Determination of Nucleic Acid Backbone Conformation by ^{1}H NMR†

Seong-Gi Kim, Leh-Jame Lin,‡ and Brian R. Reid*
Departments of Chemistry and Biochemistry, University of Washington, Seattle, Washington 98195
Received September 3, 1991; Revised Manuscript Received January 10, 1992

ABSTRACT: In DNA or RNA duplexes, the six-bond C3′–O3′–P–O5′–C5′–C4′–C3′ backbone linkage connecting adjacent residues contains six torsion angles (ϵ, ζ, α, β, γ, δ) but only four protons. This seriously limits the ability to define the backbone conformation by NMR using purely ^{1}H–^{1}H distance geometry (DG) methods. The problem is further compounded by the inability to assign two of the four backbone protons, namely, the poorly resolved H5′ and H5″ protons, and invariably leads to DG structures with poorly defined backbone conformations. We have developed and tested a reliable method to constrain the β, γ, and ϵ (and indirectly α and ζ) backbone torsion angles by lower-bound NOE distances to *unassigned* H5′/H5″ resonances combined with either ^{1}H line widths or the conservative use of $\sum J$ measurements; the method relies only on ^{1}H 2-D NMR data, does not involve any structural assumptions, and leads to much improved backbone convergence among DG structures. The C4′–C5′ torsion angle γ is constrained by lower-bound NOE distances from H2′ and from H6/H8 to *any* H5′/H5″, as well as by $\sum J_{H4'}$ coupling measurements in the 3.9–4.4 ppm region; δ is constrained by H1′–H4′ NOE distances and by H3′–H4′ and H3′–H2″ J couplings in COSY data; ϵ is partially constrained by H3′ line width and/or further constrained by subtracting the minimum possible $\sum J_{H3'-H}$ from the observed $\sum J_{H3'}$ (COSY) to arrive at the maximum possible $J_{H3'-P}$, which is then converted to H3′–P distance bounds. The angle β is partially constrained via H5′–P and H5″–P distance bounds consistent with the maximum H5′–P and H5″–P J couplings derived from the observed H5′ and H5″ line widths, while α and ζ are indirectly constrained by lower distance bounds on the observed (n)H1′ to (n + 1)H5′/H5″ NOEs combined with the prior partial constraints on β, γ, δ, and ϵ. The combined effects of these additional constraints in determining distance geometry structures have been demonstrated using a 12-base duplex, [d(GCCGTTAACGGC)]$_2$. Coordinate RMSDs per atom between structures refined with these constraints from random-embedded DG structures, from ideal A-DNA, and from B-DNA starting structures were less than 0.4 Å for the central 8 base pairs indicating good convergence. All backbone angles for the central 8 base pairs are very well constrained with less than 10° variation in any of the 48 torsion angles.

The determination of the structure of short DNA duplexes in solution by high-resolution nuclear magnetic resonance (NMR) spectroscopy is currently under intensive investigation in several laboratories [for reviews, see Wüthrich (1986), Reid (1987), Hosur et al. (1988), van de Ven and Hilbers (1988), and Clore and Gronenborn (1989)]. The most common method of structure determination relies heavily on dipole–dipole coupling (the nuclear Overhauser effect) between protons that are separated by less than 5 Å. The nuclear Overhauser effect (NOE) is proportional to r^{-6}, and distance information can be obtained from analyses of cross-peak intensities in two-dimensional NOE experiments. These distances can be used to generate three-dimensional structures by a variety of model building and computational approaches such as distance geometry, energy minimization, and molecular dynamics (Hare & Reid, 1986; Hare et al., 1986a,b; Nilsson et al., 1986; Nilges et al., 1987a,b; Banks et al., 1989; Nerdal et al., 1988, 1989; Baleja et al., 1990; Gochin & James, 1990; Niconowicz & Gorenstein, 1990). In addition, back-calculation methods have been used to further refine the initially generated structure using either numerical integration (Banks et al., 1989; Nerdal et al., 1989) or relaxation matrix methods (Boelens et al., 1988, 1989; Borgias & James, 1990; Nikonowicz et al., 1990) to simulate the NOESY spectra.

The two major difficulties in determining the structure of DNA and RNA by NMR are (i) insufficient distance information to define the backbone conformation, i.e., α, β, γ, δ, ϵ, and ζ angles (P–O–C5′–C4′–C3′–O–P–) and (ii) the lack of long-range information, i.e., inadequate global structural information. The backbone conformation is important, not only in itself but also because of its influence on the overall helical twist of the duplex. First, the δ (C5′–C4′–C3′–O) angle is part of the sugar conformation which is relatively easily determined by NMR methods; once the sugar conformation is known, either from H–H distances or H–H J coupling, δ is fixed. Hosur et al. (1986) reported methods for determining the sugar pucker conformation using proton–proton J coupling values obtained from a 2-D homonuclear correlation experiment (COSY). Even without determining the actual H–H coupling constant values, the relative intensity of the corresponding COSY cross-peak qualitatively reflects the magnitude of the coupling constant. In the case of the C2′-endo sugar conformation (B-form DNA), the H2″–H3′ and H3′–H4′ cross-peaks are either absent or very weak due to very small coupling constants (less than 1.0 Hz). However, in the C3′-endo sugar conformation (A-form DNA) the H2″–H3′ and H3′–H4′ cross-peaks are very strong, due to J values of ≥8.0 Hz. Alternatively, the sugar conformation can be determined from NOESY distance data (Chary et al., 1989); in the case of the C2′-endo conformer, the H2″–H4′ distance is greater than 4.0 Å, producing a very weak NOESY cross-peak, while this cross-peak intensity is strong ($d \leq 2.8$ Å) in the case of the C3′-endo conformation. The β (P–O–C5′–C4′) and γ (O–C5′–C4′–C3′) angles can, at least in

† This work was supported by National Institutes of Health Grant GM32681.
* To whom correspondence should be addressed at the Department of Chemistry.
‡ Present address: Institute of Biomedical Sciences, Academia Sinica, Taiwan.

