Structural Polymorphism and Dynamism in the DNA Segment GATCTTCCCCCCGGAA: NMR Investigations of Hairpin, Dumbbell, Nicked Duplex, Parallel Strands, and i-Motif

Shanteri Singh, P. K. Patel, and R. V. Hosur*

Chemical Physics Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India

Received April 8, 1997; Revised Manuscript Received July 15, 1997[⊗]

ABSTRACT: Structure and dynamism in a DNA segment d-GATCTTCCCCCCGGAA have been investigated by nuclear magnetic resonance (NMR) spectroscopy. At neutral pH, the molecule exists largely as a dumbbell, formed by the association of two hairpins with sticky ends. A small percentage of hairpin is also detectable. The stem of the dumbbell, which is 12 base pairs long, has two nicks separated by 4 base pairs. The three-dimensional structures of the dumbbell and also of a 12-mer duplex, the sequence of which is identical to that of the stem of the dumbbell, have been determined by NMR in conjunction with restrained molecular dynamics calculations. It is observed that the presence of nicks causes minor changes in the structure of the duplex. Fraying at the nicks is much less than at the ends of a regular duplex. The loop shows very few nuclear Overhauser effects, which is a reflection on the greater dynamism in its structure. At lower pH, the molecule undergoes a transition to an i-motif type of structure with two parallel stranded duplexes involving $C-C^+$ pairing, interdigitating each other. The structure is highly stable, with a melting temperature >60 °C.

Our knowledge of DNA structure and function is undergoing rapid enrichment in recent years with discoveries of new forms of DNA and elucidation of three-dimensional structures of different synthetic oligonucleotides with known base sequences [see reviews in James (1996)]. In addition to the most common form of DNA, namely the duplex, several other forms such as hairpins, cruciforms, three-stranded structures such as three-way junctions, triplexes, four-stranded structures such as Holliday junctions, G-quadruplex, and i-motifs, characteristic folded structures such as DNA aptamers etc. (Wyatt et al., 1994), have been discovered and investigated in detail. Hairpin, dumbbell, and cruciform DNAs occur as a consequence of inverted repeats in a single strand of the DNA and are known to play crucial roles in genetic recombination and regulation. Sequences capable of hairpin formation are often seen near gene regulatory and promoter sites in DNA, and they are stabilized by interaction of the loops with specific enzymes (Cheng & Tinoco, 1994). The process of genetic expression requires unwinding and opening of the duplex to single-stranded DNA (ss-DNA), and then depending on the base sequence and environmental conditions, these ss-DNAs could form secondary structures that are specifically recognized by proteins during the course of the biological functions.

The process of genetic recombination in living cells is found to proceed via triplex formation (Panayotatos & Fontaine, 1994; Stevens et al., 1993; Wells et al., 1988; Htun & Dahlberg, 1989; Broitman & Fresco, 1989). The RecA protein binds a ss-DNA to form a filament, which then recognizes a duplex DNA having a specific base sequence, and this is driven by triplex formation between the bound single strand and the incoming duplex DNA. Triplex formation has also important gene targeting applications (Goodman & Nash, 1989; Mather et al., 1989; Cooney et al., 1988; Doan et al., 1987). For example, specific reagents attached to ss-DNA can be targeted to specific sites in the duplex DNA. Similarly, nicks and mismatches in duplex DNA are very significant from the point of view of ligation and repair mechanisms (Karmer et al., 1984; Lu et al., 1984; Cleverys et al., 1983; Dohet et al., 1985). Failures in correction cause frame shift mutations leading to alterations in genetic expressions (Loeb & Kunkel, 1982; Fowler et al., 1974) and behaviors.

DNA packaging inside the cell requires extensive folding of the DNA chains. While a significant cause of such effects may be interaction between proteins and DNA, it is now recognized that DNA-DNA interaction can also cause substantial stabilization of the folded structures. For example, at the telomeric end of the genome, which is rich in G-stretch (Ahmed et al., 1994; Zimmerman et al., 1975; Zimmerman, 1976; Willinger et al., 1993), it is believed that the chain folds to form G-quadruplexes, which are highly compact structures. Likewise, in stretches that are rich in C-sequences, $C-C^+$ pairing based i-motif can occur, and this is again a very compact structure (Ahmed et al., 1994; Willinger et al., 1993; Gehring et al., 1993; Chen et al., 1994).

A molecular level understanding of all these phenomena requires a detailed knowledge of the three-dimensional structures of the individual unusual DNA forms. Besides, structural transitions between the various forms of DNA would have important consequences *in vivo*. With these views in mind, we have studied a particular DNA sequence, which has a great potential to form different structures under different conditions and thus provides a good system to investigate many of the features described above. The sequence under investigation is d-GATCTTCCCCCCGGAA. Under normal conditions of neutral pH and ionic strength, this molecule can form hairpin and dumbbell structures as

[⊗] Abstract published in *Advance ACS Abstracts,* October 1, 1997.

S0006-2960(97)00819-2 CCC: $14.00

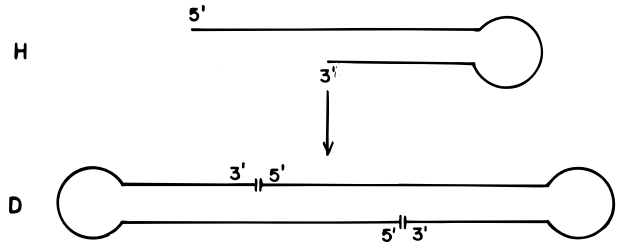

FIGURE 1: Schematic hairpin and dumbbell structures formed by the 16-mer 5′-GATCTTCCCCCCGGAA-3′.

shown in Figure 1. There have been extensive nuclear magnetic resonance (NMR) studies in the literature on different aspects of hairpin stability, sequence dependence, loop size, loop comparison, etc. [see reviews: Van de ven and Hilbers (1988) and Patel et al. (1987)]. Hairpin is expected to occur at low DNA concentration (submillimolar), and at higher (millimolar) concentrations two hairpins would associate via their sticky ends to form a dumbbell. The central portion of the dumbbell is a 12 base pair duplex but with two nicks in the middle. The sequence in the stem has another interesting feature. It has a six unit long purine stretch GGAAGA followed by a six unit long pyrimidine stretch TCTTCC, which is complementary to the purine stretch. Thus, it may be anticipated that at lower pH the molecule may associate further to form triplex structures. Besides, the 16-mer sequence has a stretch of six cytidines which could form $C-C^+$ pairs at lower pH and thus form a parallel stranded duplex or its extension to an i-motif, as has been seen recently in TC_5 and other cytidine stretches.

MATERIALS AND METHODS

DNA Synthesis. Two oligonucleotides, d-GATCTTC-CCCCCGGAA (16-mer) and d-GGAAGATCTTCC (12-mer), were synthesized at 10 μmol scale on an automated DNA synthesizer (ABI381) using phosphoramadite chemistry (McBride et al., 1983). The DNAs were purified by polyacrylamide gel electrophoresis. The samples for NMR were prepared by dissolving lyophilized DNA in either 90% H_2O plus 10% D_2O mixtures or 99.99% D_2O. Phosphate buffer (10 mM) was used, and the salt concentration (NaCl) was varied from 50 to 200 mM for studying salt dependence. However, for the final structural studies, the salt concentration was 100 mM.

