Solution Structure of a Homopyrimidine:Homopurine Dodecamer Encoded by the HIV-1 Envelope Gene: NMR and Molecular Simulation Studies†

Patrick Sodano,[§,||] Brigitte Hartmann,[⊥] Thierry Rose,[||] Simon Wain-Hobson,[#] and Muriel Delepierre*[,||]

Laboratoire de Résonance Magnétique Nucléaire, CNRS URA 1129, Unité de Rétrovirologie Moléculaire, Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex 15, France, and Laboratoire de Biochimie Théorique, CNRS URA 77, Institut de Biologie Physico-chimique, 13 rue Pierre et Marie Curie, 75005 Paris France

Received November 10, 1994; Revised Manuscript Received February 7, 1995[⊗]

ABSTRACT: The solution structure of the nonpalindromic dodecanucleotide homopyrimidine:homopurine, d(5′-TTTCTCCTTTCT):d(5′-AGAAAGGAGAAA), was determined by two-dimensional nuclear magnetic resonance spectroscopy combined with molecular simulation. The dodecamer sequence studied was found within the HIV-1 enveloppe sequence and had all four Gs substituted by A in two hypermutants. A set of low-energy B-DNA conformations satisfying the quantitative NOE data were obtained. These highly related structures had neither peculiar helical parameters for the base pairs nor axis curvature. Analysis of the dihedral angles ($\epsilon-\zeta$) suggests that the A stretches flanking the GpA dinucleotides were more flexible.

The replication of retroviruses occurs in the virtual absence of any polymerization proof reading. While base substitutions dominate the mutation spectrum, occasionally major genetic changes may occur. One of the most remarkable is G⟹A hypermutation where between 10 and 60% of hundreds of G residues may be monotonously substituted by A (Vartanian et al., 1991, 1994). G⟹A hypermutation arises when reverse transcription occurs in the presence of highly asymmetric dNTP concentrations (Martinez et al., 1994). While the phenomenon is more frequent during RNA-dependent DNA synthesis, it may occur during DNA-dependent DNA synthesis. Although G⟹A hypermutation may affect many retroviruses most data has been accumulated for human immunodeficiency virus type 1 (HIV-1).[1] While hypermutation of any region of the genome has now been reported, the original study described concerned itself with a region from the viral envelope gene (Vartanian et al., 1991).

In vivo G⟹A substitutions are preferentially located within oligopurine regions suggesting that the structure of such sequences might be conducive to efficient formation of, or elongation beyond, multiple G:T mismatches produced by asymmetric dCTP/dTTP concentrations. Substitutions within the GpG dinucleotide were suggested to occur by base mispairing, while those within a GpA context by strand slippage by a mechanism termed displacement mutagenesis (Kunkel & Alexander, 1986).

During the last decade NMR has proven to be a powerful tool for solution structure determination of proteins and nucleic acids [see Wüthrich (1986), Clore and Gronenborn (1987, 1989), and van de Ven and Hilbers (1988) for reviews]. Interproton distances for those spatially closed are extracted from 2D NOESY experiments, while torsion angles may be deduced from COSY type experiments. This has allowed the determination of a number of oligonucleotide duplexes; however, a homopurine:homopyrimidine has not been reported. As the purification of chemically synthesized oligoribonucleotides proved particularly difficult it was decided to start with an oligodeoxypyrimidine:oligodeoxypurine duplex as fewer problems were encountered with oligodeoxypurine purification. The dodecamer sequence studied, d(AGAAAGGAGAAA), was found within the HIV-1 envelope sequence. Two hypermutants had all four Gs substituted by A.

MATERIALS AND METHODS

Sample Preparation. The dodecamers d(AGAAAGGAGAAA) and d(TTTCTCCTTTCT) were synthesized on an Applied Biosystems DNA Synthesizer following standard procedures. Each separate strand was chromatographed over Chelex 100 resin to remove dicationic ions and then dissolved in either 0.35 mL of D_2O or H_2O solution containing 14 mM NaCl, 13 mM $MgCl_2$, 100 mM KCl, and 0.05 mM EDTA at pH 7.0 to a final concentration of 2.4 mM. The absorbance at 260 nm was measured for each strand so that they could be mixed in equal concentration to minimize the amount of single strands present. The complementary

† This work was supported by grants from the Institut Pasteur, the Agence Nationale pour la Recherche sur le SIDA, and the Centre National de la Recherche Scientifique. P.S. acknowledges the Agence Nationale pour la Recherche sur le SIDA for financial support.

* To whom correspondence should be addressed.

§ Present address: Centre de Biophysique Moléculaire, 1A avenue de la Recherche Scientifique, 45071 Orléans France.

|| Laboratoire de Résonance Magnétique Nucléaire, CNRS, Institut Pasteur.

⊥ Laboratoire de Biochimie Théorique, CNRS.

Institut Pasteur.

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

[1] Abbreviations: HIV-1, human immunodeficiency virus type 1; FID, free induction decay; 1D, 2D NMR, one- and two-dimensional nuclear magnetic resonance spectroscopy; 2Q-COSY, double quantum-filtered correlated spectroscopy; TOCSY, two-dimensional correlated spectroscopy; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; di(X;Y) intraresidue distances between protons X and Y; ds(X;Y) sequential distance between protons X and Y where X is in the 5′ direction relative to Y; rms, root mean square; rmsd, root mean square deviation of atomic coordinates between two sets; BCK-NOESY, back-calculation of NOESY spectra; ϵ and ζ, dihedral angles of backbone defined by atoms (C4′-C3′-O3′-P) and (C3′-O3′-P-O5′), respectively. The bases of the pyrimidine and purine dodecamers are numbered from 1 to 12 and from 13 to 24, respectively, both from the 5′ to 3′ direction.

0006-2960/95/0434-6900$09.00/0

strands were annealed by incubation in a water bath at 80 °C for 10 min and then allowed to cool slowly. The final sample concentration was 1.2 mM at pH 7.0.

NMR Spectroscopy. NMR experiments were carried out on a Varian Unity 500 spectrometer and processed on a Sun workstation. Data processing was carried out with the VNMR 4.1 program.

Two-dimensional data sets for 2Q COSY, NOESY and TOCSY spectra were collected in the phase-sensitive mode with the hypercomplex scheme (States et al., 1982) and a 1.9 s recycle delay. T_1 relaxation experiments gave T_1 values less than 1.8 s for the slowest proton to relax with the exception of adenine H2 aromatic protons which have much longer relaxation time, that is, between 3 and 4 s. The sample temperature was adjusted to 15 °C for all experiments. Two TOCSY experiments using MLEV17 (Bax & Davies, 1985; Griesenger et al., 1988) to produce isotropic mixing were carried out with mixing times of 80 and 100 ms. A period of 35 ms was used for double-quantum excitation in the 2Q-COSY.

For all NOESY experiments 700 complex pairs were collected in the t_1 dimension and 2048 data points in the t_2 dimension. Two mixing times, 150 and 500 ms, were chosen for assignment with 32 scans for each t_1 value. Most of the cross-peaks had higher intensities in the 500 ms spectrum compared to the 150 ms one. Moreover, the 500 ms spectrum showed a greater number of peaks due to spin diffusion. Hence, the NOESY spectrum obtained with the 500 ms mixing time was used for assignment purposes as a particularity of DNA is that spin diffusion effects are always in the direction of the assignment pathway. Indeed, the cross-relaxation effect is most effective around the deoxyribose rings within individual nucleotides and additional sequential cross-relaxation is in the 5′ to 3′ direction (Wüthrich, 1986).

For the structure calculations, six NOESY spectra were collected successively in a single period, with mixing times of 40, 70, 100, 150, 200, and 250 ms and 64 scans for each t_1 value. These spectra were then zero-filled to a final matrix of 2048 × 4096 points. Prior to a Fourier transform, the matrices were multiplied by a shifted sine-bell window function for resolution enhancement. The volume integral of each resolved NOE cross-peak was estimated by summation of the intensity in an area around the center of the cross-peak. The distances between protons were estimated using the extrapolation method (Baleja et al., 1990a–c, 1991). First, to convert NOE buildup rates into distances, the two-spin approximation was used. The distances were derived from the volume intensity for each NOESY mixing time using the equation $V_i/V_j = (R_j/R_i)^6$. A cytosine H5/H6 interproton distance of 2.45 Å was used as the reference distance for all NOEs involving nonexchangeable protons, with the exception of methyl groups which were discarded for structure calculations. Each derived distance was plotted against mixing time. The extrapolation to zero mixing time gives a distance which can be used as input for structure calculations. This method ensures a first-order correction for spin diffusion effects.