0006-2960/92/0431-3564$03.00/0

principle, be determined from the J couplings between phosphorus and H5′/H5″, and between H4′ and H5′/H5″, respectively. However, due to the small chemical shift differences between the H4′, H5′, and H5″ resonances, even assigning the H5′ and H5″ peaks is often not possible, making it difficult to restrict the β and γ angles from coupling constants. Altona (1982) reported that the γ angle could be determined from the sum of $J_{H4'-H5'}$ and $J_{H4'-H5''}$ (these two coupling constants total less than 4 Hz when $\gamma = 60°$). In practice, however, the sum of $J_{H4'-H5'}$ and $J_{H4'-H5''}$ cannot be determined in moderate-sized DNA and RNA molecules due to the additional H4′ coupling with H3′. The ϵ (C4′–C3′–O–P) angle can be related to the vicinal J coupling constant between H3′ and phosphorus, which itself can be determined by a ^{31}P–^{1}H correlation experiment. Sklenar and Bax (1987) and Gorenstein and co-workers (Powers et al., 1990; Roongta et al., 1990) observed that no H3′–phosphorus J coupling constants exceeded 6.0 Hz in three different DNA sequences. Even a specific $J_{H3'-P}$ of, say, 5.1 Hz is consistent with four possible ϵ conformations, namely, two B_I and two B_{II} conformers, both of which have been found in crystal structures (Dickerson & Drew, 1981; Prive et al., 1991). The α (O–P–O–C5′) and ζ (C3′–O–P–O) angles cannot be determined directly from either NOE or J coupling information. The chemical shift of phosphorus may be related to the backbone conformation (Roongta et al., 1990; Powers et al., 1990), but the relationship is not accurately known.

This paper reports a combination of straightforward and reliable methods that we have found to lead to quite narrow constraints for the backbone torsion angles by using sums of J coupling constants, line widths, and dipole–dipole interactions in simple proton NOESY and COSY spectra. The resulting allowed torsion angle ranges have been transformed to distance bounds for use in conjunction with distance geometry refinement methods.

As part of an ongoing project to study context effects on AT base pair structure, a series of DNA duplexes containing internal A_nT_n and T_nA_n sequences (where n is 1, 2, or 3) are currently being investigated and compared by NMR and distance geometry (DG) structural methods in this laboratory. The methods reported below for constraining the polynucleotide backbone were derived, refined, and tested in studies on the T_2A_2 sequence $[d(GCCGTTAACGGC)]_2$, which contains the *Hpa*I restriction site GTTAAC. However, the approach described is equally applicable to any duplex sequence with a moderately resolved proton NMR spectrum.

Materials and Methods

Sample Preparation. The DNA dodecamer containing the *Hpa*I restriction sequence, $[d(GCCGTTAACGGC)]_2$, was synthesized and purified using methods described previously (Kintanar et al., 1987). The DNA sample (28 mg) was dissolved in 0.4 mL of buffer containing 0.7 mM EDTA, 30 mM sodium phosphate, and 75 mM NaCl. The sample was repeatedly lyophilized to dryness and finally dissolved in 0.4 mL of 99.996 % D_2O to produce a final DNA concentration of ca. 9.6 mM.

NMR Spectroscopy. The NMR experiments were performed at 500 MHz either on a Bruker AM-500 spectrometer or on a home-built NMR spectrometer (Gladden and Drobny, manuscript in preparation). Four NOESY spectra with mixing times of 40, 80, 120, and 200 ms were collected using the phase-sensitive method (States et al., 1982) within a single 4-day time period, without removing sample from the spectrometer. In each NOESY spectrum, the mixing time was randomly varied over ±10% of the designated mixing time to eliminate zero-quantum coherence. For all NOESY experiments, 1024 complex points in $t2$ and 400 points in $t1$ were collected with a relaxation delay (RD) of 2.0 s and a spectral width (SW) of 4386 Hz. The exclusive COSY (E. COSY) spectrum (Griesinger et al., 1986, 1987) was recorded in the phase-sensitive mode with time-proportional phase incrementation (Drobny et al., 1979; Marion & Wüthrich, 1983). To increase sensitivity and resolution in $t2$, the E.COSY spectrum was acquired with 2048 complex points in $t2$ and with 48 scans per $t1$ experiment. The NMR data were transferred to an IRIS 4D computer and processed using the FTNMR and FELIX programs (Hare Research, Inc., Woodinville, WA). NOESY data sets were zero-filled to 2048 points in each dimension and apodized using 3 Hz of Gaussian broadening in $t2$ and a sine-squared 90° phase-shifted function in $t1$, then Fourier-transformed. E. COSY data were processed using 6 Hz of exponential narrowing and 6 Hz of Gaussian broadening in the $t2$ domain and a skewed-sinebell window function in $t1$. Spectral resolution was enhanced by zero-filling to 4096 points in $t2$ (1.07 Hz per point) and 2048 points in the $t1$ domain. In all spectra, $t1$ ridges were attenuated by multiplying the first $t1$ slice by 0.5 (Otting et al., 1985).

The resolved cross-peak volumes corresponding to each detectable proton–proton dipole–dipole interaction were obtained from the NOESY spectra (vide infra). Assuming a two-spin initial rate approximation, that is, no spin diffusion at short mixing times, the distance (r_{ij}) can be estimated relative to a fixed distance (r_{ref}) from the relative initial cross-relaxation rates R_{ij} and R_{ref} using the relationship $r_{ij} = r_{ref}(R_{ref}/R_{ij})^{1/6}$. The cytosine H5–H6 distance of 2.46 Å was used as a reference (r_{ref}). From the E.COSY spectrum, J coupling constants and the sum of J coupling constants ($\sum J$) for each proton were determined.

Distance Geometry Calculations. The concepts and details of the distance geometry algorithm DSPACE have been described previously (Hare et al., 1986a,b; Hare & Reid, 1986; Nerdal et al., 1988, 1989; Banks et al., 1989). All distance geometry calculations were performed on an IRIS 4D workstation. To initiate the distance geometry calculation, upper and lower bounds were assembled for all distances implicit in the primary structure and stored in a bounds matrix, along with the experimentally determined distance bounds. The entire bounds matrix containing all the available distance information was subjected to several smoothing procedures including the triangle and inverse triangle inequalities (Crippen, 1981; Havel & Wüthrich, 1985). Several different distance matrices were then generated by choosing random distances between the lower and upper bounds for each matrix element, to produce several trial matrices, each of which was converted to a metric matrix and embedded in 3-D space (Crippen, 1981; Havel et al., 1983). The initially embedded structure was refined by a combination of conjugate gradient refinement and a simulated annealing algorithm (Metropolis et al., 1953; Kirkpatrick et al., 1983), which uses the deviation from the distance bounds as its primary penalty parameter (Nerdal et al., 1988). Canonical B-form and A-form DNA structures (Arnott & Hukins, 1972a,b) were also used as trial starting structures and were randomized prior to refinement by a combination of simulated annealing and conjugate gradient refinement.

Calculation of Helical Parameters. Backbone angles for all structures were calculated with the program NEWHELIX90 (kindly provided by R. E. Dickerson).

