

A note on sequence-dependence of DNA structure

A. Galat

Department of Chemistry, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

Received July 28, 1989/Accepted in revised form October 3, 1989

Abstract. A circular dichroism study was conducted on the solution structure of several different oligonucleotides, whose X-ray structures have been solved. It is suggested that in aqueous solution the oligonucleotides can form structures that maintain geometrical elements which are typical of B-DNA, A-DNA, and their intermediate forms. It is shown that 5'GGATGGGAG:5'CTCCCATCC, which forms an A-DNA helix in the crystal state (McCall et al. 1986), in aqueous solution maintains an A-DNA like structure at temperatures below 10°C. At temperatures between 10°C and 25°C it shows a tendency to form an intermediate structure between A-DNA and B-DNA. Also, it is shown that TFE does not cause a transition from B-DNA to A-DNA helix in short DNA fragments, but instead disrupts the helix.

Key words: DNA structure, A-DNA, B-DNA, CD spectra, molecular dynamics simulation

Introduction

A number of studies has shown that DNA can adopt a variety of different structures which are determined by the solvent, salts, temperature, time, and DNA sequence (Leslie et al. 1980; Wang et al. 1982; Shakked and Rabinovich 1986; Fairall et al. 1989). Also, it has been shown that DNA can alter its structure with changes in water content (Lindsey et al. 1988). For example, using Raman scattering spectroscopy and X-ray diffraction crystallography differences are observed between the crystal structures of DNA and their counterparts in aqueous solution (Benevides et al. 1986, 1988; Wang et al. 1989).

There are generally three techniques used to study DNA conformation in solution, namely NMR, CD

and Raman scattering. Raman spectroscopy allows determination of sugar ring conformation characteristic of A-DNA (predominantly the C3'-endo sugar pucker) or B-DNA (predominantly the C2'-endo sugar pucker). Also, it allows for the establishment of the conformation of the DNA backbone. However, B-DNA may also adopt the C3'-endo sugar pucker in base steps which have smaller than average global helical twist and smaller than average distance between the phosphorus atoms. On the other hand, A-DNA may adopt conformations for sugar moieties other than the canonical C3'-endo, especially in base steps which have higher than average global helical twist and distance between the phosphorus atoms longer than 5.8 Å. A number of NMR-elucidated structures of short DNA duplexes apparently contain helical parameters typical of A-DNA structure (Nilsson et al. 1986; Nilges et al. 1987).

CD spectroscopy can supply information on the extent of base-base interaction in the helical stack, overall stability of the helix, and can show gross differences in the helix structure, among others. The shape and magnitude of the Cotton effects of short fragments of DNA are dependent on its sequence, salt and temperature. The sequence-dependence of Cotton effects usually becomes negligible for long natural double-stranded DNAs. However, in short oligonucleotides it is usually difficult to give a precise interpretation of CD results.

NMR-elucidated structures of DNA are not simple in refinement (van Gunsteren et al. 1986; Nilsson et al. 1986; Patel et al. 1987) since the difference in H-H distances in model A- and B-DNA helices are not of a great enough magnitude (Wüthrich 1986) and the error of NOE calculated distances is about 0.5 Å. This means that the atomic resolution of an NMR-elucidated structure is usually between 3 to 5 Å. Moreover, it has been well documented that DNA structures change upon formation of a complex with protein (McClarín et al. 1986; Thomas et al. 1989) and drugs

(Wang et al. 1987). Also, various footprinting experiments have revealed long-range readjustment in DNA structure resulting from single base pair substitution, several bases away from the drug binding site (Murray and Martin 1985).

In this communication the author attempts to discuss sequence-dependent variations of double stranded helices of A-DNA and B-DNA and their intermediate forms which may exist in aqueous solution.

Materials and methods

Synthesis and purification of oligonucleotides

The oligonucleotides shown in Table 1 were synthesized using phosphoramidite chemistry (Sinha et al. 1984). Crude oligonucleotides were purified on a 20% polyacrylamide gel containing 7 M urea. The purified oligonucleotides were desalted on a SEP-PAK reverse phase C18 column (Waters Associates) (Lo et al. 1984). Equimolar quantities of each strand of non-self-complementary duplex were mixed at room temperature in the appropriate buffer, heated to 10°C over the melting point of the duplex (usually above 65°C), and allowed to equilibrate for 15 min, and then allowed to cool down over 30 min. These samples were used for circular dichroism and thermodynamic measurements.

Melting curves and data analysis

Absorbance measurements were made with a Perkin-Elmer spectrophotometer in standard low salt buffer (0.1 M NaCl, 10 mM sodium phosphate, 0.1 mM EDTA, pH 7.5). Each measurement was made over the range 5°–85°C. The temperature was raised at a rate of not more than 1°C/min, while the absorbance was monitored at 260 nm. The oligonucleotide concentration was estimated from separate measurements on samples at 85°C at which the oligonucleotide was totally melted. The molar absorptivities of the oligonucleotides were estimated based on the extinction coefficients for the four deoxyribonucleotides (Nadeau and Gilham 1985). The fraction, α , of molecules in the duplex state is obtained from the following expression:

$$\alpha = \frac{\varepsilon_s - \varepsilon(T)}{\varepsilon_s - \varepsilon_h}, \quad (1)$$

where ε_h and ε_s are average absorptivities of the single strands and the duplex, and $\varepsilon(T)$ is the absorptivity measured as a function of temperature. The equilibrium constant (K) for the formation of a self-com-

Table 1. DNA oligonucleotides studied by circular dichroism

DNA sequence	Code	X-ray structure
5'GTGTGTG CACACAC-5'	(7)	
5'GTGTACAC CACATGTG-5'	(8A)	(Jain et al. 1987)
5'GCACGTGC CGTGCACG-5'	(8B)	
5'GGATGGGAG CCTACCCTC-5'	(9)	(McCall et al. 1986)
5'GTGTGCATAT CACACGTATA-5'	(10)	
5'CGCGAATTCGCG GCGCTTAAGCGC-5'	(12A)	(Dickerson and Drew 1981)
5'CGCAAATTTGCG GCGTTTAAACGC-5'	(12B)	(Coll et al. 1987)
5'CGCAAAAAAGCG GCGTTTTTTCGC-5'	(12C)	(Nelson et al. 1987)
5'GGAAAATTTTCC CCTTTTAAAAGG-5'	(12D)	

plementary duplex is,

$$K = \frac{\alpha}{[2(1-\alpha)^2 C_s]} \quad (2)$$

and for the non-self-complementary duplex,

$$K = \frac{2\alpha}{[(1-\alpha)^2 C_s]}, \quad (3)$$

where C_s is the total oligonucleotide strand concentration (Borer et al. 1974). Thermodynamic parameters were first estimated from linear regression of α vs. $1/T$ using the equation;

$$K = \exp(-\Delta H^0/RT + \Delta S^0/R) \quad (4)$$

and later optimized using the Powell non-linear least-squares minimizer (Powell 1965). The parameters optimized were ΔH^0 , ΔS^0 , m_d and n_d ; m_d and n_d are linear functions of temperature for the absorptivity of the duplex, where $\varepsilon(T) = m_d T + n_d$ (Nadeau and Gilham 1985).

Circular dichroism measurements

Circular dichroism spectra are averages of three recordings on two different samples. The CD spectra shown in Figs. 1, 2, and 3 are the original, scaled dichrograph recordings. CD spectra were measured in 0.1 M NaCl–0.01 M phosphate buffer containing 0.1 mM EDTA at 6°C and room temperature (20°C) with a Jobin Yvon Mark V dichrometer. The circular

dichroism values are expressed in $M^{-1} \text{ cm}^{-1}$ and were calculated on the basis of mean residue concentration. The CD spectra were analysed with a recently described program (Galat 1986).