Molecular Simulation. Molecular modeling has been carried out with version 7.0 of JUMNA (Junction Minimization of Nucleic Acids: Lavery, 1988; Ramstein & Lavery, 1988; Hartmann et al., 1989, 1993; Poncin et al., 1992; Zakrzewska, 1992; Mauffret et al., 1992) on a Silicon Graphics Crimson work station. Details of the force field have been published elsewhere (Lavery et al., 1984, 1986a,b). Phosphorus net charge was fixed at −0.5e to mimic counterion screening. Dielectric effects were computed using a sigmoidal distance-dependent function (Hingerty et al., 1985; Lavery et al., 1986a,b). The JUMNA algorithm uses the helicoidal variables (helical parameters and backbone dihedrals) to describe the nucleic acid structure during minimization. This approach reduces the number of variables to be treated by a factor of 10 compared to Cartesian coordinate molecular mechanics. Furthermore, it facilitates energy minimization, diminishes the number of local minima, and increases the amplitude of the conformational changes which can be obtained. Similar results using Cartesian coordinate representations can only be achieved with molecular dynamic simulations which are much more costly. Having reached a conformation compatible with the experimental constraints, additional options in JUMNA allow the study of controlled deformations and the generation of 1D or 2D adiabatic maps to explore the conformational space around the optimal conformation (Mauffret et al., 1992; Hartmann et al., 1993). Such studies of flexibility, in common with constrained molecular dynamic simulations, give an idea of the range of conformations compatible with the experimental constraints. However, these estimates are necessarily dependent on the quality of the force field employed and on the modeling of the environment surrounding the molecule studied.

Comparisons of oligomer conformations are made using CURVES which calculates the optimal helical axis and complete set of helicoidal parameters (Lavery & Sklenar 1988, 1989). These parameters obey the Cambridge conventions (Dickerson et al., 1989).

Modeling Strategy. NMR measurements indicate that the conformation of the oligomer studied belonged to the B-DNA family. A range of B-type conformations was thus firstly chosen as starting points for modeling: (i) two fiber diffraction models (Arnott et al., 1973, 1980), (ii) five polypurine:polypyrimidine models already obtained (Poncin et al., 1992) by an extended search of the conformational hypersurface which led to a set of conformational substates with similar stabilities but considerably different helical and backbone parameters, and (iii) four starting points of minimal energy obtained from mapping the sugar puckers of the oligomer studied while maintaining mononucleotide symmetry within each strand. This strategy seems justified since the NMR data exhibit very homogeneous H1′–H4′ and H1′–H3′ distances within each strand. Surface energy mapping techniques were already used (Lavery & Hartmann, 1994) to investigate the sugar conformational space and to support our view that sugars play a fundamental role in determining the local conformation of the double helix. All these starting points differed from one another by rmsd values of 0.8–2.4 Å. They have been energy minimized without geometrical constraints and then minimized again introducing the NOESY distance constraints to yield the final structures, which are compared and discussed below.

Lastly, we performed minimizations starting with A-DNA models and, for this particular case, a complete set of distance constraints in which very large backbone angles bounds were included (a range of 60° around the canonical values for α:g$^-$, β:t, γ:g$^+$, ϵ:t and ζ:g$^-$). The main function of these additional torsion constraints is to prevent different crankshaft

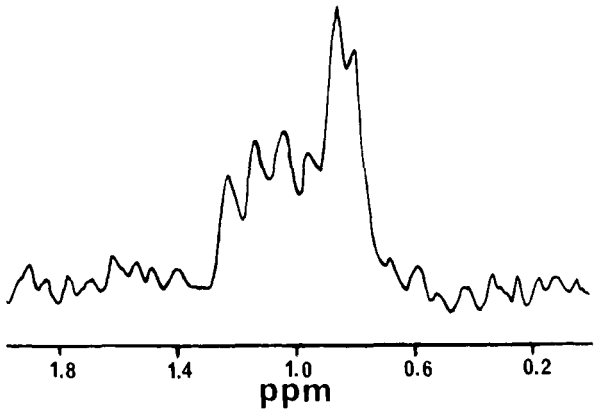

FIGURE 1: ^{31}P spectrum of the duplex at 25 °C.

transitions occuring during minimization. Thus, A-form starting points test the ability of the methodology to converge under restraints to a B-structure directly pointed out by the experimental data.

The experimental conditions employed did not allow us to obtain more than 180 proton−proton distances. A total of 149 distances were kept after removing NMR data referring to the 3′ and 5′ terminal (nucleotides 1, 12, 13, and 24). The consistency of NMR data was systematically checked with respect to proton−proton distance plots calculated for different sugar puckers and glycosidic angles using nucleotides having the standard bond length and angle geometries employed by JUMNA. These plots were very similar to those published by Wüthrich (1986).

NOESY Back-Calculation. Back-calculations of NOESY spectra were computed using the iterative relaxation matrix approach (IRMA) procedure (NMRchitect software package, 2.3 release, Biosym Technologies, San Diego). IRMA is a method based on NMR relaxation theory which generates a refined list of NOE distance restraints from a model structure and a generally incomplete set of NMR data (Boelens et al., 1988, 1989; Koning et al., 1991). A new matrix for relaxation rates was built from the current structure, and the theoretical matrix of NOE intensities was calculated. An average was then calculated from both theoretical and experimental matricies of NOE intensities. A new restraint file was deduced and used to constrain a new molecular simulation. IRMA progressively refines the NMR restraints, generating gradually improved molecular conformations. This iterative process is terminated when convergence is achieved, that is, when the *R* factor fails to drop significantly over several iterations (Gonzales et al., 1991) usually after 2−4 steps.

RESULTS AND DISCUSSION

Assignment of Nonexchangeable Protons. Assignment of individual resonances followed well-defined strategies, namely, assignment of individual spin systems for each nucleotide unit through their scalar interactions (TOCSY, 2Q-COSY) and thereafter sequential identification of the units through dipole−dipole interactions (NOESY). The first step is conformation independent while the second step is not, rendering it necessary to define a conformational type. A straightforward way to do this is to run a ^{31}P spectrum. In the ^{31}P spectrum all the resonances are clustered in a very narrow bandwidth less than 0.65 ppm at 25 °C (Figure 1), characteristic of a right-handed double-stranded helix and ruling out the presence of Z-DNA. The absence, at short mixing times, of base to H3′ sugar proton NOE signals rules out right-handed A type DNA, and the presence of strong intraresidue di(6,8; 2′) and intrastrand sequential ds(2″; 6,8) cross-peaks argues in favor of a regular B-DNA helix. Hence, proton resonances were next assigned by following the procedures developed for the sequential assignment of right-handed helices.

The individual spin systems of the deoxyribose moieties have been identified using clean TOCSY in D_2O (Griesenger et al., 1988). The problem of overlapping resonances in the H2′ and H2″ region was overcome by employing the remote connectivities of the H1′ protons to the H2′/H2″ protons observed in the double quantum spectrum acquired in D_2O. Indeed, a connectivity corresponding to the $\omega_{H2'} + \omega_{H2''}$ frequency can be observed at each H1′ position. Most of the connectivities between H3′ and H1′ protons were observed in the TOCSY experiment due to transfer from H1′ to H3′. These assignments were confirmed by the use of a NOESY experiment recorded with a mixing time of 500 ms. At this mixing time, all H6 or H8 base protons exhibit NOEs to two H3′ protons corresponding to the intraresidue di(H6,H8;H3′) and interresidue ds(H3′;H6,H8) connectivities. H4′ protons were assigned from both direct connectivities with H3′ protons and relayed transfer from H1′ in either the TOCSY experiment or the NOESY experiment run at long mixing time. The H5 and H6 base protons of the four cytosines and the H6 and methyl protons of the eight thymines were easily identified by means of the TOCSY experiment. The H2 base protons of the eight adenines were identified using an inversion−recovery experiment (Assa-Munt & Kearns, 1984).

As the double-stranded dodecanucleotide studied was composed of a homopurine strand and a homopyrimidine strand, the sequential resonance assignment strategy of the nucleotide protons differed from one strand to the other. While the homopurine strand assignment followed the usual procedure, the homopyrimidine strand required extensive use of spin diffusion effects. However, as mentioned above, spin diffusion cannot be a serious handicap and was very useful to observe long-range NOEs.

Assignment of the Homopyrimidine Strand. At the outset, the chemical shifts of the four cytosine H5 and H6 resonances, as well as the eight thymine methyl resonances, were identified from TOCSY. They represent excellent starting points for the sequential assignment procedure. In the next step, the intranucleotide and internucleotide NOEs between the base and sugar protons in the NOESY spectrum were used to assign the spin networks to individual nucleotides. In addition to the typical sequential correlations ds [d1 (base)i; (base)i+1], d2 [(H1′)i−1; (base)i], d3 [(H2′,H2″)i−1; (base)i], d4 [(H3′)i−1; (base)i] and the two typical intraresidue correlations di(H1′;H6) or di(H2′/H2″;H6), the following sequential and intraresidue correlation ds(H5;H5), ds(H6;H5), ds(H1′;H5), and di(H5;H1′) were initially carefully analyzed in the 500 ms spectrum. Although for B-DNA the ds(H5;H5) interproton distance is quite large, about 5 Å, these connectivities could be observed in the NOESY experiment using mixing time as long as 500 ms.

