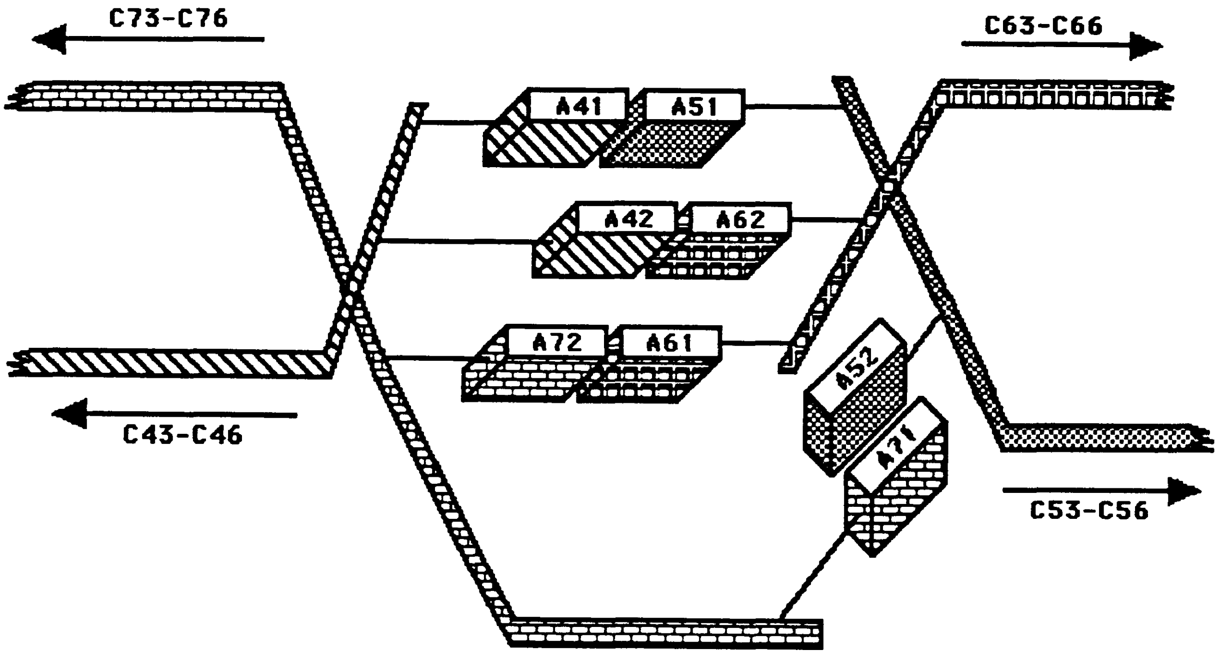

Figure 3-5(a): the schematic drawing of A-cluster 2.

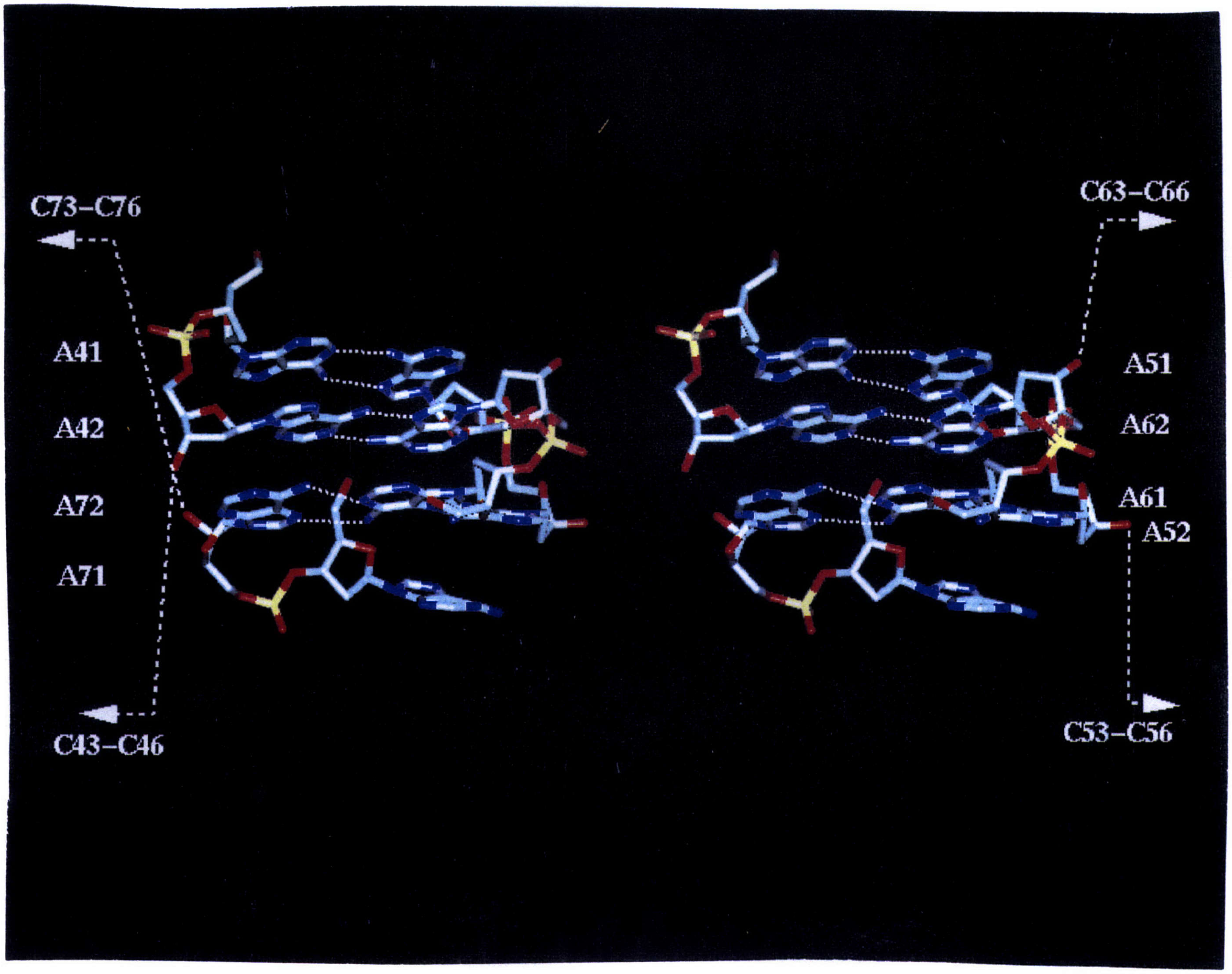

Figure 3-5(b): the stereo view of A-cluster 2.

3.3.4 Three Modes of $A{\cdot}A$ Base Pairing

Close inspection of Figure 3-4 and Figure 3-5 reveals that the polarities of these seven base pairs are not the same. In Figure 3-4, base pairs $A1^* - A12^*$ and $A2^* - A11^*$ are antiparallel, while base pairs $A20^* - A30^*$ and $A21^* - A31^*$ are parallel. In Figure 3-5, base pairs A51-A41 and A61-A72 are parallel, while A62-A42 is antiparallel. Out of the four parallel $A{\cdot}A$ base pairs, there exist all three possible different $A{\cdot}A$ base pairing modes. Figure 3-7 shows the base pair of $A20^* - A30^*$. It is a symmetric $A{\cdot}A$ N7-amino base pairing. Figure 3-8 is the base pair of $A21^* - A31^*$, a symmetric $A{\cdot}A$ N1-amino base pairing mode. Base pairs of A51-A41 and A61-A72, adopt another paring mode, which is asymmetric N1-amino, N7-amino, as shown in Figure 3-9. All the antiparallel $A{\cdot}A$ base pairs adopt the asymmetric N1-amino, N7-amino base pairing mode, as illustrated in Figure 3-10.

The glycosyl conformations in the structure are anti for all the cytosine residues, which is in agreement with previously solved d(C_4) and d(AACCC). Out of the sixteen adenine residues, eight are anti, five are syn, and three have almost clinal conformations. Several modes of sugar puckers are present in the structure. For cytosines, the most occurring one is C4'-exo, appearing 9 times, followed by C2'-endo and C3'-endo, appearing 6 times apiece. For adenines, the most frequent sugar puckers are C3'-exo and C2'-endo. The torsion angles and sugar puckers are summarized in Table 3.7.

3.3.5 Comparison with Previous Results

Compared with the previously solved cytosine-rich structures, d(AACCCC) reveals many interesting features. First, I-motif, the four-stranded, intercalated cytosine segment is still a predominant feature of the structure. It is an extremely stable feature and it is interesting to note that the crystals were grown over a wide range of pH, ranging from pH 5.0 to pH 8.0. The formation of $C{\cdot}C^+$ base pairs depends on hemi-protonation of the cytosines [1, 2, 30, 13]. In poly d(C), the hemi-protonated structure was stable up to pH 7. The fact that crystals of d(AACCCC) can grow at pH 7.5 and pH 8.0 indicates that the stable nature of the tetraplex and the packing

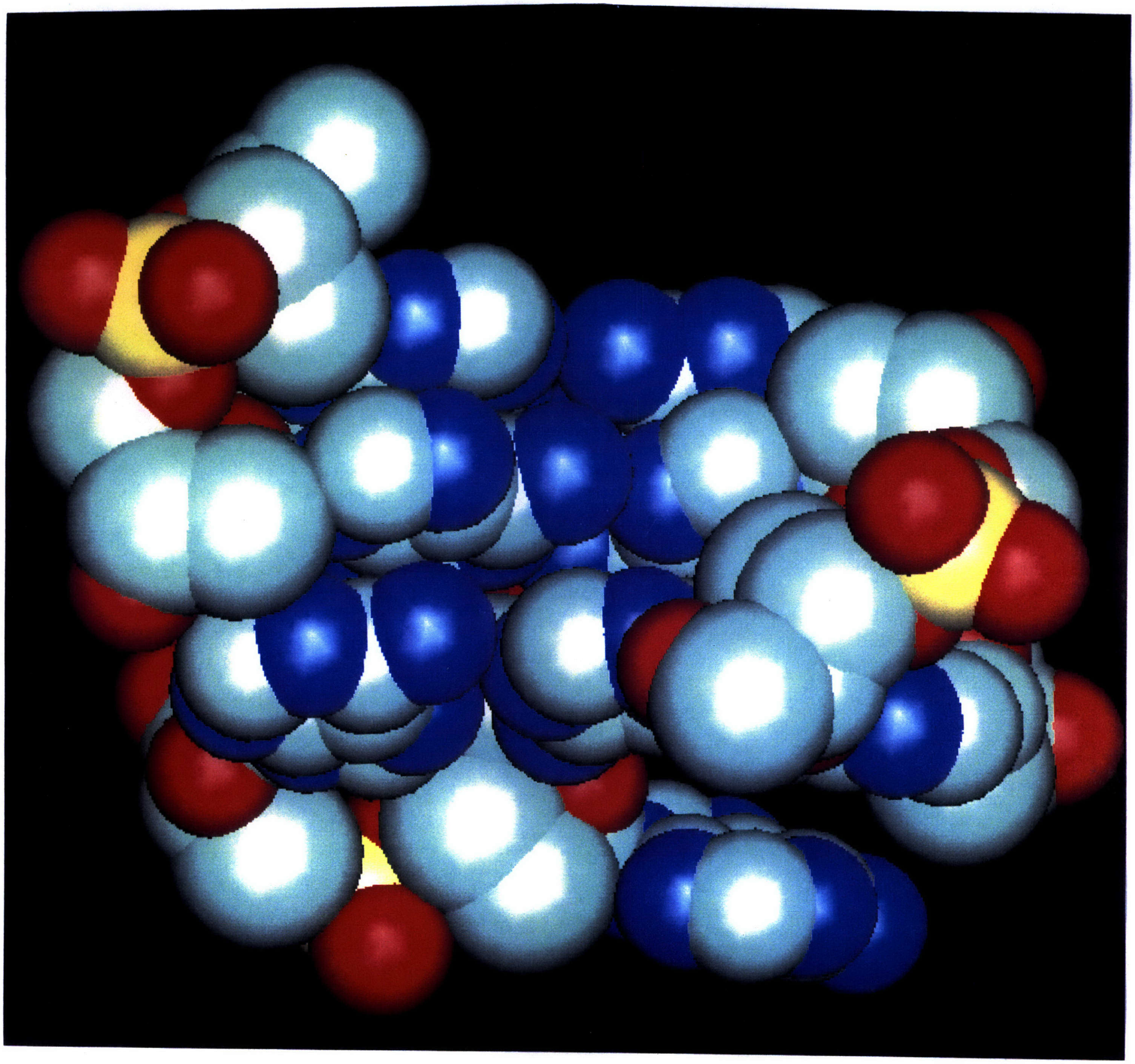

Figure 3-5(c): the Van der Waals representation of A-cluster2.

