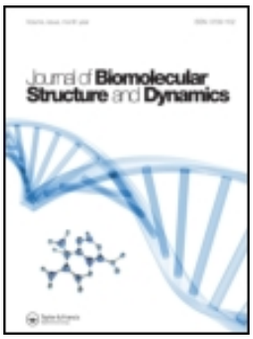

Journal of Biomolecular Structure and Dynamics

ISSN: 0739-1102 (Print) 1538-0254 (Online) Journal homepage: https://www.tandfonline.com/loi/tbsd20

A Kinked Model for the Solution Structure of DNA Tridecamers with Inserted Adenosines: Energy Minimization and Molecular Dynamics

M. Hirshberg, R. Sharon & J. L. Sussman

To cite this article: M. Hirshberg, R. Sharon & J. L. Sussman (1988) A Kinked Model for the Solution Structure of DNA Tridecamers with Inserted Adenosines: Energy Minimization and Molecular Dynamics, Journal of Biomolecular Structure and Dynamics, 5:5, 965-979, DOI: 10.1080/07391102.1988.10506443

To link to this article: https://doi.org/10.1080/07391102.1988.10506443

Journal of Biomolecular Structure & Dynamics, ISSN 0739-1102
Volume 5, Issue Number 5 (1988), ©Adenine Press (1988).

A Kinked Model for the Solution Structure of DNA Tridecamers with Inserted Adenosines: Energy Minimization and Molecular Dynamics

M. Hirshberg[1], R. Sharon[1] and J. L. Sussman[2]
[1]Department of Chemical Physics
[2]Department of Structural Chemistry
The Weizmann Institute of Science
Rehovot 76100, ISRAEL

Abstract

Structural modelling techniques using energy minimization and molecular dynamics have been employed to generate kinked models for the solution structure of two DNA tridecamer sequences containing inserted adenosines: d(CGCAGAATTCGCG)$_2$ and d(CGCAGAGCTCGCG)$_2$. These models are consistent with NMR studies of these sequences in solution. The overall shapes of the two models are similar, consisting of three B-DNA sections: two outer segments on the same side of the central portion, with the additional adenosines acting as wedges to kink the structure. An alternative scheme for the hydrogen bond pairing at the kink site is suggested as a way for the additional adenosines to be stabilized in the duplex.

Introduction

The role of DNA containing inserted bases has been of long standing interest in studies of frameshift mutagenesis (1,2). Detailed structural information on the way a DNA duplex accommodates such extra bases should give us some insight into the mechanism of genetic mutations. The conformation and the thermodynamic stability of such aberrant DNA have been examined by various *experimental* techniques such as NMR (3,4,5), calorimetry (6) and temperature jump (7) as well as *theoretically* (5,8).

The presence of additional bases in an otherwise self-complementary sequence could result in two main structural motifs - a single stranded hairpin loop or a duplex. Which of these two, or both are found in solution appears to be very sequence dependent (9,10). It has been shown, however, that inserting an additional adenosine between residues 3 and 4 of the dodecamer sequences d(CGCGAATTCGCG)$_2$ or d(CGCGAGCTCGCG)$_2$ preserves the duplex in solution. This may result in partial misalignment of the two strands, or the inserted bases could remain unpaired, either stacking into the duplex or looping out. Three different possible duplex models for the almost self-complementary tridecamer d(CGCAGAATTCGCG)$_2$, a sequence with one additional adenosine at position 4 of each strand, are shown schematically in Figure 1. The first model is a misaligned structure 13 base pairs long with five mismatches. The two other models consist of normal Watson-Crick pairing with the additional adenosine either stacking

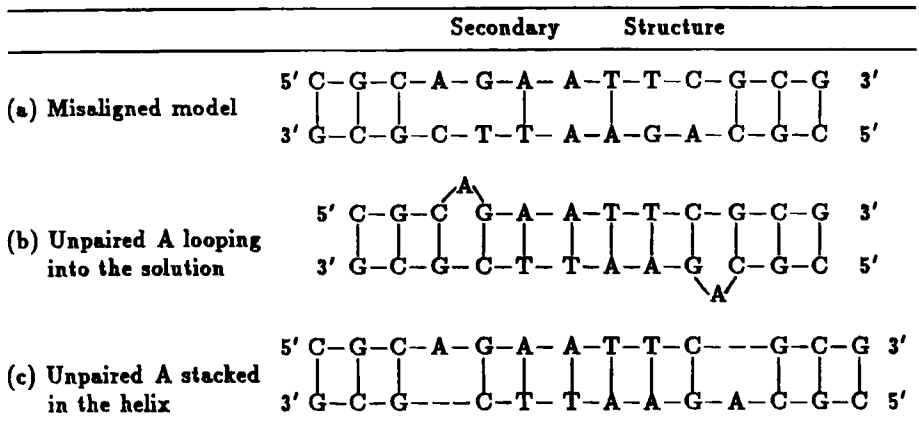

Figure 1: Schematic representation of three possible secondary structures for the DNA tridecamer 5′d(CGCAGAATTCGCG)3′ containing additional adenosines.

into the double helix or looping out towards the solution. The misaligned structure is unlikely, as it has been shown experimentally (11) and theoretically (12), that DNA containing one or two mismatches is less stable than the corresponding self-complementary duplex. NMR and calorimetric studies in solution on this tridecamer (6) showed that the additional adenosines, in an otherwise self-complementary 12 base pair DNA, appeared to be stacked into the duplex and that the duplex itself adopted a regular B-DNA structure. Similar results were obtained via 2D NMR studies on the sequences $d(CGCAGAGCTCGCG)_2$ (4), and d(CGCGAAATTTACGCG) (5) which also indicated, in each case, that the additional adenosines were stacked into the double helix rather than looping out. In contrast, studies on DNA sequences containing additional cytidines (13) or thymidines (3) showed that the pyrimidines tend to loop out toward the solution, away from the double helix. X-ray crystallographic studies on $d(CGCAGAATTCGCG)_2$ (L. Joshua-Tor and J. L. Sussman unpublished results) indicate that the conformation of the additional adenosines in the crystalline state is different from the conformation observed in solution.

