Solution Structure of the CpG Containing d(CTTCGAAG)$_2$ Oligonucleotide: NMR Data and Energy Calculations Are Compatible with a BI/BII Equilibrium at CpG[†]

A. Lefebvre,[‡] O. Mauffret,[‡] E. Lescot,[‡] B. Hartmann,[§] and S. Fermandjian*,[‡]

Département de Biologie Structurale, URA 147 CNRS, Institut Gustave Roussy, P.R.2, 39 rue C. Desmoulins, F-94805 Villejuif Cedex, France, and Laboratoire de Biochimie Théorique, URA 77 CNRS, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, F-75005 Paris, France

Received March 14, 1996; Revised Manuscript Received July 8, 1996[⊗]

ABSTRACT: We report the analysis of the solution structure of the DNA duplex d(CTTCGAAG)$_2$ compared to that of d(CATCGATG)$_2$, the two oligonucleotides being related by the permutation of residues 2 and 7. An earlier study has demonstrated the malleability of CpG in the tetrad TCGA of d(CATCGATG)$_2$ [Lefebvre et al. (1995) *Biochemistry 34*, 12019−12028]. Conformations of d(CTTCGAAG)$_2$ were evaluated by (a) two-dimensional NMR, including proton and phosphorus experiments, (b) adiabatic mapping of the conformational space, (c) restrained molecular mechanics undertaken with sugar phase angle, $\epsilon-\zeta$ difference angle, and NOE distances as input, and (d) back-calculation-refinement against NOE spectra at various mixing times. d(CTTCGAAG)$_2$ like d(CATCGATG)$_2$ exhibits a B-DNA conformation. However, significant differences are noted between the two oligonucleotides, extending up to the central CpG step, although this step resides in the same TCGA tetrad in both sequences. In structures obtained with refined NMR data, CpG adopts, for instance, a greater twist and a higher guanine phase within d(CTTCGAAG)$_2$ compared to d(CATCGATG)$_2$. In the former oligonucleotide, the structure of CpG resembles strikingly that found in the ACGT tetrad of the cAMP responsive element [Mauffret et al. (1992) *J. Mol. Biol. 227*, 852−875]. Moreover, two conformers with CpG either in the BII state (ϵ, ζ = g$^-$, t) or in the BI state (ϵ, ζ = t, g$^-$) are found equally stable for d(CTTCGAAG)$_2$. The energy barrier from BI to BII comes to only 5.7 kcal/mol, and the path of the transition is very short. When calculations on d(CTTCGAAG)$_2$ are performed taking the BI/BII equilibrium into account, the agreement with both the ^{1}H and ^{31}P data is found better than in the case with a single conformation taken alone. The BI/BII equilibrium may also occur in d(CATCGATG)$_2$, but the amount of BII conformer is now found weaker compared to its analogue. The ability of the CpG phosphate groups to adopt the BII conformation could provide a satisfying explanation for the high mutation rates observed at these sites.

The CpG oligonucleotide is known for its high propensity to undergo mutations, particularly transitions to TpG or CpA, in species with methyltransferases (for example, vertebrates) or without (for example, drosophila). In vertebrates, CpG is underrepresented by a factor of 5 compared to statistical expectations (Grippo et al., 1968; Bird et al., 1980), and mutations at such sites are implicated in about one-third of genetic diseases (Rideout et al., 1990; Cooper & Youssoufian, 1988). For example, for patients with colon cancer, 49% of mutations found on the P53 gene occur on CpGs, and four of the five codons which are mutation hotspots on this gene are CpG sites (Hollstein et al., 1996). The usual explanation for such hotspots is that in the human genome about 90% of the CpG cytosines are methylated on position 5 (Bird, 1978), and that the resulting 5-Me-cytosine can easily mutate into thymine through oxidative deamination (Coulondre et al., 1978).

Yet, the particular structural properties exhibited by CpG in DNA could also play a role in this CpG suppression.

To date, numerous structural studies have demonstrated that CpG is able to adopt various conformations, in response to sequence context (Baïkalov et al., 1993; Grzeskowiak et al., 1991; Mauffret et al., 1992; Lefebvre et al., 1995b; Privé et al., 1991; Yanagi et al., 1991). For example, when present in the ACGT tetrad, the CpG step adopts a high twist (≈42°), and a guanine sugar with a phase of about 180°, while in the TCGA tetrad it instead adopts a low or medium twist (30° to 35°) with a guanine phase of about 160° (Lefebvre et al., 1995b; Grzeskowiak et al., 1991; Baïkalov et al., 1993).

The CTTCGAAG palindromic sequence that contains a TCGA tetrad in its center appears of particular interest, as it has been identified as a mutation hotspot in several genes. For example, it contains one of the main hotspots responsible for type A hemophilia on the factor VIII gene (Tuddenham et al., 1994). It is also the main UV-induced mutation hotspot in the suppressor tRNA gene *supF* (24% of all mutations) (Parris et al., 1994). More recently, Levy et al. (1996) have demonstrated that if the place of this sequence is changed within the *sup*F gene, the UV-induced mutation rate remains exactly the same as at the parent site, underlining the very particular behavior of this sequence with respect to mutations.

The problem of sequence context on the CpG structure was investigated through comparison of the d(CTTCGAAG)$_2$ oligonucleotide [denoted (CT)] with the related d(CATCGATG)$_2$ oligonucleotide [denoted (CA)], one being obtained

[†] This work was supported by the Association for International Cancer Research (St. Andrews University, U.K.). A.L. has a fellowship from the Institut de Formation Supérieure Biomédicale.

* To whom correspondence should be addressed. Telephone: 1 45 59 49 85. Fax: 1 46 78 41 20. E-mail: sfermand@igr.fr.

[‡] Institut Gustave Roussy.

[§] Institut de Biologie Physico-Chimique.

[⊗] Abstract published in *Advance ACS Abstracts,* September 1, 1996.

S0006-2960(96)00629-0 CCC: $12.00

from the other by the permutation of second and seventh residues. Results revealed the surprising structural behavior of the oligonucleotide (CT) which presents TT and AA repeats on the 5′ and 3′ sides, respectively, of CpG (El Antri et al., 1993a). Although the (CT) and (CA) oligonucleotides have exactly the same base composition, the (CT) duplex was found much less stable than its counterpart (CA). Their circular dichroism spectra are also different: they resemble for (CA) the spectra of $(dA\text{-}dT)_n$ oligonucleotides, and for (CT) the spectra of $(dC\text{-}dG)_n$ oligonucleotides (El Antri et al., 1993a). Moreover, the central CpG dinucleotide displays a markedly low-field ^{31}P resonance (−3.9 ppm) at 25 °C for (CT), contrasting with that more typical (−4.05 ppm) observed for (CA). The temperature dependence of this ^{31}P resonance is also interesting: only in the case of (CT) it manifests an upfield shift from 30 °C to 55 °C, accompanied by a severe line broadening. Above 55 °C, it joins the cluster of the other phosphorus resonances, and moves downfield (El Antri et al., 1993a). Such a behavior is also found for the ^{31}P resonances of mismatched base pairs (Nikonowicz & Gorenstein, 1990; Gorenstein, 1994) or extrahelical bases (Nikonowicz et al., 1989), but is not displayed by (CA).

The main question of this study was whether the CpG step of (CT) presents a single BI (ϵ, ζ = t, g^-) conformation characterized by a high $\epsilon-\zeta$ value (around −30°) or a BI/BII equilibrium with a large proportion of BII conformation (ϵ, $\zeta = g^-$, t), as shown by mismatches (Roongta et al., 1990; Gorenstein, 1994). In a first step, using JUMNA energy minimization with refined nuclear magnetic resonance (NMR)[1] constraints, we determined a (CT) structure whose CpG backbone has a BI conformation ($\epsilon-\zeta = -20°$). This (CT) structure strongly differs from the (CA) structure previously reported in Lefebvre et al., (1995b), and is consistent with the CD and NMR data (El Antri et al., 1993a). Second, we assessed the relevance of a BI/BII equilibrium at the CpG step in both (CA) and (CT). Energetic mapping of a BI to BII transition showed that such an equilibrium is more favorable in (CT) than in (CA), in terms of energy, path length, and agreement with NMR data. The latter include NOE volumes and ^{31}P chemical shifts which were entirely compatible with a BI/BII equilibrium in (CT). For (CA), a mixture of three different conformers including BI and BII better reproduced the NMR data.