Simultaneous use of all the connectivities mentioned above made the assignment procedure much easier than with the usual ds(H1′;H6) and di(H1′;H6) connectivities. This was

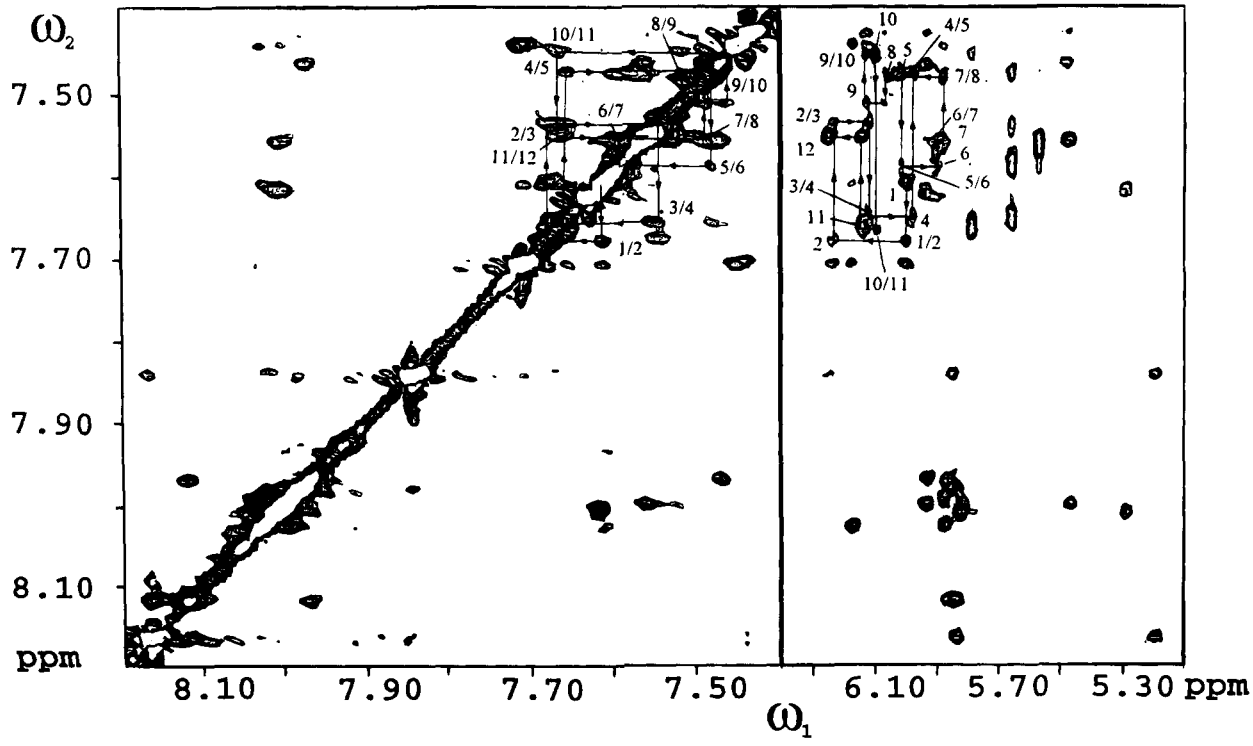

FIGURE 2: NOESY spectrum of the dodecamer in D_2O recorded at 15 °C with a mixing time of 500 ms and showing ds(H5;H5), ds(H6;H5), ds(H1′;H5), and di(H5;H1′) connectivities. Peaks labeled with a single number correspond to intraresidue cross-peaks while peaks labeled with two numbers correspond to sequential interresidues cross-peaks.

quite convenient as the base protons H6 resonate in the same region as the H8 purine base proton, while base proton H5/methyl resonate each in a distinct region allowing a better spreading of the information. As shown on Figure 2, use of only H1′;H6 connectivities would have rendered the assignment impossible.

The assignment procedure of the homopyrimidine residues was initiated by analysis of the methyl group region for methyl−methyl ds(H5;H5) interactions, which readily revealed the two stretches of three sequential thymines (T1−3 and T8−10). The two stretches could be distinguished via sequential interactions with the cytosine base protons. The T8T9T10 stretch is surrounded by two cytosine residues C7 and C11 while the T1T2T3 stretch is adjacent to cytosine C4. There are six CpT or TpC dinucleotide units in the sequence which can be found by matching the observed effect between thymine methyl groups and either the cytosine H6 or H5 base proton regions. Thus, T3C4, C7T8, T10C11, and C11T12 were unambiguously assigned. The unfortunate overlap of C4- and C6-H5 base protons precluded assignment of T5 via this procedure. However, as seven out of the eight thymines were assigned, the remaining thymine methyl group was assigned to T5. All cytosine H6/H5 assignments, with the exception of C4 and C6, were confirmed from the TOCSY experiment via their through bond interactions.

Then the assignment proceeded using the ds(H6;H5) and di(H6;H5) dipolar interactions, the latter being deduced from through bond connectivities. Starting from the T1 intraresidue connectivity di(H6;H5) the connection with T2 was established via the sequential NOESY cross-peak ds(H6;H5). Figure 3 shows the NOE network in solid lines. The assignment was straightforward up to residue C4. The observation of a ds(H6;CH3) connectivity between the T5 methyl group and a H6 base proton led to the assignment of this H6 base proton to C4. Simultaneous observation of ds(H6;H5) between residues T5 and C6 and ds(H6;H5) connectivities between C6 and C7 removed the ambiguity concerning C4 and C6 assignments. From here on, the connectivity scheme di(H6;H5) followed by ds(H6;H5) could be applied and pursued to the 3′ nucleotide T12.

The sequential assignment procedure for the H1′ deoxyribose protons relied on the intraresidue di(H1′;H5) and sequential ds(H1′;H5) connectivities. In B-DNA for XpC dinucleotides, the sequential distance ds(H1′;H5) is about 4.3 Å while the intraresidue distance is approximately 6.2 Å. Although, these distances are relatively long, using a mixing time of 500 ms for the NOESY experiment allowed observation of both connectivities (Figure 4). Self-connectivities, which represent the longest distances, were easily distinguished from the sequential connectivities by comparing the 500 ms NOESY spectrum with a spectrum obtained with a shorter mixing time (150 ms). Only sequential connectivities which correspond to the shortest distances can then be seen. On this basis, the H1′ proton assignment was straightforward. Starting from the T2 nucleotide, a through space interaction can be seen between its own methyl group and an H1′ proton in the 150 ms NOESY spectrum which is assigned, from the above considerations, to the H1′ proton of the T1 deoxyribose moiety. From this starting point and following the above procedure, all H1′ sugar protons were assigned by concomitant use of the 500 ms mixing time NOESY for di(H1′;H5) and the 150 ms mixing time NOESY spectrum for ds(H1′;H5).

Sequential Assignment of the Homopurine Strand. The eight adenine H2 base protons were located in the spectrum via measurement of the longitudinal relaxation time T_1. Since these protons are not scalar coupled to any nonexchangeable protons, they have characteristically long T_1 values compared to all other protons in the duplex. Therefore, spectra containing only these protons could be obtained, providing

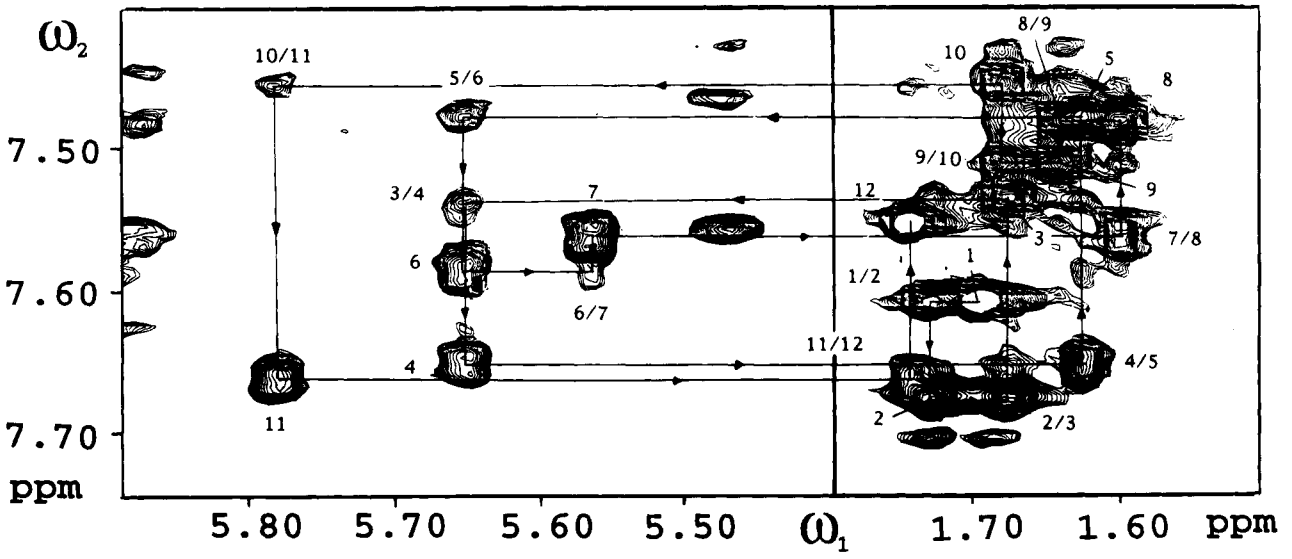

FIGURE 3: NOESY spectrum of the dodecamer in D_2O recorded at 15 °C with a mixing time of 500 ms and showing ds(H6;H5) and di(H6;H5) dipolar interactions in the homopyrimidine strand. The sequential assignment is depicted in solid line.