Molecular mechanics studies on a similar sequence containing *one* additional "stacked in" adenosine at position 4 of one strand [d(CGCAGAATTCGCG) d(CGCGAATTCGCG)] were carried out by Keepers *et al.* (8). They refined one model with the inserted adenosine stacking inside the double helix and performed an extensive analysis of the thermodynamic properties of this model. They concluded that the perturbations of the overall double helix were small and confined to the vicinity of the inserted adenosine. They also suggested that the actual time average structure may differ from their refined structure by including forms with the additional base looping out into the solution.

We present here models for two sequences of DNA $d(CGCAGAATTCGCG)_2$ and $d(CGCAGAGCTCGCG)_2$, containing additional adenosines, which were constructed using energy minimization and molecular dynamics techniques. They are consistent with the experimental results, including the interbase distances at the

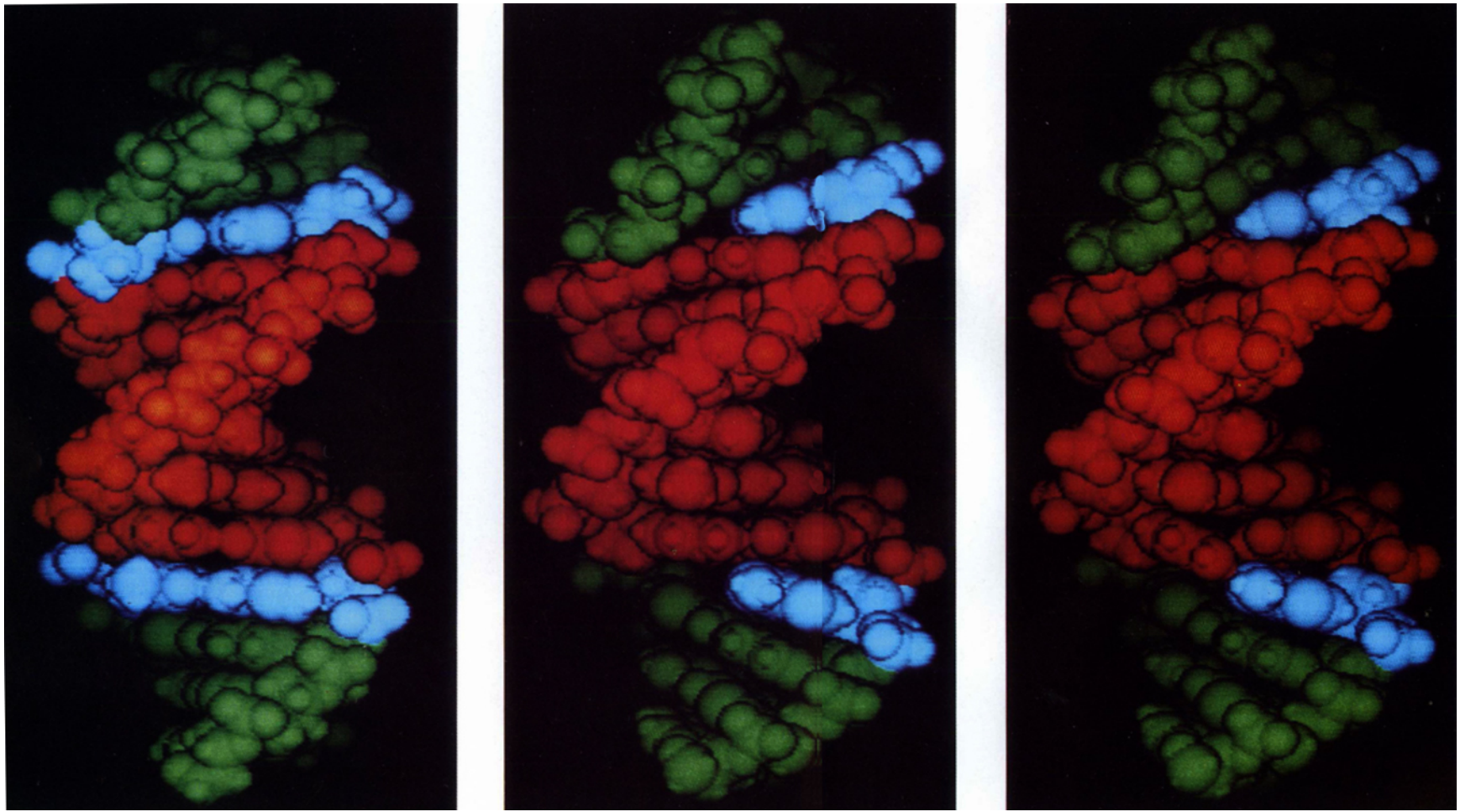

Figure 5: Two kinked models for the DNA tridecamer 5′d(CGCAGAATTCGCG)3′ and a model for the tetradecamer 5′d(CGCAGAATTCTGCG)3′ shown as CPK space filling models with the computer graphics program IMDAD (M. Levitt, personal communication). Left) tetradecamer; center) 18°-18°; right) 18°-18°(2HB).

insertion sites (4). As the two sequences are very similar, differing only in the central two base pairs, i.e. AATT versus AGCT, and as the resulting structures are similar, we describe the analysis of only d(CGCAGAATTCGCG)$_2$ in detail, followed by a comparison between them.

It is important to emphasize that due to its high degree of flexibility the actual conformation of a DNA segment depends on its particular environment (14). Consequently, the models we present here, which are based on, and consistent with, solution studies may adopt different conformations in the crystalline state.

Simulations on a system containing a DNA segment and counter ions surrounded by water molecules give a more realistic picture than in vacuum simulations. Nevertheless, for the "stacked in model", the additional bases are inside the double helix and are therefore less exposed to the environment. Thus, for this type of models, it is reasonable to start the studies with in vacuum simulations, which require less computer resources, and to continue with solution simulations on the models generated from these studies.

Methods and Model Construction

The model for the tridecamer containing two additional adenosines was constructed using unconstrained and constrained energy minimization and molecular dynamics. Using energy minimization techniques, one attempts to derive the equilibrium structure of the model molecule by varying the atomic coordinates until the potential function is minimized. Energy minimization methods, which are applicable to macromolecules lead only to local minima of the potential energy. When performing molecular dynamics, the system acquires kinetic energy which allows it to explore larger regions of the conformational space, and thus it may escape from local minima surrounded by not too high energy barriers ($\sim$KT). The method we use, the so called "annealing" methods, which we first used in oligopeptide simulations (S. Lifson and R. Sharon, 1979, unpublished results), consists of alternating short molecular dynamics runs, followed by energy minimization.