Results and Discussion

Sugar Geometries. The sugar torsion angles, ν, can be

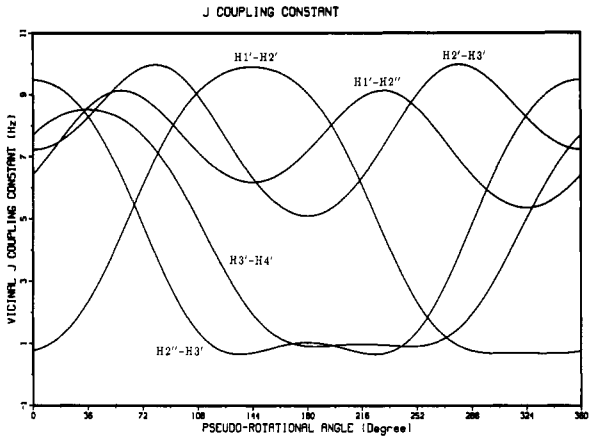

FIGURE 1: The 3-bond vicinal J coupling constants in deoxyribose as a function of the pseudorotational phase angle. The modified Karplus equation (Haasnoot et al., 1980) was used to generate J as a function of P.

determined from the pseudorotational phase angle P and the puckering amplitude T_m as $\nu_j = T_m \cos [P + 144(j - 2)]$ where j is 0–4 (Altona & Sundaralingam, 1972). The relationships between the dihedral angles ϕ and ν are as follows: $\phi_{H1'-H2'} = \nu_1 + 120$, $\phi_{H1'-H2''} = \nu_1$, $\phi_{H2'-H3'} = \nu_2$, $\phi_{H2''-H3'} = \nu_2 + 120$, and $\phi_{H3'-H4'} = 120 - \nu_3$. In the following calculations, the value of T_m was taken to be 38.5° (Westhof & Sundaralingam, 1980). From these H–C–C–H dihedral angles, and 3-bond coupling constant, ^3J, can be estimated using either a simple Karplus equation (Davies, 1978; Hosur et al., 1986) or a more complicated Karplus equation which takes into account the electronegativity and orientation of the substituents (Haasnoot et al., 1980; van de Ven & Hilbers, 1988) namely

$$^3J_{HH} = P_1 \cos^2 \phi + P_2 \cos \phi + P_3 + \sum_i \Delta\chi_i\{P_4 + P_5 \cos^2 (\xi\phi + P_6|\Delta\chi_i|)\},$$

where $P_1 = 13.86$, $P_2 = -0.81$, $P_3 = 0.0$, $P_4 = 0.56$, $P_5 = -2.32$, and $P_6 = 17.9$; χ_i is the electronegativity difference between substituent i and a proton (e.g., 1.3 for oxygen, 0.4 for carbon, and 0.9 for nitrogen); and ξ is a relative orientation factor, i.e., ±1 (Haasnoot et al., 1980). Theoretical $^3J_{HH}$ values using this relationship are shown in Figure 1; these are reasonably similar to those obtained using the simple Karplus relationship $J = 10.2 \cos^2 \phi - 0.8 \cos \phi$ (Davies, 1978; Hosur et al., 1986). The experimental spin–spin coupling constants can be obtained directly from the 2-D homonuclear correlation experiments such as E. COSY and P.E. COSY. The sugar conformation can then be estimated from an analysis of the coupling constants (Hosur et al., 1986; Celda et al., 1989). The E.COSY spectrum of $[d(GCCGTTAACGGC)]_2$ reveals that $^3J_{H1'-H2'}$ is greater than $^3J_{H1'-H2''}$ for every nucleoside residue; this limits all deoxyribose pseudorotational phase angles to 90°–190° [see Figure 1 and Hosur et al. (1986) for corresponding values using the simple Karplus equation]. The exact J coupling constants for H2′–H3′, H2″–H3′, and H3′–H4′ cannot easily be determined due to passive couplings, including phosphorus couplings. However, since the intensities of the corresponding cross-peaks depend directly on the magnitude of the coupling constants, a careful inspection of the intensities of the H2′–H3′, H2″–H3′, and H3′–H4′ multiplets allows the sugar pucker to be qualitatively constrained into restricted ranges. In the case of the C2′-endo (P = 162°) sugar geometry, the H3′–H4′ and H2″–H3′ cross-peak intensities are either very weak or absent (J less than 1.0 Hz). In contrast, in the C1′-exo (P = 126°) sugar conformation, the H2″–H3′ cross-peak is still very weak or absent (less than 1.0 Hz) while the H3′–H4′ J coupling now becomes ca. 3 Hz. From inspection of the H3′–H4′ cross-peak region in $[d(GCCGTTAACGGC)]_2$ (Figure 2) and the H2′/H2″–H3′ region (data not shown), the sugar conformations of all the pyrimidines, as well as the purine G11, exhibit very weak or nonexistent H2″–H3′ couplings but quite respectable H3′–H4′ couplings, placing them close to P = 126°, i.e., C1′-exo ($_1E$), while G4, A7, A8, and