MATERIALS AND METHODS

Sample Preparation. The $d(CTTCGAAG)_2$ octamer duplex (CT) was synthesized using the solid-phase phosphoramidite procedure on an Applied Biosystem 381 B automatic apparatus. It was purified by reversed-phase HPLC using a Waters μ-Bondapack C18 column, followed by dialysis in water. After lyophilization, the oligonucleotide was dissolved at ≈3.5 mM concentration in a phosphate buffer containing 0.2 mM EDTA at pH 7, ionic strength 0.1. Sample concentration was calculated using 8500 $M^{-1}\cdot cm^{-1}\cdot base^{-1}$ as the extinction coefficient at 260 nm, according to Fasman (1975).

NMR Spectroscopy. Acquisitions were all obtained at 25 °C. The ^1H NMR experiments were acquired at 500 MHz on an AMX-500 Brüker spectrometer, in the phase-sensitive mode using the TPPI technique (Marion & Wüthrich, 1983), with a spectral width of 5000 Hz and a relaxation delay of 2 s. Data were processed with the Felix software (Biosym Inc.) on a Silicon Graphics Indigo R4000 workstation.

The quantitative 2D NOESY experiments in ^2H_2O were recorded at four different mixing times (50, 100, 200, and 300 ms) in a single span of 4 days without removing the sample from the probe. A total of 512 experiments were performed with 1024 complex points for each FID. The reference phase of the transmitter was adjusted to give pure absorption spectra with a very flat base line. NOESY data sets were multiplied with shifted sine and sine-squared bells in dimensions 1 and 2, respectively, and zero-filled in both dimensions to lead to a final 2K × 2K real matrix. Base line correction was performed using a convolution method (Dietrich et al., 1991). Translation of experimental NOE intensities into distances was realized with the distance extrapolation method, in order to take the spin diffusion effects partially into account (Baleja et al., 1990a,b; Fedoroff et al., 1994; Mauffret et al., 1992). The intranucleotide H1′−H2′ distances were taken as reference distances.

Pure phase absorption double-quantum-filtered COSY spectra were recorded using 4096 complex points in the acquisition dimension, and 800 experiments in the t1 dimension. Data were processed with sine bells 60° shifted in the t2 dimension and 30° shifted in the t1 dimension. The spectral resolution was enhanced by zero-filling to produce a 4K × 2K real matrix. DQF-COSY cross-peak simulations were done with SPHINX and LINSHA programs (Widmer & Wüthrich, 1987), as described elsewhere (Mauffret et al., 1992).

Clean TOCSY experiments were performed with standard methods, with a spin lock pulse consisting of a MLEV17 pulse train surrounded by trim pulses of 1 ms (Bax & Davies, 1985). Mixing times ranged from 50 to 150 ms; 400 experiments with 2048 points for each FID were recorded.

Simulation Methodology. Determination of the 3D structures was carried out through molecular modeling using the JUMNA algorithm (Lavery & Sklenar, 1988; Lavery, 1988; Lavery et al., 1995; Sanghani et al., 1995). JUMNA is able to handle sugar puckering, torsion angles, and distance constraints, with respect to a fixed value or with specified upper and lower bounds (Mauffret et al., 1992; Lefebvre et al., 1995b). Constraint violations are prevented by a quadratic penalty energy term, with a force constant of 6 or 12 $kcal\cdot mol^{-1}\cdot A^{-2}$ for distance constraints and of 1000 $kcal\cdot mol^{-1}\cdot deg^{-2}$ for angular constraints.

BI to BII transitions were induced by imposing chosen values for the $\epsilon-\zeta$ dihedral difference of the studied dinucleotide junction *via* harmonic constraints. The force constant was the same as the one used for other dihedral constraints: 1000 $kcal\cdot mol^{-1}\cdot deg^{-2}$. $\epsilon-\zeta$ was varied from −100° to +200° by steps of 5°. In the same way, sugar transitions were provoked by varying the pseudorotation angle constraint from 130° to 200° by steps of 6°. A scan of the guanine sugar phase of the infinite oligonucleotide $(TCGA)_n$ led to one minimum at a phase value of about 160° accompanied by a shoulder at a phase value of 180° (Lefebvre et al., 1995b). A scan of the CpG guanine phase angle of the complete $d(CTTCGAAG)_2$ oligonucleotide also gave one single minimum, at a phase value of 180°, as well as a shoulder at a phase value of about 160°. Scans of the

[1] Abbreviations: NMR, nuclear magnetic resonance; COSY, 2D homonuclear shift correlated spectroscopy; NOE, nuclear Overhauser effect; NOESY, 2D homonuclear nuclear Overhauser effect correlated spectroscopy; CD, circular dichroism; T_m mixing time.

phase difference P(G)−P(C) produced exactly the same minima as scans of the guanine phase alone.

The 3D structures were analyzed with the CURVES program (Lavery & Sklenar, 1988, 1989) which provides a rigorous way to obtain local structural parameters and the overall helical axis locus for irregular structures of nucleic acids. Due to substantial fraying, the terminal base pairs do not present a fixed conformation. Thus, they were not taken into account either for the calculation of the global axis or for rms deviation calculations.

Molecules were visualized with Insight II software (Biosym Technologies Inc.).

Initial Structures. In agreement with the protocol previously described (Lefebvre et al., 1995b; Sodano et al., 1995), we used several experimental structures, such as crystal and NMR structures, as initial structures: the B73 structure (Arnott & Hüskin, 1973), two other crystal structures containing the TCGA sequence, as well as the structures calculated for d(CATCGATG)$_2$ and d(GTACGTAC)$_2$ oligonucleotides (Lefebvre et al., 1995b). The four different minima resulting from adiabatic mapping (see Simulation Methodology) were also used as starting structures.

These structures were all in the B-DNA domain with sugars in the wide C2′-endo range, but presented a wide range of helicoidal parameters. For example, CpG twist varied from 30° to 42°, and central guanine sugar phase ranged from 150° to 180°.

Constraints on the Conformation around the C3′−O3′ and O3′−P Backbone Bonds. $^3J_{H3'-P}$ coupling constants were measured through proton-detected heterocorrelation experiments with selective pulse on the H3′ protons (Sklenar & Bax, 1987), as described in a previous paper (Lefebvre et al., 1995a). Obtained values are also reported in the latter study. Related torsional angles ϵ and ζ can be deduced from the $^3J_{H3'-P}$ coupling constants through the relationship $J = 15.3 \cos^2 \theta - 6.1 \cos \theta + 1.6$ (Lankhorst et al., 1984). As the phosphorus line width was about 6 Hz, coupling constants of the order of 4 Hz were overestimated by 0.5−1 Hz (Neuhaus et al., 1985). This led us to diminish the lower bond of the ϵ constraints by about 5° in previous studies (Mauffret et al., 1992; Lefebvre et al., 1995b). In this latter paper, we have observed a good correlation between the calculated $\epsilon-\zeta$ values and the experimental ^{31}P chemical shifts in various oligomers: $\epsilon-\zeta = 254.5 + (72.8)\delta_{^{31}P}$, where the phosphorus chemical shifts are referenced to trimethyl phosphate at 0 ppm (Lefebvre et al., 1995b). Using either $\epsilon-\zeta$ derived from the phosphorus chemical shifts or ϵ estimated from the $^3J_{H3'-P}$ coupling constants led to exactly the same structure. It appears that the use of the $\epsilon-\zeta$ constraints provided by the phosphorus chemical shifts is an alternative to avoid the uncertainty due to line width and also allows us in turn to obtain a lower final energy, since constraints on angle differences are less stringent than constraints on a single torsion angle.