NMR Measurements. All NMR measurements were carried out either on an AMX-500 spectrometer or on a UNITY-plus 600 spectrometer. For the experiments in H_2O, water suppression was achieved with either the Jump and Return sequence (Plateau & Gueron, 1982) or the Watergate sequence (Piotto et al., 1992). Two-dimensional clean TOCSY (Griesinger et al., 1988) spectra were recorded with mixing times in the range 50−100 ms. Two-dimensional NOESY (Jeener et al., 1979) spectra were recorded with mixing times in the range 50−400 ms. In all cases 2048 t_2 points were collected, with the number of t_1 points varying between 400 and 512. For nuclear Overhauser effect (NOE) buildups, data were collected with identical experimental parameters. Heteronuclear $^1H-^{13}C$ correlation spectra at natural abundance of ^{13}C were recorded with the (heteronuclear multiple quantum coherence (HMQC)) approach (Muller, 1979). Four hundred t_1 experiments with 2048 t_2 points for each were performed, and signal averaging was done with 128 scans for each t_1 value.

Molecular Dynamics Calculations. Simulated annealing and restrained molecular dynamics calculations were carried out using the X-PLOR (3.1) (Brunger, 1992) package on Silicon Graphics and DEC-alpha computer systems. The following protocol was used. A template structure with good geometries was taken as a starting structure, and 200 steps of Powell minimization with a very low force constant on the van der Waals term ($K_{vdw} = 0.002$ kcal mol^{-1} Å^{-2}) was then carried out to improve the covalent geometry prior to dynamical simulated annealing. Annealing was performed in three steps. Step 1 comprised 20 ps dynamics with a time step of 5 fs and with very small weights on geometry and van der Waal's terms. In step 2, 10 ps dynamics was carried out at 1000 K by tightening the weight on the geometry. The initial velocities were chosen from Maxwell distribution at 1000 K. In step 3, the temperature was lowered in steps of 50 K, and at each temperature the system was equilibrated for 1 ps. At the final temperature of 300 K, 1 ps dynamics was carried out with $K_{vdw} = 4$ kcal mol^{-1} Å^{-2}. This was followed by 200 cycles of Powell minimization. The new coordinates so obtained were used as starting coordinates for subsequent NOE-based refinement of the structures. This process was repeated several times (50−100) with different initial structures, so as to scan the conformational space properly. The NOE force constant was set to 50 kcal mol^{-1} Å^{-2} in all of the above calculations.

Experimental NOE Intensities. Cross peaks in all of the NOESY spectra recorded with different mixing times were integrated to monitor the dependence of the NOE intensities on the mixing time and thereby decide on suitable mixing times for initial distance estimation and for NOE-based refinement. The peak intensities were all referenced to a CH5−CH6 NOE cross peak; the different CH5−CH6 cross peaks had only a marginal variation in the intensities.

Constraints. Constraints, which are derived from experimental data, define the conditions that an acceptable structure of the molecule has to satisfy. When these are used in a general conformational search, the conformational space gets restricted and results in saving of computational time. We have used H-bond constraints (three H-bonds for each G-C pair and two for each A-T pair), noncrystallographic symmetry constraints for maintaining the equivalence of the strands in the duplex, distance constraints derived from the NOESY spectra at short mixing time (40−50 ms), and planarity constraints for the base pairs. Each H-bond is defined by two distance constraints.

Relaxation Matrix Calculations. NOE intensities for any given structure were calculated by relaxation matrix approach using the program SIMNOE (Nibedita et al., 1992, Majumdar & Hosur, 1992), as described earlier. The calculated and experimental NOE intensities were compared to decide on the quality of the structure, and the DNA structures were iteratively refined until a satisfactory fit resulted. A NOESY spectrum recorded with a mixing time of 200 ms was used for such comparison. Calculations were performed with different correlation times, and we observed that >90% of the peaks in the stem portion of the dumbbell and also in the entire duplex having the same sequence as the stem of the dumbbell could be fitted with "the single correlation time model" within a precision of 25%; correlation times of 8 and 15 ns were used for the dumbbell and the duplex, respectively. The 25% limit for an acceptable fit was set considering the facts that there would be inaccuracies in peak

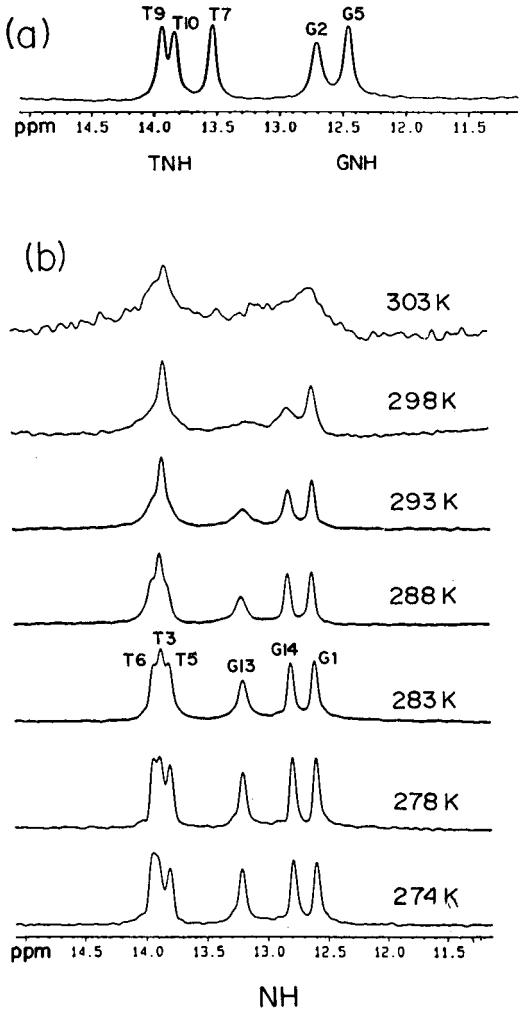

FIGURE 2: 500 MHz imino proton spectra of the 12-mer DNA d-GGAAGATCTTCC at 293 K (a) and of the 16-mer (b) in 90% H_2O + 10% D_2O at pH 8.25, 0.1 M NaCl. In (b), the spectra are shown as a function of temperature.

integrations and there are approximations in the method of calculation. Many sophisticated approaches such as time averaging of distance constraints and calculation of average NOEs have been proposed in the literature to take into account the molecular motions occurring in the solution media (Withka et al., 1992; Bonvin et al., 1993, 1994; Yip & Case, 1989; Lefevre et al., 1987). However, all of them still have approximations and no substantial changes in the overall conclusions about the structures have been revealed in the past. Therefore, it suffices to note here that the general structures derived from a simplistic approach have scope for dynamical fluctuations.

