INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONES: 224281/2/3/4/5/6
CABLE: CENTRATOM - TELEX 460392-1

SMR/111 - 10

SECOND SUMMER COLLEGE IN BIOPHYSICS

30 July - 7 September 1984

SINGLE CRYSTAL X-RAY DIFFRACTION STUDIES OF NUCLEIC ACIDS

1. Introduction to X-ray diffraction.
2. Nucleic acids structure I: RNA.
3. Nucleic acids structure II: A- and B-DNA.
4. Nucleic acids structure III: left-handed Z-DNA.
5. Nucleic acids structure IV: drug and DNA interactions.

A. WANG
Department of Biology
Massachusetts Institute of Technology
Cambridge, MA. 02139
U.S.A.

(1)

Many tRNA molecules have been crystallized, but only a few of them solved.

Yeast $tRNA^{Phe}$			
	Orthorhombic	$P2_12_12$	Rich Kim
	Monoclinic	$P2_1$	Klug
Yeast $tRNA^{Asp}$			Moras, Ebel
yeast $tRNA^{fmet}$			Sieglev
E. coli $tRNA^{fmet}$			Rich
Yeast $tRNA^{Gly}$			Wright

They are crystallized with metal ions (Mg^{++}) and polyamines.

Spermidine : $H_3\overset{+}{N}-(CH_2)_3-\overset{+}{N}H_2-(CH_2)_4-\overset{+}{N}H_3$

Spermine : $H_3\overset{+}{N}-(CH_2)_3-\overset{+}{N}H_2-(CH_2)_4-\overset{+}{N}H_2-(CH_2)_3-\overset{+}{N}H_3$

(2)

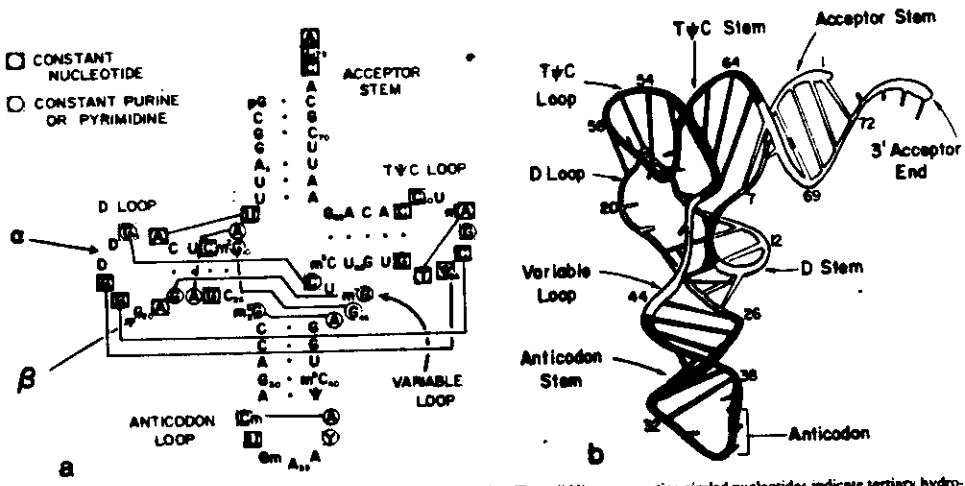


Fig. 1. (a) Cloverleaf diagram of the nucleotide sequence of yeast $tRNA^{Phe}$. The solid lines connecting circled nucleotides indicate tertiary hydrogen bonding between bases. Solid squares around nucleotides indicate that they are constant; dashed squares indicate that they are always purines or pyrimidines. The regions α and β in the D loop contain one to three nucleotides in different tRNA sequences. (b) The folding of the ribose phosphate backbone of yeast $tRNA^{Phe}$ is shown as a coiled tube; the numbers refer to nucleotide residues in the sequence. Hydrogen-bonding interactions between bases are shown as cross rungs. Tertiary interactions between bases are solid black. Bases that are not involved in hydrogen bonding to other bases are shown as shortened rods attached to the backbone.

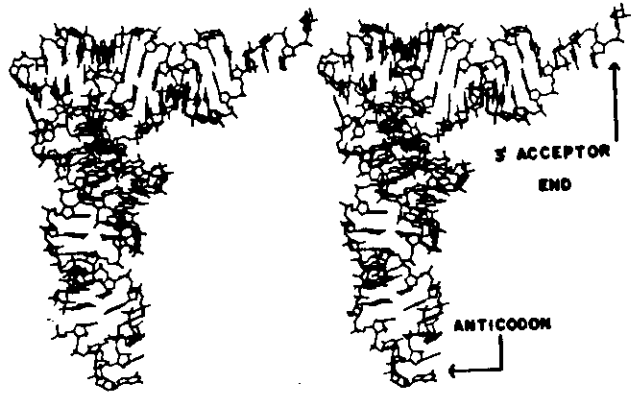


Fig. 2. Stereoscopic diagram of yeast $tRNA^{Phe}$ drawn by using the ORTEP program (37). This diagram can be seen in three dimensions by using stereoscopic glasses (24). However, the three-dimensionality of the diagram can be seen without glasses if the reader simply lets the eye muscles relax and allows the eyes to diverge slightly, so that the two images are superimposed.

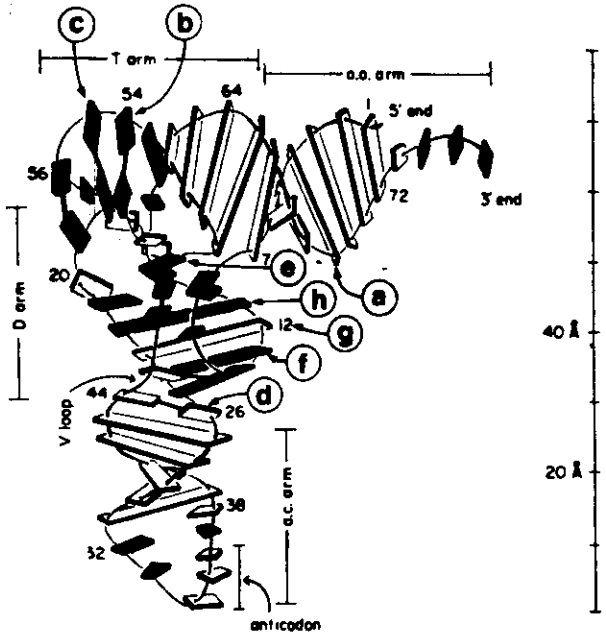
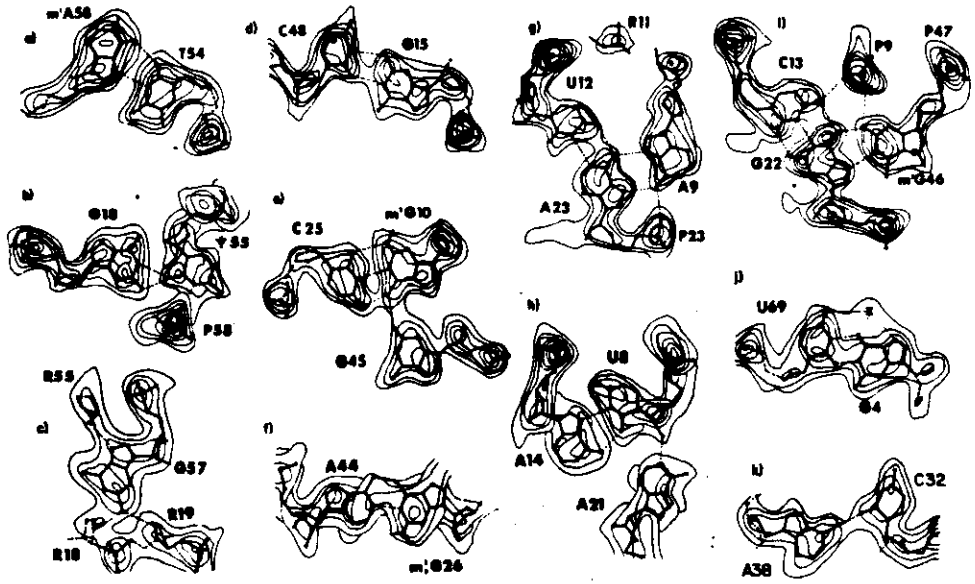


Figure 15-3 Illustration of base-pairing and stacking interactions in yeast $tRNA^{Phe}$. Base-pairs stabilizing secondary and tertiary structure are drawn as bent or fused slabs or by connecting two slabs with rods. Invariant bases in full color and semi-invariant bases in shaded color cluster in anticodon loop, CCA end of acceptor stem, and at outer corner and hinge regions of the "L" (compare tertiary interactions marked in color in Figure 15-2). Note stacking of bases 34 to 38 in the anticodon loop, consisting of the anticodon Gm_{34}–A_{35}–A_{36}. Letters a to g indicate where base-pairs and triplets displayed in Figure 15-4 are located. From (1030,1046).



(5)

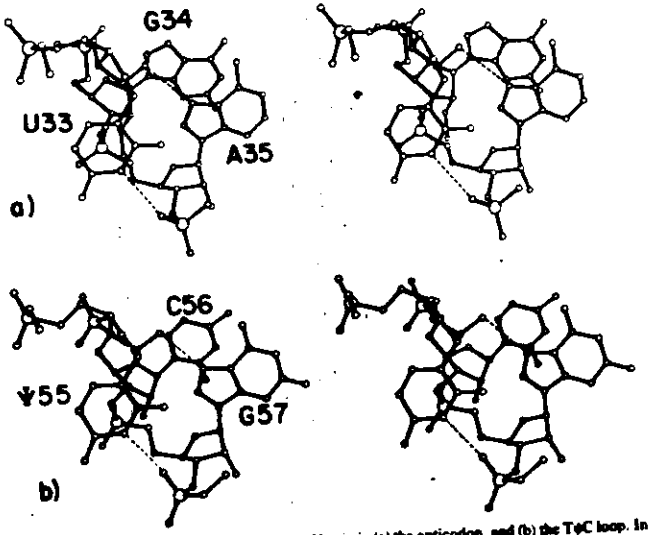
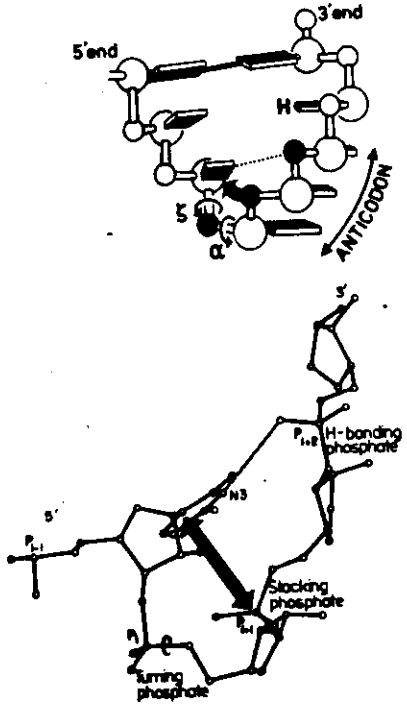


Fig. 4. Stereoscopic views of the uridine or U turn in (a) the anticodon, and (b) the TψC loop. In both cases, the turn is viewed from inside the molecule. It can be seen that phosphate groups lie directly below U33 and ψ55. The turn is stabilized by two hydrogen bonds, which are shown as dashed lines. Phosphorus atoms are drawn as large circles. The ribose phosphate backbone is darkened to clarify the folding of the chain.



(6) Why are we interested in DNA structure ?

- Genetic Control
- Structural role, High Organization
- superhelical density regulation
- Repair
- Recombination
- Replication
- Transcription

⇒ Information is coded in 3-D structure of DNA.

DNA is highly polymorphous, it can exist in many conformational states as judged by the large number of different fiber diffraction patterns.

Two types, A and B, were originally observed.

They can interconvert, i.e. A ⇄ B, depending upon the environment of the molecule.

In general, dehydration favors A-DNA.

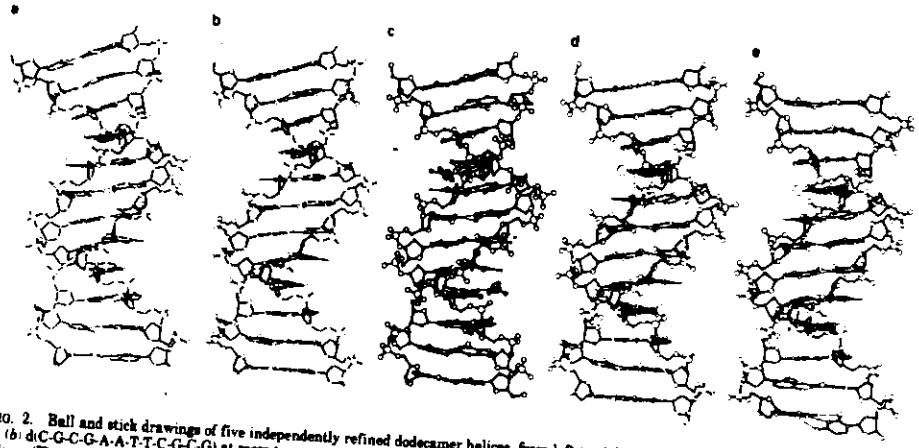


FIG. 2. Ball and stick drawings of five independently refined dodecamer helices, from left to right: (*a*) d(C-G-C-G-A-A-T-T-C-G-C-G) at 16 K (14); (*b*) d(C-G-C-G-A-A-T-T-C-G-C-G) at room temperature (12, 15–19); (*c*) d(C-G-C-G-A-A-T-T-C-G-C-G) complexed with *cis*-diamminedichloroplatinum(II) (cisplatin) (20); (*d*) d(C-G-C-G-A-A-T-T-^{Br}C-G-C-G) in 80% methylpentanediol at 7°C, or MPD7 (13, 21–24). Crossed spheres in the major groove of *c* are platinum, and large spheres in the major groove of *d* and *e* are bromines. Other atoms in order of decreasing radius are P, O, N, and C. In this view directly into the minor groove at the d(A-A-T-T) center, strand 1 begins with base C-1 at upper left and ends with G-12 at lower left; strand 2 begins with C-13 at lower right and ends with G-24 at upper right. Bending of the helix axis occurs approximately in the plane of the page. The overall bend, as measured for one full turn from base pair C-1-G-24 to C-11-G-14, has the following values: (*a*) 22°; (*b*) 18°; (*c*) 17°; (*d*) 14°; (*e*) 3°. Mean base pairs per turn for the five helices are 9.24, 9.29, 9.09, 9.36, and 9.70. All five of these drawings may be seen in stereo in ref. 25.

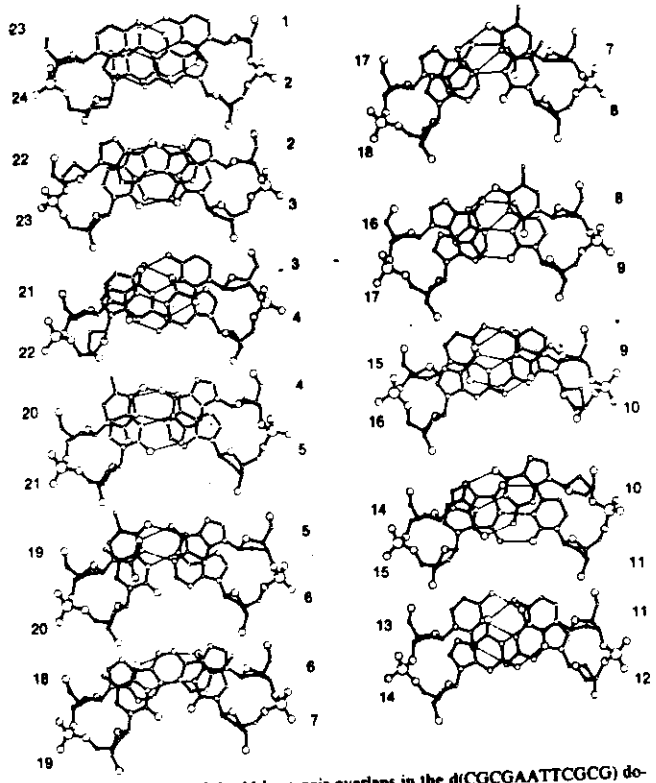
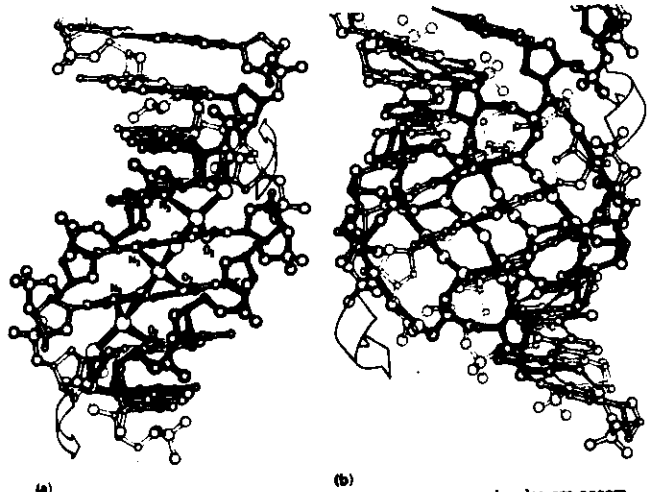


Figure 11-5. Views of the 11 base-pair overlaps in the d(CGCGAATTCGCG) dodecamer. Base sequences are indicated in color in one strand from 1 to 12, in the other from 13 to 24. Note the pattern of base overlap and sugar conformation which is rather regular in the central hexamer d(GAATTC) (steps 4 to 8) and more irregular in the outer base-pairs; compare steps 9 (A-DNA like) and 10 (D-DNA like). All base-pairs except the terminal ones appear twice and are shown in identical orientations toward the viewer. From (895).

B-DNA Dodecamer d(CGCGAATTCGCG). From (899,899a).

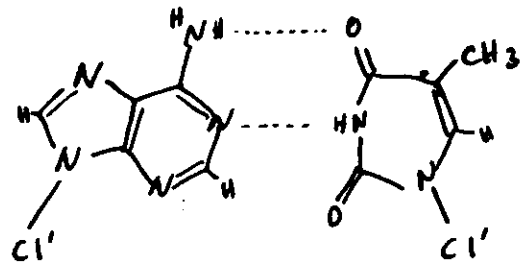
(A)

Residue	Angles, degrees χ	φ	β	γ	δ	ε	ζ	Sugar pucker	Adjacent P-atom separation (Å)
C1	−105	—	—	174	157	−141	−144	$C_{2'}$-*endo*	
G2	−111	−66	170	40	128	−186	−98	$C_{1'}$-*exo*	6.64
C3	−135	−63	172	59	98	−177	−88	$O_{4'}$-*exo*	6.47
G4	−93	−63	180	57	156	−155	−153	$C_{2'}$-*endo*	6.83
A5	−126	−43	143	52	120	−180	−92	$C_{1'}$-*exo*	6.88
A6	−122	−73	180	66	121	−186	−89	$C_{3'}$-*exo*	6.90
T7	−127	−57	181	52	99	−186	−86	$O_{4'}$-*endo*	6.29
T8	−126	−59	173	64	109	−189	−89	$C_{3'}$-*exo*	6.87
C9	−120	−58	180	60	129	−157	−94	$C_{3'}$-*exo*	6.70
G10	−90	−67	169	47	143	−103	−210	$C_{2'}$-*endo*	6.55
C11	−125	−74	139	56	136	−162	−90	$C_{2'}$-*endo*	7.05
G12	−112	−82	176	57	111	—	—	$C_{3'}$-*exo*	
C13	−128	—	—	56	137	−159	−125	$C_{2'}$-*endo*	
G14	−116	−51	164	49	122	−182	−93	$C_{1'}$-*exo*	6.62
C15	−134	−63	169	60	86	−185	−86	$O_{4'}$-*endo*	6.45
G16	−115	−69	171	73	136	−186	−98	$C_{2'}$-*endo*	7.12
A17	−106	−57	190	54	147	−183	−97	$C_{2'}$-*endo*	6.77
A18	−108	−57	186	48	130	−186	−101	$C_{2'}$-*endo*	6.71
T19	−131	−58	174	60	109	−181	−88	$C_{3'}$-*exo*	6.70
T20	−120	−59	179	55	122	−181	−94	$C_{3'}$-*exo*	6.70
C21	−114	−59	185	45	110	−177	−86	$C_{3'}$-*exo*	6.17
G22	−88	−67	179	50	150	−100	−188	$C_{2'}$-*endo*	6.60
C23	−125	−72	139	45	113	−174	−97	$C_{3'}$-*exo*	6.68
G24	−135	−65	171	47	79	—	—	$C_{2'}$-*endo*	
Mean	−117	−63	171	54[a]	123	−169	−108		
±SD	14	8	14	8	21	25	34		
B DNA[b]	−119	−61	180	57	122	−187	−91		
B_1DNA[c]	−102	−41	136	38	139	−133	−157		
A_1DNA[c]	−154	−90	−149	47	83	−175	−45		



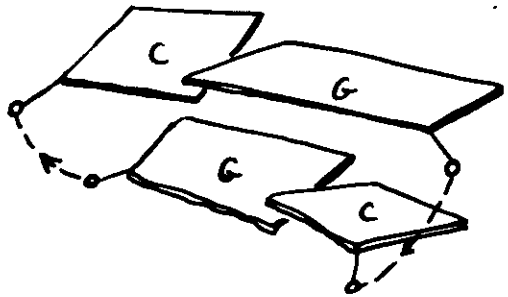
(a) (b)

Figure 17-8. (a) Drawing shows schematically how water molecules are accommodated in the *minor* groove of sequence AATT of the B-DNA-type dodecamer (1168). The first hydration layer is hydrogen bonded with pyrimidine O_2 and purine N_3 and the second layer completes the tetrahedral coordination shell around water molecules. (b) In A-DNA, filaments of water molecules are formed across the *major* groove and stitch phosphate groups together (893).

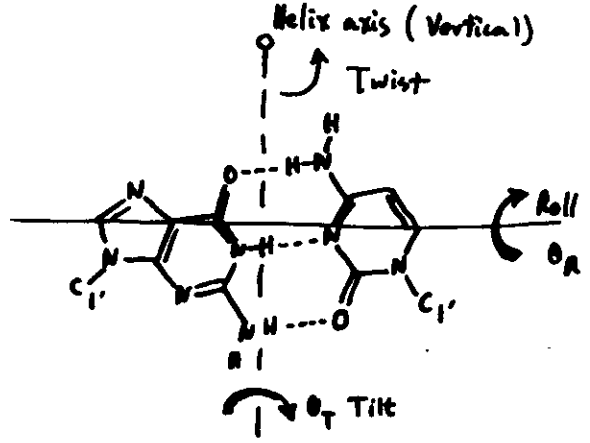


Minor Groove

A-T Base Pair

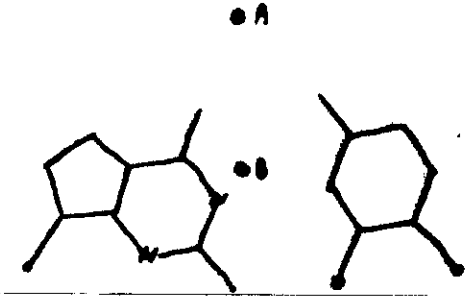


propellar twist

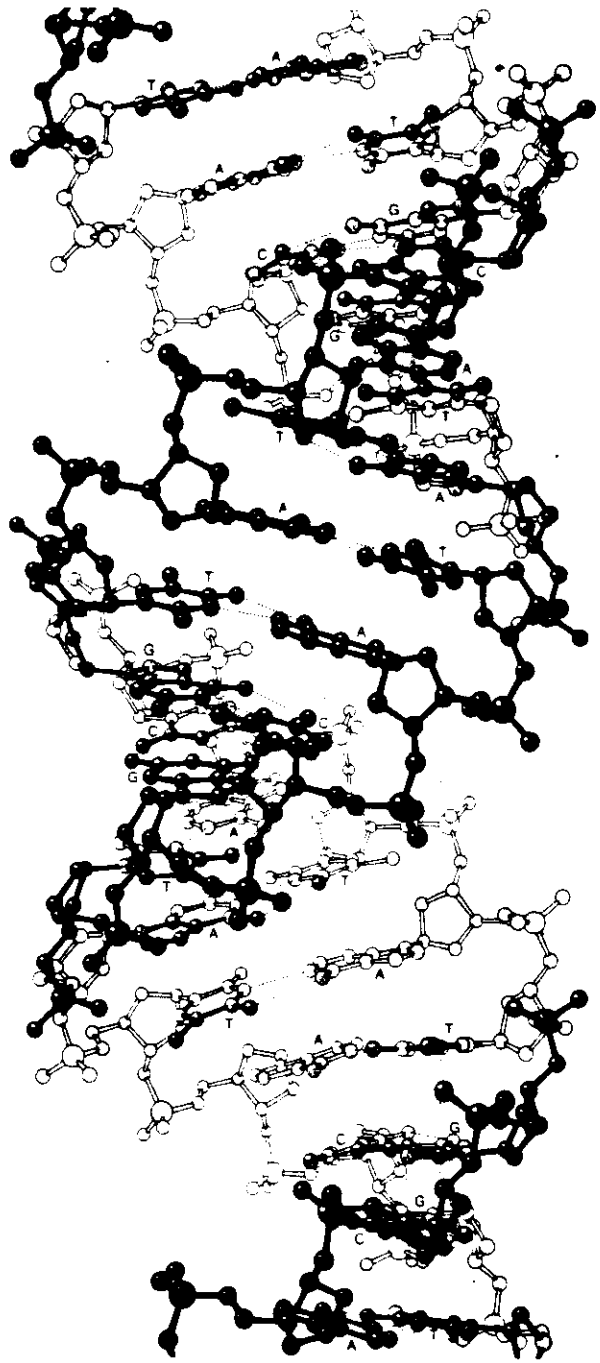


Mean Helix Parameters for A- & B-DNA

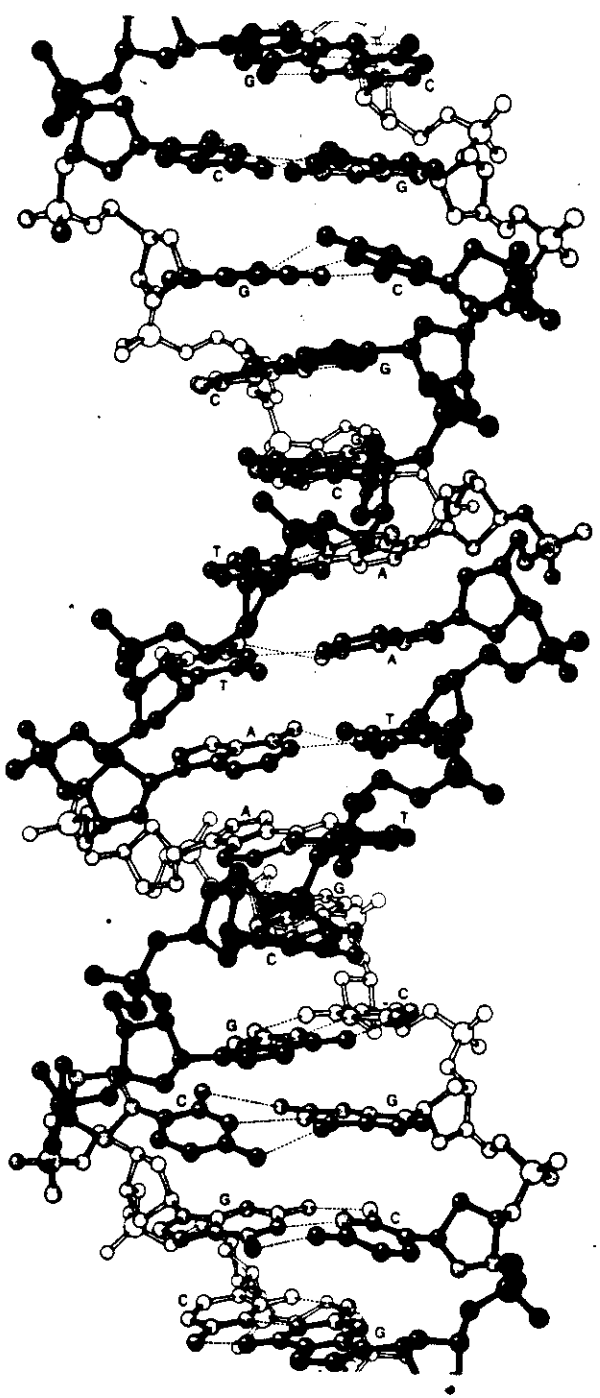
	A	A_f	B
Helical Twist (°)	33 ± 6	32.7	36 ± 4
observed range	16-44		27-42
Base-Pairs/turn	10.9		10.0
Helix Rise/Residue (Å)	2.92	2.56	3.36
Base Inclination (°)	13	20	-2
Propeller Twist	15		12
Base Roll	6		-1
Displacement D (Å)	4.0	4.5	-0.2
Sugar Pucker		C3'endo	



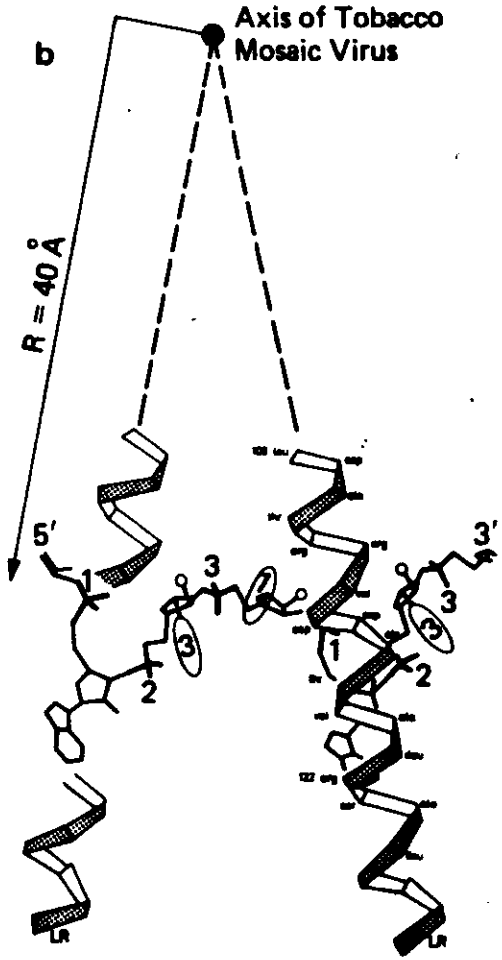
Helix position relation to bp in A, B and Z-DNA



A-DNA HELIX is seen in this perspective drawing based on the computer-generated stereoscopic pair on the opposite page. The stereo pair was generated by extending the central six bases of the octamer *GGTATA-CC*, the structure of which was determined by Kennard, Shakked and Viswamitra. The base pairs are not idealized; they are as they are observed in the crystal-structure analysis. Note how phosphate groups on opposite chains face each other across major groove.



B-DNA HELIX depicted in this drawing was generated from repetition of the central 10 base pairs of the dodecamer *CGCGAATTCGCG*. The drawing shows the large propeller twist and how the twist improves the stacking of the bases along each backbone chain.



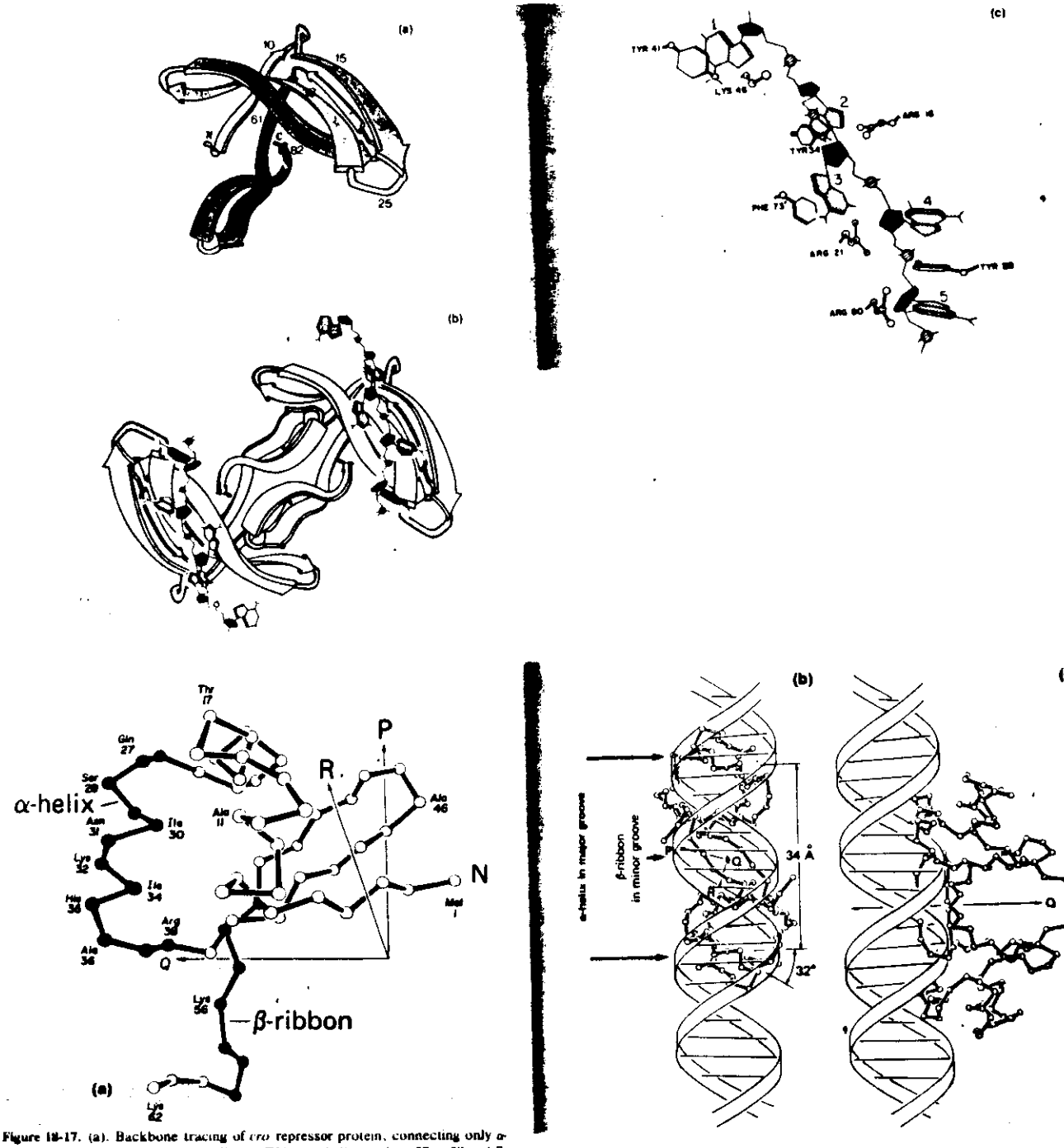


Figure 18-17. (a). Backbone tracing of *cro* repressor protein, connecting only α-carbon atoms. Presumed parts binding to DNA are helix residues 27 to 38 and C-terminal residues 53 to 62 (filled black spheres), the latter forming an antiparallel β-pleated ribbon with another *cro* molecule related by dyad Q. Two other dyads, P and R, augment this dimer to form a tetramer. Tail residues 63 to 66 "wag" and cannot be seen in electron density. They may act as "feelers" for recognizing DNA. (b,c). Model fitting of *cro* repressor dimer to B-DNA. Dyad axis Q coincides with dyad axis in B-DNA. (b). Helices and β-pleated ribbon interacting with B-DNA grooves are drawn in heavy lines and dyad axes P, Q, and R are indicated; view is along Q, with protein "behind" DNA helix. (c). Model is rotated by 90°, dyad Q is now in the plane of the paper, and N termini of the two *cro* molecules are marked. From (1278).