Back-Calculation of Theoretical NOESY Spectra and Iterative Refinement. Back-calculations of theoretical NOESY spectra from 3D structures were performed in order to refine the agreement with experimental NOE data (Boelens et al., 1989; Kim & Reid, 1992a,b; Lefebvre et al., 1995b). NOESY spectra back-calculations were performed with the corresponding subroutine of the IRMA program, integrated in the Insight II environment (Biosym Inc.). All spin diffusion pathways are taken into account by a complete relaxation matrix calculation (Boelens et al., 1989). A single interproton correlation time of 3.2 ns was used. This parameter as well as the Z-leakage factor was evaluated with the procedure described by Banks et al. (1989). Agreement with experimental NOESY data was evaluated through the classical R factor: $R = [\sum|V_{exp} - V_{calc}]/\sum V_{exp}$; as well as through a derived $R_{1/6}$ factor, that is better adapted to NMR data: $R_{1/6} = [\sum|V_{exp}{}^{1/6} - V_{calc}{}^{1/6}|]/\sum(V_{exp}{}^{1/6})$, where summations run over all peaks of interest and all mixing times. Individual R factors for a chosen peak and a selected mixing time were also used. The iterative refinement of distance constraints was then performed by comparing experimental and back-calculated NOE intensities and adjusting distance constraints to $d = d_{mol}(V_{exp}/V_{calc})^{1/6}$ as described in Lefebvre et al. (1995b).

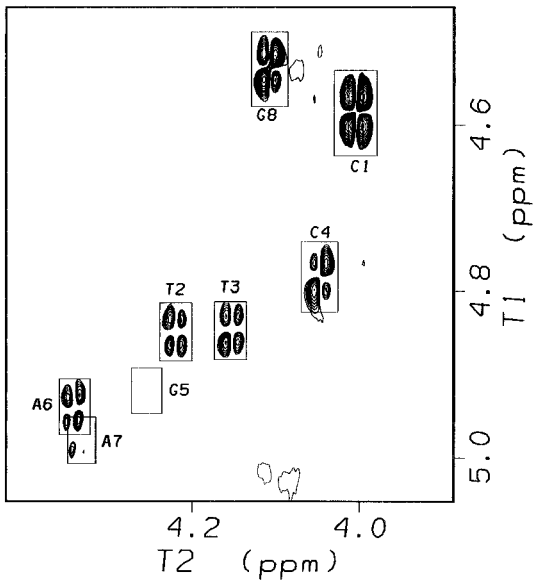

FIGURE 1: H3′−H4′ region of a 500 MHz DQF-COSY spectrum of d(CTTCGAAG)$_2$ (CT) in ^2H_2O/phosphate buffer, pH 7, at 25 °C.

RESULTS AND DISCUSSION

Sugar Ring Conformation. We used a combination of intrasugar coupling constants and H1′−H4′ distances to assess the pseudorotation angle values. For the d(CTTCGAAG)$_2$ (CT) oligonucleotide, the intensities of the H3′−H4′ COSY cross-peaks were found medium for the pyrimidines T2, T3, and C4 and for the purine A6, while that for A7 was very weak and that for G5 even nonexistent (Figure 1). These were consistent with the intensities of the H1′−H4′ cross-peaks observed in the clean TOCSY spectrum (Kim & Reid, 1992a) (not shown). Evaluation of the H1′−H4′ distances *via* the distance extrapolation method (Mauffret et al., 1992), and simulation of the H1′−H2′/H2″ and H3′−H2′/H2″ DQF-COSY cross-peaks with the SPHINX and LINSHA programs (Rinkel & Altona, 1987; Widmer & Wüthrich, 1987; Gochin et al., 1990; Mauffret et al., 1992), provided more precise pseudorotation angle evaluation (Table 1).

Interproton Distances. We examined particularly the H6/H8/H5 to H1′/H2′/H2″/H3′ intra- and internucleotide distances. Although NMR experimental data strongly suggest that the H6/H8 to preceding (5′) H2″ distances could provide a good mean to measure the structural effects occurring within dinucleotide steps (Fedoroff et al., 1994), these were

Table 1: Pseudorotation Angle and $\epsilon-\zeta$ Constraints Used in the Structure Minimization of d(CTTCGAAG)$_2$ (CT)

residue	phases (deg)	step	$\epsilon-\zeta$ (deg)	
T2	145−165	Tp_2T	−84	−54
T3	135−155	Tp_3C	−71	−41
C4	135−155	Cp_4G	−44	−14
G5	160−180	Gp_5A	−58	−28
A6	145−165	Ap_6A	−70	−40
A7	145−165			

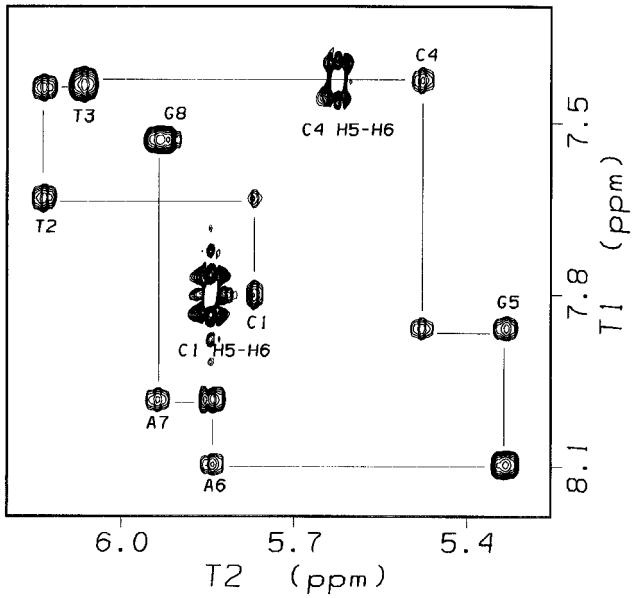

FIGURE 2: H6/H8−H1′/H5 region of a 500 MHz NOESY spectrum of d(CTTCGAAG)$_2$ (CT) in ^2H_2O/phosphate buffer, pH 7, at 25 °C, with a mixing time of 100 ms.

not taken into account in present calculations as their use led to high energy loss and strong structure distortions as related in previous reports (Lefebvre et al., 1995b; Mujeeb et al., 1992; Köning et al., 1991). The sugar conformations were constrained *via* their pseudorotation angles. The resulting intrasugar distances were found in good agreement with the NMR data; thus, the intrasugar distance constraints were not added.

Due to its oligonucleotide sequence, made of four pyrimidines followed by four purines, (CT) displayed NOESY spectra with strong peak overlaps. Thus, numerous interproton distances could not be determined, especially those implicating the H2′ and H2″ protons. Fortunately, the base and H1′ protons of residues C4 and G5 were well resolved (Figure 2), and this permitted us to collect an almost complete set of distances for the important central CpG step, including the C4 H1′/H2′/H2″/H3′/H6 to G5 H8 internucleotide distances. Among these, the H1′-to-H8 and H2″-to-H8 distances are in good agreement with the values published by Fedoroff et al. (1994) for oligonucleotides containing CpG.

ϵ and ζ Backbone Torsion Angles. The torsional angles derived from phosphorus chemical shifts were in good agreement with those determined from $^3J_{H3'-P}$ coupling constants (Lefebvre et al., 1995a). Constraints on the angle difference $\epsilon-\zeta$ deduced through the relationship $\epsilon-\zeta = 254.5 + (72.8)\delta_{^{31}P}$ (Lefebvre et al., 1995b) are reported in Table 1.

Minimization with the CpG Phosphate in the BI Conformation. Starting from a wide range of initial structures, we first carried out a free minimization without constraints. We obtained two different structures, named hereafter CTBI and CTBIb (Table 2), corresponding to the main minimum and the shoulder, respectively, of the central guanine phase scan on the d(CTTCGAAG)$_2$ oligonucleotide. Comparison of phase values was not sufficient to discriminate between the CTBI and CTBIb structures, as both were compatible with the ^{1}H NMR data. On the contrary, introduction of $\epsilon-\zeta$ constraints led to convergence (Table 2), because the CTS2 $\epsilon-\zeta$ value was too low with respect to the CpG $\epsilon-\zeta$ constraint derived from the ^{31}P chemical shift.

