Two-Dimensional ^1H and ^{31}P NMR Spectra and Restrained Molecular Dynamics Structure of an Extrahelical Adenosine Tridecamer Oligodeoxyribonucleotide Duplex†

Edward Nikonowicz, Vikram Roongta, Claude R. Jones, and David G. Gorenstein*

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907

Received March 24, 1989; Revised Manuscript Received June 19, 1989

ABSTRACT: Assignment of the ^1H and ^{31}P NMR spectra of an extrahelical adenosine tridecamer oligodeoxyribonucleotide duplex, $d(CGCAGAATTCGCG)_2$, has been made by two-dimensional ^1H–^1H and heteronuclear ^{31}P–^1H correlated spectroscopy. The downfield ^{31}P resonance previously noted by Patel et al. (1982) has been assigned by both ^{17}O labeling of the phosphate as well as a pure absorption phase constant-time heteronuclear ^{31}P–^1H correlated spectrum and has been associated with the phosphate on the 3′ side of the extrahelical adenosine. $J_{H3'-P}$ coupling constants for each of the phosphates of the tridecamer were obtained from the ^1H–^{31}P J-resolved selective proton-flip 2D spectrum. By use of a modified Karplus relationship the C4–C3′–O3–P torsional angles (ϵ) were obtained. There exists a good linear correlation between ^{31}P chemical shifts and the ϵ torsional angle. The ^{31}P chemical shifts and ϵ torsional angles follow the general observation that the more internal the phosphate is located within the oligonucleotide sequence, the more upfield the ^{31}P resonance occurs. Because the extrahelical adenosine significantly distorts the deoxyribose phosphate backbone conformation even several bases distant from the extrahelical adenosine, ^{31}P chemical shifts show complex site- and sequence-specific variations. Modeling and NOESY distance-restrained energy minimization and restrained molecular dynamics suggest that the extrahelical adenosine stacks into the duplex. However, a minor conformation is also observed in the ^1H NMR, which could be associated with a structure in which the extrahelical adenosine loops out into solution.

Single-base insertions into DNA duplexes have increasingly become a topic of structural investigation (Fink & Crothers, 1972; Fresco & Alberts, 1960; Hare et al., 1986; Lee & Tinoco, 1978; Morden et al., 1983; Patel et al., 1982; Roy et al., 1986, 1987; Joshua-Tor et al., 1988; Miller et al., 1988; Kalnik et al., 1989). Interest in the fine structure of extrahelical bases has been prompted by the naturally occurring frame-shift mutations and the issues of how the cells repair and recognize such structures.

An extrahelical base at the interior of a duplex oligonucleotide may be accommodated by either stacking into the helix or looping out into solution or possibly even causing a disruption of the duplex to form a hairpin. The hairpin loop structure appears to be favored in the oligonucleotide d(CGCGAATTACGCG) possessing an extrahelical adenosine at position 9 (Roy et al., 1986). The authors were able to follow the hairpin loop to duplex transition by monitoring the thymidine methyl resonances by 1D NMR.

A 15-residue oligonucleotide, d(CGCGAAATTTACGCG), related to the Roy et al. tridecamer, and the oligonucleotide d(CGCAGACGTCGCG) have been shown by 2D NMR to be stabilized in the duplex form with the extrahelical adenosine in each stacking into the duplex [Hare et al. (1986a,b) and Roy et al. (1987), respectively]. Earlier, through the use of various 1D NMR techniques, it was shown that the unpaired central pyrimidine residue, cytidine, of the duplex sequence $d(CA_3CA_3G)\cdot d(CT_6G)$ loops out into solution rather than stacking into the duplex (Morden et al., 1983). Recently, the crystal structures of two oligonucleotides containing an extrahelical adenosine (Joshua-Tor et al., 1988; Miller et al., 1988) have shown that the unpaired residue in each case loops out rather than stacking into the duplex. The latter of these two duplexes is the same sequence as the pentadecamer mentioned above; the former is the tridecamer d(CGCAGAATTCGCG). This tridecamer sequence had previously been studied in solution by Patel and co-workers using a combination of calorimetric and 1D NMR techniques, and it was concluded that the extrahelical adenosine stacks into the duplex (Patel et al., 1982; Pardi et al., 1982; Patel, 1982).

^{31}P NMR. ^1H NMR is now able to provide detailed information on the three-dimensional structure and dynamics of oligonucleotide duplexes and nucleic acid complexes (Van De Ven & Hilbers, 1988); however, of the six torsional angles that largely define the backbone structure, only the four involving the deoxyribose ring have been shown to be amenable to analysis by ^1H NMR techniques. Indeed, even for modest-sized oligonucleotide duplexes, determination of these torsional angles via three-bond coupling constants is quite difficult, and most structural studies have relied upon constraints solely derived from the NOESY spectra (i.e., intraresidue and interresidue proton–proton distances less than 5 Å). Our laboratory has proposed that ^{31}P NMR spectroscopy is capable of providing information on the most important remaining two torsional angles involving the phosphate ester bonds. We have noted that ^{31}P chemical shifts can potentially provide a probe of the conformation of the phosphate ester backbone in nucleic acids and nucleic acid complexes (Gorenstein, 1978, 1981, 1984; Gorenstein & Findlay, 1976;

†Supported by NIH (GM36281 and AI27744), the Purdue University Biochemical Magnetic Resonance Laboratory, which is supported by NIH (Grant RR01077 from the Biotechnology Resources Program of the Division of Research Resources), the NSF National Biological Facilities Center on Biomolecular NMR, Structure and Design at Purdue (Grants BBS 8614177 and 8714258 from the Division of Biological Instrumentation), and the National AIDS Research Center at Purdue (AI727713).

0006-2960/89/0428-8714$01.50/0

Gorenstein & Goldfield, 1984; Gorenstein et al., 1987a,b). These studies suggested that the ^{31}P resonance of a phosphate diester in a *g,g* or g^-,g^- conformation should be several ppm upfield of the ^{31}P signal of an ester in a *g,t* or *t,t* conformation (Gorenstein, 1984). In combination with the measured P–H3′ coupling constant and a Karplus-type relationship we can now also directly determine the C4–O3′ torsional angles (and indirectly the P–O3′ torsional angles; Gorenstein et al., 1988; Sklenar & Bax, 1987). Thus both ^{31}P chemical shifts and ^{31}P coupling constants are now able to provide information on the deoxyribose phosphate backbone conformation of duplex DNA.

With the recent development of methodologies to assign individual ^{31}P resonances and measure P–H3′ coupling constants of oligonucleotides (Fu et al., 1988; Gorenstein et al., 1987a,b, 1988; Lai et al., 1984; Ott & Eckstein, 1985a,b; Petersheim et al., 1984; Schroeder et al., 1986; Shah et al., 1984a; Sklenar & Bax, 1987) we and others also have been able to begin to understand some of the factors responsible for ^{31}P chemical shift variations in oligonucleotides (Cheng et al., 1987, 1982; Lai et al., 1984; Ott & Eckstein, 1985b; Patel, 1974, 1979; Schroeder et al., 1986; Shah et al., 1984a). As discussed above, our laboratory has suggested that one of the major determinants of ^{31}P chemical shifts is the main-chain torsional angles of the individual phosphodiester groups along the oligonucleotide double helix. Phosphates located toward the middle of a B-DNA double helix should assume the lower energy, stereoelectronically (Gorenstein, 1987) favored g^-,g^- conformation, while phosphodiester linkages located toward the two ends of the double helix, where increased flexibility of the helix is more likely to occur, should adopt a mixture of g^-,g^- and t,g^- conformations. [The notation for the P–O ester torsion angles follows the convention of Seeman et al. (1976) with the ζ, P–O3′, angle given first, followed by the α, P–O5′, angle.] Because the g^-,g^- conformation corresponds to a more upfield ^{31}P chemical shift, while a t,g^- conformation corresponds to a lower field chemical shift, internal phosphates in oligonucleotides would be expected to be upfield of those nearer the ends. Although several exceptions have been noted, this positional relationship is commonly observed in oligonucleotides where ^{31}P chemical shift assignments have been determined (Gorenstein, 1978; Gorenstein et al., 1976; Ott & Eckstein, 1985a,b; Schroeder et al., 1986, 1987). However, variation of ^{31}P chemical shifts with position is only part of the story.

Eckstein and co-workers (Connolly & Eckstein, 1984; Ott & Eckstein, 1985a,b) and our laboratory (Gorenstein et al., 1988; Schroeder et al., 1986, 1987) have recently suggested that ^{31}P chemical shifts are also sensitive to sequence-specific structural variations of the double helix as proposed by Calladine (1982) and Dickerson (1983). The ^{31}P chemical shifts of duplex B-DNA phosphates correlate reasonably well with some aspects of the Dickerson/Calladine sum function for variation in the helical twist (or helical roll) of the individual base steps in the oligonucleotides. In the B-form DNA double helix with ~10 base pairs per turn (360°) of the helix each base pair is rotated ca. +36° (helix twist) with respect to the nearest-neighbor base pair. However, analysis (Dickerson, 1983) of X-ray crystallographic structures of a number of duplexes has revealed large variations in local helix structure, with helix twist varying from 25° to 45°. Correlations between experimentally measured P–O and C–O torsional angles and results from molecular mechanics/dynamics energy minimization calculations show that these results are generally consistent with the hypothesis that sequence-specific variations in ^{31}P chemical shifts are at least partly attributable to sequence-specific changes in the deoxyribose phosphate backbone (Gorenstein et al., 1988).

Patel and co-workers (Patel et al., 1982) have shown that the extrahelical adenosine in the tridecamer d(CGCAGAATTCGCG) provides some very interesting ^{31}P spectral shifts. Whereas the ^{31}P spectral dispersion is <0.7 ppm in normal B-DNA double helices, new signals are shifted upfield and downfield from the "normal" double-helical phosphate ^{31}P signals of the parent dodecamer duplex. In this paper, ^{31}P NMR and two-dimensional ^{1}H/^{1}H and ^{1}H/^{31}P NMR methods have been used to assign these anomalous, "site-specific" ^{31}P signals. The tridecamer thus allows an analysis of position-specific, sequence-specific, and site-specific effects on ^{31}P chemical shifts. Finally, the NOESY-derived ^{1}H–^{1}H distances have been used in conjunction with restrained molecular dynamics calculations to examine the structure of the extrahelical adenosine-containing tridecamer duplex.