RESULTS AND DISCUSSION

NMR Spectra at Neutral pH. Figure 2 shows the imino proton spectra for the 16-mer and 12-mer sequences recorded in 90% H_2O plus 10% D_2O at 293 K. The fact that three G imino and three T imino proton resonances are seen in the 16-mer clearly indicates that a dumbbell structure is formed in the solution. In the case of the 12-mer duplex, the terminal G imino resonance is not seen because of fraying. The peaks could be readily assigned on the basis of chemical shifts and NOEs in the 2D NOESY spectra following a well-established procedure for duplexes (Wüthrich, 1986). As an illiustration, Figure 3 shows the selected regions along with assignments

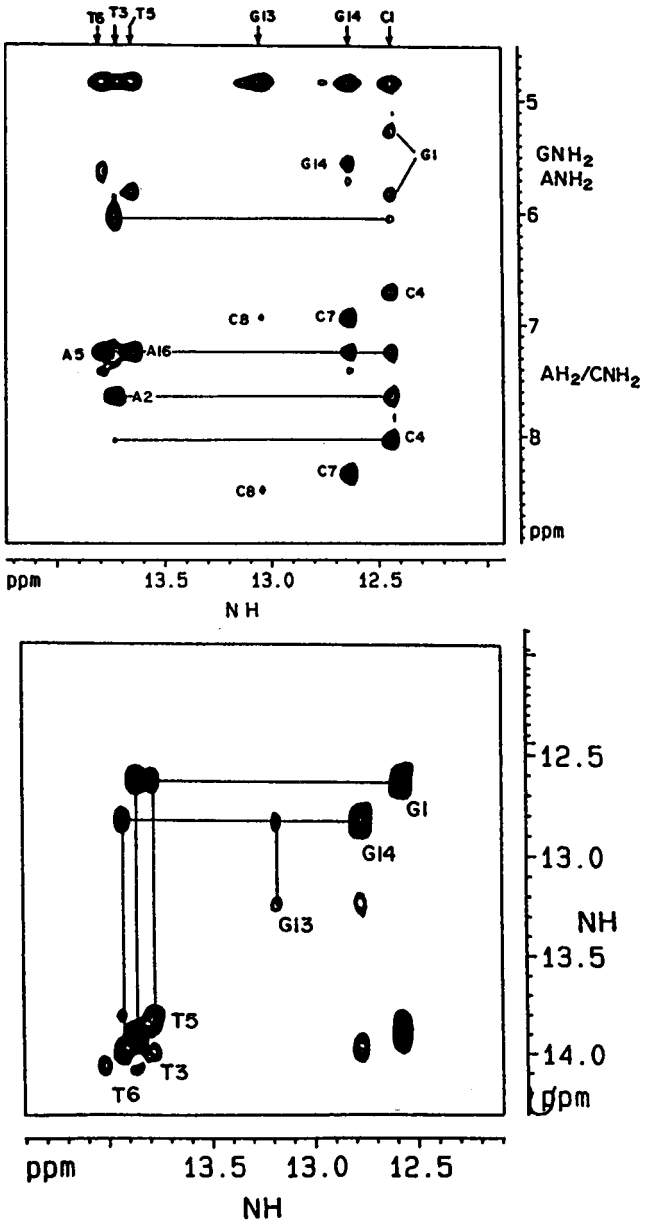

FIGURE 3: Selected regions from a NOESY spectrum in H_2O of the 16-mer showing the connectivities and the assignments (pH 8.2, temperature = 10 °C, mixing time = 150 ms).

from a NOESY spectrum of the dumbell in H_2O. It is interesting to note that a T5NH−G1NH connectivity through the nick in the stem is observed in the spectrum.

The spectra in D_2O showed excellent dispersion of peaks in both molecules, and the peaks in TOCSY and NOESY spectra could again be readily assigned following standard procedures for right-handed duplexes [see review in Hosur et al. (1988)]. Some of the ambiguities arising from the overlap of peaks were resolved by using the $^1H-^{13}C$ *J*-correlated spectrum (Radha, 1996). In the NOESY spectra of the 16-mer, several cross peaks were observed between the protons on the first nucleotide G1 and the last nucleotide A16, which again was in conformity with the dumbbell structure mentioned above. Figure 4 shows a particular portion of the NOESY spectrum of the dumbbell displaying two such connections and also the sequential connectivities in the molecule. This spectrum also serves to illustrate the quality of the spectra used for NOE-based structure refinement to be described later.

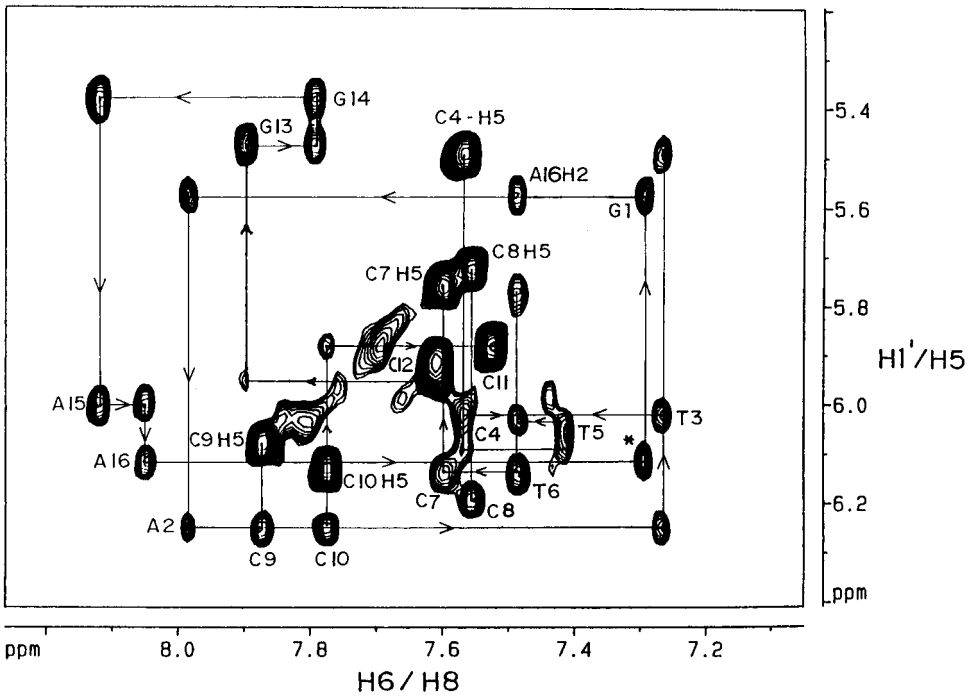

FIGURE 4: H8/H6−H1′ cross-peak region of the NOESY spectrum of the 16-mer dumbbell in D_2O at 293 K, pH 8.2, recorded with a mixing time of 150 ms. Sequential connectivities for the stem portion of the dumbbell have been drawn in. Such connectivities were not seen in the loop region. This spectrum also shows two connections between G1 and A16 (*), between which there is no covalent linkage. This establishes good stacking of the base pairs despite the nick at the above position.