DNA structure

The Cartesian coordinates of model DNA structure were built from cylindrical polar coordinates of DNA components (Arnott et al. 1969; Arnott and Hukins 1972) with the use of the CORGEN program (Galat 1989a). The Cartesian coordinates of given atom (i) in the n th residue are calculated according to the following equations:

$$X_i = R(i) * \cos(\phi_i + n * \Omega) \quad (5)$$

$$Y_i = R(i) * \sin(\phi_i + n * \Omega) \quad (6)$$

$$Z_i = Z_i + n * D_z, \quad (7)$$

where $R(i)$ is the radial part of polar cylindrical coordinates and ϕ_i is its angular component, whereas D_z is the global helical rise (3.4 Å for B-DNA and 2.55 Å for A-DNA) and Ω is the global helical twist (36° for B-DNA and 32.7° for A-DNA, respectively). All of the missing atoms are added to DNA structure by the ATOM FIX procedure by the CORGEN program using standard bond lengths, bond angles and torsion angles.

The oligonucleotide structure was optimized with the CHARMM program using the adapted basis Newton-Raphson (ABNR) method (Brooks et al. 1983). The potential energy in CHARMM is of the form

$$E = E_b + E_a + E_t + E_i + E_{vdw} + E_{el} + E_{hb}, \quad (8)$$

where E is the potential energy with respect to the Cartesian coordinates of the atoms, E_b is the bond energy, E_a is the bond angle energy, E_t is the torsion energy, E_i is the improper torsion angle energy, E_{vdw} is the van der Waals energy, E_{el} is the electrostatic energy, and E_{hb} is the hydrogen bond energy. All energy components are expressed in kcal/mol (Brooks et al. 1983). The X-ray diffraction structure of 5'GGATGGGAG:5'CTCCCATCC elucidated to 3 Å resolution by McCall et al. (1986) was acquired from the recent edition of the Protein Databank in Brookhaven PDB tape. It was optimized with 2,000 steps of ABNR. Molecular dynamics simulations were performed with the CHARMM program as previously described (Tidor et al. 1983). Dynamics trajectories were analysed in terms of helical parameters using the CORDAN program (Galat 1989b). Main-chain conformation angles are defined as $P^2-O5'^\beta-C5'^\gamma-C4'^\delta-C3'^\epsilon-O3'^\zeta$ with zero at the fully eclipsed position and positive angles by clockwise rotation of the further

pair of atoms. Base roll (ϱ), global helical twist (Ω), and propeller twist (ω) are expressed in degrees, whereas global helical rise (D_z) and slide of base pairs (D_y) are expressed in Å. Fluctuations of helical parameters are denoted by an f . The calculations were performed in part with VAXstation 3500 and Convex C1 computers.

Results

A. Conformation and stability of the nonamer 5'GGATGGGAG:5'CTCCCATCC in solution

The circular dichroism spectra of the duplex and its constituent strands were measured in aqueous solution (see Fig. 1). The CD spectra showed a large positive Cotton effect in the near UV which was followed by a somewhat smaller Cotton effect centered at about 238 nm. Such a profile for the CD spectra differs from the CD spectra of RNA and tRNA samples (Johnson 1985) and is not typical of B-DNA. However, in a previous study on ribo(CG)₃ (Adamiak et al. 1985), it was shown that the CD spectra of a short RNA fragment does not exhibit asymmetrically split profiles for the long wavelength exciton band. This is believed to be an indication of A-DNA type of structure (Aboul-ela et al. 1988). The shape and magnitude of the CD curve of DNA reflects exciton interaction between electronic transitions in nucleic acid bases. Hence, these interactions depend largely on the extent of intra-strand base overlap in the DNA molecule and only partially on inter-strand base overlap. The observed CD profile of the nonamer in aqueous solution shows that the exciton splitting is non-conservative with higher rotatory strength in the longer wavelength part of the CD band. Moreover, the CD spectrum recorded at 6°C showed an even more asymmetrical

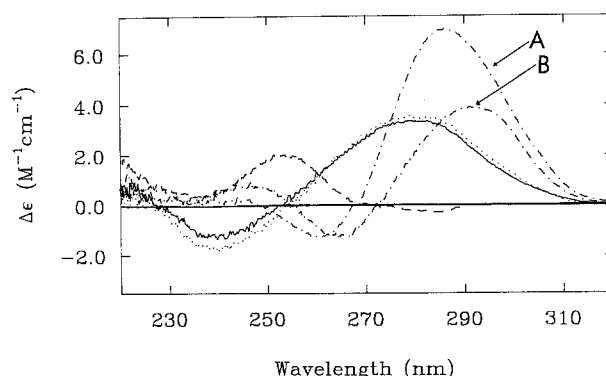


Fig. 1. The CD spectra of the 5'GGATGGGAG:5'CTCCCATCC duplex at 6°C (solid line), 20% methanol (dotted line), 5'GGATGGGAG in 67% aqueous solution of TFE (dashed line), 5'CTCCCATCC in 67% aqueous solution of TFE (A) and duplex in 67% aqueous solution of TFE (B)

Table 2. Spectroscopic enthalpies, entropies and melting temperatures for coil to helix transition in the series of oligonucleotides

Structure	ΔH^a [kcal/mol]	ΔS [e. u.]	T_m [°C]	Strand concentration [μM]
(7)	-65.6	-188.4	35.0	12.02
(8A)	-38.8	-100.8	40.0	10.77
(8B)	-57.0	-154.0	49.0	10.46
(9)	-52.2	-139.4	42.0	7.23
(10)	-42.6	-107.5	51.0	9.53
(12A)	-48.6	-128.7	49.5	11.96
(12B)	-52.0	-137.0	50.5	7.51
(12C)	-51.9	-137.6	50.0	14.02
(12D)	-47.3	-126.5	51.5	9.12

^a The mean values of ΔH and ΔS were determined from two different experiments and they did not differ by more than 5%

CD profile than that observed at 20°C. Addition of methanol (up to 20%) to aqueous solutions of the oligomer did not influence its CD spectrum. This may indicate that small increases in hydrophobicity of the solution do not alter the conformation of DNA. The melting profile of the oligomer yielded thermodynamic parameters as shown in Table 2.

Recently it has been suggested (Aboul-ela et al. 1988), based on CD and NMR studies, that the nonamer adopts a B-DNA structure in solution. Also it was suggested that at 70% TFE, a transition occurred from the B-DNA to the A-DNA form. In my study no melting of the duplex was observed in the temperature range 5°–85°C in aqueous solution containing 67% TFE and 0.05 M sodium salt. This suggests that the helix was disrupted by high concentrations of TFE. In solutions with a high content of TFE the cytosine-rich strand exhibits a large Cotton effect ($\Delta\epsilon$ 7.0 M⁻¹ cm⁻¹) (Fig. 1) whereas the guanosine-rich strand shows rather small Cotton effects. Higher than average Cotton effects which occur above 300 nm can be due to partially-protonated cytosine residues (see Edwards et al. 1988); note that an aqueous solution of TFE is slightly acidic. Obviously, the large CD band in 70% TFE (Aboul-ela et al. 1988) is due to a melted duplex instead of an 'A-DNA conformation'. Such transitions are known to occur even for RNA oligomers, i.e., addition of TFE to a buffered solution of r(CGCGCG)₂ at a TFE/H₂O ratio of 1/1 (v/v) causes a total collapse of the double helix (Adamiak et al. 1985). Comparison between the CD spectra of single stranded nonamer's components with that of the duplex shows that the latter is dominated by the Cotton effects of the cytosine rich strand. The CD spectrum of the duplex shows, however, a larger half-band width for the long wavelength component associated with the $\Pi \rightarrow \Pi^*$ exciton band as compared to the CD

spectrum of the cytosine-rich single strand. This broadening of the CD curve indicates the formation of a double helix. Moreover, the CD spectrum of the duplex at low temperature shows higher asymmetry of the long wavelength Cotton effect and is more similar to the CD spectrum of A-DNA. Based on the CD spectra it is suggested that at low temperature (within the range 5°–10°C), the nonamer shows a tendency to form a structure related to A-DNA. At temperatures between 10°–25°C there may exist intermediate forms in equilibrium between A- and B-DNA and over 25°C the duplex begins to melt.