Experimental Procedures

Synthesis and Sample Preparation. The self-complementary tridecamer duplex d(CGCAGAATTCGCG)$_2$ was synthesized by a manual modification of the phosphite triester method on a solid support (Lai et al., 1984; Schroeder et al., 1987; Shah et al., 1984a,b). The porous glass derivatized dG support was synthesized as described (Gait, 1984). We have introduced the ^{17}O label into the phosphoryl group by replacing the I_2/H_2O in the oxidation step of the phosphite by $I_2/H_2{}^{17}O$ (Monsanto $H_2{}^{17}O$, 51.5% ^{17}O, 35.6% ^{18}O, 12.9% ^{16}O) as previously described (Schroeder et al., 1987; Shah et al., 1984b). After cleavage from the support and deprotection, the resulting oligonucleotides were purified by reverse-phase HPLC as previously described (Gorenstein et al., 1984) followed by dialysis. The 13-mer was then passed through a cation-exchange column (Dowex-50× K^+ form), dialyzed, lyophilized, and redissolved in 400 μL of D_2O. The samples used in the COSY and NOESY experiments were prepared by dissolving approximately 460 OD units of the tridecamer in 400 μL of D_2O (99.995%) buffer containing 100 mM potassium phosphate, 100 mM KCl, and 0.1 mM NaN_3 at pH* 7.1 (uncorrected pH meter reading). The sample used in the ^{31}P–^{1}H PAC experiment contained approximately 520 OD units of 13-mer dissolved in 0.4 mL of D_2O containing 100 mM KCl and 1 mM NaN_3 at pH* 7.1.

NMR. The COSY and NOESY spectra were recorded on a Nicolet NT-470 (470-MHz ^{1}H) spectrometer at ambient temperature. The COSY spectrum was acquired with 2048 points in the t_2 dimension and 256 t_1 increments. The data were apodized with an unshifted sine bell before Fourier transformation in both dimensions. The phase-sensitive TPPI (Marion & Wüthrich, 1983) NOSEY spectrum was acquired with 2048 points in the t_2 dimension and 256 t_1 increments. Mixing times of 50, 100, 150, 200, 250, 300, 400, 500, and 700 ms were used. The 400-ms mixing time NOESY data were treated with a small negative exponential and a small Gaussian multiplication in the t_1 and t_2 dimensions.

The ^{31}P 1D NMR spectra, the ^{31}P melting profiles, the ^{1}H/^{31}P correlation PAC spectra, and the ^{1}H/^{31}P heteronuclear 2D *J*-resolved spectrum were run on a Varian XL-200A NMR spectrometer (200 MHz, ^{1}H). All experiments were run in 99.998% D_2O. Typical ^{31}P 1D NMR parameters were as follows: sweep width was 172 Hz; acquisition time was 2.98 s; block size was 1K zero-filled to 16K; pulse width was 7.0 μs; spectra were resolution enhanced with a Gaussian apodization function; the number of acquisitions was between 2000 and 3000. The ^{31}P spectra were referenced to an external

sample of 85% phosphoric acid which was recorded at each temperature. The spectra are reported relative to trimethyl phosphate (TMP) at 0.000 ppm, which is 3.456 ppm downfield of 85% phosphoric acid.

A $^{31}P/^{1}H$ pure absorption phase constant-time (PAC) version (Fu et al., 1988; Jones et al., 1988) of the Kessler–Griesinger long-range heteronuclear correlation (COLOC) experiment (Kessler et al., 1984) was conducted on the tridecamer at ambient temperature. The PAC spectrum provides chemical shift correlation between the H3′, H4′, and H5′ protons of the deoxyribose rings and the three- or four-bond coupled phosphorus. Coupling occurs between the phosphorus and the H3′ proton on the 5′ side and the H4′ and H5′ protons on the 3′ side of the dinucleotide fragment. The data set was acquired with 256 points in the ^{31}P dimension and 45 t_1 increments and then zero-filled to 512 × 256. The data set was multiplied by a combination of an increasing exponential and Gaussian function before Fourier transformation in both dimensions. ^{1}H NMR resonances are externally referenced to 3-(trimethylsilyl)propanesulfonic acid sodium salt (DSS), which is 4.76 ppm upfield of the internal HDO reference, and ^{31}P NMR resonances are referenced to an external sample of 85% phosphoric acid.

The Bax–Freeman selective 2D J-resolved long-range correlation experiment with a Dante sequence for the selective 180° pulse (Sklenar & Bax, 1987) was performed on the tridecamer to correlate the ^{31}P chemical shift with the phosphorus–H3′ coupling constant. The ^{31}P–^{1}H *J*-resolved spectra were recorded at 18–70 °C. The data sets were acquired with 256 points in the ^{31}P dimension and 32 t_1 increments and then zero-filled to 512 × 128. Resolution enhancement of the Gaussian type was applied before Fourier transformation in both dimensions. *J* values were measured from peak center to peak center.

The observed three-bond coupling constant is used with a proton–phosphorus Karplus relationship to measure the H3′–C3′–O–P torsional angle θ from which we have calculated the C4′–C3′–O–P torsional angle ϵ ($=\theta - 240°$). The relationship $J = 15.3 \cos^2 \theta - 6.1 \cos \theta - 1.6$ was determined by Lankhorst et al. [1984; see also Sklenar and Bax (1987)].

NOESY Distance Restrained Structure. Two initial model built structures for the extrahelical adenosine tridecamer duplex were considered. A base loop-out structure similar to the X-ray structure of other extrahelical base duplexes and a base insert structure were first modeled by using the MIDAS program. The AMBER program was then used to calculate an energy-minimized structure for the initial models. The structures were then reminimized with the incorporation of 218 NOESY distance constraints (from the 50-ms mixing time spectrum) and then subjected to 15 ps of restrained molecular dynamics.

RESULTS

Nonexchangeable Proton Assignments of the Tridecamer. Assignment of the proton signals of the tridecamer d-(CGCAGAATTCGCG) was accomplished through analysis of two-dimensional COSY and NOESY spectra via a sequential assignment methodology (Broido et al., 1984; Feigon et al., 1983a,b, Hare et al., 1983; Scheek et al., 1984; Schroeder et al., 1987). The COSY experiment, collected in D_2O, pH = 7.2, 0.1 M KCl, and 22 °C, was used to locate the H6 resonances of the four cytidines through the off-diagonal cross-peaks of the scalar-coupled H5/H6 protons. Additionally, the sugar ring proton, H1′–H2′/2″ and H2′/2″–H3′, connectivities could be traced in the COSY spectrum (spectra not shown are available upon request).

However, assignment of most of the nonexchangeable proton resonances of the tridecamer duplex was accomplished by inspection of 2D NOESY spectra collected at mixing times of 150 and 400 ms under the previous conditions (Figure 1). Although a B-DNA-type geometry was not initially assumed, the sequential assignment of this tridecamer followed that for a right-handed DNA helix (Hare et al., 1983; Scheek et al., 1984; Weiss et al., 1984). In B-DNA, each pyrimidine H6 or purine H8 base proton is situated such that it is able to give rise to an NOE with the H1′ proton of its own deoxyribose ring as well as the H1′ proton of the 5′-flanking sugar (region A, Figure 1). The 5′-terminal base is easily identified since only a single intraresidual base–H1′ NOE cross-peak is possible. Further, the 3′-terminal H1′ proton is identified by its lack of an NOE cross-peak to a 3′-flanking base proton, thus allowing the base–H1′ region to be sequentially assigned from either end. It should be noted here that this sequential assignment is not interrupted through the C3–A3′–G4 stretch [numbering is based upon the Dickerson dodecamer, d-(CGCGAATTCGCG), with the extrahelical adenosine between residues 3 and 4 labeled as residue 3′] containing the extrahelical adenosine although the A3′ H8–G4 H1′ connectivity is weak. The C9–G10 step, that which is directly opposite the extrahelical adenosine, does not produce the expected C9 H1′–G10 H8 interaction at 150 ms (however, the interaction is seen in the 250- and 400-ms mixing time NOESY spectra; see Figure 1B). This crucial cross-peak thus could represent a secondary NOE arising from spin diffusion, and interpretations of NOESY spectra at these long mixing times are therefore suspect.

Additional NOESY cross-peaks from the H5 of cytidines 3, 9, and 11 to the H6/H8 base proton of the $i - 1$ residue are also present. Interstrand connectivities for the H2 protons of A5 and A6 are also present (Figure 1B). The base–base region, 7.0–8.2 ppm, contains several sequential connectivities consistent with the base–sugar ring proton regions of the spectrum. Of particular interest is the presence of a C3 H6–A3′ H8 cross-peak and the absence of a C9 H6–G10 H8 cross-peak.

Identification of the H1′ resonances allows further assignment of protons H2′, H2″, and H3′. Fortunately, only the H1′ protons of C1 and C12 were nearly degenerate, making most of the other eleven H1′–H2′,2″ and 3′ connectivities straightforward. An example of the base to H2′, 2″ region is shown in Figure 1C. Stereospecific assignments of the H2′ and H2″ protons were achieved by analysis of cross-peak intensities. Due to its closer spatial proximity in B-DNA, the H2″ proton will generate a more intense NOESY cross-peak than that produced by the interaction of H2′ with H1′ (Feigon et al., 1983; Scheek et al., 1984).

Finally, it was possible to identify the H4′ resonances and a few of the H5′/5″ resonances through their NOESY connectivities with the H1′ and H3′ sugar protons (Hare et al., 1983; Scheek et al., 1984; Wemmer et al., 1984). These interactions were confirmed by inspection of a 50-ms mixing time NOESY spectrum where little spin diffusion is expected. The proton assignments are listed in Table I.

