Structural Behavior of the CpG Step in Two Related Oligonucleotides Reflects Its Malleability in Solution†

A. Lefebvre,‡ O. Mauffret,‡ B. Hartmann,§ E. Lescot,‡ 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 April 17, 1995; Revised Manuscript Received July 3, 1995⊗

ABSTRACT: We report on the determination of the solution structure of two sequence-related oligonucleotides, d(GTACGTAC)$_2$ and d(CATCGATG)$_2$. Results have been obtained by using a combined approach of (a) two-dimensional NMR, including proton and phosphorus experiments, (b) restrained molecular mechanics performed with sugar phase angle, backbone ϵ angle, and NOE distances as input, and (c) back-calculation refinements against the NOE spectra at various mixing times. The two oligonucleotides adopt the B-DNA structure with, however, noticeable differences centered on their core sequence and especially the CpG step. Due to the permutation of its flanking residues, the CpG step modifies its twist values and backbone ϵ value; globally, the CpG step appears more flexible within the tetranucleotide TCGA than ACGT. The solution structure of d(GTACGTAC)$_2$ differs from the previously reported X-ray structure, which was found to be A-form throughout [Takusagawa, F. (1990) *J. Biomol. NMR 3,* 547−568]. On the other hand, in the X-ray structure of d(CCAACGTTGG)$_2$ [Privé et al. (1991) *J. Mol. Biol. 217,* 177−199] the structure of the ACGT sequence is similar to that found in solution d(GTACGTAC)$_2$. Similarly, the central TCGA tetranucleotide of d(CATCGATG)$_2$ presents a solution structure analogous to that observed on the X-ray structures of both d(CGATCGATCG)$_2$ [Grzeskowiak, et al. (1991) *J. Biol. Chem. 266,* 8861−8883] and d(CGATCGmeATCG)$_2$ [Baïkalov, et al. (1993) *J. Mol. Biol. 231,* 768−784]. At the end we discuss the possible biological significance of the particular structures exhibited by the ACGT and TCGA tetranucleotides.

The three-dimensional structures and dynamics of oligonucleotides in solution have been essentially estimated by nuclear magnetic resonance (NMR)[1] (Wüthrich, 1986; Wijmenga et al., 1993; Van de Ven & Hilbers, 1988). The NOE intensities have been used as input for modeling methods, such as distance geometry, restrained molecular mechanics, and dynamics, to estimate the conformational properties of oligonucleotides (Wijmenga et al., 1993; Van de Ven & Hilbers, 1988). However, since in the DNA double helix no long-range distance constraints similar to those found in proteins are available, a good definition of the oligonucleotide structure remains difficult. Certainly, the overall double-helical conformation is known from other methods such as circular dichroism (Gray et al., 1992) and infrared spectroscopy (Taillandier & Liquier, 1992). However, the true questions regard fine details of the sublocal structure. This places high demands on the quality and efficiency of the algorithm applied and requires the use of scalar coupling constants combined with NOE intensities as constraints in molecular modeling studies (James, 1994).

A great deal of effort has been devoted to obtaining accurate NMR constraints. Full relaxation matrix calculations with programs such as MARDIGRAS (Borgias & James, 1990), MORASS (Post et al., 1990), and IRMA (Boelens et al., 1989) allow spin diffusion effects to be taken into account and thus provide accurate interproton distances from 2D NOE cross-peak intensities. Quantitative simulation of cross-peak patterns in coherence transfer experiments may help to determine the coupling constant values in the case of broad resonances (Widmer & Wüthrich, 1986, 1987). Despite these improvements, one must keep in mind the averaging processes due to the multiple exchanging conformations (Schmitz et al., 1993; Pearlman, 1994; Pearlman & Kollman, 1991) and to rapid internal motions. The latter effects have been treated explicitly using specific motional models or order parameters but require additional parameters that are generally difficult to obtain, thus precluding a more detailed analysis (Lane, 1990; Koning et al., 1991; Borer et al., 1994).

As is often repeated (Ulyanov et al., 1992; van de Ven & Hilbers, 1988; Metzler et al., 1990; Wijmenga et al., 1993), the available NMR constraints represent only a subset of the total number of constraints that would be necessary to uniquely define the 3D structure. A large conformational space has to be explored in order to locate a global minimum (or minima) consistent with all the available stereochemical and NMR constraints. Efficient algorithms specifically

† This work was supported by Fondation Médicale pour la Recherche, Ligue Nationale contre le Cancer, and Association pour la Recherche contre le Cancer. A. L. is a Fellow of the Institut de Formation Supérieure Biomédicale.

‡ Institut Gustave Roussy.

§ Institut Biologie Physico-Chimique.

⊗ Abstract published in *Advance ACS Abstracts,* August 15, 1995.

[1] Abbreviations: COLOC, X−^{1}H shift correlation by long-range coupling via ^{1}H−^{1}H decoupling; COSY, 2D homonuclear shift correlated spectroscopy; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, 2D homonuclear nuclear Overhauser effect correlated spectroscopy; ppm, part per million; TOCSY, total correlated spectroscopy.

0006-2960/95/0434-12019$09.00/0

adapted for nucleic acids have been developed (Lavery, 1988; Ulyanov et al., 1992) to attain this goal.

Such an approach is developed here in order to define the influence of the neighboring bases on the structural parameters of the CpG step within DNA. Recently, Grzeskowiak et al. (1991) analyzed the high-resolution crystal structures (1.5 Å) of the following two decamers : d(CGATCGATCG)$_2$ and d(CCAACGTTGG)$_2$. The CpG and CpA steps exhibit behaviors strongly influenced by the preceding and following steps in the sequence. In particular the twist pattern of the tetranucleotide ACGT (low−high−low) in d(CCAACGTTGG)$_2$ is inverse compared to that of TCGA (high−low−high) in d(CGATCGATCG)$_2$, reflecting the sensitivity of the CpG sructure toward its flanking steps and thus its extreme malleability. Such a property inherent to CpG has been also pointed out through our previous studies performed on oligonucleotides in solution (El Antri et al., 1993a,b).

The propensity of CpG to adopt different conformations within the DNA double helix could be essential for the specific recognition of DNA sequences by gene regulatory proteins, such as GCN4 or c-Jun. These strongly associate to the promoter domain of the cAMP-responsive element (CRE) which contains CpG as the central step in its sequence (Ellenberger et al., 1992; König & Richmond, 1993; Mauffret et al., 1992). The relative deformability of CpG facilitates the formation of a stable complex by providing a low-energy barrier to any structural adjustments upon binding to the protein. Moreover this step behaves as a hot spot for mutations (El Antri et al., 1993a; Cooper & Youssoufian, 1988; Laird & Jaenisch, 1994), and the question is whether its particular structural and dynamic properties interfer in this process.

Here, we compare the structural properties of CpG in tetranucleotides ACGT and TCGA pertaining to the octanucleotides d(GTACGTAC)$_2$ (which we will abbreviate as TAK) and d(CATCGATG)$_2$ (abbreviated as PERMUT), respectively. Structures of duplexes in solution were assessed with a NMR restraint molecular mechanics approach. Part of the NMR results were published earlier (El Antri et al., 1993a).

MATERIALS AND METHODS

Sample Preparation. The two DNA octamers, d(GTACGTAC)$_2$ and d(CATCGATG)$_2$ were synthesized using the solid phase procedure on an Applied Biosystems 381 B automatic apparatus. The resulting oligonucleotides were purified by reversed phase HPLC on a C18 Waters μBondapak C18 column followed by dialysis in water and lyophilization. Oligonucleotide concentrations were calculated using 8200 M^{-1} cm^{-1} $base^{-1}$ as the extinction coefficient value at 260 nm, according to Fasman (1975).

Each oligonucleotide was dissolved at ~3.5 mM concentration in a phosphate buffer containing 0.2 mM ethylenediaminetetraacetic acid (EDTA) at pH 7, ionic strength I = 0.1.

For the experiments in ^2H_2O, the samples were lyophilized three times in increasing grades of ^2H_2O and finally taken up in 0.4 mL of 99.99% ^2H_2O.

NMR Spectroscopy. (A) ^1H NMR Experiments. The spectra were acquired at 500 MHz on an AMX 500 Bruker spectrometer and processed using the Felix Software (Biosym Technologies) on a Silicon Graphics INDIGO R4000 workstation. All the NMR experiments were conducted at 25 °C.