Table 2: Energy and R and $R_{1/6}$ Factors at the Different Steps of d(CTTCGAAG)$_2$ (CT) Modeling

molecule	constraints	energy (kcal/mol)	R	$R_{1/6}$
CTBI	−	−399.4	0.27	0.062
CTBIb	−	−393.7	0.25	0.060
CTS1	phases	−397.6	0.27	0.062
CTS2	phases	−393.7	0.26	0.059
CTSE	phases + ϵ	−397.3	0.27	0.062
CTSED	phases + ϵ + distances	−392.4	0.27	0.062
CTraf	phases + ϵ + refined distances	−393.9	0.22	0.054

With application of the distance constraints, subsequently submitted to iterative distance refinement, convergence was obtained after three cycles. Energies as well as R and $R_{1/6}$ factors of the successive structures are given in Table 2.

Minimization of the refined structure, CTraf, with releasing of all constraints again led to the main minimum (CTBI). NMR constraints induced an energy cost of 5 kcal/mol, and an improvement of only 5% of the R factor (0.008 for $R_{1/6}$). The minimized structure obtained without constraints (CTBI) was thus already in good agreement with the experimental data, and particularly the sugar puckers and interproton distances. This is illustrated in Figure 5 through the superposition of the CTBI and CTraf structures. Conformational parameters of the molecule CTraf are included in Figure 3.

Significant differences appeared for the central TCGA tetranucleotide, between the present structure of d(CTTCGAAG)$_2$ and that of d(CATCGATG)$_2$ previously reported (Lefebvre et al., 1995b). The value of the CpG $\epsilon-\zeta$ angle difference was found less negative in (CT) (−19°) compared to (CA) (−53°), in agreement with both the ^{31}P chemical shift and the $J_{H3'-P}$ values (Lefebvre et al., 1995a). Similarly, in (CT), the G5 sugar adopted a phase angle of 180° with an amplitude of 37°, compared to the phase angle of 165° and the amplitude of 41° observed in (CA). Transition of the G5 sugar pucker was associated with both an increase in the amplitude of the preceding C4 residue [46° in (CT) *vs* 40° in (CA)] and a change of the T3 residue which exhibited a lower amplitude [33° in (CT) *vs* 38° in (CA)] and a greater phase [155° in (CT) *vs* 139° in (CA)]. However, the other backbone torsion angles remained quite classical in (CT): 57° ± 2° for γ, −66° ± 2° for α, and 172° ± 3° for β, although some small differences in β and γ angles were noted at the central CpG step between (CA) and (CT).

Some striking differences were also observed among the helicoidal parameters. Permutation of the second and the seventh base pairs strongly changed the general structural profile, and while (CA) presented homogeneous parameters along its double helix, leading to a so-called "flat" profile,

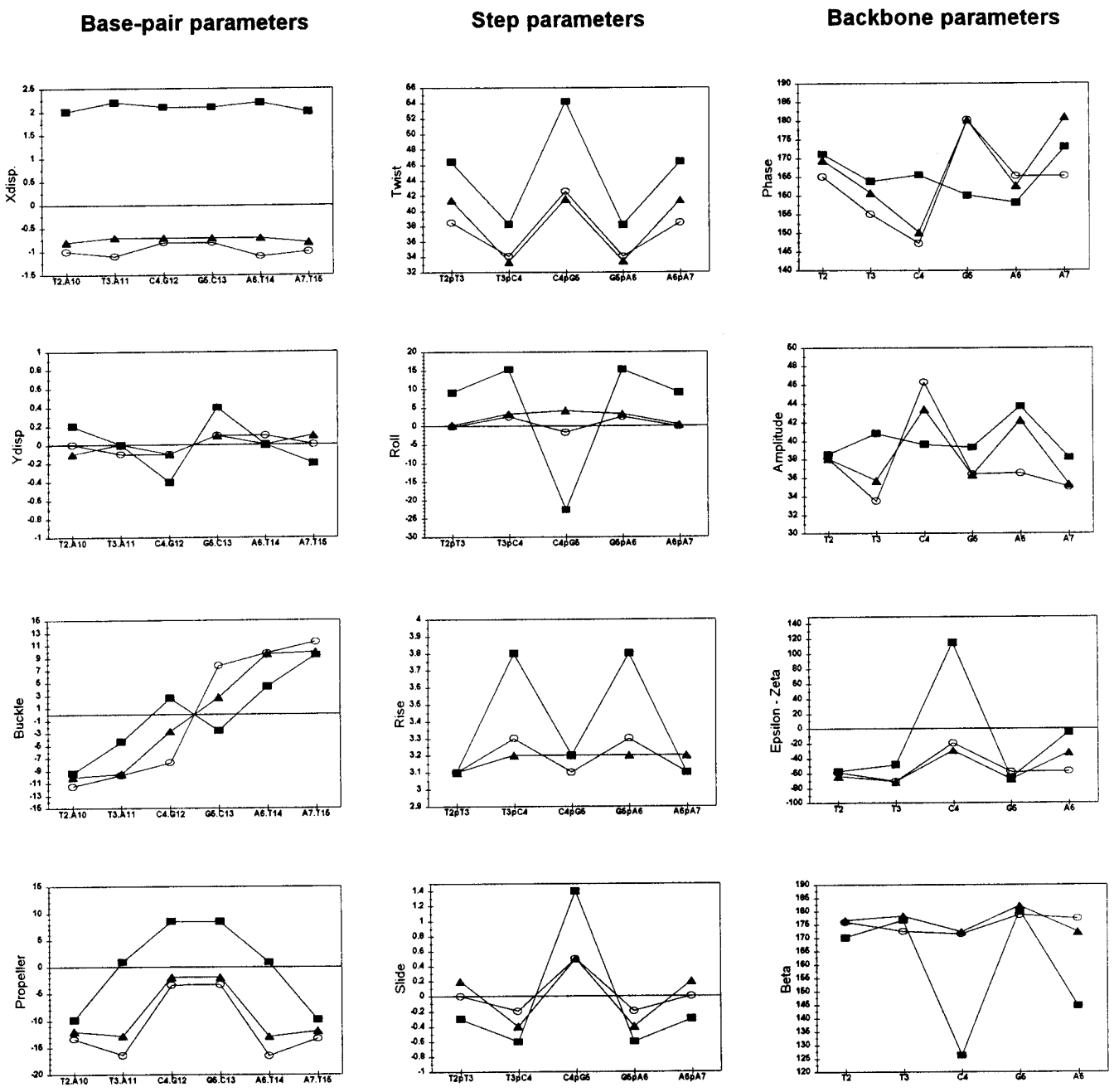

FIGURE 3: Conformational parameters for the CTraf (○), CTBI (▲), and CTBII (■) structures. Translational parameters are in angstroms, and rotational parameters are in degrees.

(CT), on the contrary, displayed an alternating profile. In the latter oligomer, for example, CpG adopted a high twist (43°) (correlated with a large backbone $\epsilon-\zeta$ value), bracketed by two low twist values (34°) in the TCGA tetrad. This alternating profile was also obvious for other parameters, such as the slide (−0.2 Å, +0.5 Å, −0.2 Å), the roll (2.5°, −2°, 2.5°), or the propeller twist (−16°, −3°, −3°, −16°). On the other hand, (CA) presented only medium twist values along its TCGA tetrad (37°, 35°, 37°).

There is thus strong evidence indicating that the conformation of CpG undergoes significant change in going from (CA) to (CT). This is especially surprising as this step has the same adjacent residues in the two sequences. Moreover, while the TCGA tetrad structure within (CT) differs from that found in (CA), it resembles more the global structure of the ACGT tetrad observed in both d(CATGACGTCATG)$_2$ (Mauffret et al., 1992) and d(GTACGTAC)$_2$ [named (GT)] (Lefebvre et al., 1995b). This demonstrates the danger of uniquely considering the neighbors in evaluating the properties of a dinucleotide within a DNA double helix.