^{31}P Resonance Assignments of the Tridecamer. The assignment of the resonances in the tridecamer ^{31}P spectrum is based upon a combined oxygen isotope labeling method (Ott & Eckstein, 1985a; Petersheim et al., 1984; Shah et al., 1984b) and a pure absorption phase constant-time $^{1}H/^{31}P$ heteronuclear correlated spectrum, PAC (Fu et al., 1988). Our labeling method is a simple modification of the I_2/H_2O oxidation step in the solid-phase phosphoramidite oligonucleotide synthesis

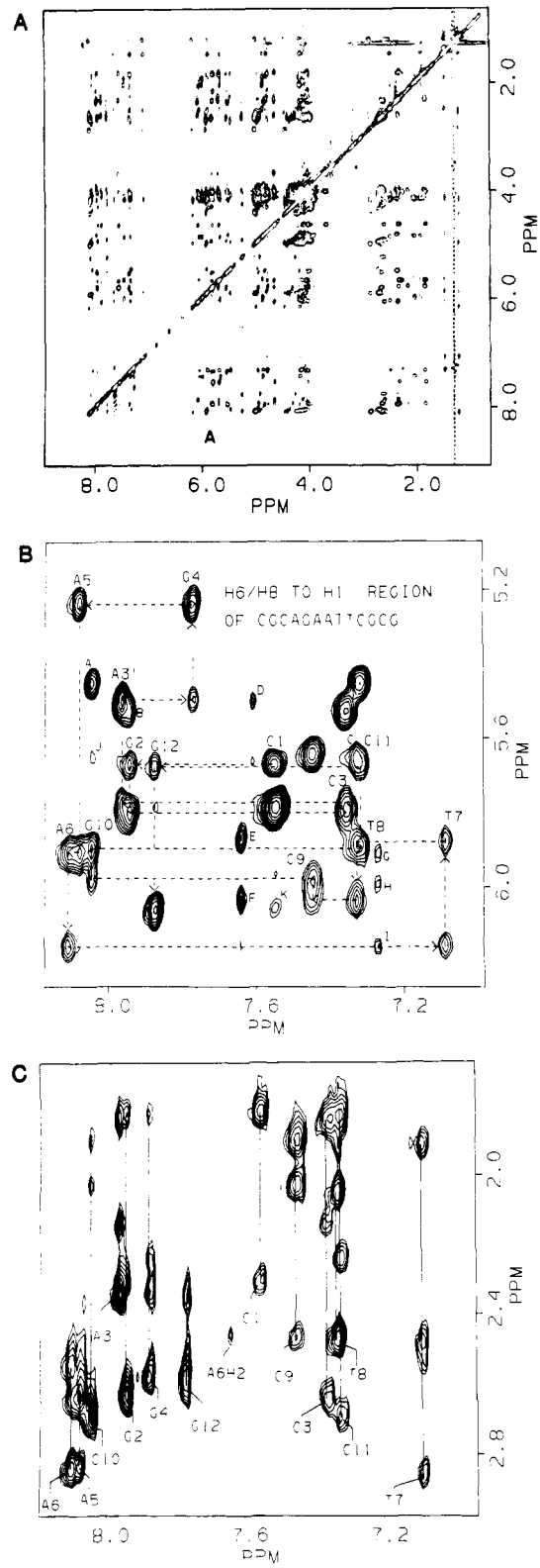

FIGURE 1: (A) Pure absorption phase ^{1}H/^{1}H NOESY NMR spectrum of duplex tridecamer, at 470 MHz, 400-ms mixing time. (B) Expansion of region labeled "A" in panel A. The sequential assignment of the base and deoxyribose H1′ protons is diagrammed. Labeled cross-peaks correspond to G10 H8–C11 H5 (A), G2 H8–C3 H5 (B), T8 H6–C9 H5 (C) A3′ H2–A3′ H1′ (D), A6 H2′–T7 H1′ (E), interstrand A6 H2–T8 H1′ (F), A5 H2–A5 H1′ (G), interstrand A5 H2–C9 H1′ (H), A5 H2–A6 H1′ (I), G10 H8–C9 H5 (J), and C1 H6–G12 H1′ (K). (C) Base to H2′,H2″ region (150-ms mixing time).

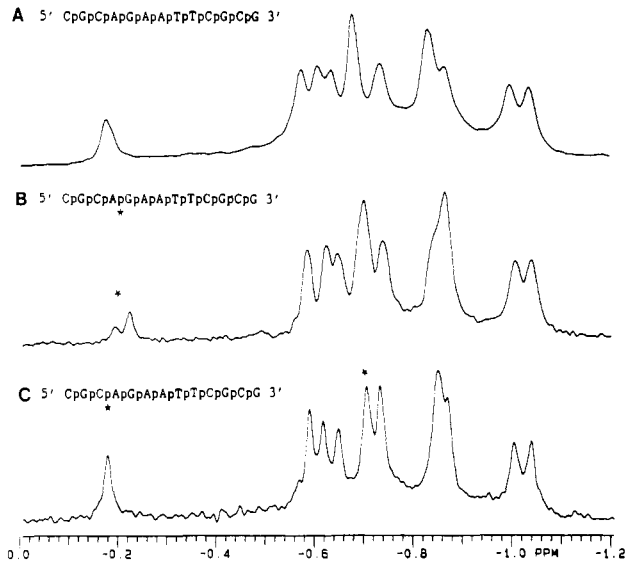

FIGURE 2: (A) ^{31}P NMR spectra of tridecamer. Examples of site-specific ^{17}O labeling of two of the phosphates of the tridecamer at position 3′ and 3 (shown by an asterisk) are shown in (B) and (C), respectively.

scheme with $I_2/H_2{}^{17}O$, thus directly attaching a quadrupolar ^{17}O to a specific phosphorus atom. Inspection of the ^{31}P spectrum reveals an approximate 50% reduction of intensity of the corresponding phosphorus resonance, since ^{17}O-enriched water contains only approximately 50% $H_2{}^{17}O$. Figure 2 shows the ^{31}P spectra of the unlabeled tridecamer and two ^{17}O-labeled tridecamers. It is interesting to note that the resonance of the labeled phosphate is observed as two reduced-intensity, resolved peaks associated with ^{16}O (unlabeled) and ^{18}O-labeled phosphorus resonances. (The $H_2{}^{17}O$ sample also contains both $H_2{}^{16}O$ and $H_2{}^{18}O$.) This can most easily be seen in Figure 2B where the ^{18}O-labeled phosphate ^{31}P signal is shifted slightly upfield relative to the remaining ^{16}O phosphorus resonance (Shah et al., 1984b). As can be seen from the figure, assignment of the CpA and ApG phosphorus resonances is straightforward.

Due to the time and expense involved in labeling each phosphorus atom of the sequence with ^{17}O, we assigned the other ^{31}P signals through a heteronuclear correlation experiment between the assigned ^{1}H signals and the ^{31}P resonances. Rather than using conventional 2D ^{1}H/^{31}P HETCOR NMR spectroscopy (Lai et al., 1984; Pardi et al., 1983), which suffers from poor sensitivity, a heteronuclear version of the "constant-time" coherence-transfer technique (Bax & Morris, 1981), referred to as COLOC and originally proposed for ^{13}C/^{1}H correlations (Kessler et al., 1984), was chosen. We have used a pure absorption phase version of the COLOC sequence (PAC) previously modified and optimized for the ^{31}P to 3′ proton of oligonucleotides in our laboratory (Fu et al., 1988) in conjunction with the 3′ proton assignments from the NOESY spectrum to assign the remaining ^{31}P resonances. The PAC spectrum of the tridecamer is shown in Figure 3. Table II lists the ^{31}P assignments of the 12 phosphates in the tridecamer backbone.

^{31}P Melting Curve. The temperature dependence of the ^{31}P chemical shifts is shown in Figure 4. As previously reported (Patel et al., 1982), and as expected (Gorenstein, 1981, 1984), the main cluster of resonances moves downfield with increasing temperature. Figure 4B shows representative examples of the ^{31}P spectra at various temperatures. It is interesting to note the diminishing intensity of the most downfield resonance, that which is 5′ to G4 (3′ to A3′), starting at around 38 °C and ending at 46 °C, and then again reemerging by 50 °C and gradually shifting upfield into the main ^{31}P cluster. Above 65 °C, the tridecamer is expected to be in a completely sin-

Table I: Nonexchangable Proton Chemical Shifts (ppm) of d(CGCAGAATTCGCG)

residue	H8/H6	H5/CH3/H2	H1′	H2′	H2″	H3′	H4′	H5′/H5″
C1	7.54	5.78	5.67	1.82	2.31	4.64	4.02	3.66, 3.69
G2	7.94		5.82	2.62	2.66	4.94	4.30	3.97, 4.04
C3	7.36	5.52	5.78	1.84	2.14	4.76	4.01	
A3′	7.96	7.60	5.49	2.33	2.36	4.83	4.18	3.89
G4	7.76		5.23	2.58	2.64	4.91	4.25	3.97
A5	8.07	7.27	5.91	2.63	2.84	5.02	4.40	4.10, 4.13
A6	8.10	7.64	6.16	2.55	2.86	4.98	4.45	4.22, 4.24
T7	7.08	1.21	5.88	1.91	2.47	4.79	4.17	
T8	7.32	1.47	6.04	1.90	2.48	4.86	4.14	
C9	7.44	5.64	5.98	1.91	2.01	4.85	4.18 (5)	
G10	8.04		5.89	2.65	2.70	4.96	4.35	3.95, 4.08
C11	7.32	5.45	5.67	1.83	2.25	4.78	4.06	
G12	7.87		6.06	2.59	2.33	4.63	4.12	

Table II: ^{31}P Chemical Shifts (ppm) in the Tridecamer d(CGCAGAATTCGCG)

5′ residue	resonance no.[a]	^{31}P chemical shift (ppm)	^{1}H chemical shift H3′ (PAC)[b]	^{1}H chemical shift H3′ (NOESY)[c]
C1	6	−4.099	4.67	4.64
G2	3	−4.054	4.95	4.94
C3	5	−4.054	4.80	4.76
A3′	1	−3.647	4.86	4.83
G4	8	−4.239	4.93	4.91
A5	10	−4.298	5.06	5.02
A6	12	−4.470	5.01	4.98
T7	11	−4.430	4.83	4.79
T8	7	−4.183	4.89	4.86
C9	9	−4.277	4.88	4.85
G10	4	−4.054	4.98	4.96
C11	2	−3.989	4.80	4.78

[a] Resonances numbered from low field to high field. [b] Chemical shift determined from PAC spectrum. [c] Chemical shift determined from NOESY spectrum.

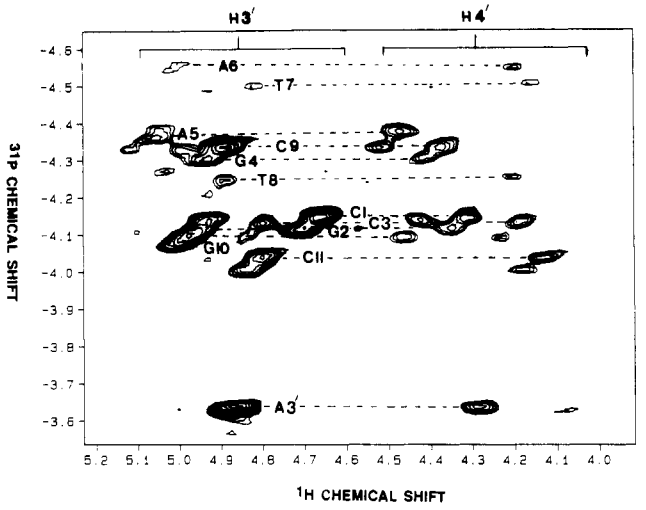

FIGURE 3: Two-dimensional ^{31}P–^{1}H PAC heteronuclear correlation NMR spectrum of duplex tridecamer at 200 MHz (^{1}H). The ^{31}P signals are shown along one axis, and the H3′ and H4′ proton signals are shown along the second axis.

gle-stranded form (Patel et al., 1982; Pardi et al., 1982; Patel, 1982).