In a more extensive CD study on solution structure of DNA fragments by Fairall et al. (1989) it has been suggested that the nonamer has an intermediate form between A-DNA and B-DNA. Moreover, the structure of the duplex is dependent on the presence of divalent cations and spermidine (Fairall et al. 1989). These factors are known to exert a stabilizing effect on the DNA double helix. Thus, temperature induced changes in the CD spectra reported in this work and lack of such changes reported by Fairall et al. (1989) points to higher stability of the nonamer in solution used in the crystallization of the duplex (McCall et al. 1986).

B. CD spectra of the heptamer 5'GTGTGTG : 5'CACACAC duplex and related sequences

The CD spectra of the 5'GTGTGTG:5'CACACAC duplex showed small asymmetric CD bands in the near UV region with another Cotton effect centered at about 240 nm (see Fig. 2). Also the CD spectra of single strands are shown. It is apparent that 5'CACACAC shows CD bands which are twice as strong as those of the duplex. Thus the cytosine-rich strands show a tendency to exhibit strong Cotton effects which can mimic CD bands of A-DNA. The CD spectrum for a self-complementary 5'GTGTACAC octamer duplex (8A) is shown in Fig. 2. The octamer is closely related in sequence to the heptamer, and adopts an intermediate form between A-DNA and B-DNA in the crystal state (Jain et al. 1987). The CD spectrum of 8A has a distinct band profile which is not typical of canonical B-DNA, with somewhat higher intensity for the long wavelength part of the $\Pi \rightarrow \Pi^*$ exciton band. A self-complementary decamer duplex 5'GTGTGCACAC which has a G-C base step dividing the GT stream from the CA stream, exhibits a different CD spectrum from the heptamer, and 8A. A more unusual CD profile was observed for the self-complementary 5'GCACGTGC octamer duplex (Fig. 2), whose sequence is closely related to 7, 8A and 10. The CD spectrum shows a low rotatory strength for the

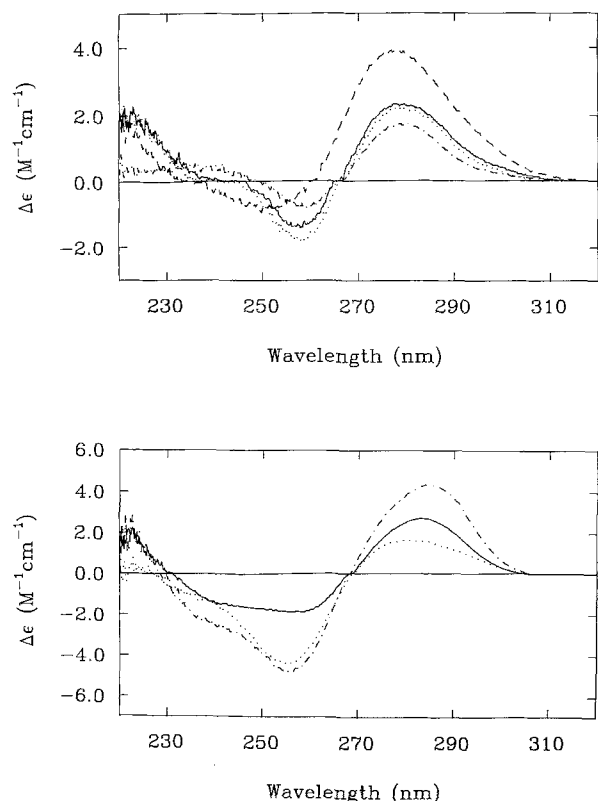


Fig. 2. Circular dichroism spectra of 5'GTGTGTG:5'CACACAC (upper panel) duplex in the aqueous solution (continuous line), duplex with 20% MeOH (dotted line), single stranded 5'GTGTGTG in aqueous solution (dotted-dashed line) and single stranded 5'CACACAC in aqueous solution (dashed line); (lower panel) self-complementary 5'GTGTACAC duplex (continuous line), self-complementary 5'GCACGTGC duplex (dotted line) and self-complementary 5'GTGTGCACAC duplex (dotted-dashed line)

long wavelength part of the exciton CD bands and a large negative Cotton effect centered about 255 nm.

The difference in the $\Delta\epsilon$ values of the long wavelength CD bands of the heptamer duplex and the octamer 5'GTGTATAT duplex can be related to differences in the DNA sequence. Thus, the CD spectra shown in Fig. 2 clearly indicate that the heptamer and octamer 8A have similar, non-standard B-DNA structures in aqueous solution, whereas 8B and 10, which have G-C bp, are in B-DNA like structures.

The helix stability of 8A is smaller than that of 8B and the latter effect is related directly to the sequence, namely 8B has more G-C base pairs than 8A. The heptamer showed, however, higher coil-to-helix transition enthalpy than the octamers, and the dodecamers. Since in our experiments the strand concentrations were similar the differences in ΔH and ΔS must be due to intrinsic properties existing within each oligonucleotide duplex, such as different curvature of the helix, hydration shells and salt-dependent helix stabilization, among others.

C. CD spectra of dodecamers whose X-ray structures have been elucidated

Figure 3A shows the CD spectrum of the self-complementary dodecamer 5'CGCGAATTCGCG duplex whose X-ray structure has been solved (Dickerson and Drew 1981). The spectrum shows quasi-symmetric CD bands with somewhat higher rotatory strength for the short wavelength part of the spectrum. Figure 3B shows the CD spectrum of another dodecamer 5'CGCAAATTTGCG whose X-ray structure was recently established (Coll et al. 1987). Both dodecamers show similar CD profiles since they do not differ significantly in sequence, and they both exist in the B-DNA structure in the crystal state (Dickerson and Drew 1981; Coll et al. 1987). There is, however, a somewhat higher intensity of the negative Cotton effect in the former compared to the latter. This may be related to differences in the exciton splitting of the $\Pi \rightarrow \Pi^*$ electronic transitions. Namely, in the latter sequence there is a 5'TG base step and 5'CA base step while the former has a 5'GA base step and a 5'TC base step. In Fig. 3C, the CD spectrum of 5'CGCAAAAAAGCG:5'CGCTTTTTTTGCG (Nelson et al. 1987) is shown to exhibit symmetric splitting of the long wavelength CD profile. A very similar CD spectrum was exhibited by self-complementary 5'GGAAAATTTTCC. Essentially they show very similar CD spectra although their base content is not the same. Thus it is possible that both duplexes exist as structurally related B-DNA forms in aqueous solution and in the solid state.