Figure 3-5: Adenine-cluster 2.

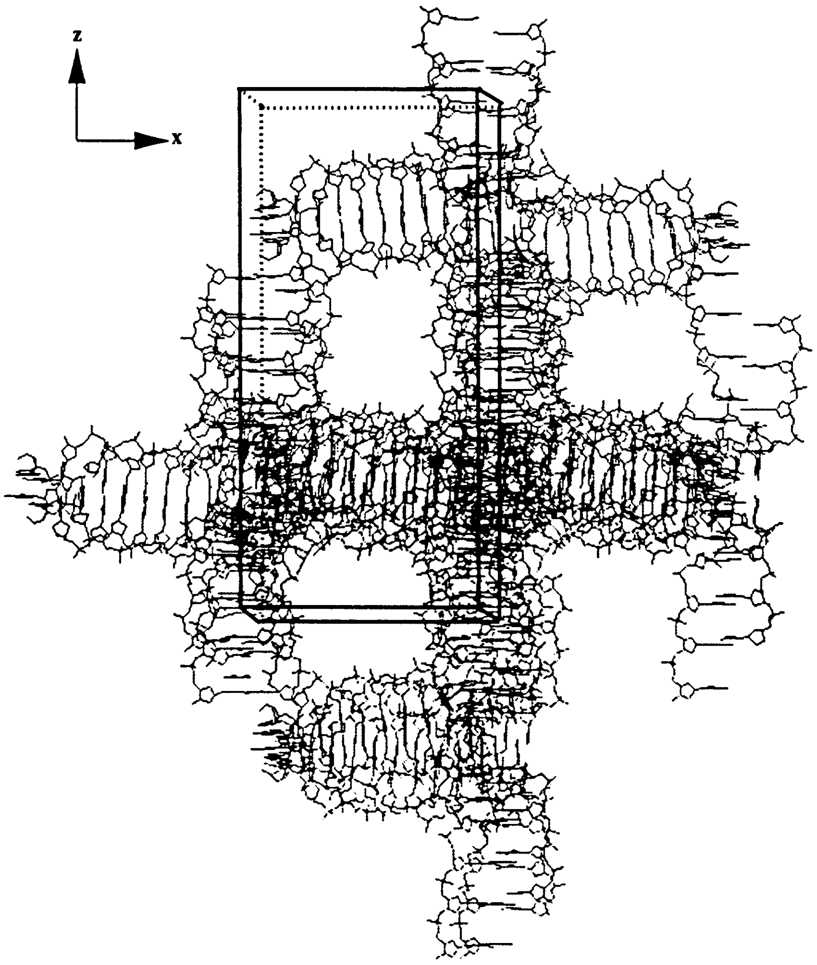

Figure 3-6: Crystal Packing of d(AACCCC): a 3-D Network.

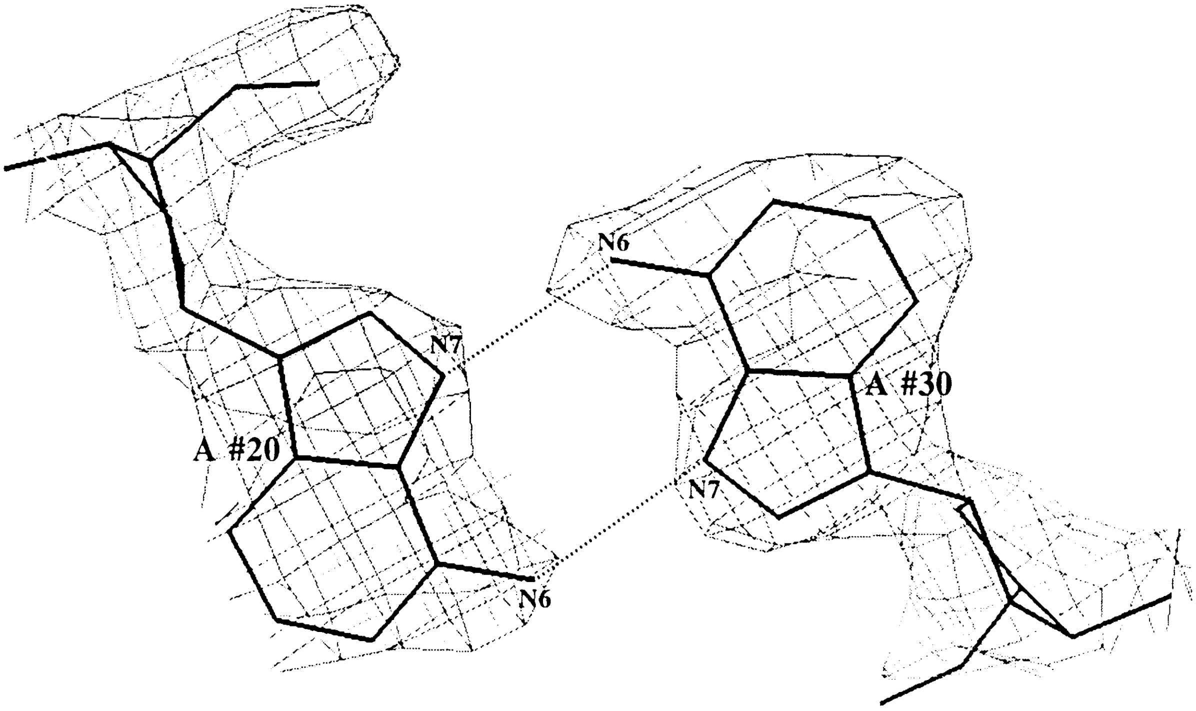

Figure 3-7: Base Pairs of Parallel A20*-A30*.

forces raised the pK for hemi-protonation to a higher value. This reinforces the possibility that the tetrahymena telomere may adopt I-motif *in vivo* at physiological pH, possibly at the presence of its binding proteins.

Besides the general structural similarity in I-motif, we have found many variations. One notable difference is the presence of two different conformations of cytosine tetraplexes, first observed in d(AACCC). In one tetraplex, like other previously reported C-tetraplexes, the utmost base pairs are from the 5' end of each strand; while in the other, the utmost base pairs are from the 3' end of each strand. It suggests that the two conformations are comparably favorable energetically, leaving open the possibility that the telomere sequences might adopt either one of the two conformations, depending on the conformation of the non-cytosine residues.

Despite the apparent rigidity of the cytosine tetraplex conformations, each individual strand varies considerably from structure to structure. The average twists between covalently linked cytosines vary from 12.4^o of d(C_4) to 16.6^o of d(AACCCC), and to 20.8^o of d(AACCC). When the common I-motif portion of the structures are

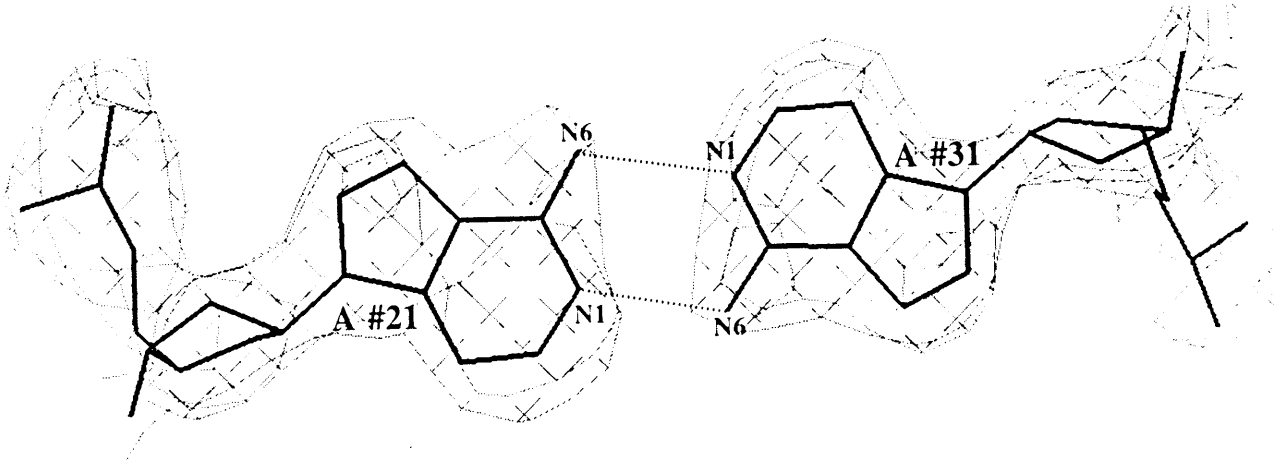

Figure 3-8: Base Pairs of Parallel A21*-A31*.

superimposed, the r.m.s. differences are quite considerable, especially among the sugar-phosphate backbones. Table 3.8 summarizes the r.m.s. differences between tetraplexes in d(AACCCC) and those from d(CCCC) and d(AACCC). In all cases, the tetraplexes exhibit considerable differences from structure to structure, and the differences are mainly due to those of sugar-phosphate backbones. The cytosine bases are very stable and almost superimposable. This is not all that surprising, given the less flexible nature of the $C{\cdot}C^{+}$ base pairing due to three strong planar hydrogen bonds. In contrast, the sugar-phosphate backbones are intrinsically more flexible, partly due to their lack of torsional restraints and partly due to strong electrostatic repulsion between phosphate group in the narrow grooves. These flexible aspects of the cytosine tetraplex might be important for telomere sequences to adopt different conformations at different biological conditions.