J-Correlated Spectra. Since one of the ultimate goals of this project is the elucidation of the fine structure of the tridecamer in solution, it is important to obtain as many constraints as one can in order to generate as "valid" a structure as possible from the NMR data. Unfortunately, ^{1}H/^{1}H NOESY does not provide any significant constraint on the conformation of the phosphate ester backbone (Van De Ven & Hilbers, 1989). One parameter which has yet to be fully exploited in oligonucleotide NMR solution structural determinations is the C–O–P–O–C backbone torsional angles. Two of these torsional angles, ϵ (C4′–C3′–O–P) and β (C4′–C5′–O–P), have three-bond ^{1}H–^{31}P couplings. Unfortunately, due to the complexity of the 5′,5″ region, only the C4′–C3′–O–P torsional angle (ϵ) is easily accessible by a 2D heteronuclear *J*-resolved spectrum. The heteronuclear proton-flip experiment (Bax & Freeman, 1982) as optimized for ^{31}P–H3′ *J* correlations (Sklenar & Bax, 1987) was performed, without reverse detection, on the tridecamer at 30 °C. This temperature was chosen since it is well below the reported melting temperature of 58 °C (Patel et al., 1982; Pardi et al., 1982), yet high enough that the ^{31}P spectrum is well resolved (Figure 4B). The *J*-resolved ^{31}P/^{1}H spectrum is shown in Figure 5. The H3′–P coupling constants for 9 of the 12 resonances have been determined (C3, G10, and C11 coupling constants could not be separately resolved, but as a group exhibited a *J* value of 5.2 Hz). Table III lists the P–H3′ coupling constants. We have utilized these coupling constants and a proton–phosphorus Karplus relationship (see Experimental Procedures) to determine the H3′–C3′–O–P torsional angle θ and thus the C4′–C3′–O–P torsional angle ϵ. Up to four different torsional angles (−180° – +180°) may be derived from the same coupling constant, and we assume that the torsional angle ϵ closest to the crystallographically observed value of ca. −169° (Saenger, 1984) is the correct value. As shown by Dickerson (Dickerson, 1983; Dickerson & Drew, 1981), there is a strong correlation (R = −0.92) between torsional angles ζ and ϵ in the crystal structures of a dodecamer [ζ may be calculated from the relationship (Dickerson, 1983; Dickerson & Drew, 1981) $\zeta = -317 - 1.23\epsilon$]. Thus, assuming this correlation of ζ and ϵ exists for other duplex structures in solution as well, and from the measured coupling constants, we can calculate the C3′–O3′–P–O5′ (ζ) torsional angle. A

Table III: ^{31}P Chemical Shifts, P–H3′ Coupling Constants, and ϵ and ζ Torsional Angles

5′ residue	chemical shift (ppm)	*J*(H–P) (Hz)	torsional angles C–O3′ (ϵ) (deg)	torsional angles P–O3′ (ζ) (deg)
C1	−4.099	6.6	−156.41	−126.18
G2	−4.054	5.2	−163.63	−117.37
C3	−4.054	5.2	−163.63	−117.37
A3′	−3.647	5.5	−162.08	−119.26
G4	−4.239	3.2	−174.60	−103.98
A5	−4.298	1.7	−185.51	−90.67
A6	−4.470	1.9	−183.71	−92.88
T7	−4.430	1.7	−185.51	−90.67
T8	−4.183	4.2	−168.91	−110.93
C9	−4.277	6.6	−156.41	−126.18
G10	−4.054	5.2	−163.63	−117.37
C11	−3.989	5.5	−162.08	−119.26

Table IV: Selected Distances (Å) from the Restrained Molecular Dynamics Refinement[a]

atom pair	50-ms NOESY	stack model	SI	SII	loop model	LI	LII
C3 H1′–A3′ H8	4.20	6.18	4.84	4.83	7.16	3.38	3.39
		6.29	4.90	4.76	7.15	3.93	3.81
A3′ H2–A3′ H1′	4.39	4.54	4.85	4.52	4.59	4.69	4.67
		4.54	4.50	4.36	4.59	4.57	4.56
A3′ H8–A3′ H1′	3.46	3.90	3.78	3.57	3.91	3.82	3.83
		3.91	3.90	3.68	3.91	3.91	3.90
A3′ H8–A3′ H3′	3.89	2.91	4.30	2.88	4.29	4.57	4.56
		2.89	3.29	4.02	4.30	4.37	4.34
A3′ H8–C3 H2″	3.88	3.65	4.20	4.45	6.07	4.63	4.54
		3.63	3.14	3.86	6.04	4.48	4.62
A3′ H1′–A3′ H3′	3.46	3.79	3.82	3.59	3.88	3.93	3.91
		3.78	3.79	3.70	3.87	3.77	3.78
G4 H8–A3′ H1′	4.30	4.60	5.10	4.80	6.70	4.83	4.77
		4.63	4.60	4.66	6.70	4.82	4.98
G4 H8–A3′ H2″	3.26	2.53	2.77	2.77	5.75	2.75	2.96
		2.63	2.73	3.20	5.75	2.68	2.72

[a] Distances from both strands are reported as top and bottom entries. For SI and SII, both entries represent 3/4 stacked structures. For LI and LII the top and bottom entries represent the 2/3 and "triplex" structures, respectively.

comparison of the variation of both ζ and ^{31}P chemical shifts for the tridecamer sequence as well as the parent dodecamer duplex d(CGCGAATTCGCG) is shown in Figure 6 [the ^{31}P assignments of the dodecamer are from Ott and Eckstein (1984a), and the coupling constants are from Sklenar and Bax (1987)]. The correlation coefficient between ζ (or ϵ) and ^{31}P chemical shifts varies from 0.74 (at ambient temperature; Figure 7) to 0.89 (at 50 °C). We have also confirmed that $J_{H3'-P}$ coupling constants and ^{31}P chemical shifts are strongly correlated for several other oligonucleotide duplexes (Gorenstein et al., 1989, 1988).

Molecular Modeling of Tridecamer Duplex. Using the distances derived from the 2D NOESY spectrum of Figure 1, in conjunction with restrained molecular mechanics and dynamics calculations, we derived a structure for the duplex tridecamer. Distances were calculated by integrating the cross-peaks and utilizing the two-spin approximation of short mixing times (Wüthrich, 1986). Although volumes were measured for a complete set of mixing times (50, 100, 200, 300, 400, 500, and 700 ms), the volumes for a single 50-ms TPPI NOESY were used to calculate the intra- and interproton distances from the observed NOE cross-peaks (Nikonowicz et al., unpublished results). (Selected measured distances from adjacent base pairs to the extrahelical adenosine are provided in Table IV.) The distances derived from a 150-ms NOESY spectrum (data not shown) are similar to those of the 50-ms spectrum, demonstrating that the 50-ms NOEs were largely primary NOEs only partially affected by spin diffusion. [A complete relaxation matrix calculation analysis of the NOESY data is in progress, and a comparison of the structures derived from both the two-spin approximation at various mixing times and a complete relaxation matrix (Keepers & James, 1984) approach will be described in a forthcoming paper by Nikonowicz et al.]

Two alternative models for the tridecamer were considered, which we define as the stacked or looped-out conformation. Using the molecular modeling program MIDAS (Ferrin & Langridge, 1980), we model built a standard B-DNA structure for the parent dodecamer duplex, d(CGCGAATTCGCG)$_2$. The linkage between C3 and G4 was broken and an extrahelical adenosine inserted either in a loop-out orientation or with the adenosine base stacked into the double helix. After editing the file and using MIDAS to create a symmetrically related extrahelical adenosine on the opposite strand, the two different initial models (Figure 8A) were energy minimized by using the molecular mechanics/dynamics program AMBER (Weiner & Kollman, 1981). The NOESY-derived distances were then used in a distance-constrained energy minimization (AMBER) of the tridecamer duplex structures. Instead of a simple harmonic potential error function to restrain the NMR-derived distances, we have modified AMBER so as to provide a flat well harmonic function which better reflects the intrinsic accuracy of these NOESY distance restraints (Clore et al., 1985; Gorenstein et al., 1989; Powers et al., submitted for publication). The left and right force constants in the flat well harmonic potential for the NOESY distance constraint term were set to 10 kcal/(mol·Å^2), and the permitted errors were ±15% in the NOESY distances.

Table V: Calculated Energies for the Tridecamer during Restrained Molecular Dynamics Structural Refinement Starting from either the Stacked (S) or Looped-Out (L) Models

structure	E_{tot}[a] (kcal/mol)	E_{tot} (no constraints)[b] (kcal/mol)	$E_{constraint}$[c] (kcal/mol)
S(3/4–3/4)		−650	
S(3/4–3/4)I	−882	−950	33
S(3/4–3/4)II	−851	−931	32
L(loop/loop)model		−702	
L(2/3–triplex)I	−879	−952	40
L(2/3–triplex)II	−901	−971	37

[a] Total energy including NOESY distance constraint term. The structures during the last 3 ps of a 5-ps restrained molecular dynamics simulation (300 K) were averaged and then subjected to distance-constrained molecular mechanics (MM) energy minimization, until the root mean square gradient was less than 0.1. The constraint energies from the latter calculation are reported in the last column. Roman numerals in the structure identifier represent successive 5-ps MD/MM calculations. [b] Total energy not including NOESY distance constraints. The structures during the last 3 ps of a 5-ps restrained molecular dynamics simulation (300 K) were averaged and then subjected to MM energy minimization without distance constraints. [c] NOESY distance constraint energies. See footnote *a* for basis of calculation.