The quantitative 2D NOE experiments in ^2H_2O were recorded at different mixing times (50, 100, 200, 300, and 400 ms) and conducted in a single span of 4 days without removing the sample from the probe. Time proportional phase incrementation (TPPI) was used for quadrature detection in F1. A spectral width of 5000 Hz was used with the carrier frequency positioned at the HDO absorbance, which was presaturated during the relaxation delay of 2 s and during the mixing time. This delay is enough to provide the relaxation of all the protons, except the adenine H2 protons. For these protons, additional experiments were recorded with a relaxation delay of 6 s to ensure sufficient recovery of their magnetization between successive scans and thus a better quantification of cross peaks implicating this proton. This allowed us in turn to test the validity of the relaxation delay (2 s) used for the other protons. A total of 400 experiments were performed with 1024 complex points acquired for each FID. The reference phase of the transmitter was adjusted to give pure absorption spectra, presenting a flat baseline. After multiplication of NOESY data sets with shifted sine and sine-square bells in dimensions 1 and 2, respectively, data were zero-filled in the two dimensions to yield a final 2K × 2K real matrix. Baseline correction was performed with a baseline convolution method (Dietrich et al., 1991). The measurements of experimental intensities and translation of these intensities into distances were made by the distance extrapolation method, which permits one to take partially into account the spin diffusion effects (Mauffret et al., 1992; Baleja et al., 1990a,b; Fedoroff et al., 1994).

Pure phase-absorption double-quantum COSY spectra of the octamers were recorded by using time proportional phase incrementation for the separation of −1 and +1 quantum transitions (Marion & Wüthrich, 1983) using 4096 complex points in t_2 and 700 in t_1. The spectral width was 4000−5000 Hz, the number of scans 48−64, and the repetition delay 2 s. COSY data were processed using sine bells 60° shifted in the t_2 dimension and 30° phase-shifted in the t_1 dimension. The spectral resolution was enhanced by zero-filling to produce a 4096 × 2048 real matrix. For the simulations of the COSY cross peaks, the SPHINX and LINSHA programs (Widmer & Wüthrich, 1987) were used as described elsewhere (Mauffret et al., 1992).

Clean TOCSY experiments were performed using standard methods with a spin lock pulse of 125 ms consisting of an MLEV17 pulse train surrounded by trim pulse of 1 ms (Bax & Davies, 1985).

(B) ^{31}P NMR Experiments. The 1D and 2D experiments were recorded on either a MSL 300 or an AMX 500 spectrometer, both equipped with a selective ^{31}P detection probe. For assignment purposes, we currently record Kessler−Griesinger long-range heteronuclear correlation (COLOC) (Kessler et al., 1984; Gorenstein et al., 1988) or ^1H detection (Bax & Davies, 1985) experiments. The assignments of phosphorus resonances for the oligonucleotides TAK and PERMUT were previously reported (El Antri et al., 1993b).

For measurements of $^{31}P-^1H$ coupling constants, two kinds of experiments were performed: either the 2D J-resolved experiments selective for H3′ protons (Roongta et al., 1990) or the ^1H detected heterocorrelation experiments

with selective pulses on the H3′ protons (Sklenar & Bax, 1987). The three-bond coupling constants $^3J_{H3'-P}$ were then used through a proton−phosphorus Karplus-type relationship to extract the H3′−C3′−O−P torsional θ angle related to the C4′−C3′−O−P ϵ angle. The selected relationship $J = 15.3 \cos^2 \theta - 6.1 \cos \theta + 1.6$ was previously reported by Lankhorst et al., (1984).

Simulation Methodology. Determination of the 3D structures for the two octamers was carried out through molecular modeling using the JUMNA (JUnction Minimization of Nucleic Acids) algorithm (Lavery & Sklenar, 1988; Lavery, 1988; Poncin et al., 1992; Mauffret et al., 1992). JUMNA permits the direct use of helicoidal variables to describe the nucleic acid structure during minimization. The number of variables to be treated is reduced by a factor of 10 compared to Cartesian coordinate molecular mechanics, and this greatly facilitates energy minimization, diminishing the number of local minima and increasing the amplitude of conformational changes that can be studied. Similar results using Cartesian coordinate representations can only be achieved with molecular dynamics simulations, which are much more costly. The JUMNA algorithm is also able to handle distance constraints, either with respect to a chosen fixed *d* value or within specified upper and lower bounds (Mauffret et al., 1992). In both cases violations of constraints are prevented by a simple quadratic penalty term using a force constant of either 6 or 12 kcal mol^{-1} Å^{-1} or any other values as will be described in the text. It is also possible to constrain sugar puckering in terms of phase, amplitude, or both these variables in combination.

The obtained structures were analyzed with the CURVES program (Lavery & Sklenar, 1988, 1989), which provides a rigorous way to obtain local structural parameters and overall helical axis locus for irregular structures of nucleic acids (Lavery & Sklenar, 1988). The output files from the program CURVES include two sets of helical parameters: the "global helical parameters" defined relatively to a global helical axis and the "local helical parameters" defined relatively to a local helical axis. Unless specified otherwise, we considered in our analysis the helical parameters calculated with reference to the global axis.

Note also that the analysis excludes the terminal base pairs of each oligomer, which are subject to fraying in solution. The determination of the global axis was made by taken into account only the six central base pairs. These base pairs are the only ones considered for the root-mean-square (rms) calculations unless specified otherwise.

Generation of Initial Structures. In order to obtain the starting points for JUMNA, adiabatic maps for the two central sequences ACGT and TCGA were systematically generated (Lavery & Hartmann, 1993; Sodano et al., 1995). We worked with infinite regular polymers so that the complications due to end effects of oligonucleotides could be eliminated. This was easily achieved with JUMNA by imposing helicoidal symmetry and by optimizing the energy per unit cell within the studied polymer. We have already shown that sugar conformations play a fundamental role in determining the local structure of the double helix (Poncin et al., 1992; Mauffret et al., 1992). The energy surface was consequently investigated by sugar pucker variables. The maps obtained were independent of the starting conformation (see thereafter). For sugar puckers around C2′-endo (P = 120−190°), three minima were found for $(ACGT)_n$ and only one for $(TCGA)_n$. These minima display similar stabilities but differ considerably in terms of their helicoidal parameters.

For the two octamers, minimizations were first carried out without constraints, starting from a range of initial conformations. For the TAK oligonucleotide, nine structures were selected as starting conformations: the B-73 structure (Arnott & Hukins, 1973), the crystal structure of a decamer with the TACGTA sequence (Privé et al., 1991), the NMR structure obtained for $d(GTACGTAC)_2$ with NMR constraints (Baleja et al., 1990a), one DG NMR structure obtained with the constraints of phase, ϵ angles, and distances used in this work [supplemented with a recommended range of values based on experimental data as described in Kim et al.(1992)] and distance geometry algorithm (DG-II algorithm, Biosym Technologies) determined in our laboratory; one structure with the central tetramer in the conformation found for the CRE sequence by a combined JUMNA NMR constraint approach that we earlier conducted on two dodecamers (Mauffret et al., 1992); the three structures obtained by adiabatic mapping for $(ACGT)_n$ (see above); and finally the structure obtained by adiabatic mapping for $d(CATCGATG)_2$. These structures present rms ranging from 0.9 to 3.1 Å. A large range of helicoidal parameters is encountered; for example the central twist (at the CpG step) lies between 30 and 45°.

For the $d(CATCGATG)_2$ oligonucleotide, eight starting structures were selected: the B-73 structure, two crystal structures with TCGA sequence, one DG NMR structure obtained with the same kind of constraints as for the $d(GTACGTAC)_2$, the structure resulting from adiabatic mapping for $(TCGA)_n$ (see above), and the three structures obtained by adiabatic mapping for $(ACGT)_n$. For minimizations starting with crystallographic models, we used a complete set of distance constraints in which very large backbone angle bounds were included (a range of 40° around the canonical values). The main function of these additional constraints is to avoid different crankshafts present in crystallographic structures (ϵ/ζ and α/γ). The rms for these PERMUT structures varies from 0.5 to 2.8 Å, and as with TAK, large variations in twist angles and sugar puckers are present.

Calculation of Theoretical NOESY Spectra and Back-Calculation Iterative Refinement. To refine the agreement between the theoretical NOEs calculated from minimized structures and the experimental NOEs and then to correct the inaccurate initial distance estimates (the distance extrapolation method incorporates spin diffusion effects only partially), we applied iterative NOESY back-calculation refinements (Kim & Reid, 1992a,b; Boelens et al., 1989).