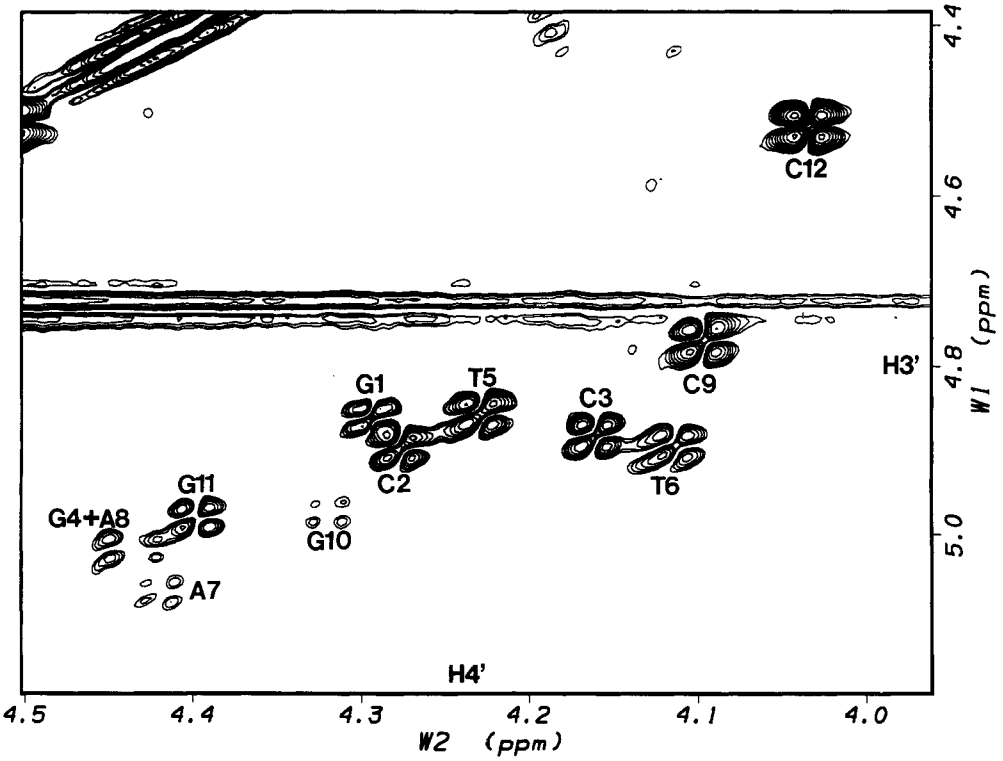

FIGURE 2: The E. COSY spectrum of the H3′ (4.4–5.2 ppm) to H4′ (4.0–4.5 ppm) region of $[d(GCCGTTAACGGC)]_2$. The spectrum was collected into 2048 × 400 points with sample spinning and zero-filled to 4096 × 2048 points. The sum of J couplings on H4′ was measured by the separation of peaks in the ω_2 domain.

G10 have much weaker, but quite detectable, H3′–H4′ couplings in combination with nonexistent H2″–H3′ couplings, placing P in the 140°–162° ${}^{2}_{1}T$ to ${}^{2}E$ range (i.e., around C2′-endo). Since DSPACE does not employ angle constraints, these dihedral angle constraints were converted to atom–atom distance constraints using DSPACE template files.

The sugar conformation can also be determined independently from the NOESY spectrum and used to corroborate the COSY conclusions. The H1′–H4′ and H2″–H4′ distances show the most variation with P and provide the most diagnostic information concerning sugar conformers (Chary & Modi, 1988; van de Ven & Hilbers, 1988; Celda et al., 1989). For P values greater than 126°, the H2′–H4′ NOESY cross-peak (≤3.85 Å) should be more intense than the H2″–H4′ cross-peak (≥3.85 Å). For $P \geq 144°$, the H1′–H4′ cross-peak (≥3.0 Å) should be less intense than the H1′–H2′ cross-peak (3.00 Å ± 0.05 for P = 70°–220°) in the linear initial-rate regime of the NOE buildup [this work, and van de Ven and Hilbers (1988)]. The absolute values of the H1′–H4′ distances reported by Wüthrich (1986) are shorter than the values derived in the present study and those reported by van de Ven and Hilbers, probably due to the lack of a furanose ring closure boundary condition in the earlier calculation [see Chary and Modi (1988)]. The sugar geometry of each residue determined by the NOESY distances agrees well with that determined from the E. COSY coupling criteria. It should however be noted that if the H4′ overlaps H5′ or H5″ peaks [especially from the (n + 1) residue], the H1′–H4′ cross-peak may contain additional intensity, making accurate H1′–H4′ cross-peak integration and distance calculation difficult; in such cases the COSY approach is less prone to error and is the method of choice.

Glycosidic Torsion Angle (χ). The χ angle can only be determined from intranucleotide distances between the base H6/H8/H5 protons and sugar H1′/H2′/H2″/H3′ protons since there are no base–sugar J couplings. Of these distances only the base–H1′ proton distance is directly dependent on the χ angle alone; the other distances are dependent on the combination of both χ and the sugar conformation. A minimum of two independent base to sugar proton distances plus the sugar conformation are required to determine the χ angle. For the duplex $[d(GCCGTTAACGGC)]_2$, the experimentally determined intranucleotide values of 2.2–2.5 Å for all H6/H8–H2′ distances, 3.4–3.7 Å for all H6–H1′ distances, and 3.5–4.0 Å for all H8–H1′ distances calculated from the initial rate (40, 80 ms) NOESY data, constrain all 12 χ angles to lie in the range between −100° and −140° [for χ–distance relationships, see Wüthrich (1986)]. The intranucleotide base–H2″ distance estimate is significantly contaminated by spin diffusion via the H6/H8–H2′–H2″ pathway, even at short mixing times, due to the short 1.8-Å geminal distance and was not used as a direct distance estimate in the present study. The H6/H8–H3′ apparent cross-relaxation rate is also contaminated by spin diffusion magnetization transfer via the H6/H8–H2′–H3′ pathway. These spin diffusion pathways invariably cause the calculated base proton to H2″/H3′ distances to be shortened in B-type DNA and using them produces incorrect values for the χ angle.

γ Angle. In principle the most direct approach to determining the γ torsion angle is through either J coupling or distance measurements between H4′ and H5′/H5′m protons (Altona, 1982). In practice, the severe overlap of H4′, H5′, and H5″ resonances in the 4.1–4.5 ppm region, as well as the existence of the passive couplings of H4′, H5′, and H5″, renders the measurement of $J_{H4'-H5'}$ and $J_{H4'-H5''}$ impractical, even for relatively short duplexes. Altona (1982) has used a theoretically calculated sum for $J_{H4'-H5'}$ and $J_{H4'-H5''}$ to determine the γ angle; the sum $J_{H4'-H5'} + J_{H4'-H5''}$ is more than 12.5 Hz in the case of γ values in the 180° or 300° range and less than 4 Hz in the case where γ is ca. 60°. However, $J_{H4'-H5'} + J_{H4'-H5''}$ cannot be determined experimentally in normal nonselective COSY experiments due to other passive couplings. We have found that the most rigorous and effective way to constrain γ is to use the sum of J couplings of H4′ combined with analysis of lower bound NOE distances to any protons in the 4.1–4.5 ppm H5′/H5″ region.

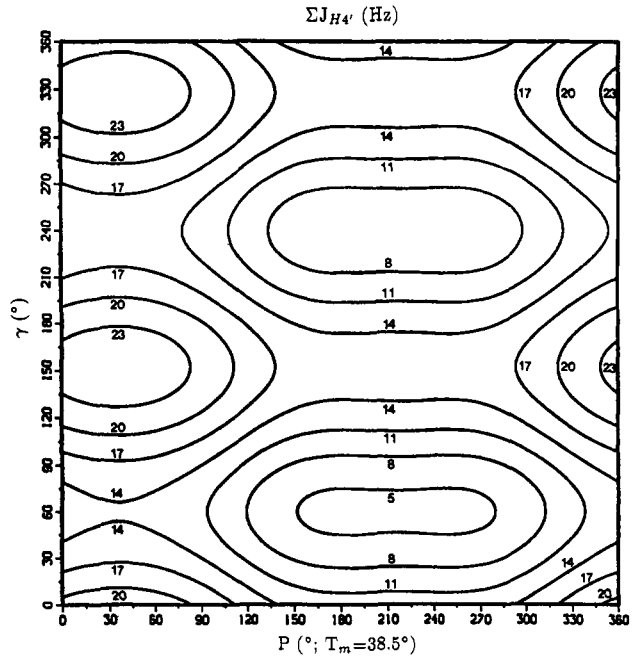

FIGURE 3: Contour map of the sum of J couplings on H4′ as a function of the pseudorotational and γ angles. This sum of J couplings includes $J_{H4'-H3'}$, $J_{H4'-H5'}$, and $J_{H4'-H5''}$.