Structure Refinement of the 13-mer. Starting from the two initial model built structures for the extrahelical adenosine tridecamer [base loop-out (L) and stacked (S) structures], the AMBER program was used to calculate an energy-minimized structure (SI and LI); see Table V. We have used 206 50-ms NOESY constraints (103 per strand, calculated with the two-spin approximatation method, where the C H5/H6 volumes were used as internal rulers) in the flat well modified restrained molecular dynamics (MD) calculation. An additional 12 imino hydrogen bond constraints were incorporated into the distance constraints to reflect observed imino proton signals in a 90% H_2O/10% D_2O spectrum (Nikonowicz, unpublished results). A total of 15 ps of 300K MD was carried out on each minimized initial structure in 5-ps steps. The left

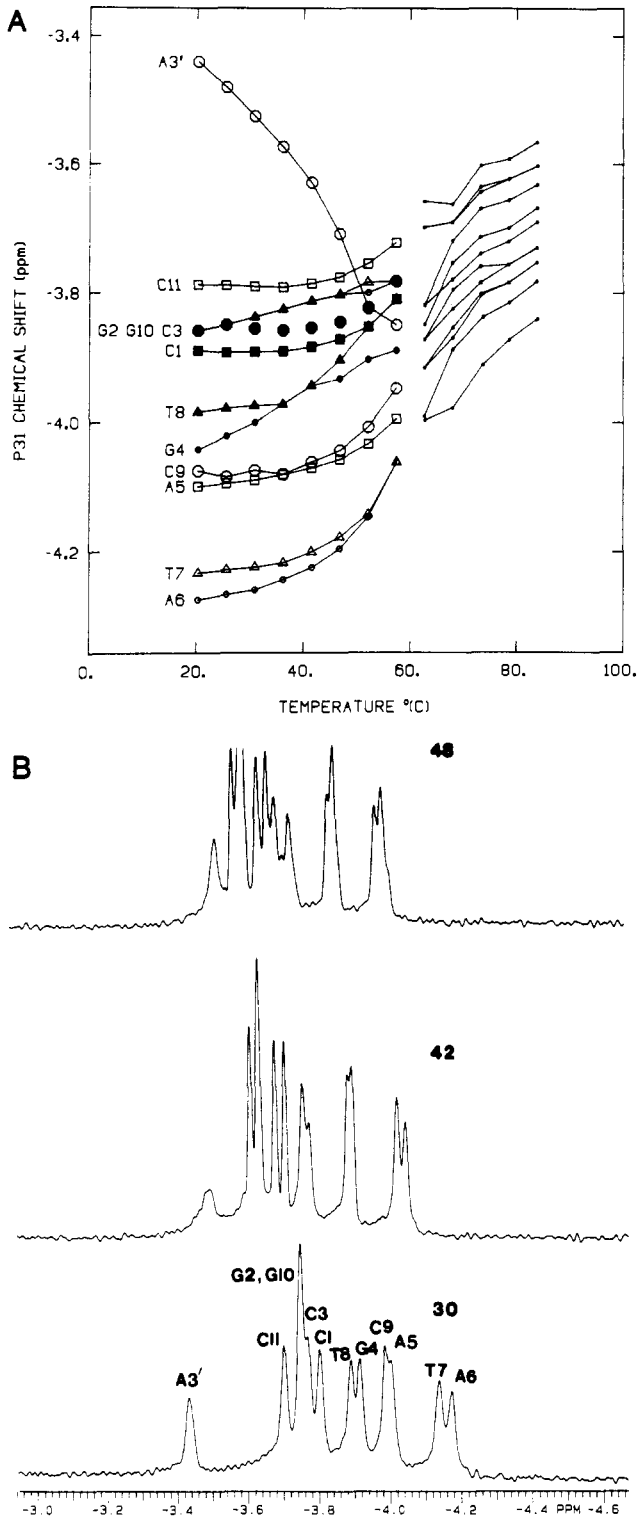

FIGURE 4: (A) Temperature dependence of ^{31}P chemical shifts of duplex tridecamer. Because of overlap of the signals in the melting region (ca. 55–60 °C) it is not possible to identify the individual peaks at higher temperatures, and it has been assumed that the relative order of the peaks remains the same. (B) ^{31}P spectrum of the tridecamer at indicated temperatures. ^{31}P signals are labeled in the 30 °C spectrum.

and right force constants in the flat well harmonic potential for the NOESY distance constraint term were set to 10 kcal/(mol·Å^2), and the error limits were ±15% in the NOESY distances. After each 5 ps of restrained molecular dynamics, structures from the last 3 ps of a simulation were averaged and then subjected to distance-constrained energy minimization. The energy-minimized, restrained MD averaged structures starting from both the looped-out and stacked models are shown in Figure 8B (structures from the first 5-ps cycle). One of the looped-out extrahelical adenosines has moved in between base pairs G2·C11 and C3·G10 on one half of the nonsymmetrical LI duplex (shown in detail in Figure 8C and identified in Table V as the 2/3 stacked structure), and the other extrahelical adenosine on the other half of the duplex has formed a "triplex" base triplet H-bonding structure with C3 and G10 (shown in Figure 8C and identified in Table V as the triplex structure). It is intriguing that this "triplex" does not grossly disobey the constraints for the 50-ms NOESY distances (Table IV) and is energetically allowed (Table V), suggesting that the time-averaged or ensemble-averaged structure for the 13-mer might include a population of such base triplets. Although tRNA and triple-strand DNA is known to contain this type of H-bonding triplets (Saenger, 1984), A-C-G triplets of the type shown in our structure are not known. It is also rather remarkable that the initial restrained MD calculations would produce an extrahelical insert between base pairs G2·C11 and C3·G10. There appears to be sufficient flexibility in the sugar phosphate backbone for the A3′ to coil back upon itself to produce this highly unusual structure. Note as shown in Table V that the L(2/3–triplex)II structure (initial loop model following three 5-ps restrained MD cycles) has the lowest unconstrained energy (−971 kcal/mol) of any of the structures. However, the constraint energy derived from NOESY distance violations for the L-(2/3–triplex)II structure is slightly higher (37 kcal/mol vs 33 kcal/mol) than that for the S(3/4–3/4)I structure (initial stacked model following two 5-ps restrained MD cycles). Thus surprisingly, both the triplex and the 2/3 stacked structures are energetically quite good and satisfy the 50-ms two-spin distance constraints. Indeed, at this point one is unable to confidently distinguish between the "triplex," the G2/C3 stacked, or the C3/G4 stacked structures solely on the basis of 50-ms derived distances and restrained molecular dynamics calculations.

It is premature to speculate on the relative merits or validity of any of these refined structures based upon the 50-ms restrained molecular dynamics calculations. It is clear, however, that the 50-ms NOESY distances can rule out the initial loop out structure (Table IV). We are presently working toward a refined structure of the tridecamer using a complete relaxation matrix analysis of the 150-ms NOESY data as well as longer molecular dynamics simulations. (Details of the NOESY distance-constrained structures will be described in a forthcoming paper by Nikonowicz et al.)

Discussion

Spectral Assignments. We have studied the tridecamer sequence d(CGCAGAATTCGCG) with particular emphasis on the two-dimensional proton and proton/phosphorus spectra. We have sequentially assigned nearly all nonexchangeable protons of the tridecamer, with the exception of some of the 5′ and 5″ sugar ring protons (Table I), via the two-dimensional NOESY technique (Broido et al., 1984; Feigon et al., 1983a,b; Hare et al., 1983; Scheek et al., 1984; Wemmer et al., 1984). Sugar proton to base proton connectivities are consistent with that expected for a right-handed helix B-type DNA (Weiss et al., 1984).

The base protons have produced NOESY cross-peaks to their own H1′, H2′, H2″, and H3′ protons as well as to those protons of the 5′-neighboring nucleotide. The 4′ protons were assigned via their spatial connectivity with the previously assigned 1′ and 3′ protons. These assignments as well as some 5′/5″ assignments were confirmed by using the 50-ms mixing

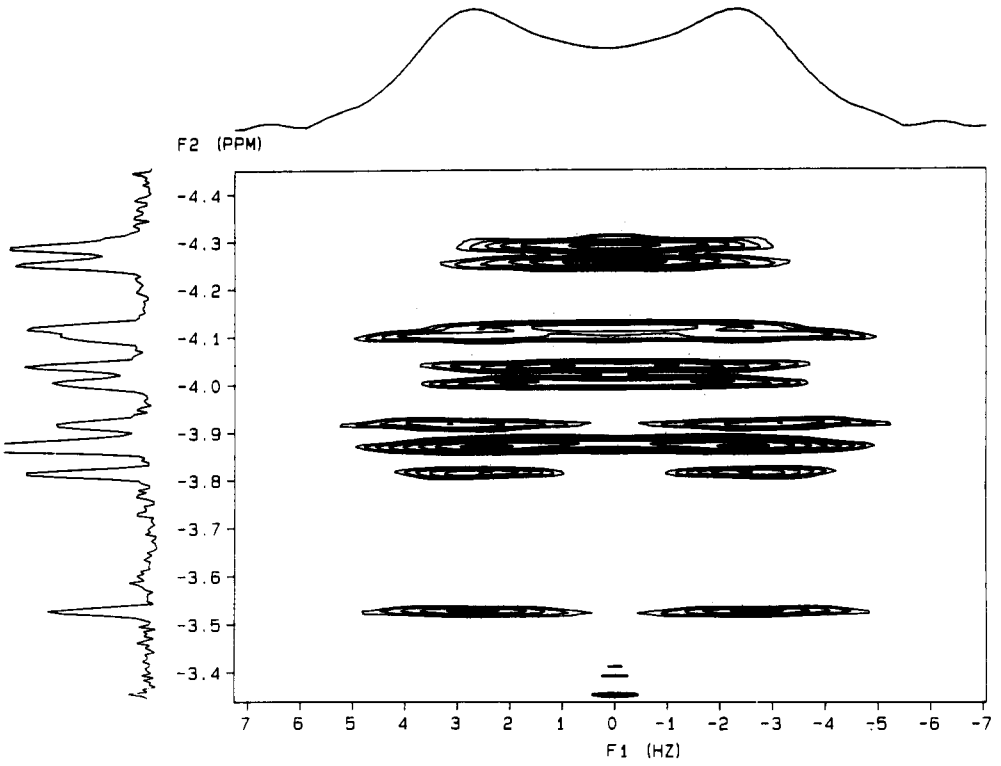

FIGURE 5: 2D *J*-resolved ^{31}P/^{1}H spectrum of the tridecamer at 30 °C. The 1D decoupled ^{31}P NMR spectrum is also shown along one axis, and the H3′ coupled doublets are shown along the second dimension.