A further remarkable feature is the good correlation for (CT) between the twist angle and the ^{31}P chemical shift values. Such a correlation was also found for (CA) and for the (GT) oligonucleotide (Lefebvre et al., 1995b), thus confirming the interdependence of helix geometry and backbone structure in DNA duplexes. This correlation can be extended to our previous CD data (El Antri et al., 1993b). As the positions and intensities of CD signals are closely related to the helix geometry, we propose that the differences which have been observed between (CT), (GT), and (CA) (El Antri et al., 1993b) express structural differences in the helices mainly related to the twist angles (Lefebvre et al., 1995b). A particularly good correlation is found between the average twist value and the position of the positive CD

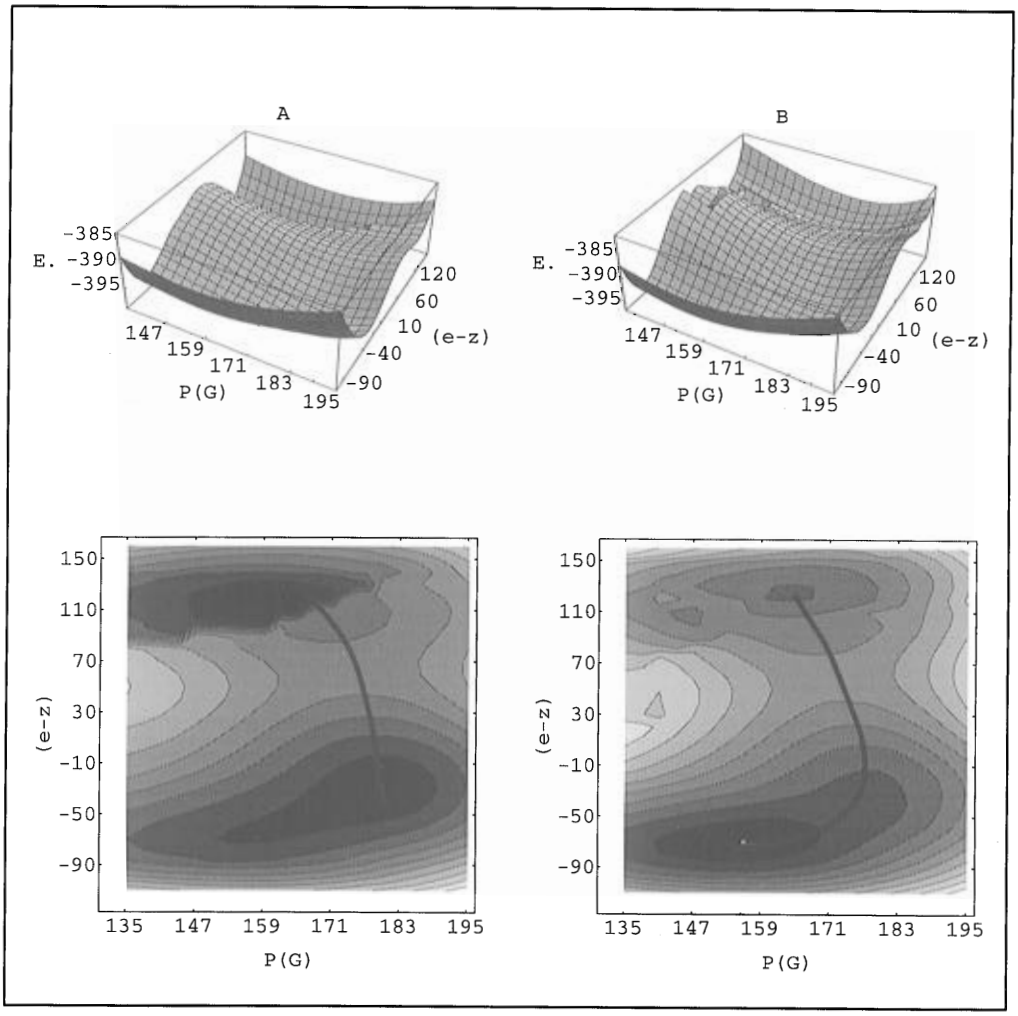

FIGURE 4: Energy maps as a function of the G5 phase angle and $\epsilon-\zeta$ values for $(CTTCGAAG)_2$ (A) and $d(CATCGATG)_2$ (B).

band, in agreement with the previous findings of Johnson et al. (1981). Thus, (CT) and (GT) which showed an average twist of 37.5° had their positive CD band at ≈280 nm, *vs* 260 nm for (CA) which had a twist average value of only 34°.

Transition to a CpG BII Phosphate Conformation. In the case of (CT), the phosphate group of CpG presented a strongly downfield-shifted resonance (−3.9 ppm), together with an atypical temperature variation, mimicking that shown by mismatched base pairs, especially the wobble G·T (Roongta et al., 1990). This effect reflects either a high $\epsilon-\zeta$ value or a BI/BII equilibrium at the CpG step backbone (Gorenstein et al., 1994; Roongta et al., 1990). The latter hypothesis is in complete agreement with the observed line broadening of the CpG phosphorus resonance.

After having investigated the possibility of an average BI structure, we wished to test the relevance of the second hypothesis (i.e., BI/BII equilibrium) in (CT), and also in (CA), for comparison. Still in agreement with the ^{31}P chemical shift data, all structures containing a BII conformation at any step except CpG were found to be much less stable, and, in addition, the corresponding BI to BII transitions presented high energy barriers.

The energy maps for (CA) and (CT) obtained as a function of the G5 pseudorotation angle and the $\epsilon-\zeta$ values of CpG are given in Figure 4. Paths to transit from the energy minimum BI to that for BII are also indicated on Figure 4.

Table 3: Energetic Characteristics of the BI to BII Transition for $d(CTTCGAAG)_2$ (CT) and $d(CATCGATG)_2$ (CA)[a]

	oligonucleotide			
	(CT)	(CA)		
equilibrium	CTBI/CTBII	CABI/CABII	CABI/CAsh	CAsh/CABII
$E(BI) - E(BII)$	0	−3.75	−1.9	−1.9
energy barrier	5.7	7	0	5.1

[a] Energies are in kilocalories per mole.

Energy differences between minima and the energy barriers between these BI and BII structures are given in Table 3.

At first glance, (CA) and (CT) appear to display very different behavior with respect to the BI to BII transition at CpG (Figure 4). (CT) presented two well-defined minima, corresponding to CpG, either under the BI conformation (named CTBI) or under the BII conformation (named CTBII). In comparison, the minima were less well-defined for (CA). We also detected one minimum in the BI region (bottom of the map), for a G5 phase of about 160° (named CABI), this being accompanied by a shoulder with a G5 phase of about 180° (named CAsh), in agreement with our previous results from adiabatic mapping (Lefebvre et al., 1995b). At the same time, the BII region of (CA) (top part of the map) presented two energy minima, with a G5 phase of either 165° (named CABII) or 147° (named CABIIb). The reality of the CABIIb minimum will not be further discussed, since, first, its high energy renders its presence in solution

very unlikely (+4.5 kcal/mol with respect to the main BI minimum CTBI) and, second, its helicoidal parameters were very similar to those of CABII, making it difficult to discriminate between these structures.

The difference between (CA) and (CT) was also visible from an energetic point of view (Table 3). The energy barrier which has to be crossed to achieve the BI to BII transition was lower in (CT) than in (CA) (5.7 kcal/mol *vs* 7 kcal/mol). Moreover, in (CT), the BII minimum appeared as stable as the BI minimum, while in (CA), it was less stable than both the main BI minimum and the shoulder [E(BI) − E(BII) = −3.75 kcal/mol]. Also, as shown in Figure 4, the transition path is quite direct and very short for (CT), contrasting with that required in the case of (CA) for the transition from the main BI minimum (PG5 = 160°) to any of the BII minima, because the molecule has to pass through the CAsh structure.