The sum of H4′ J couplings ($\sum J_{H4'} = J_{H4'-H3'} + J_{H4'-H5'} + J_{H4'-H5''}$) depends on the sugar conformation P and the γ torsion angle, and this mutual dependence is shown in Figure 3 in the form of a contour plot. From Figure 2, the E. COSY outer peak-to-peak separations in the H4′ dimension, and hence the $\sum J_{H4'}$ values, for all residues in $[d(GCCGTTAACGGC)]_2$ were found to lie in the range 6.9–8.5 Hz, except for G1 which was 10.1 Hz. Thus, from Figure 3, all internal γ angles can be constrained to either 60° ± 40° or 240° ± 40°; i.e., there are two allowed ranges of γ consistent with the observed coupling, and these conclusions are valid without any prior information on the sugar P values. The 60° ± 40° γ values swing the geminal H5′/H5″ protons away from the helix into the solvent, while the 240° ± 40° values swing the H5′/H5″ protons in toward the helix axis, bringing them close to the H6/H8 base protons and to H2′ protons (ca. 2.6 Å). Thus the 2-fold ambiguity in γ from $\sum J_{H4'}$ analysis can easily be resolved by consulting the H6/H8–H5′/H5″ and H2′–H5′/H5″ regions of the NOESY spectra (shown in Figure 4A,B). The immediate realization that there are no H6/H8 cross-peaks to *any* H5′/H5″ resonances that are as large as the H6/H8–H1′ cross-peaks (3.2–3.9 Å) and that all H2′–H5′/H5″ cross-peaks are weaker than H2′–H1′ cross-peaks (ca. 3.0 Å regardless of P) effectively eliminates the 240° ± 40° γ range from consideration. Thus, the combined $\sum J_{H4'}$ and NOESY analysis constrains all internal γ values in $[d(GCCGTTAACGGC)]_2$ to the range 20°–100°; this corresponds to distance bounds of 2.28–2.70 Å for all H4′–H5′ and H4′–H5″ distances, and such generic bounds were used in the

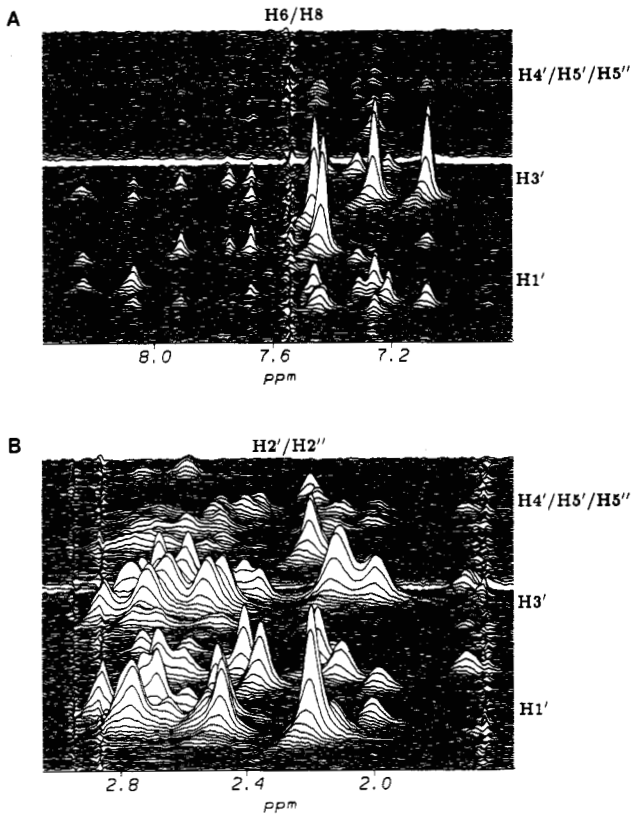

FIGURE 4: Stack plots of (A) the H8/H6/H5–H1′/H3′/H4′/H5′/H5″ region and (B) the H2′/H2″–H1′/H3′/H4′/H5′/H5″ region of the 80-ms NOESY spectrum of [d(GCCGTTAACGGC)]$_2$. Typical distances determined by scaling to the reference H5–H6 cross-peak are 3.6–4.0 Å for H8 to its own H1′ and 2.9–3.0 Å for H2′ to H1′.

distance file in the present study. Somewhat tighter constraints can be generated by using individual $\sum J_{H4'}$ values for each residue (e.g., the C3 $\sum J_{H4'}$ value is actually 6.9 Hz rather than 8.5 Hz) and by using the independent P values for individual sugars in conjunction with Figure 3; the latter input is especially effective when P lies in the 120°–155° range, leading to γ values of 60° ± 5° when $\sum J_{H4'}$ is ca. 8 Hz and P is ca. 120°.

Although the 4-bond J coupling constant ^4J between H4′ and phosphorus depends on the value of both γ and β angles, this long-range coupling constant was not included in the present analysis due to an ambiguity in the relationship between $^4J_{H4'-P}$ and the γ angle. Hall and Malcolm (1972a–c) reported that the maximum $^4J_{H4'-P}$ is ca. 2.7 Hz for a planar W conformation and that $^4J_{H4'-P}$ decreases for all other conformations to 0.0 Hz. Altona and co-workers (Mellema et al., 1984) reported $^4J_{H4'-P}$ values of ≤3 Hz in tetramer DNAs. Again our approach is a very conservative one in that the inclusion of a 1–3-Hz $^4J_{H4'-P}$ coupling would further restrict γ to an even narrower region around 60° since the allowed sum of 3-bond H–H couplings on H4′ would be further reduced by subtracting the 4-bond H4′–P coupling contribution.