It has been well documented that the stability of the double helix depends on salt concentration, strand concentration, buffer composition (Marky et al. 1983) and sequence (Nadeau and Gilham 1985). These oligonucleotides differ from each other in their helix stabilities (see Table 2). However, this variation is not related to the sequence in a straightforward fashion and may be partially due to errors associated with spectral measurements.

D. The optimized A-DNA crystal structure of 5'GGATGGGAG:5'CTCCCATCC

The crystal structure of the nonamer has been solved at 3 Å resolution (McCall et al. 1986). It shows characteristic helical parameters for an A-DNA structure, i.e., a large displacement of base-pairs from the imaginary helix axis, a substantial base pair inclination with respect to the helix axis with base-pair slide greater than 1 Å, and smaller Ω and D_z values as compared to B-DNA structure. The base numeration is as follows: 5'G₁G₂A₃T₄G₅G₆G₇A₈G₉(+) : 5'C₁₀T₁₁C₁₂C₁₃C₁₄A₁₅T₁₆C₁₇C₁₈(-).

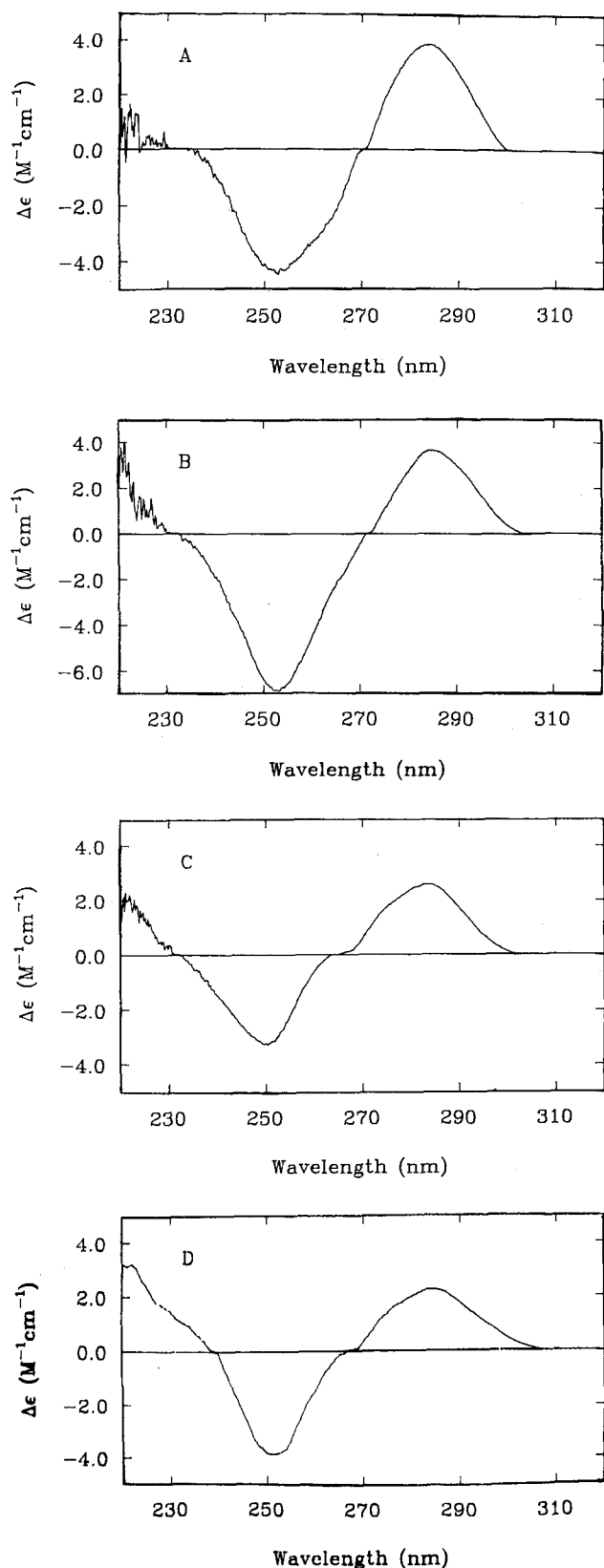


Fig. 3A–D. CD spectra of the oligonucleotides (A) self-complementary 5'CGCGAATTCGCG duplex, (B) self-complementary 5'CGCAAATTTGCG duplex, (C) 5'CGCAAAAAGCG:5'CGCTTTTTCG duplex, (D) self-complementary 5'GGAAAATTTTCC

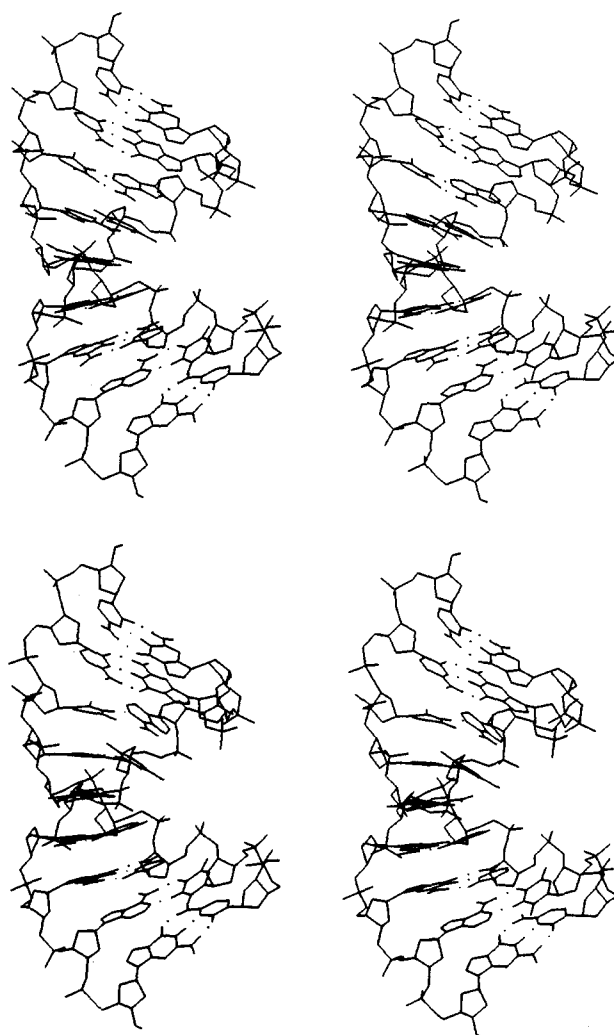


Fig. 4. Stereodrawings of the optimized (upper panel) and 20-ps optimized average dynamics structure (lower panel) of (A) model A-DNA structure of 5'GGATGGGAG:5'CTCCCATCC duplex, the total potential energy of the optimized structure is $E = -564$ kcal/mol and the optimized average dynamics structure $E = -584$ kcal/mol

Molecular mechanics and dynamics simulation methods may provide some information on sequence-dependence of DNA structure. However, no explicit salt and water molecules were included in the calculations presented in this work. Thus, the calculated structures show variations of helical parameters which are valid only for the intramolecular interactions within the double helix. A more realistic picture of sequence-dependent variations of helical parameters and helix stability could be derived from computer simulations in which one would include explicitly "the real buffer content", i.e. Mg^{2+} , Na^+ and Cl^- (about 0.1 M), spermidine and water molecules, among others. Because the crystal structure at 3 Å resolution does not permit precise establishment of all helical parameters, the X-ray structure was optimized *in vacuo* and is shown in Fig. 4 (upper panel). It has an

equally variable distribution of sugar puckers as determined in the original X-ray crystal structure. For example, G7 and A8 are in the C2'-endo and C3'-exo conformation, respectively. C10 and C18 retained the C2'-endo pucker whereas C13 and C17 remained in the C3'-exo pucker. The separation between phosphorous atoms was somewhat higher than in the canonical A-DNA structure to accommodate sugar pucker different than the C3'-endo. Their mean separation is 5.95 Å (*SD* 0.65 Å) whereas for the optimized structure it is 5.9 Å (*SD* 0.89 Å). Thus, A-DNA structure does not have to adopt the uniform C3'-endo pucker.