All the simulations were performed using ENCAD, Energy Computations Analysis and Dynamics package, which is an upgraded version of EREF, extensively used in macromolecular structure refinement (M. Levitt, 1986 personal communication). The potential function used by ENCAD contains terms for bond stretching, bond angle bending, hindered bond twisting, van der Waals interactions and hydrogen bond interactions. The electrostatic interactions were neglected and no solvent molecules were included. The potential function is described in detail by Levitt (15-16).

Our initial model for the tridecamer d(CGCAGAATTCGCG)$_2$ was based on NMR and calorimetric studies (4-6) as well as preliminary X-ray studies (17-18). These studies indicated that the tridecamer was in the B form. Furthermore, the NMR studies also showed the additional adenosine to be stacked into the double helix. Therefore, we built a B-DNA double helix (19). for the tetradecamer sequence d(CGCAGAATTCTGCG)$_2$, a sequence with two extra thymidines opposite to the

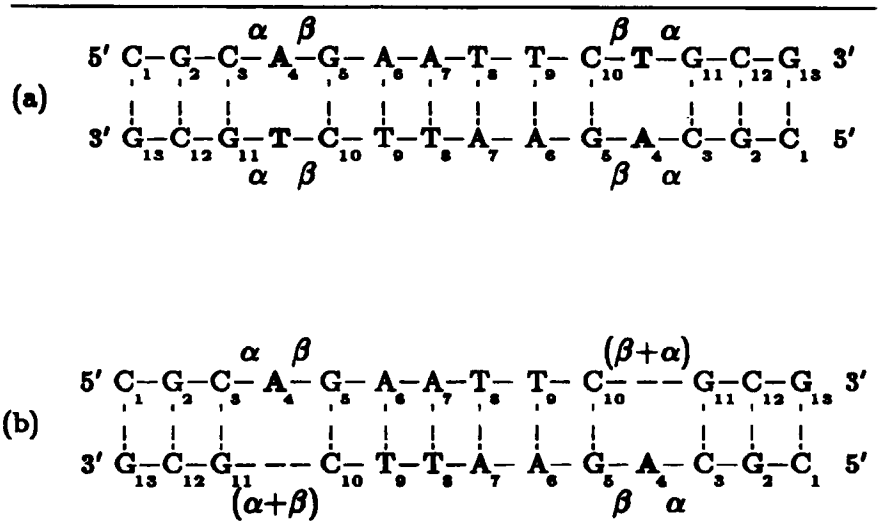

Figure 2: Schematic representation of the initial model for the DNA tridecamer 5′d(CGCAGAATTCGCG)3′. α, β are the helix rotation angles per residue, before minimization, in the vicinity of the additional adenosines. Eight models were generated: (α,β) = [(12°,24°), (15°,21°), (18°,18°), (21°,15°), (24°,12°), (24°,24°), (30°,30°), (36°,36°)]. a) The starting tetradecamer sequence generated with two thymines (T) opposite to the additional adenosines (A); b) the tridecamer before minimization. (For the sequence 5′d(CGCAGAGCTCGCG)3′ A_7T_8 are replaced with G_7C_8).

inserted adenosines (Figure 2a). These two thymidines were then deleted from the model, the gap was spanned by a stretched O3′-P covalent link (Figure 2b), and the structure was subjected to energy minimization using the computer program ENCAD until an equilibrium was reached.

For the B-DNA structure based on fiber diffraction studies (19) the helical parameters are 36° rotation and 3.38Å axial rise per base pair. Thus, in our initial model for the tridecamer, designated the 36°-36° model, the helical rotation angle between C_{10} and G_{11} was 72° (α + β in Figure 2b), significantly affecting the stacking energy between these bases. In order to check the effect of the initial model on the final optimized structure, we constructed other models with different helical rotation angles around the additional bases, such that the sum α + β after deleting the two extra thymidines was equal to 60, 48 or, in five distinct models, 36° - the standard value for B-DNA. For each model, a tetradecamer double helix was generated using B-DNA helical parameters (19), except for the helical rotation angles α between base pairs 3 and 4 and β between base pairs 4 and 5. The thymidines were then removed and the structure was subjected to energy minimization. Each of the refined α-β models was further subjected to two separate annealing simulations. The molecular dynamics parts were done at 300K (room temperature) and 330K, respectively, with time steps of 0.002 psec.

A second set of structures designated α-β(2HB) with the additional adenosines hydrogen bonded to the diagonally flanking cytidines, C_{10}, were also generated. Each of the α-β (2HB) was constructed from the corresponding refined α-β model using constrained

energy minimization, with constraints on the hydrogen bond distances between both pairs A_4 and C_{10}, followed by minimization with no constraints. This resulted in a symmetric structure with two hydrogen bonded adenosines.

Structural and energy analysis was performed using ENCAD. The local and the global helical parameters were calculated using a program of best molecular fit, BMF (J.L. Sussman, in preparation) and MODBROLL (R.E. Dickerson, personal communication). The structures obtained were viewed on a Liacom Color Raster Graphics System connected to a VAX 11/780 computer using the molecular graphics program IMDAD (M. Levitt, personal communication).

Results and Discussion

The 18°-18° Model

All models gave similar but not identical structures. Table I summarizes the total energy of the various structures after initial minimization, after constrained minimization and after "annealing", together with the energies after minimization of the two self-complementary B-DNA sequences: the dodecamer $d(CGCGAATTCGCG)_2$ and the tetradecamer $d(CGCAGAATTCTGCG)_2$ used as controls. The most stable structure after "annealing" was obtained from the 18°-18° model (columns 4 and 5 in Table I). Also, the 18°-18°(2HB) is the most stable structure among the α-β (2HB) models (column 3 in Table I). Thus, we have chosen the structures which were obtained from models with initial helical rotation (α,β)=(18°,18°), as our representative models. The RMS deviations between 18°-18° model and all the other α-β models are given in Table II. The three principal stages in the model building of the 18°-18° model are shown in Figure 3.