While the above analysis requires an additional well-digitized E. COSY (or phase-sensitive COSY, DQF-COSY) spectrum to be run, somewhat less restrictive constraints on γ can be derived without COSY data via indirect coupling arguments or direct NOESY distance lower bounds, simply by reexamining a NOESY data set. For example, the H4′ NOESY line width contains information on $\sum J_{H4'}$. For a complex multiplet, the full-width at half-height ($\nu_{1/2}$) consists of the $\sum J$ splittings plus some contribution from the natural line width ($1/\pi T_2^*$). The natural linewidth ($1/\pi T_2^*$) can be approximated by the line width of an uncoupled proton such as a guanine or adenine H8 resonance. The $1/\pi T_2^*$ of H8 was found to be 10 ± 1 Hz (without using any window function). The contribution of $1/\pi T_2^*$ to the $\nu_{1/2}$ of a complex peak depends on the splitting of the multiplet components of the coupled protons. Simulation of the multiplet components of H4′ with combinations of the allowed pseudorotational phase angles and γ angles revealed that the contribution of $1/\pi T_2^*$ to $\nu_{1/2}$ can vary between 30 and 70% of the 10-Hz natural line width, i.e., ca. 3.0–7.0 Hz. For most sugars in [d(GCCGTTAACGGC)]$_2$ the H4′ $\nu_{1/2}$ determined from ω_2 slices of the NOESY spectrum (with no window function used in the ω_2 domain) lies in the range 11–15 Hz. For the case where $\nu_{1/2}$ is 14 Hz, subtracting the 3-Hz minimum contribution by $1/\pi T_2^*$ from the $\nu_{1/2}$ leaves the $\sum J$ contribution to H4′ as no more than 11 Hz; this alone restricts γ to the 10°–110° range when used in conjunction with the lack of H5′/H5″ NOEs to H6/H8/H2′ resonances. This conservative analysis yields a somewhat wider range of allowed γ angles, but individual analysis of each specific residue produces a more restricted range of individual γ angles. The usefulness of this approach is that, in the absence of adequately digitized E. COSY data (vide infra), the analysis merely requires reexamination of existing NOESY data.

Finally, γ can be partially constrained direclty in distance space, without resorting to any coupling–torsion angle arguments by using lower bound NOE distance constraints to *any* protons in the 3.8–4.6 ppm region, i.e., without H5′/H5″ assignments. Although the γ angle can be determined by two intranucleotide distances between stereospecific H5′/H5″ protons and sugar protons (Gronenborn et al., 1984), rigorous assignments of specific H5′/H5″ resonances are difficult in moderate-sized DNAs. Despite the lack of specific H5′ and H5″ assignments, lower bounds for all intranucleotide H6/H8–H5′/H5″ and H2′/H2″–H5′/H5″ distances in the molecule can be generated by examining the 7.0–8.4 to 3.8–4.6 ppm zone (Figure 4A) and the 1.5–3.0 to 3.8–4.6 ppm zone (Figure 4B), respectively. The distance corresponding to the strongest peak in the former region was used as the closest possible approach (lower bound) for all H6/H8–H5′/H5″ distances; thus without any H5′/H5″ assignments it can be deduced that no H6/H8–H5′ or H6/H8–H5″ proton pairs are closer than 3.8 Å. Considering the (n)H6/H8 distances to the (n − 1)H5′/H5″, (n)H5′/H5″, and (n + 1)H5′/H5″ protons, this produces 68 additional lower bound distance constraints in a symmetrical dodecamer duplex. A similar analysis of the H2′/H2″–H5′/H5″ NOESY region reveals that all H2′–H5′, H2′–H5″, H2″–H5′, and H2″–H5″ distances are greater than 3.3 Å, allowing this lower bound to be used for every residue. Including the (n − 1) and (n + 1) nearest-neighbor H5′ and H5″, this generates 136 additional lower bound constraints. The combined effect of these additional distance constraints prevents the H5′/H5″ geminal pair from swinging over the furanose ring; i.e., they constrain γ to the 0°–120° range. It is worth noting that these are rather conservative lower bounds distance constraints on γ. Instead of using these general lower distance bounds, the individual determination of each specific residue H6/H8–H5′/H5″ and H2′/H2″–H5′/H5″ lower distance bounds (even without stereospecific H5′ and H5″ assignments) can further restrict the γ angle of each residue [for distance–γ relationships, see Wüthrich (1986)]. Implicit in this analysis is the assumption of an H5′/H5″ correlation time similar to that of other protons, and the only way to experimentally overestimate these lower bounds would be the existence of large amplitude na-

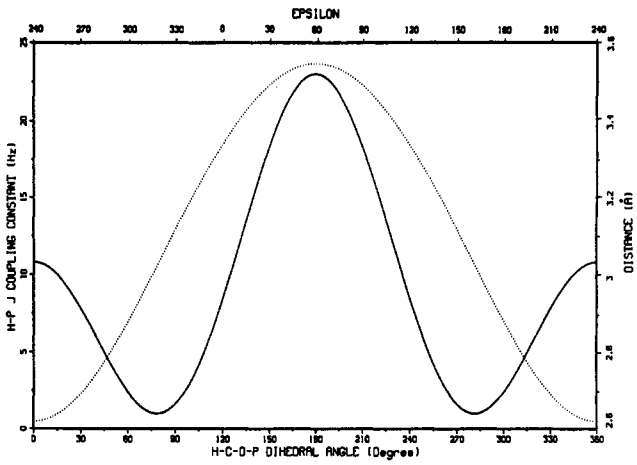

FIGURE 5: Vicinal J coupling constants and distances between H and P (H–C–O–P) as a function of ϵ and the H–C–O–P torsion angle. The solid line represents J_{H-P}, and the dotted line is the distance between phosphorus and proton atoms.

nosecond internal motions of the H5′/H5″ protons.

ϵ and β Angles. The backbone angles ϵ and β are related to the ^{31}P–^{1}H 3-bond J coupling (J_{HCOP}) according to $J_{HCOP} = 15.3 \cos^2 \phi - 6.1 \cos \phi + 1.6$ where ϕ is the dihedral angle H–C–O–P (Lankhorst et al., 1984; Sklenar & Bax, 1987; van de Ven & Hilbers, 1988). The couplings and corresponding proton–phosphorus distances as a function of the dihedral angle are shown in Figure 5. $J_{P-H3'}$ can be determined directly from a heteronuclear correlation experiment (Sklenar & Bax, 1987; Nikonowitz & Gorenstein, 1990; Roongta et al., 1990; Powers et al., 1990), but for $J \leq 11$ Hz there are four possible solutions for the ϵ torsion angle (Figure 5), and hence $J_{P-H3'}$ alone is of limited value. There are no experimental methods for determining an exact value for ϵ, and various laboratories have resorted to either a comparison with crystal structures or energy minimization methods to select an ϵ angle in solution (Sklenar & Bax, 1987; Baleja et al., 1990; Nikonowitz & Gorenstein, 1990; Roongta et al., 1990; Powers et al., 1990). The former are susceptible to distortion by crystal packing forces, while the latter suffer from inadequate force field parameterization and are often calibrated on the former. For the β angle the vicinal $J_{P-H5'}$ and $J_{P-H5''}$ couplings and the geminal couplings $J_{H5'-H5''}$ cannot be determined in moderate size DNA or RNA molecules due to overlapping H5′ and H5″ resonances.