The 2D spectra of the duplex exhibited interesting features. There were intense intrastrand sequential H1′−H1′ cross peaks at mixing times 70 ms and above, which are normally not seen in a duplex. However, careful investigations with different concentrations of DNA and under different salt concentration revealed that the molecule undergoes end to end stacking in solution, resulting in a large molecular mass. The structure of the duplex itself did not seem to change because of stacking. This was evident from the observations that (i) on dilution only the terminal nucleotide protons exhibited chemical shift changes and (ii) the relative intensities of the peaks at any mixing time were very similar. Thus, the observed H1′−H1′ cross peaks were attributed to efficient spin diffusion in the stack of duplexes, and therefore these peaks were not converted into distance constraints in the structure calculation exercise. However, these intensities were indeed used for comparison of calculated and experimental intensities.

Structure Calculations on the DNA Dumbbell. The structure of the 16-mer dumbbell was calculated following the protocols described under Materials and Methods. In this exercise, a 5′-terminal phosphate was included with a view to generate structures for the dumbbell that, on the one hand, would satisfy the experimental constraints and the stem portion, on the other hand, would truly represent a duplex with nicks in the middle. We felt this was more important than generating structures for the dumbbell without the phosphate, since this way we would be getting biologically relevant information. The scientific validity for this modification is that, in the end, when the experimental constraints are satisfied, dropping the phosphate from the structures would produce acceptable structures for the molecule that we have synthesized.

The constraint set consisted of 234 intrastrand $^1H-^1H$ distance constraints (117 per strand), 60 H-bond distance constraints, 12 base pair planarity constraints, and noncrystallographic symmetry (X-PLOR 3.1) constraints. Simulated annealing, restrained molecular dynamics, and NOE refinement starting with >50 initial structures resulted in 8 convergent structures; the one with the best NOE fit is shown in Figure 5. The peak-to-peak NOE fit for this structure is shown in Figure 6a. It is seen that for most of the peaks, calculated (S) and experimental (E) intensities match within 30%; for six peaks the deviations are larger. Four of these peaks, namely, C11H6−C11H3′ (peak 1), G13H1′−G13H4′ (peak 8), G13H1′−G13H3′ (peak 9), and G13H8−C12H1′ (peak 14), belong to the loop region, and for the other two peaks, the quantification is inaccurate due to peak overlaps. Figure 6b shows RMSD comparison of the eight structures at every nucleotide step in the molecule. An analysis of the torsion angles in the eight structures in the form of dials is given in Figure 7. All of these data reflect the uniqueness of the structure in different regions of the 16-mer dumbbell. The smaller the range of the RMSD, the more unique is the structure and vice versa. Thus, we notice that the dumbbell structure is rather uniquely determined in the stem portion of the molecule even though the stem has two nicks, four base pairs apart. In contrast, the loop regions are poorly characterized. This is a consequence of insufficient number of NOE peaks obtained from this portion of the molecule, which in turn is a consequence of greater dynamism. The overall direction of the loops with respect to the stem is quite varied. Analysis of the various structural parameters for the stem duplex revealed that this part of the dumbbell belonged to the B-DNA family.

Structure Calculations on the Stem Duplex. For the purpose of evaluating structural consequences of nicks in a duplex, we separately determined the structure of the intact duplex corresponding to the stem of the dumbbell. Considering that the duplex undergoes end-to-end aggregation, we

FIGURE 5: Stereoview of the 16-mer dumbbell structure having the best NOE fit.

tried to investigate the structure at low concentration of the DNA. However, at these concentrations the signal to noise ratio was very poor and, consequently, the NOE quantifications were very poor. Since we also observed that the structure of the duplex itself did not change due to aggregation, we decided to use the high concentration data itself but use a higher correlation time for fitting the NOE intensities. For the structure calculations, an identical protocol as for the dumbbell was used. The constraint set included 226 intrastrand $^1H-^1H$ distance constraints (113 per strand), 60 H-bond distance constraints, and 12 base pair planarity constraints. Here again, noncrystallographic symmetry constraints were used to maintain the equivalence of the two strands. A NOESY spectrum recorded with a mixing time of 200 ms was used for relaxation matrix based structure refinements, and we used a single correlation time of 15 ns for the NOE calculations. The final results are presented in Figure 8. The best of the NOE fits shown in Figure 8a indicates that the duplex structure is well characterized by the NOEs. Similarly, the residue-wise RMSD plot depicted in Figure 8b also indicates that the duplex structure is more uniquely defined by the NOEs than the dumbbell. Ten convergent structures were used for the calculation of these RMSDs.

Analysis of the various structural parameters for the duplex revealed that, overall, the molecule's topology is of the B-type but with some sequence-dependent variations.

Effects of Nicks and Loops on the Stability and Structure of the Duplex in the Dumbbell. The data in Figure 2 reveal that the nicks and loops have significant effects on the stability of the base pairs and, consequently, on the overall stability of the central duplex in the dumbbell. Specifically, it is seen that the G13−C8 pair in the dumbbell, which is sequence-wise equivalent to the G1−C12 pair in the duplex, forms a more stable base pair in the dumbbell. In fact, in the duplex, the fraying effect has completely broadened the imino proton of G1 in the first GC pair. This indicates that the loops shield the G13−C8 base pair from the solvent water and thus provide a better stability to this base pair. Furthermore, the G1−C4 pair, which marks the nick in the

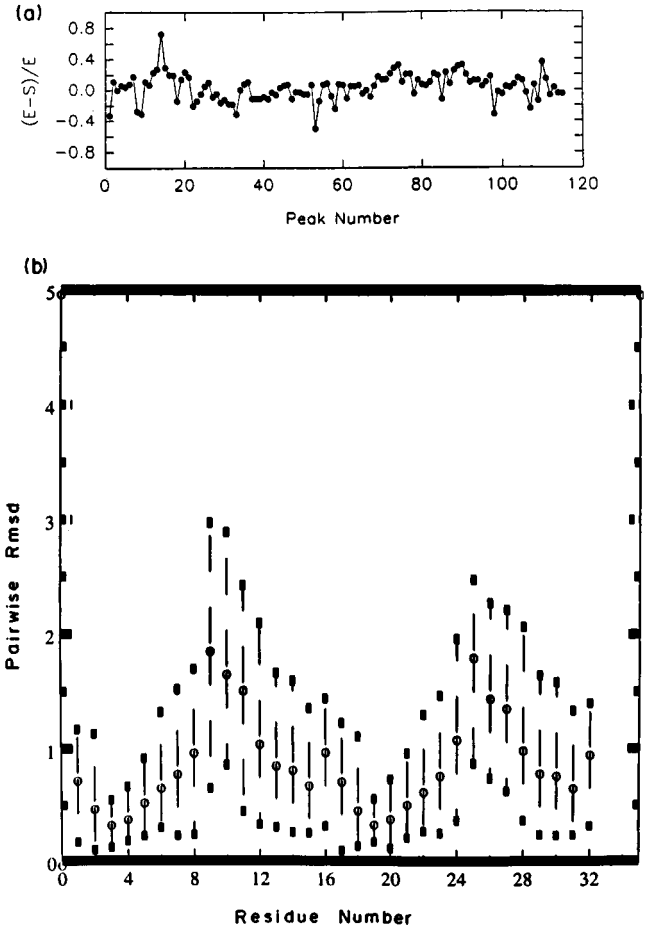

FIGURE 6: (a) Peak to peak fit between the calculated (S) and experimental (E) NOE intensities for the structure in Figure 5. For the other structures too, the NOE fits were within 30% for >90% of the peaks. The maximum deviations were around 70−80% for a few peaks which belonged to nucleotides near the loops in the dumbbell. (b) Pairwise RMSD comparison of the eight convergent structures of the dumbbell, at every nucleotide step. Vertical bars indicate the range of RMSD deviations among the eight structures. It is observed that the absolute values and the ranges for the RMSDs are quite small at all the nucleotides in the stem portion of the dumbbell. The loops are rather poorly defined.