Introduction of additional adenosines to an otherwise self-complementary B-DNA, results in local changes mainly in the relative orientations of the planes of the flanking base pairs next to the inserted adenosines. The two base pairs C_3-G_{11} and G_5-C_{10} open up away from the additional adenosine, and the tilt angle formed between the best planes through these base pairs, containing their long axes (C6/C8-C8/C6), is 30°. Thus, the global structure has two kinks, one at each insertion site, and can be divided naturally into three parts: a central segment consisting of six base pairs and two outer segments, each three base pairs long. A comparison between atomic coordinates of the three segments of the 18°-18° model, taken separately, with the corresponding segments of the tetradecamer, give RMS deviations of 0.6Å and 0.1Å for the central and outer segments respectively, indicating that each region has the same overall shape as in the tetradecamer. The kink angle between the helical axes of the central part and each of the outer parts is 34.5°. A schematic representation of the three segments is given in Figure 4a. The kink may also be visualized by comparing the CPK model of the self-complementary tetradecamer (after minimization) and the 18°-18° model (Figure 5a and 5b).

Usually, the helical rotation is defined as the projected angle between C1′-C1′ (or C6/C8-C8/C6) vectors in successive base pairs (20). This method is obviously not

Table I
Total Energy (Kcal/mole) for the different starting models.

Model	EM[1]	EM[2]	2HB[3]	MD 300K[4]	MD 330K[5]
12°-24°	-217.2	(-8.4)	-221.2	-218.0	-220.3
15°-21°	-198.4	(-7.6)	-208.4	-214.6	-224.3
18°-18°	-215.0	(-8.3)	-226.0	-228.2	-236.0
21°-15°	-204.6	(-8.9)	-202.8	-215.0	-215.4
24°-12°	-217.4	(-8.4)	-222.0	-224.9	-220.7
24°-24°	-212.3	(-8.2)	-223.6	-222.2	-226.5
30°-30°	-197.0	(-7.6)	-213.9	-223.8	-215.2
36°-36°	-199.7	(-7.7)	-201.6	-199.3	-209.1
dodecamer[6]	-205.6	(-8.6)	-	-	-
tetradecamer[7]	-250.2	(-8.9)	-	-	-

1. **Energy after minimisation.**
2. **Energy after minimisation per nucleotide.**
3. **The α,β(2HB) model, obtained from the energy minimized model by constrained minimisation.**
4. **Molecular dynamics at 300K, obtained from the minimized model.**
5. **Molecular dynamics at 330K, obtained form the minimized model.**
6. **Self-complementary dodecamer d(CGCGAATTCGCG).**
7. **Self-complementary tetradecamer d(CGCAGAATTCTGCG).**

Table II
RMS Deviations between the 18°-18° model and the other models.

Model	12°-24°	15°-21°	21°-15°	24°-12°	24°-24°	30°-30°	36°-36°
RMS A	0.41	0.66	0.76	0.36	0.75	1.64	1.89

applicable for the tridecamer sequence as the C1′-C1′ vectors at the inserted sites can not be defined. We propose here an alternative method for the calculation of the helical rotation angles for a self-complementary double helix, as well as for a double helix containing additional bases.

For two successive base pairs the best molecular fit (BMF) is calculated for the ribose atoms (including O3′ and O5′ and N1/N9 of the two base pairs). The rotation matrix by which the coordinate set of one base pair is superimposed on the coordinate set of a second base pair may be represented in this case by a single rotation angle around a helical axis. This angle is the helical rotation angle between the two successive base pairs. A comparison between helical rotation angles calculated using the C1′-C1′ vectors and the BMF method is given in Table III for the

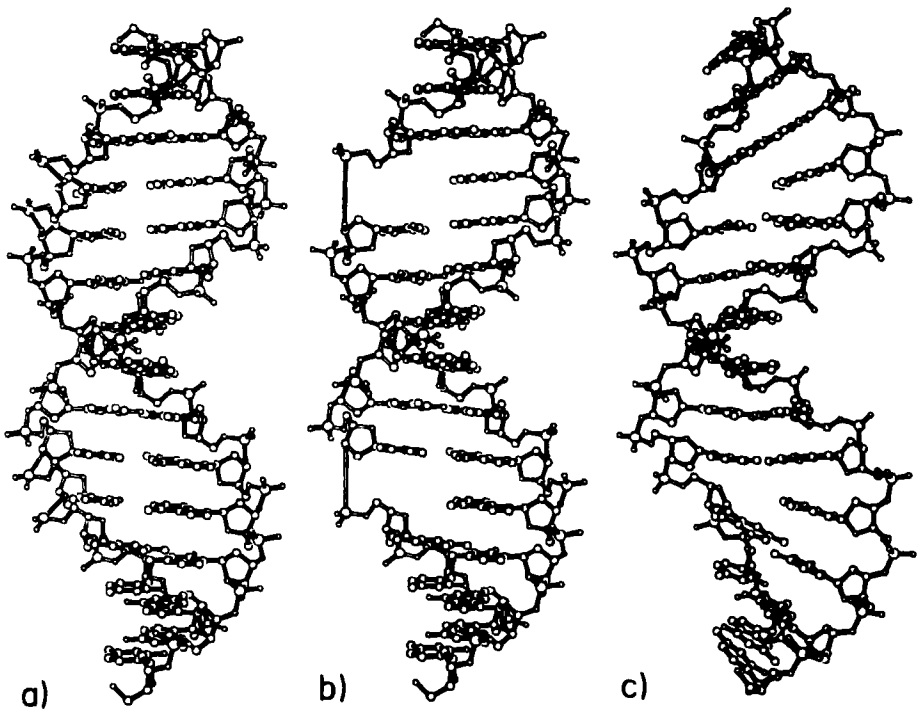

Figure 3: PLUTO drawings (S. Motherwell, personal communication), for the three principal stages in the 18°-18° model construction, of the tridecamer 5′d(CGCAGAATTCGCG)3′. a) The initial tetradecamer model, a sequence with two T's opposite to the extra A's; b) the structure after the two extra T's were removed before energy minimization; c) the relaxed structure after energy minimization.