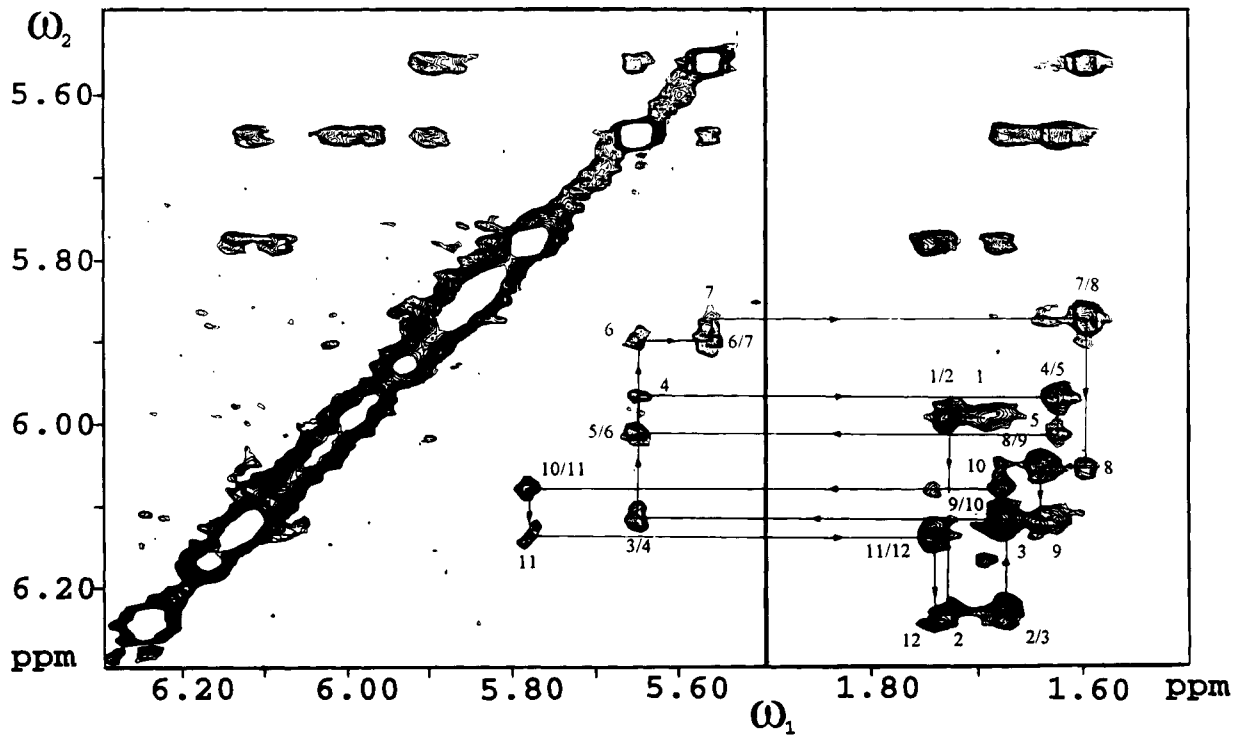

FIGURE 4: NOESY spectrum of the dodecamer in D_2O recorded at 15 °C with a mixing time of 500 ms and showing di(H1′;H5) and sequential ds(H1′;H5) connectivities for the homopyrimidine strand.

the appropriate delay was chosen. These protons were assigned using two different approaches. First, for the two stretches of three sequential adenines, the H2 base protons were identified through direct dipolar connectivities due to strong sequential intrastrand interactions between these protons (this distance is approximately 3.6 Å). For adenines, it has been shown previously that the H2 base proton can give rise to three types of NOE with sugar H1′ protons, two intrastrand resonances, that is, with its own H1′ and with the H1′ of the 3′ flanking residue and one interstrand resonance (Grütter et al., 1988). The intensity of these NOEs can show important variations as a consequence of sequence-dependent variations of the local conformation along the duplex (Chuprina et al., 1991, 1993). However, since H1′-sugar protons of the homopyrimidine strand were unambiguously assigned, from A24 the self di(H1′;H2) connection led to a unique solution for the assignment of the adenine H2 proton based on interstrand NOE interactions. These assignments confirmed the sequential assignment and showed that the molecule is in a double-stranded helical state under the present experimental conditions. From here on, the sequential assignment procedure followed the usual scheme based on di(H1′;H8) and ds(H1′;H8) connectivities. Such a connectivity pathway is shown in Figure 5 and could be traced all the way down the purine strand. Each adenine H1′ sugar proton assignment was corroborated by the observation of an intranucleotide connectivity di(H1′;H2) as well as the sequential ds(H2; H1′) connectivity. Finally, the proposed assignment was cross-checked using the sequential base-to-base interactions.

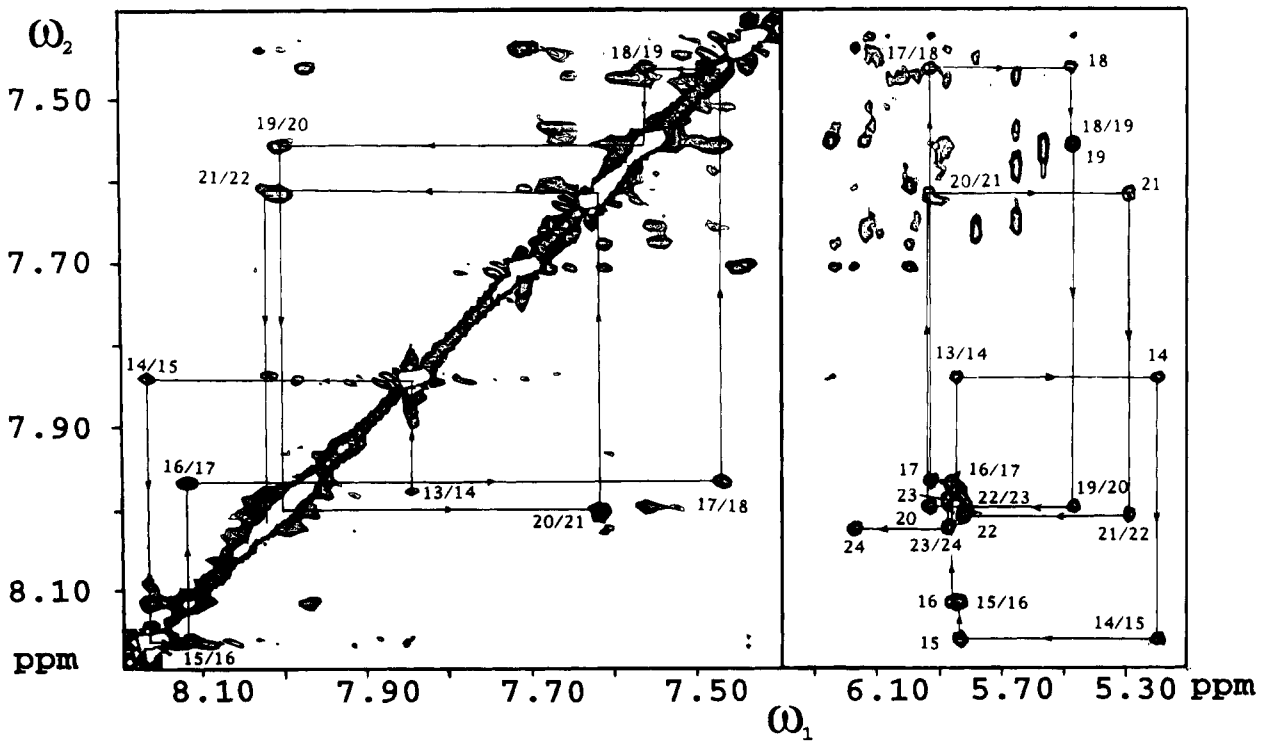

FIGURE 5: NOESY spectrum of the dodecamer in D_2O recorded at 15 °C with a mixing time of 500 ms and showing di(H1′;H8) and ds(H1′;H8) connectivities for the homopurine strand.