E. X-ray diffraction average dynamics structure of 5'GGATGGGAG : 5'CTCCCATCC

Molecular dynamics simulations were performed on the optimized X-ray diffraction structure of the helix (McCall et al. 1986). This technique allows one to sample local minima existing around the optimized structure and supplies information on the intramolecular interactions in those minima represented by the changes of helical parameters. A 20-ps dynamics simulation caused alterations of oligonucleotide geometry (the helical parameters are summarized in Tables 3 and 4). The optimized average dynamics structure is shown in Fig. 4 (lower panel). It is apparent that a number of sugar units did not change their pucker and for example G7, A8, C13 and C17 remained in the C3'-exo pucker whereas C10 and C18 were in the C2'-endo pucker. The values of Ω remained 31.7° (*SD* 8.6°) and D_z was 2.41 Å (*SD* 0.57 Å). The mean separation of phosphorus atoms remained 5.9 Å (*SD* 0.78 Å). This confirms again that A-DNA does not have to adopt the canonical all-C3'-endo conformation. Other sugar puckers may be present owing to optimization of different helical parameters. The Ω value for the T4-A15/G5-C14 base step increased to 34.1° ($f=6.7^\circ$) but it decreased for the following base step to 22.9° ($f=5.4^\circ$). The fluctuations of some of the helical parameters for the T4-A15/G5-C14 base step are shown in Fig. 5. The base roll angle (ϱ) for the G2-C17/A3-T16 base step decreased to 8.2° (27.7° in X-ray) whereas it increased for the following base step to 16.2° (vs X-ray crystal structure 8.9°). Furthermore, there are two negative base roll angles in the crystal structure for the G5-C14/G6-C13 and A8-T11/G9-C10 base steps, whose signs remained unchanged in the average dynamics structure. This is another departure from the canonical A-DNA structure where all of the base roll angles are positive. The negative base roll angle is compensated by considerable slide ($D_y=2.8$ Å for the G5-C14/G6-C13 base step and $D_y=2.6$ Å for the G6-C13/G7-C12 base step). Also a number of sugar units occupy higher energy conformations than the C3'-endo (see Table 3).

Table 3. Amplitude of pucker (τ_m), pseudorotation phase angle (P), δ torsion angle, and sugar conformation in the average dynamics structure simulated with the X-ray coordinates of 5'GGATGGGAG:5'CTCCCATCC

Sequence	τ_m	P	δ	Conformation
5'-G1	48.0	22.2	73.8	C3'-endo
G2	44.6	6.7	77.0	C3'-endo
A3	45.8	21.2	70.9	C3'-endo
T4	47.7	7.8	70.2	C3'-endo
G5	44.3	-0.6	83.5	C3'-endo
G6	44.1	1.3	81.6	C3'-endo
G7	41.0	181.2	156.9	C3'-exo
A8	41.5	211.7	166.2	C3'-exo
G9	47.5	164.4	72.3	C2'-endo
5'-C10	44.5	175.4	159.3	C2'-endo
T11	48.1	6.4	70.9	C3'-endo
C12	48.0	8.1	72.8	C3'-endo
C13	40.2	211.3	165.9	C3'-exo
C14	47.4	13.3	73.6	C3'-endo
A15	45.2	0.2	78.5	C3'-endo
T16	47.3	8.9	72.3	C3'-endo
C17	40.7	189.4	159.2	C3'-exo
G18	44.4	174.2	158.2	C2'-endo

Table 4. Helical parameters of base steps and base pairs in the average dynamics structure of 5'GGATGGGAG:5'CTCCCATCC simulated with the optimized X-ray diffraction structure [23]

B/P	Step	ϱ	Ω	D_z	D_y	ω
G1-C18						-11.7
G2-C17	1	-7.3	22.0	2.45	1.64	-24.7
A3-T16	2	8.2	32.0	3.16	1.08	-31.4
T4-A15	3	16.3	38.1	2.14	1.23	-34.2
G5-C14	4	9.4	34.1	2.02	1.46	-8.9
G6-C13	5	-1.7	22.9	1.57	2.79	-6.2
G7-C12	6	7.0	36.1	3.13	2.27	-15.4
G8-C11	7	8.2	30.1	2.67	1.40	-6.9
G9-C10	8	-7.0	38.4	2.10	2.24	-20.4
Average		2.4	31.7	2.41	1.58	-17.7

Base roll (ϱ), global helical twist (Ω), and propeller twist (ω) are expressed in degrees, whereas global helical rise D_z , slide of base pairs D_y are expressed in Å

F. B-DNA model of 5'GGATGGGAG : 5'CTCCCATCC

Based on NMR data it has been suggested that in aqueous solution the nonamer exists in the B-DNA form (Aboul-ela et al. 1988). A model B-DNA struc-