Earlier studies from this laboratory (Banks et al., 1989; Nerdal et al., 1989) attempted to use only NOE-derived distance constraints to elucidate DNA structure, without any J-coupling ϵ torsion angle constraints. The resulting back-calculated distance geometry structures frequently produced surprising backbone phosphodiester conformations not previously observed. Many phosphodiester linkages assumed a "tucked-in" phosphate geometry in which the phosphorus atom points inward toward the helix interior. Such "tucked-in" structures (g^+, g^-) are consistent with all explicitly measurable NOE distances and bond angle, bond length, and Van der Waals constraints but were not subjected to dihedral angle energy minimization. However, the H3′–C3′–O3′–P torsion angle in such "tucked-in" phosphates is ca. 180° (ϵ = 60°) which, regardless of which set of Karplus-type parameters (Lankhorst et al., 1984; Hosur et al., 1988) is used, would lead to H3′–P coupling constants of 20 Hz or more (see Figure 5). Since the experimental H3′ overall line widths (including H–H couplings and the natural line width contribution) were all clearly less than 20 Hz, this phosphate conformation is obviously incorrect.

The H3′ line width ($\nu_{1/2}$) in DNA molecules consists of (i) proton–proton J couplings, i.e., $J_{H3'-H2'}$, $J_{H3'-H2''}$, and $J_{H3'-H4'}$, (ii) proton–phosphorus vicinal coupling, and (iii) some contribution from $1/\pi T_2^*$. Simulation of the H3′ multiplet lineshape for the various combinations of torsion angles reveals that the contribution to the multiplet $\nu_{1/2}$ from the 10-Hz natural line width is not less than 3 Hz. The $\nu_{1/2}$ for H3′ of G10 in [d(GCCGTTAACGGC)]$_2$ (using ω_2 domain data) was found to be 16 Hz. Thus, the sum of J couplings is no more than ca. 13 Hz after subtracting the minimal natural line width contribution. The sum of the three proton–proton J couplings on H3′ varies with sugar conformation but, depending on which Karplus values are used, the minimum allowed value for $\sum J_{H3'-H}$ is either 7.1 Hz (Rinkel & Altona, 1987) or 5.5 Hz (Hosur et al., 1986); hence the maximum allowed H3′–P coupling constant for residue G4 is 6–7.5 Hz. The experimental overall H3′ line widths for all residues in [d-(GCCGTTAACGGC)]$_2$ were observed to be in the range 13–18 Hz. Consequently, the maximal H3′–P coupling constants must lie in the 4.5–9.5-Hz range (using the most conservative 5.5-Hz minimal sum of proton–proton couplings) or in the 2.9–7.9-Hz range using the Rinkel–Altona $\sum J_{H3'}$ minimum of 7.1 Hz.

The sum of J coupling constants can be determined more accurately from COSY data than from gross line width, since the complications from uncertainties in the natural line width contribution do not enter into the analysis. Figure 6A shows a contour plot of the sum of J coupling constants for DNA H3′ as a function of the pseudorotational phase angle P and the ϵ torsion angle; Figure 6B shows the corresponding sum of J couplings for RNA H3′ protons. The smaller RNA $\sum J_{H3'}$ values reflect the absence of any H3′–H2″ coupling in RNA. Experimentally, the $\sum J_{H3'}$ value for G10 in [d-(GCCGTTAACGGC)]$_2$ was determined from the E. COSY experiment (using the ω_2 domain) and was found to be 10.0 Hz. Note that the measurement of J couplings from ω_1 domain profiles is less accurate due to the lower digital resolution (2.1 Hz/pt). A reasonably constrained range for ϵ can thus be derived from Figure 6A, together with the prior independent information on the pseudorotation phase angle. In general, the ϵ angle is more tightly constrained in the C1′-exo region (P = 125°–140°) than in the C2′-endo region (P = 160°–170°). Since the minimal sum of proton–proton J coupling constants is ca. 7 Hz (Rinkel & Altona, 1987) and the experimental $\sum J_{H3'}$ value of G10 is 10 Hz, the maximum possible G10 H3′–P coupling is ca. 3 Hz. Experimentally, the $\sum J_{H3'}$ values for all residues in [d(GCCGTTAACGGC)]$_2$ were found to lie in the 9.0–12.0-Hz range and hence the maximum possible H3′–P coupling constants are all in the 2–5-Hz range. The allowable ranges for ϵ (in the least constrained case where $\sum J_{H3'}$ = 12 Hz) are either 130°–200° (a B_I type phosphate) or 280°–350° (a B_{II} type phosphate). This conservative analysis yields a wider range of allowed ϵ angles, but the individual analysis of each specific residue produces a more restricted range of individual ϵ angles. These indirectly derived proton–phosphorus J coupling constants (≤5.0 Hz) agree well with those measured directly via the ^{1}H–^{31}P correlation experiments of Sklenar and Bax (1987) and Gorenstein and co-workers (Nikonowitz & Gorenstein, 1990; Roongta et al., 1990; Powers et al., 1990), which yielded 4 possible ϵ angles, but similarly could not distinguish between B_I and B_{II} phosphate conformations. Sklenar and Bax (1987) rejected the B_{II} phosphate form exclusively on the basis of theoretical energy considerations; i.e., they claim that the B_{II} conformation is energetically less stable in solution and only