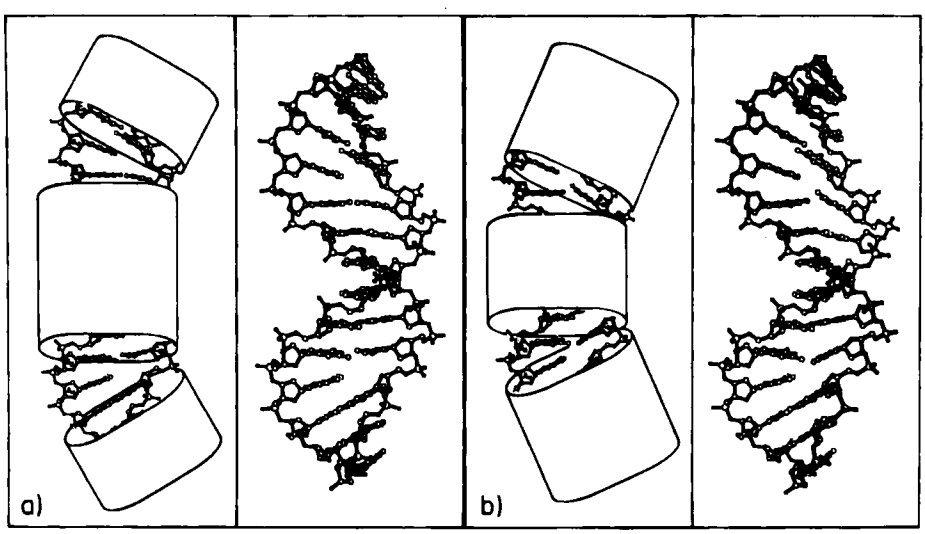

Figure 4: Two kinked models for the DNA tridecamer 5′d(CGCAGAATTCGCG)3′ shown as PLUTO drawings and schematically as three segments which are symbolized by cylinders: a) 18°-18°; b) 18°-18°(2HB).

Table III
Helical rotation angles (°) for the tetradecamer.

	A	B	C	D	E
				5′	3′
C-G				C	G
⟨ ⟩	31.5	32.9	31.8	⟨ 34.7	28.5 ⟩
G-C				G	C
⟨ ⟩	28.4	28.4	29.7	⟨ 26.8	38.8 ⟩
C-G				C	G
⟨ ⟩	31.9	31.3	32.0	⟨ 38.2	33.1 ⟩
A-T				A	T
⟨ ⟩	30.7	30.1	30.6	⟨ 42.4	21.6 ⟩
G-C				G	C
⟨ ⟩	31.1	30.9	30.6	⟨ 23.5	35.5 ⟩
A-T				A	T
⟨ ⟩	36.2	36.4	35.6	⟨ 34.9	35.2 ⟩
A-T				A	T
⟨ ⟩	31.9	31.5	34.2	⟨ 41.5	41.5 ⟩
T-A				T	A
⟨ ⟩	35.8	36.1	35.9	⟨ 35.3	34.9 ⟩
T-A				T	A
⟨ ⟩	30.6	31.1	30.8	⟨ 35.5	23.5 ⟩
C-G				C	G
⟨ ⟩	30.2	30.0	30.5	⟨ 21.6	42.3 ⟩
T-A				T	A
⟨ ⟩	31.6	31.3	32.1	⟨ 33.1	38.2 ⟩
G-C				G	C
⟨ ⟩	28.6	28.3	29.6	⟨ 38.8	26.8 ⟩
C-G				C	G
⟨ ⟩	32.2	32.9	31.8	⟨ 28.5	34.7 ⟩
G-C				G	C
				3′	5′

A. Helical rotation per base pair, calculated relative to a global helix axis using the C1 -C1 vector.
B. Helical rotation per base pair, calculated relative to local helical axes using the C1 -C1 vector.
C. Helical rotation per base-pair, calculated relative to local helical axes using the BMF method.
D. Helical rotation per base, calculated along one strand using the BMF method.
E. Helical rotation per base, calculated along the second strand using the BMF method.

Table IV
Helical parameters for the 18°-18° model.

A			B	C
	Rotation (°)	Rise (A)	Rotation (°)	Rotation (°)
			5′	3′
C-G			C	G
⟨ ⟩	31.3	3.2	⟨ 33.7	28.3 ⟩
G-C			G	C
⟨ ⟩	30.4	3.2	⟨ 26.7	39.4 ⟩
C-G			C	G
⟨ ⟩			⟨ 32.4	⟩
⟨ ⟩	48.7	5.1	A	37.3 ⟩
⟨ ⟩			⟨ 27.9	⟩
G-C			G	C
⟨ ⟩	30.4	3.8	⟨ 38.5	36.8 ⟩
A-T			A	T
⟨ ⟩	33.6	3.4	⟨ 28.8	33.0 ⟩
A-T			A	T
⟨ ⟩	32.2	3.0	⟨ 38.9	38.9 ⟩
T-A			T	A
⟨ ⟩	33.6	3.4	⟨ 33.0	28.8 ⟩
T-A			T	A
⟨ ⟩	30.6	3.8	⟨ 36.8	38.5 ⟩
C-G			C	G
⟨ ⟩			⟨	27.9 ⟩
⟨ ⟩	48.7	5.1	⟨ 37.3	A
⟨ ⟩			⟨	32.4 ⟩
G-C			G	C
⟨ ⟩	30.4	3.2	⟨ 39.4	26.7 ⟩
C-G			C	G
⟨ ⟩	31.1	3.2	⟨ 28.3	33.8 ⟩
G-C			G	C
			3′	5′

A. Helical rotation and rise per base pair, calculated using the BMF method.
B. Helical rotation per base, calculated along one strand using the BMF method.
C. Helical rotation per base, calculated along the second strand, using the BMF method.

optimized tetradecamer sequence. Furthermore, using this method, we could also calculate the helical rotation angles for single strands, and the results are given in columns D and E in Table III. The helical rotation angle per base pair is a property of the double helix itself and both methods give similar results. On the other hand,

Table V
Average Torsion Angles (°).

	α	β	γ	δ	ε	ζ	χ
tetradecamer[1]	-47	214	36	156	-205	-95	143
tetradecamer[2]	-67 (3)	170 (3)	63 (3)	101 (20)	-172 (7)	-81 (7)	103 (14)
18°–18° [2]	-68 (3)	171 (4)	63 (3)	101 (20)	-172 (8)	-80 (7)	103 (14)
18°–18°(2HB)[2]	-68 (3)	170 (3)	64 (3)	97 (22)	-172 (9)	-79 (8)	100 (16)

Average backbone torsion angles and their standard deviation in parenthesis. The torsion angles are defined by :
α (O3'-P-O5'-C5'), β (P-O5'-C5'-C4')
γ (O5'-C5'-C4'-C3'),δ (C5'-C4'-C3'-O3')
ε (C4'-C3'-O3'-P), ζ (C3'-O3'-P-O5')
χ (C2'-C1'-N1-C2) for pyrimidins.
χ (C2'-C1'-N9-C4) for purines.