Structural Domains of Transfer RNA Molecules

The ribose 2′ hydroxyl which distinguishes RNA from DNA plays a key role in stabilizing tRNA structure.

Gary J. Quigley and Alexander Rich

DNA and RNA are the major macromolecules used for the transfer of information in biological systems. They differ from each other principally by the presence of a hydroxyl group on the 2′ position of the ribose sugar in RNA. This systematic difference in the polynucleotide backbone is likely to be related to significant structural and functional differences. DNA is used solely as an information carrier; certain kinds of RNA are also used in this role, including viral RNA and messenger RNA. However, other classes of RNA such as ribosomal and transfer RNA seem to play a significant structural role as well. Ribosomal RNA is believed to form a three-dimensional lattice on which ribosomal proteins are placed in the assembly of the functioning organelle. In the transfer RNA molecule, only the three anticodon nucleotides play a direct role in the transfer of genetic information through their interaction with the three nucleotides in the codon of messenger RNA. Nonetheless, transfer RNA molecules have 74 to 91 nucleotides (*1*, *2*). The vast bulk of these are involved in forming a complex three-dimensional structure whose functional role is only partly understood at the present time. For several years, we have been studying the three-dimensional structure of yeast phenylalanine transfer RNA ($tRNA^{Phe}$). This work has led to an understanding of the three-dimensional conformation of this molecule and to some understanding of the three-dimensional structures of tRNA molecules as a class (*3*).

In this article, we review aspects of the three-dimensional structure of yeast $tRNA^{Phe}$ and, in particular, discuss some of the more recent structural findings, which are based on a refinement of the molecular structure. This has led to the recognition that this tRNA molecule has several domains which essentially repeat certain structural features or motifs. In a striking fashion, the refinement has revealed the critical role played by the ribose 2′ hydroxyl group in stabilizing important features of the folding of the RNA polynucleotide chain in the nonhelical regions of the molecule. This has provided us with some perspective on the manner in which nature has utilized RNA molecules to form complex three-dimensional structures. It is likely that some of these features will be seen in the future as we come to understand the three-dimensional structure of other types of RNA, including ribosomal RNA.

Transfer RNA plays a central role in protein synthesis. An amino acid is attached enzymatically to the 3′ end of the molecule, which then enters the ribosome and forms a specific interaction with a codon base triplet of messenger RNA. The growing polypeptide chain is transferred to the amino acid, and the complex of messenger RNA, tRNA, and polypeptide chain is moved to an adjacent site in the ribosome. The polypeptide chain is then transferred to the amino acid attached to a subsequent tRNA which has entered the ribosome, and the original tRNA is released from the ribosome. We would like to understand the molecular basis of these key events in the expression of genetic information.

In 1965, Holley *et al.* (*4*) sequenced the first tRNA and noted that it could be folded into a cloverleaf arrangement in which the stem regions contained complementary bases. Since then, the sequences of some 75 different tRNA molecules have been determined (*1*, *2*) and all of them can be organized into this same general cloverleaf folding, which is illustrated for yeast $tRNA^{Phe}$ in Fig. 1a (*5*). From these results has come an awareness that a large number of the positions in the cloverleaf sequence are occupied by constant or invariant nucleotides (enclosed in a solid square in Fig. 1a). Other positions are semi-invariant, occupied exclusively by purines or by pyrimidines (shown by the dashed squares in Fig. 1a). The stem and loop regions are named as indicated in Fig. 1a; the anticodon bases (numbered 34 to 36) interact with messenger RNA during protein synthesis. During aminoacylation, an amino acid is esterified onto the 3′ terminal ribose of adenosine 76, which is at the end of the acceptor stem.

The constant features associated with all tRNA molecules include the seven base pairs in the acceptor stem, three or four base pairs in the D stem, and five base pairs and a seven-residue loop in both the anticodon and TψC arms. Eighty percent of tRNA's have either four or five nucleotides in the variable loop, while the remainder have very large loops containing 13 to 21 nucleotides (*1*, *2*). In addition, two regions in the D loop have variable numbers of nucleotides. These are labeled α and β in Fig. 1a and flank the two constant guanosine residues in the D loop. These each contain from one to three nucleotides in different tRNA species.

Polynucleotide Chain Folding

Very little was known about the three-dimensional conformation of tRNA, but the discovery in 1968 (*6*) that it was possible to crystallize tRNA molecules made it apparent that the structure would eventually be elucidated. The difficulty in obtaining crystals of sufficient quality and resolution to give a clear picture of the polynucleotide chain slowed the solution of the structure. However, in 1971 it was discovered that the addition of spermine to yeast $tRNA^{Phe}$ would produce orthorhombic crystals of that tRNA which had x-ray diffraction patterns with a resolution close to 2 Å (*7*). Spermine-stabilized yeast $tRNA^{Phe}$ was subsequently found to form well-ordered monoclinic crystals as well (*8*). The folding of the polynucleotide chain in this tRNA was discovered in 1973 through an analysis at 4 Å resolution of the orthorhombic crystals (*9*). An unusual structure emerged from this study. The molecule has a bent or L-shaped appearance in which the acceptor and TψC stems form one leg of the L; the D stem and anticodon stem form the other leg, as shown in Fig. 1b. In this conformation, the anticodon is about 75 Å away from the amino acid acceptor end. Further details of the three-dimensional folding of the molecule were revealed in subsequent studies at 3 Å resolution of the orthorhombic (*10*) and monoclinic (*11*) crystal forms. A number of tertiary interactions were revealed involving base-base hydrogen bonding. The yeast $tRNA^{Phe}$ molecules appeared very similar in these two different lattices, although some significant differences remained. However, on further interpretation of the monoclinic crystal structure at 2.5 Å resolution (*12*), the differences were found to be much fewer than had been suggested from the initial interpretation. At the present time the principal difference in the conformation of the molecule in these two lattices involves the position of the 3′ terminal residue A76. Two independent preliminary refinement analyses have been carried out on the orthorhombic data (*13*, *14*), giving rise to two closely related sets of coordinates. Similarly, two sets of preliminary coordinates have been published from the analysis of the monoclinic crystal (*15*, *16*). A comparison of three preliminary sets of coordinates has recently been published (*17*).

A significant feature of the tertiary interactions was that they involved many of the nucleotide positions which are invariant or semi-invariant as shown in Fig. 1. Analysis of the structure suggested that the polynucleotide folding of yeast $tRNA^{Phe}$ was a good model for understanding the three-dimensional folding of all tRNA's since the structure could accommodate the nucleotide variations in different tRNA sequences (*3*). Furthermore, it became apparent that this structure could account for the physical and chemical properties in solution of not only $tRNA^{Phe}$ but other tRNA's as well (*1*).

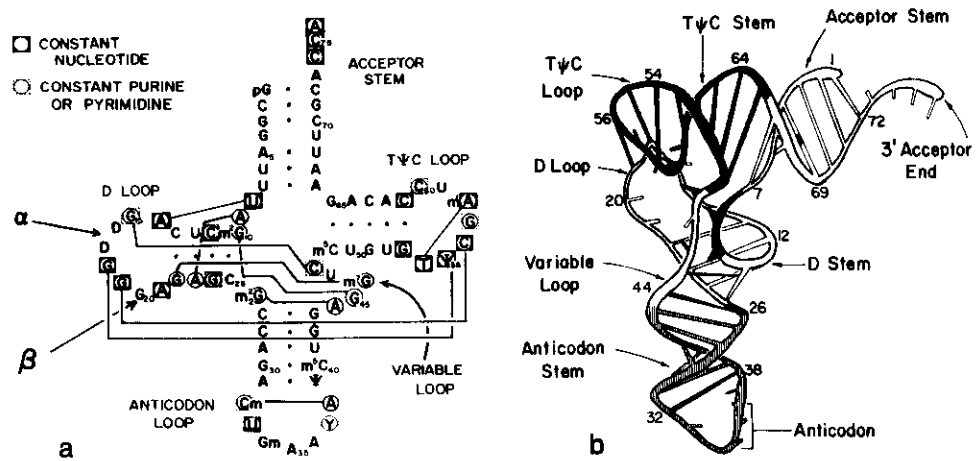


Fig. 1. (a) Cloverleaf diagram of the nucleotide sequence of yeast $tRNA^{Phe}$. The solid lines connecting circled nucleotides indicate tertiary hydrogen bonding between bases. Solid squares around nucleotides indicate that they are constant; dashed squares indicate that they are always purines or pyrimidines. The regions α and β in the D loop contain one to three nucleotides in different tRNA sequences. (b) The folding of the ribose phosphate backbone of yeast $tRNA^{Phe}$ is shown as a coiled tube; the numbers refer to nucleotide residues in the sequence. Hydrogen-bonding interactions between bases are shown as cross rungs. Tertiary interactions between bases are solid black. Bases that are not involved in hydrogen bonding to other bases are shown as shortened rods attached to the backbone.

Dr. Quigley is a research associate and Dr. Rich is William Thompson Sedgwick professor of biophysics in the Department of Biology, Massachusetts Institute of Technology, Cambridge 02139.

Molecular Refinement

Recently we have carried out a refinement at 2.5 Å resolution of the molecular coordinates of yeast $tRNA^{Phe}$ in the orthorhombic lattice. In the refinement we used a version of the real-space refinement program of Diamond (*18*), modified to accommodate polynucleotides rather than polypeptides. It was applied in a manner similar to that used by Huber *et al.* (*19*). In these procedures, the approximately 5000 positional degrees of freedom (3 for each of 1700 atoms) are reduced to 500 degrees of freedom. This is done by fixing bond distances and angles of the ribose, phosphate, and base residues to those determined very accurately from structures of small molecules. Positional adjustments are made by shifting the variable dihedral angles of the molecule to change the conformation. Within these constraints, the molecular model is altered to optimize its fit to the electron density map of the structure as calculated from the observed diffraction data and model phase information. The improved model produces improved phase information, which is applied to the observed diffraction data to produce an improved map. This procedure is iterated, and, as the map and model improve, solvent molecules and ions are added as their positions become clear. The agreement in the fitting of the molecular model to the observed x-ray diffraction data is measured by a residual, R (*20*), which decreases as the agreement between the observed and calculated data improves. For our initial molecular model, the residual at 3 Å resolution was 0.46. Some preliminary refinement was carried out and a preliminary set of coordinates was published with a residual value of 0.33 at 3.0 Å resolution and of 0.35 at 2.7 Å resolution (*13*). With continued refinement, the R value is now 0.28 at 3.0 Å resolution for 5823 reflections, and 0.29 at 2.5 Å for 7895 reflections (*21*). As the refinement proceeded, it was observed that many potential hydrogen-bonding atoms moved into positions such that hydrogen bonding was implied. At the present stage in the refinement, 91 intramolecular hydrogen bonds have been identified and they have an average distance of 2.90 Å. Our estimate of the standard deviation in atomic positions is about 0.2 to 0.3 Å. The relative deviation is much lower, of course, in the purines and pyrimidines and in the sugar phosphate backbone because their internal distances are constrained in the refinement program. However, the uncertainties in the positions of the noncovalently bonded groups make it necessary to use a knowledge of the geometry as well as interatomic distances in evaluating interactions such as hydrogen bonding. In a few cases, however, we are uncertain, either because the distance is somewhat long or because the geometry is awkward. We have called these probable hydrogen bonds in the comprehensive listing in Table 1. Some hydrogen-bonding interactions may be revised with further refinement, but we do not anticipate a large number of changes. These newly refined coordinates have allowed us to make a number of additional observations concerning the detailed conformation of the molecule. The three-dimensional form of the entire molecule is shown in Fig. 2. Sections of the 2.5Å electron density map [$2F_o - F_c$ map (*19*)] together with portions of the model in its refined position are shown in Fig. 3. Although these newer coordinates are not included here, they are available from the authors on request and have been deposited in the protein data bank (*22*).

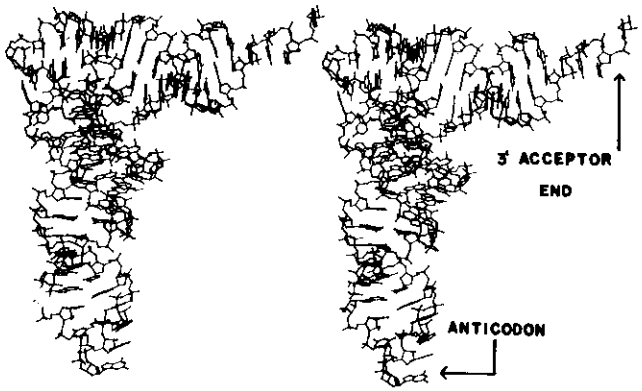


Fig. 2. Stereoscopic diagram of yeast $tRNA^{Phe}$ drawn by using the ORTEP program (*37*). This diagram can be seen in three dimensions by using stereoscopic glasses (*24*). However, the three-dimensionality of the diagram can be seen without glasses if the reader simply lets the eye muscles relax and allows the eyes to diverge slightly, so that the two images are superimposed.

Many Types of Hydrogen Bonding

Figure 1b is a schematic diagram illustrating the folding of the polynucleotide chain in yeast $tRNA^{Phe}$. A coiled tube represents the backbone, and the cross rungs represent the hydrogen-bonding interactions between the bases. As inferred from the cloverleaf diagram, the stem regions are in the form of double helical segments of polynucleotide chain. The black cross rungs in Fig. 1b represent a variety of tertiary base-base hydrogen-bonding interactions, which serve to stabilize the molecule beyond the secondary hydrogen-bonding interactions found in the double helical stem regions. With the exception of the G4 · U69 base pair, all of the base-base hydrogen-bonding interactions in the double helical stems are of the Watson-Crick type (*23*). One of the striking features of the three-dimensional structure of yeast $tRNA^{Phe}$ is the variety of alternative hydrogen-bonding interactions seen in the nonhelical regions; these utilize potentials for hydrogen bonding many of which were hitherto seen only in the structure of synthetic polynucleotides or in the crystal structures of nucleic acid constituents.

An overview of the three-dimensional form of the molecule is seen in the stereoscopic diagram of Fig. 2 (*24*). The structure is dominated by two base-stacking domains, which involve all but 5 of the 76 purine and pyrimidine bases in the molecule. The unstacked bases are in the two variable regions of the D loop (α, D16 and D17; β, G20), in the variable loop (U47), and at the 3′ terminus (A76). Base stacking is the major stabilizing feature of the molecule and is likely to be preserved in the three-dimensional structure of all tRNA molecules (*3*). The horizontal stacking domain contains the acceptor and TψC stems supplemented by two bases from the D loop and all but two of the bases in the TψC loop. The vertical domain includes the two remaining bases from the TψC loop, the D stem supplemented by hydrogen-bonded bases at either end, and the anticodon stem to the anticodon at the bottom of the molecule.

A more detailed view of some of the bases in the molecule is shown in Fig. 3, which illustrates 11 regions of the electron density map, with the corresponding refined molecular model drawn with heavier lines. Most of the parts of Fig. 3 contain single sections of the electron density map, but some contain other sections 1 Å apart where features with some depth are illustrated. In high-resolution (~ 1 Å) studies of small molecules, an electron density map shows individual atoms as isolated peaks. However, with data at 2.5 Å resolution, the electron density map shows clusters of atoms so that the individual peaks which occur are associated with the bases, the riboses, and the electron-dense phosphate groups, as seen in Fig. 3. In general, bases that are hydrogen bonded are rep-

Table 1. Tertiary hydrogen bonding in yeast phenylalanine tRNA. Standard symbols are used for referring to bases and their atoms: R stands for a purine and Y for a pyrimidine; S is used for the ribose sugar residue, and it is generally followed by an atom designation such as O2′; P stands for the phosphorus atom, and the atoms O1 and O2 are phosphate oxygens. An asterisk indicates probable but not certain hydrogen bonds.

Interaction	Comments	Structural role
	A. Base-base interactions	
1. U8 · A14	Reversed Hoogsteen pairing, explains constant U8 and A14	Stabilizes bend between residues 7 and 8
2. A9 · A23	Poly(rA) pairing of A9 with A23 in major groove of D stem	Stabilizes sharp bend of chain between residues 9 and 10
3. m^2G10 · G45	Single H bond from G45 N2 to m^2G10 O6 in major groove of D stem	Maintains interaction between variable loop and D stem
4. G15 · C48	*Trans* pairing with two H bonds, explains constant R15 and Y48, *trans* pairing required by parallel chain directions	Stabilizes joining of D stem with TψC by stacking; keeps variable loop and D loop together
5. G18 · ψ55	G18 N2 and N3 both within H-bonding distance of ψ55, may partially explain role of constant G18 and ψ55	Maintains interaction of D loop and TψC loop
6. G19 · C56	The only tertiary Watson-Crick base pair partially explains constant G19 and C56	Forms outermost corner of molecule; stabilizes interaction of D and TψC loops
7. G22 · m^7G46	Protonated m^7G46 donates two H bonds to G22 in major groove of D stem	Interaction of extra loop and D stem is also stabilized by electrostatic bond due to charged m^7G46
8. m_2^2G26 · A44	Two H bonds, involve m_2^2G26 N1 and O6 with A44 N1 and N6, propeller twist of bases	Maintains the continuity of stacking interactions from the D stem to the anticodon stem
9. Cm32 · A38	One H bond from A38 N6 to Cm32 O2, may explain constant Y32 in anticodon loop	Stabilizes the anticodon loop
10. T54 · m^1A58	Reversed Hoogsten pairing, partially explains constant T54 and role of constant A58	Maintains stacking in TψC loop
	B. Base-backbone interactions	
1. A9 N6–P23 O2	Type of H bonding found in double helical poly(rA)	Stabilizes sharp bend between residues 9 and 10
2. (a)*C11 N4–S9 O2′	H bond in major groove of D stem, may explain constant Y11-R24 in D stem	Stabilizes sharp bend between residues 9 and 10
or		
(b)*G10 N7–S9 O2′	H bond in major groove of D stem; position 10 is usually a G residue	Stabilizes sharp bend between residues 9 and 10
3. C13 N4–P9 O2	H bond from major groove of D stem to phosphate	Ties residues 8 and 9 to D stem
4. *G18 N2–S58 O1′	One H bond to furanose ring O	Stabilizes interaction of D and TψC loops
5. A21 N1–S8 O2′	A21 N1 faces S8; explains constant A	Stabilizes D loop and folding of chain
6. U33 N3–P36 O2	Explains constant U33	Stabilizes sharp U turn in anticodon loop
7. *A35 N7–S33 O2′	Single H bond to non-Watson-Crick side of central base of anticodon	Stabilizes anticodon conformation
8. *m^7G46–S9 O5′	H bond from variable loop to phosphate 9	Ties residues 8 and 9 to variable loops, helps neutralize positive charge on m^7G46
9. ψ55 N3–P58 O2	Partially explains constant ψ55	Stabilizes sharp U turn of TψC loop
10. G57 N7–S55 O2′	Partially explains constant R57	Stabilizes sharp turn of TψC loop
11. G57 N2–S18 O2′	G57 amino group is near S18	Stabilizes interaction of D and TψC loops by augmenting stacking interaction of G57
12. G57 N2–S19 O1′	G57 amino group is near S19	Stabilizes interaction of D and TψC loops by augmenting stacking interaction of G57
13. C61 N4–P60 O1	Explains constant G53 · C61	Stabilizes TψC loop stacking on TψC stem and positions bases U59 and C60 away from other TψC loop stacking interactions
	C. Backbone-backbone interactions	
1. S7 O2′–P49 O2	Maintains continuity of interrupted ribose phosphate chain	Stabilizes double helical stacking between acceptor and TψC stems
2. *S16 O2′–P18 O2	Holds residue 17 away from molecule	Maintains conformation of α region of molecule
3. *S17 O2′–P19 O2	Holds residue 18 in position to bind to ψ55	Maintains conformation of α region of molecule
4. *S45 O2′–P10 O1	Joins variable loop to sharp bend at residues 9 and 10; likely to supplement interaction B-2a above	Stabilizes juncture of anticodon stem and D stem
5. S46 O2′–P48 O2	Residue 47 bulges out of variable loop; this provides chain continuity between residues 46 and 48	Maintains conformation of central part of variable loop
6. S47 O2′–P50 O1	Joins bulge in variable loop to TψC stem	Stabilizes the bend in polynucleotide chain between residues 48 and 49
7. S48 O2′–S59 O1′	Joins variable loop to TψC loop	Stabilizes stacking between U59 and C48 · G15 base pair
8. S58 O2′–P60 O2	Residues 59 bulges out of TψC loop	Stabilizes stacking of U59 and C60 perpendicular to other bases of TψC loop; also stabilizes m^1A58 stacking on C61

resented by peaks of electron density, which are often separated from each other but sometimes run together at the lower levels of electron density. Figure 3 illustrates nine of the ten total tertiary base-base hydrogen-bonding interactions that are not of the Watson-Crick type. In addition, two sections show other features of interest.

Secondary interactions in tRNA molecules are those involved in the double helical stem regions of the cloverleaf diagram. Interactions involving bases in the nonhelical regions are generally described under the heading of tertiary interactions, and prominent among these are hydrogen bonds. Three types of tertiary hydrogen-bonding interactions are involved in stabilizing the three-dimensional structure of yeast $tRNA^{Phe}$. These include base-base hydrogen bonding, base-backbone hydrogen bonding, and backbone-backbone hydrogen bonding. Table I has a list of all the tertiary hydrogen-bonding interactions in these three classes which are observed at the present stage of our refinement analysis of the orthorhombic crystal structure. Several of these hydrogen bonds have been cited earlier (*10–17*). In addition to the interactions of which we are fairly confident, Table I includes some interactions which we consider probable but which have a slightly long hydrogen-bonding distance or a somewhat unfavorable geometry. These may be resolved on further refinement. Table I also includes comments on the various interactions as well as brief descriptions of their structural roles. Two features are apparent from Table I. The first is the great diversity of hydrogen-bonding interactions involving the bases. The second is the high frequency with which 2′ hydroxyl groups are involved in hydrogen-bonding interactions, especially in the nonhelical segments of the molecule where the polynucleotide chain has a folded or irregular conformation.

Helical Domains

The most prominent feature in the structure of yeast $tRNA^{Phe}$ is the existence of the double helical stem segments, most of which have a conformation close to the 11-fold type A RNA helix (*25*). In this form of the double helix, the base pairs are tilted approximately 15° from the helix axis and are displaced away from the axis. This creates the effect, visible in Fig. 2, of a very deep major groove in the helical stems and a very shallow minor groove. There is considerable local variation in the helical parameters of the different segments of the molecule, which undoubtedly reflect the local environment. For example, a minor perturbation is found in the acceptor stems associated with the presence of a G · U base pair. This base pair is of the wobble type (*26*). Its electron density is illustrated in Fig. 3j. In this conformation, the G is slightly farther away from the helix axis than it would be in a normal Watson-Crick base pair, while the U residue is slightly closer to it. The electron density map shows a peak of low density in the major groove of the helix, which may be due to an ion occupying that position. The peak of electron density is actually somewhat elongated in the major groove, and this may be the position of one of the spermine molecules in the orthorhombic crystal lattice. Several other counterions and solvent molecules have been visualized in the electron density map and they will be discussed more fully elsewhere.

The helical regions of the acceptor, TψC, and anticodon stems are all fairly close to a conventional RNA-A helix

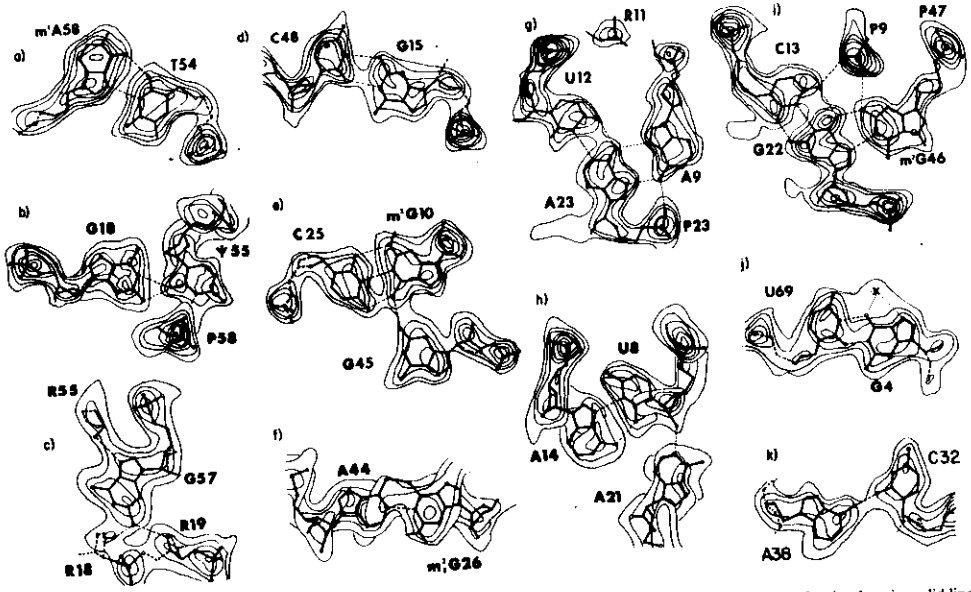


Fig. 3. Electron density sections of the refined model of yeast $tRNA^{Phe}$. Curved lines represent contours of electron density; heavier solid lines represent the molecular model as defined by the refined atomic coordinates. Hydrogen bonds between the bases are shown as thin, dashed lines. Heavier dashed lines are used to show segments of the polynucleotide chain which are outside the plane of the section. Oxygen atoms are shown as small open circles, nitrogen atoms as small solid circles. The phosphorus atoms are slightly larger solid circles. Sections b, f, and h show more than one section, as the segments of the molecule that are illustrated do not all lie in one plane. In these cases, parallel sections 1 Å apart are used. The position marked X in section j corresponds to an ion or solvent molecule which may be coordinating with the bases.

with nearly 11 residues per turn; however, the D helix deviates from this with closer to 10 residues per turn (*27*). This undoubtedly is associated with the large number of interactions which are found in the major groove of this double helix, as discussed below.

The large differences between the RNA double helix and the DNA double helix are due to the steric influence of the 2′ hydroxyl group present in RNA. In tRNA there appear to be a few cases where a hydrogen bond is possible between the O2′ of one ribose and the O1′ of the next ribose. However, the hydrogen-bonding capacity of the 2′ hydroxyl groups is largely unutilized in the helical regions of $tRNA^{Phe}$.

The Uridine Turn (U Turn)

One of the striking features which has emerged from the refinement analysis is the similarity between two regions of the molecule where the polynucleotide chain turns an abrupt corner. This is immediately clear from Fig. 4, which shows a stereoscopic view of the sharp turns in the anticodon loop and the TψC loop. Three important stabilizing interactions are seen. In both cases, the uracil residues (U33 and ψ55) play a key role in defining the character of the turn; hence we call this the uridine turn or U turn in the polynucleotide chain. As shown in Fig. 4a, the polynucleotide backbone starts with phosphate 33 at the upper left, and the chain passes away from the reader and diagonally down to the lower right. The chain changes direction quite abruptly at phosphate 34, largely through the rotation of the P-O5′ torsion angle, as pointed out previously (*28*). The net effect of this change in the direction of the chain places phosphate 35 directly under the pyrimidine ring of uracil 33, so that these are in van der Waals contact. The turn of the chain is stabilized in this conformation by two hydrogen bonds, one of which goes from N3 of uracil 33 to phosphate 36. The other hydrogen bond, which at the present stage of the analysis is slightly long, is between the O2′ of ribose 33 and the imidazole nitrogen (N7) of adenine 35.

A similar conformation is found at the end of the TψC loop (Fig. 4b), where phosphate 57 is in van der Waals contact with the uracil ring of ψ55 while nitrogen atom N3 of ψ55 forms a hydrogen bond with phosphate 58. At the same time, the 2′ hydroxyl group of ribose 55 forms a hydrogen bond to N7 of guanine 57. It appears that the stabilization for this sharp turn derives from three different interactions: (i) the hydrogen bond between the uracil ring and the phosphate group three residues farther down the chain, (ii) the hydrogen bond between the 2′ hydroxyl group and the imidazole nitrogen of the purine ring, and (iii) the van der Waals contact between the phosphate ion and the π electrons of the uracil ring, which may be stabilized through an ion-induced dipole interaction. In both cases, this phosphate group may be regarded as a capping interaction at the end of the large column of bases which form the two long stacking domains of the molecule described above.

The uridine turn is likely to be a component of all RNA's since most of the key components are invariant. Thus, the uridine residues in both cases (uracil 33 and pseudouracil 55) are constant for tRNA's involved in polypeptide chain elongation. Similarly, in the TψC turn, position 57 is always occupied by a purine residue, so that N7 is a constant feature in that position.

In the case of the anticodon loop, the center base of the anticodon A35 can be occupied by any nucleotide, not just a purine. When position 35 is occupied by a uracil or a cytosine, only very small shifts appear necessary for O2′ of ribose 33 to form a hydrogen bond to the non-Watson-Crick side of either uracil O4 or cytosine N4. If a hydrogen bond is found in the latter case, the amino group of cytosine at N4 would donate its hydrogen for the bond, and the hydroxyl would then accept rather than donate its hydrogen. It should be stressed, however, that the final decision with regard to the hydrogen bonding to the center base of the anticodon from the 2′ hydroxyl of ribose 33 will have to await the results of further refinement studies and a knowledge of other tRNA structures.

We are impressed by the fact that the detailed geometry of these turns is so similar in both the anticodon and TψC loops, even though other features of the loops are different. Within the short stretch of three nucelotides, starting with uridine, the polynucleotide chain undergoes a 180° change in direction. Even though the change in the direction occurs largely through a dihedral rotation within one residue, the stabilization for the turn involves interactions with the nucleotides on either side. It will be of interest to see if uridine turns also occur in other RNA structures.

A somewhat similar turn of the polynucleotide chain occurs between residues 9 and 10, but while the O2′ of ribose 9 may be in a position to hydrogen bond to N4 of C11, the turn is more open than the uridine turn. The only other example of the chain reversing direction is in the D loop, where the change takes place gradually over the whole loop.

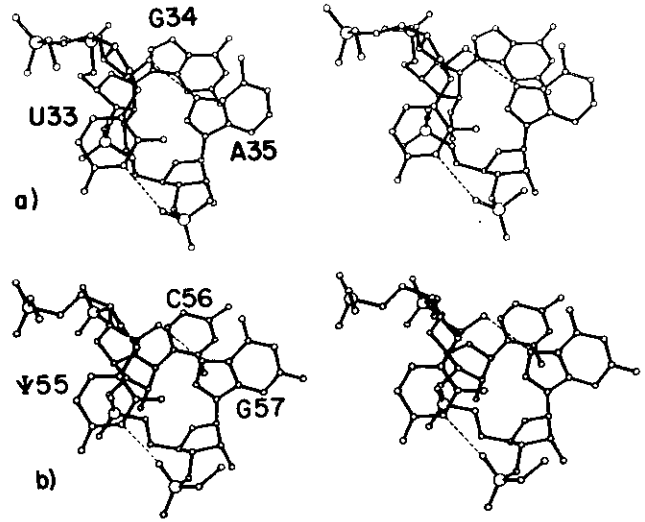


Fig. 4. Stereoscopic views of the uridine or U turn in (a) the anticodon, and (b) the TψC loop. In both cases, the turn is viewed from inside the molecule. It can be seen that phosphate groups lie directly below U33 and ψ55. The turn is stabilized by two hydrogen bonds, which are shown as dashed lines. Phosphorus atoms are drawn as large circles. The ribose phosphate backbone is darkened to clarify the folding of the chain.

The Arch Conformation

One of the interesting features which became apparent on tracing the polynucleotide chain at 3 Å resolution was the fact that the chain bulges in various regions; that is, arches out in such a way as to exclude certain residues from stacking interactions with adjacent nucleotides. In the course of the refinement analyses, we have come to recognize that there appears to be a common structural mechanism for stabilizing these arch domains in which the 2′ hydroxyl group of an adjoining nucleotide plays a crucial role. An example of this can be seen in Fig. 5, which shows a corner of the molecule near the inside of the L, where there is a joining of the acceptor and TψC stems and the variable loop extending down toward the bottom of the diagram. Yeast $tRNA^{Phe}$ has five nucleotides in its variable loop, four of which are involved in base-base hydrogen bonding. One of them, U47, is essentially excluded from the center of the molecule, and that base projects out away from the body of the molecule, where it is readily susceptible to chemical modification (*1*).

The structural mechanism for stabilizing this looping out is seen in Fig. 5. The central feature is the hydrogen bonding of the 2′ hydroxyl group of ribose 46 on one side of residue 47 to the phosphate group of residue 48 on the other side (Table 1, C-5). This has the net effect of pulling together the ribose phosphate backbone in such a way that nucleotide 47 bulges out, projecting its uracil residue away from the rest of the molecule. This conformation is likely to be found in tRNA's that have five nucleotides in the variable loop. It is noteworthy that a substantial number of tRNA's have four residues in the extra loop. As suggested earlier (*3*), this may be brought about by the exclusion of residue 47, with joining of the ribose phosphate chain between positions 46 and 48. This joining could easily be facilitated by small rotations of the ribose and phosphate residues in 46 and 48 without substantial distortion of the remainder of the polynucleotide chain.