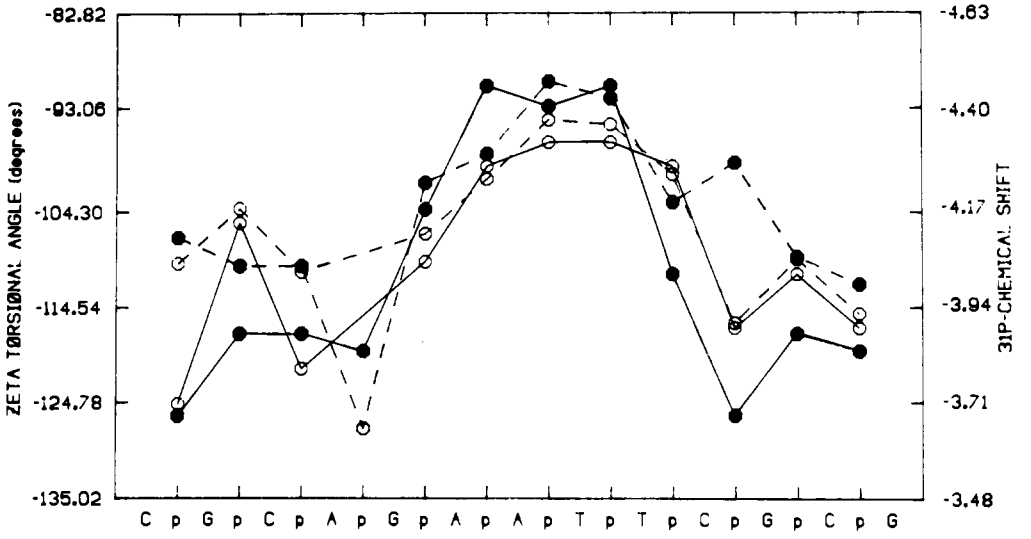

FIGURE 6: Comparison of ^{31}P chemical shifts (solid curves) and P–O3′ ester torsional angle ζ (dashed curves) for the tridecamer vs phosphate (●) and the 12-mer d(CGCGAATTCGCG) (○) vs position along the 5′–3′ strand for duplex tridecamer. ζ torsional angle calculated from the $J_{H3'-P}$ coupling constants derived from Figure 5, the calculated ϵ torsional angle from the Karplus relationship, and the correlation between ϵ and ζ ($\zeta = -317 + 1.23\epsilon$) for both sequences are also shown (solid curves).

time NOESY where only the 3′(*i*)–4′(*i*) interactions were seen (spectrum not shown).

In general, the chemical shifts of the base protons are consistent with those of Patel et al. (1982) which were based upon 1D NOE spectra and chemical shift perturbations relative to related oligonucleotide sequences. The extrahelical adenosine H8 and H2 protons were identified by comparison with the parent dodecamer sequence. Our assignment of the A3′ H8 proton is in agreement with that of Patel's [shown in Figure 4 of Patel et al. (1982); however, it is inconsistent with that described in Figure 3 of that paper].

Stacking of Extrahelical Adenosine. Evidence of the stacked nature of the extrahelical adenosine was first presented on the basis of an observed downfield shift of the H8 and H2 protons of A3′ (numbering is based upon the Dickerson dodecamer with the extrahelical adenosine labeled as residue 3′) upon melting of the duplex DNA into the single-strand form [T_m = 57.5 and 55 °C for H8 and H2 (Patel et al., 1982)]. Several pieces of information gathered here support this view.

Although the intensity of the A3′ H1′–G4 H8 cross-peak is weak, the H8/H6–H1′ sequential connectivities are observed for all base steps in the 150-ms mixing time NOESY up until the C9 H1′–G10 H8 base step. At this point a cross-peak is found only in the 250- and 400-ms NOESY. Although this cross-peak may arise from spin diffusion (see Results), this suggests an increased spatial separation of the C9 H1′ and G10 H8 protons, possibly due to a wedging effect of the facing A3′ residue. While not all of the base to base connectivities can

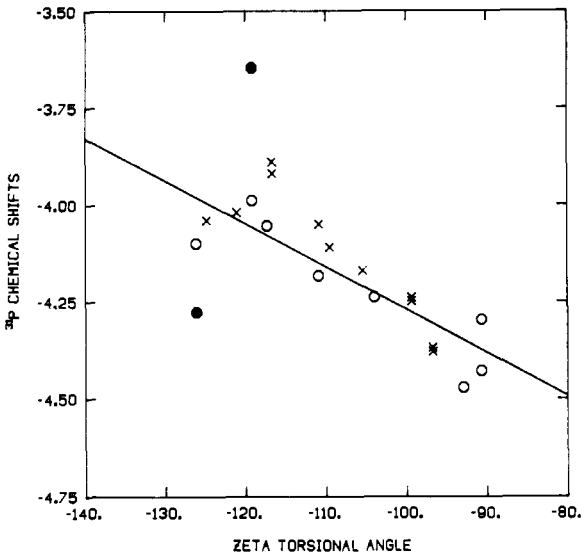

FIGURE 7: Plot of ^{31}P chemical shift vs P–O3′ ester torsional angle ζ for the tridecamer (O) and the parent Dickerson dodecamer d-(CGCGAATTCGCG)$_2$ (×). The unusual ^{31}P signals (A3′ and C9) are identified by the filled circles.

be drawn in the 150-ms NOESY spectrum, the C3 H6–A3′ H8 cross-peak is present.

The H2 proton of A3′ exhibits through-space interactions as well, the strongest being with that of its own H1′. However, it does not appear that the base is oriented in a syn conformation since the intensity of the H8–H1′ intraresidual cross-peak, although strong relative to most of the other 12 base–H1′ interactions, is weak relative to the H5–H6 interactions of the four cytidines. In addition to the intraresidual cross-peak, the A3′ H2 proton also appears to exhibit an interresidual interaction with the C3 H2′ proton at 250 ms.

Finally, the G4 H1′ resonance is shifted ~0.21 ppm upfield relative to the G4 H1′ resonance of the parent dodecamer sequence (Hare et al., 1983). This indicates a movement of the H1′ proton into the shielding portion of the ring-current cone of either G4 or A5. "Wedging" of the extrahelical adenosine into the stacked duplex could produce such an effect, while looping out of the adenosine would be expected to produce the opposite effect. The NOESY distance restrained molecular dynamics calculations also confirm that the stacked-in structure or at the very least nonlooped structures (Figure 8B) are preferred. These results in solution are thus directly opposite to those of the recently published crystal data (Joshua-Tor et al., 1988) in which it was shown that the extrahelical adenosine loops out.

^{31}P Assignments. We have introduced ^{17}O labels into the phosphoryl groups by replacing the I_2/H_2O in the oxidation step of the phosphite by $I_2/H_2{}^{17}O$ (Gorenstein et al., 1984; Ott & Eckstein, 1985a; Schroeder et al., 1987). The quadrupolar ^{17}O nucleus (generally 40%–50% enriched) broadens the ^{31}P signal of the directly attached phosphorus so that only the high-resolution signal of the remaining 50%–60% nonquadrupolar broadened phosphorus is observed (Petersheim et al., 1984; Tsai, 1984; Gorenstein et al., 1984). In this way, two of the ^{31}P resonances were identified (Figure 2). We note here that the downfield-shifted resonance is unambiguously that of the phosphorus atom attached to the 3′ end of the A3′ nucleoside. The remaining ^{31}P peaks were assigned via a 2D pure absorption phase constant-time (PAC) heteronuclear correlation NMR spectrum (Fu et al., 1988). Since the 3′ and 4′ resonances of the tridecamer had previously been assigned via the COSY and NOESY experiments, it was possible to assign the ^{31}P peaks through their connectivities via scalar coupling to the 3′ and 4′ protons (Lai et al., 1984; Pardi et al., 1983).

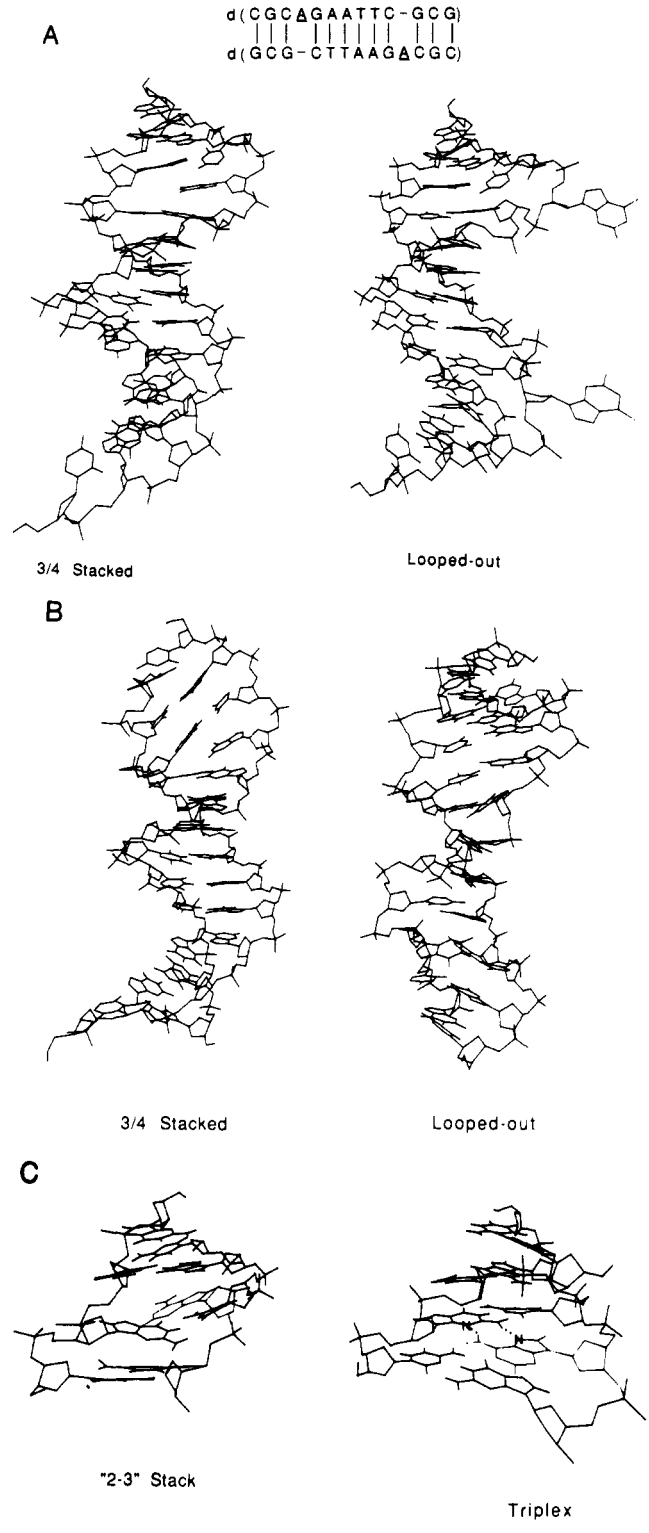

FIGURE 8: NOESY distance-restrained molecular dynamics structures of the extrahelical tridecamer d(CGCGCATAATCGCG)$_2$ starting from the base loop-out or 3/4 base stacked-in models (A), and after the first 5-ps iteration of NOESY distance-restrained molecular dynamics refinement using two-spin approximation 50-ms NOESY distances (B). Expanded views of both ends of the restrained MD loop-out model structure showing the 2/3 stacked half and the triplet half (C).