1) Initial model (Arnott, 1972) before minimization.
2) Structure after energy minimization.

the rotation angle per base is very sensitive to the orientation of the base itself via the vector C1′-N1/N9, and thus offers an insight into the properties of the individual strands.

Table IV gives the helical rotation angles per base and per base pairs for the 18°-18° model as calculated using the BMF method. The helical rotation between C_3-G_{11} and G_5-C_{10} in the self-complementary dodecamer is 32.5°, while 48.7° is observed in the 18°-18° model (column A in Table IV). Consequently, insertion of an additional adenosine results in unwinding of the double helix by 16°. A similar magnitude of unwinding has been observed in drug intercalation (21).

Each of the three segments has an overall shape of "B like" DNA. The helical axis of each segment passes through the middle of the base pairs, and the base pairs are perpendicular to this axis (Figure 5). The helical rotation angles and the axial rises are typical to B-DNA (Table IV). The average values for the main chain torsion angles and the sugar-base glycosidic bond angles for the 18°-18° model and for the optimized tetradecamer are given in Table V. The torsion angles are similar to those obtained by M. Levitt (16) and by P. Kollman (22) and differ by ~20° from the B-DNA values obtained from fiber diffraction studies (23). The torsion angles are dependent on the nature of the potential, nevertheless, the overall shape of the DNA is not very sensitive to the backbone torsion angles.

The 18°-18° (2HB) Model

In the 18°-18° model, the additional adenosine is stacked into the double helix and acts like a spacer between the outer and the central segments (Figure 6a). The structures obtained after "annealing" have the same overall shape as the 18°-18° model, nevertheless, the geometry at the kink site is different, offering an alternative way by which the additional adenosine might be stabilized inside the double helix. Specifically, a hydrogen bond is formed between the imino nitrogen N1 of the additional adenosine A_4 and the amino proton of N4 of the diagonally flanking cytidine C_{10}, coupled with the breaking of the hydrogen bond between this amino proton and the carbonyl O6

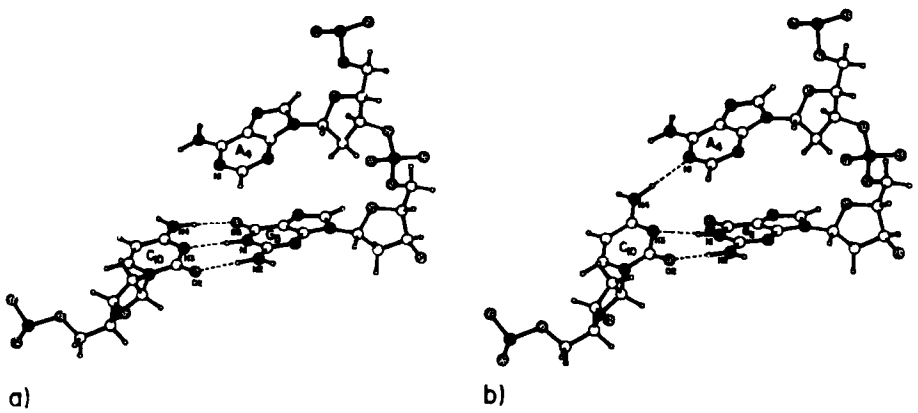

Figure 6: PLUTO drawings of the DNA tridecamer 5'd(CGCAGAATTCGCG)3' in the vicinity of the inserted adenosines (A_4 and G_5C_{10}), perpendicular to the additional adenosine. a) The 18°-18° model - the additional adenosine is unpaired; b) the 18°-18°(2HB) model - the additional adenosine is hydrogen-bonded to the diagonally flanking cytidine (C_{10}). The hydrogen bonds are dashed and are less then 2.2Å.

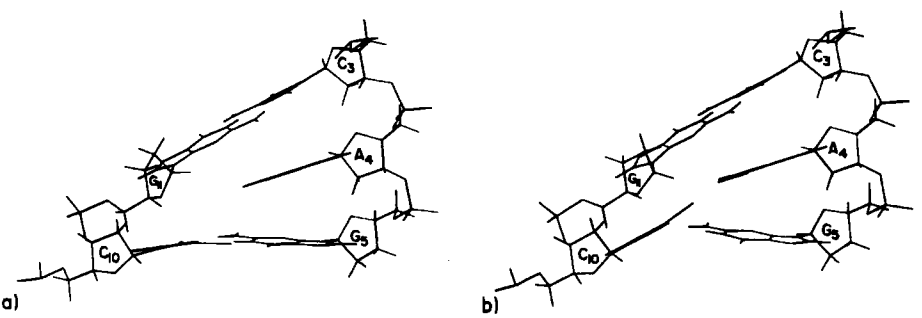

Figure 7: PLUTO drawings of the DNA tridecamer 5'd(CGCAGAATTCGCG)3' in the vicinity of the inserted adenosines (C_3G_{11}, A_4 and G_5C_{10}), perpendicular to the helix axis. a) the 18°-18° model - the additional adenosine is unpaired; b) the 18°-18° (2HB) model - the additional adenosine is hydrogen-bonded to the diagonally flanking cytidine (C_{10}).

of G_5 (Figure 6b). A new model containing two such hydrogen bonds, between the A_4 in each strand and the corresponding C_{10} on the opposite strand, was constructed from the 18°-18° model by constrained minimization as described in the model construction section. The new structure designated the 18°-18°(2HB) model is more stable by 11Kcal/mole than the 18°-18° model and the RMS deviation between this structure and the original 18°-18° model is 0.5Å.

Comparison between the 18°-18°(2HB) and the 18°-18° models shows that the hydrogen bond between A_4 and C_{10} is formed mainly by reorientation of C_{10} towards A_4 and away from G_5, which is a result of conformational changes in the δ torsion angle of G_{11}. The reorientation of C_{10} is emphasized in Figure 7, where the additional adenosine and the flanking base pairs are shown for the two models.

The differences in the detailed geometry at the insertion site do not result in major changes in the overall shape of the DNA. The 18°-18°(2HB) model has the same two

kinks as the 18°-18° model but the C_{10} now belongs to the outer segment rather than to the central one. A schematic representation of three segments of this model is given in Figure 4b, while Figure 5c gives the CPK model of the 18°-18°(2HB) model.