There are several examples of arches somewhat similar to that seen in residue U47. Another example is found in the TψC loop, where two of the bases, U59 and C60, are oriented so that they do not stack with the remainder of the bases in the TψC loop, but in a direction virtually at right angles to these (Fig. 6). The bases 59 and 60 initiate the long vertical stacking domain (Fig. 2), which extends down to the anticodon of the molecule. This is achieved partially by creating a bulge for residue U59 (Fig. 6), which is expelled from the stacking interactions of the remainder of the TψC loop through formation of a hydrogen bond between O2′ of ribose 58 and phosphate 60 (Table 1, C-8), leaving U59 stacked on the pair C48 · G15.

Two other probable examples of arch domains are found in the α region of the D loop, where dihydrouracil residues are found in positions 16 and 17. Here we find that D17 is projected out away from the rest of the polynucleotide chain through a hydrogen bond between O2′ of ribose 16 and phosphate 18 (Table 1, C-2) in a manner similar to that described above for U47 and U59. It is interesting that a similar hydrogen bond between the O2′ of ribose 17 and phosphate 19 (Table 1, C-3) has the effect of projecting guanine 18 back toward the molecule, where it interacts with the base ψ55. Since the α region of the D loop has one to three nucleotides, it is clear that this arch domain is used to add additional nucleotides without actually having the nucleotides buried in the body of the molecule. This structural feature is correlated with the fact that nucleotides in the α and β regions are characteristically reactive in chemical modification experiments (*1*).

The arch domain is commonly characterized by the use of an altered ribose conformation. Instead of the normal 3′-*endo* conformation, the ribose residue which donated its O2′ hydroxyl for the hydrogen bonding is usually found in the 2′-*endo* conformation, as in riboses 46, 58, and 17.

It is also possible to have larger bulge or arch domains. Figure 5 shows examples of two other types of bulges. The 2′ hydroxyl of ribose 47 forms a hydrogen bond with phosphate 50 (Table 1, C-6) encompassing two nucleotides. In this case, the polynucleotide chain has changed direction between residues 50 and 49, which are in the acceptor stem, and residues 48 and 47, which are in the variable loop. This change in the direction of the polynucleotide chain is stabilized by the participation of a 2′ hydroxyl-phosphate hydrogen bond in a manner similar to that seen in the examples above.

A further example of a very large arch or looping out is seen in Fig. 5. Residues 50 and 49 are part of the TψC stem, while residue U7 is part of the acceptor stem. Even though the bases in these two stems are stacked on each other so that they have the appearance of a continuous double helix, the polynucleotide chain is interrupted between residues 7 and 49 with an extremely large bulge which, in fact, constitutes the entire D arm, anticodon arm, and variable loop. This very large arch is similarly stabilized by a hydrogen bond between ribose 7 and phosphate 49 (Table 1, C-1), and ribose 7 is also in the 2′-*endo* conformation. The hydrogen bond uniting ribose 7 and phosphate 49 has the effect of maintaining the continuity of the ribose phosphate chain along one side of the double helix, even though the actual ribose phosphate chain bulges out away from the double helix to form a large segment of the molecule.

In view of the fact that it appears possible to have a very large arching domain, one is tempted to ask whether or not this would be a model of how the large variable loops are integrated into the tRNA structures. Approximately 20 percent of the tRNA sequences have variable loops containing 13 to 21 nucleotides (*1*, *2*). These may be accommodated in much the same way as the single nucleotide bulge of U47, which is shown in Fig. 5. In other words, simply increasing the number of nucleotides in that arch might be a structural basis of the large variable stem and loop structure.

Thus we see that the 2′ hydroxyl group which forms hydrogen bonds to phosphates has the property of stabilizing bulges in the polynucleotide chain which are used in the three-dimensional structure of yeast $tRNA^{Phe}$ in a variety of ways. It is likely that similar arches or looping out of single or multiple residues in which this stabilizing mechanism is used will be found in other RNA structures. Such looping out is commonly seen in the proposed helical stretches of ribosomal and messenger RNA. Indeed, the crucial role of the 2′ hydroxyl group in stabilizing this conformation in the nonhelical regions may be one of the important distinguishing features which separates RNA from DNA. The absence of this stabilizing structure for looping out in DNA may be important in maintaining the fidelity of replication and transcription.

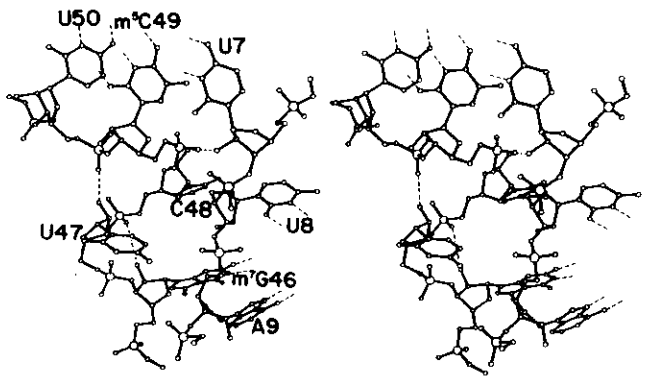


Fig. 5. Stereoscopic view of the juncture between the acceptor and TψC stems and the variable loop near the core of the molecule. This region is near the inside of the L-shaped molecule where it turns the corner. The axis of the acceptor stem and TψC stem is horizontal across the top of the diagram. The molecule is roughly in the same orientation as that shown in Fig. 2 except that the acceptor stem is rotated toward the reader. The coiling of the polynucleotide chain is stabilized by several hydrogen bonds between 2′ hydroxyl groups and phosphate residues, as described in the text. Dashed lines attached to the bases imply that they are hydrogen bonding to other bases, which are not shown.

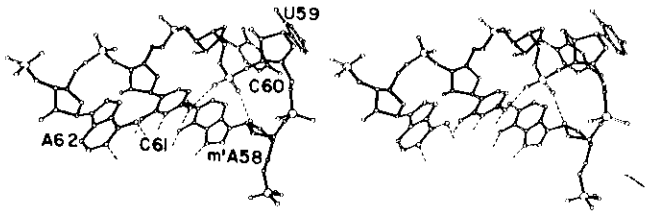


Fig. 6. Coiling of the polynucleotide chain in the TψC loop shown in a stereoscopic diagram. Residues 61 and 62 are part of the TψC stem; m¹A58 is part of the TψC loop. It can be seen that the polynucleotide chain loops out in such a way that residues 59 and 60 do not stack with the rest of the bases. This conformation is stabilized by hydrogen bonding between O2′ of ribose 58 and phosphate 60 as well as between C61 and phosphate 60.

The Corner of the Molecule

A number of additional structural features of interest have emerged in the course of the refinement, and these will be discussed by surveying the molecule from top to bottom in the orientation seen in Fig. 2.

The form of the TψC loop and its interactions with the D loop are likely to be the same in all tRNA's responsible for polypeptide chain elongation because of the large number of invariant nucleotides in this region.

Two features characterize the TψC loop: the sharp bend (U turn) where the polynucleotide chain reverses direction, and the arch structure, which ejects residue U59 from the loop and also makes possible the stacking of pyrimidine C60 outside the loop. The loop is held tightly together by a variety of interactions. One involves the amino group of C61 forming a hydrogen bond to phosphate 60, as shown in Fig. 6 (Table 1, B-13). This is of interest because the base pair C61 · G53 is constant in all tRNA sequences, suggesting that this interaction is important. It appears to be used in stabilizing the turning out of residues 59 and 60 from the remainder of the loop.

The other residues of the loop are characterized by strong stacking interactions at four levels. The first level is the base pair m¹A58 · T54 (Fig. 3a). The importance of stacking interactions in stabilizing these loops is seen in the interesting changes in tRNA of *Thermus thermophilus*, a thermophilic bacteria. Its tRNA's have an elevated melting temperature compared to those of mesophilic organisms (*29*) and systematically have 5-methyl-2-thiouracil instead of thymine. Instead of having an NH. . .O hydrogen bond, as in the pair m¹A58 · T54, there is probably an NH. . .S hydrogen bond (*30*). The higher melting temperature is probably due to the fact that substitution of the sulfur atom in the 2 position of T54 places more polarizable electrons at that site, thereby increasing the strength of the stacking interactions. This suggests that one of the major events which initiates the melting of tRNA's may frequently involve disruption of this corner of the molecule.

In the next stacking level the residue ψ55 of the TψC loop is found hydrogen bonded to G18 of the D loop (Fig. 3b). The O2 atom of ψ55 is within hydrogen bonding distance of both N1 and N2 of G18. Although these distances are not identical, the O2 of ψ55 may be bonding to one or both of these G18 nitrogen atoms. In addition, N3 of ψ55 is hydrogen bonding to phosphate 58 (Figs. 3b and 4b), where it is one of the stabilizing interactions of the uridine turn of that loop.

The electron density of a section through G57 (Fig. 3c) shows that it is stabilized by three hydrogen bonds, one of which involves the 2′ hydroxyl of ribose 55 bonded to N7 of G57. As mentioned previously, this is in accord with the fact that position 57 always contains adenine or guanine. In addition, a purine in position 57 maximizes the stacking overlap in its intercalation between the constant guanines 18 and 19 of the D loop, thereby stabilizing the joining of the two loops. The N2 amino group of G57 is hydrogen bonding to O1′ of ribose 19 and O2′ of ribose 18. These two hydrogen bonds are able to form because the polynucleotide chain of the D loop is fully extended between residues 18 and 19 where they surround the intercalated G57. This elongation is associated with the 2′-*endo* conformation in both ribose 18 and ribose 19. However, when an adenine is present in position 57 in other tRNA's, these two hydrogen bonds would not be formed. In this context it would be interesting to ask whether the tRNA's with adenine residues in position 57 have, in general, a somewhat lower melting temperature than those with guanine there. In view of the fact that thermally stable tRNA's are produced by stabilizing interactions in the TψC loop, this is a possibility.

The outermost corner of the molecule (Fig. 2) is formed by a slightly distorted Watson-Crick pairing between guanine 19 and cytosine 56. This distortion may be related to the extended conformation of residue 19.

The work of Erdmann and colleagues (*31*) has suggested that the TψC loop opens up in the ribosome where it may interact with the 5*S* RNA. From this point of view, it would be interesting to know whether there are instabilities in the structure which suggest that the conformation of this corner of the molecule is less stable than other regions. It is difficult to make a clear assessment of instability from the knowledge of the structure itself. The detailed fitting together of the elements of the D loop with the TψC loop suggests that loop disengagement might be followed by a spontaneous opening of the TψC loop. This question could be subjected to an experi-

mental analysis. If one isolates the 3′ half of the tRNA molecule, it includes the entire TψC stem and loop, but not the D loop. Does the half-molecule still maintain the uridine turn, as illustrated in Fig. 4b? It should be possible to carry out a spectroscopic analysis which would answer this question, and which might shed light on the detailed mechanism of the participation of the TψC loop in protein synthesis.

The Complex Core Region

There is one region in the center of the tRNA molecule where four polynucleotide chain segments interact; we have called this the core region. It includes the four base pairs of the D stem together with residues 14 and 15 of the D loop; residues 8 and 9, which span these bases; and the nucleotides of the extra loop which interact with the D stem. In many ways, this is the most complex portion of the molecule, in that a large number of intricate interactions from these four different chain segments contribute to the total organization of the molecule. The interactions in this area include not only base-base interactions, but also base-backbone and backbone-backbone interactions, as listed in Table 1. Several sections of Fig. 3 illustrate interactions in the core region. At the very top of the core region (using the orientation of Fig. 1b) there is a *trans* Watson-Crick pairing between C48 and G15 (Fig. 3d). Immediately below this is a reverse Hoogsteen interaction between U8 and A14 (Fig. 3h). The constant nucleotide A21 is in the same plane and is in contact with ribose 8 through a hydrogen bond from the ribose 2′ hydroxyl group to N1 of A21 (Table 1, B-5). The next lower level is a complex interaction of all four chains, as seen in Fig. 3i, in which the paired bases of the D stem (C13 · G22) are forming hydrogen bonds in their major groove with m^7G46 and phosphate 9. In addition, a probable hydrogen bond is found between the amino group of m^7G46 and phosphate 9. Since m^7G46 has a positive charge, it is likely that the charge neutralization by phosphate 9 adds to the stabilization, as is suggested in chemical modification experiments (*32*).

Figure 3g shows the interactions of adenine 9 in the major groove of the U12 · A23 base pair of the D stem. Adenine 9 interacts with A23 in a manner similar to that found in the poly(rA) double helix (*33*), including the hydrogen bond between the amino group of A9 and phosphate 23. In the level immediately below in the D stem is a pairing of cytosine 11 with guanine 24. As has been pointed out earlier (*12*), the O2′ of ribose 9 can hydrogen bond to the N4 of C11 (Table 1, B-2b) even though the hydrogen bond is slightly long. This bonding also facilitates a hydrogen bond between the ribose 45 O2′ hydroxyl group and phosphate of G10 (Table 1, C-4). However, O2′ of ribose 9 is also in a reasonable position for forming a hydrogen bond to the imidazole N7 of G10 (Table 1, B-2a). Further refinement will have to be carried out before we can say with certainty which of these two interactions is likely to be found in this molecule. Either conformation would assist in stabilizing the turn in the polynucleotide chain between residues 9 and 10.

The bottom base pair of the core region includes the single hydrogen-bonding interaction of G45 with m^2G10 of the D stem (Fig. 3e). In this position G45 is not in the same plane as the base pair of the D stem to which it is hydrogen bonded, but is stacked on the adjacent A44.

One of the striking characteristics of the core region is the extent to which there is a high density of interactions connecting different parts of the molecule. These interactions include not only the close packing due to size and shape and the stacking of the bases, but also the large number of hydrogen bonds, which provide a great deal of the specificity to the packing. The core region contains 25 percent of the nucleotides in yeast $tRNA^{Phe}$, yet it has almost 50 percent of the tertiary hydrogen bonds listed in Table 1. Many of the interactions in the core region involve invariant or semi-invariant nucleotides, so it is likely that all tRNA molecules have certain common features in the core region. However, as discussed earlier (*3*), tRNA's with a different sequence are likely to have a number of small variations of the basic arrangements seen in yeast $tRNA^{Phe}$.

One base pair involving A44 and m^2_2G26 (Fig. 3f) connects the core region to the anticodon stem. At an early stage in the refinement analysis, it appeared that residue A44 might be in a *syn* conformation (*13*); however, further analysis showed clearly that it is in the *anti* conformation, as seen in Fig. 3f. Several sections of the electron density map are used in drawing Fig. 3f since there is a marked propeller-like twisting between the bases A44 and m^2_2G26, as first described at 3Å resolution (*10*). This twisting is a necessary consequence of the van der Waals interactions due to the dimethylamino group on the guanine residue, which prevents these two bases from being planar. The net effect of the propeller stacking is that it allows A44 to stack on the first base pair of the anticodon stem, while m^2_2G26 can stack quite heavily on the first pair of the D stem. These two stacking interactions would not be possible if the bases did not have this propeller-like twist; thus, dimethylation of the guanine residue effectively maximizes stacking interactions in this joining region.

The Anticodon and Protein Synthesis

The anticodon stem is a typical RNA-A helix. At the present stage in our refinement, however, we have detected an anomaly in the bottom base pair of the anticodon stem A31 · ψ39. The electron density on the ψ residue does not make clear whether the two atoms involved in hydrogen bonding to the adenine are O4 and N3 or O4 and N1. In the former, more normal pairing, the O2 is in the usual position for Watson-Crick pairing in a *syn* conformation. However, the electron density in the O2 region is weak. In the second pairing, ψ39 is in an *anti* conformation and O2 is no longer projecting toward the adenine but away from it, where the electron density is stronger. Further work on the refinement of this structure will be necessary to resolve the conformation in this region. It is interesting to note that an A · ψ pair frequently occurs at the base of the anticodon stem, and the ψ is found exclusively in position 39 and never in position 31. Furthermore, a U is never found in position 39 (*2*).

Immediately below this base pair is an unusual interaction between A38 and C32, for which the electron density is shown in Fig. 3k. It can be seen that the O2 of C32 is close enough to A38 to form a hydrogen bond even though the geometry is not optimal. This interaction is of interest since position 32 is always occupied by a pyrimidine and both C and U have oxygen atoms in the 2 position. Position 38 does not always have adenine, however, even though it usually has a base which can donate a hydrogen atom. Analysis of other tRNA's will be necessary before it is possible to evaluate the generality of this interaction.

Two views of the anticodon together with the hypermodified purine (Y37 base) on its 3′ side are shown in Fig. 7. Figure 7a is a stereoscopic view as seen from the inside of the molecule. Figure 7b is a diagram viewed from the outside

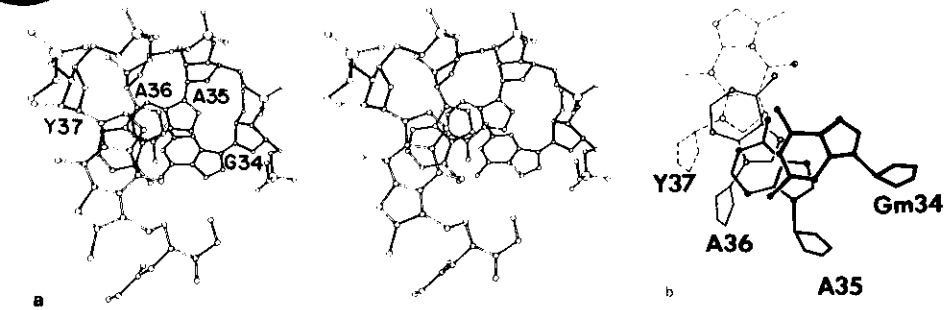


Fig. 7. (a) Stereoscopic view of the anticodon bases 34 to 36 and the hypermodified base Y37 as viewed from the interior of the molecule. It can be seen that the stacking of residues Y37, A36, and A35 is somewhat greater than that of G34 and A35. (b) Diagram illustrating the stacking of anticodon bases 34 to 36 on the hypermodified Y base as viewed from the exterior of the molecule. Oxygen atoms are stippled circles, and nitrogen atoms unstippled circles. Not all of the side chain of the Y base is included. It can be seen that the anticodon takes the form of a right-handed helix.

of the molecule, which shows the three anticodon bases plus the Y base projected onto the plane of A35. From Fig. 7, a and b, it can be seen that residue A36 is well stacked on the Y37 base, and A35 is well stacked on A36. However, G34 is only partially stacked on A35, but this may be due to intermolecular interactions. In the orthorhombic lattice a twofold axis passes near the anticodon such that G34 of one molecule is stacked on G34 of a neighboring molecule. We cannot be certain of the extent to which this affects the detailed stacking of the anticodon.

The significant feature of the anticodon, however, is that the bases are stacked in a right-handed helical array where they are presumably ready to interact with the codon. The anticodon does not have the conformation of one half of a typical RNA-A helix. The helical twist is tighter, having 7 to 8 residues per turn rather than 10 to 11 (*27*). This raises a fundamental question. When the anticodon combines with the codon, does the anticodon loop undergo a conformational change and adopt another conformation—for example, that of an RNA-A helix—or does the codon of the messenger RNA adopt instead the conformation of the anticodon and form a fragment of a tighter double helix with approximately eight residues per turn? We do not know the answer as yet, but it is of interest that when two tRNA molecules with complementary anticodons combine with each other, the anticodon conformation has only a slight change in stacking (*34*).

As discussed above, the turn of the polynucleotide chain at the anticodon is stabilized by the interactions of the uridine turn (Fig. 4a), so that it is entirely possible that the conformation is maintained when bound to the codon. If this is so, then it appears to be impossible for two tRNA's to form codon-anticodon complexes with adjacent codons if the messenger RNA retains the form of an uninterrupted helix, since the adjacent Y base effectively occupies space which the continuation of the codon helix must occupy. However, two adjacent codons could hydrogen bond to two tRNA molecules if there was a disruption in the helical conformation of the messenger RNA in this region. It has been suggested that the messenger RNA might become unstacked between codons while it is being read (*35*). However, we will have to wait for the results of further experiments before questions of this type can be answered.

The anticodon conformation with G34 at the very apex of the molecule leaves that residue in a position where it may be capable of assuming the different conformations required for the wobble pairing (*25*). This possibility for the anticodon conformation was first pointed out by Fuller and Hodgson (*36*).

How could G34 interact in wobble pairing? The wobble interaction implies that the guanine residue could form either a normal Watson-Crick G · C pair, or a G · U pair such as that seen in Fig. 3j. This suggests that two different conformations of the residue G34 may be used, one of which is involved in the normal Watson-Crick interaction and the other in the wobble interaction. The difference in the position of G34 involves a displacement of the base away from the helical axis of approximately 2 Å. We have found that a small conformational change in the region of residue ribose 34 makes possible such a shift in the position of G34.

Summary

In this article, we have described various detailed features of the conformation of yeast $tRNA^{Phe}$ revealed by recent refinement analysis of x-ray diffraction data at 2.5 Å resolution. The gross features of the molecule observed in the unrefined version have been largely confirmed and a number of new features found. The unique role of the ribose 2′ hydroxyl groups in maintaining a series of nonhelical conformations in this RNA molecule has become apparent. Many of these features are a direct consequence of the geometry of the ribose phosphate backbone of RNA molecules, and these may also be found in structured regions of other RNA species as well. Special attention has been directed toward two conformational motifs revealed by this analysis. These include the striking similarity between the TψC and anticodon hairpin turns in the polynucleotide chain, which are stabilized by the participation of uridine in the U turn. In addition, there is frequent occurrence of an arch conformation in the polynucleotide chain which is stabilized by hydrogen bonds from 2′ hydroxyl residues to phosphate groups across the base of the arch. The importance of the 2′ hydroxyl interactions in defining tertiary structure is illustrated by the fact that, in the nonhelical regions, almost half of the ribose residues are involved in O2′ hydrogen-bonding interactions which stabilize the conformation of the molecule.

(27)

Two regions which may have considerable functional significance during protein synthesis are described in detail. One involves the joining of the TψC and D loops, which may undergo conformational change in the ribosome during protein synthesis. The other region is the anticodon, which seems conformationally poised, ready to interact with a single-stranded polynucleotide messenger RNA. Analysis of this end of the molecule suggests ways in which the anticodon may interact with the message, although as yet not enough is known to understand how two tRNA molecules interact with adjoining codons on the message. The next goal in this research effort is clearly that of trying to relate the detailed structural conformation of tRNA molecules to their important biological functions in the transmission of genetic information during protein synthesis.

References and Notes

1. For a recent review of tRNA sequences, chemical modifications, and solution chemistry, see A. Rich and U. L. Rajbhandary, *Annu. Rev. Biochem.* **45**, 805 (1976).
2. For a convenient compilation, see B. G. Barrell and B. F. C. Clark, *Handbook of Nucleic Acid Sequences* (Joynson-Bruvvers, Oxford, 1974).
3. S. H. Kim, J. L. Sussman, F. L. Suddath, G. J. Quigley, A. McPherson, A. H. J. Wang, N. C. Seeman, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4970 (1974).
4. R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, A. Zamir, *Science* **147**, 1462 (1965).
5. The notation for the nucleotide bases is as follows: A, adenosine; T, thymidine; C, cytidine; U, uridine; G, guanosine; ψ, pseudouridine; D, dihydrouridine; and Y, highly modified purine. Also, m denotes methyl and poly(rA) is polyriboadenylate.
6. The early crystallization work is reviewed by F. Cramer, *Prog. Nucleic Acid Res. Mol. Biol.* **11**, 391 (1971).
7. S. H. Kim, G. J. Quigley, F. L. Suddath, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 841 (1971).
8. T. Ichikawa and M. Sundaralingam, *Nature (London)* **236**, 174 (1972); A. D. Mirzebekov, D. Rhodes, J. T. Finch, A. Klug, B. F. C. Clark, *ibid.* **237**, 27 (1972); S. H. Kim, G. Quigley, F. L. Suddath, A. McPherson, D. Sneden, J. J. Kim, J. Weinzierl, A. Rich, *J. Mol. Biol.* **75**, 421 (1973).
9. S. H. Kim, G. J. Quigley, F. L. Suddath, A. McPherson, D. Sneden, J. J. Kim, J. Weinzierl, A. Rich, *Science* **179**, 285 (1973).
10. S. H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. Wang, N. C. Seeman, A. Rich, *ibid.* **185**, 435 (1974).
11. J. D. Robertus, J. E. Ladner, J. T. Finch, D. Rhodes, R. S. Brown, B. F. C. Clark, A. Klug, *Nature (London)* **250**, 546 (1974).
12. J. E. Ladner, A. Jack, J. D. Robertus, R. S. Brown, D. Rhodes, B. F. C. Clark, A. Klug, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4414 (1975).
13. G. J. Quigley, N. C. Seeman, A. H. J. Wang, A. Rich, *Nucleic Acid Res.* **2**, 2329 (1975).
14. J. L. Sussman and S. H. Kim, *Biochem. Biophys. Res. Commun.* **60**, 89 (1976).
15. J. E. Ladner, A. Jack, J. D. Robertus, R. S. Brown, D. Rhodes, B. F. C. Clark, A. Klug, *Nucleic Acid Res.* **2**, 1629 (1975).
16. C. Stout, H. Mizuno, J. Rubin, T. Brennan, S. Rao, M. Sundaralingam, *ibid.* **3**, 1111 (1976).
17. J. L. Sussman and S. H. Kim, *Science* **192**, 848 (1976).
18. R. Diamond, *Acta Crystallogr. Sect. A* **27**, 436 (1971).
19. R. Huber, D. Kukla, W. Bode, P. Schwager, K. Bartels, J. Diesenhofer, W. Steigemann, *J. Mol. Biol.* **89**, 73 (1974).
20. $R = \Sigma \| F_o | - | F_c \| / \Sigma | F_o |$, where F_o and F_c are observed and calculated structure factors.
21. The R values are based on data with an intensity greater than two estimated standard deviations above background.
22. T. F. Koetzle, *Acta Crystallogr. Sect. B* **29**, 1746 (1973).
23. F. H. C. Crick and J. D. Watson, *Proc. R. Soc. London Ser. A* **223**, 80 (1954).
24. Simple stereoscopic viewers may be purchased from Taylor-Merchant Corp., New York, and Hubbard Scientific Co., Northbrook, Ill. A professional metal viewer may be purchased from Abrams Instrument Co., Lansing, Mich.
25. S. Arnott, D. W. L. Hukins, S. D. Dover, *Biochem. Biophys. Res. Commun.* **48**, 1392 (1972).
26. F. H. C. Crick, *J. Mol. Biol.* **19**, 548 (1966).
27. Helix parameters were calculated by N. C. Seeman as described in J. M. Rosenberg, N. C. Seeman, R. O. Day, A. Rich, *Biochem. Biophys. Res. Commun.* **69**, 979 (1976).
28. S. H. Kim and J. L. Sussman, *Nature (London)* **260**, 645 (1976).
29. K. Watanabe, T. Oshima, M. Saneyoshi, S. Nishimura, *FEBS Lett.* **43**, 59 (1974).
30. U. Thewalt and C. E. Bugg, *J. Am. Chem. Soc.* **94**, 8892 (1972).
31. V. A. Erdman, M. Sprinzl, O. Pongs, *Biochem. Biophys. Res. Commun.* **54**, 942 (1973); D. Richter, V. A. Erdmann, M. Sprinzl, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3326 (1974).
32. W. Wintermeyer and H. G. Zachau, *FEBS Lett.* **58**, 306 (1975).
33. A. Rich, D. R. Davies, F. H. C. Crick, J. D. Watson, *J. Mol. Biol.* **3**, 71 (1961).
34. H. Grosjean, D. G. Soll, D. M. Crothers, *ibid.* **103**, 499 (1976).
35. A. Rich, in *Ribosome*, M. Nomura, A. Tissieries, P. Lengyel, Eds. (Cold Spring Harbor Laboratory, New York, 1974), p. 171.
36. W. Fuller and A. Hodgson, *Nature (London)* **215**, 817 (1967).
37. C. K. Johnson, *ORTEP* (Report ORNL-3794, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1965).

38. We thank A. H. J. Wang, M. Teeter, and N. C. Seeman for helpful discussions and P. Bellin for expert technical assistance. This research was supported by grants from the National Institutes of Health, the National Science Foundation, the National Aeronautics and Space Administration, and the American Cancer Society.

(28)

Right-handed and Left-handed Double-helical DNA: Structural Studies

A.H.-J. WANG, S. FUJII, J.H. VAN BOOM,* AND A. RICH

*Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; *Gorlaeus Laboratories, University of Leiden, 2300 RA Leiden, The Netherlands*

The availability of chemically synthesized oligonucleotides of defined sequence has changed the nature of nucleic acid structural studies from fibers to single-crystal X-ray analyses. Studies on nucleic acid fibers have intrinsic limitations because their X-ray diffraction patterns are of only limited resolution; the amount of structural information one can obtain from them is correspondingly limited. Solution of a fiber diffraction pattern makes it impossible to obtain many interesting structural details.

Unlike fibers, single crystals often diffract to atomic (~1 Å) or near-atomic resolution. Solution of a crystal structure makes it possible to visualize the molecular structure directly without any assumption or model building. For many years there have been single-crystal X-ray studies of nucleic acid bases, nucleosides, and monomeric nucleotides. Several single-crystal diffraction studies have been carried out in which the purine-pyrimidine base pairs have been shown to have a variety of hydrogen bonding (for review, see Voet and Rich 1970). However, the first visualization of the double helix at atomic resolution occurred about 10 years ago with the solution of the structures of rApU (Rosenberg et al. 1973; Seeman et al. 1976) and rGpC (Day et al. 1973; Rosenberg et al. 1976), which were found to form right-handed double-helical fragments of the type that had been anticipated for RNA molecules. This work on ribonucleotide fragments laid the groundwork for a more extensive, subsequent analysis of deoxyoligonucleotides and even DNA-RNA hybrids. In this paper we review some of the more recent work dealing with DNA structures from chemically synthesized oligonucleotides.

DNA from A to Z

DNA is known to form a number of right-handed double-helical structures. The two major types, which have been known for almost 30 years, were first visualized by X-ray diffraction studies of DNA fibers (Franklin and Gosling 1953). If the fibers were allowed to dry, they produced an A-type diffraction pattern, and if the fibers remained hydrated, the pattern was a B-type. There are a number of differences between A-type and B-type conformations: the tilting of the base pairs relative to the helix axis, the displacement of the base pairs away from the helix axis, and, finally, the pucker of the furanose ring. The latter of these is illustrated in Figure 1, which shows the five-member ring of the deoxyribose sugar drawn so that the plane defined by the atoms C1′, O1′, and C4′ is horizontal. The two other ring atoms, C2′ and C3′, can be oriented in two different classes of positions. Those in which the C2′ atom is above the horizontal plane are called C2′-*endo*, whereas if atom C3′ is above that, it is called C3′-*endo*. The major effect of these differences in sugar pucker is that they alter the separation between the phosphate groups attached to the sugar by more than 1 Å. In the C3′-*endo* conformation, the phosphate groups are closer together, whereas in the C2′-*endo* conformation, they are further apart. This results in the introduction of an elastic element in the backbone of polynucleotides. These differences in distance are widely used in naturally occurring nucleic acid structures. For example, in the three-dimensional structure of yeast phenylalanine transfer RNA ($tRNA^{Phe}$), a majority of the 76 nucleotides are in the C3′-*endo* conformation. However, 8 of them are in the C2′-*endo* conformation (Quigley and Rich 1976), and they occur in specific positions, such as where the polynucleotide chain has to be extended to ac-

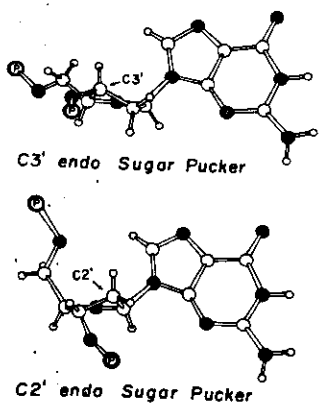


Figure 1. Diagram illustrating the two major families of furanose ring pucker in ribonucleotides and deoxynucleotides. The C1′-O1′-C4′ atoms are drawn in a horizontal plane. Atoms above that plane are in the *endo* conformation. Note that in the C2′-*endo* conformation the phosphate groups are further apart than in the C3′-*endo* conformation.

commodate or surround an intercalating base. They also occur when the chain is stretched out, e.g., in the segments of the backbone connecting the acceptor stem to the dihydro U stem. The energy difference between these two different types of pucker is relatively small, and accordingly one might anticipate that a molecule would have a mixture of puckers, especially in solution, where the molecule is free to vibrate. Generally, ribonucleotides tend to have the C3′-*endo* conformation, mostly because of the steric effect of O2′. However, DNA molecules can have either a C2′-*endo* or C3′-*endo* conformation. In the A-DNA family, the C3′-*endo* conformation predominates, whereas in B-DNA there is mostly C2′-*endo* pucker.

The DNA families are of types A, B, C, and D, which generally include right-handed double helices with Watson-Crick pairing (Leslie et al. 1980; Wells et al. 1980). The more recently discovered, left-handed Z-DNA helices (Wang et al. 1979, 1981) also occur in a family of conformations, as discussed below. In the last few years, a number of single crystals of helical oligonucleotides have been formed containing 4 bp or more in the A, B, and Z conformations, and their three-dimensional structures have been determined by X-ray diffraction analysis. Table 1 lists the characteristics of several single crystals of DNA fragments.

Molecules with Alternating Conformations

Although we tend to describe DNA structures as being either A-DNA or B-DNA, it is important to recognize that these represent only idealized simplifications of molecules and that the actual conformation a particular molecule adopts reflects to varying extents an individual sequence. In a crystal lattice, we generally observe the lowest energy conformation for that sequence. The deviations from ideality may be considerable, as for example, if we consider the structure of the octamer d(GpGpCpCpGpGpCpC) (Wang et al. 1982a). Although this molecule is predominantly in the A-DNA form, it actually has a central segment containing the two different families of sugar pucker so that the molecule has an interesting mixture of A-DNA and B-DNA features. Figure 2 illustrates some of the features of this structure in comparison to A- and B-DNAs. It was found that the molecules in the crystal underwent a small temperature-dependent conformational change so that two different structures were solved at −8° and −18°C. The upper part of Figure 2 shows a van der Waals diagram of the molecule viewed from the side. The center two panels show the conformations obtained at the two different temperatures, whereas the left and right panels are idealized models of A-DNA and B-DNA, respectively, constructed using the same sequence. The bottom of the figure shows views down the axis of skeletal models.