It has been suggested that the phosphorus 3′ to the C9 nucleoside, that opposite the extrahelical adenosine, was the "anomalous" downfield peak at −3.647 ppm (Patel et al.,

1982); however, this is not the case.

Previously (Gorenstein et al., 1986), we have shown through a sulfur labeling technique that the phosphorus on the strand opposite the extra dA residue was not responsible for the downfield ^{31}P signal at −3.647 ppm. In the sulfur labeling procedure, a sulfur atom replaces one of the phosphoryl oxygens of a single ^{31}P atom during the synthesis of the oligonucleotide. The thiophosphoryl ^{31}P resonance is shifted 50–55 ppm downfield, thus completely removing one resonance from the main cluster. Although introduction of a thiophosphoryl group can produce small ^{31}P chemical shift perturbations at neighboring phosphates (Eckstein, 1984), the downfield shifted ^{31}P peak at −3.65 ppm was not disturbed by this procedure, indicating that Patel et al.'s (1982) original suggestion for its identity was incorrect.

In the extrahelical adenosine sequence d-(CGCGAAATTT*A*Cp*GCG) (Roy et al., 1987), it was determined through a $^{31}P/^1H$ reverse detected HETCOR experiment (Sklenar et al., 1986) that a similarly downfield-shifted resonance was not the phosphorus atom on the strand opposite the the extra dA residue or indeed even 3′ to the extrahelical adenosine at position 11, but rather was the phosphorus 3′ to the 3′-flanking nucleotide (i.e., 3′ to C12 in the sequence above; the downfield phosphate is indicated by p*). In this case, the extrahelical adenosine was concluded to be stacked also. It is quite interesting that, in both of these extrahelical adenosine sequences, the downfield-shifted signals arise from phosphates at different sites from the perturbing base [the downfield signal in our sequence, d-(CGC*A*p*GAATTCGCG), is the labeled phosphate p* which is 3′ to the extrahelical adenosine].

Additional peaks are present in the PAC spectrum (Figure 3) but do not appear in the one-dimensional ^{31}P spectrum. It is not known at this time whether they arise from small quantities of impurities or if they represent a minor conformation(s) of the tridecamer phosphate backbone whose presence is "amplified" as a result of weaker $^1H/^1H$ parisitic interactions, larger $^{31}P/^1H$ coupling constants, or longer transverse relaxation times which will produce more intense PAC cross-peaks (Jones et al., 1988). Note that a minor component hairpin loop should be expected to have longer transverse ^{31}P relaxation times (at $^1/_2$ the molecular weight of the duplex, the hairpin loop will have narrower line widths at half-height), which should selectively enhance the intensity of the PAC cross-peaks. It is unlikely that the former explanation is the case since both reverse-phase and ion-exchange HPLC shows the sample to be pure. However, the NOESY spectral data, even at 400-ms mixing times, do not show any significant unassigned cross-peaks. One piece of data supporting the latter explanation, however, does exist in the methyl region of the one-dimensional proton spectrum. At 22 °C, the T7 methyl resonance exhibits a shoulder whereas at <15 °C the shoulder no longer is present. In light of previous reports (Roy et al., 1986, 1987), it is possible this could be indicative of a second conformation such as the hairpin loop or two different conformations for the extrahelical adenosine (i.e., stacked into the duplex vs looped out).

Positional and Site- and Sequence-Specific Variation of ^{31}P Chemical Shifts and Backbone Torsional Angles. As mentioned above, one factor that will affect ^{31}P chemical shifts is the degree of conformational constraint imposed by the duplex geometry (Connolly & Eckstein, 1984; Gorenstein, 1981; Gorenstein et al., 1988; Ott & Eckstein, 1985a). Note that the ^{31}P chemical shift of phosphates 3–9 (Figure 6) in both the parent dodecamer and the tridecamer moves upfield the more interior the phosphate. Base pairs closer to the ends of the duplex are less constrained to the stacked, base-paired geometry. This "fraying" at the ends imparts greater conformational flexibility to the deoxyribose phosphate backbone, and thus phosphates at the ends of the duplex will tend to adopt more of a mixture of g^-,g^- and t,g^- conformations. Interior phosphates are more constrained to the polymer P–O g^-,g^- conformation.

This "positional" ^{31}P chemical shift effect is apparently superimposed on site- and sequence-specific effects (Connolly & Eckstein, 1984; Gorenstein, 1981; Gorenstein et al., 1988; Ott & Eckstein, 1985a; Schroeder et al., 1987, 1986; Patel, 1982). The minor groove clash steps (pyrimidine–purine base steps) have been generally associated with a relatively downfield ^{31}P chemical shift. However, as can be seen from the ^{31}P chemical shift pattern shown in Figure 6, the minor groove clash steps at phosphate positions 3 and 7 appear to follow the positional effect. The relative upfield shift of phosphates 1, 2, and 9 may also attributed to the sequence-specific effect, although they more likely originate from the site-specific perturbation in the backbone geometry resulting from the extrahelical adenosine.

Two of the most important parameters controlling ^{31}P chemical shifts in phosphate esters are the P–O torsional angles [in nucleic acids the P–O5′ (α) and P–O3′ (ζ) torsional angles (Gorenstein, 1987; Gorenstein & Kar, 1975) and the C–O5′ (β) and C–O3′ (ϵ) torsional angles] (Giessner-Prettre et al., 1984; Ribas-Prado et al., 1979), although the P–O torsional angle may be more important. Using the selective 2D *J*-resolved correlation experiment, we can measure the three-bond H3′–C3′–O–P coupling constant (listed in Table III) for each of the phosphates in the tridecamer (Gorenstein et al., 1988; Sklenar & Bax, 1987).

The data of Figures 6 and 7 support our hypothesis that the position-specific, site-specific, and sequence-specific effects on ^{31}P chemical shifts are largely attributable to variations in the helical parameters and the backbone torsional angles (at least for ζ and ϵ). Unwinding or winding the double helix changes the backbone δ, α, and ζ torsional angles, and these backbone changes presumably are responsible for the variations in the ^{31}P chemical shifts of oligonucleotides. It is significant that the ^{31}P chemical shifts for phosphates removed from the extrahelical adenosine position show a significant correlation with J(P–H3′) (removing the two data points for the A3′ and C9 phosphates gives a correlation coefficient of −0.91 for the tridecamer data plotted in Figure 7; interestingly, the correlation coefficient is also a very good −0.90 for the parent dodecamer data shown in the plot as well). However, the A3′ and C9 phosphates fall off this correlation, presumably arising from distortion of the phosphodiester backbone from the extrahelical adenosine. The large downfield shift for the A3′ phosphate is similar to the effect of binding an intercalating drug to the duplex which unwinds the duplex (Gorenstein, 1981; Patel, 1979). This unwinding and lengthening of the deoxyribose phosphate backbone allows the drug or extrahelical adenosine to insert with a base to base separation of 6.7 Å to accommodate the additional heterocycle drug or base (although in the tridecamer the adenosine acts as a wedge to only partially open the stacked base pairs). This opening is apparently only partially accomplished by the phosphate switching from the normal B-DNA conformation ("B_I") in which both P–O torsional angles ζ and α are in the g^- conformation to a "B_{II}"-like conformation ($\zeta = t$, $\alpha = g^-$). The ^{31}P signal of a phosphate in a t,g^- B_{II} conformation is predicted to be ~1.5 ppm downfield from the g^-,g^- phosphate in the B_I

conformation (Gorenstein, 1981), and we expect to observe a large downfield shift of the ^{31}P signal upon stacking of the adenosine into the duplex. It is interesting, however, that the signal that is perturbed the most downfield is the A3′pG4 phosphate and that this phosphate does not fall on the correlation of ^{31}P chemical shift and ζ (Figure 7). In addition, the C9pG10 phosphate opposite the extrahelical base is shifted considerably *upfield* relative to the ^{31}P signal of the C9pG10 phosphate in the parent dodecamer and also does not fall on the correlation of Figure 7. The ζ torsional angle for both the A3′ and C9 phosphates are more B_{II}-like (i.e., ζ is more translike than the other phosphates). In fact, the C9 phosphate is the most B_{II}-like of any of the phosphates, consistent with the expected trans-type conformation expected of a base-intercalated base step. Yet the ^{31}P chemical shift of this phosphate is not downfield as expected on the basis of the torsional angle effect alone. This suggests that the ^{31}P chemical shift is perturbed by some other factor(s) at this site other than ζ and ϵ torsional angle changes alone. This could reflect changes in torsional angles α (P–O5′) and β (C–O5′), which are quite variable in the A-DNA structures although they are quite constant in normal B-DNA structures (Sanger, 1984). Alternatively, as has previously been shown, ^{31}P chemical shifts are also very sensitive to bond angle distortions as well (Gorenstein, 1975, 1984). It is quite reasonable to assume that the extrahelical adenosine not only unwinds the C9pG10 base step (as evidenced by the change in the ζ torsional angle) but also introduces some bond angle distortion to further widen the base step to accommodate the extrahelical adenosine in a stacked-in geometry. Widening of the ester O–P–O bond angle indeed is expected to produce an *upfield* shift (Gorenstein, 1975, 1984), and it is presumably this upfield bond angle effect that counterbalances the expected downfield shift of this phosphate in the translike conformation.

^{31}P Melting Curves. As previously reported (Patel et al., 1982), the ^{31}P peaks of the main cluster move downfield (Figure 4A) as the sample is heated. In a right-handed B-type DNA duplex, the PO ester conformation is g^-,g^- (gauche, −60°). As is expected, heating of the duplex to above the T_m produces increased conformational flexibility along the phosphate backbone. This is in agreement with previous observations (Gorenstein et al., 1976, 1984; Gorenstein, 1984) that the nongauche conformation of the PO ester resonates downfield of the g^-,g^- form.