Back-calculation of NOESY spectra from various structures was performed using the corresponding subroutine of the program IRMA, integrated in the INSIGHTII environment (Biosym technology), as discussed by Boelens et al. (1989). This algorithm takes into account all the spin diffusion pathways. A single correlation time was used for every interproton vector, namely, 2 and 4 ns for TAK and PERMUT, respectively. This parameter and the *Z*-leakage factor were determined using the procedure described by Banks et al. (1989).

The theoretical NOESY spectra of a given structure were then calculated at various mixing times, and the calculated cross-peak volumes V_{ij}^{s} compared to the experimental cross-

peak volumes V^e_{ij} at the corresponding mixing time. To enable this comparison, a scaling factor $S(T_m)$ was determined at each mixing time by comparing the nonterminal cytosine H5−H6 reference cross-peak volumes in the calculated and experimental spectra, i.e. $S(T_m) = \sum V^e_{ref}(T_m)/V^s_{ref}(T_m)$. All simulated NOE intensities were then multiplied by this scaling factor at the corresponding T_m.

The fit of the calculated NOE intensities to the experimental data is then evaluated by calculating the NOE residual R factor: $R = (\sum V^e_{ij} - V^s_{ij})/\sum V^e_{ij}$, where the sums run over four mixing times (50, 100, 200, and 300 ms) and over all resolved cross peaks among the intranucleotide H6/H8 to H1′/H2′/H2″/H3′ and H1′ to H2′/H2″ connectivities, and the internucleotide H6/H8 to H1′/H2′/H2″/H5/H6/H8 connectivities (giving ~110 cross peaks).

The distance constraints were then submitted to an automated refinement procedure (Kim & Reid, 1992b). The algorithm used modifies the distance bounds according to the difference between calculated and experimental cross-peak intensities. If V^e_{ij}, the upper distance bound is set to d^{new}_{ijinf}, where $d_{ij\ mol}$ is the distance between protons i and j in the molecular structure considered and the lower bound is kept unchanged. If $V^e_{ij} < V^s_{ij}$, the lower distance bound is set to $d^{new}_{ijsup} = d_{ij\ mol}(V^s_{ij}/V^e_{ij})^{1/6}$ and the upper bound is kept unchanged.

A new run with JUMNA was then performed with the new constraints set. This refinement procedure was repeated until the R factor showed no further improvement.

RESULTS

Sugar Ring Conformation Determination. Recent molecular mechanics studies, confirmed by experimental results, underline the crucial role of the deoxyribose conformation to fashion the DNA structure (Poncin et al., 1992; Mauffret et al., 1992). The determination of sugar puckering is thus essential for fine structure determination of double-stranded DNA fragments. To derive the approximate range of pseudorotation, we used a strategy based on (a) the coupling constants $J_{1'2'}$, $J_{1'2''}$ measured from high-resolution DQF-COSY experiments, (b) the examination of DQF-COSY cross-peak patterns as those H1′(w1) − H2′(w2) (Rinkel & Altona, 1987; Widmer & Wüthrich, 1987; Gochin et al., 1990; Mauffret et al., 1992), and (c) determination of H1′−H4′ distances from NOESY experiments registered at low mixing times (Chuprina et al., 1993). While the individual values for the coupling constants and interproton distances may not be completely correct, they, taken together, restrict the sugar conformation to a fairly narrow range. For the two oligonucleotides, the purines and the pyrimidines present a clear difference in their H1′(w1) − H2′(w2) cross-peak patterns, as well as in the NOE-derived H1′−H4′ distances. P values are found to be smaller and more heterogenous for the pyrimidines than for the purines, as illustrated in Figure 1 for TAK. Note for example the difference between the fine structures of the T3 and the C4 H1′-H2′ cross-peak multiplets. The P angle values derived from the above procedure for the nonterminal residues are reported in Table 1.

Proton−Phosphorus Coupling Constants. The constraints for the ϵ angles have been obtained through the $^1H-^{31}P$ heteronuclear measurements (2D J-resolved). The values of $^3J_{H3'-P}$ were measured for all the nonterminal residues and converted into the corresponding ϵ (C4′−C3′−O−P) values (Table 1) using the proton−phosphorus Karplus relationship of Lankhorst et al. (1984). Although one coupling constant may correspond to four different ϵ values (0−360°), only the rotamers g− and *trans* (ϵ: −60 and 180°) are energetically feasible (Gorenstein et al., 1988; Gorenstein, 1992). Because of a strong correlation between the ϵ angle and the ζ (C3′−O3′−P−O5′) angle, the conformations BI (ϵ, ζ: t, g−) and BII (ϵ, ζ : g−, t) are generally defined (Gorenstein et al., 1988). The typical BII conformation is rejected on the basis of the ^{31}P chemical shifts, which are predicted to be at very low field for BII (Nikonowicz & Gorenstein, 1990; Chou et al., 1992), a characteristic which is not observed in either d(CATCGATG)$_2$ or in d(GTACGTAC)$_2$.

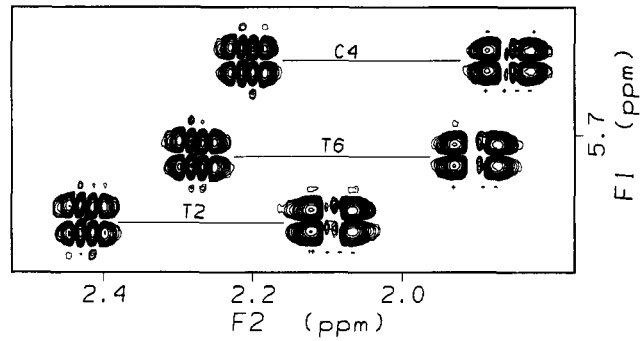

FIGURE 1: Pyrimidine H2′/H2″−H1′ part of a 500 MHz DQF-COSY spectrum of the TAK oligonucleotide at 25 °C. Note the sign and frequency differences between the innner components of multiplets.

Table 1: Constraints Applied during the Energy Minimization of TAK and PERMUT to the P Sugar Pseudorotation Angle and the ϵ Torsion Angle[a]

	TAK					PERMUT			
	P (deg)		ϵ (deg)			P (deg)		ϵ (deg)	
residue	min	max	min	max	residue	min	max	min	max
T2	150	165	−177	−171	A2	160	180	−179	−167
A3	170	180	−177	−171	T3	135	150	−169	−157
C4	140	155	−167	−161	C4	135	150	−167	−155
G5	170	180	−177	−171	G5	165	185	−176	−164
T6	135	150	−177	−171	A6	160	180	−179	−167
A7	170	180			T7	135	150		

[a] The terminal residues were kept free.

The ϵ values used as constraints for the molecular modeling are also given in Table 1. Remarkably, it is the central CpG step in the two oligonucleotides that presents the largest ϵ angle.

Determination of Distance Constraints from NOE Experiments. The distances included as initial constraints are the H6/H8/H5−H1′/H2′/H2″/H3′ inter- and intraresidue distances, together with the base−base proton distances. The interproton distances were estimated using the distance extrapolation method (Baleja et al., 1990a−c; Mauffret et al., 1992), as a first-order correction for spin diffusion effects:

$$r_{ij} = \lim_{T_m \to 0} r_{ref}(T_m)[V_{ref}(T_m)/V_{ij}(T_m)]^{1/6}$$

The reference distance r_{ref} used was the non terminal cytosine H5−H6 distance, taken equal to 2.5 Å (Gronenborn & Clore, 1985). The input precision (difference between lower and upper bound) was 0.8 Å for distances superior to

3 Å, 0.4 Å for distances between 2.5 and 3 Å, and 0.2 Å for distances less than 2.5 Å. In the case of the interproton distances for which spin diffusion effects are known to be large [namely, the H6/H8−H3′ and the H6/H8−H2″ interresidue distances (Kim et al., 1992)], the upper bounds were augmented. Thanks to a good dispersion of NOE cross peaks for both oligonucleotides, the number of distance constraints was approximatly the same for each residue, namely, three or four intraresidue distance constraints and four sequential interresidue constraints.