Assignment of Exchangeable Protons. Observation of complete NOE pattern along the oligomer, at 15 °C, is direct evidence that the imino protons are bonded in Watson–Crick base pairs, and that the DNA formed a double-helical structure. Comparison of such data with those obtained at 25 °C shows that, at this temperature, the duplex had an opening-closure dynamic along half of its length. This observation led us to limit experiments to 15 °C.

The most direct sequential pathway to assign exchangeable protons is based on through-space interactions between the imino protons of adjacent base pairs. However, an interesting alternative in our case is the presence of eight adenine residues that allow the establishment of intra-base-pair correlations from the thymine imino protons to the adenine H2 protons within an AT base pairs and then following sequentially through-space interactions from this H2 base proton to the adjacent imino proton. Thus, the A16 H2 base proton gives NOE interactions with the imino protons of three thymines which should correspond to T8, T9, and T10, while the A15 and A17 H2 base protons give dipolar interactions with only two imino protons which should correspond to T9, T10 and T8, T9, respectively. Hence these three thymine imino protons were assigned, and the procedure was continued through analysis of dipolar interactions between the imino proton of the adjacent base pair from the T8 to the G18 and from the T10 to the G14 imino protons. The same strategy was used to assign T2, T3, and T5 imino protons from A23, A22, and A20 H2 base protons followed by the ds(imino;imino) from T3 to G21 and T5 to G19. The imino protons of the frayed terminal residues could not be detected. All imino proton assignments were cross-checked using relayed information under conditions of extensive spin diffusion (500 ms NOESY) mainly from the guanine imino protons to the cytosine H5 base protons via the amino protons. Chemical shifts for the protons of the two strands are given in Table 1.

Table 1: Chemical Shift Assignments (ppm) of the Nonexchangeable and Exchangeable Protons of the Dodecamer at 15 °C

	H6 H8	H2 H5	H1′	H2′	H2″	H3′	H4′	H5′/ H5″	imino	amino
T1	7.61	1.70	5.99	2.24	2.60	4.73	4.14	3.78*		
T2	7.68	1.73	6.23	2.32	2.65	4.92	4.28	4.15*		
T3	7.54	1.68	6.12	2.3	2.59	4.93	4.27	NA[a]	13.88	
C4	7.65	5.65	5.97	2.15	2.54	4.77	4.23	NA		
T5	7.48	1.63	6.02	2.22	2.53	4.88	4.25	NA	13.885	
C6	7.59	5.65	5.90	2.25	2.45	4.83	4.2	NA		
C7	7.56	5.57	5.88	2.14	2.5	4.75	4.19	NA		
T8	7.49	1.6	6.06	2.2	2.6	4.88	4.21	NA	14.09	
T9	7.51	1.65	6.13	2.15	2.65	4.90	4.19	NA	14.03	
T10	7.45	1.68	6.08	2.16	2.53	4.92	4.18	NA	13.86	
C11	7.67	5.78	6.15	2.3	2.5	4.87	4.23	NA		
T12	7.56	1.75	6.25	2.3	2.3	4.57	4.08	4.17		
A13	7.98	7.85	5.85	2.34	2.50	4.79	4.17	3.65		
G14	7.84		5.19	2.66	2.66	4.96	4.30	NA	12.63	8.49
A15	8.16	7.28	5.84	2.67	2.85	5.07	4.43	NA		
A16	8.12	7.14	5.85	2.63	2.86	5.07	4.43	NA		
A17	7.97	7.43	5.93	2.48	2.83	5.02	4.44	NA		
G18	7.47		5.47	2.38	2.62	4.94	4.32	NA	12.83	
G19	7.56		5.47	2.45	2.63	4.94	4.32	NA	12.68	8.39
A20	8.00	7.63	5.93	2.52	2.78	5.03	4.38	NA		
G21	7.62		5.29	2.45	2.51	4.95	4.30	4.13	12.61	8.28
A22	8.01	7.35	5.82	2.50	2.75	5.00	4.38	NA		
A23	7.99	7.45	5.88	2.5	2.78	5.01	4.38	NA		
A24	8.03	7.71	6.17	2.34	2.51	4.66	4.23	4.13		

[a] NA, protons not assigned.

Proton Chemical Shift Variation with Temperature. Figure 6 shows the chemical shift dependence of the thymine methyl groups and of some of the well-resolved adenine H2 protons with temperature from 5 to 45 °C. Experiments at higher temperatures were carried out but are not reported here as the large increase in line width between 45 and 60 °C precluded assignment of these protons and at higher temperature all the protons were overlapping. The line width

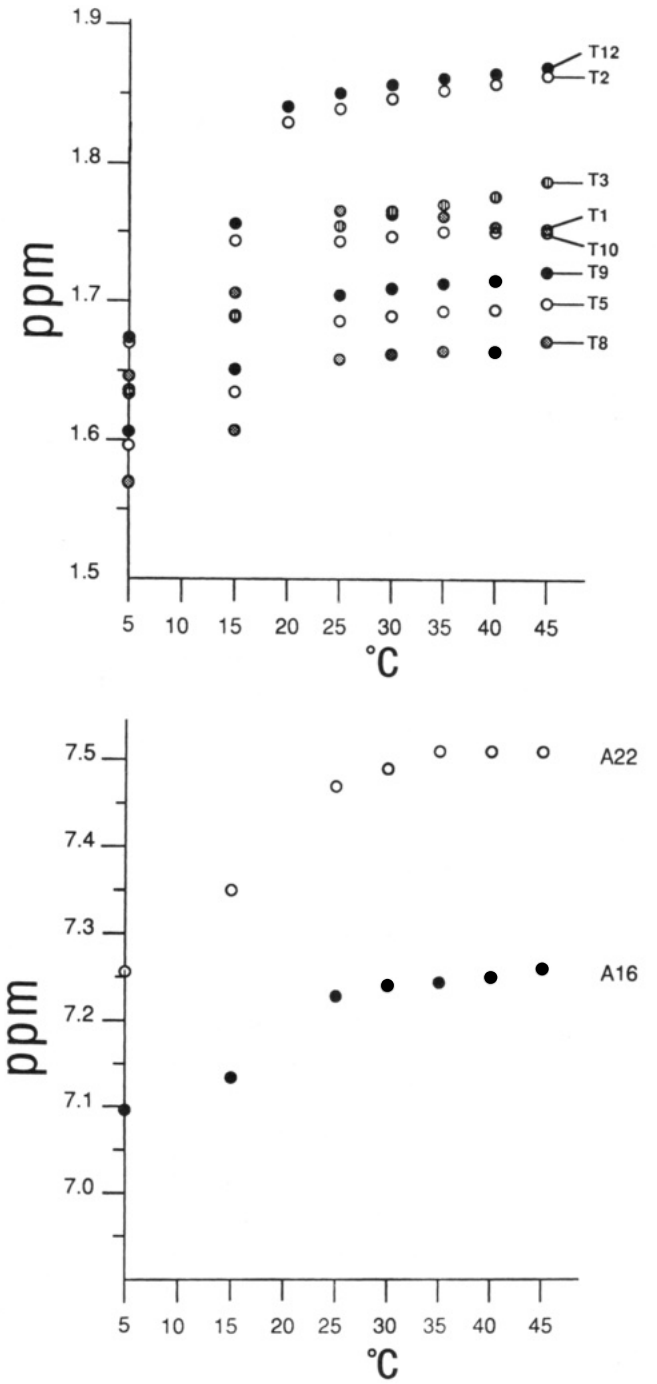

FIGURE 6: Chemical shift dependance of thymine methyl group with temperature.

increase with temperature was not uniform and T1, T2, and T12 methyl groups signals were much less affected than the other signals in agreement with the fraying of the double helix. The melting point of the dodecamer was about 50 °C. This rather low T_m can result from the relatively high percentage of AT bases pairs in the duplex since hydrogen bonding in AT pairs is weaker than in GC. In addition, at low temperature a phase transition could be observed. The large line width observed for the methyl resonances, at 0 °C, argues in favor of DNA aggregation. The dodecamer sequence under investigation can potentially form triple-stranded structures; thus spectra at low temperature were carefully analyzed for signals characteristic of this type of structure in particular in the imino proton region. However, even in the aggregated state new resonances could not be observed excluding triple-helix formation.