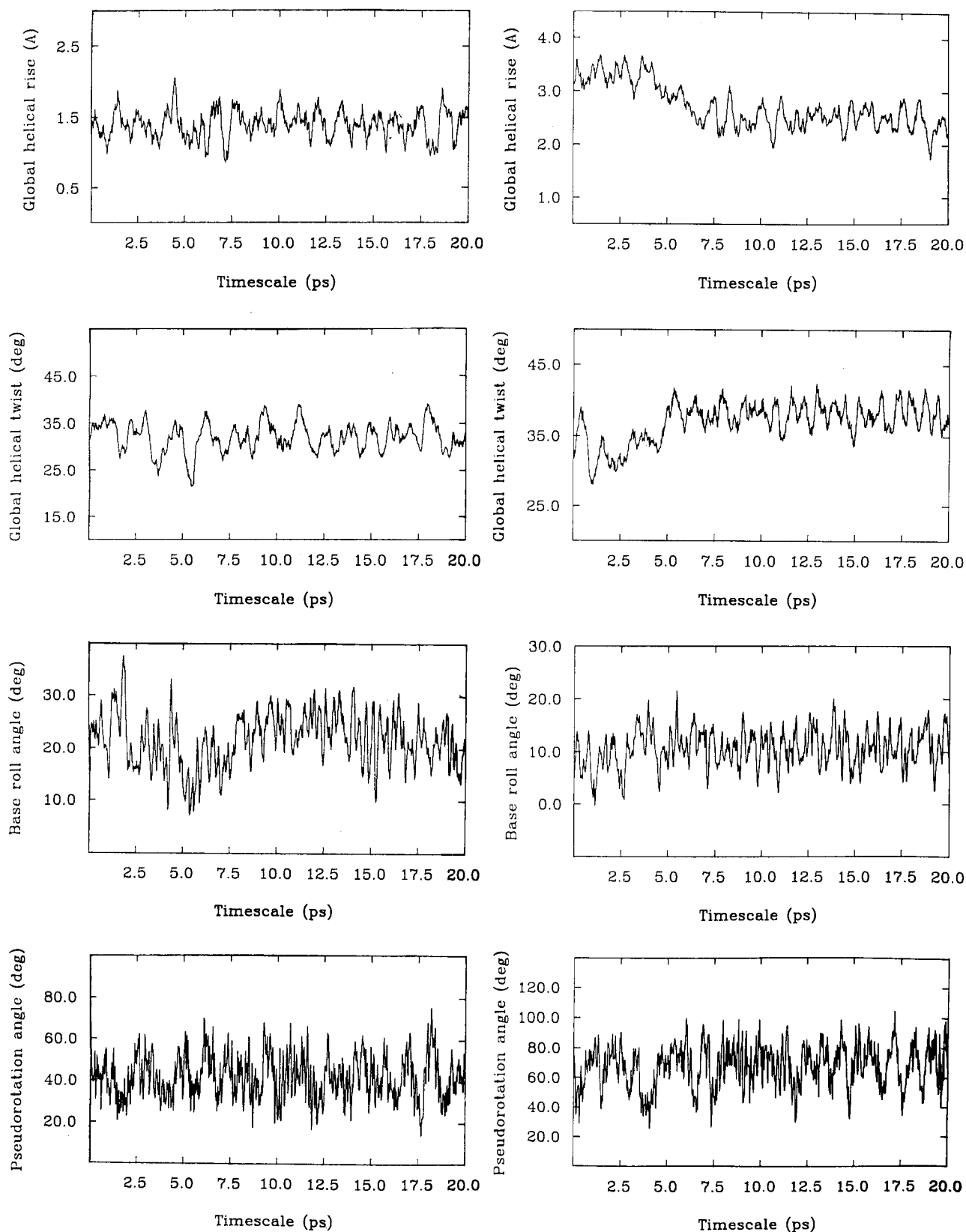


Fig. 5. Fluctuation of helical parameters of A-DNA and B-DNA forms of 5'GGATGGGAG:5'CTCCCATCC during 20-ps dynamics simulation; fluctuations of A-DNA parameters are shown columnwise in the *left panel* whereas fluctuations of B-DNA parameters are shown in the *right panel*, respectively

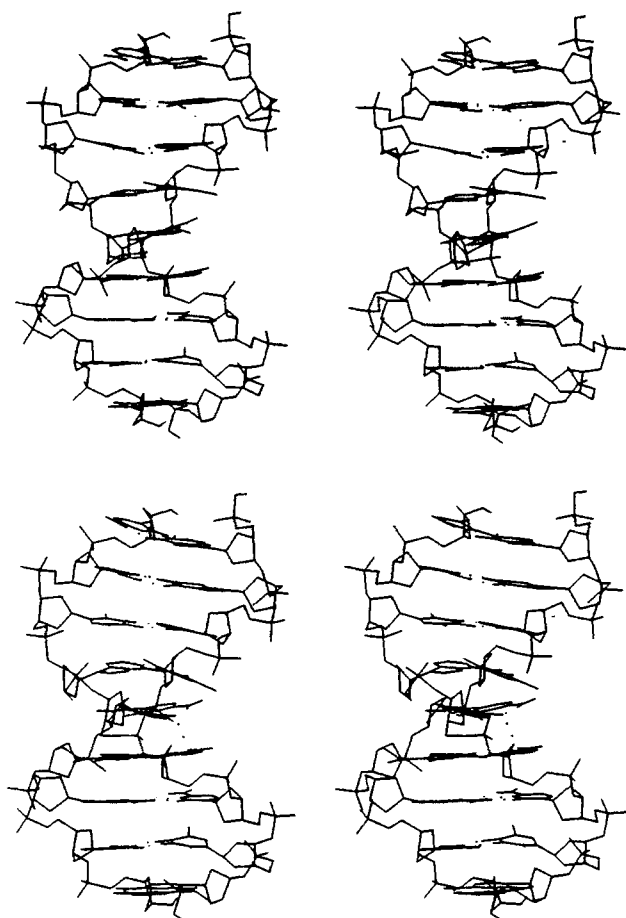


Fig. 6. Model B-DNA structure where the total potential energy of the optimized structure is $E = -601$ kcal/mol (upper panel) and the optimized average dynamics structure (lower panel) -609 kcal/mol, respectively

ture of the duplex was subjected to 2,000 steps of ABNR minimization followed by 20-ps dynamics simulation. The optimized and average dynamics structures of the DNA are presented in Fig. 6. The dynamics trajectory was analysed with the CORDAN program (Galat 1989b) in terms of geometrical parameters and some of these are summarized in Tables 5 and 6. The mean value of Ω in the average dynamics structure is higher than in the canonical B-DNA structure. Moreover, the 20 ps trajectory shows variable fluctuations of Ω for different base steps, i.e., for the A3-T16/T4-A15 base step $\Omega = 27.7^\circ$ ($f = 3.5^\circ$) whereas for the following base step $\Omega = 36.7^\circ$ ($f = 12.0^\circ$). The mean value of D_z was somewhat smaller than in the canonical B-DNA and showed considerable fluctuations for the T4-A15/G5-C14 base step $D_z = 2.50$ Å ($f = 1.38$ Å) whereas for the preceding base step $D_z = 3.63$ Å ($f = 0.21$ Å). This latter result illustrates that there are considerable fluctuations of geometry of T4 due to inter-strand purine clashes in the tri-

Table 5. Amplitude of pucker (τ_m), pseudorotation phase angle (P), δ torsion angle, and sugar conformation in the average dynamics structure of B-DNA model of 5'GGATGGGAG:5'CTCCCATCC

Sequence	τ_m	P	δ	Conformation
5'-G1	45.9	174.7	159.1	C2'-endo
G2	42.8	188.6	159.8	C3'-exo
A3	39.9	194.9	159.4	C3'-exo
T4	48.7	57.1	68.3	C4'-exo
G5	46.4	166.8	154.7	C2'-endo
G6	40.8	185.9	156.6	C2'-endo
G7	41.2	190.6	157.2	C2'-endo
A8	44.0	180.3	160.2	C2'-endo
G9	42.9	178.9	158.1	C2'-endo
5'-C10	42.3	185.1	161.1	C2'-endo
T11	51.6	44.3	64.8	C4'-exo
C12	42.0	182.9	156.5	C2'-endo
C13	43.3	190.9	162.5	C3'-exo
C14	55.0	39.3	60.1	C4'-exo
A15	50.9	40.3	61.2	C4'-exo
T16	42.2	182.6	156.3	C2'-endo
C17	40.6	187.2	158.2	C2'-endo
C18	42.9	189.5	161.7	C2'-endo

Table 6. Geometrical parameters of base pair steps and base pairs in the average dynamics structures of model B-DNA 5'GGATGGGAG:5'CTCCCATCC

B/P	Step	ϱ	Ω	D_z	D_y	ω
G1-C18	1	5.1	42.4	3.13	-0.22	-15.0
G2-C17	2	-0.2	42.2	2.98	0.52	-7.3
A3-T16	3	-1.1	30.1	3.63	-0.42	-11.7
T4-A15	4	9.9	25.1	3.82	0.57	-10.5
G5-C14	5	-6.5	43.2	2.24	-0.01	-17.6
G6-C13	6	2.0	43.0	2.86	-0.08	-10.3
G7-C12	7	2.7	39.0	3.58	-0.71	-8.9
A8-T11	8	-3.1	39.2	2.77	-0.24	-13.0
G9-C10						-5.3
Mean value		1.1	38.0	3.13	1.95	-11.1

See footnote to Table 4

angle A3-A15-G4 (see Fig. 5). The base-pairs of the $G_6G_7A_8G_9:C_{10}T_{11}C_{12}C_{13}$ part of the duplex showed substantial displacement from the imaginary helix axis which suggests that this part of the oligomer has tendency to drift into an A-DNA structure. The T4-A15/G5-C14 base step shows the high average base roll angle ($\omega = 17.9^\circ$, $f = 9.0^\circ$). The average phosphorus atom separation was maintained at 6.7 Å, and 4 out of 18 sugar units were in a sugar pucker other than the

C2'-endo and C3'-exo. This fact can account for the observation (Aboul-ela et al. 1988) that about 20% of sugar units are in the C3'-endo conformation in the duplex dissolved in aqueous solution at room temperature.