There are several differences between A-DNA and B-DNA. In B-DNA the bases occupy the center of the molecule, and they are almost perpendicular to the axis. In A-DNA the base pairs are removed from the central axis of the molecule, and they have a considerable tilt in the opposite direction. In B-DNA the rise per residue along the axis is 3.34 Å, which is the thickness of the unsaturated purine-pyrimidine base pair. Due to the tilting of the bases in A-DNA, the rise per residue is 2.56 Å along the axis. The diameter of the A-DNA molecule is slightly larger. In A-DNA the minor groove is flat and wide, whereas the major groove is deep and extends through the central axis of the molecule. The tilt of the base pairs brings the phosphate groups closer together on opposite strands across the major groove of A-DNA than they are in B-DNA. These differences are indicated in Figure 2 by the arrow at the right of each structure, showing the distance along the helix axis between the phosphate groups at either end of the molecule. The vertical distance between the phosphates

Table 1. Various Types of Single-crystal DNA Structures

Compound	Cell constants a Å	b Å	c Å	Space group	Resolution (Å)	Current R value	Reference
A-type							
d(GGCCGGCC)							
(−8°C)	42.06	42.06	25.17	$P4_12_12$	2.25	0.17	Wang et al. (1982a)
(−18°C)	40.51	40.51	24.67	$P4_12_12$	2.25	0.16	Wang et al. (1982a)
d(CCCCGGGG)	42.1	42.1	25.1	$P4_12_12$			Wang et al. (1982a)
r(GCG)d(TATACGC)	24.20	43.46	49.40	$P2_12_12_1$	2.0	0.16	Wang et al. (1982b)
d(GGTATACC)	44.97	44.97	41.76	$P6_1$	2.25	0.18	Shakked et al. (1981)
d(ᵗCCGG)	41.4	41.4	26.7	$P4_12_12$	2.1	0.21	Conner et al. (1982)
B-type							
d(CGCGAATTCGCG)	24.87	40.39	66.20	$P2_12_12_1$	1.9	0.18	Wing et al. (1980)
d(CGTACG): daunomycin complex	27.92	27.92	52.89	$P4_12_12$	1.5	0.20	Quigley at al. (1980)
Z-type							
d(CGCGCG)	17.88	31.55	44.58	$P2_12_12_1$	0.9	0.13	Wang et al. (1979)
d(m⁵CGm⁵CGm⁵CG)	17.76	30.57	45.42	$P2_12_12_1$	1.3	0.16	S. Fujii et al. (in prep.)
d(CGCGCGCG)	31.27	31.27	43.56	$P6_5$	1.6	0.19	S. Fujii et al. (in prep.)
d(CGCATGCG)	30.90	30.90	43.14	$P6_4$	2.5	0.16	S. Fujii et al. (in prep.)
d(CGCG)	31.25	31.25	44.06	$P6_5$	1.5	0.19	Crawford et al. (1980)
d(CGCG)	19.50	31.27	64.67	$C222_1$	1.5	0.20	Drew et al. (1980)

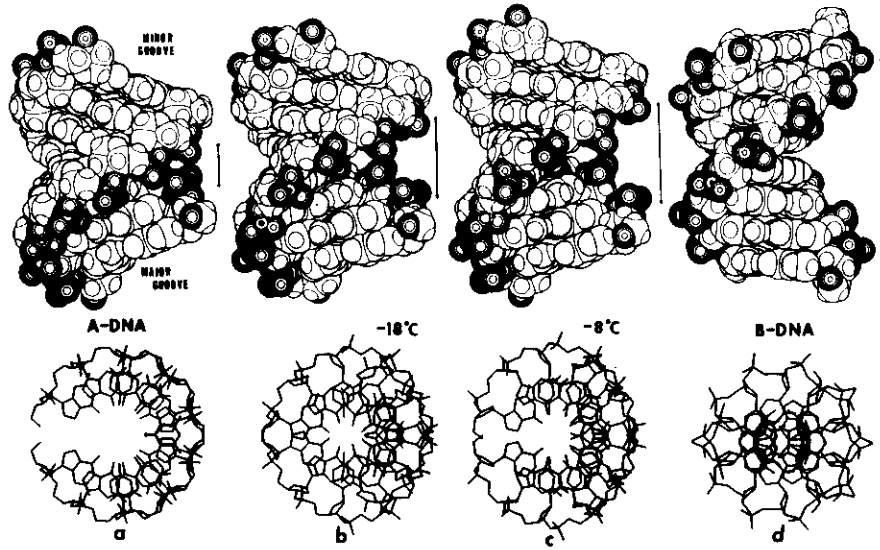


Figure 2. Molecular structure of d(GGCCGGCC). The center diagrams (*b* and *c*) show the form of the octamer that is observed experimentally at −18°C and at −8°C, respectively. Models of pure A-DNA (*a*) and B-DNA (*d*) are shown on either side for comparison. The upper diagrams are van der Waals drawings of the four different DNA structures. The helix axis is vertical, and a horizontal twofold axis is located in the plane of the paper. Arrow to the right of each structure represents the vertical distance between the terminal phosphates of the two chains. The phosphorous and oxygen atoms have a heavier shading than the other atoms. The tilt to the bases from the horizontal is 19° in A-DNA and −7° in B-DNA, whereas base pairs in *b* and *c* are tilted approximately 14° and 12°, respectively. At the bottom, skeletal views down the helix axes are shown for the four structures.

at the end of the two chains is much smaller in A-DNA than in B-DNA.

In the two octamer structures, we find that they have a conformation that is neither A nor B. It is largely A-DNA, but it has some features that place it intermediate between the two. The tilt of the bases is 14° and 12° in the −8°C and −18°C structures, respectively, compared with a tilt of 19° for A-DNA and −7° for B-DNA. Likewise, the arrow at the right of the central two molecules is longer than that seen around the A-DNA model but shorter than that found in B-DNA. The skeletal models at the bottom of Figure 2 show that there is a hole down the axis of A-DNA. That hole is absent in B-DNA because the bases are located on the helix axis. The two crystal structures at −18° and −8°C have a conformation intermediate between A-DNA and B-DNA, although it is closer to A-DNA. The octamer A-DNA end view is only a part of a circle because there are 11 bp per turn. B-DNA has 10 bp per turn, and the end view projection of that molecule looks more circular. The central two octamer structures have an intermediate conformation in which the ends of the chains are closer together than in A-DNA but are not as closed as in B-DNA.

The reason that the octamer structure is not pure A-DNA is found by careful examination of the sugar pucker of successive residues. The numbering of the self-complementary octamer structure goes from G1 to C8 down one strand and then from G1* to C8* along the other strand. Thus, residues G1 and G2 are paired to C8* and C7* of the opposite strands. All four of these residues have conformations close to A-DNA. However, the central segment of 4 bp, C3 through G6, consists of residues that have alternating ring puckers, i.e., alternatively near C2′-*endo*, C3′-*endo*, C2′-*endo*, C3′-*endo*. This difference in ring pucker is reflected in the distances between adjacent phosphates. The P-P distance surrounding residues C3 and C5 are closer to that seen for B-DNA than A-DNA. This is shown in Figure 3, in which the stacking or overlap of successive base pairs are drawn with the heavier lines indicating the base pair on top. In this figure we are looking at a view perpendicular to the base pair. The left column shows the idealized stacking for pure A-DNA, whereas the right-hand column shows the similar stacking for B-DNA. The central column shows the observed stacking of the −8° octamer. Close comparison of the center-column stacking diagrams reveals that they are closer to that found in A-DNA than to B-DNA, although the stacking pattern is not identical. However, the conformation of the sugar-phosphate backbone is alternating for the central 4 bp. This can be seen by looking at the distances between C1′ sugar atoms of adjacent nucleotides along the polynucleotide backbone. The distance C1′-C1′ in

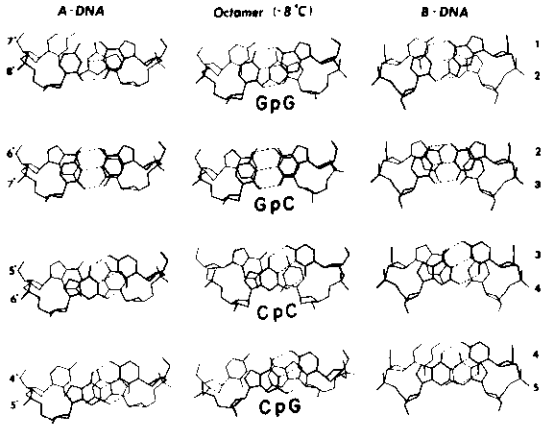


Figure 3. Stacking of successive base pairs is illustrated for idealized A-DNA and B-DNA and the experimental octamer structure d(GGCCGGCC) that was determined at −8°C. The base pairs drawn with the heavier lines are closer to the reader than the base pairs drawn with the thin lines. The sequences listed under the center diagram also apply to the two diagrams on either side. The sequence refers to the right-hand side of the diagram, and the left-hand side has the complementary sequence. In the upper level, the base pair 1-8* is stacked above the base pair 2-7*. In the second level, base pair 2-7* is stacked above base pair 3-6*. It should be noted that although the backbone is regular in the A-DNA and B-DNA columns, it is quite irregular in the central column, which shows the actual structure. This is because the conformation differs considerably from the idealized or averaged conformation, which has been obtained from fiber diffraction studies.

the sequence G2pC3 is shorter than the same distance in C3pC4. Similarly, the C1′-C1′ projection distance in C4pG5 is also shorter than the distance in C3pC4. The backbone is thus alternately compressed and expanded as one goes down the polynucleotide backbone chain. There is a twofold axis between the base pairs C4·G5* and G5·C4*, so the alternating compression and expansion of the backbone continues through the center of the molecule. These differences are also reflected in differences in the twist angle between adjacent base pairs. This is the angle between the line C1′-C1′ of 1 bp with the next base pair. The variations in the twist angle are plotted in Figure 4 as a function of sequence. It can be seen that they vary from 45° to 16°, and these are to be compared with a constant twist angle of 33°, which is found in the idealized A-DNA structure, and 36° found throughout the B-DNA structure.

We do not know if this alternating A-DNA conformation is found in a particular sequence. It is interesting that the octamer d(CCCCGGGG) listed in Table 1 crystallizes in virtually the identical lattice as the −8°C and −18°C octamers. Furthermore, examination of sample precession diffraction patterns from crystals of that octamer shows that it is virtually identical to that of the octamer with the sequence d(GGCCGGCC). These two octamers both have the sequence CCGG in their central 4 bp. This leaves open the possibility that this alternating A conformation may be favored by a sequence involving two pyrimidines followed by two purines, or it may be favored by the sequence CCGG. If so, it is possible that this is what is recognized by the restriction endonuclease *Hpa*II and *Msp*I, which cleave this particular sequence. At the present time, we must have an open mind regarding the relationship of a conformation found in the analysis of a single-crystal structure and the conformation found in solution or in vivo. One cannot argue that the crystallization environment per se is one that is very far removed from that which is found in vivo. The molecule is heavily hydrated, and for

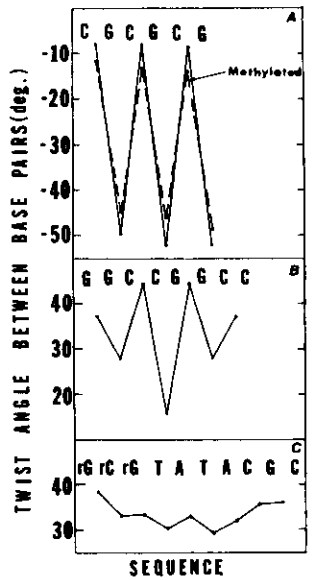


Figure 4. Twist angle between adjacent base pairs is plotted for four different crystal structures. The points are positioned between the bases listed in the sequence, and each point represents the twist angle between the 2 bp on either side of the point. The twist angle is the angle between adjacent lines connecting C1′ atoms of nucleoside base pairs. (*A*) Twist angles for methylated and nonmethylated Z-DNA crystals. (*B*) Twist angles for the DNA octamer d(GGCCGGCC); note that in the center of the sequence, the twist angles vary widely as different conformations are found in the middle of the molecule. (*C*) Twist angles of the DNA-RNA hybrid decamer structure; these twist angles are relatively constant. By comparison, the twist angles for idealized B-DNA would be 34° between all adjacent base pairs, and for A-DNA the angle would be 36°.

the octamer, more than 80 water molecules have been found that surround the molecule in all directions, except where adjacent molecules interact with it to form the contacts that build up the lattice. Mixed puckers of sugar residues in nucleic acid structures are not rare. These were first found by Sobell and his colleagues in the study of intercalative systems (Jain et al. 1977; Tsai et al. 1977). Intercalators are frequently accommodated by adopting mixed sugar puckers in the nucleotides surrounding the intercalator molecule, as mentioned above in yeast $tRNA^{Phe}$. In general, what is found in a crystal lattice is the lowest energy state of the system. In the present sequences, this is represented by a conformation in which adjacent residues adopt slight conformational modifications.

The evidence concerning the actual conformation of DNA found in biological systems is fragmentary, at best. For natural DNA, the fiber diffraction patterns tend to emphasize only those areas that are periodic, and it blurs out segments that are less than fully regular. Diffraction studies of DNA in bacteriophage (North and Rich 1961) or in spermatozoa also suffer from the same difficulty. The results of many crystallographic studies and the comparison of these studies with what is observed in solution in general tends to support the concept that what is observed in crystal structures represents a stable conformation that is likely to be found in solution as well.

Lattice Interactions

The manner in which these A-DNA octamer fragments form a lattice is quite interesting. The flat, shallow minor groove is used as one of the component building blocks for organizing the molecules into a lattice. In particular, the planar C-G base pair at the end of the octamer is found to stack on the flat surface in the minor groove of an adjacent octamer in a manner shown in Figure 5A. In this stereodiagram, the octamer itself is drawn without any shading, but the terminal 2 bp of adjacent molecules are drawn with shaded atoms. It clearly shows the manner in which the adjacent molecules use the flat surface of the minor groove to stabilize, through van der Waals forces, the interactions between adjacent molecules to build up a lattice. Attention has been drawn to the fact that the size and shape of the purine-pyrimidine base pair in this case is rather similar to the size and shape of the benzo[a]pyrene nucleus, which is a powerful carcinogen (Wang et al. 1982a). That carcinogen is known to bind covalently to the N2 position of guanine (Jeffrey et al. 1976). In the present lattice, the planar purine-pyrimidine base pair rests in a position close to the guanine N2. Thus, this mode of interaction may also be a model for the interaction of planar carcinogen molecules with DNA.

In surveying the crystals that form in an A-DNA conformation in Table 1, we might ask why these have crystallized as A-DNA in contrast to the dodecamer studied by Wing et al. (1980), which crystallizes as B-DNA. The feature that is characteristic of all of the nucleotide fragments that so far have crystallized as A-DNA is the fact that they all contain the sequence GG. Is it possible that this dinucleotide favors an A-DNA-type stacking, which may introduce a tendency to form A-DNA in regions nearby? It is clear that additional oligonucleotide structures will have to be solved before it will be possible to answer a question of this type.

An A Conformation in a DNA-RNA Hybrid

During DNA replication, a single strand of DNA is used as a template for assembling complementary deoxynucleotides. For one of the two strands of the double helix, DNA replication takes an unusual form, Okazaki fragments (Ogawa and Okazaki 1980), in which segments of DNA are assembled and later joined to form an intact final strand. The synthesis of DNA Okazaki fragments is primed at the 5′ end by a small number of ribonucleotides, which initiate the synthesis. Synthesis is then continued by covalently joined deoxynucleotides. In an attempt to learn something of the structure of this initiating complex and to learn more generally something about DNA-RNA hybrid molecules, the structure of an oligonucleotide was solved that contains three ribonucleotides at the 5′ end connected covalently to seven deoxynucleotides (Wang et al. 1982b). The compound synthesized was r(GCG)d(TATACGC). This molecule is self-complementary and forms two DNA-RNA hybrid segments with 3 bp surrounding a central region of 4 bp of double-helical DNA. All three parts of the molecule adopt a conformation that is close to that seen in A-DNA or in the 11-fold RNA double helix.

Figure 6 shows three side views of the hybrid molecule. In these diagrams the ribo components are shaded and the 2′-hydroxyl groups are black. These three views, which are 90° apart, illustrate the typical A-DNA geometry of the complex. The major groove is deep and narrow and the minor groove is broad and shallow. The base pairs have an average twist angle of 33° but, as is shown in the bottom panel of Figure 4, there is some variation among the residues even though the differences are not very large. The base pairs have a rise along the axis of 2.6 Å, and in this helix there are 10.9 residues per turn. The base pairs are considerably removed from the axis of the molecule and are tilted approximately +20° from the axis of the molecule. Figure 7 shows a stereoscopic view down the axis of the molecule. There is an empty space nearly 2 Å in diameter along the axis of the molecule. It is interesting that the hybrid molecule has a fairly regular geometry even though it is made of components containing both ribonucleotides and deoxynucleotides. The molecule has an almost perfect A-type geometry with all of the ribose and deoxyribose sugars in the C3′-*endo* conformation. There are no structural discontinuities between the hybrid helix and the DNA helix itself. There is a slight discontinuity associated with the structure of the central 4 bp containing the sequence d(TATA). In contrast to the six hybrid GC base pairs at either end, which

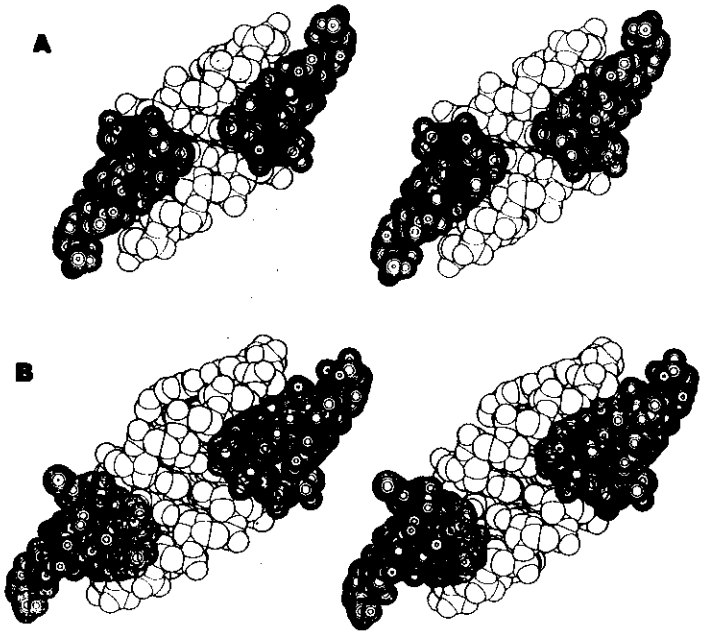


Figure 5. Stereodiagrams illustrating the interaction between molecules in the crystal lattices of the octamer d(GGCCGGCC) (*A*) and the decamer hybrid r(GCG)d(TATACGC) (*B*). In both cases, a complete molecule is shown with unshaded atoms. The shaded atoms represent the base pairs at the ends of the adjoining molecules. Two or three base pairs from the adjoining molecules are shown with shaded atoms in order to illustrate the packing of the terminal, planar base pairs on the flat minor-groove surface of the unshaded molecule in the center.

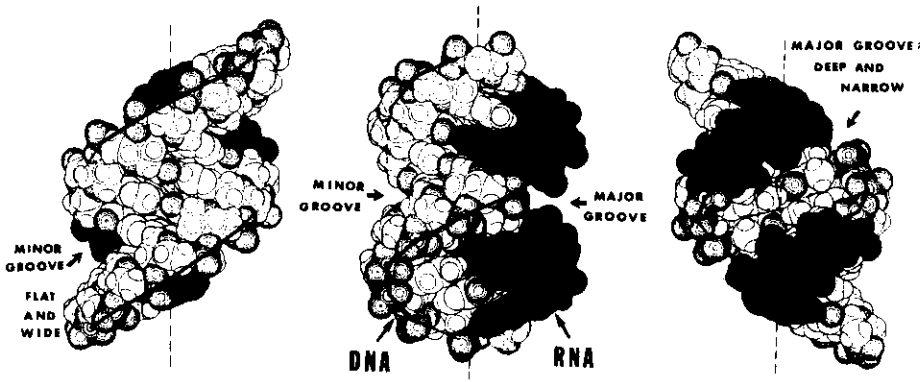


Figure 6. Space-filling drawing of the DNA-RNA hybrid r(GCG)d(TATACGC). The molecule is shown in three different orientations 90° apart from each other about the dashed vertical axis. At the left the view is into the broad, flat minor groove of the molecule. Phosphate groups on opposite strands are 8.6 Å apart across the minor groove. The heavy black lines are drawn from phosphate to phosphate to show the flow of the polynucleotide chains. The ribonucleotides are shaded, and the ribose 2′ oxygen atoms are black. The center figure, rotated 90°, shows both the minor groove on the upper part of the molecule and the major groove at the lower right. The shaded ribonucleotide backbone segments are close to each other on either side of the major groove. The bases are tilted 19° from the vertical axis. In the center view, the tilt reverses itself in going from the upper part to the lower part of the molecule. The view on the right is rotated 90° from the center of the molecule, and the view is down into the deep and narrow major groove of the molecule. The phosphate groups are 4.7 Å away from each other across this groove. Oxygen atoms are solid circles; nitrogen atoms are stippled; phosphorous atoms have spiked circles.

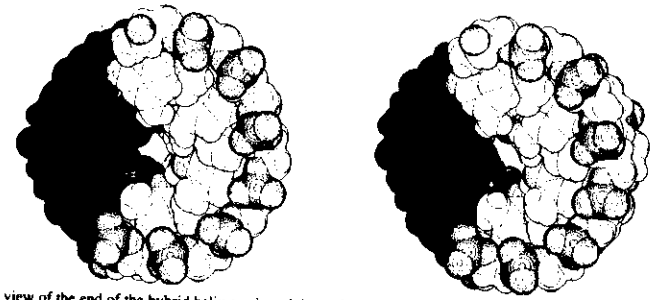


Figure 7. Stereo view of the end of the hybrid helix as viewed down the helix axis. The axis of the molecule has an empty space approximately 2 Å in diameter. The base pairs form a prominent, right-handed, staircaselike arrangement with the phosphate groups tipped into the major groove of the helix so that they form an interrupted railing for the staircase. Note that the ribonucleotide 2′ oxygen atoms are located on the outside of the molecule.

have normal hydrogen bond lengths, the four AT base pairs have somewhat longer NH–O hydrogen bond lengths than are normally seen. In addition, there are some discontinuities in the orientation of the helix axis in the d(TpA) segments.

These discontinuities in the TATA region may be an expression of structural features that are an inherent part of the TATA sequence possibly associated with stacking modifications. This might be of physiological interest, since the base sequence TATA is one of the common signals used in the initiation of transcription by RNA polymerase.

The major systematic difference between RNA and DNA is the presence of a 2′-hydroxyl group on the ribonucleotide sugar. In the ribocytidine residues, these appear to play a conformational role by forming hydrogen bonds to water molecules, which in turn are hydrogen-bonded to the O2 of the same cytidine residue. This is illustrated in the stereodiagram of Figure 8 for the two ribo-C residues in the sequence. The 2′-hydroxyl group appears to play a significant role in stabilizing the conformation of the ribonucleotide segment through the use of a bridging water molecule. The hydroxyl groups found in the four riboguanosine residues are also hydrogen-bonded to water molecules, but these water molecules are bonded to adjacent molecules and are used in stabilizing the lattice. However, in the dinucleoside monophosphate rGpC (Seeman et al. 1976), the 2′-hydroxyl groups of the guanosine sugars are hydrogen-bonded to a water molecule, which in turn is hydrogen-bonded to the N3 atom of guanine. It is thus reasonable to believe that such bridging water molecules might be a constant feature of RNA double helices in solution. This is an example of what is undoubtedly a generalization in the structure of nucleic acids—namely, that the total conformation of the molecule is determined in part by the disposition of the various components of the polynucleotide chain but is also determined in part by water molecules, especially those in the primary hydration shell, which may stabilize particular conformational features.

The packing of the hybrid A-DNA segments in the crystal lattice uses the same type of molecular interactions as those found in the organization of the deoxy octamer lattice described above. Figure 5B shows a stereodiagram of the stacking interactions using space-filling models. The hybrid decamer is shown with unshaded atoms, and 3 bp of the stacking molecules adjoining it are shown with shaded atoms. It can be seen that the packing motif is one in which the terminal planar purine-pyrimidine base pair stacks on the flat surface of the minor-groove A-type double helix in the

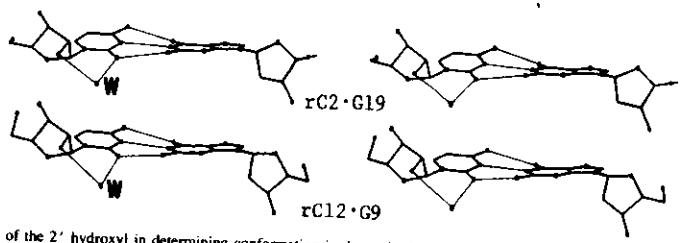


Figure 8. Role of the 2′ hydroxyl in determining conformation is shown in these stereo skeletal drawings of two different guanosine-cytosine base pairs in the hybrid helix. The nucleotides are numbered 1 to 10 on one chain and 11 to 20 on the other chain, both in the 5′ to 3′ direction. The figures show a similar conformation of the cytosine ribose residues at the left in both base pairs. A bridging water molecule (W) at the lower left in both diagrams is hydrogen-bonded to both the 2′-hydroxyl group of the ribose and the O2 of cytosine. It is likely that this bonding helps to stabilize the orientation of the base relative to the sugar.

same way as seen for the octamer stacking in Figure 5A. This stacking motif appears to be used in all of the A-type crystal structures described in Table 1.

At the present time, six different crystals have been identified in which A-DNA fragments are formed in crystalline lattices. This is to be compared with the fact that only one pure B-DNA crystal has formed a lattice. The reason for the higher frequency of A-type crystals is not that A-DNA is intrinsically more stable than B-DNA. This illustrates the somewhat biased selection found in crystallographic studies. The only conformations that can be analyzed are those that actually form crystals, i.e., those that have stable intermolecular interactions strong enough to build a three-dimensional lattice. The high frequency of A-type conformations relative to B derives from the stability of the stacking of a planar purine-pyrimidine base pair on the broad, shallow minor groove of A-DNA. These interactions are conducive to building a lattice. In B-DNA, the grooves are considerably deeper, and this type of intermolecular stacking is less probable. The opportunities to build up lattices containing B-DNA are apparently much less frequent than those for building lattices with fragments of A-DNA.

The uniformity of the hybrid molecule in forming an A-type structure is undoubtedly due to the influence of the ribonucleotides in the structure. For ribonucleotides, the C3′-*endo* pucker is more stable than the C2′-*endo* conformation. Although there is a slight preference in favor of a C2′-*endo* conformation in deoxynucleotides, the difference in energy is not sufficiently great so that the DNA can adopt a C3′-*endo* conformation with relative ease. In the hybrid molecule, it is quite likely that the C3′-*endo* conformation of the central DNA duplex segment is due to the influence of the hybrid regions at the ends where the C3′-*endo* conformation is maintained by the three ribonucleotides. This interpretation is strengthened by a nuclear magnetic resonance study of the same molecule composed entirely of deoxynucleotides (J.-R. Mellema et al., in prep.). That study revealed that the molecule has a typical B-DNA conformation when it is composed only of deoxynucleotides, but inserting the three ribonucleotides makes the molecule shift to an A conformation, which is maintained even in solution.

It thus appears as if the RNA backbone has forced its geometry upon the DNA backbone to form a stable A-type structure. In this conformation the DNA-RNA hybrid remains distinctive because of the 2′ oxygen atoms in the ribonucleotide segment that are positioned on the outer surface of the molecule. This means that proteins or other molecules interacting with this structure can readily detect the difference between an RNA fragment forming an A helix and a deoxynucleotide segment with the same conformation. What is unknown at present is how far into a DNA duplex a DNA-RNA hybrid helix would propagate an A-type conformation. This question will have to be answered by the analysis of similar molecules with longer DNA duplex regions.

In the octamer d(GGCCGGCC) and the DNA-RNA hybrid, we find both of them in an A-DNA conformation, but we explain it using two different mechanisms. For the pure deoxy octamer, we have speculated that it may be the sequence of stacking of two adjacent deoxyguanosine residues that favors an A sequence. In the DNA-RNA hybrid, it is clearly the influence of the ribonucleotide that favors the C3′-*endo* conformation, which forces the entire molecule to adopt an A conformation. Throughout, it should be recognized that the A conformation and B conformation are always in dynamic equilibrium and that the actual conformation seen in a crystal structure is simply one of these two stable energy states. Due to the fact that the crystal is carrying out a sampling operation, it leaves open the possibility that other conformations in equilibrium with this may be represented in solution to a significant extent. The strength of single-crystal X-ray analyses is that they provide firm information about molecular conformation. The limitation of the analysis is that it focuses on one conformation and does not provide information about the other conformations that are in equilibrium with it in solution. It is always important to remember this point in thinking about the overall conformation of nucleic acids in biological systems.

DNA Twists the Other Way: Left-handed Z-DNA

The conformational changes in going from B-DNA to A-DNA are small in comparison to the molecular rearrangement that occurs when DNA adopts the left-handed Z-DNA conformation. This left-handed structure was found in an X-ray diffraction analysis of a single crystal of a hexanucleoside pentaphosphate with the sequence $(dC\text{-}dG)_3$ (Wang et al. 1979). Its molecular structure has many differences compared with the familiar right-handed B-DNA. The structure is favored by sequences that have an alternation of purines and pyrimidines. The reason for this is seen in Figure 9, which compares the conformation of deoxyguanosine in Z-DNA with that in B-DNA. As pointed out above, B-DNA has a C2′-*endo* sugar pucker, and A-DNA (as well as RNA structures) has a C3′-*endo* sugar pucker. In Z-DNA the situation is even more complex, since both puckers are found within the same structure in a systematic manner. The deoxyguanosine residues have the C3′-*endo* pucker (Fig. 9), whereas the deoxycytidine residues have the C2′-*endo* pucker. A further significant change is the rearrangement in the orientation of the purine residue relative to the sugar-phosphate chain. In Z-DNA the guanine residues adopt the *syn* conformation, whereas the cytosines have an *anti* conformation. This is in contrast to B-DNA, where all residues have the *anti* conformation. Theoretical analyses (Haschemeyer and Rich 1967), as well as physical studies of *syn* and *anti* conformations in solution, have shown that they are energetically equivalent in purine-containing nucleotides, but pyrimidine-containing nucleotides have a preference for the *anti* conformation. There is some energy loss in having a pyrimidine adopt an *anti* conformation, but it should be pointed out that this loss is not very large. In Z-DNA the asymmetric unit is two nucleotides, in contrast to the single nucleotide found in B-DNA. The two nucleotides with a *syn* purine and an *anti* pyrimidine are the asymmetric unit in the molecule. The Z-DNA helix has 12 bp per turn of the helix, which adopts a length of 44.6 Å. This is in contrast to the 10 bp per turn occupying a distance of 34 Å in right-handed B-DNA. The diameter of the Z-DNA helix is only 18 Å compared with the somewhat thicker 20 Å found in B-DNA.

In Z-DNA it was possible to assign precise parameters associated with the structure almost immediately upon its discovery. The reason, as shown in Table 1, is that the crystal diffracted to a resolution of 0.9 Å (i.e., atomic resolution); thus, every atom is visualized in the electron density map. This includes not only the 6 bp of the oligonucleotide, but also the large number of water molecules as well as spermine and magnesium ions. In the crystal lattice, the hexamer Z-DNA fragments are stacked upon each other in such a manner that they make a structure in which the base pairs have a stacking that runs continuously through the crystal parallel to the *c* axis, and around it the sugar-phosphate chains form a continuum, except that they are interrupted by the absence of a phosphate group every six residues. The structure has enough regularity so that it is visualized in the form of a continuous helix. The Z-DNA helix is shown in Figure 10A, in which the helical groove is shaded and the heavy black line that forms a zigzag around the groove shows the position of the phosphate groups along the serrated edge of the chain. There is only one groove in Z-DNA compared with two in B-DNA. The Watson-Crick base pairs that form the outer convex surface of the molecule in Z-DNA correspond to the major groove of the B-DNA helix.

From a topological point of view, one cannot go from B-DNA to Z-DNA by simply turning the helix around the other way. In addition to the rotation of the helix in the opposite direction, the base pairs must "flip over," since they have an orientation relative to the backbone

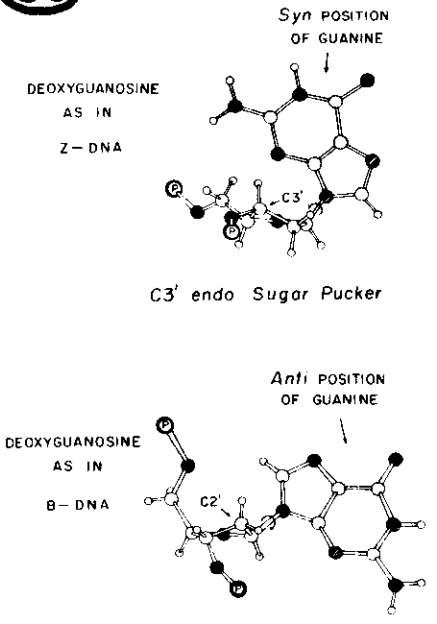


Figure 9. Diagram illustrating the conformation of deoxyguanosine in Z-DNA and B-DNA. In Z-DNA (*top*) the guanosine residues have the C3′-*endo* sugar pucker, and the guanine base is found in the *syn* conformation. Both of these are different from the conformation found in B-DNA (*bottom*). However, the deoxycytidine residues of Z-DNA have the C2′-*endo* sugar pucker and the *anti* position of cytidine.