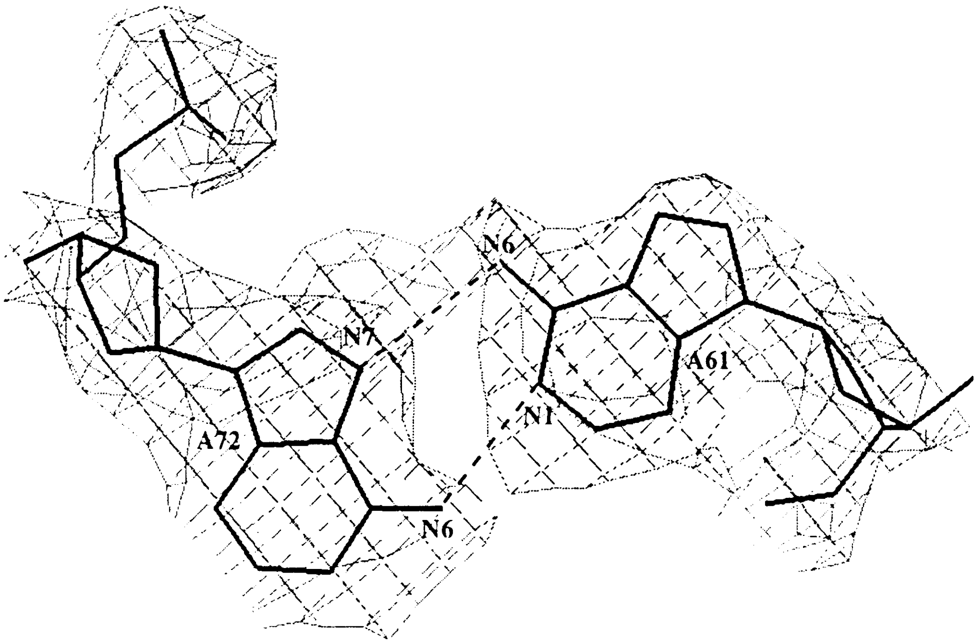

Figure 3-9: Base Pairs of Parallel A61-A72.

3.3.6 The Bridging of A-clusters

If the cytosine tetraplex reveals variations among different structures, the major variable, however, occurs at the non-cytosine part of the structure. Unlike the other telomeric sequence solved, namely the metazoan telomere d(TAACCC) [27], where the adenine/thymine segment of the structure folds back on itself to form a stable loop, the adenines in this structure and d(AACCC) adopt an entirely different conformation. In this case, the adenines, adopting three different kinds of $A{\cdot}A$ base pairs, are the central lattice building block. There are two adenine residues per strand. In cytosine tetraplex 1, which points in x-direction, there exist four pairs of adenine residues. As seen in Fig. 1, these four pairs project away from the central C-tetraplex, and the planes of the bases are perpendicular to the z-axis. Each pair, along with six other symmetry-related adenine residues, form an A-cluster stacking along z, connecting two symmetry-related cytosine tetraplex 2, thus forming continuous stacking along the z-direction, as illustrated in Figure 3-11. Therefore, the original

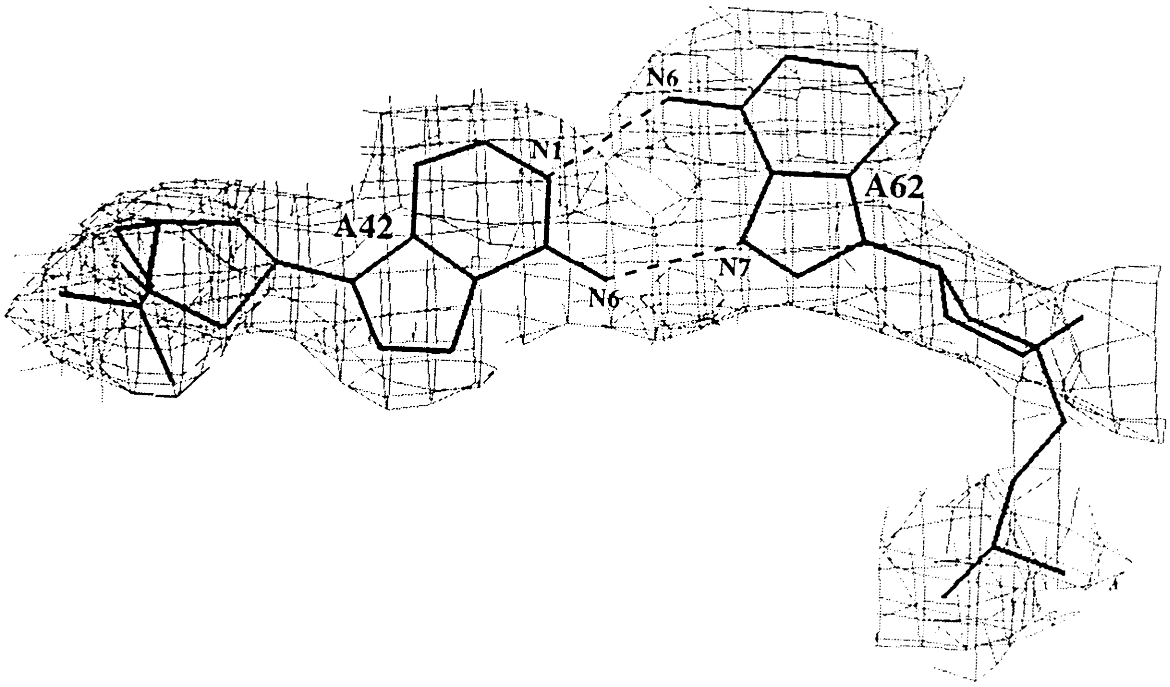

Figure 3-10: Base Pairs of Antiparallel A42-A62.

Therefore, the original four pairs of adenine residues from tetraplex 1, are involved in four A-clusters at different locations (the four A-clusters involved are symmetry-related), creating four continuous columns of z stacking. In a similar manner, the eight adenine residues from the strands forming cytosine tetraplex 2, join C-tetraplex 1 in x-direction, forming four continuous stacking columns in x-direction, as seen in Figure 3-12. In contrast to the rather rigid cytosine tetraplex, the non-cytosine part of the telomeric sequences clearly shows a great deal of variability and versatility of forming different structural conformations.

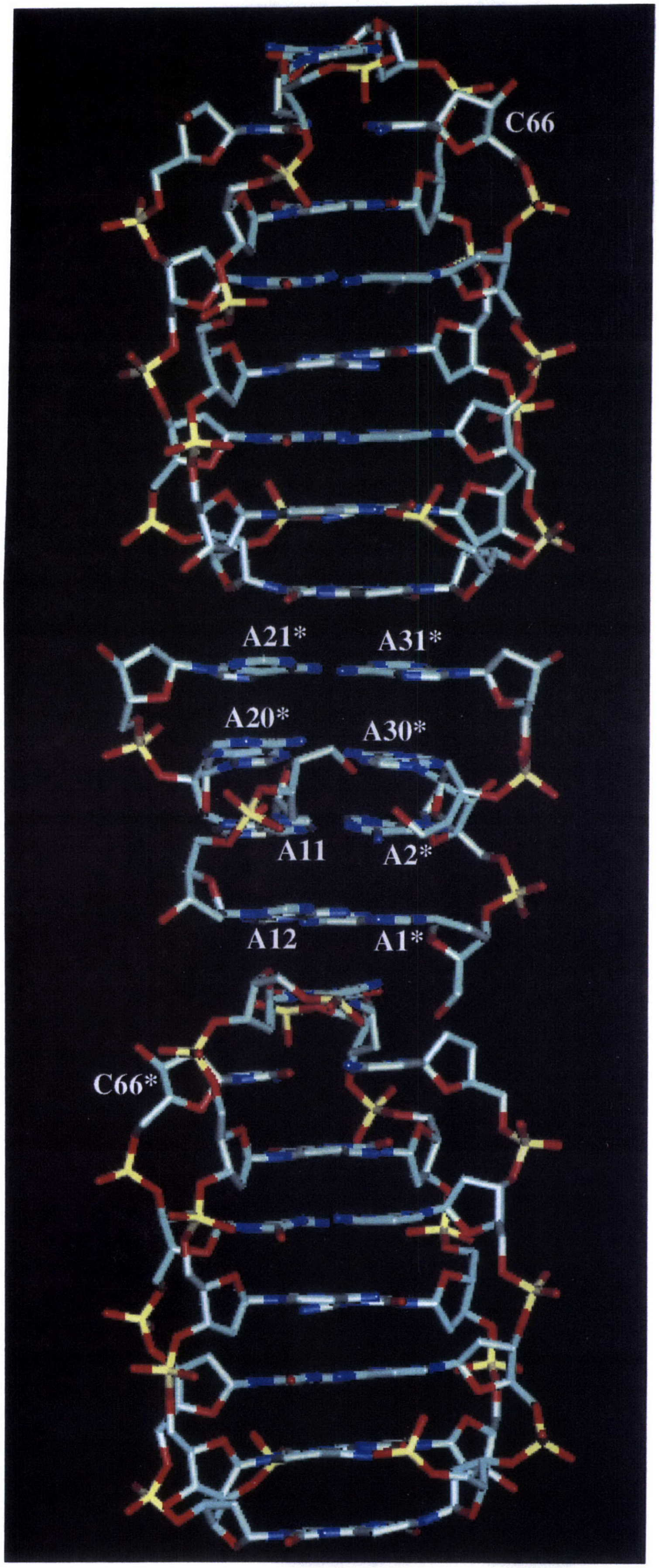

Figure 3-11: Stacking of A-cluster 1 and C-tetraplexes 2.

Table 3.7: Glycosidic Torsion Angles And Sugar Puckers of d(AACCCC)