Since the sugar conformation was constrained through *P*, no intrasugar distances were introduced as additional constraints. The distances implicating methyl groups were not taken into account because, as is well-known, cross-peak intensities depend not only on distances but also on the actual geometry , i.e., on the polar angles in the molecular frame of the internuclear vector (Kim & Reid, 1992a; Tropp, 1980; Tropp & Redfield, 1981).

As pointed out by several groups (Kim et al., 1992; Boelens et al., 1989; Baleja et al., 1990b,c; Fedorov et al., 1994; Cuniasse et al., 1987; Weisz et al., 1992; Mauffret et al., 1992), a rather systematic discrepancy is observed for the H6/H8−H2″ interresidue distances (d_s[H6/H8−H2″]) between the NMR−derived distances (2.7 Å on average) and the crystallographic distances (2.2−2.3 Å on average). Thus, we conducted two series of minimizations: (i) the d_s[H6/H8−H2″] distances derived from NOE measurements and (ii) the d_s[H6/H8−H2″] distances recalibrated from classical B-DNA (2.3 Å) as described in Cuniasse et al. (1987).

Application of the Constraints. In the minimization procedure, the constraints were applied to each starting structure in three successive steps in reference to the sugar phase angles, the ϵ torsion angles, and finally the interproton distances, in order to take into account the relative importance of each type of constraints. Four sets of constraints were thus defined: (1) set S, sugar pseudorotation phase angle constraints; (2) set SE, sugar and ϵ angle constraints; (3) set SED_{nmr}, same as set SE but supplemented by interproton distance constraints, the d_s[H6/H8−H2″] distances being directly derived from NOE measurements; (4) Set SED_{rec}, set SE but supplemented by interproton distance constaints, the d_s[H6/H8-H2″] distances being recalibrated at mearly 2.3 Å (Cuniasse et al., 1987).

NOESY Back-Calculation Refinements. In the final stage of the structural refinement, spin diffusion was taken into account by using the full relaxation matrix approach. This yields more accurate distance constraints, judging by comparison of back-calculated values with experimental data. The refinement procedure led to two new sets of constraints, RAF_{nmr} (NMR-derived d_s[H6/H8−H2″) and RAFrec (recalibrated d_s[H6/H8−H2″]). In RAFrec the d_s[H6/H8−H2″] constraints were not modified during the refinements. For the two oligonucleotides, the refinement was repeated until *R*-factor convergence, which occurred after about three cycles. Finally, two structures were obtained for each octamer, which we will call $TakRAF_{nmr}$ and TakFIN for $d(GTACGTAC)_2$, and $PermutRAF_{nmr}$ and PermutFIN for $d(CATCGATG)_2$.

Energy and R Factor. The energy and the *R* factor of the molecules for all minimizations are presented in Table 2a for $d(GTACGTAC)_2$ and in Table 2b for $d(CATCGATG)_2$. Results correspond to both NMR-derived distances and recalibrated d_s[H6/H8−H2″] distances. The main results for each oligonucleotide are presented in the following.

Table 2: Conformational Energy and *R*-Factor Values at Different Steps of the Structure Refinement Protocol

constraint set	structures obtained	energy (kcal/mol)	*R* factor
	(a) TAK		
S	TakS1	−396.9	0.53
	TakS2 (rms = 0.5 Å)[a]	−397	0.51
SE	TakSE1	−396.9	0.52
	TakSE2 (rms = 0.4 Å)[a]	−396.9	0.50
SED_{nmr}	$TakSED_{nmr}$	−394.4	0.38
SED_{rec}	$TakSED_{rec}$	−396.1	0.48
RAF_{nmr}	$TakRAF_{nmr}$	−390.8	0.26
RAF_{rec}	TakFIN	−391.9	0.28
−	TakREL[b]	−398.1	0.55
	(b) PERMUT		
S	PermutS1	−389.3	0.49
	PermutS2 (rms = 0.29Å)[a]	−389.0	0.47
SE	PermutSE	−388.9	0.47
SED_{nmr}	$PermutSED_{nmr}$	−385.3	0.3
SED_{rec}	$PermutSED_{rec}$	−387.8	0.45
RAF_{nmr}	$PermutRAF_{nmr}$	−380.7	0.26
RAF_{rec}	PermutFIN	−384.0	0.37
REL	PermutREL[c]	−390.9	0.46

[a] The rms are calculated for the six central base pairs. [b] TAKREL is the unique structure obtained by energy minimization without constraints of TakS1, TakS2, TakSE1, TakSE2, $TakSED_{nmr}$, $TakSED_{rec}$, $TakRAF_{nmr}$, and TakFIN. [c] PermutREL is the unique structure obtained by energy minimization without constraints of PermutS1, PermutS2, PermutSE1, $PermutSED_{nmr}$, $PermutSED_{rec}$, $PermutRA_{nmr}$ and PermutFIN.

(A) $d(GTACGTAC)_2$. The rms between starting and final structures range from 0.4 to 2.8 Å. Application of sugar-puckering constraints sets to the eight different starting structures resulted in two significantly distinct conformations (rms = 0.5 Å), presenting, however, essentially equal energies (−397 and −396.9 kcal/mol). For this oligonucleotide, addition of ϵ constraints (set SE) did not improve the structure convergence while the addition of distance constraints yielded a unique structure, namely, $TakSED_{nmr}$ or $TakSED_{rec}$ (rms = 0.24 Å), according to the type of constraints used, SED_{rec} or SED_{nmr} (depending on whether the d_s[H6/H8−H2″] was recalibrated or not). The distance constraint refinement further improved this convergence and allowed a gain of about 10−20% on the *R* factor, with an energy cost of only 4 kcal/mol (Table 2a). The structural parameters of the two final structures, TakFIN and $TakRAF_{nmr}$ were very close (rms = 0.22 Å), and only the structure with the lowest energy, TakFIN, will be further discussed.

(B) $d(CATCGATG)_2$. Again, two structures resulted from energy minimization of the eight starting structures under the sugar pseudorotation angles constraints. Although the rms difference between these two structures was only 0.3 Å, they differed significantly one from another in the twist values characterizing the central CpG step: 35° in PermutS1 vs 41° in PermutS2. A difference of ~12° in the phase angle of the G5 residue is attached to this twist difference. In contrast to the TAK oligonucleotide, subsequent application of ϵ and distance constraints promoted convergence to the rather similar structures $PermutSED_{rec}$ and $PermutSED_{nmr}$. For instance, these present almost identical central twist value (37°) and G5 phase angle (165°). Upon application of every constraint, it was shown that the internucleotide NOEs in

Table 3: Conformational Energy and *R*-Factor Values Obtained after Distance Refinement with a Force Constant of 1000 kcal mol^{-1} $Å^{-1}$ and a Constraint Set Containing the NMR-Derived d_s[H6/H8−H2″] Distances, for the TAK Oligonucleotide

constraint set	energy (kcal/mol)	*R* factor	Δ*E* (kcal/mol)
RAF_{nmr}	−321.1	0.14	+78

the central C4pG5 step and its next A6pT7 step, together with the phosphorus−proton coupling constants in the central CpG, were the main drivers of this convergence. The A6pT7 step appears implicated because of a correlation between the pseudorotation phases of the linked G5 and A6 residues, any change in the first entailing a modification in the second.

Back-calculation refinement of NOE distance constraints allowed further improvement of the *R*-factor value, with rather low energy cost (~5 kcal/mol), leading to $PermutRAF_{nmr}$ and PermutFIN. As in the case of TAK, these two final conformations presented very similar structural parameters (rms = 0.40 Å). The energy and *R*-factor excesses were once again due mainly to violation of the d_s[H6/H8−H2″] distance constraints. Here also, only the structure with the lowest energy, namely PermutFIN, will be retained, although it can be noted that the final *R* factor (0.37) in PermutFIN is not as low as the one obtained in TakFIN. The quality of spectra in terms of signal-to-noise ratio being rather similar for the two oligonucleotides, a higher residual factor found for PermutFIN may reflect our difficulty to account for the largest time-averaging effects affecting this compound. Perhaps, the $d(CATCGATG)_2$ duplex should be better described by an equilibrium of conformers, in contrast to $d(GTACGTAC)_2$.