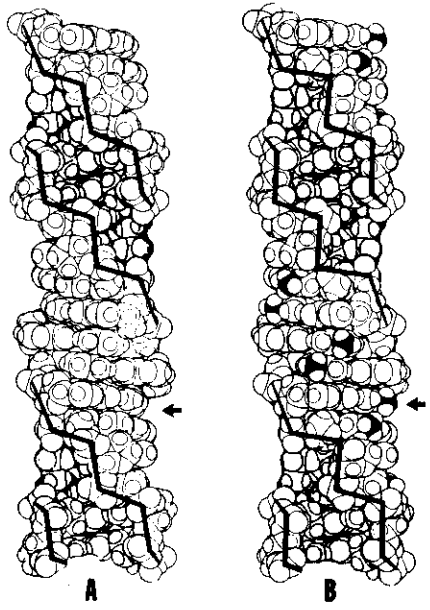


Figure 10. Two van der Waals drawings showing the structure of Z-DNA in both its unmethylated (*A*) and methylated (*B*) forms as determined in single-crystal structures of $(dC\text{-}dG)_3$ and $(m^5dC\text{-}dG)_3$, respectively. van der Waals drawings are used and the groove in the molecule is shown by the shading. The black zigzag line goes from phosphate group to phosphate group to show the arrangement of the sugar-phosphate backbone. In the methylated polymer (*B*), the methyl groups on the C5 position of cytosine are drawn in black. Arrow illustrates that a depression found in the unmethylated polymer (*A*) is filled by methyl groups in *B*. The methyl group indicated by the arrow is in close contact with the imidazole ring of guanosine above it and the C1 and C2′ atoms of the sugar ring.

that is the opposite of that found in B-DNA. Thus, the transformation from B-DNA to Z-DNA is not a simple one but requires considerable complex modifications.

The transition from B-DNA to Z-DNA was first observed in solution experiments in which it was noted that the CD of poly(dG-dC) nearly inverted when the concentration of NaCl in the solution was raised (Pohl and Jovin 1972). In solution there exists an equilibrium in which both Z-DNA and B-DNA are represented. The equilibrium constant is determined by many factors, including base sequence. It is interesting that the crystals of Z-DNA were formed in a solution that had a low concentration of cations, so that the majority species in the solution was B-DNA (Wang et al. 1979). However, the crystals that nucleated were Z-DNA, and as these crystals grew, the equilibrium shifted, until finally all of the material had changed to Z-DNA in the crystal lattice.

Figure 11 shows an end view of Z-DNA, using space-filling atoms. This shows a cross-section of 3 bp as viewed down the axis. The helix twists in a clockwise direction moving toward the reader. Thus, on the left side the three phosphate groups are visible, whereas only one is seen on the right. It can be seen that the groove in Z-DNA is fairly deep, and the axis is found very close to the O2 oxygen of the cytosine residue. In this diagram it can be seen that the imidazole ring of guanine projects on the outside of the molecule, where it is readily accessible to other substances.

Z-DNA can also form in oligonucleotides containing AT base pairs. This has been shown in the crystal structure of the octamer d(CGCATGCG), which adopts the same crystal lattice and structure as d(CGCGCGCG) (S. Fujii et al., in prep.).

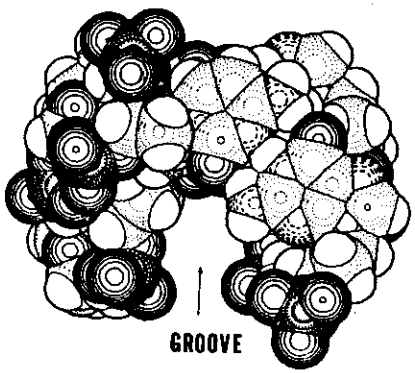


Figure 11. End view of Z-DNA in which 3 bp are drawn in a van der Waals diagram. The groove extends almost down to the helical axis, which passes through the cytosine O2 atom. As the bases come toward the reader, they rotate in a clockwise direction, thereby exposing three phosphate groups on the left. Oxygen atoms are shaded with circles, phosphorous atoms have spiked circles, and nitrogen atoms are indicated by dotted circles.

Z-DNA Stabilization by Cytosine Methylation

One of the principal modifications of the nucleic acids, especially in the higher eukaryotes, is methylation of cytosine on the 5 position where that cytosine precedes a guanine. Extensive studies have been carried out that have shown that methylation of DNA is generally associated with turning off transcription; conversely, demethylation is associated with the initiation of transcription or the expression of the gene (Razin and Riggs 1980; Ehrlich and Wang 1981). One of the first features that was apparent in the initial Z-DNA work was the fact that CG sequences played an important role in the Z conformation. Accordingly, it is reasonable to ask to what extent methylation might modify the distribution between right-handed B-DNA and left-handed Z-DNA. Behe and Felsenfeld (1981) have addressed this problem directly by synthesizing a polymer, poly(dG-m^{5}dC), that is fully methylated. They found that methylation of the cytosine residue on the 5 position had a profound effect in altering the equilibrium between B-DNA and Z-DNA in solution. If one has poly(dG-dC) and poly(dG-m^{5}dC) in 50 mM NaCl, both molecules are in the right-handed B-DNA conformation. To convert poly(dG-dC) to left-handed Z-DNA, the magnesium concentration must be raised to 760 mM. However, for the fully methylated polymer, 0.6 mM $MgCl_2$ is adequate to convert it to Z-DNA. There has been a decrease by three orders of magnitude in the amount of magnesium ions needed to stabilize Z-DNA if the polymer is methylated.

We have recently solved the structure of the hexamer (m^{5}dC-dG)$_3$ (Fujii et al. 1982), which is the same hexamer as that in which the original Z-DNA helix was discovered. The crystallographic data for this are presented in Table 1, and Figure 10 compares these B and Z forms. The overall form of the methylated Z-DNA molecule is similar to that seen in the unmethylated molecule. This result is reasonable in view of the fact that antibodies to nonmethylated Z-DNA can also recognize Z-DNA formed by the methylated polymer (Nordheim et al. 1981). However, there have been some subtle changes in the geometry of the molecule, principally associated with the fact that the methyl group is very close to the carbon atoms C1′ and C2′ of the adjacent guanosine residue. This has made a small change in the helix twist angle of Z-DNA, as shown in Figure 4. It can be seen that the methylated polymer has a twist angle for CpG that is $-13°$, whereas the average value for the unmethylated polymer is $-9°$. There is a similar change averaging 4° for the sequence GpC. The net effect of these changes in twist angle is that the methyl groups on the surface of the molecule are drawn closer together in the methylated polymer than they would be if they were attached to the unmethylated polymer without a change in twist angle.

The stabilization of Z-DNA by methylation is due to two factors. One of these is that the methyl group fills a vacancy or depression on the surface of Z-DNA. This is shown in Figure 10, in which the arrow near the unmethylated polymer (A) points to a depression at the side of the molecule, whereas in the methylation polymer (B) that depression is filled by the methyl group. The methyl group is in van der Waals contact with the imidazole ring of guanine immediately above it in Figure 10B and also with the C1′ and C2′ carbon atoms of the same guanosine. In effect, these form a small hydrophobic patch on the surface of the molecule. The methyl group fills a depression in the molecule that would otherwise be filled by water molecules. This is in contrast to the situation in B-DNA, where the methyl group projects out in the solvent from the major groove of the double helix. The effect of methylation is thus twofold: (1) to destabilize B-DNA by interposing a hydrophobic group into the water and (2) to stabilize Z-DNA by closing off the access of water molecules into a hydrophobic region and thereby allowing the molecule to be stabilized by hydrophobic bonding. The structural studies thus account for the considerable change in the equilibrium between B-DNA and Z-DNA. It also provides a structural rationale for why methylation of DNA might favor the formation of small segments of Z-DNA.

Z-DNA in Biological Systems

The structural studies on Z-DNA suggest that there exists a stable conformation of DNA in the left-handed form that may exist in vivo. Biological studies have provided evidence for the existence of these structures (Nordheim et al. 1981; H.J. Lipps et al., in prep.). What is not clear at the present time is the size of the smallest Z-DNA that might exist in the middle of B-DNA. Frequently, methylation occurs in CpG sequences in DNA in which the surrounding sequences do not have large stretches of alternating purine-pyrimidines. One wonders whether these small dinucleotides could actually form Z-DNA in the middle of a B-DNA strand. The answer to this might come through further structural studies in which the nature of the B-Z interface is trapped in a crystal lattice. If we knew the detailed folding of B-DNA when it joins to Z-DNA, we might then have a better appreciation of the processes that stabilize this conformational change. Further structural studies might provide a clue for interpreting the profound effect of methylation of CG sequences on gene expression.

CONCLUSION

Recent structural studies on right- and left-handed fragments of DNA have considerably broadened our perspective about the conformational vitality of DNA. The high prevalence of A-DNA structures existing in crystalline lattices suggests that these structures may also exist in biological systems and may be stabilized by particular sequences or by interactions with proteins. In the same way, the discovery of Z-DNA in a crystal structure has led to its discovery in biological systems. We now have an entirely different image of DNA than was available a few years ago. The molecule may now be appropriately regarded as one that is in dynamic equilibrium among a number of different conformations. The detailed conformation assumed by the molecule is one that will be influenced by nucleotide sequence, ions in the surrounding medium, the particular proteins that bind to the molecule, and the degree of negative supercoiling stress that the molecule feels (Nordheim et al., this volume). All of these factors will be important in defining a highly dynamic system in which a number of different conformations are found.

Are there other conformations of DNA yet to be discovered? The answer to this is unclear at present. Although DNA has been studied extensively for more than three decades, only recently have we become aware of the fact that the molecule is dynamically active, and this awareness may lead to even further discoveries in the future.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institutes of Health, the American Cancer Society, and the National Aeronautics and Space Administration. We thank Dr. Gary Quigley for assistance in providing a variety of computer programs that were used both in solving these structures and in drawing various figures describing them.

REFERENCES

BEHE, M. and G. FELSENFELD. 1981. Effects of methylation on a synthetic polynucleotide: The B-Z transition in poly(dG-m^{5}dC)·poly(dG-m^{5}dC). *Proc. Natl. Acad. Sci.* **78:** 1619.

CONNER, B.N., T. TAKANO, S. TANAKA, S. ITAKURA, and R.E. DICKERSON. 1982. The molecular structure of d(CpCpGpG). a right-handed A-DNA double helix. *Nature* **295:** 294.

CRAWFORD, J.L., F.J. KOLPAK, A.H.-J. WANG, G.J. QUIGLEY, J.H. VAN BOOM, G. VAN DER MAREL, and A. RICH. 1980. The tetramer of d(CpGpCpG) crystallizes as a left-handed double helix. *Proc. Natl. Acad. Sci.* **77:** 4106.

DAY, R.O., N.C. SEEMAN, J.M. ROSENBERG, and A. RICH. 1973. A crystalline fragment of double helix: The structure of the dinucleoside phosphate guanylyl-3′, 5′-cytidine. *Proc. Natl. Acad. Sci.* **70:** 849.

DREW, H., T. TAKANO, S. TANAKA, K. ITAKURA, and R.E. DICKERSON. 1980. High-salt d(CpGpCpG), a left-handed Z-DNA double helix. *Nature* **286:** 567.

EHRLICH, M. and R.Y.-H. WANG. 1981. 5-Methylcytosine in eukaryotic DNA. *Science* **212:** 1350.

FRANKLIN, R.E. and R. GOSLING. 1953. Molecular configuration in sodium thymonucleate. *Nature* **171:** 740.

FUJII, S., A.H.-J. WANG, G.A. VAN DER MAREL, J.H. VAN BOOM, and A. RICH. 1982. Molecular structure of (m^{5}dC-dG)$_3$: The role of the methyl group on 5-methyl cytosine in stabilizing Z-DNA. *Nucleic Acids Res.* (in press).

HASCHEMEYER, A.E.V. and A. RICH. 1967. Nucleoside conformations: An analysis of steric barriers to rotation about the glycosidic bond. *J. Mol. Biol.* **27:** 369.

JAIN, S.C., C.-C. TSAI, and H.M. SOBELL. 1977. Visualization of drug-nucleic acid interactions at atomic resolution. II. Structure of an ethidium/dinucleoside monophosphate crystalline complex, ethidium: 5-iodocytidylyl (3′-5′) guanosine. *J. Mol. Biol.* **114:** 317.

JEFFREY, A.M., K.W. JENNETTE, S.H. BLOBSTEIN, I.B. WEINSTEIN, F.A. BELAND, R.G. HARVEY, H. KASAI, I.

Miura, and K. Nakanishi. 1976. Benzo[a]pyrene-nucleic acid derivative found *in vivo*: Structure of a benzo[a]pyrenetetrahydrodiol epoxide-guanosine adduct. *J. Am. Chem. Soc.* **98:** 5714.

Leslie, A.G.W., S. Arnott, R. Chandrasekaran, and R.L. Ratliff. 1980. Polymorphism of DNA double helices. *J. Mol. Biol.* **143:** 49.

Nordheim, A., M.L. Pardue, E.M. Lafer, A. Muller, B.D. Stollar, and A. Rich. 1981. Antibodies to left-handed Z-DNA bind to interband regions of *Drosophila* polytene chromosomes. *Nature* **294:** 417.

North, A.C.T. and A. Rich. 1961. X-ray diffraction studies on bacterial viruses. *Nature* **191:** 1242.

Ogawa, T. and T. Okazaki. 1980. Discontinuous DNA replication. *Annu. Rev. Biochem.* **49:** 421.

Pohl, F.M. and T.M. Jovin. 1972. Salt-induced cooperative conformational change of a synthetic DNA: Equilibrium and kinetic studies with poly(dG-dC). *J. Mol. Biol.* **67:** 375.

Quigley, G.J. and A. Rich. 1976. Structural domains of transfer RNA molecules. *Science* **194:** 796.

Quigley, G.J., A.H.-J. Wang, G. Ughetto, G. van der Marel, J.H. van Boom, and A. Rich. 1980. Molecular structure of an anticancer drug-DNA complex: Daunomycin plus d(CpGpTpApCpG). *Proc. Natl. Acad. Sci.* **77:** 7204.

Razin, A. and A.D. Riggs. 1980. DNA methylation and gene function. *Science* **210:** 604.

Rosenberg, J.M., N.C. Seeman, R.O. Day, and A. Rich. 1976. RNA double-helical fragments at atomic resolution: II. The crystal structure of sodium guanylyl-3′, 5′-cytidine nonahydrate. *J. Mol. Biol.* **104:** 145.

Rosenberg, J.M., N.C. Seeman, J.J.P. Kim, F.L. Suddath, H.B. Nicholas, and A. Rich. 1973. Double helix at atomic resolution. *Nature* **243:** 150.

Seeman, N.C., J.M. Rosenberg, F.L. Suddath, J.J.P. Kim, and A. Rich. 1976. RNA double-helical fragments at atomic resolution. I. The crystal and molecular structure of sodium adenylyl-3′, 5′-uridine hexahydrate. *J. Mol. Biol.* **104:** 109.

Shakked, Z., D. Rabinovich, W.B.T. Cruse, E. Egert, O. Kennard, G. Sala, S.A. Salisbury, and M.A. Viswamitra. 1981. Crystalline A-DNA—The X-ray analysis of the fragment d(GGTATACC). *Proc. R. Soc. Lond. B* **213:** 479.

Tsai, C.-C., S.C. Jain, and H.M. Sobell. 1977. Visualization of drug-nucleic acid interactions at atomic resolution. I. Structure of an ethidium/dinucleoside monophosphate crystalline complex, ethidium: 5-iodouridylyl (3′-5′) adenosine. *J. Mol. Biol.* **114:** 301.

Voet, D. and A. Rich. 1970. The crystal structure of purines, pyrimidines and their intermolecular complexes. *Prog. Nucleic Acid Res. Mol. Biol.* **10:** 183.

Wang, A.H.-J., S. Fujii, J.H. van Boom, and A. Rich. 1982a. Molecular structure of the octamer d(G-G-C-C-G-G-C-C): Modified A-DNA. *Proc. Natl. Acad. Sci.* **79:** 3968.

Wang, A.H.-J., S. Fujii, J.H. van Boom, G.A. van der Marel, S.A.A. van Boeckel, and A. Rich. 1982b. Molecular structure of r(GCG)d(TATACGC): A DNA-RNA hybrid helix joined to double helical DNA. *Nature* **299:** 601.

Wang, A.H.-J., G.J. Quigley, F.J. Kolpak, G. van der Marel, J.H. van Boom, and A. Rich. 1981. Left-handed double helical DNA: Variations in the backbone conformation. *Science* **211:** 171.

Wang, A.H.-J., G.J. Quigley, F.J. Kolpak, J.L. Crawford, J.H. van Boom, G. van der Marel, and A. Rich. 1979. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282:** 680.

Wells, R.D., T.C. Goodman, W. Hillen, G.T. Horn, R.D. Klein, J.E. Larson, U.R. Müller, S.K. Neuendorf, N. Panayotatos, and S.M. Stirdivant. 1980. DNA structure and gene regulation. *Prog. Nucleic Acid Res. Mol. Biol.* **24:** 167.

Wing, R.M., H.R. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R.E. Dickerson. 1980. Crystal structure analysis of a complete turn of B-DNA. *Nature* **287:** 755.

Conformational Flexibility of the Polynucleotide Chain

Alexander Rich, Gary J. Quigley, and Andrew H.-J. Wang
Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

One of the widely appreciated features of protein structures is the fact that the polypeptide chain can fold into different conformations. This includes tightly folded conformations such as found in the α helix or more extended conformations as is found in the β sheet. It is less widely appreciated that the polynucleotide chain also has conformational flexibility which results in significant modifications of the chain extension. Many of these differences are associated with different puckers of the five-membered furanose ring which is a component of both the DNA and RNA chains. Here we illustrate changes in the sugar ring conformation in a variety of structures, taking examples in oligonucleotides as well as in polynucleotides. The property of polynucleotide sugar rings to alter their pucker should be looked upon as a degree of conformational freedom. It is important to have this kind of flexibility in order to form complex structures beyond the simple double helix.

The Double Helix and Ring Pucker

Double helices can be formed with both DNA and RNA molecules. These double helices have similar features in that the bases are in the center of the molecule with the familiar Watson-Crick pairing, either adenine-thymine pairs in DNA or adenine-uracil in RNA. The sugar phosphate chains are on the outside of the double helix. Although the two types of double helices are similar, they differ in one fundamental feature; the pucker of the five-membered furanose ring is different in ribose and deoxyribose.[1] The differences in ring pucker are illustrated in Figure 1. The furanose ring is viewed edge on in a plane which is defined by the ring oxygen O1′ and carbons C1′ and C4′; C1′ has attached to it the ring nitrogen (N) of the purine or pyrimidine; C4′ has attached to it the additional carbon atom C5′ as indicated in the diagram. Both of these constituents, the base (N) and C5′, are located on the upper side of the ring in Figure 1. It is clear that the principal difference in the conformation in the sugar ring is whether the carbon atom C2′ or C3′ is above the plane of the ring. Atoms which are above the plane of the ring in Figure 1 are considered to be in the *endo* conformation.[2] Thus the conformation C2′ *endo* has C2′ on the upper side of the ring and this correspondingly forces C3′ to be on the lower side of the ring. In the other conformation C3′ is *endo* and C2′ is on the lower side of the ring. The major conformation of a deoxyribose chain is one in which the furanose ring has a C2′

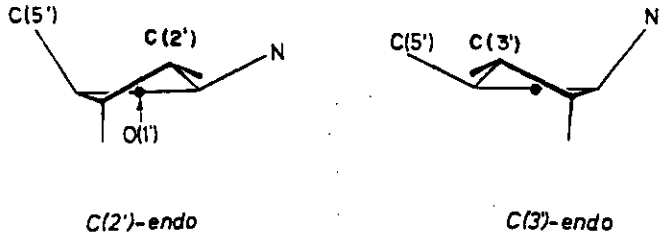


Figure 1. Diagram illustrating two different conformations of the ribofuranose ring. The plane of the ring is defined by three atoms: C1′ to which is attached the glycosidic bond indicated by N, O1′ and C4′ which is attached to the atom C5′. In this diagram, we are looking edge on at the plane defind by these three atoms. The remaining two atoms in the ring, C2′ and C3′ are located either above or below the plane of the ring. Atoms located above the plane of the ring are in the *endo* position. On the left C2′ is in the *endo* position and on the right C3′ is in the *endo* position. Although two examples of ring pucker are shown in this diagram, there are actually a number of intermediate states in which the displacement of C2′ or C3′ is not as great as that illustrated here. The detailed nomenclature for furanose ring pucker is complex[2]; here we elect to use only the simplified *endo* conformations.

endo conformation. In contrast, the normal polyribose chain conformation is C3′ *endo*. The principal reason for the difference is associated with the additional space occupied by the oxygen atom attached to C2′in ribose which is absent in the deoxy series where only a hydrogen atom is attached to C2′. It should be emphasized that even though these are the normal conformations of these sugars in their respective double helices, the energy barriers involved in changing sugar conformation are not very great.[3]

Because of the C2′ *endo* conformation of deoxyribose, the form of the DNA double helix is such that the base pairs are found on the helix axis in the familiar B form of the DNA. The bases occupy the axis of the molecule so that the familiar double helical DNA has a solid, rod-like appearance with bases in the center and two helical grooves running down the molecule. These are the major and minor grooves which occupy the space between the deoxyribose-phosphate chains (see Color Plates 2 and 6).

In contrast, the C3′ *endo* conformation of ribose makes significant modifications in the RNA double helix. The base pairs are no longer perpendicular to the helix axis, but are tilted at about 15-18°. Furthermore, they are set back away from the helix axis. In fact, there is a clear space in the center of the molecule of approximately 3 Angstroms in diameter. The RNA double helix thus looks more like a molecular ribbon wrapped around an imaginary cylinder 3 Angstroms in diameter in which there is now a very deep major groove and a comparatively shallow minor groove on the outside of the helix.

These forms of the familiar double helices are not invariant. Reducing the water content of the medium by adding alcohol readily converts the familiar B form of DNA into the A form in which deoxyribose adopts the C3′ *endo* conformation

which is normally found in RNA double helices. The double helix then changes its conformation and looks more like double helical RNA (see Color Plates 3 and 7). It is noteworthy that in the conversion from the B to the A form of DNA, there is over an Angstrom difference in the phosphate-phosphate distance along each polynucleotide chain.

The reason for the change in phosphorus-phosphorus distance can be seen schematically in Figure 2 which shows the conformation of two forms of an adenosine nucleotide in a ribose polynucleotide chain. In the upper figure, the ribose is in the normal C3′ *endo* conformation and the phosphorus-phosphorus distance is near 5.9 Angstroms. In the lower part of Figure 2, the ribose is in a C2′ *endo* con-

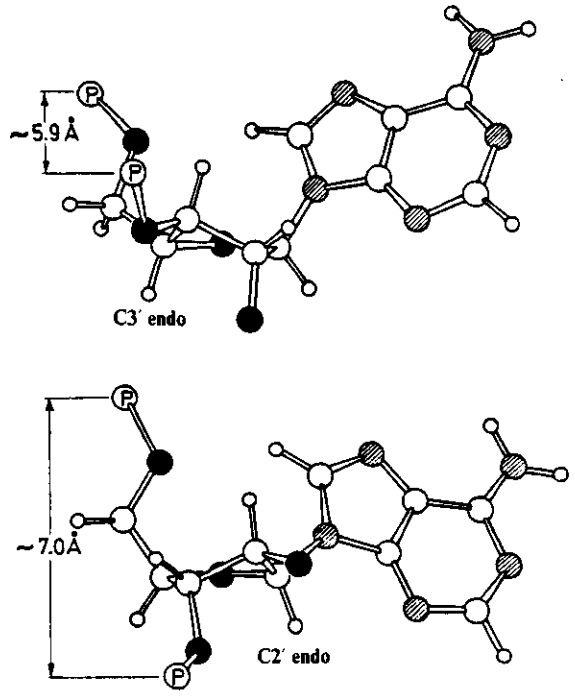


Figure 2. Illustration of two different conformations of adenylic acid. In the upper diagram with the C3′*endo* conformation, the two phosphate groups are both above the plane of the ribose ring and are approximately 5.9 Angstroms apart. In the lower diagram, the C2′ *endo* conformation, the phosphate attached to O3′ is located below the plane of the ribose ring and the phosphates are now approximately 7 Angstroms apart. These two conformations are associated with considerable differences in the extension of the sugar phosphate chain.

formation (normally found in DNA) and the phosphorus-phosphorus distance is near 7 Angstroms. The change in distance is largely associated with a change in the pucker of the ring, although there are also small changes in the other dihedral angles in the backbone. The polynucleotide chain thus has a degree of conformational freedom which allows it to change the degree of extension in a significant way. This flexibility is used in a variety of ways in polynucleotide structures.

Polynucleotide Chain May Turn Corners With Changes in Ring Pucker

There are many examples in which the conformation of the sugar ring is changed. In some cases, the changes in pucker are associated with a change in the direction of a polynucleotide chain. An example of this is seen in Figure 3 which shows the crystal structure of adenyly-(3′,5′)-uridine (ApU) which formed a crystalline complex with 9-amino acridine.[4] In this structure, there are stacked columns of planar molecules in which the 9-amino acridine molecule alternates with an adenine-uracil base pair hydrogen bonding through the ring nitrogen of N7 of adenine. Uracil is at one end of the chain and the backbone has an extended conformation in which the molecule actually turns a corner so that its adenine residue is then hydrogen bonded to still another uracil residue (Figure 3). In the course of making this turn, one of the sugar residues adopts the unusual C2′ *endo* conformation as shown in the diagram. As seen at the right of the diagram, the uridine ribose O1′ is found at the bottom while it is found at the top of the adenosine ribose. This illustrates the extent to which the chain has radically changed direction in this structure. A somewhat similar conformation is seen in the structure of adenylyl-(3′,5′)-adenosine (ApA) which

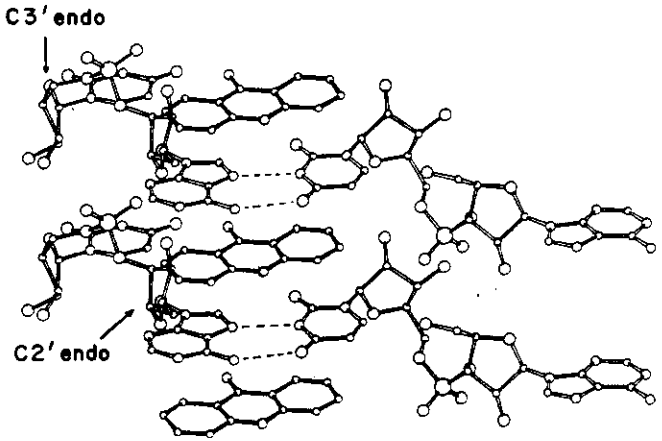


Figure 3. The crystal structure of the complex of ApU and 9 amino acridine. The polynucleotide chain of ApU is in an extended conformation and the chain turns a sharp corner. This can be seen by the different orientation of the two ribose residues in the oligonucleotide. The change in direction is associated with a change in ring pucker as shown.

44

crystallizes with proflavin in a similar manner.[5] However, it is not obligatory for the ribose in RNA to change pucker in coiled regions of the molecule. The examples cited above illustrate the fact that changes in pucker may be seen in some nucleotide structure determinations.

Tables I and II cite the structure of oligonucleotides which have crystallized by themselves (Table I) or together with intercalators (Table II). They are listed with description of the sugar conformation in various parts of the backbone. Although we have cited two examples in which oligonucleotides change sugar pucker where there is a change in the direction of the polynucleotide chain, there are several

Table I
Oligonucleotides

		5′ end	3′ end	Reference
DNA fragments	pdTpdT	C2′ endo	C2′ endo	15
	d-(pApTpApT)	C3′ endo (A)	C2′ endo (T)	16
RNA fragments	ApU	C3′ endo	C3′ endo	17
	GpC	C3′ endo	C3′ endo	18
	GpC	C3′ endo	C3′ endo	19
	ApA*	C3′ endo	C3′ endo	6
	A*pA*	C3′ endo	C3′ endo	6
	UpA	C3′ endo	C3′ endo	20, 21

Table II
Intercalator Complexes

	Nucleotide	Intercalator	Sugar Pucker 5′ end	Sugar Pucker 3′ end	Reference
DNA fragments	dCpG	TPH*	C3′ endo	C2′ endo	7
RNA fragments	CpG	Acridine orange*	C3′ endo	C2′ endo	8
	rIodo UpA	Ethidium*	C3′ endo	C2′ endo	9
	rIodo CpG	Ethidium*	C3′ endo	C2′ endo	10
	rIodo CpG	9-Amino acridine*	C3′ endo	C2′ endo	11
	rIodo CpG	Acridine orange*	C3′ endo	C2′ endo	12
	rIodo CpG	Ellipticine*	C3′ endo	C2′ endo	12
	CpG	Proflavine*	C3′ endo	C3′ endo	13
	rIodo CpG	Proflavine*	C3′endo	C3′ endo	12
Non-helical	ApU	9-Amino acridine*	C2′ endo	C3′ endo	4
	ApA	Proflavine*	C2′ endo	C3′ endo	5

oligonucleotides in which there has been no change in the pucker even though the chain is in a fairly extended conformation. An example is seen in the structure of the trinucleotide ApApA.[6]

The Double Helix Changes Conformation Upon Intercalation

One of the conformational changes frequently encountered in double helical nucleic acids is that associated with the binding of a planar intercalator molecule which lodges between the base pairs. It does this without substantial disruption of the helix. Intercalation has two important effects: it introduces a gap between adjacent base pairs which are then separated by 6.8 Angstroms instead of the normal 3.4 Angstroms. This is due to the planarity of the intercalator which usually has unsaturated rings with π electrons which are 3.4Å thick. In addition, there is an unwinding of the double helix.

A series of structures have been solved involving intercalators, mostly in the ribonucleotide series but with one in the deoxynucleotide series (Table II). The structure of a double helical fragment of DNA together with an intercalator is shown in Figure 4. Here we see the structure of deoxy CpG which accommodates a terpyridine-platinum intercalator in the midst of a double helical fragment.[7] As seen in the diagram, the 5′ end of the double helical segment adopts the unusual C3′ *endo* conformation while the 3′ end maintains the C2′ *endo* conformation which is normal for double helical DNA.

Figure 5 shows the structure of the ribose dinucleoside phosphate CpG together with the intercalator acridine orange.[8] It can be seen that the 3′ end of the double helical RNA fragment has adopted the unusual C2′ *endo* conformation while the 5′ end

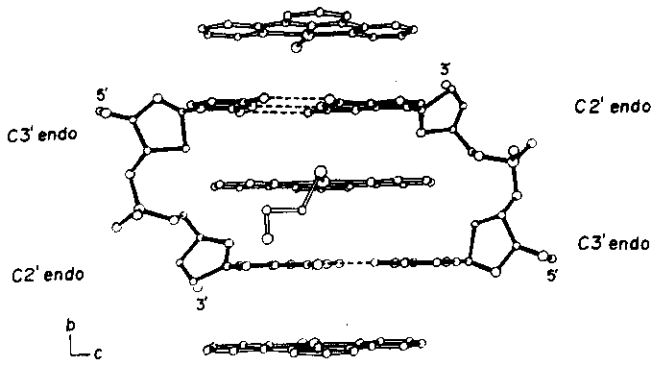


Figure 4. The structure of the deoxy CpG-platinum terpyridine intercalator complex. It can be seen that the DNA double helix has a planar intercalator lodged between the base pairs. This is associated with an unusual ring pucker on the 5′ side of the oligonucleotide segment. This conformation allows an extension of the polynucleotide chain.

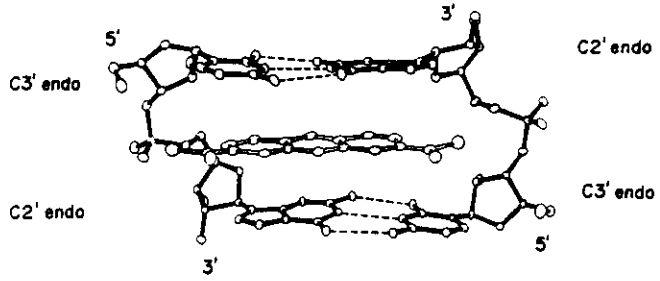


Figure 5. The structure of the CpG-acridine orange intercalator complex. It can be seen that there is an unusual conformation at the 3′ end of the oligonucleotide chain surrounding the intercalator. This mixed pucker conformation is associated with an extension of the polynucleotide chain.

maintains the normal C3′ *endo* conformation. A large number of intercalator structures have been solved in the ribonucleotide series as listed in Table II. Sobell and his colleagues were the first to point out that intercalation is associated with a modification of the pucker of the ribonucleotide chain on the 3′ side of the intercalator.[9-12] Although intercalation is generally associated with conformations similar to those seen in Figure 5, a number of alternative conformations are listed at the bottom of Table II. These are usually associated with the intercalators proflavin[9-12-13] and 9-amino acridine[4] both of which have the property of forming hydrogen bonds between the intercalator and the phosphate of the dinucleoside phosphate. In both cases, other modes of pucker are found. For example, in the complex of proflavin with the ribose CpG fragments, both residues have the normal C3′ *endo* conformation.[13]

For "simple" intercalators in which there is no hydrogen bonding to the phosphate residue, it is possible to make an interesting generalization about the way double helical DNA and RNA accommodate intercalator addition. This is illustrated schematically in Figure 6. At the left the DNA double helix is shown diagrammatically with the normal C2′ *endo* conformation in all residues, except for those on the 5′ side of the intercalator where the unusual C3′ *endo* conformation is adopted in a manner analogous to that which is illustrated in Figure 4. On the right the diagram shows the way in which double helical RNA accommodates an intercalator. All of the residues are in the normal C3′ *endo* conformation except for those residues on the 3′ side of the intercalator which adopt the unusual C2′ *endo* conformation. However, it can be seen that in the region immediately surrounding the intercalator (enclosed in the dashed line) the conformation of both the RNA and DNA chain are similar. Thus both molecules elongate to accommodate an intercalator by adopting a similar conformation. These conformational changes, as described by Sobell[9-12] explain the nearest neighbor exclusion. If a DNA or RNA double helix is saturated with an intercalator, the most that can be accommodated is one intercalator for every two base pairs. The reason for this is probably associated with the necessity for mixed pucker on either side of the intercalator.

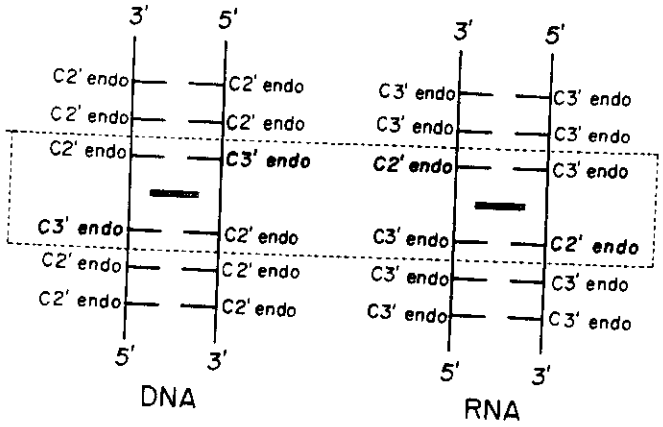


Figure 6. A schematic diagram illustrating the manner in which either DNA or RNA double helices change the pucker of the sugar residues in the region immediately surrounding a simple intercalator. Although the changes are different in the two types of double helices, the conformation in the region surrounding the intercalator enclosed by the dashed rectangle is the same in both cases.