Strand	Residue	α	β	γ	δ	ϵ	ζ	χ	Pseud	Pucker
a1	A1	n.a.	n.a.	193.4	170.2	232.7	192.4	316.4	192.9	C3'-exo
	A2	277.3	142.7	46.8	70.7	81.2	299.0	272.5	15.7	C3'-endo
	C3	258.7	268.3	158.8	74.5	243.2	286.9	185.0	71.3	C4'-exo
	C4	101.5	182.0	208.6	137.1	228.7	241.6	265.4	165.8	C2'-endo
	C5	278.7	161.7	54.5	82.2	165.7	286.2	255.8	84.6	O4'-endo
	C6	291.8	194.8	54.3	80.9	n.a.	n.a.	253.2	50.0	C4'-exo
a2	A11	n.a.	n.a.	65.5	142.3	198.4	263.4	252.1	160.2	C2'-endo
	A12	291.9	173.0	55.3	133.9	280.8	286.5	282.6	146.6	C2'-endo
	C13	303.8	178.4	53.0	76.5	225.1	214.5	214.9	28.9	C3'-endo
	C14	145.8	260.6	170.8	74.5	246.2	251.6	217.0	32.4	C3'-endo
	C15	316.4	140.0	74.5	71.8	231.6	285.3	219.5	48.1	C4'-exo
	C16	270.7	74.3	185.9	150.8	n.a.	n.a.	229.3	183.9	C3'-exo
a3	A20	n.a.	n.a.	57.1	148.3	181.5	270.7	261.5	182.3	C3'-exo
	A21	298.7	172.5	54.1	126.3	267.8	292.9	265.1	127.0	C1'-exo
	C22	296.7	199.6	71.4	144.5	253.7	295.2	237.5	157.5	C2'-endo
	C23	251.5	85.0	168.1	84.2	217.0	293.2	217.1	82.6	O4'-endo
	C24	177.4	136.6	166.4	81.0	169.3	295.6	249.0	22.7	C3'-endo
	C25	165.4	201.7	165.3	95.2	n.a.	n.a.	241.8	89.4	O4'-endo
a4	A30	n.a.	n.a.	64.9	134.1	175.2	280.0	249.9	155.2	C2'-endo
	A31	129.8	170.4	191.8	149.7	279.7	299.5	272.5	182.9	C3'-exo
	C32	297.6	189.5	65.5	144.7	260.7	291.6	231.9	147.8	C2'-endo
	C33	318.0	278.4	235.0	143.1	215.6	237.1	244.3	224.5	C4'-endo
	C34	301.4	151.2	60.5	72.1	216.8	279.0	228.8	52.7	C4'-exo
	C35	281.1	92.0	182.0	134.3	n.a.	n.a.	219.7	167.4	C2'-endo
b1	A41	n.a.	n.a.	173.0	163.1	181.2	276.8	90.3	174.1	C2'-endo
	A42	275.7	168.8	70.6	124.0	252.0	279.3	295.2	128.0	C1'-exo
	C43	301.4	175.4	51.4	145.2	219.0	298.4	236.1	163.4	C2'-endo
	C44	140.6	146.9	175.6	88.7	189.2	256.4	246.2	50.7	C4'-exo
	C45	162.9	229.3	176.1	76.8	203.1	257.2	226.0	76.0	O4'-endo
	C46	165.2	214.2	174.1	70.4	n.a.	n.a.	231.8	61.1	C4'-exo
b2	A51	n.a.	n.a.	327.3	140.2	261.7	141.4	63.7	205.3	C3'-exo
	A52	283.1	304.9	207.8	153.6	266.9	297.4	46.7	177.4	C2'-endo
	C53	282.2	210.1	63.3	56.9	58.1	64.4	206.2	16.5	C3'-endo
	C54	216.9	208.5	68.0	89.2	179.9	285.6	240.6	32.9	C3'-endo
	C55	288.1	200.1	61.2	81.4	170.8	274.7	246.7	44.0	C4'-exo
	C56	178.2	195.4	168.5	83.4	n.a.	n.a.	226.1	19.5	C3'-endo
b3	A61	n.a.	n.a.	59.4	161.1	208.6	263.2	76.2	186.0	C3'-exo
	A62	292.1	143.4	51.4	107.4	251.7	277.6	284.8	113.5	C1'-exo
	C63	307.3	170.2	62.3	148.9	232.2	309.5	199.6	190.1	C3'-exo
	C64	145.0	174.0	160.1	86.9	204.8	270.6	237.8	40.4	C4'-exo
	C65	299.6	185.2	57.6	139.4	218.0	208.4	283.7	152.9	C2'-endo
	C66	304.9	142.1	55.0	71.5	n.a.	n.a.	234.8	45.5	C4'-exo
b4	A71	n.a.	n.a.	190.6	83.4	152.0	146.7	76.6	101.4	O4'-endo
	A72	291.5	144.5	317.0	166.1	270.3	276.5	314.0	196.6	C3'-exo
	C73	292.5	205.5	64.4	151.8	266.0	275.8	199.1	191.9	C3'-exo
	C74	294.6	143.5	52.3	116.2	194.8	287.3	252.7	119.7	C1'-exo
	C75	193.9	76.0	181.2	150.2	198.5	240.9	246.4	205.8	C3'-exo
	C76	169.7	214.2	179.2	115.8	n.a.	n.a.	242.5	127.6	C1'-exo

Backbone torsion angles for the bonds in the backbone P-O5'-C5'-C4'-C3'-O3'-P are α, β, γ, δ, ϵ and ζ, respectively, and the glycosidic angle is χ. (a) is calculated by the program NEWHEL93.

Table 3.8: r.m.s. Differences Between Cytosine Tetraplexes in d(A_2C_4) and Those From d(C_4) and d(A_2C_3)

r.m.s. difference	Tetraplex 1 in d(A_2C_4) vs Tetraplex 1 in d(C_4)	Tetraplex 1 in d(A_2C_4) vs Tetraplex 1 in d(A_2C_3)	Tetraplex 2 in d(A_2C_4) vs Tetraplex 2 in d(A_2C_3)
Tetraplex (Å)	1.26	1.47	1.22
Base (Å)	0.45	0.40	0.38
Backbone (Å)	1.55	1.88	1.55

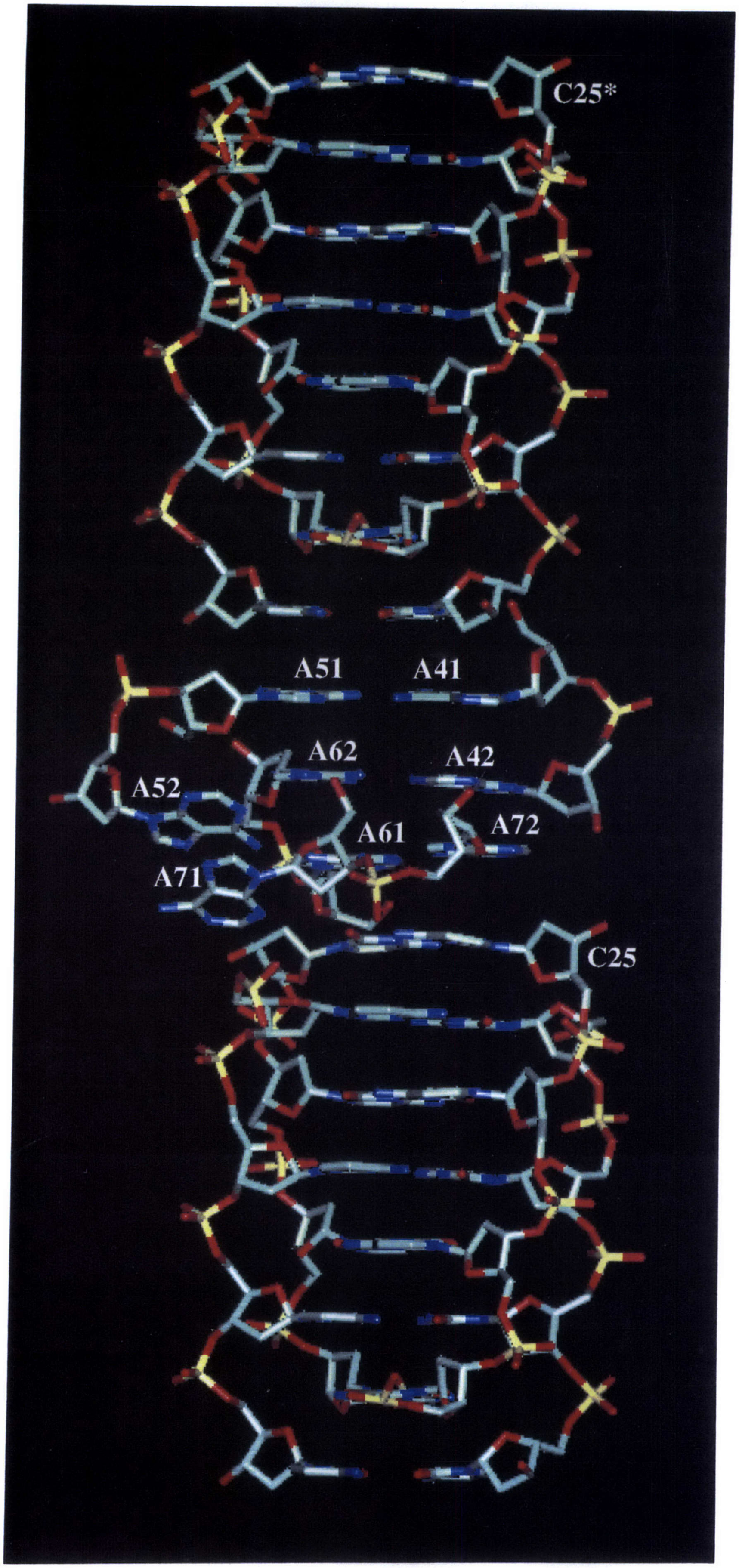

Figure 3-12: Stacking of A-cluster 2 and C-tetraplexes 1.

3.4 DISCUSSION

Sequences containing stretches of cytosines and adenines are found in telomeres [16] and also occur in segments scattered throughout the genome. They may also exist in large RNAs such as group I and group II introns, and ribosomal and spliceosomal RNAs. Recent crystal structure of the P4-P6 domain of the *Tetrahymena thermophila* intron [31, 32] revealed adenosine platforms which contribute to key components of the domain tertiary structure. The crystal structure of *Tetrahymena* telomeric sequence d(AACCCC) shows two different novel adenine-clusters that play a key role in building the crystal lattice and stabilizing the structure. The abundance of adenosine residues in internal loops of many RNAs and the ability of A-clusters observed in this structure to form stabilized tertiary structures suggests the possibility that A-clusters, like the adenosine platforms observed in a group I intron fragment, could be a motif present in large RNAs to facilitate folding and long-range tertiary interactions.

This crystal structure shows that, *in vitro*, the telomeric sequence adopts a very different structural conformation from the standard B-DNA. Does this structural conformation occur *in vivo?* We do not have the answer yet. The fact that both C-rich sequences and its complement G-rich sequences could form tetraplexes makes it possible that the two structures could act in concert and/or one would promote the formation of the other. It may play an important role in DNA self recognition, which is essential in many biological activities.

Figure Legends

Figure 3-1. Cytosine tetraplex 1 of the structure d(AACCCC). (Left) A schematic diagram illustrating the overall configuration of tetraplex 1. Two strands that are parallel and form hydrogen bonds between their cytosine bases are colored black, while the other two are colored white. (Right)View of the tetraplex 1 through its major groove. The major groove is wide and open. The center of the molecule is composed of intercalating cytosine residues held together by $C{\cdot}C^+$ base pairs. Note that there are two adenine residues at the 5' end of each strand, and they are projecting away from the center of the molecule. The outermost $C{\cdot}C^+$ base pairs of the tetraplex are from the 3' end of each strand.

Figure 3-2. cytosine tetraplex 2 of the structure d(AACCCC). (Left) A schematic diagram illustrating the overall configuration of tetraplex 2. Two strands that are parallel and form hydrogen bonds between cytosine bases are colored black, while the other two are colored white. Residues with asterisks represent symmetry related residues (Equivalently, we could have chosen the asymmetric unit in such a way that four strands in the asymmetric unit would form tetraplex 2). (Right) View of the tetraplex 2 through its major groove. The intercalating motif here is very similar to that of tetraplex 1. However, the outermost $C{\cdot}C^+$ base pairs of the tetraplex are from the 5' end of each strand.

Figure 3-3. Two adjacent layers of $C{\cdot}C^+$ base pairs from tetraplex 2 along with two water molecules that are within $3.5\AA$ from the base pairs. We are looking down the axis of the molecule, which is the z-axis. Unlike structures such as d(AACCC) and d(TAACCC), the broad grooves of this structure are essentially flat, and the phosphates are not bent over. There is also no water molecule bridging the cytosine N4 amino group with the phosphate oxygens on the opposite side of the groove. The absence of this feature shows the variance of cytosine tetraplexes.

Figure 3-4. A-cluster 1 of d(AACCCC). (a) A schematic diagram illustrating the formation of A-cluster 1 and its relation with the rest of the cytosine residues of the strands. There are two parallel $A{\cdot}A$ base pairs, namely, A20*-A30* and A21*-A31*. The other two $A{\cdot}A$ base pairs, A2*-A11 and A1*-A12, are antiparallel. Every cytosine portion of four strands combines with three other symmetry related cytosines strands (not shown) to form tetraplex 1. Therefore, there are four cytosine tetraplexes 1 connected by A-cluster 1. (b)View of A-cluseter 1 connecting four cytosine strands which belong to four different cytosine tetraplexes. It consists of four stacking $A{\cdot}A$ base pairs. It stacks on two cytosine tetraplexes 2 (not shown), from top and bottom, forming continuous stacking along z. (c) Van der Waals representation of A-cluster 1.