Discussion

5'GGATGGGAG:5'CTCCCATCC nonamer duplex constitutes a part of the internal control region in the *Xenopus* 5S RNA gene which binds transcription factor IIIA (Rhodes and Klug 1986). This part of DNA showed anomalous cleavage by DNaseI (Rhodes and Klug 1986), which has been rationalized in terms of non-B-DNA structure of some segments of the gene (possible A-DNA-like structure). Moreover, there have been suggestions that certain sequences show a tendency to adopt a structure with A-DNA elements and this may constitute an important factor in formation of nucleosomes (Calladine and Drew 1986).

Although in the solid state the duplex exists in the A-DNA structure (McCall et al. 1986), it has a number of geometrical elements which are not typical for the A-DNA helix. For example, the sugar units of G7, A8, C10, C13, C17 and G18 occupy the C1'-exo or C2'-endo conformations instead of the C3'-endo, as in the canonical A-DNA (McCall et al. 1986). However, B-DNA sugar pucker was also observed in the central four base pairs in a self-complementary duplex d(GGCCGGCC) which forms the A-DNA structure in the crystal state (Wang et al. 1982). The base roll angles for the base steps G5-C14/G6-C13 and A8-T11/G9-C10 in the nonamer have negative values which is another departure from the canonical A-DNA structure. Negative base roll angles, however, were observed in a number of base steps in the modified A-DNA crystal structure of d(GCCCGGGC) (Heinemann et al. 1988). Thus, these facts show that A-DNA structure can adopt a number of helical parameters typical of B-DNA.

In the crystal structure of the nonamer a number of base steps show considerable sliding, which can directly account for diminished exciton interaction between bases in comparison to canonical A-DNA. The Cotton effects of DNA are strongly dependent on the base overlap but the latter depends exclusively on the sequence which also determines the structure of the oligomer in solution. Superposition of the Cotton effects of the single strands does not reproduce the CD spectrum of the duplex in terms of positions and width of the CD bands. However, it does reproduce it in terms of the magnitude of the Cotton effects. This suggests that base-base overlap causes enlargements of band width and shift of the CD maxima but it does not affect significantly the magnitudes of the CD bands.

Twenty ps dynamics simulations did not provide a unified structure for the oligonucleotide, but the B-DNA model structure showed a tendency to adopt certain helical parameters typical of A-DNA, whereas the A-DNA structure had a tendency to adopt helical parameters typical of the B-DNA form. Substantially longer dynamics simulations of solvated duplex with the inclusion of explicit salt content might lead both structures into one unified structure. However, such a task would require one to overcome substantial difficulties in simulation of complex solvent-solute mixtures. However, a complete agreement cannot be expected between the helical parameters established in the X-ray structure (McCall et al. 1986) (usually 2.5 M salt concentration and low water activity), NMR-elucidated structure (Aboul-ela et al. 1988) and under model conditions in the vacuum simulation. Comparative results obtained from these techniques show that the DNA structure is determined globally by conditions of the phase and locally by the sequence.

The CD spectra of dodecamers show that these duplexes exist in the B-DNA form and their solution structures are similar to each other as are their crystal counterparts. There is higher variability of DNA structures in alternating GT-AT sequences (oligomers 7, 8A, 8B and 10). The octamer 8A exhibits CD spectra which are not typical of either the A-DNA or B-DNA helix. However, the heptamer shows CD spectra similar to 8A, and it is thus possible that these oligonucleotides can form a non-standard B-DNA helix in aqueous solution. The present CD study shows that not all of the oligonucleotides studied in aqueous solution can be assigned to the classical B-DNA family. This suggests that the structure of DNA oligomers is flexible in aqueous solution and can adopt a variety of conformations which lie between the canonical B-DNA and A-DNA structures (see Fairall et al. 1989). Such a DNA conformational spectrum is dependent on solvent, salt, temperature, sequence and DNA length. Also, one should be aware of the edge-effect artefacts in short DNA helices.

The alterations in DNA structure may be produced by synchronized changes of several parameters in the helix, i.e., unwinding of the helix by a few degrees with simultaneous sliding of base pairs, displacement of base pairs from the imaginary helix axis, and opening of most base steps towards the minor groove (positive base roll angle) must take place during the transition from B-DNA to A-DNA. Reverse motions to those given above should occur during a transition from A-DNA to B-DNA. Such transitions are time and temperature dependent and can be induced by solvents, salts and interacting molecules. Each of these conformational transitions causes alterations in the minor and major groove widths and their exposure to the solvent, which is the factor that may determine the

final structure of DNA (Conner et al. 1982). In fact, these helical movements partially occurred during short dynamics simulations of B-DNA and A-DNA forms of the nonamer which is a part of TFI_{II}A. This may suggest that the sequence of the nonamer clearly shows preference for the helix with a number of elements typical for an intermediate form between B-DNA and A-DNA. This tendency can be rationalized by the presence of an array of guanosine residues which are known to crystallize in structures related to A-DNA (McCall et al. 1985). Moreover, the DNA structure show sequence-dependent fluctuations of the helical parameters, which are influenced by the flanking sequences. Thus, the question as to whether the "original" crystal, or solution structure of short oligonucleotide segments remains unaltered if the segment constitutes a part of a larger DNA sequence, remain to be elucidated.

Acknowledgements. I would like to express my appreciation to Prof. Gerald D. Fasman for access to the Jobin Yvon Mark V dichrograph and to Prof. I. H. Goldberg and Prof. M. Karplus for their financial support. This work was supported in part by an award from the Merck Company.