Conformational Flexibility in Nucleic Acid Macromolecules: Yeast tRNA^{Phe}

Examples of changes in the type of sugar pucker can be seen in the three-dimensional structure of yeast phenylalanine transfer RNA (tRNA^{Phe}). Although most of the seventy-six nucleotides in this transfer RNA molecule adopt the C3′ *endo* conformation which is normal for a ribonucleotide chain, several residues are found to adopt the less common C2′ *endo* conformation. This is frequently associated with an interesting type of structural accommodation which is similar to those interactions described above in the oligonucleotide structures. Figure 7 is a diagram (see also Color Plate 1) which shows secondary and tertiary hydrogen bonding found in the yeast tRNA^{Phe}. This schematic diagram is useful in interpreting Figures 8 through 11 which illustrate conformational changes in various parts of the molecule.[14]

Changes in Polynucleotide Direction

The principal sites using C2′ *endo* conformation in yeast tRNA^{Phe} are listed in Table III. Nine examples are cited. These occur in two principal situations. In five cases (residues 7, 17, 18, 48 and 60) the C2′ *endo* conformation is adopted when the polynucleotide chain undergoes a distinct change in direction. For example, nucleotides 1 through 7 are involved in the double helical acceptor stem but residues 8 and 9 are extended and provide a linker region which attaches one end of the acceptor stem with the beginning of the D stem. In this region, there are several conformational changes, one of which is associated with the C2′ *endo* conformation of ribose 7 where the polynucleotide chain changes direction. Other examples are

48

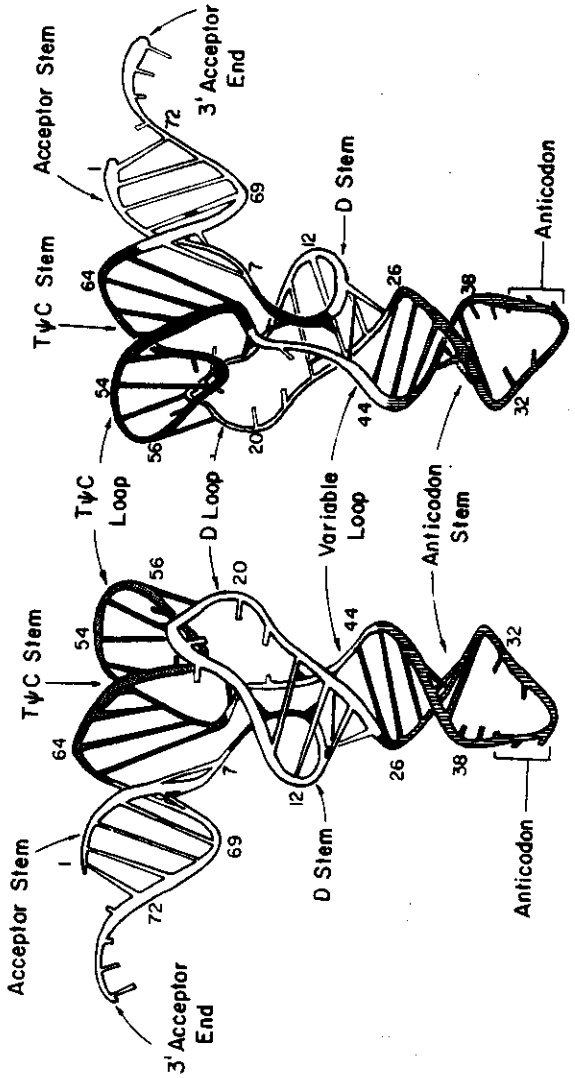


Figure 7. A schematic diagram showing two side views of yeast tRNA^{Phe}. The ribose-phosphate backbone is depicted as a coiled tube, and the numbers refer to nucleoside residues in sequence. Shading is different in different parts of the molecules, with residues 8 and 9 in black. Tertiary hydrogen-bonding interactions between bases are shown as solid black rungs, which indicate either one, two or three hydrogen bonds. Those bases that are not involved in hydrogen bonding to other bases are shown as shortened rods attached to the coiled backbone.

found in regions where the backbone loops out away from the center where the bases are stacked. The dihydrouracil residues 16 and 17 are extended out away from the remainder of the molecule and their bases are not stacked with the other bases in tRNA. The backbone has a "bulge" or looping-out at that point (Figure 7). In order to accommodate this extended conformation, residues 17 and 18 adopt the C2′ *endo* conformation. In addition, ribose 47 in the varible loop is extended and its base thrusts away from the center of the molecule. This is associated with a C2′ *endo* conformation in ribose 48. Still another example is found in the T loop where bases 59 and 60 are excluded from the stacking of the other bases in the T stem and loop (Figure 7). This exclusion involves a C2′ *endo* conformation in ribose 60. All of these conformational changes are thus associated with an abrupt change in the direction of the polynucleotide chain. This directional change is accommodated by adoption of a C2′ *endo* conformation and there is a substantial change in the phosphorus-phosphorus distance along the polynucleotide chain.

It should be pointed out, however, that there are many other examples in the loop regions of $tRNA^{Phe}$ in which the polynucleotide chain changes direction but the C2′

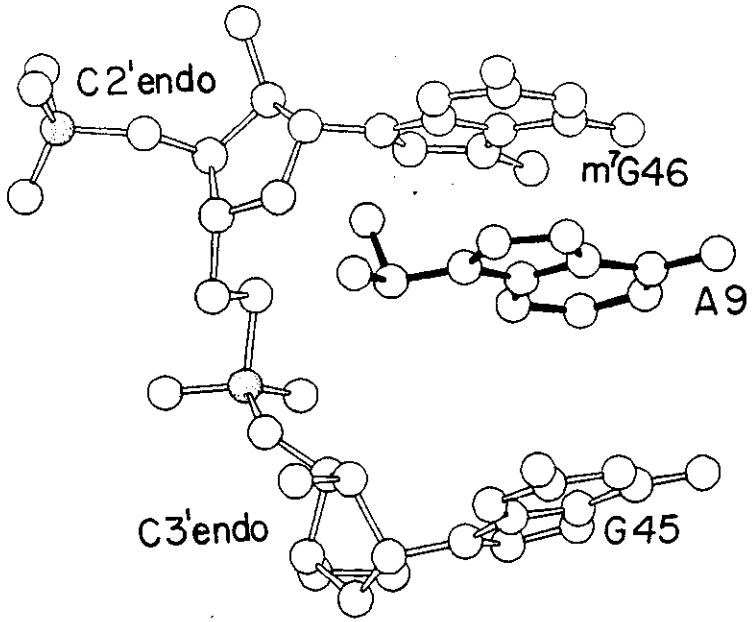


Figure 8. Intercalation in yeast $tRNA^{Phe}$. Adenine of residue 9 intercalates between the bases of 45 and 46. This intercalation is accommodated by a change in the ribose pucker at the 3′ end of the segment. The ribose of m⁷G46 is in the unusual C2′ *endo* conformation.

50

endo conformation is not used. Thus, the change in pucker is a structural accommodation which may be adopted in the molecule, especially where the chain is extended, but it is not generally used when the polynucleotide chain undergoes a change in direction.

Intercalation in Yeast $tRNA^{Phe}$

There are two parts of the molecule in which extensive intercalation is found involving pairs of nucleotides. These are in the central region with intercalation involving nucleotides 8 and 9 as well as in the corner of the molecule where the T and D loops interact. Figure 8 shows the conformation adopted by the sugars of residues G45 and m⁷G46 which has the adenine ring of A9 intercalated between them. Residues 45 and 46 are not involved in a double helix; nonetheless, the conformation of m⁷G46 is C2′ *endo* in a manner analogous to the conformational changes which are seen for double helical ribonucleotide fragments surrounding intercalators. The C2′ *endo* conformation is adopted at the 3′ end of the intercalator region in Figure 8. Another example is shown in Figure 9 where residues U8 and A9 are found in the extended

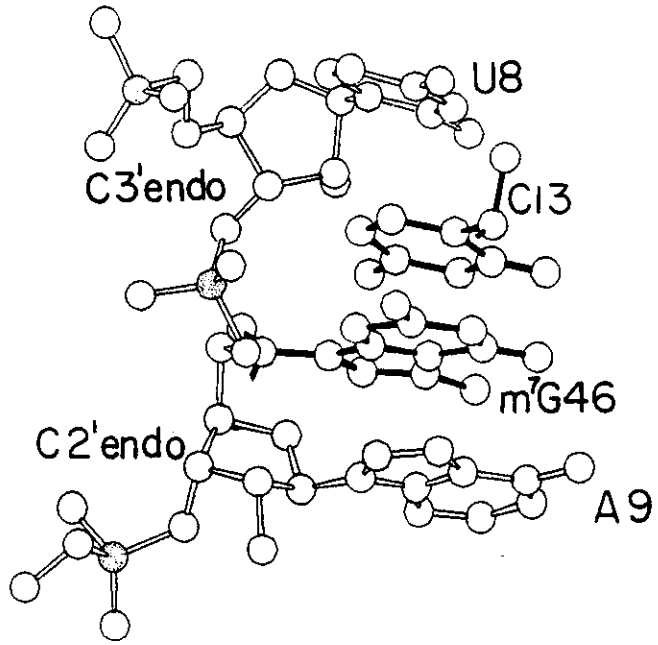


Figure 9. The nucleotides U8 and A9 have intercalated between them the bases C13 and m⁷G46. These bases are both hydrogen bonded to G23 in the tertiary structure of yeast $tRNA^{Phe}$. This intercalation is accommodated by a change in the conformation of ribose of A9.

segment which connects the acceptor stem with the D stem. It can be seen that A9 at the 3′ end of the segment has adopted the C2′ *endo* conformation even though it is not in a complementary double helical structure.

Figure 9 shows that intercalated between U8 and A9 are the bases m'G46 and C13 both of which are hydrogen bonded to G22. The base pair C13•G22 is part of the D stem. In Figure 9, it can be seen that the nucleotides U8 and A9 accommodate the additional distance associated with intercalation by adopting the C2′ *endo* conformation in residue A9.

By comparing Figures 8 and 9, it can be seen that they are both portions of the same structure in which there are two interacting segments of the polynucleotide chain, both of which intercalate around each other. The polynucleotide chains are interleaved between each other so that each intercalates into the opposite member of the pair.

A similar pair of interleaved structures are found near the corner of the molecule where the bases of nucleotides 18 and 19 interact with the bases of 57 and 58. In Figure 10, it can be seen that m¹A58 and G57 are spread apart with residue G18 in-

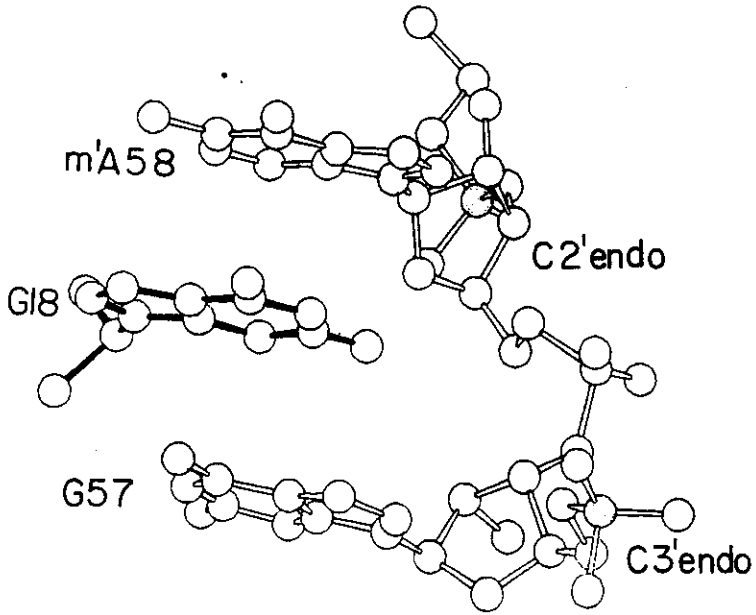


Figure 10. Diagram showing the intercalation of guanine 18 between G57 and m¹A57. The ribose of m¹A58 is in the C2′ *endo* conformation to accommodate the intercalation.

52

tercalated between them. This intercalation is associated with C2′ *endo* conformation found in m¹A58. In this case, it also adopts the C2′ *endo* conformation at the 3′ end of the oligonucleotide segment which surrounds the intercalating guanosine of G18. This is quite similar to the double helical RNA intercalator models which are described in Table II, as the bases G57 and m¹A58 are hydrogen bonded to other residues, although not in Watson-Crick pairs.

An interesting variant is seen in Figure 11 which shows the intercalation of G57 between the bases of G18 and G19. This intercalating interaction is one of the important stacking interactions which stabilize the corner of the tRNA molecule and helps maintain the interaction of the T loop and D loop. However, in this example, the less common C2′ *endo* conformation is found in both riboses of G18 and G19. By analogy with the simple intercalator structures, one would expect a C2′ *endo* conformation to be found in G19 but not in G18. Here it is likely that the unusual C2′*endo* conformation of G18 is not associated so much with the intercalation of G57 but is probably related to the fact that residues 16 and 17 are excluded from the molecule as described above so that their bases do not stack with the others. Another

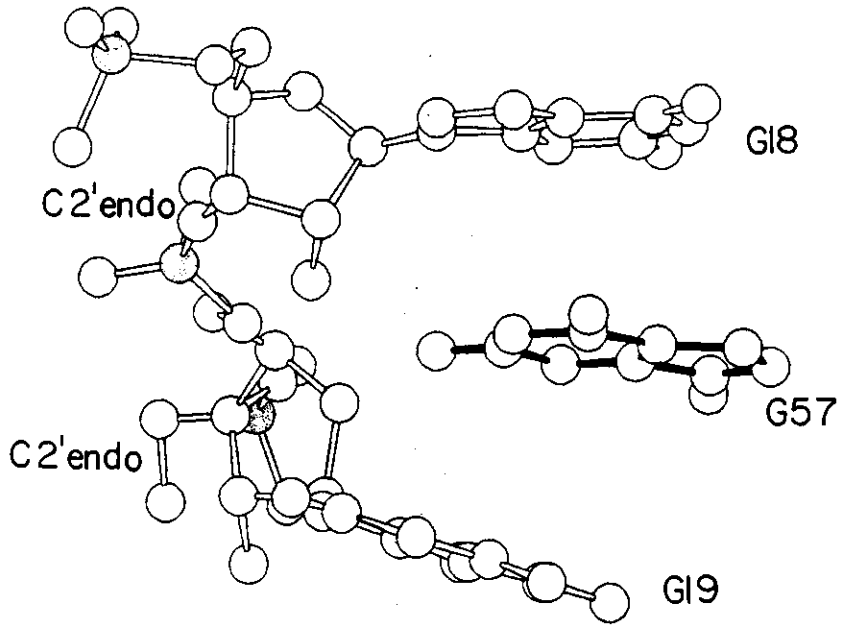


Figure 11. The intercalation of G57 between G18 and G19 is illustrated. Unlike the previous examples, the riboses of G18 and G19 are both in the unusual C2′ *endo* conformation. The C2′ conformation of G18 is probably associated with the unusual conformation of residues 16 and 17 immediately adjoining G18.

Table III
C2′ *Endo* Conformations in Yeast tRNA^{Phe}

Residue number	Description
Intercalation:	
9	8-9 intercalation by C13
19	18-19 intercalation by G57
46	45-46 intercalation by A9
58	57-58 intercalation by G18
Change of Direction:	
7	Extended segment at bend in chain, juncture between acceptor stem and 8-9 connection
17	Looping out of backbone at residues 16 and 17
18	Looping out of backbone
48	Looping out of variable loop backbone at U47; juncture between variable loop and T stem
60	Extended segment at bend in chain, bases 59 and 60 excluded from T stem and loop stacking

interesting feature associated with this is the fact that the bases of G18 and G19 are not only separated from each other by a distance of 6.8 Angstroms, which would be necessary for the intercalation of G57, but they are also translated relative to each other so the bases are no longer strictly on top of each other. This translation of G18 relative to G19 in the plane of the bases, can only be accommodated if both residues are in the C2′ *endo* conformation. It is likely that this feature is also related to the fact that the bases of residues 16 and 17 are excluded from stacking with the remainder of the bases in the tRNA structure. Some tRNA molecules contain only one nucleotide in this region which is excluded from the stacking, instead of the two seen in yeast tRNA^{Phe}. It is possible that in these cases, with only one residue, that G18 may have the normal C3′ *endo* conformation since the translation of G18 relative to G19 in the plane of bases may not be required.

Discussion

In the above, we have discussed the conformational flexibility found in the polynucleotide chains. This flexibility is inherent in the fact that there is a furanose ring in the chain which can adopt two different conformations which are associated with a varied extensibility of the chain. Structural studies on simple oligonucleotides reveal that conformational changes occur in regular double helical structures associated with intercalation as well as in other structures where polynucleotide chain undergo an abrupt change in direction. In the three-dimensional structure of yeast tRNA^{Phe}, examples are shown of both types of conformational changes.

Changes in pucker due to intercalation in general follow trends which are seen in intercalation in simple double helical RNA structures even though the tRNA examples do not involve simple RNA double helices. This implies that the adoption of a C2′*endo* conformation together with an associated extension of the polynucleotide chain is not solely limited to double helical intercalation, but may in fact be of a more general nature as shown by the examples cited above.

Nucleic acid molecules have considerable conformational flexibility. We have only a hint of this flexibility in studying the structure of the double helix itself. However, when one begins to study the interaction of double helical structures with other molecules, especially intercalators, we see that there is a method for accommodating them involving changes in extensibility. In complex globular polynucleotides such as the transfer RNA molecule, a variety of changes in pucker are observed. The molecule adopts unusual conformations associated either with chain extension, changes in direction of the polynucleotide chains or with the accommodation of other bases which intercalate into the chain even though the chains are not involved in a double helical array. As the structure of more complex polynucleotide structures are solved, it is our expectation that this will be found to be a completely general feature. This element of conformational flexibility associated with changes in ring pucker is likely to be a constant feature of polynucleotide chains when they interact with other molecules including proteins, as well as when they interact with other polynucleotide chains.

Acknowledgements

This research was supported by grants from The National Institutes of Health, The National Science Foundation, The National Aeronautics and Space Administration and the American Cancer Society. We thank J. Simpson for help in preparing the manuscript. A. H. J. W. is supported in part by the M. I. T. Center for Cancer Research (Grant No. CA-14051).

References and Footnotes

1. Arnott, S. and Hukins, D. W. L., *Biochem. Biophys. Res. Comm. 47*, 1504-1509 (1972).
2. Sundaralingam, M. in *Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions*, Ed by M. Sundaralingam and S. T. Rao, University Park Press, Baltimore, Maryland, pp. 487-524 (1974).
3. Yathindra, N. and Sundaralingam, M. in *Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions*, Ed by M. Sundaralingam and S. T. Rao, University Park Place, Baltimore, Maryland, pp. 649-676 (1974).
4. Seeman, N. C., Day, R. O., and Rich, A., *Nature 253*, 324-326 (1975).
5. Neidle, S., Tayler, G., Sanderson, M., Shieh, H.-S. and Berman, H. M., *Nucleic Acids Res., 5*, 4417-4422 (1978).
6. Suck, D., Manor, P. C., and Saenger, W., *Acta. Crysta. B32*, 1727-1737 (1976).
7. Wang, A. H.-J., Nathans, J., van der Marel, G., van Boom, J. H. and Rich, A., *Nature 276*, 471-474 (1978).
8. Wang, A. H.-J., Quigley, G. J. and Rich, A. *Nucleic Acids Res.* (in press).
9. Tsai, C. C., Jain, S. C. and Sobell, H. M., *J. Mol. Biol. 114*, 301-315 (1977).
10. Jain, S. C., Tsai, C. C. and Sobell, H. M., *J. Mol. Biol. 114*, 317-331 (1977).
11. Sakore, T. D., Jain, S. C., Tsai, C. C., and Sobell, H. M., *Proc. Natl. Acad. Sci. USA, 74*, 188-192 (1977).

12. Sobell, H. M., in *The International Symposium on Biomoleculr Structure, Conformation, Function and Evolution*. (Ed. R. Srinivasan), Pergamon Press, Oxford, New York (in press).
13. Neidle, S., *et al.*, *Nature 269*, 304-307 (1977).
14. Quigley, G. J. and Rich, A., *Science 194*, 796-806 (1976).
15. Camerman, N., Fawcett, J. K. and Camerman, A., *Science 182*, 1142-1143 (1973).
16. Viswamitra, M. W., Kennard, O., Jones, P. G., Sheldrick, G. M., Salisbury, S., Falvello, L. and Shakked, Z., *Nature 273*, 687-688 (1978).
17. Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Kim, J. J. P., and Rich, A., *J. Mol. Biol. 104*, 109-144 (1976).
18. Rosenberg, J. M., Seeman, N. C., Day, R. O., and Rich, A., *J. Mol. Biol. 104*, 145-167 (1976).
19. Hingerty, B., Subramanian, A., Stellman, S. D., Sato, T., Broyde, S. B., and Langridge, R., *Acta Crysta. B32*, 2998-3013 (1976).
20. Sussman, J. L., Seeman, N. C., Kim, S. H. and Berman, H. M., *J. Mol. Biol. 66*, 403-421 (1972).
21. Rubin, J., Brennan, T. and Sundaralingam, M., *Biochem. 11*, 3112-3129 (1972).

3 The Structure of the Z Form of DNA

ANDREW H.-J. WANG
ALEXANDER RICH
Massachusetts Institute of Technology
Cambridge, Massachusetts

57

CONTENTS

1. INTRODUCTION

A helix has handedness; it can twist. The DNA double helix, a chiral assembly, can exist in both right-handed and left-handed conformations. This idea was considered shortly after the formulation of the double helix by Watson and Crick in 1953. In this early period attention was focused on the question whether the X-ray diffraction patterns of DNA fibers were consistent with the initial proposal. During the decade following, investigators produced ample experimental evidence that, with some modification of details, the Watson–Crick right-handed helical model was essentially consistent with the experimental X-ray data (Arnott, 1970). Although DNA fiber analysis provided physical data in agreement with the Watson–Crick structure, it was understood by workers in the field that it did not unambiguously prove the structure. The limited resolution of the fiber X-ray diffraction patterns simply required too many assumptions to interpret the data. One constraint in particular was that the structure was perfectly regular with an asymmetric unit consisting of one nucleotide. Agreement between model and fiber diffraction data was reasonable, however, and this eventually led to the general adoption of the right-handed B-DNA conformation as the predominant form in biological systems.

Substantial interest was inspired, however, by observations that under certain conditions, alterations in conformation appeared. The most important of these was the A-DNA pattern, originally derived from air-dried fibers (Franklin and Gosling, 1951). The greater wealth of diffraction data contained in the A pattern carried with it more extensive structural information. The relative ease with which the A form converted to the B pattern reinforced the idea that these were two conformational variants, and it was shown that both structures could still be accommodated by a right-handed double helix.

Although the possibility of left-handed helices was considered, particularly in the period immediately following the description of the double helix in 1953, few systematic studies of left-handed helices were conducted during the next two decades. It might have been anticipated that left-handed double helical DNA regions would facilitate fork movement and chain separation during DNA replication, but flexibility as a mechanism for generating structural recognition signals along the DNA molecule or for regulating superhelical topology was not discussed seriously. It is only rather recently that there has been increased emphasis on alternative conformational models for DNA (Hopkins, 1981; Sasiskeharan, 1982). Part of the stimulus arose from the postulate that DNA might not exist as a continuously coiled double helix but that it might adopt a side-by-side conformation (Rodley, et al., 1976; Gupta, et al., 1980). Confronted with these new challenges, it became apparent that fiber diffraction, with its limited information content, could no longer resolve questions regarding the detailed conformations that DNA is capable of assuming.

2. OLIGONUCLEOTIDES AND SINGLE CRYSTALS

Chemical methods developed over the past few years for synthesizing DNA make it possible to produce oligonucleotides in sufficient quantity for crystallization experiments. Unlike fibers, single crystals can diffract

58

to high resolution, and they provide large numbers of reflections or experimental measurements. This makes it possible to solve structures and observe fine details of DNA conformation in which very little or no interpretation is required. In an atomic resolution (<1.0 Å) electron density map, every atom is seen directly and no assumptions are required to obtain bond angles, distances, ring pucker, and other features. Studies of single crystals bear the promise of resolving issues regarding the fine details of DNA conformation.

Single crystal X-ray studies of nucleic acid bases, nucleosides, and monomeric nucleotides have been carried out for many years. Single crystal diffraction studies have been conducted in many laboratories that show the purine-pyrimidine base pairs to have many types of hydrogen bonds (Voet and Rich, 1970). The first visualization of the double helix at atomic resolution, however, occurred only 10 years ago with the solution of the structures of rApU (Rosenberg et al., 1973; Seeman et al., 1976) and rGpC (Day et al., 1973; Rosenberg et al., 1976), which were found to form right-handed double helical fragments of the type that had been anticipated for RNA molecules. This work on ribonucleotide fragments laid the groundwork for the subsequent and more extended analysis of deoxyoligonucleotides.

In our research, we have collaborated with J. H. van Boom and his colleagues at Leiden University who developed many of the methods for synthesizing oligonucleotides. Several years ago, van Boom and his colleagues synthesized all four of the self-complementary deoxynucleotide tetramers containing only guanosine and cytosine residues. It seemed likely that these sequences might produce stable oligonucleotide double-helical fragments that could then be crystallized. Three of the four were crystallized and one of them, d(CpGpCpG), crystallized with apparent ease. This suggested that it might be profitable to explore the hexamer d(CpGpCpGpCpG) seen in Figure 1. It was discovered that crystals could be formed readily using magnesium and spermine as cations (Wang et al., 1979). These crystals diffracted to 0.9 Å resolution. It was clear that solution of this crystal structure would not only make it possible to visualize the details of the double helix but would also yield information about hydration and ions.

The hexamer crystal structure was solved by the conventional multiple isomorphous replacement method. Various crystalline heavy metal cation derivatives were obtained to solve the structure. The asymmetric unit of

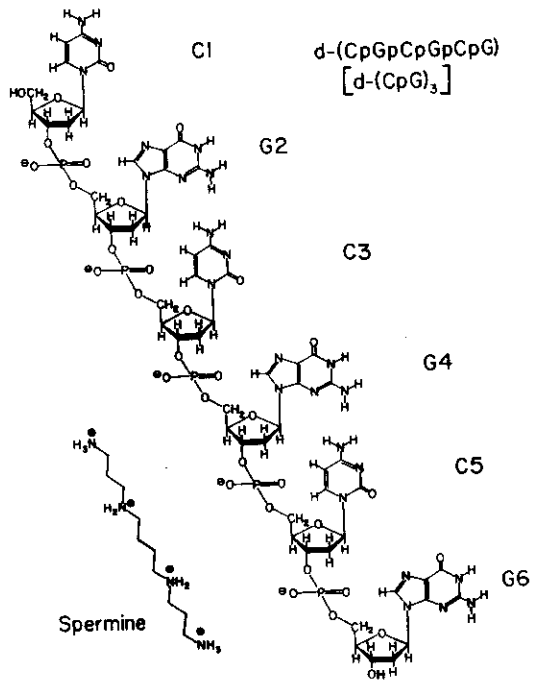


Figure 1. Chemical structure of the hexanucleoside pentaphosphate d(CpGpCpGpCpG) and spermine. Note that only the amino group hydrogen atoms of spermine are shown.

the crystal had one double-helical fragment of DNA containing six base pairs, two spermine molecules, a hydrated magnesium ion, and 62 water molecules. The asymmetric unit had a molecular weight of almost 5,000 daltons and thus was as large as some of the smaller proteins.

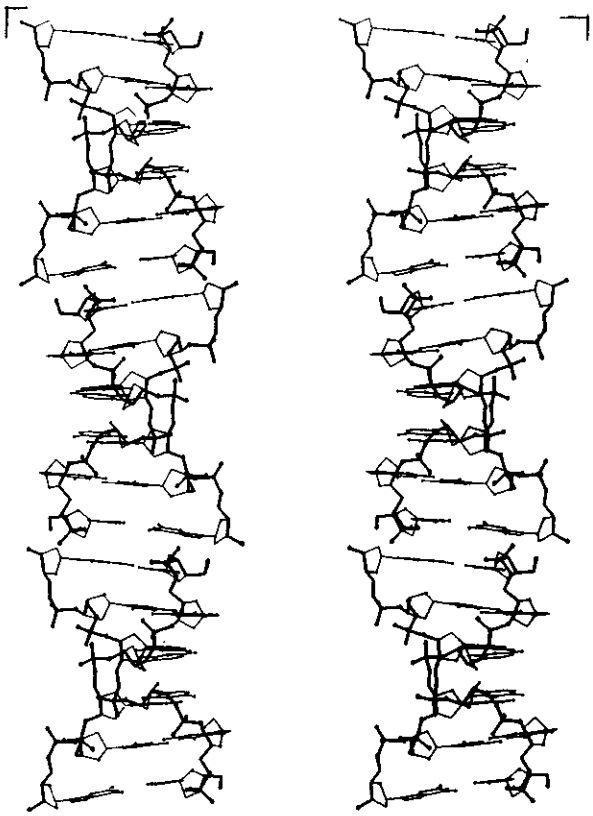


Figure 2. Stereo diagram of three molecules in the Z-DNA crystal lattice as they are stacked along the *c* axis in what looks like a continuous double helix. However, a phosphate group is absent every sixth residue along the chain. Successive base pairs along the helix have the deoxyribose rings orientated so that O1′ is alternately pointing up or down and the same orientation is found in both sugar residues across the base pair. This indicates that the helix asymmetric unit is two nucleotides rather than one as in B-DNA.

3. THE DOUBLE HELIX TURNS LEFT-HANDED

The structure was quite unusual (Figure 2). It was a left-handed double helix. However, it was not a simple left-handed helix but one in which the alternating purine-pyrimidine sequence of the oligonucleotide was used in a unique way. A van der Waals drawing is shown in Figure 3 alongside a drawing of right-handed B-DNA. A comparison is listed in

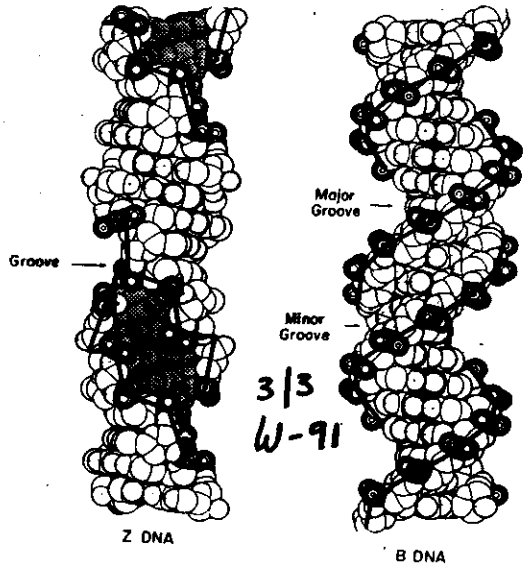


Figure 3. Van der Waals side views of Z-DNA and B-DNA. The irregularity of the Z-DNA backbone is illustrated by the heavy lines, which go from phosphate to phosphate residues along the chain. This includes positions where the phosphate residues are missing in the crystal structure but would be occupied in a continuous double helix. The shaded groove in Z-DNA is quite deep. In contrast, B-DNA has a smooth line connecting the phosphate groups and it has two grooves.

Table 1. The left-handed helix has the phosphates arranged in a zigzag array; hence the name Z-DNA. The Z-DNA helix has 12 base pairs per turn of the helix, which adopts a length of 44.6 Å. This is in contrast to the 10 base pairs per turn with a pitch of 34 Å in right-handed B-DNA. The same Z structure has been found in crystals of tetramers d(CpGpCpG) (Drew et al., 1980; Crawford et al., 1980) as well as in other hexamer crystals (Wang et al., 1981) and in fibers (Arnott, et al., 1980).

We could be confident that the structure was correct because it was solved at atomic resolution. Individual atoms were seen distinctly separated from neighboring atoms and their assignment into structure was straightforward. This is not necessarily the case for medium or lower resolution single crystal analyses. For example, at 3 Å resolution one might have been considerably less confident about the form of the molecule.

The structure of left-handed Z-DNA differs in numerous and significant ways from right-handed B-DNA. Why had the structure not been anticipated? The answer lies in the fact that a number of profound changes occur in the molecule that make it different from the right-handed double helix. One does not form Z-DNA simply by reconstructing the right-

Table 1

	B-DNA	Z-DNA[a]
Helix sense	Right-handed	Left-handed
Residues/turn	10	12 (6 dimers)
Diameter	~20 Å	~18 Å
Rise/residue	3.4 Å	3.7 Å
Helix pitch	34 Å	45 Å
Base pair tilt	6°	7°
Rotation/residue	36°	−60° (per dimer)
Glycosidic torsion angle		
Deoxyguanosine	*anti*	*syn*
Deoxycytidine	*anti*	*anti*
Sugar pucker		
Deoxyguanosine	C2′-*endo*	C3′-*endo*
Deoxycytidine	C2′-*endo*	C2′-*endo*
Distance of P from axis		
dGpC	9.0 Å	8.0 Å
dCpG	9.0 Å	6.9 Å

[a] Average values are given and end effects are excluded.

handed double helix in a left-handed fashion. A number of other conformational alterations are also required.

In the Z-DNA crystal, the backbone has guanosine residues in an unusual conformation. This is illustrated in Figure 4, which shows the disposition of deoxyguanosine as seen in Z-DNA and B-DNA. In Z-DNA the guanine bases adopt the *syn* conformation. The *anti* conformation is found in all the bases in B-DNA and in the cytosine residues in Z-DNA. Nuclear magnetic resonance (NMR) studies in solution have shown that purine residues can adopt the *syn* and *anti* conformations with equal ease (Son et al., 1972). An early analysis of the rotational barriers about the glycosyl bonds suggested that purines can rotate more easily than pyrimidines into the *syn* conformation (Haschemeyer and Rich, 1967).