Figure 3-5. A-cluster 2 of d(AACCCC). (a) A schematic diagram of A-cluster 2. Note there are only three stacking base pairs. The other two bases form stacking of their own, which is tilted about 38^o with the other base pairs. Like A-cluster 1, A-cluster 2 connects four cytosine tetraplexes 2. Of the three $A{\cdot}A$ base pairs, A61-A72 and A41-A51 are parallel. A42-A62 is antiparallel. (b) View of A-cluster 2 connecting four cytosine strands which belong to four different cytosine tetraplexes. It consists of three stacking $A{\cdot}A$ base pairs. It stacks on two cytosine tetraplexes 1 (not shown), from top and bottom, forming continuous stacking along x. (c) Van der Waals representation of A-cluster 2.

Figure 3-6. View of the three dimensinal network formed by continuous stacking along x and z axes. The box shown is the unit cell of the crystal.

Figure 3-7. Base pairs of parallel A20*-A30*. It is a symmetric $A{\cdot}A$ N7-amino base paring.

Figure 3-8. Base pairs of parallel A21*-A31*. It is a symmetric $A{\cdot}A$ N1-amino base paring.

Figure 3-9. Base pairs of parallel A61-A72. It is an asymmetric $A{\cdot}A$ N1-amino N7-amino base paring.

Figure 3-10. Base pairs of antiparallel A42-A62. It is an asymmetric $A{\cdot}A$ N1-amino N7-amino base paring.

Figure 3-11. A-cluster 1 stacks on two symmetry-related cytosine tetraplexes 2, from top and bottom, creating continuous stacking in z direction.

Figure 3-12. A-cluster 2 stacks on two symmetry-related cytosine tetraplexes 1, from top and bottom, creating continuous stacking in x direction.

Appendix A

Telomere: An Introduction

The scope of this thesis is to study, structurally, the hereditary material, deoxyribonucleic acid (DNA), and more specifically, the structure of cytosine-rich telomere's DNA sequences. This topic presents two interesting and also related aspects of the problem. First, we want to know the structure of the biologically very important telomeric DNA sequences. Second, because of this unique stretch of sequence repeats at the end of the chromosome, coupled with our prior knowledge of unconventional features of cytosine-rich structures, we anticipated to observe a novel alternative conformation of DNA. Here in the following pages, I will give an introduction to the pertinent biological and structural background of the problem, so that one can appreciate how this study here fits in a broader picture.

A.1 TELOMERES CARRY OUT THREE IMPORTANT FUNCTIONS

The DNA of telomeres–the terminal DNA-protein complexes of chromosomes– differs significantly from other DNA sequences in both structure and function. It has long been established that telomeres are essential for chromosome stability and thus cell viability [19, 20]. Recent work has shown its remarkable mode of synthesis by ribonucleoprotein reverse transcriptase, telomerase [19, 20, 15, 33], which allows the complete replication of linear chromosomal DNA without the loss of terminal bases at the 5' end of each strand. It has also been shown the ability of telomeres to aid in gene-regulation and possibly serving as a "mitotic clock" for the cells of higher animals. Its remarkable ability to form unusual structures *in vitro* has also been the focus of many recent works, including this one presented here.

A.1.1 Chromosome Integrity

To a living cell, few jobs are more important than protecting the integrity of its chromosomes. When these libraries of genetic information are damaged, the cell's own survival is in jeopardy, let alone its progeny's. Ever since the work of pioneering geneticists Hermann Muller and Barbara McClintock more than fifty years ago, cell biologists have believed that at least part of the job of chromosomal protection falls to their ends, namely, telomeres. Therefore, the telomeres were originally defined functionally on the basis of early cytological and genetic studies which demonstrated that chromosomes with broken ends were unstable [19, 20]. The broken ends were able to fuse end to end, leading to dicentric, ring or other unstable chromosome forms. The contrast between this observed instability and the stability of normal chromosomal ends led to the concept of the telomere as the specialized structure at the natural end of a eukaryotic chromosome, which prevents chromosomes from fusing end-to-end, and from subsequent chromosome breakage and loss as cells divide.

A.1.2 Complete Replication

Molecular analysis has revealed the mechanism that telomeres are able to carry out another crucial function: the ability to allow the end of the linear chromosomal DNA to be replicated completely without the loss of terminal bases at the 5' end of each strand of DNA. Such loss is predicted from the properties of the machinery of conventional semiconservative replication: its ability to work only in one direction (from 5' end to 3' end), and the requirement of cellular DNA polymerases for an RNA primer. The problem of complete replication of a linear DNA molecule by conventional

DNA replication is illustrated in Figure A-1. Part a of the figure shows a DNA duplex, whose end is on the right, with 5' and 3' ends of each strand indicated. Part b of the figure shows the parental DNA is copied by a replication fork moving from the left toward the end of the molecule. Leading strand 5'-to-3' synthesis copies the bottom parental strand all the way to its last nucleotide. Discontinuous lagging strand synthesis, also in the 5'-to-3' direction, copying the top parental strand, is primed by RNA primers (zig-zag lines). Part c shows that when RNA primers are removed, the internal gaps are filled in by extension of the discontinuous DNA and ligation. A 5' gap in this newly synthesized strand is left because there is no primer allowing it to be filled in. Successve replication rounds will produce shortened daughter chromosomes.

What do telomeres have to do to avoid the loss of terminal bases at 5' end during replication? The answer is telomerase, which is unlike nearly all other enzymes that have been studied. It contains an essential RNA component in addition to the expected protein. Telomerase activities have been identified *in vitro* in ciliate [34, 35, 36, 37, 38, 39] and human cell-free extracts [40]. It has been shown that the telomerase is used to synthesize the G-rich strand of telomeres [19]. The telomerase RNAs of *Tetrahymena* and *Euplotes* have been identified, and contain a sequence, 5'-CAACCCCAA-3' and 5'-CAAAACCCCAAA-3', respectively, which is complementary to the telomeric repeats synthesized by the enzyme [34, 35]. Studies *in vitro* indicated that this complementary RNA sequence acts as the template for synthesis of the G-rich telomeric DNA strand [34, 35]. This fact was established by site-directed mutagenesis. Expression of the mutated telomerase RNA gene in transformed *Tetrahymena* cells results in the synthesis of telomeres whose sequence corresponds to the mutated template sequence [41]. Telomerase can thus be defined as an unusual ribonucleoprotein reverse transcriptase whose RNA template is an intrinsic part of the enzyme. Figure A-2 shows the current model [34, 35] for the mechanism of telomere DNA synthesis by telomerase. It illustrates that telomerase RNA binds to the DNA strand, where it serves as a template for addition of the telomeric repeats, sliding along the strand it is elongating. Another cell polymerase

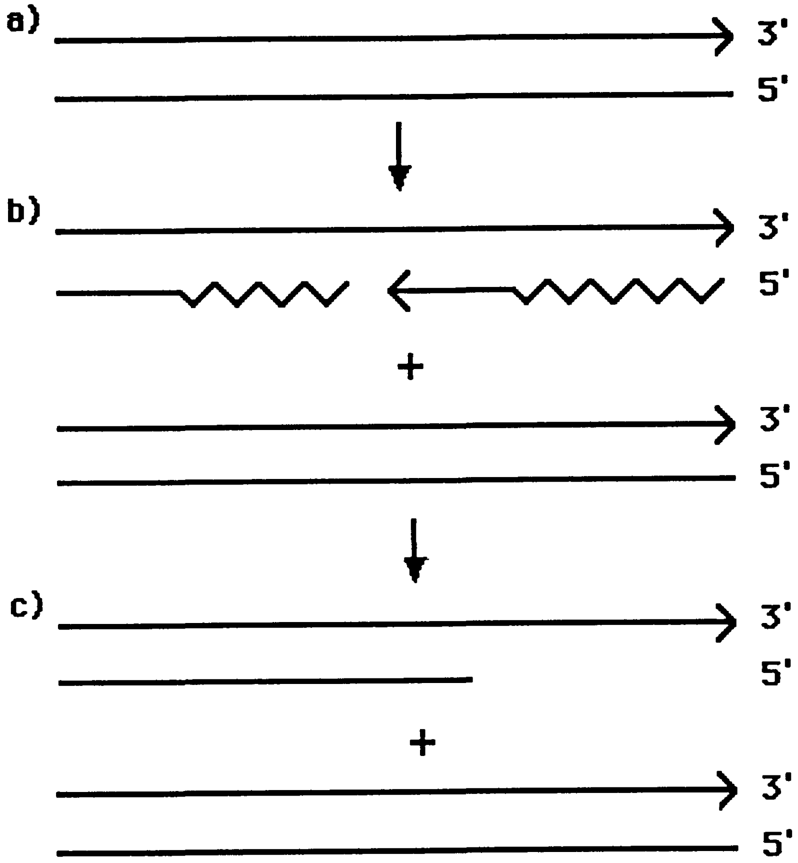

Figure A-1: The Loss of Terminal Bases at 5′ End of DNA as A Result of Conventional DNA Replication.

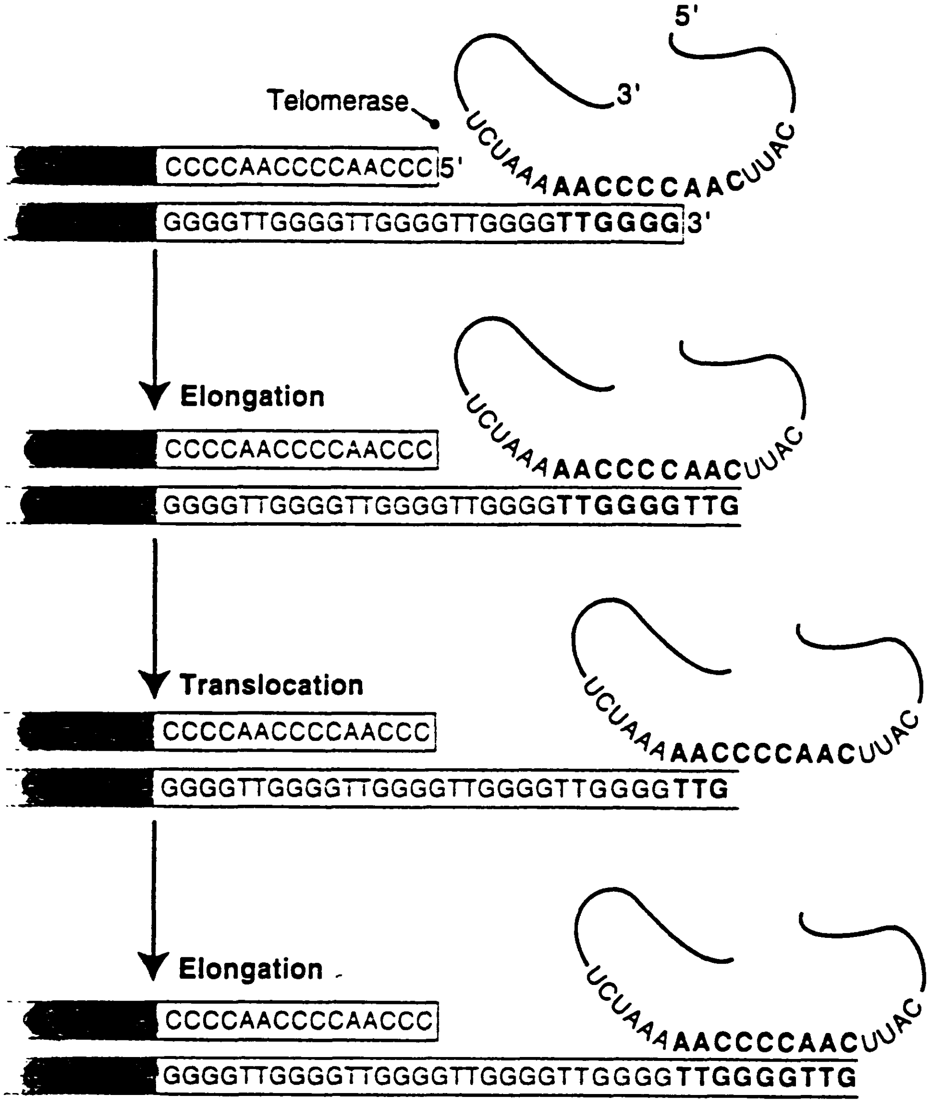

Figure A-2: Adding the Ends by Telomerase.

then synthesizes the second strand.