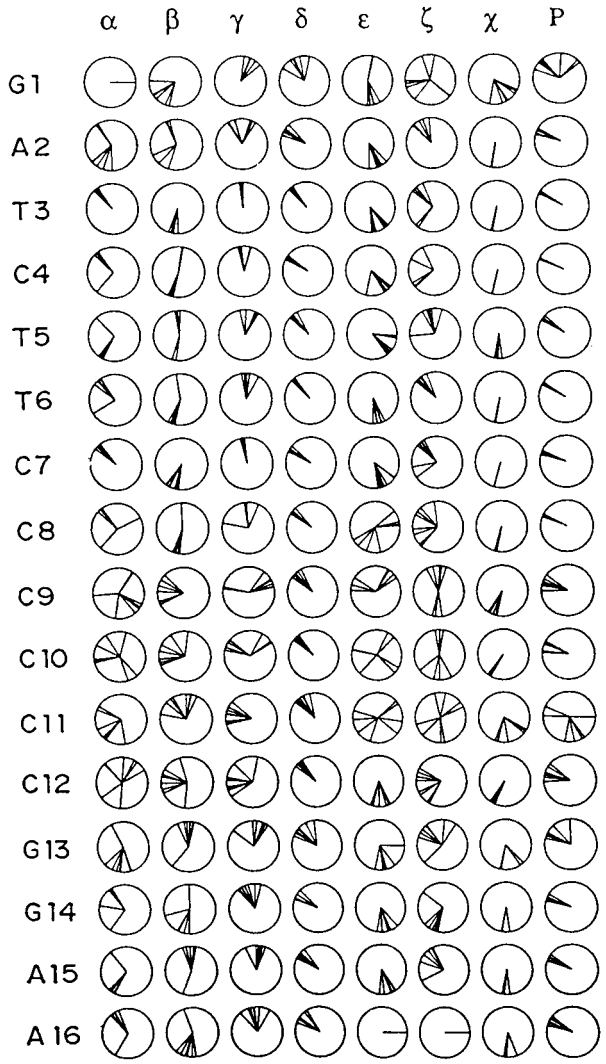

FIGURE 7: Torsion angles in the eight convergent structures of the dumbbell plotted in the form of dials. The distribution in the dials reflects the similarities and the differences in the structures, at different nucleotide levels. The dials for the nucleotide units C8−G13 and G1−A2 show large variations.

dumbbell, is fairly stable, indicating that fraying at the nick is relatively less than that at the ends of a duplex.

We also observed (Figure 2b) that the dumbbell almost completely melts away by about 35 °C, which is much lower than the melting temperature of the corresponding duplex ($T_m = 62$ °C). This decrease in the stability of the dumbbell is a consequence of the nicks in the stem portion. The nicks also produce a partial destacking of the adjacent base pairs, as can be seen from the fact that T3 imino in the dumbbell is shifted downfield by about 1 ppm as compared to the T7 imino resonance in the duplex (Figure 2). Likewise, in the D_2O spectrum G1H8 in the dumbbell is shifted quite upfield compared to that in the duplex. However, the stacking is still sufficiently good to see NOE cross peaks across the nicks. For example, in Figure 3, an NOE is seen between T5NH and G1NH protons. Similarly, A16H2−G1H1′ and G1H8−A16H1′ cross peaks through the nick are seen in the NOESY spectrum in D_2O (Figure 4). These observations indicate that the base pairs G1−C4, A2−C3 and A16−T5 in the dumbbell stack differently from in the regular duplexes.

Figure 9 provides a visual comparison of the molecular structures discussed above to highlight the effects of the nicks in the center of the duplex. These structural distortions

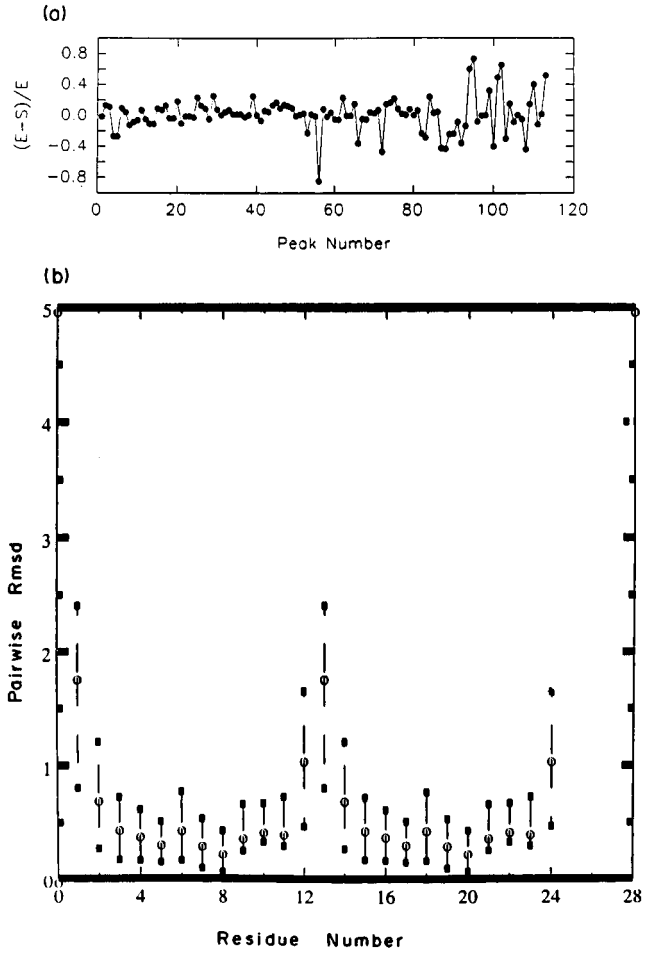

FIGURE 8: (a) Peak to peak NOE fit between calculated and experimental intensities, for the 12-mer duplex structure which had the best fit. (b) RMSD comparison of 10 convergent structures as in Figure 5.