Further Investigations Concerning d_s[H6/H8−H2″] Distances. We applied NMR derived d_s[H6/H8−H2″] constraints (2.7 Å in average) to the structure with crystallographically derived values (2.3 Å in average) to examine the real influence of these distances on the calculated structures. The force constant on distance constraints was varied until the NMR derived d_s[H6/H8−H2″] constraints were obtained. This required a force constant value as strong as 1000 kcal mol^{-1} Å. The energy and *R*-factor values for TAK with NMR-derived d_s[H6/H8−H2″] constraints and a 1000 kcal mol^{-1} Å force constant on distances, after the refinement procedure, are reported in Table 3. The good fit of the NMR-derived d_s[H6/H8−H2″] constraints entailed an excellent *R* factor, confirming that the highest *R* factor obtained with the standard force constant (12 kcal mol^{-1} $Å^{-1}$) is principally due to the bad agreement introduced by the H6/H8−H2″ internucleotide NOEs. Although the structures resulting from such a procedure present a good agreement with the experimental data, as evidenced by the low *R* factor, obtaining them required an important energy cost. The energy difference with respect to the totally unconstrained structure (78 kcal/mol) is undoubtly unacceptable according to our criteria (Mauffret et al., 1992). In addition, these structures appeared strongly distorted in view of the DNA structures commonly observed by X-ray crystallography (Sänger, 1984; Dickerson, 1994) and presented characteristics nearer to the A form, namely, a very negative value of X displacement (−3.8°), as well as small average rise and twist values (2.8 Å and 31.5°, respectively). Moreover, a noticeable bend (10°) centered at the CpG step was observed in the molecule. Such observations allowed us to reject these structures.

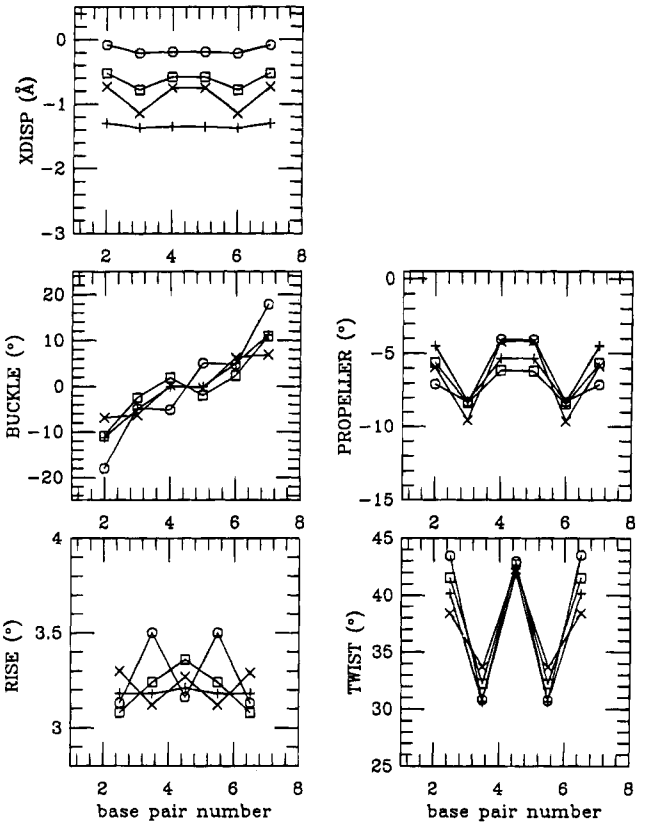

FIGURE 2: X-displacement, buckle, rise, propeller, and twist values at the different steps of the protocol for TAK: (□) TakS1, (+) $TakSED_{rec}$, (○) TakFIN, and (×) initial structures, calculated without constraints.

Evolution of the Helical Parameters for the Different Steps of Minimization. Measuring the helical parameters of the oligomers allowed us to understand the influence of the various constraints. Several of these parameters are reported in Figure 2 (TAK) and Figure 3 (PERMUT) for the successively derived structures. We selected one *axis−base pair parameter* (X displacement) , two *intrabase pair parameters* (buckle and propeller twist), and two *interbase pair parameters* (twist and rise).

(A) $d(GTACGTAC)_2$. One striking feature for this oligomer (Figure 2) was that the main characteristics (buckle, propeller twist, twist) of the final structure were already visible in the structures determined with the only sugar constraints. We noted also that the tilt and roll values were rather weak in all structures (not shown). The rise and X displacement values, in contrast, were rather weakly conserved along the different steps. If only the central tetramer ACGT is considered, we see that its final conformation expressed in terms of phase angle, twist, and ϵ torsion angles is very near one of the structures issued from the scan search procedure (SCAN1). For the four central base pairs, a rms deviation of 0.24 Å is found between SCAN1 and TakS2 and of 0.4 Å between SCAN1 and $TakRAF_{nmr}$ (where 0.25 Å is found between TakS2 and $TakRAF_{nmr}$).

(B) $d(CATCGATG)_2$. The selected helical parameters are presented in Figure 3 for $d(CATCGATG)_2$. They appear relatively well conserved along the different minimization steps, although less so than in the case of $d(GTACGTAC)_2$. For example, the twist profile of the central TCGA tetranucleotide is high−low−high in PermutS1 and PermutS2 structures and becomes medium−medium−medium in the

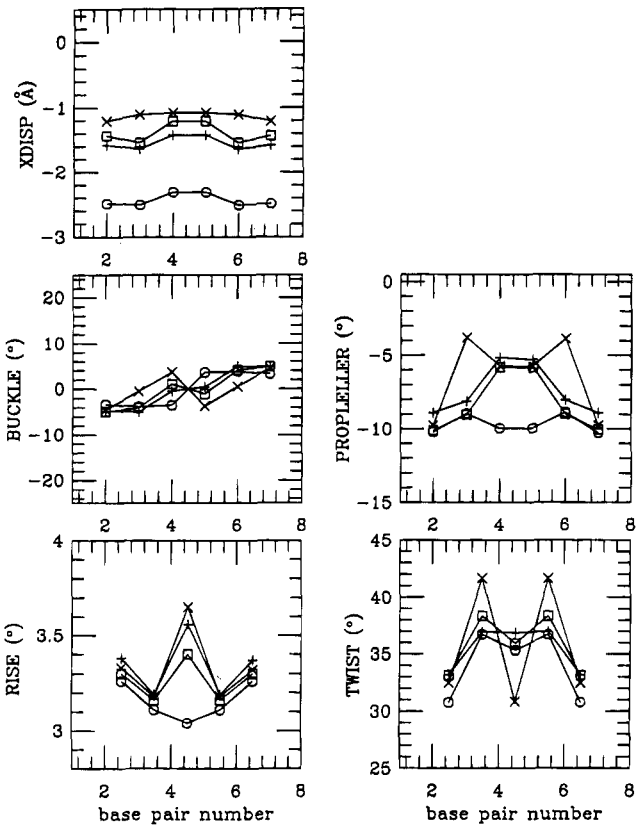

FIGURE 3: X-displacement, buckle, rise, propeller, and twist values at the different steps of the protocol for PERMUT: (□) PermutS1, (+) PermutSED$_{rec}$, (○) PermutFIN, and (×) initial structures, calculated without constraints.

final structure. This profile is in total contrast with that low−high−low displayed by the ACGT tetranucleotide in d(GTACGTAC)$_2$. Buckle (Figure 3) and tilt and roll (not shown) exhibit low values (4° maximum) for all structures and all residues. The X displacement, rise, and propeller twist values change very significantly after the refinement, in particular within the central tetramer TCGA. This is well reflected by the larger rms value (0.6−0.7 Å) found between PermutFIN and PermutSED*rec*.

DISCUSSION

The two oligonucleotides d(CATCGATG)$_2$ (PERMUT) and d(GTACGTAC)$_2$ (TAK) present the B-DNA structure, but with noticeable local differences in both the backbone geometry and in the base-pair arrangement of their central tetranucleotide, which houses the malleable CpG step. The main structural parameters are listed in (Table 4a for PERMUT and Table 4b for TAK).

Backbone Structure. In our approach, the only constraints applied to the backbone conformation corresponded to the ϵ angle derived from H3′−P coupling constants through the Lankhorst et al. (1984) curve. Other distances and angles that could be useful to restrain the backbone conformation (Kim & Reid, 1992a) were not used since they were not accurately measurable due to strong overlaps of corresponding cross peaks.