References

- Aboul-ela F, Varani G, Walker GT, Tinoco I Jr (1988) The TFI_{II}A recognition fragment d(GGATGGGAG) · d(CTCCCATCC) is B-form in solutions. *Nucleic Acid Res* 16:3559–3579
- Adamiak RW, Galat A, Skalski B (1985) Salt- and solvent-dependent conformational transition of ribo-CGCGCG duplex. *Biochim Biophys Acta* 825:345–352
- Arnott S, Hukins DWL (1972) Optimised parameters for A-DNA and B-DNA. *Biochem Biophys Res Commun* 47:1504–1509
- Arnott S, Dover SD, Wonacott AJ (1969) Least-squares refinement of the crystal and molecular structures of DNA and RNA from X-ray data and standard bond lengths and angles. *Acta Crystallogr B* 25:2192–2206
- Benevides JM, Wang AHJ, Rich A, Kyogoku Y, van der Marel GA, van Boom JH, Thomas Jr GA (1986) Raman spectra of single crystals of r(CGCGd(CGC)) and d(CCCCGGG) as models of A-DNA, their structure transitions in aqueous solution and comparison with double-helical poly(dG) · poly(dC). *Biochemistry* 25:41–50
- Benevides JM, Wang AHJ, van der Marel GA, van Boom JH, Thomas Jr GA (1988) Crystal and solution structures of the B-DNA dodecamer of d(CGCAAATTTGCG) probed by Raman spectroscopy: heterogeneity in the crystal structure does not persist in the solution structure. *Biochemistry* 27:931–938
- Borer PN, Dengler B, Tinoco I Jr (1974) Stability of ribonucleic acid double-stranded helices. *J Mol Biol* 86:843–853
- Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J Comput Chem* 4:187–217
- Calladine CR, Drew HR (1986) Principles of sequence-dependent flexure of DNA. *J Mol Biol* 192:907–918
- Coll M, Frederick CA, Wang AHJ, Rich A (1987) A bifurcated hydrogen-bonded conformation in d(AT) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc Natl Acad Sci USA* 84:8385–8389
- Conner BN, Takano T, Tanaka S, Itakura K, Dickerson RE (1982) The molecular structure of d(CpCpGpG), a fragment of right-handed double-helical A-DNA. *Nature* 295:294–299
- Dickerson RE, Drew HR (1981) Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *J Mol Biol* 149:761–786
- Edwards EL, Ratliff RL, Gray DM (1988) Circular dichroism spectra of DNA oligomers show that short interior stretches of C-C⁺ base pairs do not form in duplexes with A-T base pairs. *Biochemistry* 27:5166–5174
- Fairall L, Martin S, Rhodes D (1989) The DNA binding site of the *Xenopus* transcription factor IIIA has a non-B-form structure. *EMBO J* 8:1809–1817
- Galat A (1986) Computer aided analysis of infrared, circular dichroism and absorbance spectra. *Comp Appl Biosci* 2:201–205
- Galat A (1989a) CORGEN: A Fortran-77 program for generation of standard and non-standard DNA helices. *Comp Appl Biosci* 5:279–286
- Galat A (1989b) Analysis of dynamics trajectories of DNA and drug-DNA complexes. *Comp Appl Biosci* 5:271–278
- Gunsteren WF van, Berendsen HJC, Geurtsen WF, Zwinderman HRJ (1986) A molecular dynamics computer simulation of an eight-base-pair DNA fragment in aqueous solution: comparison with experimental two-dimensional NMR data. *Ann NY Acad Sci* 482:287–303
- Heinemann U, Lauble H, Frank R, Blocker H (1988) Crystal structure analysis of an A-DNA fragment at 1.8 Å resolution: d(GCCCGGGC). *Nucleic Acid Res* 15:9531–9550
- Jain S, Zon G, Sundaralingam M (1987) The potentially Z-DNA forming sequence d(GTGTACAC) crystallizes as A-DNA. *J Mol Biol* 197:141–145
- Johnson Jr WC (1985) Circular dichroism and its empirical application to biopolymers. *Methods Biochem Anal* 31:61–163
- Leslie AGW, Arnott S, Chandrasekaran R, Ratliff RL (1980) Polymorphism of DNA double helices. *J Mol Biol* 143:49–72
- Lindsey SM, Lee SA, Powell JW, Weidlich T, Demarco C, Lewen GD, Tao NJ, Rupprecht A (1988) The origin of the A to B transition in DNA fibers and films. *Biopolymers* 27:1015–1043
- Lo MK, Jones SS, Hackett NR, Khorana HG (1984) Specific amino acid substitution in bacteriorhodopsin: replacement of a restriction fragment in the structural gene by synthetic DNA fragment containing altered codons. *Proc Natl Acad Sci USA* 81:2285–2289
- Marky LA, Blumenfeld KS, Kozlowsky S, Breslauer KJ (1983) Salt-dependent conformational transitions in the self-complementary deoxydodecanucleotide d(CGCGAATTCGCG): evidence for hairpin formation. *Biopolymers* 22:1247–1257
- McCall M, Brown T, Kennard O (1985) The crystal structure of d(G-G-G-G-C-C-C-C). A model for poly(dG) · poly(dC). *J Mol Biol* 183:385–396
- McCall M, Brown T, Hunter W, Kennard O (1986) The crystal structure of d(GGATGGGAG) forms an essential part of the binding site from transcription factor IIIA. *Nature* 322:661–664
- McClarín JA, Frederick CA, Wang B-C, Greene P, Boyer WH, Grable J, Rosenberg J (1986) Structure of the DNA-Eco RI endonuclease recognition complex at 3 Å resolution. *Science* 234:1526–1541
- Murray V, Martin RF (1985) Comparison of the sequence specificity of bleomycin cleavage in two slightly different DNA sequences. *Nucleic Acid Res* 13:1467–1481
- Nadeau JG, Gilham PT (1985) Anomalous hairpin formation in an oligodeoxyribonucleotide. *Nucleic Acid Res* 13:8259–8274

- Nelson HCM, Finch JT, Luisi BF, Klug A (1987) The structure of an oligo(dA) · oligo(dT) tract and its biological implication. *Nature* 330:221–226
- Nilsson L, Clore GM, Gronenborn AM, Brünger AT, Karplus M (1986) Structure refinement of oligonucleotides by molecular dynamics with nuclear Overhauser effect interproton distance restraints: application to 5'd(CGTACG)₂. *J Mol Biol* 188:455–475
- Nilges M, Clore GM, Gronenborn AM, Brünger AT, Karplus M, Nilsson L (1987) Refinement of the solution structure of the DNA hexamer 5'd(GCATGC)₂: combined use of nuclear magnetic resonance and restrained molecular dynamics. *Biochemistry* 26:3718–3733
- Patel DJ, Shapiro L, Hare D (1987) DNA and RNA: NMR studies of conformations and dynamics in solution. *Q Rev Biophys* 20:35–112
- Powell MJD (1965) A method for minimizing a sum of squares of non-linear function of several variables without calculating derivatives. *Comput J* 7:303–307
- Rhodes D, Klug A (1986) An underlying repeat in some transcriptional control sequences corresponding to half a double helical turn of DNA. *Cell* 46:123–132
- Shakked Z, Rabinovich D (1986) The effect of the base sequence on the fine structure of the DNA double helix. *Prog Biophys Mol Biol* 47:159–195
- Sinha ND, Biernat J, McManus J, Koster H (1984) Polymer support oligonucleotide synthesis XVIII: use of β -cyanoethyl- N,N-dialkylamino-/N-morpholino phosphoramidate of deoxynucleosides for the synthesis of DNA fragments simplifying deprotection and isolation of the final product. *Nucleic Acid Res* 12:4539–4557
- Thomas GA, Kubasek WL, Peticolas WL, Greene P, Grable J, Rosenberg JM (1989) Environmentally induced conformational changes in B-type DNA: comparison of the conformation of the oligonucleotide d(TCGCGAATTCGCG) in solution and its crystalline complex with the restriction nuclease EcoRI. *Biochemistry* 28:2001–2009
- Tidor B, Irikura KK, Brooks BR, Karplus M (1983) Dynamics of DNA oligomers. *J Biomol Struct Dyn* 1:231–252
- Wang AHJ, Fuji S, van Boom JH, Rich A (1982) Molecular structure of the octamer d(G-G-C-C-G-G-C-C): modified A-DNA. *Proc Natl Acad Sci USA* 79:3968–3972
- Wang AHJ, Ughetto G, Quigley GJ, Rich A (1987) Interaction between an anthracycline antibiotic and DNA: molecular structure of daunomycin complexed to d(CpGpTpApCpG) at 1.2 Å resolution. *Biochemistry* 26:1152–1163
- Wang Y, Thomas GA, Peticolas WL (1989) A duplex of the oligonucleotides d(GGGGGTTTTT) and d(AAAAACCCCC) forms an A to B conformational junction in concentrated salt solutions. *J Biomol Struct Dyn* 6:1177–1187
- Wüthrich K (1986) *NMR of proteins and nucleic acids*. Wiley, New York