Further differences are found in the pucker of the sugar ring. The guanosine in Z-DNA (Fig. 4) has the C3′-*endo* conformation (which is the preferred conformation for ribonucleotides), whereas in B-DNA the C2′-*endo* conformation is adopted for all nucleotides. However, the cytidine residues in Z-DNA have the C2′-*endo* conformation of the sugar. The asymmetric unit in Z-DNA is thus a dinucleotide consisting of a cytidine residue in one conformation and a guanosine residue in a different conformation. This is seen diagrammatically in Figure 5, which shows the sugar conformations in B-DNA and Z-DNA molecules. Another feature distinguishing B-DNA from Z-DNA is the orientation of the base pairs relative to the sugar-phosphate backbone. In order to convert from B-DNA to Z-DNA it is necessary for the base pairs to turn over or flip by 180° so that the relationship of the sides of the base pairs is inverted. This is shown diagrammatically in Figure 6, in which four base pairs in the center of B-DNA form Z-DNA. This occurs schematically by an inversion of the base pairs, as shown in the diagram. This "flipping over" of the base pairs is brought about by rotating the purine residues about their glycosyl bond going from *anti* to *syn*. In the case of the pyrimidine residues to which they are paired, both the sugar and the base are turned upside down. It is this inversion of the sugar which introduces a zigzag into the sugar-phosphate backbone in Z-DNA.

These changes in nucleotide conformation result in significant modifications in the stacking relation between successive base pairs along the helix. Figure 7 shows the stacking of the CpG and GpC sequences for Z-DNA and B-DNA. The guanine residues are in the *syn* position, and this results in the imidazole ring of guanine occupying a position near the

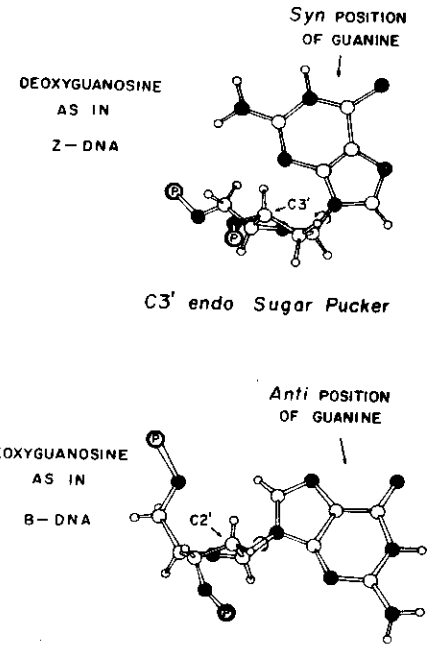


Figure 4. Conformation of deoxyguanosine in B-DNA and in Z-DNA. The sugar is oriented so that the plane defined by C1'–O1'–C4' is horizontal. Atoms lying above this plane are in the *endo* conformation. The C3' is *endo* in Z-DNA, whereas in B-DNA the C2' is *endo*. These two different ring puckers are associated with significant changes in the distance between the phosphorus atoms. In addition, Z-DNA has guanine in the *syn* position, in contrast to the *anti* position in B-DNA. A curved arrow around the glycosyl bond indicates the site of rotation.

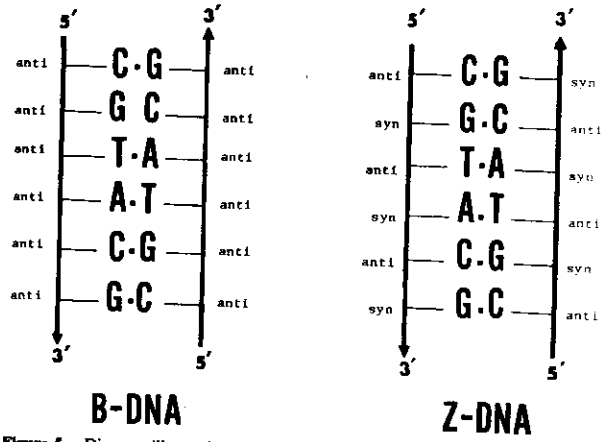


Figure 5. Diagram illustrating the conformation of nucleosides in a segment of B-DNA and Z-DNA. The purines occupying alternative positions along the chain of Z-DNA are in the *syn* conformation.

outside of the helix with the atom C8 near the periphery. This is in contrast to its position in B-DNA, where it is in van der Waals contact with the sugar-phosphate backbone. The stacking of CpG differs considerably from GpC in Z-DNA even though they are fairly similar in B-DNA. The CpG sequence is sheared so that the cytosine residues stack upon each other, but the guanine residues do not stack upon bases at all. Instead, they stack on the sugar of the next residue. This results in a stacking interaction between O1' of the sugar and the pyrimidine-purine ring. The GpC residues, on the other hand, have the bases on each strand stacking upon the bases of the base pair below in a manner similar to that found in right-handed B-DNA.

The groove in Z-DNA is fairly deep compared with that of B-DNA. Figure 8 shows a van der Waals diagram containing three base pairs of Z-DNA. The axis of the molecule is indicated by the solid dot and the

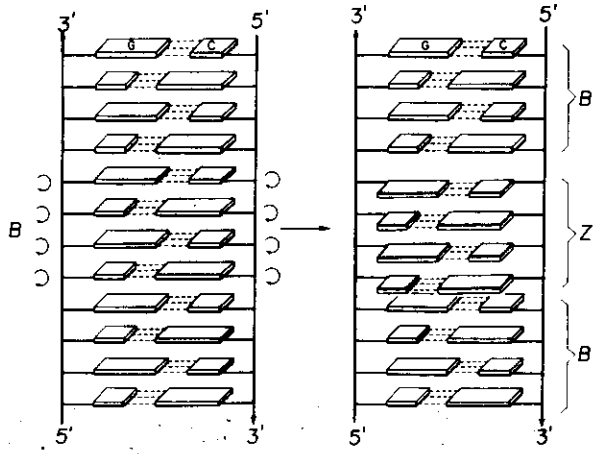


Figure 6. Diagram illustrating the change in topological relationship if a four-base-pair segment of B-DNA were converted to Z-DNA. This conversion could be accomplished by rotation of the bases relative to those in B-DNA. This rotation is shown diagrammatically by shading one surface of the bases. All of the dark shaded areas are at the bottom in B-DNA. In the segment of Z-DNA, however, four of them are turned upward. The turning is indicated by the curved arrows. Rotation of the guanine residues about the glycosyl bond produces deoxyguanosine in the *syn* conformation while for dC residues, both cytosine and deoxyribose are rotated. The altered position of the Z-DNA segment is drawn to indicate that these bases may not be stacking directly on the base pairs in the B-DNA segment.

depth of the groove can be seen quite clearly. Unlike B-DNA, the bases are displaced away from the center of the axis. The helix twists in a clockwise direction moving toward the reader. Thus, the three phosphate groups are visible on the left side while only one is seen on the right. The axis is found very close to the O2 oxygen of the cytosine residue. In this diagram, the imidazole ring of guanine is shown projecting on the outside of the molecule where it is readily accessible to other reagents, ions, or ligands. Figure 9 shows a stereo diagram of Z-DNA tilted by about 30°

so that the groove can be visualized as an indentation in the molecule. It can be seen to extend almost to the center.

End views of Z-DNA and B-DNA are compared in Figure 10. Here the purine residues are shaded and the backbone is drawn in slightly heavier lines. In B-DNA, the purine residues are clustered near the center of the axis while the sugar-phosphate backbone is organized on the outer perimeter of the molecule. In Z-DNA, on the other hand, the purine residues are located near the perimeter of the molecule with the imidazole groups pointing to the outside. This diagram also illustrates the fact that

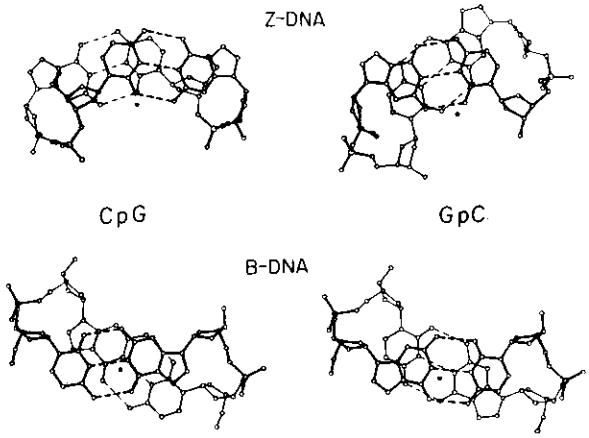


Figure 7. Stacking diagram illustrating the overlap of successive bases along Z-DNA and B-DNA chains. The base pair drawn with heavier lines is stacked above the pair drawn with lighter lines. The left-hand column represents d(CpG) sequences of both Z-DNA and B-DNA, while d(GpC) sequences are on the right. The direction of the deoxyribose–phosphate chains is the same in all these diagrams. Note that the minor groove in B-DNA is found at the top of the B-DNA diagrams while the analogous side of the base pairs is found at the bottom of the Z-DNA diagrams. The solid black dot indicates the helical axis.

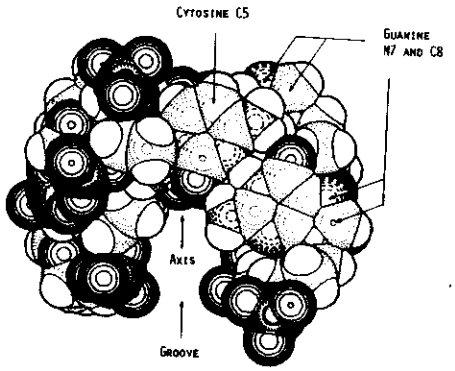


Figure 8. Van der Waals drawing of a fragment of Z-DNA as viewed down the axis of the helix. Three base pairs are shown, and the deep groove is seen to extend almost to the axis of the molecule. In these three base pairs the groove rotates clockwise, moving toward the reader. For that reason, three phosphates are visible on the left and only one on the right. The N7 and C8 atoms of guanine are near the outer wall of the molecule. The solid black dot indicates the axis of the molecule.

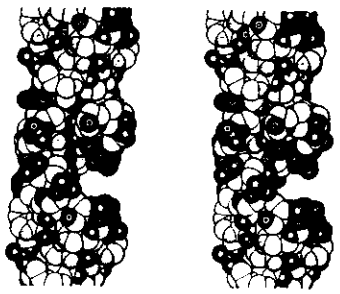


Figure 9. Van der Waals stereo drawing of Z-DNA in which the axis of the molecule is tipped away from the reader in such a way that the depth of the helical groove can be visualized. The prominent ridges of the sugar-phosphate backbone are readily visible.

Z-DNA has a slightly smaller cross-sectional diameter than B-DNA. The van der Waals diameter of B-DNA is close to 20 Å whereas Z-DNA has a diameter of only 18 Å.

The different forms of DNA are distinctive with regard to the position of the molecular axis around which the base pairs turn. This is shown in Figure 11, which illustrates the axes of A-DNA, B-DNA, and Z-DNA superimposed on a GC base pair. In B-DNA the axis is near the center of the base pair, and this accounts for the clustering of the purine residues, shaded in the center of the diagram shown in Figure 10. In Z-DNA, on the other hand, the axis is close to the O2 atom of the cytosine residue and it is no longer on the perpendicular bisector of the C1′–C1′ line in Figure 11. In both A-DNA and B-DNA, where the asymmetric unit is one base pair, the axis must be on this perpendicular *bisectrix*. In Z-DNA,

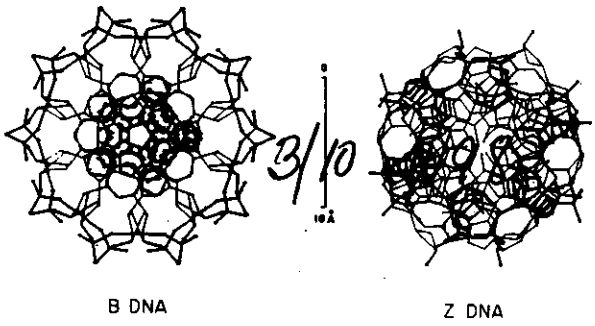


Figure 10. End views of B-DNA and Z-DNA are illustrated in which the guanine residues of one strand have been shaded. The Z-DNA figure represents a view down the complete *c* axis of the crystal structure encompassing two molecules. The shaded guanine residues illustrate the approximate sixfold symmetry. The imidazole part of the guanine residue forms a segment of the outer cylindrical wall of the molecule together with the phosphate residues. The B-DNA figure represents one full helix turn. In contrast to Z-DNA, the guanine residues in B-DNA are located closer to the center of the molecule and the phosphates are on the outside. The B-DNA diagram is drawn from idealized regular coordinates obtained from fiber diffraction studies. The actual molecule is likely to have irregularities similar to those seen in Z-DNA.

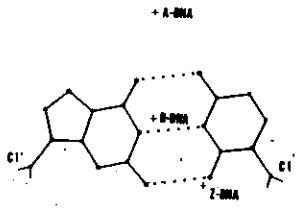


Figure 11. Diagram illustrating the positions of the axes in three major DNA conformations. The axes of A-DNA and B-DNA are found on a line perpendicular to C1′–C1′ and halfway between them. This is necessary because of the pseudodyad found in both of these structures in the plane of the base pair. The axis of Z-DNA is not on that line because the asymmetric unit consists of two bases and there is no longer a pseudodyad in the plane of the pairs. The base pairs are far removed from the A-DNA axis, so that the major groove is deep and the minor groove on the external surface in the molecule is shallow. The Z-DNA axis is also somewhat removed from the center, so that its groove is deep and the external surface of Z-DNA is what corresponds to the major groove side of the base pair. In B-DNA the axis goes through the middle of the base pairs, giving rise to two grooves of slightly unequal size.

where the asymmetric unit contains two base pairs, the axis can be anywhere—it is not confined to the perpendicular bisectrix of the C1′–C1′ vector. In A-DNA, the axis is well above the base pair, reflecting the fact that the double helix has a very deep major groove in A-DNA and a very shallow minor groove.

In a regular B-DNA helix there are 10 residues per turn of the helix, which spans 34 Å. The twist angle is the angle between adjacent base pairs going along the helix. For a 10-fold helix this would be 36°. The twist angle is constant in going from base pair to base pair along the helix in idealized B-DNA. The situation with Z-DNA is quite different. Figure 12 shows that, in contrast to the constant relationship between the base pairs, the twist angle between adjacent base pairs is the same for CpG as it is for GpC, that is, +36° (Wang et al., 1982). The sequence CpG has very little rotation between adjacent base pairs while the sequence GpC has a strong negative rotation. This is evident from looking at the stacking diagrams in Figure 7 and is shown quantitatively in Figure 12, which plots

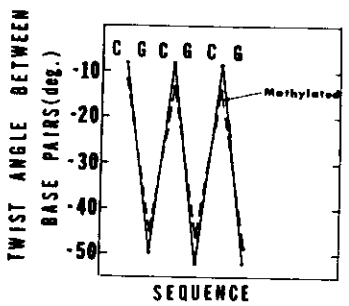


Figure 12. Twist angle between adjacent base pairs is plotted for methylated and nonmethylated Z-DNA crystal structures. The points are positioned between the bases listed in the sequence, and each point represents the twist angle between the two base pairs on either side of the point. The twist angle is the angle between adjacent lines connecting C1′ atoms of nucleoside base pairs.

the twist angle of adjacent sequences along the helix axis. The solid line shows that the twist angle is close to −8° for the CpG sequences and clost to −52° for the GpC sequences.

4. CONFORMATIONAL CHANGES IN Z-DNA

There is a pseudotwofold, or pseudodyad, axis between the base pairs which relates one chain to another. These dyads are not used in the symmetry of the crystal lattice but are a consequence of interactions within the double helix. The extent to which these are true dyads can be seen in Figure 13, which shows the conformation of one nucleotide strand (C1 to G6) compared with the other strand (C7 to G12) (Wang et al., 1981). In creating Figure 13, the two strands, which are antiparallel to each other in the crystal lattice, were rotated to show the same view of both. The two strands have a remarkably similar conformation, reflecting the internal regularity of the molecule. A significant exception is the conformation of the phosphate group between G4 and C5 on one strand and G10 and C11 on the other. The conformation of G10pC11 is similar to the

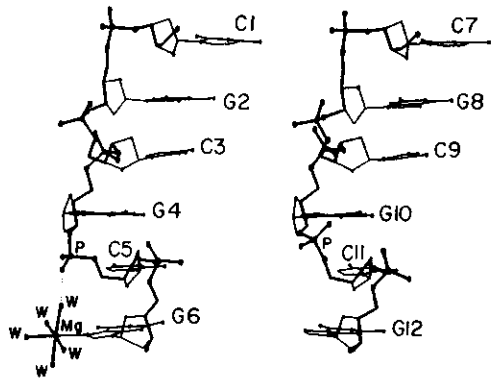


Figure 13. Diagram showing the conformation of the two independent hexanucleotide molecules in the spermine-magnesium Z-DNA crystal. In the crystal the two strands are antiparallel, but in this diagram they have been arranged in a parallel alignment to show the similarity of the two chains. The two independent chains are very similar except for the linkage G4pC5. That phosphate group on the left has rotated in such a manner that its oxygen is forming a hydrogen bond with a water molecule (W) in the octahedral coordination shell surrounding the magnesium ion, which is complexed to N7 of guanine 6. This conformation is called Z_{II}, while the conformation for the rest of the phosphate groups is called Z_I.

conformation of the other two GpC phosphates in the molecule. However, the phosphate group in G4pC5 has rotated and the phosphorus atom is approximately 1 Å away from the position it occupies in the other three GpC linkages. Further, the C3′-O3′-P-O5′ dihedral angle has rotated about 125°. In the G4pC5 step, the phosphate group is in a position where it forms a hydrogen bond to a hydrated magnesium ion, which is in turn complexed to the imidazole N7 of guanine 6, as shown in Figure 13. The magnesium ion has octahedral coordination in which one ligand is guanine N7 and the other five are water molecules. We designate the majority conformation found in linkages G2pC3, G8pC9, and G10pC11 as Z_I and the conformation found in G4pC5 as Z_{II}. Conformation Z_I is *gauche(−)trans* for the phosphodiester conformation, while Z_{II} is *gauche(+)trans* (Sundaralingam, 1969).

Analysis of several crystals suggests that left-handed Z-DNA molecules may exist in a mixture of conformations ranging from pure Z_I through mixtures of Z_I and Z_{II} to pure Z_{II}. The exact proportions of these two may depend on the ionic composition and hydration structure of the lattice.

One of the principal differences between Z-DNA and B-DNA is the distance between phosphate groups on opposite strands of the molecule. The distance of the closest approach in B-DNA is across the minor groove, where phosphate groups are 11.7 Å apart. In marked contrast, the closest phosphate groups are only 7.7 and 8.6 Å apart in Z_I-DNA and Z_{II}-DNA. These are distances between the phosphates of CpG sequences. The closest distances across the chains between phosphates of CpG and GpC sequences are 11.2 and 13.5 Å for Z_I and Z_{II}, respectively. The longer distance for Z_{II} reflects the fact that its GpC phosphates are rotated away from the groove.

The differences between Z_I and Z_{II} are shown in Figure 14 which has end views of both conformations together with B-DNA (Wang et al., 1981). As these are all idealized conformations, Z_I and Z_{II} have perfect sixfold symmetry and B-DNA has tenfold symmetry. The large number of differences among A-DNA, B-DNA, and Z-DNA are summarized in Figure 15. In B-DNA, dyad axes are located in the plane of the base pairs

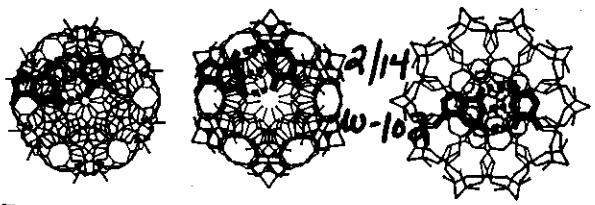


Figure 14. End views of the regular idealized helical forms of Z_I, Z_{II}, and B-DNA. Heavier lines are used for the phosphate-ribose backbone. A guanine-cytosine base pair is shown by shading. The difference in the positions of the base pairs is quite striking; they are near the center of B-DNA but at the periphery of Z-DNA.

as well as between the base pairs. In Z-DNA the dyad axis is preserved between the base pairs but no longer is there a twofold symmetry element in the plane of the base pairs.

Figure 15 further demonstrates that there are appreciable differences in the organization of the polynucleotide chains in Z-DNA compared with B-DNA. In a sense, these differences are so numerous that it is easy to understand why the Z form was not suggested by model building. In retrospect, a number of clues in the literature might now be interpreted as anticipating the Z-DNA structure. For example, the manner by which the sugar pucker could change from C3′-*endo* to C2′-*endo* and the observation of *syn* conformations in purines but not in pyrimidines were well known. Further, the possibility of forming left-handed structures did receive due consideration. Evidence for conformational changes from solution studies of polymers with alternating guanine and cytosine residues was a strong indication. Nevertheless, it would have been difficult to predict from a study of B-DNA all of the necessary changes required to produce the Z structure.

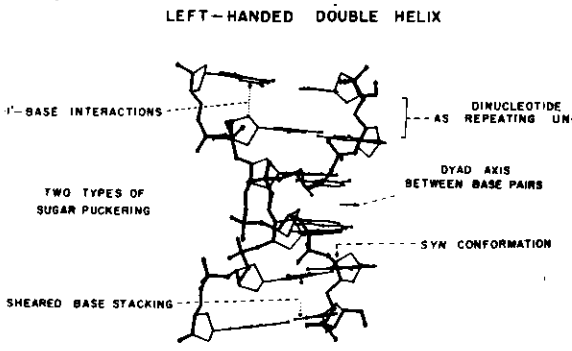


Figure 15. Schematic diagram showing a projection of the left-handed double helix of the spermine-magnesium d(CpGpCpGpCpG) hexamer. Seven structural features found in this left-handed DNA conformation are different from those found in right handed B-DNA, as indicated in the diagram and discussed in the text.

Table 2. Crystal Data of the Z-DNA Family

Compound	Cell Constants a(Å)	b(Å)	c(Å)	Space Group	Resolution (Å)	Current *R* value	Reference
Orthorhombic Lattice							
d(CGCGCG) [Mg^{2+} spermine]	17.88	31.55	44.58	$P2_12_12_1$	0.9	0.13	Wang et al., 1981
d(CGCGCG) [Mg^{2+} spermidine]	17.96	31.19	44.73	$P2_12_12_1$	1.3	0.17	Wang et al., 1981
d(CGCGCG) [Ba^{2+} spermine]	17.94	31.54	44.68	$P2_12_12_1$	1.1	0.18	Wang et al., 1981
d(CGCGCG) [Mg^{2+} (and/or Na^+)]	17.98	30.94	44.81	$P2_12_12_1$	1.8	0.13	Wang et al., 1981
d(m^{2}CGm^{2}CGm^{2}CG) [Mg^{2+} spermine]	17.76	30.57	45.42	$P2_12_12_1$	1.3	0.16	Fujii et al., 1982
d(brCGTAbrCG) [Mg^{2+}/Na^+]	17.85	30.74	44.78	$P2_12_12_1$	1.5	0.17	Wang et al., 1984
d(m^{2}CGTAm^{2}CG) [Mg^{2+}/Na^+]	17.91	30.43	44.96	$P2_12_12_1$	1.2	0.16	Wang et al., 1984
d(CGCG) [Mg^{2+}(and/or Na^+)]	19.50	31.27	64.67	$C222_1$	1.5	0.20	Drew et al., 1980
Hexagonal Lattice							
d(CGCGCGCG) [Mg^{2+}(and/or Na^+)]	31.27	31.27	43.56	$P6_5$	1.6	0.19	Fujii et al., 1984
d(CGCATGCG) [Mg^{2+}(and/or Na^+)] (MPD-isopropanol form)	30.90	30.90	43.14	$P6_5$	2.5	0.16	Fujii et al., 1984
d(CGCATGCG) [Mg^{2+}(and/or Na^+)] (MPD-$MgCl_2$ form)	31.05	31.05	43.24	$P6_5$	1.9	0.24	Fujii et al., 1984
d(CGCG) [Mg^{2+} spermine]	31.25	31.25	44.06	$P6_4$	1.5	0.19	Crawford et al., 1980
d(CGCG) [Mg^{2+}(and/or Na^+)]	30.92	30.92	43.29	$P6_4$	1.5	0.21	Crawford et al., 1980

5. LATTICES

Z-DNA structures have been found in crystalline tetramers, hexamers, and octamers. In all of these crystals (Table 2) two types of lattices are found, orthorhombic and hexagonal (Crawford et al., 1980). The most common type of orthorhombic lattice and the single hexagonal lattice are shown in Figure 16. The structure of a tetramer $(dC\text{-}dG)_2$ has been solved in the hexagonal lattice, and the hexamer $(dC\text{-}dG)_3$ in the orthorhombic lattice. There are two types of molecules in the hexagonal unit cell: those at the corners of the diamond-shaped figures, and those that are internal. The molecules at the corners are symmetry related by sixfold screw axes and the molecules inside by threefold screw axes. The c axis of 44.6 Å is such that it will just accommodate a helix with 12 base pairs per turn, as observed in the hexamer crystal (Wang et al., 1979). Thus, the symmetry used by the internal molecules (threefold screw axes) allowed precisely what one would expect with three segments of tetramer forming a left-handed helix with 12 base pairs per turn of the helix.

6. STUDIES OF Z-DNA IN SOLUTION

The earliest evidence suggesting the existence of Z-DNA was the work of Pohl and Jovin (1972). They showed that increasing the NaCl concentration of a poly(dG-dC) solution to 4 M produced a near inversion of the circular dichroism spectrum. In Z-DNA, the phosphate groups on op-

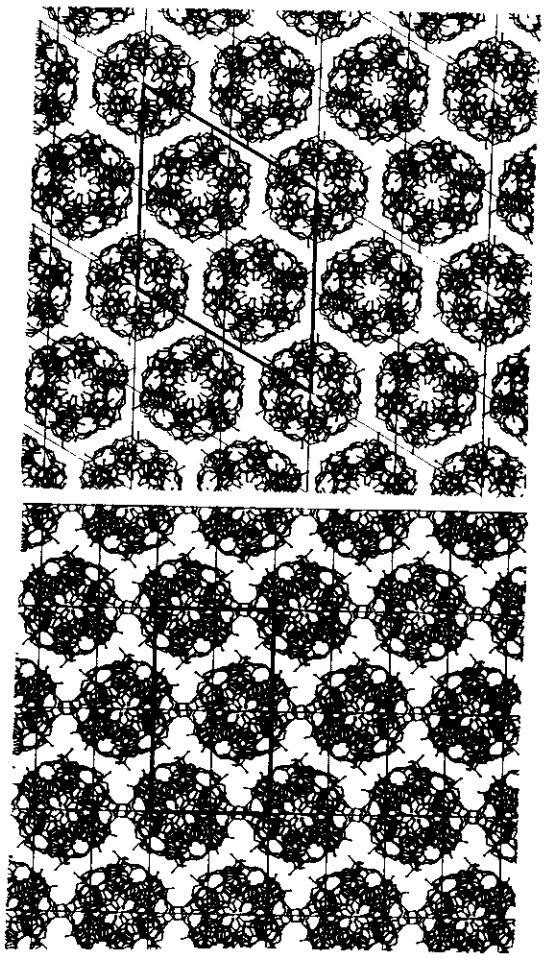


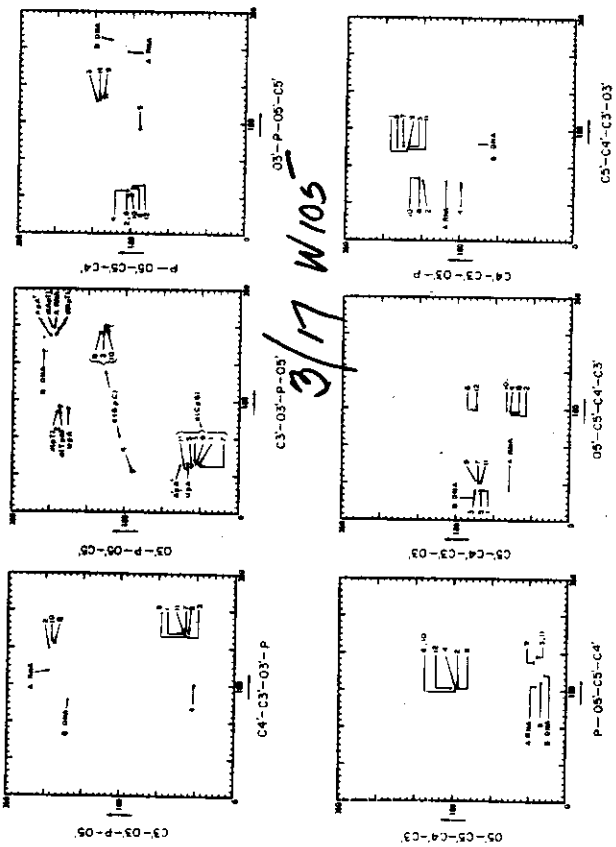
Figure 16. Diagram of molecular packing of Z-DNA molecules as viewed down the helix axis. (Left) Packing of the hexamer $d(CG)_3$ is shown in the outlined orthorhombic unit cell. (Right) Closely related packing of the tetramer $d(CG)_2$ is shown in a hexagonal unit cell. The horizontal axes in both the orthorhombic and hexagonal unit cells are close to 31.5 Å. The hexagonal unit cell is somewhat larger than the orthorhombic and it can be seen to contain two different kinds of molecules: those found at the corners of the diamond-shaped unit cell and the two that are internal. Those molecules found at the corners of the unit cell are related by sixfold screw axis symmetry whereas the internal molecules by threefold screw axis symmetry. The molecules in the orthorhombic hexamer crystal are related by twofold screw symmetry. Only the oligonucleotides are plotted while spermine ions and solvent molecules are omitted. It is interesting that all of the molecules in both lattices are completely sheathed in water, with no direct contacts between the oligonucleotides. The apparent contact between the hexamer nucleotides does not actually exist because the two parts of the molecule that appear to be in contact are actually separated by half the unit cell.

posite chains are closer together than they are across the minor groove of B-DNA. This implies that the structure is likely to be stabilized by increasing the concentration of cations. Thus it was reasonable to believe that the high salt form of poly(dG-dC) might be Z-DNA. Conclusive proof was obtained from the observation that the Raman spectra of the high salt and the low salt forms of poly(dG-dC) differ from each other in a distinctive manner (Pohl et al., 1973). Examination of the Raman spectrum of the Z-DNA hexamer crystals revealed that they had a spectrum almost identical to that of the high salt form of poly(dG-dC) and quite different from that of the low salt form (Thamann et al., 1981). Thus the high salt form of poly(dG-dC) was established to be Z-DNA.

In solution, there is an equilibrium between B-DNA and Z-DNA for poly(dG-dC), where the equilibrium point is determined by the environment. The midpoint for the conversion is 2.5 M Na^+ or 0.7 M Mg^{2+} (Pohl and Jovin, 1972). Evidence for the existence of an equilibrium is provided by the fact that the original solution of the hexamer which crystallized as Z-DNA was a low salt solution (Wang et al., 1979) and there is no evidence for Z-DNA formation under those conditions, as judged by the circular dichroism. The crystals were probably nucleated by a minor component of Z-DNA in the low-salt solution, and as the crystals grew by mass action all of the material was converted into the Z form. Recognition of this equilibrium is a useful step toward developing an understanding of the effect of various modifications of the polymer on the ability to form Z-DNA. It is fair to presume that there is a similar equilibrium in native DNA, which is strongly influenced by base sequence as well as methylations, as discussed below.

7. TORSIONAL ANGLES IN Z-DNA

We have summarized the distribution of torsion angles in Z-DNA in Figure 17. It contains plots of the various torsion angles listed against each other



gure 17. Torsion angles are shown for the various linkages in the Z-DNA crystal, $CG)_3$. The numbers refer to the nucleotides, which are 1–6 on one chain and 7–12 on e other. These are shown explicitly in Figure 13. The torsion angles are plotted against ch other in order. The conformational change that gives rise to Z_{II} (Figure 13) produces e torsion angles where residues 4 and 5 stand alone. For comparative purposes, A-RNA r A-DNA) and B-DNA are included. The shading is added to illustrate the clustering of e data.

Table 3. Dihedral Angles of DNA

	α	β	γ	δ	ε	ξ	χ
Z_I-DNA							
C	−137	−139	56	138	−94	80	−159
G	47	179	−169	99	−104	−69	68
Z_{II}-DNA							
C	146	164	66	147	−100	74	−148
G	92	−167	157	94	−179	55	62
A-DNA	−90	−149	47	83	−175	−45	−154
B-DNA	−41	136	38	139	−133	−157	−102

for the most highly refined $(dC\text{-}dG)_3$ structure. The pronounced regularity of the structure is due to the clustering together of angles in groups. Notable exceptions associated with the Z_I and Z_{II} conformation are seen in torsion angles of nucleotides 4 and 5 (see Figure 13 for the numbering system). For comparative purposes, torsion angles are listed for right-handed B-DNA and A-RNA (which is very similar to A-DNA) (Table 3). The differences are apparent and they represent basic alternatives in the way the two molecules are folded.

8. METHYLATION OF CYTOSINE STABILIZES Z-DNA

In higher eukaryotes, one of the most important modifications of DNA is methylation of cytosine on the 5 position where the cytosine is followed by a guanine. Many studies have shown that methylation of DNA is often but not always associated with cessation of transcription. Conversely, demethylation is associated with the initiation of transcription or the expression of the gene (Razin and Riggs, 1980; Ehrlich and Wang, 1981). One of the features apparent in the Z-DNA structure was that CG sequences play an important role in fixing its conformation. Accordingly, it is reasonable to ask to what extent methylation might modify the distribution between right-handed B-DNA and left-handed Z-DNA. Behe and Felsenfeld (1981) addressed this problem directly by synthesizing poly(dG-m^5dC), which is fully methylated. Methylation of the cytosine

residue on the 5 position has a profound effect in altering the equilibrium between B-DNA and Z-DNA in solution. If one has poly(dG-dC) and poly(dG-m^5dC) in 50 m*M* NaCl, both molecules are in the right-handed B-DNA conformation. In order to convert poly(dG-dC) to left-handed Z-DNA, the magnesium concentration must be raised to 760 m*M* (Pohl and Jovin, 1972). However, for the fully methylated polymer, 0.6 m*M* is adequate to convert it to Z-DNA. There is a decrease by three orders of magnitude in the amount of magnesium ions needed to stabilize Z-DNA if the polymer is methylated. Z-DNA is the form of poly(dG-m^5dC) in a physiological salt solution.