A.1.3 Gene Regulation

The newer roles for telomeres include aiding in gene regulation and possibly serving as a "mitotic clock" for the cells of higher animals. It has been shown that the telomeres shorten slightly every time the chromosomes replicate in preparation for cell division, suggesting that cells become senescent and die when the telomeres have shortened beyond a certain point [42, 43]. Compelling evidence that telomerase activity is essential for long-term viability of *Tetrahymena* comes from analysis of telomerase RNA mutations *vivo*. Overexpression of one particular mutant telomerase RNA gene in *Tetrahymena* is sufficient to cause a dominant negative phenotype characterized by telomere shortening and senescence [41]. Cell rescue has not been observed except by loss of this mutant RNA gene through recombination and/or segregating out the plasmid bearing the mutant gene [41]. In yeast, loss of function of a gene essential for long-term viability, *EST1*, causes steady and continuous telomere shortening over several cell generations, and eventually senescence [44]. Cell death is preceded by increased rates of chromosome loss. The *EST1* gene encodes a reverse transcriptase-like protein [45]. This, together with the identification of the ciliate telomerase as specialized reverse transcriptase [34, 35, 41] and the phenotype of *est1*-deficient mutants, indicates that *EST1* is a protein component of telomerase [45]. These findings imply that telomerase is essential for maintenance of telomere length and long-term viability in yeast as well as *Tetrahymena.*

The findings also suggest that telomerase may play a role in overriding the process that determine the proliferative life span of mammalian cells. Leonard Hayflick showed in the 1960s that normal cells have a limited life span in culture, with cells taken from younger individuals dividing more times before they become senescent and die than do cells from older individuals. *Tetrahymena* experiments showed that one particular mutation could change the activity so dramatically that the telomeres got shorter and shorter as a result, until the cells die. Hayflick's experiments showed normal mammalian cells are not immortal, and evidences have been found that their limited life span may be a result of telomere shortening. For example, human sperm cells–whose clock have just started ticking–have much longer telomeres than ordinary tissue cells that have gone through multiple rounds of division. In addition, telomeres shorten every time normal cells divide in culture. The reason that telomeres shorten in normal cells while stay intact in *Tetrahymena* and other ciliates is probably due to the lack of a functional telomerase in normal cells which can replace the chromosome ends lost to incomplete DNA replication.

Exactly what might cause cells to stop dividing and die when their telomeres get too short is currently unknown. There are several postulates. One is based on the finding that removing a single telomere from a yeast chromosome temporarily blocked cell division, and that this effect depends on the activity of the *RAD9* gene. Because *RAD9* halts the growth of cells whose DNA has been damaged, allowing for repairs, the result suggests that the telomeres may be part of the cell's damage sensing system. The second postulate is based on the evidence that loss of telomere DNA may lead to activation of the *p53* tumor suppressor gene, which serves the same function in mammalian cells as *RAD9* does in yeast: arresting cell growth in response to DNA damage. Another possibility is that intact telomeres somehow repress the activity of nearby gene. Some of these genes might trigger cell death once the telomeres become short enough to allow their expression.

All these findings does not mean we should hope the mammalian cells will be able to preserve their telomeres. Actually, the opposite might be true. Evidence is accumulating that the presence of telomerase in cells that normally lack it may contribute to the uncontrolled cell growth of cancer. It suggests that telomerase may be a good target for anti-cancer drugs.

A.2 TELOMERIC DNA STRUCTURES

When the whole telomere research field opened up in the 1970s, researchers studied telomeres in protozoan ciliates, single-celled organisms that propel themselves with hairlike projections called cilia. At the time, ciliates were much easier to work with because they have many more telomeres per cell than do mammal cells. The organisms have two nuclei, and during the formation of the larger of these, the so-called macronucleus, the chromosomes break up into fragments that then replicate, producing from 20,000 to as many as 10 million pieces of DNA, each of which becomes capped at both ends by telomeres. In contrast, a human cell has but 92 telomeres, two of each of the 46 chromosomes.

It is not surprising that the first telomere isolated was from the ciliate *Tetrahymena thermophila* in the early 1970s. This telomere has GGGGTT on strand, and AACCCC on the other, with the latter being one of the three structures presented in this thesis. The sequences repeat 50-70 times at the end of the chromosome. The sequence of the telomeric DNA came as a surprise at the time, since the only DNAs that had been closely examined were from viruses and bacteria, which do not have such repeated sequences.

But this rather "peculiar" fact turned out to be quite normal for telomeres. It was later shown that telomere sequences were conserved during evolution among otherwise widely divergent eukaryotes [23]. This helped allay one worry that researchers had when the first telomere was isolated from *Tetrahymena*: that the telomeric sequences detected on the chromosome fragments in the ciliate macromolecules might not be at all representative of the telomeres of the more standard chromosomes of other organisms. As it turns out, the essential telomeric DNA consists of a stretch of very simple, tandemly repeated sequence [15]. Table A.1 is a list of G-rich strand of telomeric repeat sequences in eukaryotes.

To understand the biological function of DNA, merely knowing its sequence is not enough. We must know its structural features. Nucleosides and nucleotides interact with proteins in all their metabolic or control operations, so mutual recognition of the two (or more) reactants is required. This presupposes that the partners involved have well-defined three-dimensional structures which, if we desire to understand functioning at atomic level, we must also know at the atomic level. X-ray crystallography provides the most powerful means of viewing macromolecular structures at high resolution.

Table A.1: Telomeric Repeat Sequences in Eukaryotes

Telomeric Repeat(5' to 3')	Representative species
AG_{1-8}	*Dictyostelium discoideum*
TTAGGGGG	*Cryptococcus*
$TG_{2-3}(TG)_{13}$	*Saccharomyces cerevisiae*
GGGGTT	*Tetrahymena thermophila*
TT(T/G)GGG	*Paramecium aurelia*
TAGGG	*Giardia lamblia*
TTAGGG	*Homo sapiens, Neurospora crassa,* *Trypanosoma brucei*
TTTAGGG	*Arabidopsis thallium*
TTTTAGGG	*Chlamydomonas reinhartdii*
$TTAC(A)AG_{2-7}$	*Schizosaccharomyces pombe*
TT(T/C)AGGG	*Plasmodium berghei*
CTTAGG	*Ascaris lumbricoides*

For all its apparent chemical simplicity, DNA can adopt a surprisingly range of structures *in vitro.* One of the icons in twentieth century, the right-handed antiparallel double helix, can exist in a number of conformations, with A- and B- forms being the best known. In 1979, Alex Rich and his co-workers discovered Z-DNA, a left-handed double helix [46], which really opened up the field of alternative conformations of DNA. More recently, studies have found that, under supercoiling, large stretches of homopurines-homopyrimidines can be induced to undergo partial strand separation with the formation of triple-stranded DNA [47]. The latest examples of alternative conformations of DNA, four-stranded structures, have come from the specialized the sequences found at the end of chromosomes in telomeres [17, 18].

As we discussed above, telomeres are found in which one strand contains many repeats of a predominantly guanine-rich sequence and the other strand a complementary cytosine-rich sequence. The first well-characterized four-stranded structures are the guanine quartets. The origin of this discovery goes back to the early experiments of polyinosinic acid which suggested that this molecule formed a parallel three- or four-stranded structures, as deduced from the X-ray diffraction fiber studies. Recent X-ray crystallography and NMR studies [17, 18] provided the first detailed structural view of the guanine quartets. These consists of four guanine bases, in a square-planar arrangement, which cohere by Hoogsteen base-pair interactions. Stacks of these guanine quartets, stabilized by monovalent cations, readily form *in vitro* under physiological conditions. The complexes are polymorphic: they can be formed from one, two or four separate DNA, or RNA, strands which can be in a parallel, or anti-parallel, orientation relative to one another. The connectivity of the strands can vary, as can the glycosidic conformation –syn or anti –of the guanine bases.

As a number of telomeric sequences have been determined to form guanine tetraplexes, one inevitably faces the question: Do these structures exist *in vivo*? The fact that the sequences are stable up to pH 7, coupled with evidence from the study of telomere-binding proteins [48, 49] and the dimerization of retroviral RNA genomes [34] indicates that they probably do.

For every stretch of G-rich sequence capable of forming G-quartet structures in a DNA double helix, there is a complementary strand rich in cytosine bases. What is the fate of these sequences? That is the focus of study presented in this thesis. We present here three crystal structures of cytosine-rich DNA sequences, with the last being the *Tetrahymena* telomere sequence. In these structures, we see another kind of four-stranded DNA motif, the I-motif.

The demonstration that telomere C-rich sequences adopt an I-motif conformation suggests that this structure may be of relevance *vivo.* G-tetraplex and I-motif structures could act in concert: any sequence, such as telomeric DNA, that will form a G-tetraplex on one strand potentially has the ability to form a C-tetraplex on the other strand. Indeed, one structure might promote formation of the other. G-tetraplexes have been suggested to mediate self-recognition, telomere formation, meiotic chromosome pairing and recombination. The I-motif has the potential to do the same.

Appendix B

Experimental Procedures

In this section here, I will discuss some of the common experimental procedures that were used in solving the structures presented in this thesis. In the "Materials and Methods" section of each chapter, detailed crystal growing conditions and data collection parameters are presented. Here, instead of repeating these conditions, I will give a general ovcrview of some of the methods that we used, which include crystal growth, derivative preparation and X-ray sources in data collection.