The calculated values for the backbone torsion angles in the PermutFIN and TakFIN structures, presented Table 4, indicate that the α angles lie in the −synclinal range (-64.8 ± 2° for TAK and −66.2 ± 3° for PERMUT), while most of the β angles are locked in the antiperiplanar domain (174.9 ± 5° for TAK and 176.8 ± 4° for PERMUT), and the γ angles are confined with a relatively narrow dispersion to the +synclinal region (60.9 ± 4° for TAK and 55.4 ± 3° for PERMUT). The ϵ angles are in the trans (−ap, trans) region and the ζ angles in the −anticlinal (−ac) region, both being characteristics of canonical B-DNA. All these values conform globally to the theoretical and experimental data found for oligodeoxyribonucleotide duplexes (Säenger, 1984; Dickerson, 1994; Wijmenga et al., 1993; Beveridge & Ravishanker, 1994).

Sugar Pucker. The δ-angle values, related to the deoxyribose endocyclic v torsion angles and thus to sugar puckering, are distributed in the +anticlinal (+ac) range (although very near to the +antiperiplanar domain), with a clear alternation of values along the sequence from purines to pyrimidines. These data are taken into account in the pseudorotational approach and will not be discussed individually.

For d(GTACGTAC)$_2$ (TAK), the pseudorotational angle, P, is found to be significantly different in purines and pyrimidines, although all the residues present a C2′-endo deoxyribose conformation, with P values near 175° for purines and in the range 140−165° for pyrimidines. Thus, a larger heterogeneity is observed on the P values of pyrimidines compared to purines.

For d(CATCGATG)$_2$ (PERMUT), P values lie over 165-180° for the purines and 140-150° for the pyrimidines. The range for purine P values is thus broader in d(CATCGATG)$_2$ (PERMUT) than in d(GTACGTAC)$_2$ (TAK). The present P values are consistent with most of the recently published P values [see for instance, Weisz et al. (1994), Chou et al. (1994), and Mujeeb et al. (1993)], but are larger than those reported by Baleja et al. (1990a) (i.e., 100−145°) for d(GTACGTAC)$_2$ (TAK). However, the results of Baleja et al. (1990a) were obtained from restrained molecular dynamics using the GROMOS force field. The present data confirm the tendency of the GROMOS force field to yield sugar structures with small phase angle values (Hartmann et al. 1995).

Helicoidal Parameters. As mentioned above, the two oligonucleotides display the B-DNA features. Due to the permutation of its flanking residues, the central CpG step presents a structure significantly different in TAK and PERMUT (twist, 42 vs 35°; roll : −1 vs 3.7° in TAK and PERMUT, respectively). Apart from these differences, no other salient distinction can be discerned between TAK and PERMUT for the helicoidal parameters of CpG.

We can compare the present TakFIN structure with the NMR structure calculated by Baleja et al. (1990a) for the same oligonucleotide. Despite the above-mentioned differences regarding the sugar conformations and a large rms of 0.8 Å, the CpG step adopts similar twist and roll values in the X-ray structure and the solution structure. Comparison of the TAK octamer with the CRE- (cAMP-responsive element) containing dodecamer, previously reported by us (Mauffret et al., 1992) shows that the central ACGT sequence adopts similar conformations in the two oligonucleotides. We find an rms deviation of 0.5 Å comparing this tetranucleotide and as low as 0.25 Å for the CpG step.

Comparison with Crystallographic Structures. Comparison of crystal structures with structures in solution is not straightforward as crystal packing forces play a significant role in modulating conformation. For example, crystallizing

Table 4: Structural Parameters[a]

	Xdisp	buckle	propel.		rise	tilt	roll	twist		phase	ampli	χ		ϵ	ζ
							(a) PERMUT								
A2-T10	−2.2	−3.3	−10.2	A2pT3	3.2	−0.2	−1.6	30.9	A2	178.2	37.4	−106.2	A2	−174.3	−104.3
T3-A11	−2.1	−3.9	−9.0	T3pC4	3.1	1.8	3.2	36.7	T3	139.0	38.4	−117.3	T3	−169.1	−104.9
C4-G12	−1.9	−3.5	−9.9	C4pG5	3.1	0.0	3.6	35.3	C4	148.4	39.9	−111.8	C4	−167.2	−114.2
G5-C13	−1.9	3.6	−9.9	G5pA6	3.1	−1.8	3.3	36.7	G5	164.9	42.0	−98.6	G5	−171.9	−124.5
A6-T14	−2.1	3.9	−8.9	A6pT7	3.2	0.2	−1.6	30.9	A6	167.0	35.6	−108.5	A6	−172.1	−107.1
T7-A15	−2.2	3.3	−10.3						T7	143.7	39.1	−115.8	T7	−172.7	−105.2
av	−1.8	0.0	−9.7		3.2	0.0	1.4	34.1		156.9	38.7	−109.7		−171.0	−111.0
							(b) TAK								
T2-A10	−0.3	−17.1	−6.2	T2pA3	3.1	1.4	1.5	43.5	T2	165.0	42.8	−114.9	T2	−170.7	−129.8
A3-T11	−0.3	−4.7	−7.3	A3pC4	3.5	−0.1	3.3	32.0	A3	179.3	40.4	−102.9	A3	−176.8	−112.6
C4-G12	−0.2	−4.2	−3.7	C4pG5	3.2	0.0	−0.9	42.6	C4	155.1	44.2	−115.1	C4	−163.2	−137.8
G5-C13	−0.2	4.2	−3.8	G5pT6	3.5	0.1	3.3	32.1	G5	180.1	36.9	−106.7	G5	−174.2	−107.8
T6-A14	−0.3	4.8	−7.3	T6pA7	3.1	−1.5	1.6	43.5	T6	150.1	40.4	−113.2	T6	−170.9	−124.6
A7-T15	−0.3	17.1	−6.2						A7	179.4	39.9	−100.3	A7	−180.4	−103.0
av	−0.3	0.0	−1.3		3.7	0.0	1.9	37.3		162.7	40.7	−108.9		−172.1	−116.7

[a] Translational parameters are in angstroms, and rotational parameters are in degrees.

the same duplexes in two different crystal forms causes significant changes in structural features (Lipanov et al., 1993). Nevertheless, in many instances, the X-ray crystallographic and calculation studies allowed the observation of similar sequence-specific structural variations (Dickerson, 1983; Grzeskowiak et al., 1991; Privé et al., 1991; Yanagi et al., 1991; Poncin et al., 1992; Beveridge & Ravishanker, 1994).

For example, the duplex d(GTACGTAC)$_2$ (TAK) has crystallized under the A form (Takusagawa, 1990), reflecting the packing force effects which prevent any comparison with our solution structures. On another hand, the oligonucleotide d(CCAACGTTGG)$_2$ adopts the B-DNA structure in the crystal (Privé et al., 1991; Yanagi et al., 1991) apparently without noticeable distortions arising from the packing forces. In this crystal structure, determined at 1.4 Å resolution, the cytosine and thymine sugars of the central ACGT sequence exhibit the O4′-endo conformation rather than of the C2′-endo conformation seen in the present study for TAK in solution. Despite these minor differences, the ACGT structure presents a similar low−high−low twist profile in the two environnements i.e., 30−45−30° and 32−42−32°, for crystal and solution, respectively.

For the central TCGA tetranucleotide of the d(CATC-GATG)$_2$ oligonucleotide (PERMUT), we referred to the X-ray structures of d(CGATCGATCG)$_2$ obtained at 1.5 Å (Grzeskowiak et al., 1991) and of d(CGATCGmeATCG)$_2$ at 2 Å (Baïkalov et al., 1993). The central tetranucleotide twist angles in these two crystal structures present a rather similar profile, namely, 39−30−39° and 36−35−35°, respectively, that resembles the 37−35−37° found for the PermutFIN solution structure.

The ^{31}P Chemical Shift as Conformational Parameter. It has been repeatedly shown that the ^{31}P chemical shifts provide a good probe of the phosphate ester backbone conformation in oligonucleotides (Gorenstein, 1994; El Antri et al., 1993a,b). In fact, as initially pointed out by Ott and Eckstein (1985a,b) and Schroeder et al. (1986, 1987), the ^{31}P chemical shift appears essentially to vary in response to local, sequence-specific, and induced environmental distortions in the double-helix conformation. We have reported previously that the ^{31}P chemical shifts of phosphate groups along the sequences of d(GTACGTAC)$_2$ and d(CATC-GATG)$_2$ displayed completely different profiles (El Antri et al., 1993). The chemical shift pattern was found strongly alternating for TAK and bell-shaped for PERMUT, reflecting thus the Pur/Pyr sequence effects on ^{31}P chemical shifts. It is noticeable that the TAK compound d(GTACGTAC)$_2$ presents (excluding the terminal 3′ and 5′ steps subject to fraying) a step succession, low field (TA), high field (AC), low field (CG), high field (GT), and low field (TA), that contrasts with the more gradual succession, high field (AT), low field (TC), low field (CG), low field (GA), and high field (AT), observed for PERMUT (El Antri et al., 1993a,b; Gorenstein, 1992).