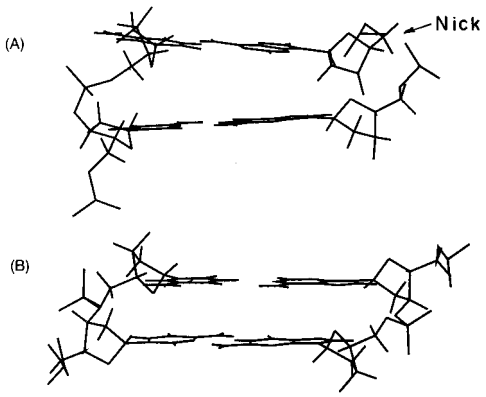

FIGURE 9: Visual comparison of the duplex structure and the stem portion of the dumbbell at the position of the nicks. Base stacking differences are seen between the two molecules. The stacking, however, is still good in the dumbbell to be able to see sequential NOE connections through the nicks.

produced by the nicks may be of significance for recognition by enzymes and other reagents *in vivo*.

pH Induced Structural Transition in the Dumbbell−i-Motif. Environment-induced structural transitions in DNA and macromolecules are known to play important roles in biological functions. In this sense H^+ concentration or pH of the environment is an important factor. Although the average physiological pH in the cellular soup is ~7, different pH conditions in the environments of specific sites in DNA

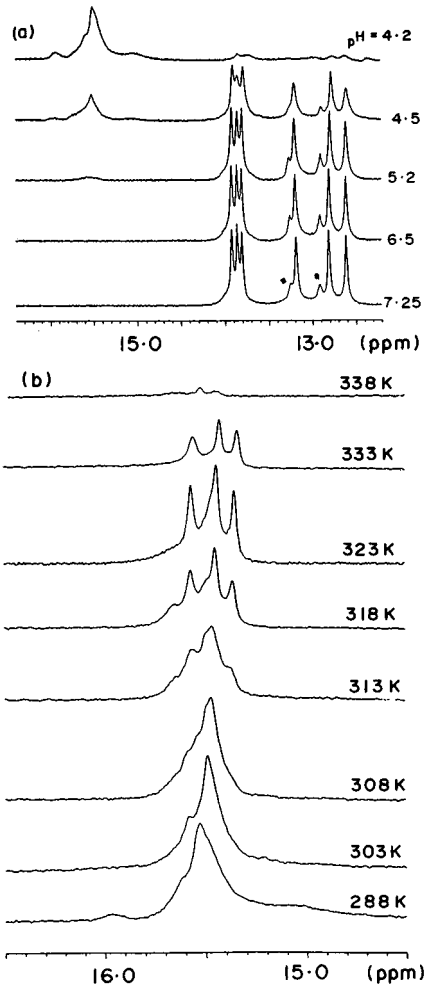

FIGURE 10: (a) Imino proton resonances of the 16-mer as a function of pH at 293 K. The asterisk indicates the peaks belonging to the hairpin. These spectra have been recorded on a 600 MHz varian unity plus spectrometer. (b) Temperature dependence of the C^+ imino peaks at 15−16 ppm, pH 4.2. These spectra have been recorded on a 500 MHz bruker AMX spectrometer, since the gradient probes on the varian machine had the limitation of going to temperatures >50 °C.

may occur due to ionizable groups present in DNA-binding proteins and other surrounding molecules inside the cell. This renders studies of pH effects on DNA structure significant.

Figure 10a shows the behavior of the imino proton resonances of the 16-mer as a function of pH of the solution. We notice that as the pH is lowered, a new set of peaks appear at 15−16 ppm and at the same time the G and T imino peaks belonging to GC and AT pairs start decreasing in intensity. Clearly a new species is emerging at the cost of the dumbbell. In Figure 10b the temperature dependence of the new peaks at 15−16 ppm at pH 4.2 is shown. Interestingly, the resonances are stable up to 60 °C, indicating that the new structure that is formed is a very stable one. One can count atleast six separate peaks at 313 K, and any new structure must account for this many peaks. The chemical shifts of these peaks suggest their origin to be C^+ iminos involved in some sort of H-bonding interactions. Protonated cytidines have been seen earlier in the literature in two situations: (1) C^+−GC triplex formation in which C^+ imino pairs with G in a Hoogsteen fashion (Wells et al., 1988; Haner & Dervan, 1990; Sklenar & Feigon, 1989; Radhakrishnan et al., 1991a,,b, 1992). In this case the amino protons of the C^+ are shifted downfield to ∼10 ppm. (2) C^+−C pairing: This has been seen in the parallel stranded duplex in the i-motif structure (Gehring et al., 1993; Leroy et al., 1993). The pairing scheme involves N3-H−N3 hydrogen bonds and three H-bonds for each C^+−C pair. The proton at the N3 position exchanges rapidly between the two equivalent sites, and therefore the amino protons on the two cytidines will be pairwise equivalent; that is, the two internal H-bonded NH_2 will be equivalent and the two external amino protons will also be equivalent. Furthermore, because of the rapid exchange of the N−H protons, the C−NH_2 proton chemical shifts will be averages of the chemical shifts in the protonated and unprotonated cytosines and will occur at ∼9 ppm for the internal H-bonded amino proton and at ∼8 ppm for the external free amino proton (Gehring et al., 1993; Leroy et al., 1993). The observation of the C-amino resonances ∼9.2 and 8.0 ppm and observation of NOE cross peaks between these amino protons and the imino protons at 15−16 ppm in the present case (Figure 11b) are consistent with the C−C^+ pairing scheme mentioned above.

The fact that we can count at least six C^+ imino resonances in the present case indicates in the first place that a parallel stranded C−C^+ paired duplex is formed. Furthermore, the high stability of the C−C^+ base pairs (Figure 10b) and distinct imino−imino, imino−sugar, amino−amino, and amino−sugar NOE cross peaks shown in Figures 11 and 12 seem to indicate that an i-motif type of structure may have been formed in which two parallel stranded duplexes running in antiparallel directions interdigitate each other; C^+NH_2−(H2′, H2″) cross peaks are characteristic of such a structure. A similar structural transition from a hairpin to an i-motif has also been reported earlier (Rohozinski et al., 1994) in an oligonucleotide containing a polycytidine stretch. This should not be surprising considering that i-motif structures are invariably formed by running C-stretches at low pH conditions (Leroy et al., 1993). The NMR data did not show any clear evidence for G-G or T-T pairing at room temperature and above in the parallel stranded duplexes. Thus, it appears that pairing and interdigitation are restricted to the six unit long C stretch in the molecule at room temperature. The amino−amino cross peaks among the H-bonded protons at 9.0−9.3 ppm (Figure 12a) suggest good stacking of the exocyclic amino groups. Likewise, the C^+ imino−sugar NOEs shown in Figure 11c add further support to the above conclusions. However, the other i-motif characteristic NOEs, namely the H1′−H1′ NOEs (Gehring et al., 1993), were not discernible in our spectra because of poor chemical shift dispersion among the H1′ protons of the cytidines. These same difficulties were encountered by Gueron and coworkers (Leroy et al., 1993) in the case of several cytidine stretches they studied, and sequence specific assignment of the sugar protons could not be obtained.

Interdigitation of the C−C^+ duplexes formed by the present 16-mer DNA can, in principle, occur in two ways, as shown below (only one strand of each duplex is shown for clarity).