It is possible that the formation of such additional hydrogen bonds between a base on one strand and a flanking base on the opposite strand may be important for the sequence determining variations in DNA structure as well as possibly playing a role in sequence dependent bendability of DNA (24-25).

Comparison Between the Two Tridecamer Sequences

The two sequences $d(CGCAGAATTCGCG)_2$ and $d(CGCAGAGCTCGCG)_2$ differ only in the central two base pair, G_7C_8 instead of A_7T_8, resulting in very similar energy minimized structures with an RMS deviation of only 0.4Å (the central base pairs were not included in the RMS calculations).

The model for $d(CGCAGAGCTCGCG)_2$ was constructed using the same procedure as described previously in the model construction section. As in the case of the $d(CGCAGAATTCGCG)_2$ sequence, the most stable structure was obtained from a starting model with $(\alpha,\beta)=(18°,18°)$ - (Figure 2). This model has the same boomerang shape, with two kinks at the inserted sites. Also, the model with hydrogen bonds between A_4 and the corresponding C_{10}, was found to be more stable compared to the corresponding 18°-18° model.

The 18°-18° model is consistent with most of the proton-proton distances measured by 2D NMR studies (4), in particular the distance between adenosine H2 proton of A_4 and the imino proton of the adjacent G_5-C_{10} base pair. This distance established the fact that the additional adenosines are stacked inside the double helix.

Analysis of the Potential Function Used

In order to asses the predicting power of the potential, several tests were performed. The energy refined structure of the dodecamer $d(CGCGAATTCGCG)_2$ was compared with the B-DNA fiber structure (19). The comparison revealed similarity in the overall shape and in the B-DNA helical parameters, while discrepancies were observed in the main chain torsion angles. These angles differ by about 20° from the B-DNA values and the sugar pucker had no preferred conformation, ranging from C2′-endo to C3′-endo. A similar picture emerged in studies of B-DNA by Kollman (22) using the AMBER force field package (26), with the exception that the sugar puckers were mainly C2′-endo with only few C3′-endo.

The dependence of the refined structure on the pseudorotation energy barrier was studied using a slightly different potential. In this potential the bond angle bending energy terms $K(\theta-\theta_0)^2$ (Levitt's ENCAD) of four of the five internal sugar bending angles C4′C3′C2′, C3′C2′C1′, O1′C4′C3′ and C2′C1′O1′ were modified by changing θ_0 from the tetrahedral value 109.5° to the corresponding mean values found in X-ray studies of nucleic acids (27). This resulted in a potential barrier of 1.27 Kcal/mole

between C2′-endo and C3′-endo. The sugar puckers of the dodecamer minimized using this potential were mostly of the C2′-endo type, however the overall structure was similar to the one produced with the unmodified Levitt potential.

As also discussed by Kollman's group (28), the calculated vibrational normal modes may pinpoint the inaccurate part of a given potential. As with the AMBER potential, the calculated normal modes fit quite well the observed spectra for frequencies below $700cm^{-1}$. Those are the modes that contribute most to thermodynamic properties, of interest to this study, i.e. the overall shape of the B-DNA molecule. The modes between $800cm^{-1}$ - $1500cm^{-1}$ which contain the C-O motions and the sugar pucker marker: $815cm^{-1}$ C3′-endo, $830cm^{-1}$ C2′-endo (29) are less well fit.

The effect of the replacement of the four tetrahedral angles of the sugar was examined using a more detailed force field, the QCFF/PI+MCA package (30) commonly used for small to medium size molecules (31-32). The replacement induced a rise in the potential barrier between C2′-endo and C3′-endo from 1.25 Kcal/mole to 2.27 Kcal/mole but the sugar pucker marker frequencies changed from $816cm^{-1}$, $832cm^{-1}$ to approximate $790cm^{-1}$ for both.

From earlier computer simulation on the sugar ring (M. Levitt, R. Sharon unpublished results), it became evident that the height of the barrier is strongly correlated to the vibrational spectra. The total energy is a sum of the bending, torsional, nonbonded and bonding energy terms. On the pseudorotation path, the bonding energy is almost constant and the nonbonded energy has a maximum at O1′ exo sugar pucker. Torsional and bending energy terms oscillate with almost the same amplitude but opposite phases and give constant contribution to the total energy. An energy barrier higher than 1.7 Kacl/mole is correlated with bending frequencies which are lower than the experimental frequencies, i.e. the bending energy term is too flat and thus does not cancel the torsional energy term oscillations. Consequently, the total energy oscillates with almost the same amplitude as the torsional potential energy.

In summary, the unmodified Levitt potential, as well as the modified one, fit well the observed spectra for frequencies below $700cm^{-1}$ and maintain the overall shape and helical parameters of B-DNA. To fit the modes between $800cm^{-1}$ - $1500cm^{-1}$ which contain the sugar pucker, a better potential field is required, as well as higher resolution X-ray crystallographic studies of B-DNA in order to check the potential.

Conclusions

We suggest that incorporation of an additional base into a B-DNA structure results in local changes which manifest themselves in a kink of about 30° at the insertion site. This result is consistent with, and based on NMR and calorimetric studies on the tridecamer sequences $d(CGCAGAATTCGCG)_2$ and $d(CGCAGAGCTCGCG)_2$ which indicate that in solution the additional adenosine residues are stacked into a B-DNA duplex (4-6). The fact that the energy minimized model is symmetric and that the two outer segments are on the same side of the central segment, is due only to the fact that there are exactly six base pairs in the central segment (half a helical turn)

and that the two additional helical bases are on opposite strands. The symmetry is likely to be only an approximation of the real structure, as was shown for the dodecamer (33) where there is symmetry in the sequence but not in the three-dimensional structure.

The hydrogen bond formed between the additional adenosine and the flanking cytidine on the opposite strand suggests an alternative way for the additional adenosine to be stabilized inside the helix. This hydrogen bond scheme is very much dependent on the nature of the flanking base pairs around the additional adenosine. Although the structure with the extra hydrogen bond is somewhat more stable, in solution the C_{10} may be disordered and may actually be flipping between the two structures.