Thus, these results again confirm the distinct behavior of CpG in (CT) with respect to (CA), although this dinucleotide has the same direct neighbors in both oligonucleotides. All the energy parameters (barrier, difference between the BI and the BII structures) and also the path length for transitions jointly show that the occurrence of a BI/BII equilibrium at the CpG step is more favorable in (CT) than in (CA).

Particular Features of the BII Conformers. The three structures obtained for (CT) are presented in Figure 5: BI minimum (CTBI), BII one (CTBII), and that from minimization with refined NMR constraints (CTraf). Their helicoidal parameters are presented in Figure 3.

The general appearance of the BI and BII minima revealed that the two BII steps, which face one another, induce a local kink at the center of the double helix, as previously described by Hartmann et al. (1993).

Examination of the helicoidal parameters confirmed this feature (Figure 3). The double helix (CT) presented a very negative roll (−22°) at CpG, accompanied by both a strong positive X displacement (+2.4 Å) and an associated inclination (−35°). As this inclination was conserved all along the double helix, thus preventing the loss of stacking interaction, no tilt angle was generated. A very high twist (54°) associated with a strong positive slide (+1.4 Å) was also observed for CpG under the BII conformation. In the backbone, the positive value of $\epsilon-\zeta$ (+155°) typical of BII was coupled to a significant decrease of β (126°); the G5 sugar phase also changed from 180° to 160°. The BII structure of (CA) presented exactly the same features: a negative roll at CpG (−30°) and a high twist (48°), as well as a negative inclination (−25°) all along the double helix. These were in perfect agreement with those previously obtained for BII structures, through energy minimization (Hartmann et al., 1993) or crystallography, at least for the roll and twist values (Heinemann & Hahn, 1992). Moreover, in our case, the β angle at CpG was determined to be smaller in CTBII than in CTBI (138° *vs* about 180°).

For CABI, CAsh, and CABII, a continuous variation was observed for several parameters. For example, the CpG twist was small in the first structure (32°), higher in the second one (42°), and very high in the last one (48°). This was also true for the slide, which was −0.6 Å in CABI, +0.4 Å in CAsh, and +1.3 Å in CABII. Note also that the CpG junction was different in the three structures: P(C) = 167°, A(C) = 37°, and P(G) = 157° in CABI, the reverse, P(C) = 157°, A(C) = 47°, and P(G) = 179° in CAsh, and P(C) = 153°, A(C) = 43°, and P(G) = 159° in CABII, like in CTBII (P, sugar phase angle; A, sugar amplitude).

NMR Arguments in Favor of a BI/BII Equilibrium at CpG. Given these strong variations among parameters, the structures of interest do not generate identical NMR data. The question is then whether the NMR data fit better the BI conformer, the BII conformer, or a BI/BII equilibrium; also other conformations are certainly also possible.

For the backbone conformation at CpG, the main difference arises from the characteristic $\epsilon-\zeta$ value. As mentioned above, the CpG ^{31}P chemical shift was only −3.9 ppm in (CT), in agreement with a significant amount of BII conformer (Roongta et al., 1990). In the case of (CA), a more classical ^{31}P chemical shift was found for CpG (−4.05 ppm). This is also compatible with the presence of the BII conformer, but in a smaller proportion than in (CT).

The drastic changes observed in the helicoidal parameters between the BI and BII conformers of CpG also have consequences on several distances amenable to NOE volume measurements. This was the case for the internucleotide distances involving CpG, which strongly depend on the twist and the roll values, and also for the H1′−H4′ intranucleotide distances for C4 and G5, which are directly related to the corresponding sugar conformations (Table 4).

Obviously, in the case of a two-state equilibrium in solution, the intensities of the experimental NOE may not represent either of the two structures, but may reflect the mixture of the two conformers. To assess the relevance of such an equilibrium, we calculated the simulated NOE volumes of both structures and then deduced, by a linear combination, the theoretical NOE spectra of mixtures corresponding to various ratios of conformers. Hence, the agreement between these theoretical volumes and experimental NOE intensities was calculated. For (CT), we analyzed the equilibrium between conformers CTBI and CTBII. Remarkably, when compared to values obtained by NOE refinement, several distances were found smaller in CTBI and greater in CTBII. For example, d(C4H6−G5H8) was 5.3 Å and 6.5 Å in CTBI and CTBII, respectively, while the value obtained after NOE refinement in CTraf lay exactly between these two distances. In the same way, d(C4H2′−G5H8) was 4.1 Å, 4.5 Å, an 4.4 Å in CTBI, CTBII, and CTraf, respectively. These observations suggested that a mixture of CTBI and CTBII would better fit the NOE volumes than a single conformer.

The agreement of individual NOE volumes with experimental data as a function of CTBI ratio is presented in Figure 6A. This concerns those distances which varied significantly between the two structures. Such a direct analysis of NOE volumes confirmed that neither of the two conformers, CTBI or CTBII, could perfectly reproduce the NOE data. On the other hand, all volumes considered presented maximum agreement with the corresponding NMR intensity for a mixture with a significant proportion of CTBII structure (Figure 6A). Thus, the C4 and G5 H1′−H4′ volumes showed a minimum R factor at low mixing time for a BII/BI ratio of about 40%. In the same way, internucleotide distances such as C4H2′−G5H8 and C4H6−G5H8 displayed a maximum agreement for a CTBII/CTBI ratio of 50−60%. It is thus likely that distances with the greatest conformational significance would be better fitted by a BI/BII equilibrium at the CpG step, rather than a single-state conformation.

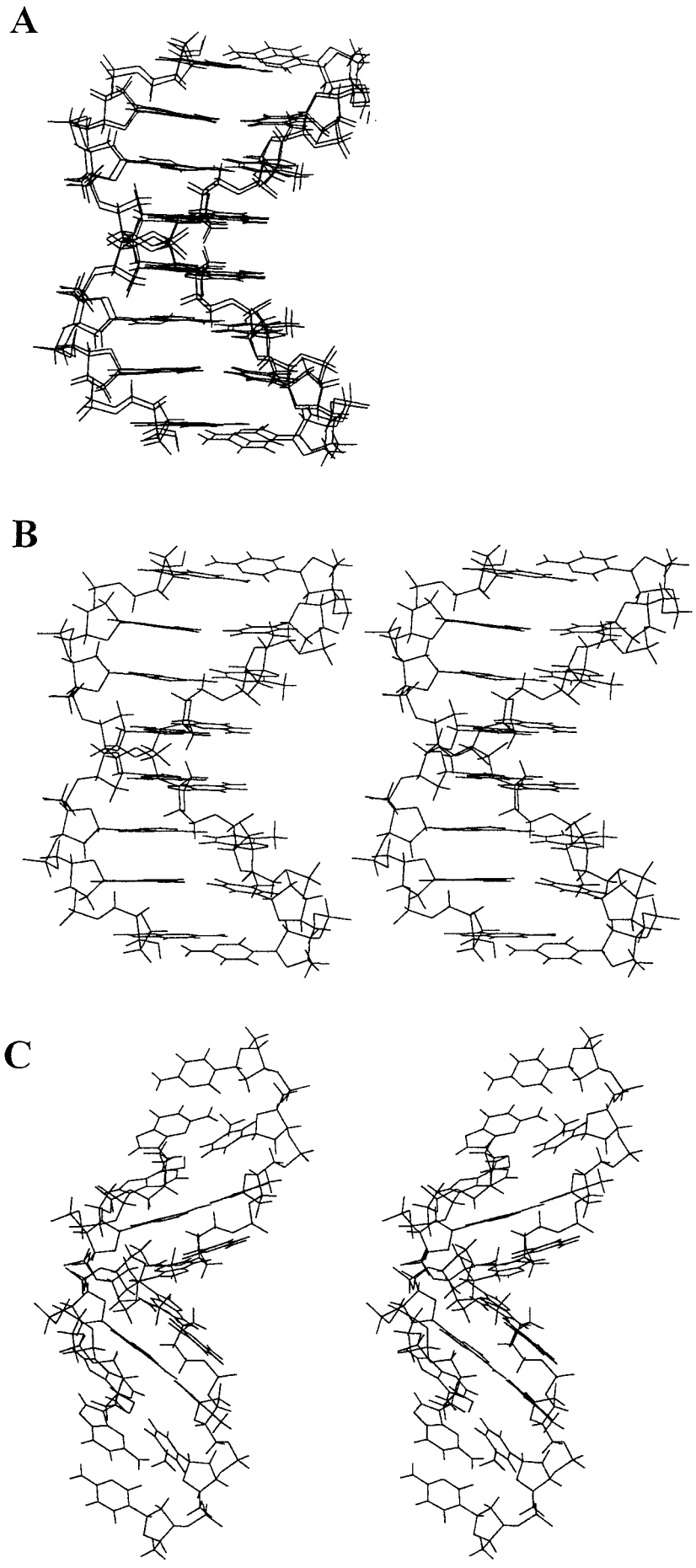

FIGURE 5: Structures of CTBI and CTraf (superimposed) (A), of CTBI (B), and of CTBII (C).