5′-G1---A2----T3----C4---- T5-----T6----C_7---C_8----C_9---C_{10}---C_{11}---C_{12}---G13---G14----A15---A16-3′
3′- A16-- A15---G14---G13--C_{12}---C_{11}---C_{10}----C_9----C_8----C_7-----T6------T5---- -C4----T3---A2----G1--5′

or

5′-G1---A2-----T3----C4---- T5-----T6-----C_7---C_8----C_9---- C_{10}---C_{11}---C_{12}--G13--G14----A15---A16-- 3′
3′- A16-- A15---G14---G13--C_{12}---C_{11}---C_{10}----C_9----C_8----C_7-----T6----T5--- -C4----T3----A2----G1--5′

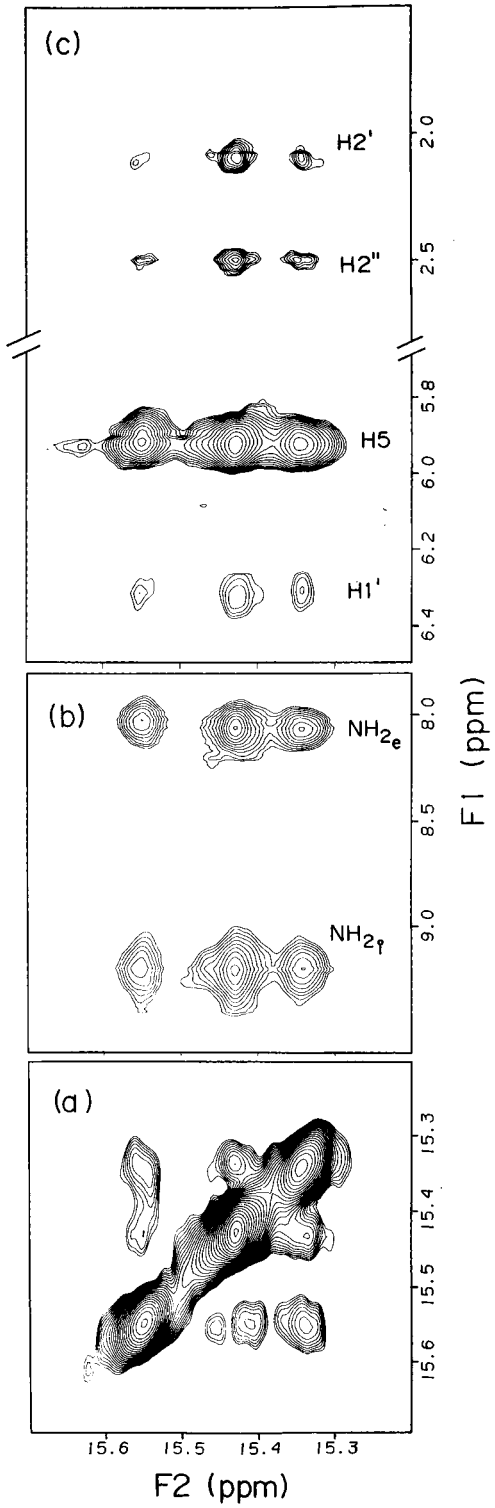

FIGURE 11: C^+NH-C^+ NH (a), $C^+NH-C^+NH_2$ (b), and C^+NH-(H1′, H2′, H2″) (c) cross peaks in the NOESY spectrum of $C-C^+$ paired structure of the 16-mer at pH 4.2, temperature = 40 °C, and mixing time = 200 ms.

Clearly, these two arrangements would be nonequivalent because the stacking neighbors for every nucleotide unit are different in the two cases. The present NMR data do not favor any of these. It is possible that both make contributions; greater than the expected number of peaks are seen in some of the imino and amino proton regions in the 2D NOESY spectra. This is in contrast to the TC_5 molecule for which only one type of stacking was observed. Therefore, in the present case, it appears that the neighboring nucleotide units provide some stability to both stacking patterns.

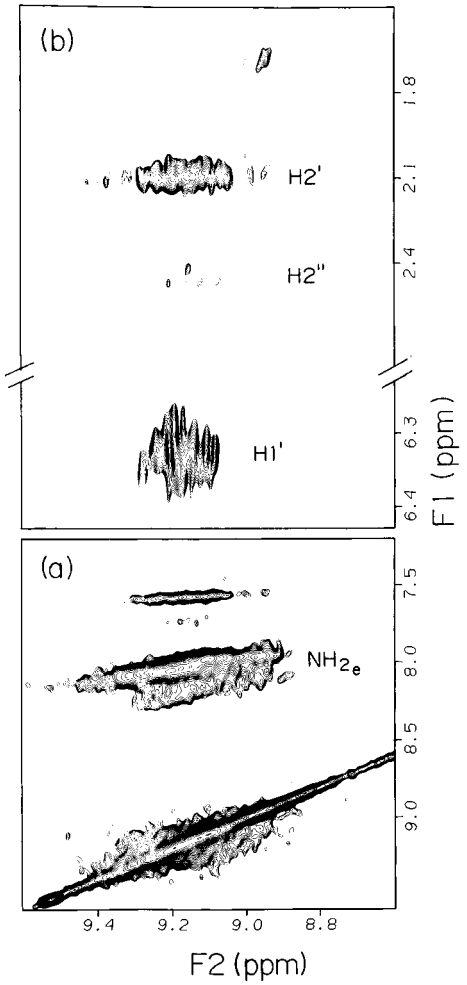

FIGURE 12: Cytidine amino−amino (a) and amino−(H1′, H2′, H2″) (b) cross peaks from the same NOESY spectrum as in Figure 11. The amino−sugar NOEs are characteristic of the i-motif structure.

Finally, a few points are worth making at this stage to put the results of the present work in proper perspective. The structure determinations here have relied entirely on the NOE data, and no input has been obtained from the coupling constants. The reason for this is that the *J*-correlated spectra having antiphase multiplets had very poor cross peaks both in the dumbbell and in the related stem duplex. This must be a consequence of large line widths vis-à-vis the $^1H-^1H$ coupling constants in the DNA backbone. Several factors may be contributing to these large line widths. In the case of the duplex, one factor we could detect was the end to end stacking of the duplexes, resulting in a large molecular mass. In the case of the dumbbell, lower stability and consequent exchange broadening would contribute to the larger line widths. Although the NOE intensity calculation algorithm does not have provisions for such fraying effects, an effective long correlation time may partly compensate for the approximations. For the same reasons we have used a rather loose NOE fit criterion. Even so, the fact that we could pick only a handful of convergent structures from simulated annealing places confidence on the "goodness" of the structures derived.

CONCLUSION

We have investigated in this paper DNA polymorphism exhibited by a particular DNA segment using NMR spec-

troscopy. Determination of the solution structure of the 16-mer DNA, which forms a dumbbell with a 12 base pair stem, having two nicks in the center and 4 base C−loops at either end, and also of the 12-mer duplex alone without the nicks independently to atomic resolution has allowed characterization of the effects of the nicks on the structure and stability of the duplex. It is noted that even in the presence of the nicks, the structure is highly ordered with good stacking of the base pairs. Internucleotide NOE connectivities are observed through the nicks as well. The stacking of the base pairs near the nicks is, however, slightly different compared to that in the normal duplex. This could have important implications for recognition by specific enzymes such as ligases, DNA polymerase, and other DNA-repair enzymes.