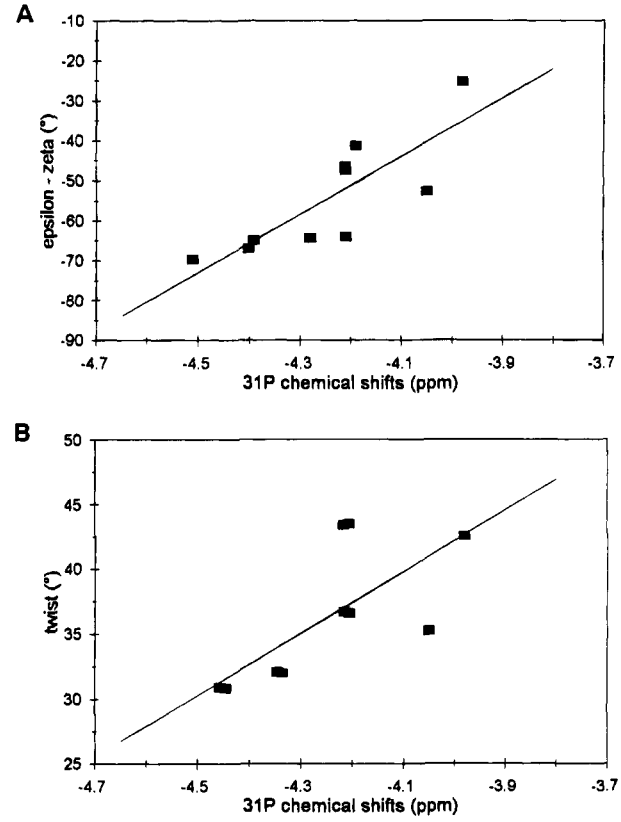

FIGURE 4: Correlation between (A) the ^{31}P chemical shifts and the calculated $\epsilon-\zeta$ values ($R = 0.82$) and (B) the ^{31}P chemical shifts and the calculated twist angle values ($R = 0.71$), for the nonterminal steps of the two oligonucleotides studied.

The exact conformational origin of the effects on ^{31}P chemical shifts is still questionable. Thus, correlation coefficients were evaluated for several structural parameters in the two oligonucleotides. For each selected parameter, 10 values corresponding to the five internal steps of the two oligonucleotides (thus excluding the terminal base pairs) were available for the analysis. As the oligonucleotides exhibit completely different conformational patterns, reflecting a rather good dispersion of values, we have a representative range of values at our disposal. As shown in Figures 4A and 4B, satisfying correlations with the ^{31}P chemical shifts were obtained for the helical twist and also for the difference $\epsilon-\zeta$, $R = 0.71$ and 0.82, respectively. It can be noted that, for the correlations with helical twist, we used the average chemical shift value of the phosphorus situated on each strand on both sides of the step considered, namely, P2/P6 and P3/P5. The classical correlation initially provided by crystal and calculated structures (Yanagi et al., 1991; Poncin et al., 1992; Heinemann et al., 1994) between twist and slide was also observed ($R = 0.96$). As the roll did not vary significantly within our two oligonucleotides, no attempt was made to correlate this parameter with the ^{31}P chemical shift.

All the above results and supplementary data from other oligonucleotide structures (Mauffret et al., 1992) conform to the previous findings of Gorenstein (1992).

CONCLUSION

The present study shows that the experimental restraints, namely, the interproton distances and torsion angles, are valuable parameters to be used as input in molecular calculations since they allow definition of a unique structure for each oligonucleotide. The agreement between the calculated structures and the experimental data, expressed in terms of R-factor and penalty energy, appears remarkably good. Due to the time-averaged nature of NMR data and the possible flexibility of molecules in solution, it was not obvious that all restraints would be satisfied simultaneously. Thus, the hypothesis that TAK presents a single major conformer in solution appears reasonable. On another hand, the weaker correspondence between the final structures and the experimental data found for PERMUT could indicate an equilibrium between several conformers or the existence of particular motions within the central part of this molecule.

Our results also implicate the sequence-dependent malleability of CpG and confirm the strong influence of flanking residues on the structure of this step. The alternating structure shown by the helix in the ACGT core of oligonucleotide TAK is reflected by both the ^{31}P chemical shift profile and the pattern of calculated backbone (ϵ) and helicoïdal (twist) parameters. There are data in the literature suggesting that the ACGT structure could be involved in specific recognition sites for DNA-binding proteins with gene regulatory functions. For example, ACGT is found in DNA functional domains, as those specifically recognized by the helix−turn−helix motif of the purine repressor LacI (Schumacher et al., 1994) and of the methionine repressor (Old et al., 1991). This tetranucleotide is also constitutive of several response elements recognized by the b-ZIP transcription factors found in yeast (GCN4) (Ellenberger et al., 1992), in mammals (CREB, c-Jun) (Gonzales et al., 1989; Halazonetis et al., 1988), or even in plants (EmBP-1) (Niu et Guiltinan, 1994).

On the other hand, we are tempted to speculate that the structural flexibility of CpG is correlated to its propensity to undergo mutations. For example, three of the major mutation hot spots detected in the factor VIII gene and responsible for hemophilia A curiously concern the TCGA sequence (codons 1941, 2209, and 2307) which in this work presents a higher flexibility than ACGT. Globally, mutations on TCGA elements represent ~36% of the CpG mutations on this gene (Tuddenham et al., 1994). It could be possible that in such elements CpG constitutes a better interaction site for drugs (Mauffret et al., 1991), mitogenic agents, and 5meC-methyltransferases.

ACKNOWLEDGMENT

The authors thank George Tevanian for very helpful advice concerning the NMR experiments and Lucie Dega for technical assistance.

REFERENCES

Arnott, S., & Hukins, D. W. L. (1973) *J. Mol. Biol. 81*, 93−105.

Baikalov, 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.

Baleja, J. D., Pon, R. T., & Sykes, B. D. (1990c) *Biochemistry 29*, 4828−4839.

Banks, K. M., Hare, D. R., & Reid, B. R. (1989) *Biochemistry 28*, 6696−7010.

Bax, A., & Davies, D. G. (1985) *J. Magn. Reson. 65*, 355−360.

Beveridge, D. L., & Ravishanker, G. (1994) *Curr. Opin. Struct. Biol. 4*, 246−255.

Boelens, R., Koning, T. M. G., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1989) *J. Magn. Reson. 82*, 290−308.

Borer, P. N., LaPlante, S. R., Kumar, A., Zanatta, N., Martin, A., Hakkinen, A., & Levy, G. C. (1994) *Biochemistry 33*, 2441−2450.

Borgias, B. A., & James, T. L. (1990) *J. Magn. Reson. 87*, 475−487.

Chou, S.-H., Cheng, J.-W., & Reid, B. R. (1992) *J. Mol. Biol. 228*, 138−155.

Chou, S.-H., Cheng, J.-W., Fedoroff, O., & Reid B. R. (1994) *J. Mol. Biol. 241*, 467−479.

Chuprina, V. P., Nerdal, W., Sletten, E., Poltev, V. I., & Fedoroff, O. Y. (1993) *J. Biomol. Struct. Dyn. 11*, 671−683.

Cooper , D. N., & Youssoufian, H. (1988) *Hum. Genet. 78*, 151−155.

Cuniasse, P., Sowers, L. C., Erijta, R., Kaplan, B., Goodman, M. F., Cognet, J., Le Bret, M., Guschlbauer, W., & Fazakerley, G. V. (1987) *Nucleic Acid Res. 15*, 8003−8022.

Dickerson, R. E. (1983) *J. Mol. Biol. 166*, 419−441.

Dickerson, R. E. (1994) in *Methods in Enzymology* (Lilley, D. M. J., & Dahlbery, J. F., Eds.) Vol. 211, pp 67−111, Academic Press, Inc., New York.

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.