Table 4: Distance Values Presenting Significant Differences between the Observed Energetic Minima of d(CTTCGAAG)$_2$ (CT) and d(CATCGATG)$_2$ (CA) Oligonucleotides

distance (Å)	CTBI	CTBII	CABI	CAsh	CABII
C4H6−G5H8	5.4	6.5	4.6	5.4	6.3
C4H2′−G5H8	4.1	4.5	3.4	4.0	4.4
C4H1′−H4′	3.0	3.2	3.2	3.1	3.3
G5H1′−H4′	3.4	3.1	3.0	3.4	3.1
G5H2′−A6H8	3.7	4.0	4.0	3.7	4.0

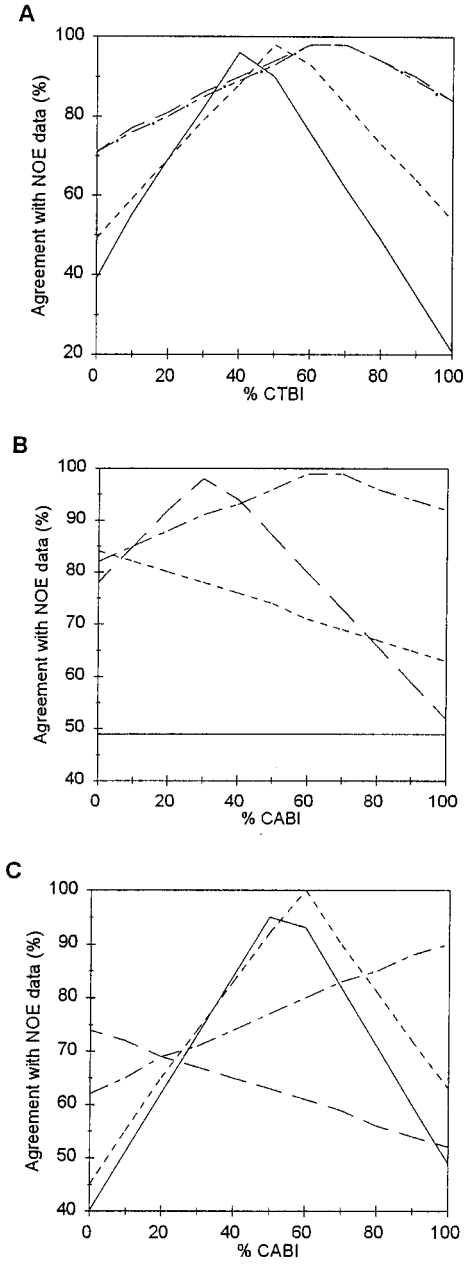

FIGURE 6: Agreement between theoretical and experimental NOE data for a CTBI/CTBII equilibrium (A), a CABI/CAsh equilibrium (B), and a CABI/CABII equilibrium (C). Distances represented are as follows: C4H6−G5H8 for a mixing time of 300 ms (—), C4H2′−G5H8 at T_m = 100 ms (- - -), C4H1′−H4′ at T_m = 50 ms (— · - —), G5H1′−H4′ (— —) at T_m = 50 ms, and G5H2′−A6H8 (— - —) at T_m = 100 ms.

The case of (CA) was a little more complicated, as we identified several distinct minima, CABI, CAsh, and CABII. Yet, as remarked in the case of the conformers CTBI and CTBII of (CT), none of these three structures were in perfect agreement with the NOE data. We started the study by examining two possible two-state equilibria: CABI−CAsh and CABI−CABII. Variations of the agreement between theoretical and experimental NOE volumes as a function of the CABI ratio are given in Figure 6B,C. The distances *d*(G5H1′−H4′) and *d*(G5H2′−A6H8) had virtually identical values in CABI and CABII, so that consideration of a BI/BII equilibrium did not improve the agreement with experimental NOE data (Figure 6B). On the contrary, a mixture of CABI and CAsh provided a much better agreement for these distances (Figure 6B). We then considered the *d*(C4H6−G5H8) and *d*(C4H2′−G5H8) internucleotide distances. This led to a different result where CABI/CAsh mixtures were unable to reproduce the NOE data, while good agreement was obtained with a CABI/CABII equilibrium for a BII/BI ratio of about 40−50% (Figure 6C).

It was thus clear for (CA) that neither CABI, CAsh, and CABII taken alone nor a two-state equilibrium combining these conformations could reproduce the NOE data. The possible existence of a three-state equilibrium corresponding to a mixture of CABI, CAsh, and CABII is quite logical. Assuming that the BII conformation does exist, this must be accompanied by the CAsh conformer, the latter being unavoidable as it lays on the transition path between the CABI and CABII conformations (Figure 4). In the same way, assuming now that the CAsh conformer is also present, the energy barrier between this latter conformer and the CABII conformer becomes comparable to that between CTBI and CTBII (Table 3); one should thus detect a proportion of the CABII conformer. Yet, both the ^{31}P chemical shift value shown by CpG and the energy difference value found between CAsh and CABII (Table 3) suggest that the BII conformer must be less abundant for (CA) compared to (CT). Indeed, we found that a mixture of 50% CAsh, 20% CABI, and 30% CABII provided an agreement with the NOE data of more than 95% when referring to either the distances *d*(G5H2′−A6H8) and *d*(C4H2′−G5H8) at 100 ms or the distance *d*(C4H1′−H4′) at 50 ms, and more than 85% with the distance *d*(C4H6−G5H8) at 300 ms.

CONCLUSION

Previous experimental data have demonstrated that the switch of d(CTTCGAAG)$_2$ (CT) into d(CATCGATG)$_2$ (CA) through permutation of residues 2 and 7 entailed noticeable changes in the structural and thermodynamic properties of the central CpG dinucleotide (El Antri et al., 1993a). Our energy minimization with NMR constraints has confirmed the above-mentioned findings and emphasized that the structural properties of CpG cannot be explained by only considering its flanking 5′ residue and 3′ residue. More distant nucleotides in the sequence must be taken into account. We also found that the (CT) oligonucleotide, which displays structural properties resembling in many aspects those of mismatch-containing oligonucleotides, is better characterized by a BI/BII equilibrium at CpG, which is in agreement with both the ^{31}P chemical shift values and NOE data simulation. Such an equilibrium may also occur in the (CA) oligonucleotide, although the proportion of BII is found now weaker. This is reflected by a higher field ^{31}P resonance, a less favorable energy, and a longer transit path from BI to BII, compared to (CT).

An interesting comparison can also be made with the structure of the d(CATGACGTCATG)$_2$ oligonucleotide

(Hartmann et al., 1993), where CpG is contained in a perfect pyrimidine-purine alternation, ACGT, instead of a two pyrimidines-two purines alternation, TCGA, as for (CT) and (CA). In that case, the conformational results taking into account the existence of a BI/BII equilibrium at CpG have been found to be incompatible with NMR data (Hartmann et al., 1993).

All together, it is likely that a pyrimidine and a purine are required at the 5′-end and 3′-end, respectively, of CpG, to generate a BII conformation, although, as revealed in this work, the next two residues also influence the BI/BII ratio.