The structure determinations have also revealed the structural consequences of the loops. We noticed that the loops shielded the last base pair from the solvent and thus reduced fraying effects at the ends of the stem. It is also clear that the loop orientation can be very different with respect to the stem, and this suggests different topological forms for hairpins in aqueous solutions. Considering that the hairpins and cruciforms can be intermediates in several DNA functions, these structural features are expected to be of significance.

The pH-dependent transformations in the 16-mer observed here are significant from the point of view of DNA folding and packing inside a cell. Local pH changes in DNA can occur due to ionizable groups in the proteins binding to DNA, and these may cause transformations depending upon the base sequences at these sites. The pH changes also affect the stabilities of the duplexes, and once again different forms may be formed depending upon the base sequences.

ACKNOWLEDGMENT

We thank the National Facility for High Field NMR for experimental and computational facility. We thank Dr. M. V. Hosur and K. K. Kannan of Bhabha Atomic Reaserch Center for X-PLOR usage. NUPARM was a kind gift from Dr. M. Bansal of Indian Institute of Science, Bangalore.

REFERENCES

Ahmed, S., Kintakar, A., & Henderson, E. (1994) *Nat. Struct. Biol. 1*, 83−88.

Bonvin, M. J. J., Rullman, J. A. C., Lamerichs, M. N. J., Boelens, R., & Kaptein, R. (1993) *Proteins 15*, 385.

Bonvin, A. M. J. J., Boelens, R., & Kaptein, R. (1994) *J. Biomol. NMR 4*, 143.

Braun, W., & Go, N. (1985) *J. Mol. Biol. 186*, 611−626.

Broitman, S. L., & Fresco, J. R. (1989) *Prog. Nucleic Acids Res. Mol. Biol.*

Brunger, A. (1992) *X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Chen, L., Cai, L., Zhang, X., & Rich, A. (1994) *Biochemistry 33*, 13540−13546.

Cheng, K. Y., & Tinoco, I. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 8705−8709.

Claverys, J. P., Majean, V., Gasc, A. M., & Sicard, A. M. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5956−5960.

Cooney, M., Czernuszewiczand, G., Postel, E. H., Flint, S. J., & Hogan, M. E. (1988) *Science 241*, 456.

Doan, T. L., Perronault, L., Praseuth, D., Habhoub, N., Decout, J. L., Thuong, N. T., Lhomme, J., & Helene, C. (1987) *Nuclic Acids Res. 15*, 7749−7760.

Dohet, C., Wagner, R., & Radman, M. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 503−505.

Fowler, R. G., Degen, G. E., & Cox, E. C. (1974) *Mol. Gen. Genet. 133*, 179−191.

Gehring, K., Leroy, J. L., & Gueron, M. (1993) *Nature 363*, 561−565.

Goodman, S. D., & Nash, H. A. (1989) *Nature 341*, 251−254.

Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870.

Haner, R., & Dervan, P. B. (1990) *Biochemistry 29*, 9761−9765.

Hosur, R. V., Govil, G., & Miles, H. T. (1988) *Magn. Reson. Chem. 76*, 218−223.

Htun, H., & Dahlberg, J. E. (1989) *Science 243*, 1571−1576.

James, T. L., Ed. (1996) *Methods in Enzymology*, Vol. 261, Academic Press, New York.

Jeener, J., Meier, B. H., Bachman, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.

Karmer, B., Karmer, W., & Fritz, H. J. (1984) *Cell 38*, 878−887.

Lefevre, J. F., Lane, A. N., & Jardetzky, O. (1987) *Biochemistry 26*, 5076−5090.

Leroy, J. L., Gehring, K., Kettani, A., & Gueron, M. (1993) *Biochemistry 32*, 6019−6031.

Loeb, L. A., & Kunkel, T. A. (1982) *Annu. Rev. Biochem. 51*, 429−457.

Lu, A. L., Welsh, K., Clark, S., Su, S. S., & Moldrich, P. (1984) *Cold Spring Harbor Symp. Quant. Biol. 49*, 589−596.

Mather, L. J., Wold, B., & Devan, P. B. (1989) *Science 245*, 725.

Majumdar, A., & Hosur, R. V. (1992) *Prog. Nucl. Magn. Reson. Spectrosc. 24*, 109−158.

Muller, L. (1979) *J. Am. Chem. Soc. 101*, 4481−4484.

Nibedita, R., Kumar, R. A., & Hosur, R. V. (1992) *J. Biomol. NMR 2*, 467−476.

Panayotatos, N., & Fontaine, A. (1994) *J. Biol. Chem. 262*, 11364−11368.

Patel, D. J., Shapiro, L., & Hare, D. (1987) *Q. Rev. Biophys. 20*, 35−112.

Piotto, M., Saudik, V., & Sklenar, V. (1992) *J. Biomol. NMR. 2*, 661−665

Plateau, P., & Gueron, M. (1982) *J. Am. Chem. Soc. 104*, 7310−7311.

Radha, P. K. (1997) *Magn. Reson. Chem. 34*, S18−S22.

Radhakrishnan, I., Gao, X., de Los Santos, C., Live, D., & Patel, D. J. (1991a) *Biochemistry 30*, 9022−9030.

Radhakrishnan, I., Patel, D. J., & Gao, X. (1991b) *J. Am. Chem. Soc. 113*, 8542−8544.

Radhakrishnan, I., Patel, D. J., & Gao, X. (1992) *Biochemistry 31*, 2514−2523.

Rohozinski, J., Hancock, J. M., & Keniry, M. A. (1994) *Nucleic Acids Res. 22*, 4653−4659

Sklenar, V., & Feigon, J. (1989) *Nature 345*, 836−838.

Stevens, S. Y., Swanson, P. C., Voss, E. W. J., & Glick, G. D. (1993) *J. Am. Chem. Soc. 115*, 1585−1586.

Van de ven, F. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1−32.

Wellinger, R. J., Wolf, A. J., & Jakian, V. A. (1993) *Cell 72*, 51−60.

Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., & Wohlraub, F. (1988) *FASEB J. 2*, 2939−2949.

Withka, J. K., Swaminathan, S., Srinivisan, J., Beveridge, D. L., & Boltan, P. H. (1992) *Science 255*, 597−599.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.

Wyatt, J. R., Vickers, T. A., Roberson, J. L., Buckeheit, R. W. J., de Baets, T. K. E., Davies, P. W., Rayner, B., Imbach, J. L., & Ecker, D. J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1356−1360.

Yip, P., & Case, D. A. (1989) *J. Magn. Reson. 83*, 643−648.

Zimmerman, S. B. (1976) *J. Mol. Biol. 106*, 663−672.

Zimmerman, S. B., Cohen, G. H., & Davies, D. R. (1975) *J. Mol. Biol. 92*, 181−192.

BI970819Y