B.1 DNA CRYSTAL GROWTH

After pure DNA samples are obtained (for a description of DNA purification, see the "Materials and Methods" section of Chapter 1), one begins to crystallize the sample. One begins with what is called the "initial trials". The most commonly used methods for initial crystal trials are the hanging-drop and sitting-drop vapor-diffusion methods. Figure B-1 shows one variation of sitting-drop setup (hanging-drop is very similar, except that the sample drop is facing down). A cross section through a single well is shown. In (A), the lips of the wells are greased where the coverslips will later be placed. High-vacuum grease can be used alone or mixed with a little silicone oil. The addition of oil makes the grease less viscous so that it flows more easily. In (B), the lower coverslip is pressed in place. Make sure there are no gaps in the grease for air to leak through. In (C), the reservoir solution is placed in the well. The common precipitants used as reservoir solution are ammonium sulfate and MPD (2-Methyl-2,4-pentanediol). The concentration of which is a variant of crystal setup conditions. In (D), solution containing DNA is put onto the lower coverslip. In (E), precipitant (often from the reservoir solution) is added to the DNA solution. The concentration of the precipitant in the DNA solution is lower than that in the reservoir. Over a period of days to months, the water in the DNA solution will gradually diffuse into the reservoir, and usually DNA will either form a crystal, which is desired, or form precipitate. (F) shows the mixing of the two layers by drawing up and down with an Eppendorf pipette. Finally, in (G), the upper coverslip is put on top to seal the well completely.

Most of these initial trials will never crystallize. It is hoped that just the right conditions will be hit upon a few of the drops. Like fine wine, DNA crystals are best grown in a temperature-controlled environment. Temperature, along with sample concentration and pH, is an important variable in crystallization conditions.

It is impossible and impractical to systematically scan every possible precipitant that has been used for growing DNA crystals. Therefore, another approach is an incomplete factorial experiment [50]. A small subset of all possible conditions are scanned in a limited number of experiments by combining a subset of solutions. These drops are scanned for crystals or promising precipitates. If anything is found, then a finer scan can be done to find better growth conditions. A successful version of this method was developed by Jancarik and Kim [51] and has been optimized to 50 conditions combining a large number of precipitants and conditions. A kit is available from Hampton Research that contains all 50 solutions premixed, so all one has to do is set up 3-5 μls of DNA sample with each of the solutions.

Once a setup is done, one needs to check the setup periodically. Never give up on a setup unless it is completely dried; it may take several months for crystals to appear. The presence of precipitant in a setup does not preclude crystallization. Often a crystal will grow from the precipitant. Nucleation is a rare event and may require a very long time to occur if one is near, but no right on, the correct crystallization conditions.

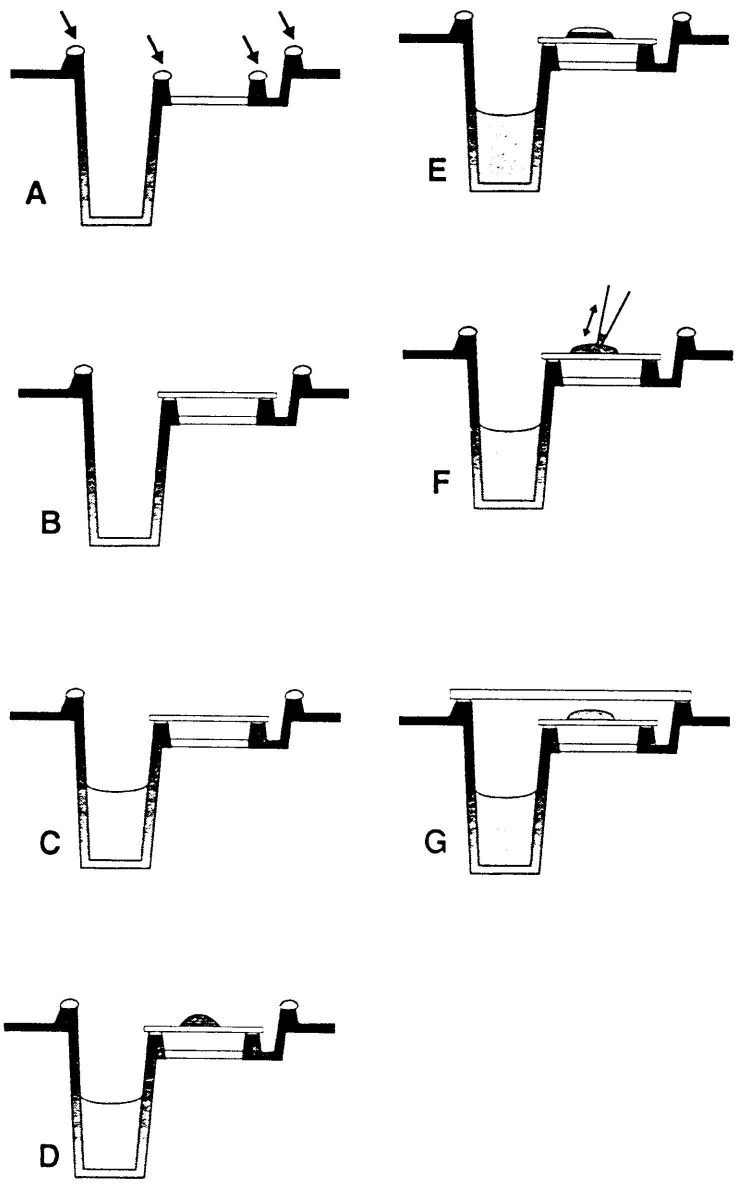

Figure B-1: Sitting-drop Crystal Setup.

B.2 DERIVATIVE CRYSTAL PREPARATION

To solve the phase problem, the most common method used is multiple isomorphous replacement. In this method a heavy atom(s) is introduced into the structure with as minimal a change to the original structure as possible. This gives phasing information by the pattern of intensity changes. A heavy atom must be used to produce changes large enough to be reliably measured. Only minimal changes, or isomorphism, are necessary because the primary assumption of the phasing equations is that the soaked crystal's diffraction pattern is equal to the unsoaked crystal's diffraction pattern plus the heavy atoms alone. More details can be found in standard crystallography texts.

In order to introduce heavy atoms or substrates into DNA crystals, they are usually soaked in an artificial mother liquor containing the reagent of choice. The compound is prepared in an artificial mother liquor solution at about 10 times the desired final concentration and then one-tenth of the total volume is layered onto the drop containing the target crystal. Diffusion occurs within several hours to saturate the crystal completely. With some heavy atoms, secondary reactions often occur that can take several days. Heavy atoms are usually introduced at 0.1-1.0 mM concentration. Many compounds will not dissolve well in the crystallization solution. In these cases it may be beneficial to place small crystals of the compound directly in the drop.

Soak time is more difficult to determine. If the soaking drop has several crystals, they can be mounted at different time intervals. Some heavy-atom reagents that are highly reactive may destroy the crystals and yet be useful if soaked for a short time. Other heavy-atom compounds undergo slow reactions that may produce a new compound that will bind. Fortunately for crystallographers, one is not as concerned with exactly what exact binds to DNA so long as something heavy binds at a few sites in an isomorphous manner. Table B.1 presents a partial list of common heavy-atom compounds and the conditions that they are normally used.

Table B.1: Useful Heavy-Atom Reagents and Conditions

Reagent	Conditions
Platinum tetrachloride	1mM, 24h
Mercuric acetate	1mM, 2-3 days
Ethyl mercury thiosalicylate	1mM, 2-3 days
Iridium hexachloride	1mM, 2-3 days
Gadolinium sulfate	100mM, 2-3 days
Samarium acetate	100mM, 2-3 days
Gold chloride	0.1mM, 1-2 days
Uranyl acetate	1mM, 2-3 days
Mercury chloride	1mM, 2-3 days
Ethyl mercury chloride	1mM, 2-3 days

B.3 X-RAY SOURCES IN DATA COLLECTION

X-rays for DNA crystallography are produced by two methods. The first method is to accelerate electrons at high voltage against a metal target, and the second is to use synchrotron radiation emitted by electrons and positrons in high-energy storage rings. Laboratory sources are limited to the former, whereas the latter is available at several international facilities.

Laboratory sources fall into two types: sealed tube sources and rotating-anode (see Figure B-2). Both produce X-rays by accelerating electrons to a high voltage of 40-50 kV at a mental target. The limiting factor in the power at which the source can operate is the rate at which heat can be removed from the target. A typical sealed tube cannot operate at more than 20 mA of current or 0.8 kW at 40 kV. A rotating anode can reach 8 kW. A brighter source is better because radiation damage to crystals is not linear with dose but is a combination of dose and time. Once the crystal is exposed, a series of chemical reactions start that eventually damage the crystal. These are triggered by the ionization resulting from the radiation. Higher powers do not linearly increase the rate at which this damage occurs, and, therefore, more useful data can be collected before the crystal is irreversibly damaged. The X-ray source in our experiments is a rotating anode generator.

The choice of the metal target determines the characteristic wavelength at which the X-rays are emitted. The most common choice is a copper target, which emits at 1.5418 $\AA$ wavelength (CuK_α). This wavelength is a good compromise between maximum-achievable resolution and absorption. The other factor in favor of copper is its superior heat-conducting properties that allow it to be used at higher power. Copper X-ray radiation is a soft X-ray: it will not penetrate very far through most materials; is absorbed quickly by air; and is scattered efficiently by air, water, and glass. The path length through air must be minimized to prevent excessive background scatter. This means placing the collimator and the beam stop as close to the crystal as possible.

In addition to laboratory sources, synchrotron radiation sources are also used. Since they may be extremely bright, they make ideal sources for characterization of small crystals. Optics at synchrotrons are also generally superior to those found in the laboratory. This is partly due to cost but also the brilliance of the synchrotron beam means that more can be thrown away and still have a very bright beam.

The X-ray radiation as it comes from the tube cannot be used for data collection without some filtering. The spectrum of a copper source consists of two main peaks at CuK_α and CuK_β and also has white radiation at both higher and lower energies than the characteristic radiation. The most common way is to use a single-crystal monochromator to filter the X-rays. A good setup with a monochromator, collimator, and beam stop can produce excellent signal-to-noise ratios.

Nowadays, the X-ray detection system attached to the X-ray source is mostly imaging plate. The system that we use in our experiments is R-AXIS II (Rigaku Automated X-ray Imaging System), which is an automated area detector diffractometer based on the Fuji imaging plate. The basic configuration of R-AXIS II is shown in

Figure B-3. In essence, it is an automated oscillation camera with reusable film and an integrated film scanner. However, the inherent properties of the imaging plate compared to ordinary X-ray film make it a far superior medium for recording X-ray events. The active area of the Fuji imaging plate is composed of an curopium doped barium halide. When X-ray strike the imaging plate, Eu^{2+} is oxidized to Eu^{3+} with the free electron being trapped in a color center. In this way latent images of the diffraction pattern are stored on the imaging plate. Measurement of the diffraction pattern is accomplished by scanning the imaging plate with a 633 nm He-Ne laser which triggers photostimulated luminescence. The light from the laser causes an electron in a color center to reduce Eu^{3+} to Eu^{2+} with atomic fluresecence occurring at approximately 390 nm. This resulting luminescence is measured by a photomultiplier and has an intensity proportional to the number of absorbed X-rays.