Figure 18 shows a van der Waals diagram of Z-DNA and the crystal structure of $(m^5dC\text{-}dG)_3$ (Fujii et al., 1982). The molecules have the same orientation. The arrow on Z-DNA points to a depression on the surface of the molecule, while on the right the methyl group of m^5C can be seen to fill the depression. Stereo views are presented in Figure 19 of the unmethylated molecule, in comparison to the methylated molecule in Figure 20. Table 4 compares features of the methylated and nonmethylated hexamer crystals.

The general form of the methylated Z-DNA molecule is similar to that seen in the unmethylated molecule. This is consistent with the observation that antibodies raised against nonmethylated Z-DNA can also recognize Z-DNA formed by the methylated polymer (Lafer et al., 1981). However, there have been some alterations in the geometry of the molecule, principally reflecting the methyl group proximity to the carbon atoms C1′ and C2′ of the adjacent guanosine residue, and this has produced minor alterations in the helix. There are slight changes as well in the relative positions of adjacent base pairs in the two structures and also some differences in the ring pucker. The major difference is in the relative flatness of the C3′-*endo* conformation of the deoxyribose ring of deoxyguanosine in the methylated polymer as compared with the nonmethylated polymer. The reason for the observed change in the ring is its interaction with the methyl group. There is also a slight change in the helical twist angle, as shown in Figure 12. The net effect of changing the twist angle is to bring the two methyl groups together. In Figure 20 it can be seen that the methyl groups on opposite strands are rather close to one another. The carbon atoms are 4.6 Å apart, almost in van der Waals contact. If one placed a methyl group on the C5 position of cytosine in the unmethylated structure,

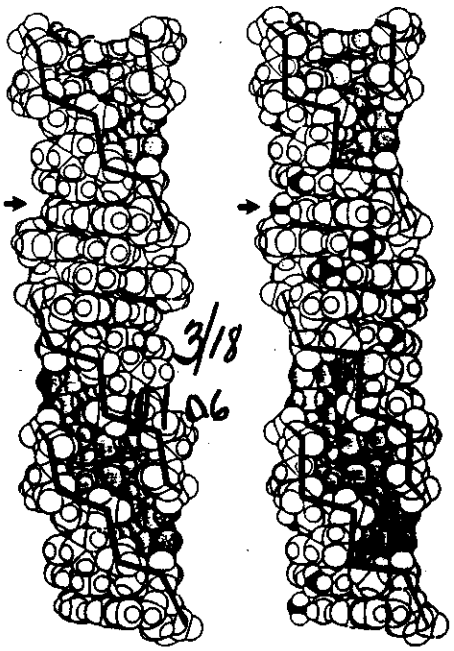


Figure 18. Two van der Waals drawings showing the structure of Z-DNA in both its unmethylated and methylated forms as determined in single-crystal structures of $(dC\text{-}dG)_3$ and $(m^5dC\text{-}dG)_3$, respectively. The groove in the molecule is shown by the shading. The black zigzag lines go from phosphate group to phosphate group to show the arrangement of the sugar-phosphate backbone. In the methylated polymer, the methyl groups on the C5 position of cytosine are drawn in black. The arrow shows that a depression found in the unmethylated polymer is filled by methyl groups. The methyl group indicated by the arrow is in close contact with the imidazole ring of guanosine above it and the C1′ and C2′ atoms of the sugar ring.

the two methyl groups would be a distance of 5.2 Å apart. Thus a shortening of almost 0.6 Å in the distance between methyl groups on opposite strands is a consequence of the change in twist angle described above.

Figures 18 and 20 show that the methyl group occupies a somewhat protected position, recessed slightly on the surface of the molecule so that it is under the imidazole group of guanine with which it is in van der Waals contact (3.4 Å). The methyl group is 3.6–3.8 Å away from C2′ of guanosine and 4.2 Å from C1′. There is thus a close contact between the methyl

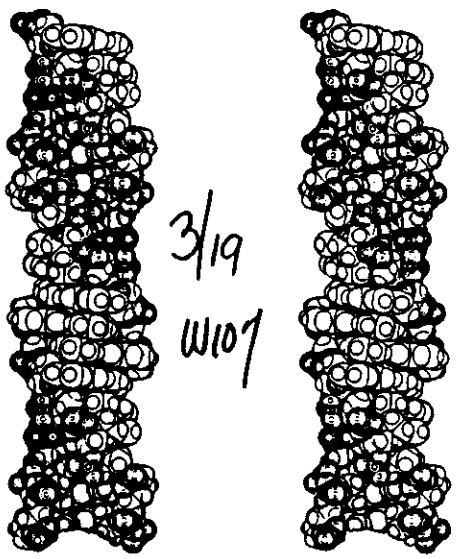


Figure 19. Stereo diagrams of a portion of the Z-DNA helix found in unmethylated hexamer crystals. Van der Waals models are drawn in which the oxygens in the backbone are indicated by circles and the phosphate groups by circles with crossed lines. A slight depression on the convex outer surface of the molecule occurs because the guanine imidazole rings project farther away from the axis than the cytosine rings.

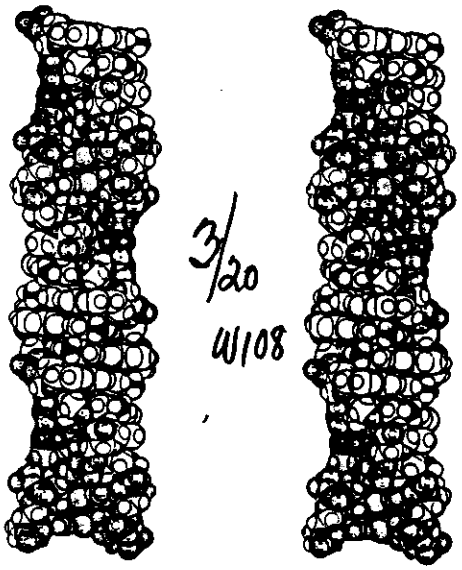


Figure 20. Stereo diagram of the methylated hexamer. The methyl groups are shaded solid black. Three hexamer segments are shown in each diagram just as they appear in the crystal. Every sixth phosphate group is missing because the molecule in the crystal is a hexanucleoside pentaphosphate. The methyl groups are close together and they fill part of the depression on the surface caused by the overhanging imidazole rings of guanine, which protrudes from the center of the molecule. This can be seen easily at the side of the molecule where the methyl groups effectively fill a depression which is visible in the nonmethylated polymer of Figure 18.

group and the sugar residue of the adjacent guanosine (5′ side) as well as its imidazole ring. The reason for the slight conformational change of the methylated hexamer compared with the nonmethylated molecule can be seen if one places a methyl group on the 5 position of cytosine on the unmethylated polymer. When such a structure is made, the distance between this methyl group and guanosine C2′ is 3.2 Å, which is too short a distance for van der Waals contact. That distance is relaxed in the actual structure of the methylated polymer to a distance of 3.6 to 3.8 Å in the various residues. In order to relieve an unacceptable van der Waals contact, the molecule has readjusted itself slightly to produce a change in the helix twist angle and flatten the guanosine sugar ring somewhat.

Table 4. Comparison of $(m^5dC\text{-}dG)_3$ and $(dC\text{-}dG)_3$

	$(m^5dC\text{-}dG)_3$	$(dC\text{-}dG)_3$
Cell constants		
a (Å)	17.76	17.88
b (Å)	30.57	31.55
c (Å)	45.42	44.58
Space group	$P2_12_12_1$	$P2_12_12_1$
Helical twist angle ($\pm \sigma$)		
CpG	−13° (1)	−8° (1)
GpC	−46° (1)	−51° (2)
Sugar conformation[a,b]		
deoxycytidine	C2′-*endo* $\delta = 141°$ ($P = 149°$, $\tau_m = 41°$)	C2′-*endo* $\delta = 146°$ ($P = 153°$, $\tau_m = 35°$)
deoxyguanosine (excluding terminal)	C3′-*endo* $\delta = 94°$ ($P = 30°$, $\tau_m = 19°$)	C3′-*endo* $\delta = 97°$ ($P = 27°$, $\tau_m = 31°$)
Glycosyl orientation[a]		
cytosine	*anti* $\chi = -157°$	*anti* $\chi = -151°$
guanine	*syn* $\chi = 69°$	*syn* $\chi = 67°$

[a] Torsional angles are defined as O3′—P—O5′—C5′—C4′—C3′—O3′—P—O5′, and χ is the glycosyl torsion angle.
[b] P and τ_m are, respectively, the phase angle of pseudorotation and the degree of pucker.

The disposition of the methyl group of 5-methylcytosine in Z-DNA, compared with the position occupied by that methyl group in right-handed B-DNA, is illustrated in the stereo diagrams of Figures 20 and 21. The black methyl groups on opposite strands are almost touching in Z-DNA (Fig. 20) and are far apart in B-DNA (Fig. 21). The change in the molecule is associated with only small alterations in the lattice. A projection down the c axis is shown in Figure 22. There are small changes, but generally the packing interactions are similar.

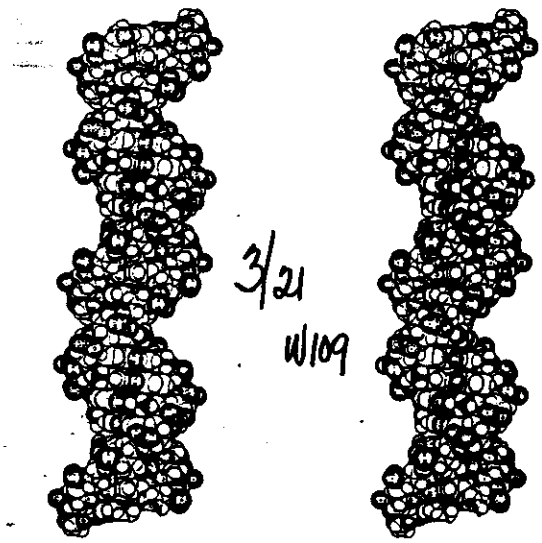


Figure 21. Stereo diagram of the methylated B-DNA with the sequence $(m^5dC\text{-}dG)_n$. Note that the black methyl groups on cytosine residues of opposite strands are far away from each other, in contrast to Figure 19 where the methyl groups are close together.

Another view comparing the methyl groups of m^5dC in Z-DNA and B-DNA is shown in the van der Waals end view of the sequences CpG and GpC in the methylated form (Fig. 23). The upper base pair has different atoms distinguished by shading, but the lower base pair shows only the outline of the bases. The methyl carbon atom on the lower base pair is shaded gray. If one looks at the sequence CpG of Z-DNA, the methyl groups on the two base pairs are fairly close together. Further, there is a limited accessibility to the methyl group by solvent water molecules, which would be found atop of the base pair in the diagram. Accessibility

to the gray methyl group of the bottom base pair is limited because of the presence of the amino group and carbonyl oxygen atom of the upper base pair. This situation is in marked contrast to the methyl group of the CpG sequence in B-DNA, shown in the lower part of Figure 23. It can be seen that the methyl group in B-DNA is thrust strongly into the solvent region so that water molecules have access to it as well as to contiguous sections of the cytosine ring. A difference in accessibility of the GpC residues is shown in the other two diagrams. In the Z conformation the methyl group of the upper base pair is in van der Waals contact with the imidazole group of the guanine below, which projects considerably further away from the center of the molecule than the methyl group. In the GpC sequence of B-DNA, the methyl is somewhat protected by the imidazole group of guanine below it but the protection is less than that seen in the Z-DNA structure.

There is also a reordering of solvent molecules around the outside of the Z-DNA helix due to the additional methyl group. Figure 24 shows the electron density found in a 3 Å thick section through a GC base pair in the Z-DNA helix. The water molecule W is hydrogen bonding to the amino group in the N4 position of cytosine. This water molecule is 2.9 Å away from the amino group. The water molecule forms an angle of 145° between W–N4 and the C4–N4 bond of the cytosine residue. The amino group is in the planar trigonal conformation, so one would expect this to exhibit an angle close to 120°. In the nonmethylated structure that same water molecule is found hydrogen bonded to the N4 position of cytosine, but in that structure the angle between the C4–N4 bond of the cytosine and the water–N4 hydrogen bond is 112°. Thus the presence of the methyl group has effectively forced the water molecule away from the position occupied by the methyl group. This demonstrates that the packing of water molecules is different in the primary hydration shell of the methylated form of Z-DNA and the nonmethylated form.

These observations strongly suggest that the Z-DNA methyl group making close van der Waals contact with both the imidazole group of guanine and the carbon atoms of the sugar residue of guanosine is stabilized by hydrophobic interactions with these residues and is somewhat shielded from surrounding water molecules. In contrast, the methyl group on C5 of cytosine in B-DNA has a much greater surface area accessible to solvent water molecules. It is likely that the stabilization of Z-DNA

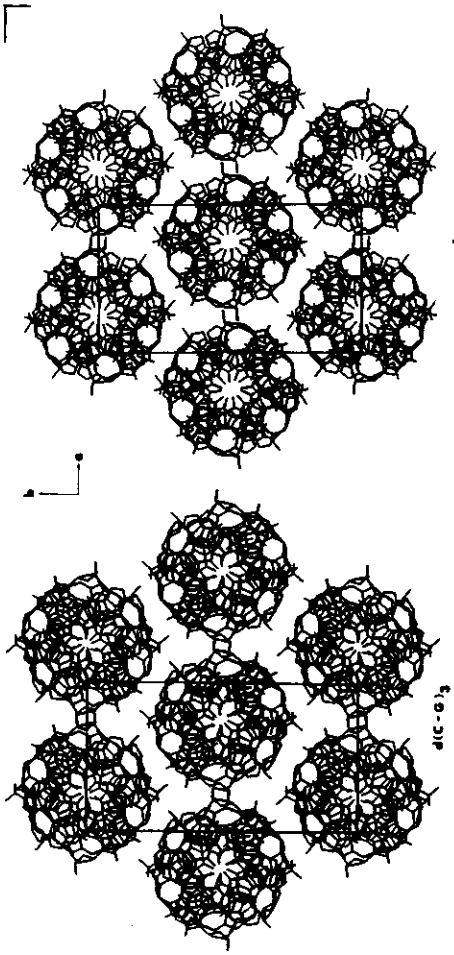


Figure 22. End views of the lattice of $d(C\text{-}G)_3$ and $d(m^5C\text{-}G)_3$. Both lattices are orthorhombic and have similar but not identical dimensions. The methylated crystal has its cytosine residues somewhat farther away from the axis compared with the unmethylated crystal.

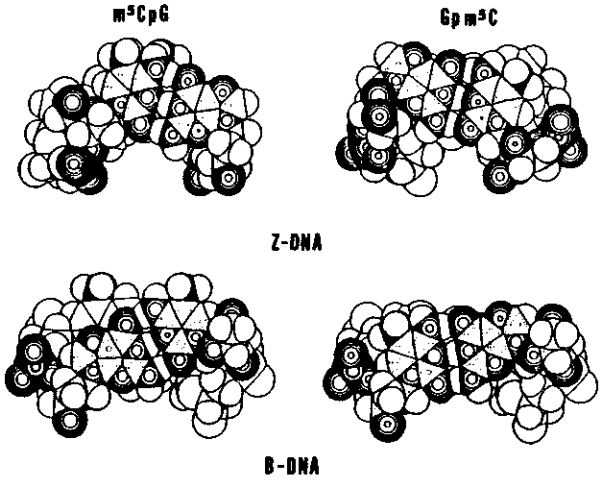


Figure 23. Fragments of Z-DNA and B-DNA are shown in van der Waals diagrams containing two base pairs with the sequences m^5CpG and Gpm^5C. The base pair closer to the reader has shaded atoms, while the base pair away from the reader has the atoms shown only in outline. The methyl group on 5-methylcytosine in the upper base pair closer to the reader is solid black, while the methyl group attached to the lower base pair is shaded gray. The two diagrams at the bottom show the methyl groups in B-DNA, which are more exposed to solvent water molecules than are the methyl groups in the upper two base pairs in Z-DNA.

on methylation relative to B-DNA is due to the destabilization of B-DNA from the methyl group in the major groove interacting with water molecules and, secondly, a stabilization of Z-DNA itself through the formation of a hydrophobic area on the surface of the molecule in which the methyl group fills a slight depression in the surface of the Z-DNA helix.

In vivo, especially for higher eukaryotes, the effects of methylation of CpG sequences in DNA are associated with an inhibition of RNA synthesis. Behe and Felsenfeld (1981) have shown that Z-DNA formation in poly(dG-dm^5C) is facilitated by small amounts of cations, especially the

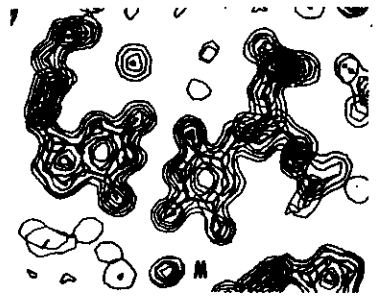


Figure 24. An electron density map is shown for a section of the methylated polymer that encloses one base pair. The electron density map covers a 3-Å-thick section of the map perpendicular to the *c* axis. The 5-methylcytosine of the base pair is on the left and the guanosine residue on the right. A water molecule is hydrogen-bonded to the amino group on the 4 position of cytosine. The presence of the methyl group nearby forces that water molecule to occupy a position closer to the line of the cytosine C4–N4 bond than is the case in the structure of the nonmethylated polymer.

polyamines. Spermine stabilizes the formation of Z-DNA for the methylated polymer in submicromolar concentrations. This supports the contention that methylation of alternating dC-dG sequences induces the formation of Z-DNA, perhaps even in short segments of DNA. These structural studies provide a rationale for understanding the mechanism for this stabilization. What is not answered at present is the question of how short a segment of Z-DNA can be formed given the stimulus that methylation of cytosine residues produces for Z-DNA formation. We should like to be able to answer the question whether the methylation of CpG sequences, which occurs *in vivo*, actually results in the formation of small stretches of Z-DNA. Structural studies show us that the destabilization of B-DNA and the relative stabilization of Z-DNA are associated with interactions in the immediate vicinity of the methyl groups themselves. It is thus conceivable that small sections of Z-DNA could form in the middle of B-DNA. In order to study this question it will be necessary to carry out a different kind of experiment in which small segments of a

DNA oligomer are methylated in the hope of trapping, in a single crystal lattice, a segment containing a B-Z-B interface.

9. ADENINE-THYMINE BASE PAIRS IN Z-DNA

Recently, it has been shown that Z-DNA forms in deoxyoligomers containing AT as well as CG base pairs (Wang et al., 1984). Through the use of antibodies specific for Z-DNA, segments of DNA in negatively supercoiled plasmids have been shown to form Z-DNA, and they contain AT base pairs (Nordheim and Rich, 1983). However, AT base pairs are less stable than CG base pairs. We have addressed the problem by solving the structure of Z-DNA with AT base pairs, specifically two hexanucleoside pentaphosphates with the general self-complementary sequence d(CGTACG) in which the cytosine residues have either methyl groups or bromine atoms on their 5 positions.

The structures of both the 5-methyl and 5-bromo derivatives of d(CGTACG) are similar to each other and likewise similar to the structure of $(dC\text{-}dG)_3$ as well as $(m^5dC\text{-}dG)_3$. The self-complementary molecules form a double helix with six base pairs. The left-handed Z-DNA double-helical fragments stack together to form an essentially continuous double helix running along the *c* axis of the unit cell. The crystal of the methylated hexamer diffracts to a resolution of 1.2 Å, and the brominated derivative to 1.5 Å. Figure 25 is a diagram of a van der Waals model of the brominated d(CGTACG), which shows the end-to-end organization of three of the double-helical molecules. The overall form of the molecule is very similar to that of the unmethylated as well as the methylated $(dC\text{-}dG)_3$.

A van der Waals view of the 5-bromocytosine derivative of d(CGTACG) is shown in Figure 25. The large black atoms are the bromine atoms. The thymine methyl groups are also shaded black but have hydrogen atoms attached to them. The carbonyl oxygen 2 atoms of the thymine groups are cross-hatched in order to illustrate their position near the axis of the molecule at the bottom of the deep groove. Unlike the cytosine O2 atoms, these oxygens do not have amino groups hydrogen-bonding to them because of the absence of an N2 amino group in adenine. The bromine atoms in Figure 24 occur as pairs on the surface of the molecule and they are found at distances of 4.56 and 4.77 Å apart. The

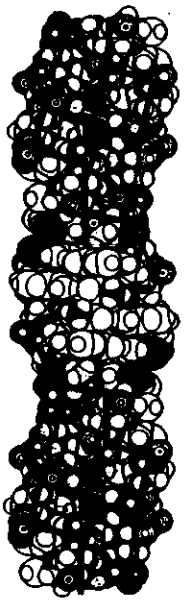


Figure 25. Stereo diagram of a van der Waals model of the methylated hexamer d(m^{5}CGTAm^{5}CG). Portions of the molecule that form the helical groove are shaded. The oxygen atoms are drawn with circles while the phosphorus atoms have spiked circles. The zigzag array of the phosphate groups bordering the helical groove is readily seen. This diagram shows three molecules stacked up as they are found in the crystal. The methyl groups are found on the outer convex portion of the molecule. The methyl group carbon atoms from the 5-methyl cytosine derivatives are cross-shatched, while those from the thymine residues are solid.

two methyl groups on the thymine residues are separated by 4.85 Å in a manner similar to that seen in the methylated derivative described above. Homologous structural features are generally found in both the methylated and brominated derivatives.

In refining the crystal structure, a large number of water molecules and solvent ions were identified. In the 5-methyl derivative of d(CGTACG), 98 solvent atoms were located. The vast majority of these peaks are water molecules, although five of the peaks have been identified as cations due to the coordination geometry they exhibit. In order to visualize the water molecules in the lattice, we present in Figure 26 six sections of the electron density map. These composites each represent

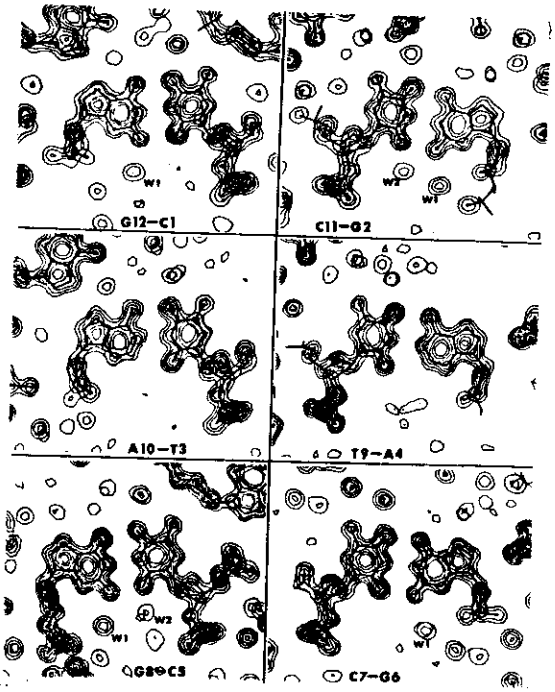


Figure 26. Sections of the electron density map are shown through the six base pairs in the double helix d(m^{5}CGTAm^{5}CG). The base pairs are oriented so that the groove is near the bottom of the diagram. Two different types of water molecules are seen in the helical groove. W_1 receives a hydrogen bond from the amino group of guanine and in turn donates a hydrogen bond to a phosphate oxygen. That water molecule lies near the plane of the bases. The second water molecule, W_2, lies halfway between the O_2 atoms of the stacked pyrimidine residues. In general it has fewer contours than W_1 due to the sectioning. The lack of ordered water molecules in the groove near the AT base pairs is readily apparent. Water molecule W_2 does not appear with such intensity in the sections at the top and the bottom of the molecule because the occupancy of these water molecules is somewhat less in the intermolecular solvent channel.

the superposition of three sections through the six base pairs in the molecule, taken in planes passing through the base pairs as well as 1 Å on either side. The axis of the molecule passes near the O2 of the pyrimidine residue and the base pairs are oriented with the helical groove of the molecule on the lower part of the diagram. Two different types of water molecules are found in the groove, W1 and W2. The water molecule W1 is in position to accept a hydrogen bond from the amino group on position 2 of guanine, and it donates a hydrogen bond to a phosphate oxygen of the same deoxyguanosine nucleotide (except for the terminal guanine G6 or G12). This bridging water molecule is in a position to stabilize the *syn* conformation of guanine. The W2 water molecule fits into the groove, where it donates two hydrogen bonds to O2 of pyrimidine residues in successive base pairs. The oxygen atom of W2 is located between the base pairs, thus it usually has less electron density in the sections that are shown in Figure 26. In the structure of both the unmethylated hexamer $(dC\text{-}dG)_3$ (Wang et al., 1979) and the methylated hexamer $(m^5dC\text{-}dG)_3$ (Fujii et al., 1982), these two water molecules are found completely occupying the helical groove of the Z-DNA. Thus there are two water molecules hydrogen-bonded to electronegative nitrogen and oxygen atoms at the bottom of the Z-DNA helical groove per base pair.

The situation is quite different in the case of the AT base pairs, as no water molecules can be seen in the groove hydrogen-bonded to the AT base pairs. The adenine residue does not have an N2 amino group and so it is not surprising that the bridging water molecule is absent. However, it was a little surprising to find that the water molecule W2 hydrogen-bonding onto the pyrimidine O2 atoms along the bottom of the groove is also missing in this region. Figure 26 also shows a variety of other water molecules found on the outer, convex portion of the molecule. The positions of some of these water molecules are also modified by the presence of the methyl group on the 5 position of cytosine (see Figure 24).

Another visualization of hydration around the Z-DNA molecule is seen in the stereodiagram of Figure 27. This shows a van der Waals model of d(m^{5}CGTAm^{5}CG) (Wang et al., 1984) together with associated solvent molecules (stippled spheres). The convex outer portion of the structure is coated with a sheath of water molecules which are binding to most of the electronegative atoms in the molecule. There are small patches of hydrophobic residues which do not have water molecules covering them. Figure 27 also shows the helical groove of Z-DNA to be filled with water

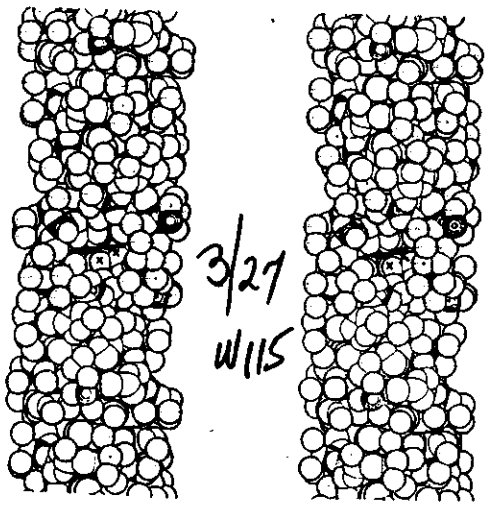


Figure 27. Stereo diagram of a van der Waals model of the hexamer d(m^{5}CGTAm^{5}CG) showing the first shell hydration. Only solvent molecules are included for which the temperature factor is lower than 45 Å^2. Water molecules are represented as spheres with stippled circles, magnesium metal ions are drawn as spheres with concentric circles, while phosphorus atoms are drawn with circles with cross striations. The helical groove is generally filled with ordered water molecules except in one region where the two AT base pairs lie. In order to visualize this section of the groove, two water molecules were removed which are bound to the phosphate groups near the periphery of the molecule.

molecules except at the position occupied by the two base pairs. In this figure we are able to inspect the bottom of the helical groove where the base pairs themselves can be seen devoid of ordered hydration structure. The bottom of the groove does not show solvent even though there are water molecules at the outer part of the groove which are hydrogen-bonded to the negatively charged phosphates. Two of these (at positions marked ×) have been removed in Figure 27 in order to make visible the

disordered solvent at the bottom of the groove. Even though the geometry of the Z-DNA molecule is not modified to a significant extent through the introduction of AT base pairs, hydration in the helical groove is significantly altered.

It is known that AT base pairs in B-DNA are somewhat less stable than CG base pairs. In Z-DNA, they are probably considerably less stable, not only because of the absence of a third hydrogen bond between the bases, but owing as well to the paucity of ordered hydration structure in the helical groove. Since the Z-DNA sequences contain information, it is likely that these differences in stability are utilized in biological systems.

10. FUTURE STUDIES

There has been a great advance from a decade or so ago in available structural information regarding the nucleic acids. Progress can be ascribed largely to the fact that synthetic oligonucleotides have yielded crystalline fragments of DNA and RNA. These have in turn permitted high-resolution diffraction studies which have taught us a great deal about the three-dimensional structure and conformation of the nucleic acids. Long helical fragments are now available comprising segments from four to over a dozen base pairs. Many of the crystal structures that have been solved are at atomic resolution. From this association of organic chemistry and X-ray crystallography has come an enormous broadening in our comprehension of the fine details of nucleic acid conformation as well as the manner by which they interact with water molecules and cations. These advances have not been confined to the traditional structures of DNA, the A and the B form, but now encompass the discovery of a novel left-handed Z form. If one were to extrapolate these studies into the future, it seems almost certain that increasingly more structural detail at high resolution will be obtained through the use of larger oligomers.

Are there other novel conformations yet to be discovered? Here again we think the answer is likely to be yes. The structural features characterizing Z-DNA are dependent on a particular preferred sequence of alternating purines and pyrimidines. Is it possible that there are other motifs involving repetitions of three or more nucleotides that produce an alternative conformational pattern? Is it possible that, through their interactions with other molecules, including proteins, the nucleic acids will be

shown capable of maintaining conformations that exist in solution even in the absence of proteins? It is always difficult to predict the future, but we now have a methodology based on the use of synthetic oligonucleotides of known sequence that are capable of forming single crystals whose structures can be solved. This powerful tool will undoubtedly lead to a continued flowering of discovery well beyond the sketchy outlines available up to now through fiber diffraction studies. Among these new results are likely to be a proliferation of new conformational states. The discovery of the Z form of DNA, for all it has told us about the structural chemistry of nucleic acids, may ultimately be seen as an even more important event for awakening us to the unrealized possibilities inherent in the double helix.

Acknowledgments. This research was supported by grants from the National Institutes of Health, the American Cancer Society, the National Science Foundation, and the National Aeronautics and Space Administration.

REFERENCES

Arnott, S. (1970). *Progress in Biophysics and Molecular Biology*. J. A. V. Butler and D. Noble, Eds. Pergamon, New York.

Arnott, S., Chandrasekarran, D. L., Birdsall, D. L., Leslie, A. G. W., and Ratliff, R. L. (1980). *Nature* **283**, 743–745.

Behe, M., and Felsenfeld, G. (1981). *Proc. Natl. Acad. Sci. USA* **78**, 1619–1623.

Crawford, J. L., Kolpak, F. J., Wang, A. H.-J., and Quigley, G. J., van Boom, J. H., van der Marel, G., and Rich, A. (1980). *Proc. Natl. Acad. Sci. USA* **77**, 4016–4020.

Day, R. O., Seeman, N. C., Rosenberg, J. M., and Rich, A. (1973). *Proc. Natl. Acad. Sci. USA* **70**, 849–853.

Drew, H., Takano, T., Tanaka, S., Itakura, K., and Dickerson, R. E. (1980). *Nature* **286**, 567–573.

Ehrlich, M., and Wang, R. Y.-H. (1981). *Science* **212**, 1350–1357.

Franklin, R. E., and Gosling, R. G. (1951). *Nature* **171**, 740–741.

Fujii, S., Wang, A. H.-J., Quigley, G. J., Westerink, H., van Boom, J. H., van der Marel, G., and Rich, A. (1984). Submitted.

Fujii, S., Wang, A. H.-J., van der Marel, G., van Boom, J. H., and Rich, A. (1982). *Nucl. Acids Res.* **10**, 7879–7892.

Gupta, G., Bansal, M., and Sasiskeharan, V. (1980). *Proc. Natl. Acad. Sci. USA* **77**, 6486–6490.

Haschemeyer, A. E. V., and Rich, A. (1967). *J. Mol. Biol.* **27**, 369–384.

Hopkins, R. C. (1981). *Science* **211**, 289–291.

Lafer, E. M., Moller, A., Nordheim, A., Stollar, B. D., and Rich, A. (1981). *Proc. Natl. Acad. Sci. USA* **78**, 3546–3550.

Nordheim, A., and Rich, A. (1983). *Nature* **303**, 674–679.

Pohl, F. M., and Jovin, T. M. (1972). *J. Mol. Biol.* **67**, 375–396.

Pohl, F. M., Ranade, A., and Stockburger, M. (1973). *Biochim. Biophys. Acta* **335**, 85–92.

Razin, A., and Riggs, A. D. (1980). *Science* **210**, 604–610.

Rodley, G. A., Scobie, R. S., Bates, R. H., and Lewitt, R. M. (1976). *Proc. Natl. Acad. Sci. USA* **73**, 2959–2963.

Rosenberg, J. M., Seeman, N. C., Day, R. O., and Rich, A. (1976). *J. Mol. Biol.* **104**, 145–167.

Rosenberg, J. M., Seeman, N. C., Kim, J. J. P., Suddath, F. L., Nicholas, H. B., and Rich, A. (1973). *Nature* **243**, 150–154.

Sasiskeharan, V. (1982). *Cold Spring Harbor Symp. Quant. Biol.* **47**, 45–52.

Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Kim, J. J. P., and Rich, A. (1976). *J. Mol. Biol.* **104**, 109–144.

Son, T.-D., Guschlbauer, W., and Gueron, M. (1972). *J. Am. Chem. Soc.* **94**, 7903–7911.

Sunderalingam, M. (1969). *Biopolymer* **7**, 821–833.

Thamann, T. J., Lord, R. C., Wang, A. H.-J., and Rich, A. (1981). *Nucl. Acids Res.* **9**, 5443–5457.

Voet, D., and Rich, A. (1970). *Prog. Nucl. Acid Res. Mol. Biol.* **10**, 183–265 (1970).

Wang, A. H.-J., Fujii, S., van Boom, J. H., and Rich, A. (1982). *Cold Spring Harbor Symp. Quant. Biol.* **47**, 33–44.

Wang, A. H.-J., Hakoshima, T., van der Marel, G., van Boom, J. H., and Rich, A. (1984). *Cell* (in press).

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

Wang, A. H.-J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H., and Rich, A. (1981). *Science* **211**, 171–176.

Watson, J. D., and Crick, F. H. (1953). *Nature* **171**, 737.