Ellenberger, T. E., Brandl, C. J., Struhl, K., & Harrison, S. C. (1992) *Cell 71*, 1223−1237.

Fasman, G. D. (1975) In *Handbook of Biochemistry and Molecular Biology: Nucleic Acids, 3rd Ed.,* CRC Press, Cleveland, OH.

Fedorov, O.-Y., Reid, B. R., & Chuprina, V. P. (1994) *Biochemistry 33*, 325−330.
Gochin, M., Zon, G., & James, T. L. (1990) *Biochemistry 29*, 11161−11171.
Gonzales, G. A., Yamamoto, K. K., Fischer, W. H., Karr, D., Menzel, P., Biggs, W., Vale, W. W., & Montminy, N. R. (1989) *Nature 337*, 749−752.
Gorenstein, D. G. (1992) in *Methods in Enzymology* (Lilley, D. M. J. and Dahlberg, J. E., Eds.) Vol. 211, pp 254−285, Academic Press, Inc., New York.
Gorenstein, D. G. (1994) *Chem. Rev. 94*, 1315−1338.
Gorenstein, D. G., Schroeder, S. A., Fu, J. M., Metz, J. T., Roongta, V. A., & Jones, C. R. (1988) *Biochemistry 27*, 7223−7237.
Gray, D. M., Ratliff, R. L., & Vaughan, M. R. (1992) in *Methods In Enzymology* (Lilley, D. M. J., & Dahlberg, J. E., Eds.) Vol. 211, pp 389−408, Academic Press, Inc., New York.
Gronenborn, A. M., & Clore G. M. (1985) *Prog. Nucl. Magn. Reson. Spectrosc. 17*, 1−32.
Grzeskowiak, K., Yanagi, K., Privé, G. G., & Dickerson, R. E. (1991) *J. Biol. Chem. 266*, 8861−8883.
Halazonetis, T. D., Georgopoulos, K., Greenberg, N. E., & Leder, P. (1988) *Cell 55*, 917−924.
Hartmann, B., Sun, J.-S., & Lavery R. (1995) In *Architecture and Activity of Nucleic Acids* (Lavery, R., and Westhof, E., Eds.) in press.
Heinemann, U., Alings, C., & Hahn, M. (1994) *Biophys. Chem. 50*, 157−167.
Hingerty, B., Richie, R. H., Ferrel, T. L., & Turner, J. E. (1985) *Biopolymers 24*, 427−439.
James, T. L. (1994) *Curr. Opin. Struct. Biol. 4*, 275−284.
Kessler, H., Griesinger, C., Zarbock, J., & Loosli, H. R. (1984) *J. Magn. Reson. 57*, 331−336.
Kim, S.-G., & Reid, B. R. (1992a) *Biochemistry 31*, 12103−12116.
Kim, S.-G., & Reid, B. R. (1992b) *J. Magn. Reson. 100*, 382−390.
Kim, S.-G., Lin, L.-J., & Reid, B. R. (1992) *Biochemistry 31*, 3564−3574.
König, P., & Richmond, T. J. (1993) *J. Mol. Biol. 233*, 139−154.
Koning, T. M. G., Boelens, R., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1991) *Biochemistry 30*, 3787−3797.
Laird, P. W., & Jaenisch, R. (1994) *Hum. Mol. Genet. 3*, 1487−1495.
Lane, A. N. (1990) *Biochim. Biophys. Acta 1049*, 189−204.
Lankhorst, P. P., Haasnoot, C. A. G., Erkelens, C., & Altona, C. (1984) *J. Biomol. Struct. Dyn. 1*, 1387−1399.
Lavery, R. (1988) in *Structure and expression* (Olson, W. K., Sarma, R. H., & Sundaralingam, M., Eds.) *Vol. 3*, 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., & Hartmann, B. (1993) *Biophys. Chem. 50*, 33−45.
Lipanov, A., Kopka, M. L., Kaczor-Grzeskowiak, M., Quintana, J., & Dickerson, R. E. (1993) *Biochemistry 32*, 1373−1389.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Mauffret, O., René, B., Convert, O., Monnot, M., Lescot, E., & Fermandjian, S. (1991) *Biopolymers 31*, 1325−1341.
Mauffret, O., Hartmann, B., Convert, O., Lavery, R., & Fermandjian, S. (1992) *J. Mol. Biol. 227*, 852−875.
Metzler, W. J., Wang, C., Kitchen, D. B., Levy, R. M., & Pardi, A. (1990) *J. Mol. Biol. 214*, 711−736.
Mujeeb, A., Kerwin, S. M., Kenyon, G. L., & James, T. L. (1993) *Biochemistry 32*, 13419−13431.
Nikonowicz, E. P., & Gorenstein, D. G. (1990) *Biochemistry 29*, 8845−8858.
Niu, K., & Guiltinan, M. (1994) *Nucleic Acid res. 22*, 4969−4978.
Old, I. G., Phillips, S. E. V., Smith, P. C., & Saint-Girons, I. (1991) *Prog. Biophys. Mol. Biol. 56*, 145−184.
Ott, J., & Eckstein, F. (1985a) *Biochemistry 24*, 2530−2535.
Ott, J., & Eckstein, F. (1985b) *Nucleic Acid Res. 13*, 6317−6330.
Pearlman, D. A. (1994) *J. Biomol. NMR 4*, 1−16.
Pearlman, D. A., & Kollmann, P. A. (1991) *J. Mol. Biol. 220*, 457−479.
Poncin, M., Hartmann, B., & Lavery, R. (1992) *J. Mol. Biol. 226*, 775−794.
Post, C. B., Meadows, R. P., & Gorenstein, D. G. (1990) *J. Am. Chem. Soc. 112*, 6796−6803.
Privé, G. G., Yanagi, K., & Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 177−199.
Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621−649.
Roongta, V. A., Jones, R., & Gorenstein, D. G. (1990) *Biochemistry 29*, 5245−5258.
Säenger, W. (1984) *Principles of Nucleic Acids Structures*, Springer-Verlag, New York.
Schmitz, U., Ulyanov, N. B., Kumar, A., & James, T. L. (1993) *J. Mol. Biol. 234*, 373−389.
Schroeder, S., Jones, C., Fu, J., & Gorenstein, D. G. (1986) *Bull. Magn. Reson. 8*, 137−146.
Schroeder, S., Fu, J., Jones, C., & Gorenstein, D. G. (1987) *Biochemistry 26*, 3812−3821.
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, in press.*
Taillandier, E., & Liquier, J. (1992) in *Methods in Enzymology* (Lilley, D. M. J., & Dahlberg, J. E., Eds.) Vol. 211, Section III, pp 307−334, Academic Press, New York.
Takusagawa, F. (1990) *J. Biomol. Struct. Dyn. 7*, 795−811.
Tropp, J. (1980) *J. Chem. Phys. 72*, 6035−6043.
Tropp, J., & Redfield, A. G. (1981) *Biochemistry 20*, 2133−2140.
Tuddenham, E. G. D., Schwaab, R., Seehafer, J., Millar, D. S., Gitschier, J., Higuchi, M., Bidichandani, S., Connor, J. M., Hoyer, L. W., Yoshioka, 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.
Ulyanov, N. B., Gorin, A. A., Zhurkin, V. B., Chen, B.-C., Sarma, M. H., & Sarma, R. H. (1992) *Biochemistry 31*, 3918−3930.
Ulyanov, N. B., Schmitz, U., & James, T. L. (1993) *J. Biomol. NMR 3*, 547−568.
Van de Ven, J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1−38.
Wang, K. Y., Borer, P. N., Levy, G. C., & Pelczer, I. (1992) *J. Magn. Reson. 96*, 165−170.
Weisz, K., Shafer, R. H., Egan, W., & James, T. L. (1992) *Biochemistry 31*, 7477−7487.
Weisz, K., Shafer, R. H., Egan, W., & James, T. L. (1994) *Biochemistry 33*, 354−366.
Widmer, H., & Wüthrich, K. (1986) *J. Magn. Reson. 70*, 270−279.
Widmer, H., & Wüthrich, K. (1987) *J. Magn. Reson. 74*, 316−336.
Wijmenga, S. S., Mooren, M. M. W., & Hilbers, C. W. (1993) in *NMR of Macromolecules* (Roberts, G. C., Ed.) pp 217−288, IRL Press, Oxford, U.K.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
Yanagi, K., Privé, G. G., & Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 201−214.

BI9508505