Does a BI/BII equilibrium have biological relevance? The occurrence of phosphate groups with BII conformations could explain the high frequency of mutations observed at some CpG steps (Rideout et al., 1990). An inquiry into the problem has been published by Mitra et al. (1995), who analyzed the behavior of T·G mispairs in various DNA contexts. They have found that local structure variations in the backbone, especially an increase of BII, could be related to the observed mutation frequencies. As emphasized in the present study, the C·G base pairs of CpG dinucleotides, although having normal Watson−Crick base-pairing, can have backbone conformations mimicking those of mismatched base pairs, such as T·G. Although the BI−BII transition is dynamic, it can be postulated that CpGs with BII conformations may be mistaken for mismatches by repair enzymes. Their "correction" could then be the cause of at least a part of the mutations usually affecting CpG sites.

ACKNOWLEDGMENT

We thank George Tevanian for very helpful advice concerning NMR experiments and Richard Lavery for reading the manuscript and fruitful discussions.

REFERENCES

Arnott, S., & Hüskin, D. W. L. (1973) *J. Mol. Biol. 81*, 193−105.
Baïkalov, I., Grzeskowiak, K., Yanagi, K., Quintana, J., & Dickerson, R. E. (1993) *J. Mol. Biol. 231*, 768−784.
Baleja, J. D., Germann, M. W., van de Sande, J. H., & Sykes, B. D. (1990a) *J. Mol. Biol. 215*, 411−428.
Baleja, J. D., Moult, J., & Sykes, B. D. (1990b) *J. Magn. Reson. 87*, 375−384.
Banks, K. M., Hare, D. R., & Reid, B. R. (1989) *Biochemistry 28*, 6996−7010.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355−360.
Bird, A. P. (1978) *J. Mol. Biol. 118*, 49−55.
Bird, A. P. (1980) *Nucleic Acids Res. 8*, 1499−1504.
Boelens, R., Konig, T. M. G., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1989) *J. Magn. Reson. 84*, 290−308.
Cooper, D. N., & Youssoufian, H. (1988) *Hum. Genet. 78*, 151−155.
Coulondre, C., Miller, J. H., Farabaugh, P. J., & Gilbert, W. (1978) *Nature 274*, 775−780.
Dietrich, W., Rudel, C. H., & Neumann, M. (1991) *J. Magn. Reson. 91*, 1−11.
El Antri, S., Mauffret, O., Monnot, M., Lescot, E., Convert, O., & Fermandjian, S. (1993a) *J. Mol. Biol. 230*, 373−378.
El Antri, S., Bittoun, P., Mauffret, O., Monnot, M., Convert, O., Lescot, E., & Fermandjian, S. (1993b) *Biochemistry 32*, 7079−7088.
Fasman, G. D. (1975) in *Handbook of biochemistry and molecular biology: Nucleic Acids*, 3rd ed., CRC Press, Cleveland, OH.
Fedoroff, O. Y., Reid, B. R., & Chuprina, V. P. (1994) *J. Mol. Biol. 235*, 325−330.
Gochin, M., Zon, G., & James, T. L. (1990) *Biochemistry 29*, 11161−11171.
Gorenstein, D. G. (1994) *Chem. Rev. 94*, 1315−1338.
Grippo, P., Iaccarino, M., Parisi, E., & Scanaro, E. (1968) *J. Mol. Biol. 36*, 195−200.
Grzeskowiak, K., Yanagi, K., Privé, G. G., & Dickerson, R. E. (1991) *J. Biol. Chem. 266*, 8861−8883.
Hartmann, B., Piazzola, D., & Lavery, R. (1993) *Nucleic Acids Res. 21*, 561−568.
Heinemann, U., & Hahn, M. (1992) *J. Biol. Chem. 267*, 7332−7341.
Hollstein, M., Shomer, B., Greenblat, M., Soussi, T., Hovig, E., Montesano, R., & Harris, C. C. (1996) *Nucleic Acids Res. 24*, 141−146.
Johnson, B. B., Dahl, K. S., Tinoco, I., Ivanov, V. I., & Zhurkin, V. B. (1981) *Biochemistry 20*, 73−78.
Kim, S.-G., & Reid, B. R. (1992a) *Biochemistry 31*, 3564−3574.
Kim, S.-G., & Reid, B. R. (1992b) *Biochemistry 31*, 12103−12116.
Köning, T. M. G., Boelens, R., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1991) *Biochemistry 30*, 3787−3797.
Lankhorst, P. P., Haasnoot, C. A. G., Herkelens, C., & Altona, C. (1984) *J. Biomol. Struct. Dyn. 1*, 1387−1399.
Lavery, R. (1988) in *Structure and expression, Vol. 3, DNA bending and curvature* (Olson, W. K., Sarma, R. H., & Sundaralingam, M., Eds.) pp 191−211. Adenine Press, New York.
Lavery, R., & Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63−91.
Lavery, R., & Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 655−667.
Lavery, R., Zakrzewska, K., & Sklenar, H. (1995) *Comp. Phys. Commun. 91*, 135−158.
Lefebvre, A., Mauffret, O., El Antri, S., Monnot, M., Lescot, E., & Fermandjian, S. (1995a) *Eur. J. Biochem. 229*, 445−454.
Lefebvre, A., Mauffret, O., Hartmann, B., Lescot, E., & Fermandjian, S. (1995b) *Biochemistry 34*, 12019−12028.
Levy, D. D., Magee, A. D., Namiki, C., & Seidmann, M. M. (1996) *J. Mol. Biol. 255*, 435−445.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Mauffret, O., Hartmann, B., Convert, O., Lavery, R., & Fermandjian, S. (1992) *J. Mol. Biol. 227*, 852−875.
Mitra, R., Pettit, B. M., & Blake, R. D. (1995) *Biopolymers 36*, 169−179.
Mujeeb, A., Kerwin, S. M., Egan, W., Kenyon, G. L., & James, T. L. (1992) *Biochemistry 31*, 9325−9338.
Neuhaus, D., Wagner, G., Vasak, M., Kägi, J. H. R., & Wüthrich, K. (1985) *Eur. J. Biochem. 151*, 257−273.
Nikonowicz, E., & Gorenstein, D. G. (1990) *Biochemistry 29*, 8845−8858.
Nikonowicz, E., Roongta, V., Jones, C. R., & Gorenstein, D. G. (1989) *Biochemistry 28*, 8714−8725.
Parris, C. N., Levy, D. D., Jessee, J., & Seidman, M. M. (1994) *J. Mol. Biol. 236*, 491−502.
Privé, G. G., Yanagi, K., & Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 177−199.
Rideout, W. M., III, Coetzee, G. A., Olumi, A. F., & Jones, P. A. (1990) *Science 249*, 1288−1290.
Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621−649.
Roongta, V. A., Jones, C. R., & Gorenstein, D. G. (1990) *Biochemistry 29*, 5245−5258.
Sanghani, S. R., Zakrzewska, K., & Lavery, R. (1995) *J. Biomol. Struct. Dyn., Proc. 9th Conversation* (in press).
Sklenar, V., & Bax, A. (1987) *J. Am. Chem. Soc. 109*, 7525−7526.
Sodano, P., Hartmann, B., Rose, T., Wain-Hobson, S., & Delepierre, M. (1995) *Biochemistry 34*, 6900−6910.
Tuddenham, E. G. D., Schwaab, R., Seehafer, J., Millar, D. S., Gitschier, J., Higuchi, M., Bidichandani, S., Connor, J. M., Hoyer, L. W., Yoshioka, A., Peake, I. R., Olek, K., Kazazian, H. H., Lavergne, J.-M., Giannelli, F., Antonarakis, S. E., & Cooper, D. N. (1994) *Nucleic Acids Res. 22*, 3511−3533.
Widmer, H., & Wüthrich, K. (1987) *J. Magn. Reson. 74*, 316−336.
Yanagi, K., Privé, G. G., & Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 201−214.

BI9606298