The temperature dependence of the most downfield-shifted peak suggests that the distorted duplex can exist in at least two different conformations. This anomalous peak begins to exhibit line broadening at 38 °C but again begins to sharpen by 46 °C, characteristic of chemical exchange. At 50 °C, the anomolous peak appears to move upfield into the main cluster of resonances. The chemical exchange of this resonance occurs between 40 and 48 °C, well below the T_m of the tridecamer. It is possible that the tridecamer can exchange between the base insert geometry (low-temperature form) and the base loop-out geometry (high-temperature form). We are currently investigating the temperature dependence to the NOESY cross-peaks to investigate this possibility.

ACKNOWLEDGMENTS

We greatly appreciate the contributions of Dr. Robert Santini and James Metz.

REFERENCES

Bax, A., & Morris, G. A. (1981) *J. Magn. Reson. 42*, 501–505.

Bax, A., & Freeman, R. (1982) *J. Am. Chem. Soc. 104*, 1099–1100.

Broido, M. A., Zon, G., & James, T. L. (1984) *Biochem. Biophys. Res. Commun. 119*, 663–670.

Calladine, C. R. (1982) *J. Mol. Biol. 161*, 343–352.

Cheng, D. M., Kan, L.-S., Miller, P. S., Leutzinger, E. E., & Ts'o, P. O. P. (1982) *Biopolymers 21*, 697.

Cheng, D. M., Kan, L., & Ts'o, P. O. P. (1987) *Phosphorus NMR in Biology* (Burt, C. T., Ed.) pp 135–147, CRC Press, Boca Raton, FL.

Clore, G. M., Gronenborn, A. M., Brunger, A. T., & Karplus, M. (1985) *J. Mol. Biol. 186*, 435–455.

Connolly, B. A., & Eckstein, F. (1984) *Biochemistry 23*, 5523–5527.

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

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761–786.

Feigon, J., Denny, W. A., Leupin, W., & Kearns, D. R. (1983a) *Biochemistry 22*, 5930–5942.

Feigon, J., Leupin, W., Denny, W. A., & Kearns, D. R. (1983b) *Biochemistry 22*, 5943–5951.

Ferrin, T. E., & Langridge (1980) *Comput. Graphics 13*, 320.

Fink, T., & Crothers, D. M. (1972) *J. Mol. Biol. 66*, 1–12.

Fresco, J. R., & Alberts, B. M. (1960) *Proc. Natl. Acad. Sci. U.S.A. 46*, 311–321.

Fu, J. M., Schroeder, S. A., Jones, C. R., Santini, R., & Gorenstein, D. G. (1988) *J. Magn. Reson. 77*, 577–582.

Gait, M. J. (1984) *Oligonucleotide Synthesis: a Practical Approach*, IRL Press, Oxford.

Giessner-Prettre, C., Pullman, B., Ribas-Prado, F., Cheng, D. M., Iuorno, V., & Ts'o, P. O. P. (1984) *Biopolymers 23*, 377.

Gorenstein, D. G. (1975) *J. Am. Chem. Soc. 97*, 898.

Gorenstein, D. G. (1978) *Jerusalem Symposium, NMR in Molecular Biology* Pullman, B., Ed.) pp 1–15, D. Reidel Publishing Co., Dordrecht.

Gorenstein, D. G. (1981) *Annu. Rev. Biophys. Bioeng. 10*, 355.

Gorenstein, D. G. (1984) *Phosphorus-31 NMR*: Principles *and Applications* (Gorenstein, D. G., Ed.) Academic Press, New York.

Gorenstein, D. G. (1987) *Chem. Rev. 87*, 1047–1077.

Gorenstein, D. G., & Kar, D. (1975) *Biochem. Biophys. Res. Commun. 65*, 1073–1080.

Gorenstein, D. G., & Findlay, J. B. (1976) *Biochem. Biophys. Res. Commun. 72*, 640.

Gorenstein, D. G., & Goldfield, E. M. (1984) *Phosphorus-31 NMR: Principles and Applications*, (Gorenstein, D. G., Ed.) Academic Press, New York.

Gorenstein, D. G., Findlay, J. B., Momii, R. K., Luxon, B. A., & Kar, D. (1976) *Biochemistry 15*, 3796–3803.

Gorenstein, D. G., Lai, K., & Shah, D. O. (1984) *Biochemistry 23*, 6717.

Gorenstein, D. G., Schroeder, S., Miyasaki, M., Fu, J., Jones, C., Roongta, V., & Abuaf, P., (1987a) *Proceedings of the 10th International Conference on Phosphorus Chemistry 30*, 567–570.

Gorenstein, D. G., Schroeder, S. A., Miyasaki, M., Fu, J. M., Roongta, V., Abuaf, P., Chang, A., & Yang, J. C. (1987b) *Biophosphates and Their Analogues—Synthesis, Structure, Metabolism and Activity* (Bruzik, K. S., & Stec, W. J., Eds.) pp 487–502, Elsevier Press, Amsterdam.

Gorenstein, D. G., Schroeder, S. A., Fu, J. M., Metz, J. T., Roongta, V. A., & Jones, C. R. (1988) *Biochemistry 27*, 7223–7237.

Gorenstein, D. G., Jones, C. R., Schroeder, S. A., Metz, J. T., Fu, J. M., Roongta, V. A., Powers, R., & Karslake, C. (1989) *Progress in Inorganic Biochemistry and Biophysics* (Gray, H., & Bertini, I., Eds.) Birkhäuser, Boston (in press).
Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., & Reid, B. (1983) *J. Mol. Biol. 171*, 319.
Hare, D., Shapiro, I., & Patel, D. J. (1986a) *Biochemistry 25*, 7445–7456.
Hare, D., Shapiro, L., & Patel, D. J. (1986b) *Biochemistry 25*, 7456–7464.
James, T. L. (1984) *P-31 NMR: Principles and Applications* (Gorenstein, D. G., Ed.) pp 349–400, Academic Press, Orlando, FL.
Jones, C. R., Schroeder, S. A., & Gorenstein, D. G. (1988) *J. Magn. Reson. 80*, 370–374.
Joshua-Tor, L., Rabinovich, D., Hope, H., Frolow, F., Appella, E., & Sussman, J. L. (1988) *Nature 334*, 82–84.
Kalnik, M. W., Norman, D. G., Zagorski, M. G., Swann, P. F., & Patel, D. J. (1989) *Biochemistry 28*, 294–303.
Keepers, J., & James, T. (1984) *J. Magn. Reson. 57*, 404.
Kessler, H., Griesinger, C., Zarbock, J., & Loosli, H. R. (1984) *J. Magn. Reson. 57*, 331–336.
Lai, K., Shah, D. O., Derose, E., & Gorenstein, D. G. (1984) *Biochem. Biophys. Res. Commun. 121*, 1021.
Lankhorst, P. P., Hassnoot, C. A. G., Erkelens, C., & Altona, C. (1984) *J. Biomol. Struct. Dyn. 1*, 1387–1405.
Lee, C. H., & Tinoco, I., Jr. (1978) *Nature (London) 274*, 609–610.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Miller, M., Harrison, R. W., Wlodawer, A., Appella, E., & Sussman, J. L. (1988) *Nature 334*, 85–86.
Morden, K. M., Chu, Y. G., Martin, F. H., & Tinoco, I., Jr. (1983) *Biochemistry 22*, 5557–5563.
Ott, J., & Eckstein, F. (1985a) *Biochemistry 24*, 253.
Ott, J., & Eckstein, F. (1985b) *Nucleic Acids Res. 13*, 6317–6330.
Pardi, A., Morden, K. M., Patel, F. J., & Tinoco, I., Jr. (1982) *Biochemistry 21*, 6567–6574.
Pardi, A., Walker, R., Rapoport, H., Wider, G., & Wüthrich, K. (1983) *J. Am. Chem. Soc. 105*, 1652.
Patel, D. J. (1974) *Biochemistry 13*, 2396–2402.
Patel, D. J. (1976) *Biopolymers 15*, 533–558.
Patel, D. J. (1979) *Acc. Chem. Res. 12*, 118–125.
Patel, D. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6424–6428.
Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Itakura, K., & Breslauer, K. J. (1982) *Biochemistry 21*, 445–451.
Petersheim, M., Mehdi, S., & Gerlt, J. A. (1984) *J. Am. Chem. Soc. 106*, 439–440.
Ribas-Prado, F., Giessner-Prettre, C., Pullman, B., & Daudey, J.-P. (1979) *J. Am. Chem. Soc. 101*, 1737.
Roy, S., Weinstein, S., Borah, B., Nickol, J., Apella, E., Sussman, J. L., Miller, M., & Shindo, H. (1986) *Biochemistry 25*, 7417–7423.
Roy, S., Sklenar, V., Appella, E., & Cohen, J. S. (1987) *Biopolymers 26*, 2041–2052.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Scheek, R. M., Boelens, R., Russo, N., Van Boom, J. H., & Kaptein, R. (1984) *Biochemistry 23*, 1371–1376.
Schroeder, S., Jones, C., Fu, J., & Gorenstein, D. G. (1986) *Bull. Magn. Reson. 8*, 137–146.
Schroeder, S., Fu, J., Jones, C., & Gorenstein, D. G. (1987) *Biochemistry 26*, 3812–3821.
Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Park Kim, J. J., & Rich, A. (1976) *J. Mol. Biol. 104*, 142–143.
Shah, D. O., Lai, K., & Gorenstein, D. G. (1984a) *Biochemistry 23*, 6717–6723.
Shah, D. O., Lai, K., & Gorenstein, D. G. (1984b) *J. Am. Chem. Soc. 106*, 4302.
Sklenář, V., & Bax, A. (1987) *J. Am. Chem. Soc. 109*, 7525–7526.
Sklenář, V., Miyashiro, H., Zon, G., Miles, H. T., & Bax, A. (1986) *FEBS Lett. 208*, 94–98.
Tsai, M.-D. (1984) *^{31}P NMR: Principles and Applications* (Gorenstein D., Ed.) Academic Press, New York.
Van De Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1–38.
Weiner, P. K., & Kollman, P. A. (1981) *J. Comput. Chem. 2*, 287–303.
Weiss, M. A., Patel, D. J., Sauer, R. T., & Karplus, M. (1984) *J. Am. Chem. Soc. 106*, 4269–4270.
Wemmer, D. E., Chou, S.-H., Hare, D. R., & Reid, B. R. (1984) *Biochemistry 23*, 2262–2268.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.