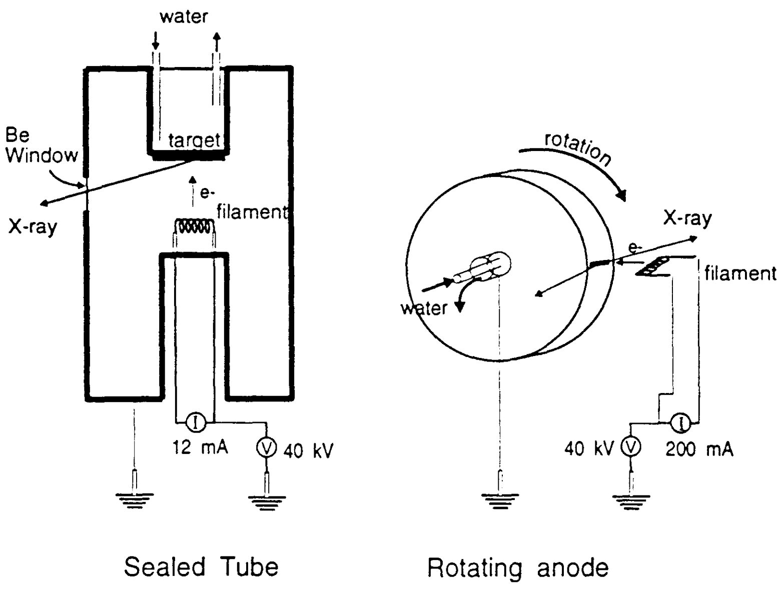

Figure B-2: Sealed Tube and Rotating Anode X-ray Sources.

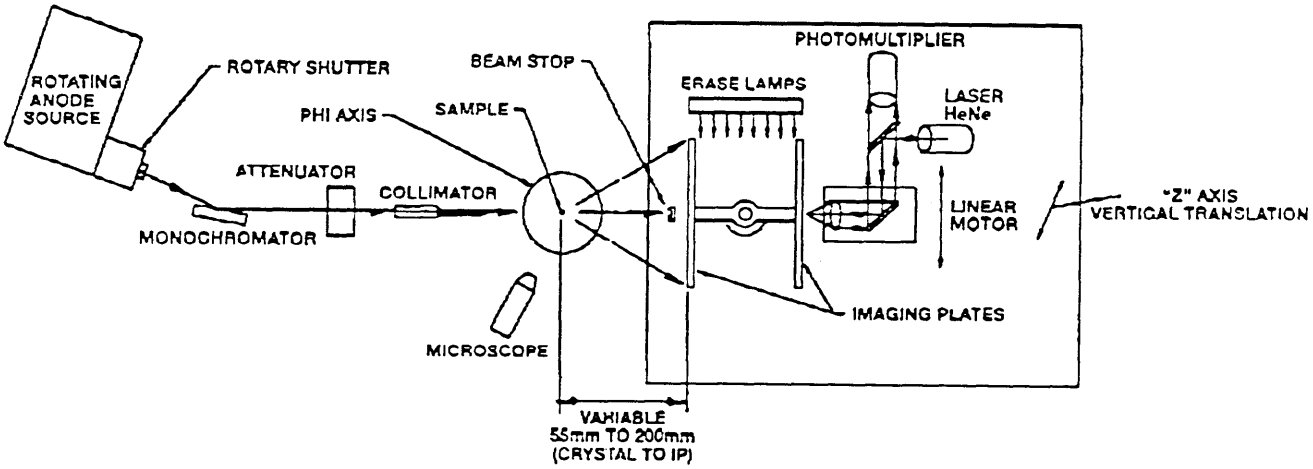

Figure B-3: R-AXIS II Design.

Bibliography

[1] Marsh, R.E., Bierstedt, R. & Eichhorn, E.L. The crystal structure of cytosine-5-acetic acid. *Acta Crystallogr,* **15**, 310-316(1962).

[2] Langridge, R. & Rich, A. Molecular structure of helical polycytidylic acid. *Nature,* **198,** 725-728(1963).

[3] Arnott, S., Chandrasekaran, R. & Leslie, A.G.W. Structure of the single-stranded polyribonucleotide polycytidylic acid. *J. Mol. Biol.,* **106,** 735-748(1976).

[4] Sarma, M.H., Gupta, G. & Sarma, R.H. A *cytosine·cytosine* base paired parallel DNA double helix with *thymine·thymine* bulges. *FEBS Lett.,* **205,** 223-229(1963).

[5] Gehring, K., Leroy, J.-L. & Gúeron, M. A tetrameric DNA structure with protonated *cytosine·cytosine* base pairs. *Nature,* **363,** 561-565(1993).

[6] Leroy, J.-L., Gehring, K. & Gúeron, M. Acid multimers of oligodeoxycytidine strands: Stiochiometry, base-pair characterization, and proton exchange properties. *Biochemistry,* **32,** 6019-6031(1993).

[7] Chen, L., Cai, L., Zhang, X. & Rich, A. Crystal Structure of a four-stranded intercalated DNA: d(C_4). *Biochemistry,* **33,** 13540-13546(1994).

[8] Terwilliger,T.C., Kim, S.H. & Eisenberg, D. *Acta Crystallogr.,* **A43,** 1-5(1987).

[9] Sheldrick, G.M. *SHELX76. Program for crystal structure determination*, University of Cambridge, England, 1976.

[10] Wang, B.-C. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.,* **115,** 90-112(1985).

[11] Jones, T.A. Interactive computer graphics: FRODO. *Methods Enzymol.,* **115,** 157-171(1985).

[12] Brünger, A.T. *X-PLOR Manual, version 2.1* Yale University, New Haven, USA, 1990.

[13] Hartman, K.A. & Rich, A. The tautomeric form of helical polyribocytidylic acid. *J. of the Amer. Chem. Soc.* **87,** 2033-2039(1965).

[14] Kistenmacher, T.J., Rossi, M. & Marzilli, L.G. A model forthe interrelationship between asymmetric interbase hydrogen bonding and base-base stacking in hemiprotonated polycytidylic acid: crystal structures of 1-methylcytosine hemihydroiodide hemihydrate. *Biopolymers,* **17,** 2581-2585(1978).

[15] Blackburn, E.H. Telomeres and their synthesis. *Science,* **249,** 489-490(1990).

[16] Blackburn, E.H. Structure and function of telomeres. *Nature,* **350,** 569-573(1991).

[17] Kang, C.H., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. Crystal Structure of four-stranded Oxytricha telomeric DNA. *Nature,* **356,** 126-131(1992).

[18] Smith, F.W. & Feigon, J. Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. *Nature,* **356,** 164-168(1992).

[19] Blackburn, E.H. & Szostak, J.W. The molecular structure of centromeres and telomeres. *A. Rev. Biochem.,* **53,** 163-194(1984)

[20] Zakian, V.A. Structure and function of telomeres. *A. Rev. Genet.,* **23,** 579-604(1989)

[21] Marx, J. Chromosome ends catch fire. *Science,* **265,** 1656-1658(1994)

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

[23] Meyne, J., Ratliff, R.L. & Moyzis, R.K. Conservation of the human telomere sequence $(TTAGGG)_n$ among vertebrates. *Proc. Natn. Acad. Sci. USA,* **86,** 7049-7053(1989)

[24] Riethman, H.C., Moyzis, R.K., Meyne, J., Burke, D.T. & Olson, M.V. Cloning human telomeric DNA fragments into *Saccharomyces cerevisae* using a yeast-artificial-chromosome vector. *Proc. Natn. Acad. Sci. USA,* **86,** 6240-6244(1989)

[25] Laughlan, G., Murchie, A., Norman, D., Moore, M., Moody, P., Lilley, D., Luisi, B. The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *et al. Science.* **265,** 520-524(1994).

[26] Kang, C.H. *et al.* Crystal structure of intercalated four stranded d(C_3T) at 1.4 Å resolution. *Proc. Natn. Acad. Sci. U.S.A.* **91,** 11636-11640(1994).

[27] Kang, C.H. *et al.* Stable loop in the crystal structure of the intercalated four-stranded cytosine-rich metazoan telomere. *Proc. Natn. Acad. Sci. U.S.A.* **92,** 3874-3878(1995).

[28] Berger, I. *et al.* Extension of the four-stranded intercalated cytosine motif by *adenine·adenine* base pairing in the crystal structure of d(CCCAAT). *et al. Nature Structural Biology* **vol.2, num.5** 416-425(1995).

[29] Brunger, A.T. Free R value: a novel statistical quantity for addressing the accuracy of crystal structures. *Nature* **355,** 472-474(1992).

[30] Inman, R.B. Transitions of DNA homopolymers. *J. Molec. Biol.* **9,** 624-637(1964).

[31] Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing. *Science* **273,** 1678-1685(1996).

[32] Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., Kundrot, C.E., Cech, T.R. and Doudna, J.A. RNA Tertiary Structure Mediation by Adenosine Platforms. *Science* **273,** 1696-1699(1996).

[33] Murray, A.W. and Szostak, J.W. Construction of artificial chromosomes in yeast. *Nature* **305,** 189-193(1983).

[34] Greider, C.W. and Blackburn, E.H. A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature* **337,** 331-337(1989).

[35] Shippen-Lenz, D. and Blackburn, E.H. Functional evidence for an RNA template in telomerase. *Science* **247,** 546-552(1990).

[36] Greider, C.W. and Blackburn, E.H. Identification of a specific telomerase terminal transferase activity in *Tetrahymena* extracts. *Cell* **43,** 405-413(1985).

[37] Greider, C.W. and Blackburn, E.H. The telomere terminal transferase of *Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51,** 887-898(1987).

[38] Zahler, A.M. and Prescott, D.M. Telomere terminal transferase activity in the hypotrichous ciliate *Oxytricha nova* and a model for replication of the ends of linear DNA molecule. *Nucleic Acids Res.* **16,** 6953-6972(1988).

[39] Shippen-Lenz, D. and Blackburn, E.H. Telomere terminal transferase activity from *Euplotes crassus* adds large numbers of TTTTGGGG repeats onto telomeric primers. *Molec. Cell. Biol.* **9,** 2761-2764(1989).

[40] Morin, G.B. The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell* **59,** 521-529(1989).

[41] Yu, G.-L., Bradley, J.D., Attardi, L.D. and Blackburn, E.H. *In vivo* alteration of telomere sequences and senescence caused by mutated *Tetrahymena* telomerase RNAs. *Nature* **344,** 126-132(1990).

[42] Shampay, J., Szostak, J.W. and Blackburn, E.H. DNA sequences of telomerase maintained in yeast. *Nature* **310,** 154-157(1984).

[43] Shampay, J. and Blackburn, E.H. Generation of telomerase-length heterogeneity in *Saccharomyces cerevisiae. Proc. Natn. Acad. Sci. U.S.A.* **85,** 534-538(1988).

[44] Lundblad, V. and Szostak, J.W. A mutant with a defect in telomere elongation leads to senescence. *Cell* **57,** 633-643(1989).

[45] Lundblad, V. and Blackburn, E.H. RNA-dependent polymerase motifs in *EST1*: tentative identification of a protein component of an essential yeast telomerase. *Cell* **60,** 529-530(1990).

[46] Wang, A. H.-J. *et al* Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282,** 680-686(1979).

[47] Frank-Kamenetskii, M.D. DNA supercoiling and unusual structures. *In DNA Topology and Its Biological Effects.* **Cold Spring Harbor Laboratory Press,** 185-215(1990).

[48] Fang, G. and Cech, T.R. The beta subunit of *Oxytricha* telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell* **74,** 875-885(1993).

[49] Liu, Z., Frantz, J.D., Gilbert, W. and Tye, B.K. Identification and characterization of a nuclease activity specific for G4 tetra-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 3157-3163(1993).

[50] Carter, C.W.,Jr., and Carter, C.W. Protein crystallization using incomplete factorial experiments. *J. Biol. Chem.* **254,** 12219-12223(1979).

[51] Jancarik, J., and Kim, S.-H. Sparse matrix sampling: A screening method for crystallization of proteins. *J. Appl. Crystalogr.* **24,** 409-411(1991).