Modeling without Constraints. Energy minimization from many different starting points, without any constraints, led to two classic B-DNA family conformations, both of very similar free energy (−558.4 and −557.8 kcal/mol), yet with distinct conformations. The two conformations differed essentially in the twist values at steps T2pT3 and T3pC4 and in the sugar puckers G14 and A15. Both conformers have sugar puckers with pseudorotation angles in the 150−180° interval for the purine strand and in the 130−160° interval for the pyrimidine strand. The two structures were tested for stability by varying their overall twist over a range of values. The energy variation corresponding to these constraints was found to be quadratic, and we thus conclude that these structures are stable local minima.

Modeling with Geometrical Constraints from NOESY. The selected remaining NMR data consist of 106 intranucleotide and 43 internucleotide distances. Among the intranucleotide distances, H1′−H4′, H3′−H4′, and H2′−H6/H8 are the only ones which vary by more than 0.3 Å in DNA oligomers as function of sequence conformation (phase and χ). Other intranucleotide distances are only weakly affected by local conformational variations and fit all the starting points well (Wüthrich, 1986). All of the internucleotide distances are always taken into account and, in general, appear to closely correlate with base sequence effects.

We have used two sets of constraints which differ from each other in the interval range around the distance calculated from NOESY. The first set uses ±10% limits around each of 149 interproton distances. With these constraints, energy minimization from all starting points led to a unique structure C. The second set of constraints uses a ±10% limit for each of the intraresidue ^{1}H−^{1}H NOE distances and ±5% limit for each of the 43 fine structure sensitive interresidue ^{1}H−^{1}H NOE distances. Constrained minimization from all starting points again led to a unique structure, named CC.

The dodecamer sequence contains two TTTCT motifs, located at each extremity of the oligomer. Experimental NOEs values within each of these motifs are very close and suggest similar conformations. Calculations were consequently performed by imposing symmetric constraints on variables involving these nucleotides and with distance constraints set at ±10%. This led to an other conformation CRep.

The conformations C, CC and CRep were classical B helices devoid of curvature (Figure 7). These three forms differ only slightly from one another (rmsd of 0.2 Å between C and CC; 0.4 Å between CC and CRep; 0.3 Å between C and CRep) but differed from the initial conformations or from the freely minimized conformations by rmsd values of 1.2−1.9 Å. The free energy of the nonconstrained structures which fit the NOE data best and the energies of the constrained conformations C, CC, and CRep are respectively −558.4, −551.0, −550.3, and −551.1 kcal/mol. Introducing constraints therefore cost between 7.3 and 8.1 kcal/mol. It should be noted that, for each set of constraints, minimizations were performed from all starting points, B-DNA, and A-DNA structures. In all cases, JUMNA calculations associated with experimental distance constraints resulted in typical B-DNA geometries, whatever the starting conformation. Moreover, within a given set of restraints, minimized structures obtained from A or B conformation are identical.

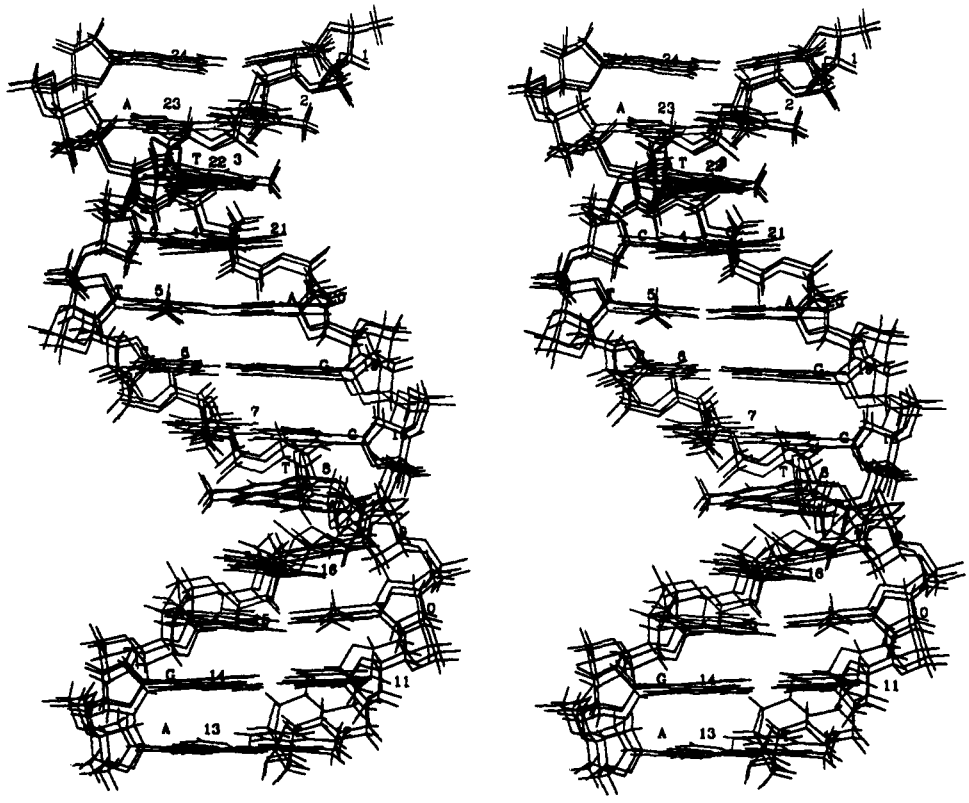

FIGURE 7: Superimposed stereopictures of the C, CC, and CRep constrained structures.

The rms differences between the experimental and calculated distances $^1H-^1H$ are 0.14, 0.13, and 0.14 Å for C, CC, and CRep, respectively. The rms differences for all constrained conformations (0.14 Å) and the best minimized conformation without constraint (0.19 Å) are roughly identical and smaller than the measured precision of the NOESY cross-peaks. Thus, we cannot select one of the three possible oligomer structures on this basis. Nevertheless, if we consider rms only for the 83 distances which present significant variations as a function of sequence within B-DNA (upper 0.3 Å), we note a slight advantage for the CC conformation. On other hand, it should be noted that ordering of the NOE distance is in very good agreement.

Another way to check the validity of structures is to compare the theoretical pattern of NOE cross-peaks built from calculated geometries with experimental NOESY data. Back-calculations were performed with the IRMA procedure for the 149 remaining interproton couplings in C, CC, and CRep structures. After the second cycle of IRMA procedure the convergence criteria were satisfied (NMRchitect User Guide 1993). The *R* factors for all mixing times (40, 70, 100, 150, 200, and 250 ms) of C, CC, and CRep are 0.54, 0.37, and 0.37, respectively. For comparison, the *R* factor of standard B-DNA Arnott conformation is 0.89. From these values, we can conclude that CC and CRep are sufficiently accurate models, even if the *R* factor criteria, as the rms criteria, do not permit selection of one and only one conformation.

Helical Parameter Analysis. The three conformations C, CC, and CRep show slight differences from one another but, overall, are close to a standard B helix, with average twist (36.5°) and rise values (3.35 Å). For the three conformations, all of the base-pair or inter-base-pair parameters (summarized in Figure 8) are very similar in the different conformations. For example, the largest twist variation between the three structures was <5° for C6pC7. In this case, the best fit with twist sensitive internucleotide NMR data again was obtained for CC, accepting that a 3.5 Å $^1H-^1H$ experimental distance is accurate to within ±0.2 Å.

Along the sequence, twist values varied between 32° and 41°. The rise is always higher than the mean value in the central C6pC7 step, as reported by Heinemann et al. (1989, 1992a,b) for the CpC steps from crystallographic studies of decamers. Tilts and rolls were small (±5° maximum), and their variations from one base pair to another were not significant. The propeller twists of the TpTpT stretches varied between −10° and −20°, in line with crystallographic results for DNA containing (A:T) tracts (DiGabrielle et al., 1989; Edwards et al., 1992).

Backbone Geometry. For all minimized conformations, we observed sugar phase values between 150° and 180° for the purine strand and between 130° and 160° for the pyrimidine strand (Figure 9). Sugar phases differ by less than 5° from one structure to another, excepting for 6, 7, 10, 11, and 20 nucleotides whose phases fluctuated between 10° and 15°. Two sugars puckers (A16 and A23) are C3′ exo, all the others are C2′ endo, characteristic of B-DNA.