The phenomena of a base forming hydrogen bonds with *two* bases on the opposite strand is more general. Recent X-ray studies on DNA sequences containing stretches of poly(dA) · poly(dT) (34-35) have demonstrated the formation of hydrogen bonds between the carbonyl O4 group of two adjacent thymines with the NH_2 amino group of the adenosine on the opposite strand. This type of bifurcated hydrogen bonds are believed to contribute to the stability, and special conformational features such as bendability of the double helix.

DNA segments are highly flexible and the actual conformation is sequence as well as environmental dependent. In this work we studied the energetical and conformational effects of an additional adenosine staking into a B-DNA duplex. We are studying now the energetical and conformational effects on the duplex, of a "looped out" adenosine. This study is being carried out in solution as the stabilization of the additional adenosine is probably due to solvation of this base by water molecules.

Acknowledgments

We acknowledge the helpful discussions with Peter Stern, Leemor Joshua-Tor, Mark Saper, and Fred Hirshfeld. We especially thank Michael Levitt for providing the ENCAD and IMDAD programs, and Sam Motherwell for the PLUTO molecular drawing program. This research was supported in part by a Grant from the U.S. Army Research Office (through its European Research Office) to J.L.S.

References and Footnotes

1. G. Streisinger, Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi and M. Inouye, *Cold Spring Harbor Symp. Quant. Biol. 31*, 77 (1966).
2. J.W. Drake, B.W. Glickman and L.S. Ripley, *American Scientist 71*, 621 (1983).
3. D.J. Patel, L. Shapiro and M. Kalnik, *J. Biol. Chem.* in press (1987).
4. D. Hare, L. Shapiro and D.J. Patel, *Biochemistry 25*, 7456 (1986).
5. S. Roy, V. Sklenar, E. Appella, J. S. Cohen, *Biopolymers 26*, 2041 (1987).
6. D.J. Patel, S.A. Kozlowski, L.A. Marky, J.A. Rice, C. Broka, K. Itakura and K.J. Breslauer, *Biochemistry 21*, 445 (1982).
7. Y.G. Chu and I. Tinoco Jr., *Biopolymers 22*, 1235 (1983).
8. J.W. Keepers, P. Schmidt, T.L. James and P.A. Kollman, *Biopolymers 23*, 2901 (1984).
9. S. Roy, S. Weinstein, B. Borah, J. Nickol, E. Appella, J.L. Sussman, M. Miller, H. Shindo and J.S. Cohen, *Biochemistry 25*, 7417 (1986).

10. M. Miller, W. Kirchhoff, F. Schwarz, E. Appella, Y.H. Chiu, J.S. Cohen, and J.L. Sussman, *Nucleic Acids Res. 15*, 3877 (1987).
11. T. Brown, O. Kennard, G. Kneale and D. Rabinovich, *Nature (London) 315*, 604 (1985) and references cited therein.
12. R. Rein, M. Shibata, R. Gardino-Juarrez and T. Keiber-Emmons in *Structure and Dynamics of Nucleic Acids and Proteins*, E. Clementi and R. Sarma (eds.) Adenine Press, N. Y., p. 269 (1983).
13. K.M. Morden, Y.G. Chu, F.H. Martin and I. Tinoco Jr., *Biochemistry 22*, 5557 (1983).
14. Y. Wang, G.A. Thomas and W. L. Peticolas in *The Fifth Conversation in Biomolecular Stereodynamics*, R.H. Sarma (ed.), Adenine Press, N.Y. pp. 61 (1987).
15. M. Levitt, *Proc. Natl. Acad. Sci. (USA) 75*, 640 (1978).
16. M. Levitt, *Cold Spring Harbor Symp. Quant. Biol. 47*, 251 (1983).
17. M.A. Saper, H. Eldar, K. Mizuuchi, J. Nickol, E. Appella and J.L. Sussman, *J. Mol. Biol. 188*, 111 (1986).
18. J.L. Sussman, L. Joshua-Tor, M. Hirshberg, M.A. Saper, F. Frolow, H. Hope and E. Appella in *International Symp. on Molecular Structure*, Beijing, China, in press (1987).
19. S. Arnott and D.W.L. Hukins, *Biochem. Biophys. Res. Commun. 47*, 1504 (1972).
20. R.E. Dickerson and H.R. Drew, *J. Mol. Biol. 149*, 761 (1981).
21. W. Saenger in *Principles of Nucleic Acid Structure*, C.R. Cantor (ed.) Springer-Verlag, New York, pp. 350 (1984).).
22. P.A. Kollman, P.K. Weiner and A. Dearing, *Biopolymers 20*, 2583 (1981).
23. S. Arnott, R. Chandrasekaran, D.L. Birdsall, A.G.W. Leslie and R.L. Ratliff, *Nature (London) 283*, 743 (1980).
24. E.N. Trifonov and J.L. Sussman, *Proc. Natl. Acad. Sci. USA 77*, 3816 (1980).
25. A.A. Travers, *TIBS 12*, 108 (1987).
26. P.K. Weiner and P.A. Kollman, *J. Comp. Chem. 2*, 287 (1981).
27. W. Saenger in *Principle of Nucleic Acid Structure*, C.R. Cantor (ed.) SpringerVerlag, New York, pp. 65-76 (1984).
28. S.J. Weiner, P.A. Kollman, D.A. Case, U. Chandra Singh, C. Ghio, G. Alagona, S. Profeta Jr. and P. Weiner, *J. Am. Chem. Soc. 106*, 765 (1984).
29. W.L. Peticolas and G.A. Thomas in *Structure and Motion: Membranes, Nucleic Acids and Proteins*, E. Clementi, G. Corongiu, M.H. Sarma and R.H. Sarma (eds), Adenine Press, New York, pp 497 (1985).
30. E. Huler, R. Sharon and A. Warshel, *QCPE* 325 (1976).
31. A. Warshel, *Computers and Chemistry 1*, 195 (1977).
32. A. Warshel, *Modern Theoretical Chemistry 7*, 133 (1977).
33. R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R.E. Dickerson, *Nature (London) 287*, 755 (1980).
34. H.C.M. Nelson, J.T. Finch, B.F. Luisi and A. Klug, *Nature (London) 330*, 221 (1987).
35. M. Coll, C.A. Frederick, A.H.J. Wang and A. Rich, *Proc. Natl. Acad. Sci. (USA) in press (1987).*

Date Received: October 4, 1987.

Communicated by the Editor Wayne Hendrickson