The phosphate conformation within the backbone of DNA can be represented by the $\epsilon-\zeta$ angle difference (Figure 10b), all negative in the three conformations [typical of a BI classical helix (Gupta et al., 1980; Prive et al., 1987)] and are coherent with the chemical shifts obtained from the phosphorous NMR experiments (Figure 1). Nevertheless, we find a tendency toward the BII state ($\epsilon-\zeta$ up to −20°) in A15pA16 and A22pA23 steps in C, CC, and CRep and at G19pA20 in CC. As we have already shown (Hartmann et al., 1993), both phase and amplitude of the sugar at the

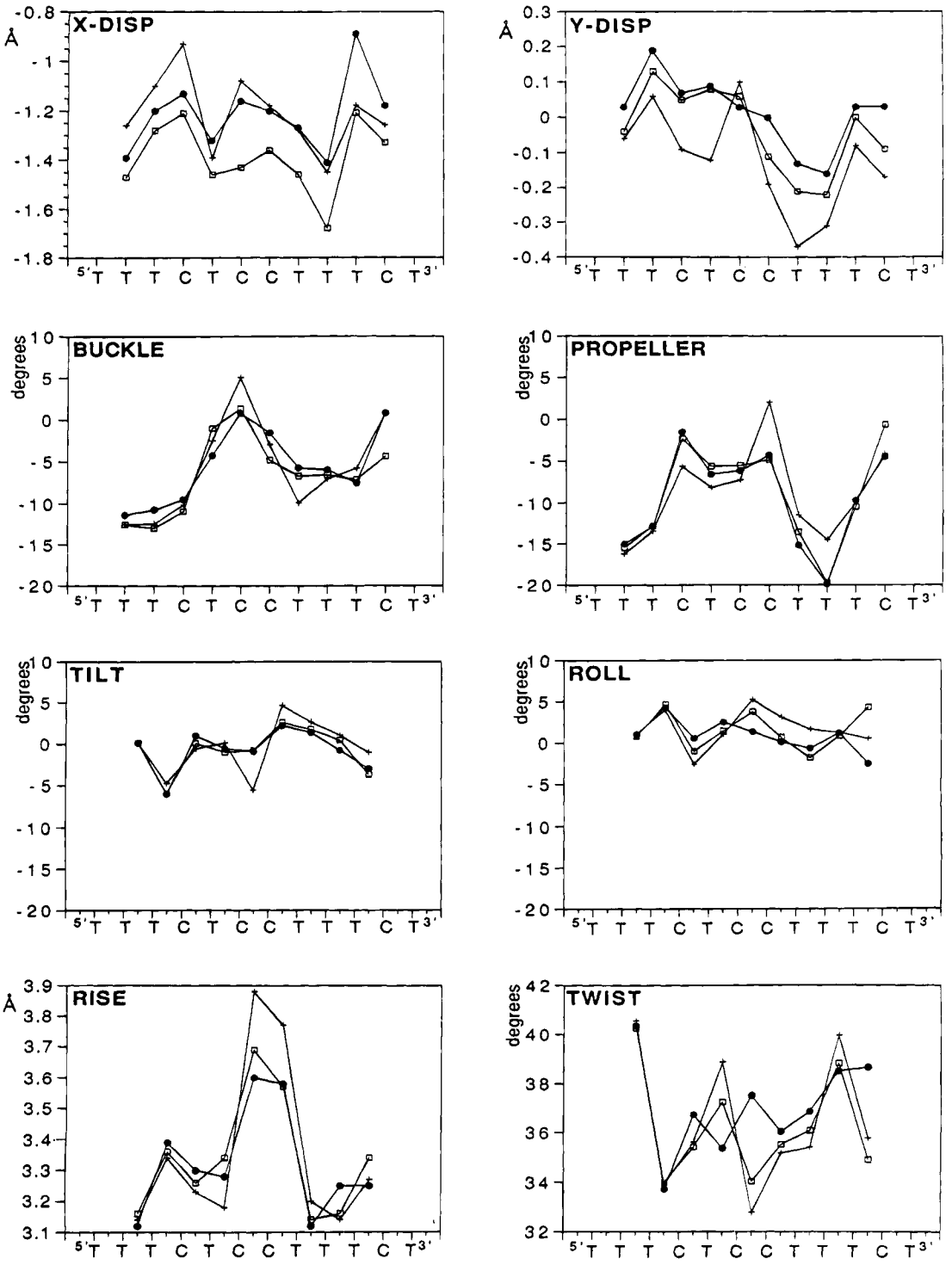

FIGURE 8: Structural parameters for the three constrained conformations (•C, +CC, CRep).

5′-side of the junction are closely coupled to the value of $\epsilon-\zeta$.

Constrained Backbone Deformation. The sugars and bases conformations are controlled by the experimental proton distances while backbone conformation depends only on modeling, as a result of the lack of torsion angle restraints. In order to provide more complete information on the phosphodiester backbone flexibility allowed by the best restrained structure, we constrained the $\epsilon-\zeta$ angle difference for all dinucleotide junctions in the CC structure. $\epsilon-\zeta$ was chosen for two reasons, first because ζ, correlated to ϵ, is the backbone most variable angle and then because $\epsilon-\zeta$ is characteristic of the phosphate position. Calculations were made using weak inter-nucleotide NMR constraints (±10% interval around the NOE $^1H-^1H$ distances). Figure 10 shows that the $\epsilon-\zeta$ angles fluctuated around a stable value, of each step for an energy cost of 2 kcal/mol. Such energy variations could be incurred simply by thermal agitation. The average amplitude of $\epsilon-\zeta$ fluctuation was ±40°, and only very little change was seen in the helical parameters. But many differences from one step to another can be noted. In particular, A15pA16 and A22pA23 have maximal dihedral rotation at 105° around a mean position (−20° in the stable form). Therefore, such variation can easily lead to a BII

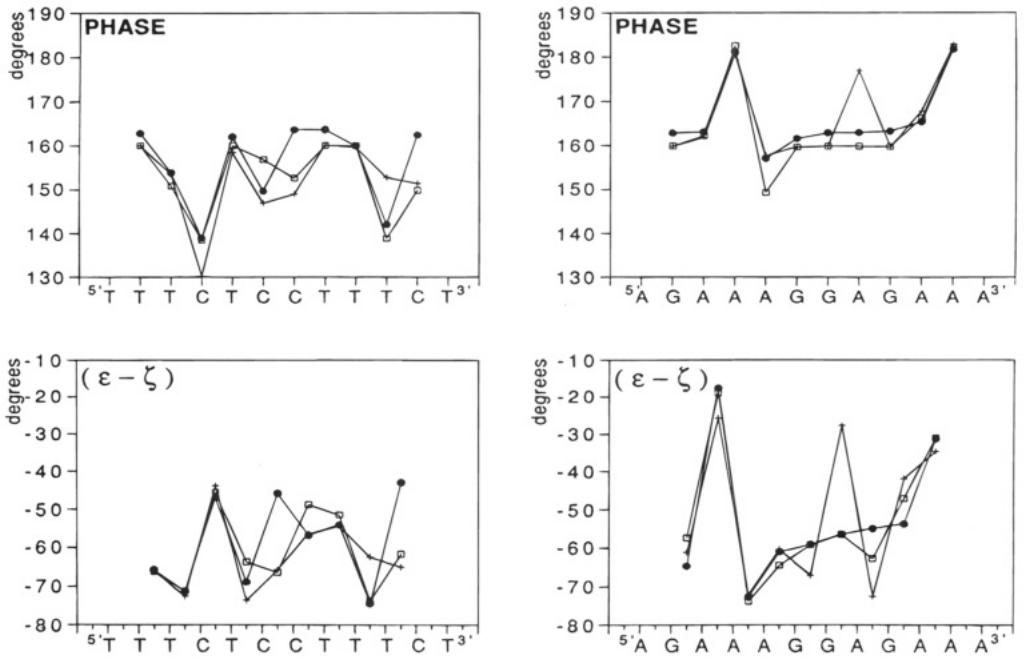

FIGURE 9: Sugar phase angles and ($\epsilon-\zeta$) dihedral angle difference for the three constrained conformations (•C, $^{+}$CC, CRep).

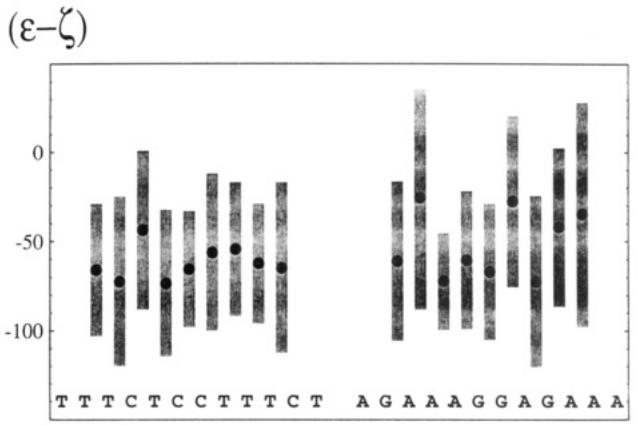

FIGURE 10: Plots of ($\epsilon-\zeta$) dihedral angle difference for the C constrained conformation. The bars shows the variations observed around the stable value for an energy cost of 2 kcal/mol.

conformation. Although no evidence of such conformation was found in the NMR data, this result must be noted. In fact, recently, many authors (Karlslake et al., 1992; Nibedita et al., 1993; El-Antri et al., 1993) have begun to appreciate the important role played by the phosphate conformation and suggested that the flexibility of the backbone may be an important component of DNA recognition.

CONCLUSION

With the strategy described in this paper, it has been possible to obtain complete assignments for the protons of a nonpalindromic dodecanucleotide homopyrimidine: homopurine, d(5′-TTTCTCCTTTCT):d(5′-AGAAAGGAGAAA). This was not trivial for the following reasons: (i) there was no symmetry to improve data quantity, in contrast to palindromic sequences, and (ii) the low melting point of the AT-rich ends of the duplex did not allow us to chose the best temperature conditions for NMR data acquisition for a molecule of such a size. Nevertheless, the solution structure determined by two-dimensional nuclear magnetic resonance spectroscopy combined with molecular simulation led to a set of low-energy B-DNA conformations satisfying the quantitative NOE data.

Imposing NMR constraints on a set of B-DNA starting points and using different upper and lower constraint bound distance errors led to three structures. The precision of NOE measurements and the fact that it was not possible to collect sufficient data do not allow us to select unequivocally one of these structures rather than another. All three conformations satisfy the NOE data and, furthermore, are very similar in terms of energy and structural parameters. Indeed, these highly related structures have neither peculiar helical parameters for the base pairs nor axis curvature. Their free energies are about 8 kcal/mol above the best freely minimized conformation. They all belong to the classical B-family and show only small deviations from regular B-DNA helix. No helix curvature was detectable nor was any peculiar feature associated with the GpA dinucleotide pairs.

In an attempt to investigate the backbone flexibility, which might play a role in protein binding (Karslake et al., 1992), we have computed the energetic cost of torsions of each $\epsilon-\zeta$ dihedral angle difference. From this systematic exploration, ApA dinucleotides, flanking the GpA steps, are found to be the most flexible (Figure 10).

The G-A hypermutation results from reverse transcription in the presence of highly asymmetric dCTP/dTTP concentrations (Vartanian et al., 1994; Martinez et al., 1994). The penchant for hypermutation in the context of GpA was attributed to substitutions occurring via a mechanism termed dislocation mutagenesis whereby the A was in fact copied twice. From this study there was no particular feature associated with the GpA dinucleotide. However, the flanking ApA dinucleotide appeared to be most flexible. It is possible that this feature could facilitate primer strand displacement.

Finally, this study reports the first structure determination in solution of a homopyrimidine:homopurine sequence known to be frequent in eukaryotic genomes. These sequences are assumed to adopt noncanonical B-DNA structures that can act as regulatory signals in gene expression (Wells et al., 1988). Moreover, they are able to form triple-

stranded structures which can be used in the control of cellular processes involving nucleic acids. Our results show that such structures may belong to the B-type duplex.

ACKNOWLEDGMENT

We thank Jean Igolen and Catherine Gouyette for oligonucleotide synthesis.

REFERENCES

Arnott, S., & Hukins, D. W. L. (1973) *J. Mol. Biol. 81,* 93−105.
Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W., & Ratliff, R. L. (1980) *Nature 283,* 743−745 and coordinates communicated to us by S. Arnott.
Assa-Munt, N., & Kearns, D. R. (1984) *Biochemistry 23,* 791−796.
Baleja, J. D., & Sykes, B. D. (1991) *J. Magn. Reson. 91,* 624−629.
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.
Bax, A., & Davies, D. G. (1985) *J. Magn. Reson. 65,* 355−360.
Bebenek, K., Abborts, J., Roberts, J. D., Wilson, S. H., & Kunkel, T. A. (1989) *J. Biol. Chem. 264,* 16948−16956.
Boelens, R., Koning, T. M. G., & Kaptein, R. (1988) *J. Mol. Struct. 173,* 299− 311.
Boelens, R., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1989) *J. Magn. Reson. 82,* 290−308.
Chuprina, V. P., Lipanov, A. A., Fedoroff, O. Y., Kim, S.-G., Kintanar, A., & Reid, B. R. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88,* 9087−9091.
Chuprina, V. P., Sletten, E., & Fedoroff, O. Y. (1993) *J. Biomol. Struct. Dyn. 10,* 693−707.
Clore, G. M., & Gronenborn, A. M. (1987) *Protein Eng. 1,* 275−288.
Clore, G. M., & Gronenborn, A. M. (1989) *Crit. Rev. Biochem. Mol. Biol. 24,* 479−564.
Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., von Kitzing, E., Wang, A. H.-J., & Zhurkin, V. B. (1989) *J. Mol. Biol. 205,* 787−791.
DiGabrielle, A. D., Sanderson, M. R. M., & Steitz, T. A. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86,* 1816−1820.
Edwards, K. J., Brown, D. G., Spink, N., Skelly, J. V., & Neidle, S. (1992) *J. Mol. Biol. 226,* 1161−1173.
El-Anrti, S., Bittoun, P., Mauffret, O., Monnot, M., Convert, O., Lescot, E., & Fermandjian, S. (1993) *Biochemistry 32,* 7079−7088
Gonzales, C., Rullmann, J. A. C., Bonvin, A. M. M. J., Boelens, R., & Kaptein, R. (1991) *J. Magn. Reson. 91,* 359−664.
Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110,* 7870−7872.
Grütter, R., Otting, G., Wüthrich, K., & Leupin, W. (1988) *Eur. Bipophys. J. 16,* 279−286.
Gupta, G., Bansal, M., & Sasisekharan, L. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77,* 6486−6490.
Hartmann, B., Malfoy, B., & Lavery, R. (1989) *J. Mol. Biol. 207,*433−444.
Hartmann, B., Piazzola, D., & Lavery, R. (1993) *Nucleic Acids Res. 21,* 561−568.
Heinemann, U., & Alings, C. (1989) *J. Mol. Biol. 210,* 369−381.
Heinemann, U., & Hahn, M. (1992) *J. Biol. Chem. 267,* 7332−7341.
Heinemann, U., Alings, C., & Bansal, M. (1992) *EMBO J. 11,* 1931−1939.
Hingerty, B., Richie, R. H., Ferrel, T. L., & Turner, J. E. (1985) *Biopolymers 24,* 427−439.
Karslake, C., Botuyan, M. V., & Gorenstein, D. G. (1992) *Biochemistry 31,* 1849−1858.
Koning, T. M. G., Boelens, R., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1991) *Biochemistry 30,* 3787−3797.
Kunkel, T. A., & Alexander, P. S. (1986) *J. Biol. Chem. 261,* 160−166.
Lavery, R. (1988) in *Structure and Expression, Vol. 3, DNA Bending and Curvature* (Olson, W. K., Sarma, R. H., Sarma, M. H., & Sundaralingam, M., Eds). p 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. (1994) *Biophys. Chem. 50,* 33−45.
Lavery, R., Zakrwewska, K., & Pullman, A. (1984) *J. Comput. Chem. 5,* 363−373.
Lavery, R., Sklenar, H., Zakrzewska, K., & Pullman, B. (1986a) *J. Biomol. Struct. Dyn. 3,* 989−1014.
Lavery, R., Parker, I., & Kendrick, J. (1986b) *J. Biomol. Struct. Dyn. 4,* 443−461.
Martinez, M. A., Vartanian, J.-P., & Wain-Hobson, S. (1994) *Proc. Natl. Acad. Sci. U. S. A. 91,* 11787−11791.
Mauffret, O., Hartmann, B., Convert, O., Lavery, R., & Fermandjian, S. (1992) *J. Mol. Biol. 227,* 852−875.
Nibedita, R., Ajay Kumar, R., Hosur, R. V., & Govil, G. (1993) *Biochemistry 32,* 9053−9064.
NMRchitect User guide (1993) version 2.3 Biosym Technologies, San Diego, CA.
Poncin, M., Hartmann, B., & Lavery, R. (1992) *J. Mol. Biol. 226,* 775−785.
Prive, G. G., Heinemann, U., Chandrasegaran, S. Kan, L. S., Kopta, M. L., & Dickerson, R. E. (1987) *Science 238,* 498−504.
Ramstein, J., & Lavery, R. (1988) *Proc. Natl. Acad. Sci. U.S A. 85,* 7231−7235.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48,* 286−292.
Van de Ven, J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178,* 1−38.
Vartanian, J.-P., Meyerhans, A., Åsjö, B., & Wain-Hobson, S. (1991) *J. Virol. 65,* 1779−1788.
Vartanian, J. P., Meyerhans, A., Sala, M., & Wain-Hobson, S. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91,* 7231−7235.
Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., & Wohlrab, F. (1988) *FASEB J. 20,* 2939−2949.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids,* John Wiley and Sons, New York.
Zakrwewska, K. (1992) *J. Biomol. Struct. Dyn. 9,* 681−695.

BI942619H