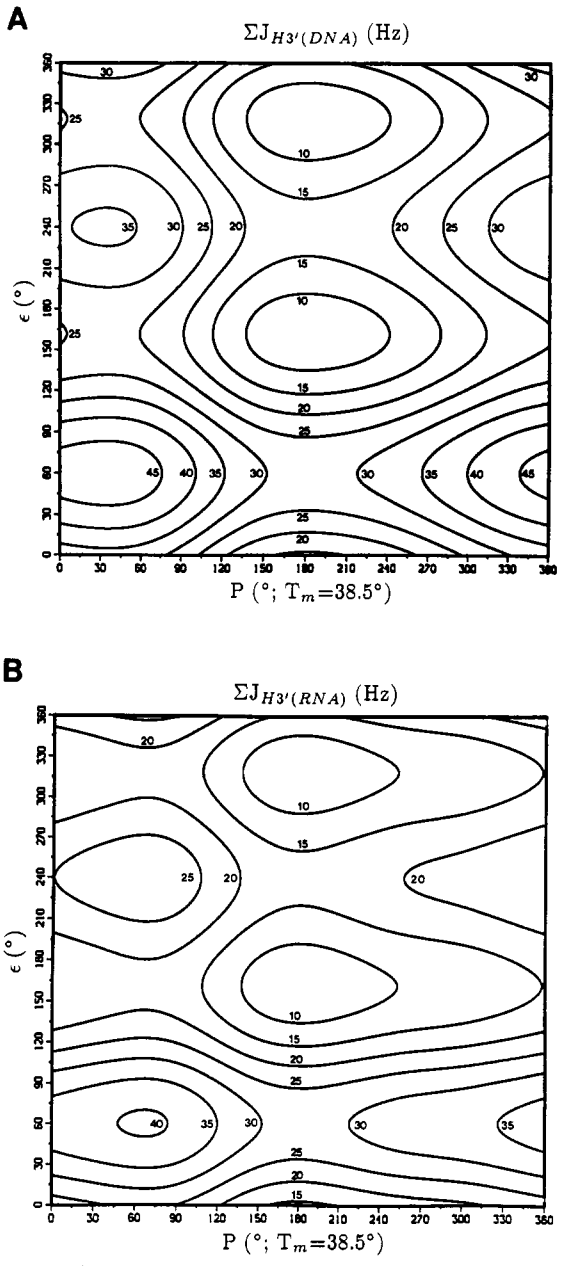

FIGURE 6: Contour maps of the sum of J coupling constants for H3′ in (A) DNA and (B) RNA as a function of the pseudorotational phase and ϵ angles. The sum of J coupling constants includes $J_{H3'-H4'}$, $J_{H3'-H2'}$, $J_{H3'-H2''}$ (for plot A), and $J_{H3'-P}$.

exists in the solid state, where it is stabilized by packing forces. Experimentally, these two phosphodiester conformations cannot be distinguished directly from NMR data; in the present work we did not impose a specific B_I or B_{II} conformation but merely used the corresponding H3′–P distance bounds of 2.80–3.25 Å, which are common to both B_I and B_{II}, as input distance constraints in the DSPACE algorithm. Interestingly, the structure of $[d(GCCGTTAACGGC)]_2$ determined in the present study was always found to contain only B_I phosphate conformations (vide infra), despite the lack of any direct input favoring the B_I conformation over the B_{II} conformation. Presumably, the combination of additional constraints on the backbone angles β, γ, δ, and $\alpha + \zeta$ (vide infra), together with the side-chain nearest-neighbor distance constraints, indirectly screens out the B_{II} backbone geometry.

Baleja et al. (1990) used H2′–H3′ DQF-COSY data as a consistency check for $\sum J_{H3'}$ and H3′–P coupling constants; they also observed H3′–P coupling constants of less than 5 Hz for all residues in their DNA sequence. The ϵ angles in their structure were not, however, experimentally constrained by the H–P coupling constant but were independently locked into the B_I conformation in solution based on crystallographic data. The E. COSY H3′–H4′ $\sum J_{H3'}$ coupling analysis described here is similar to the DQF-COSY H2′–H3′ analysis used to check the consistency of their final structure by Baleja et al. (1990); however, the present approach is greatly improved by the use of the two-dimensional plot (Figure 6A,B), which more effectively accounts for the ϵ angle dependencies and uses the results as distance bounds input to structure determination rather than as a post facto consistency check. Our analysis does not require knowledge of the individual J couplings but only the sum $\sum J_{H3'}$ in order to limit the range of ϵ and the H3′–P distance. Similarly in RNA, the ϵ angle can be determined from Figure 6B without ^{31}P–^{1}H correlation data, if either the H2′–H3′ or the H3′–H4′ cross-peak is resolved. We are currently involved in RNA and chimeric RNA–DNA duplex structural studies. While the H2′, H3′, H4′, H5′, H5″ region of RNA is spectrally quite overlapped, in many instances either the H2′–H3′ or H3′–H4′ E. COSY cross-peak is adequately resolved to assess a reasonable estimate of $\sum J_{H3'}$ to use in conjunction with Figure 6B to establish the H3′–P limits for several residues in the RNA segment.

In principle the P–O–C5′–C4′ torsion angle β can be determined from the $J_{H5'-P}$ and $J_{H5''-P}$ coupling constants, but in practice it is virtually impossible to measure these couplings in moderate-sized duplexes due to the chemical shift redundancy of the H4′, H5′ and H5″ protons. For the same reason, even the sum of J coupling constants for H5′ and H5″ cannot usually be determined from COSY data; i.e., $\sum J$ analysis (vide supra) cannot be used. However, the line width of the H5′ and H5″ multiplet resonances can easily be determined from NOESY data at longer mixing times (200–500 ms). At these long mixing times, easily detectable NOEs are observed by spin diffusion between intraresidue H6/H8 and H5′/H5″ protons, between H2′ and H5′/H5″ protons, and between interresidue H1′ and H5′/H5″ protons. In addition, the individual H5′ and H5″ line widths can be determined from individual ω_1 slices containing H6/H8/H1′/H2′ to H5′/H5″ NOEs. The overall H5′ line width ($\nu_{1/2}$) is composed of (i) proton–proton J couplings including the geminal $J_{H5'-H5''}$ and vicinal $J_{H5'/H5''-H4'}$ couplings, (ii) proton–phosphorus coupling, and (iii) some contribution from natural line width. Simulation of the multiplets for H5′ and H5″ with various torsion angles reveals that the contribution from the 10-Hz natural line width to the overall line width varies from 2 to 9 Hz depending on the β and γ angles. Line widths of the H5′/H5″ peaks were found experimentally to all be less than 28 Hz in NOESY data analysis of $[d(GCCGTTAACGGC)]_2$. After subtraction of the minimal line width contribution (2 Hz), the maximum sum of J couplings is 26 Hz. For these sums of coupling analyses, 2-D contour plots of $\sum J$ for H5′ and H5″ were calculated and are shown as a function of β and γ angles in panels A and B of Figure 7, respectively. In these simulations, the geminal H5′–H5″ coupling constant was taken to be 12 Hz (Mellema et al., 1984). From Figure 7A and the observation that $\sum$H5′ and $\sum$H5″ couplings are *both* found experimentally to be less than 26 Hz, β cannot be less than 105°; similarly, from Figure 7B β cannot be greater than 255°, regardless of the γ value. Although in an isolated P–O–CH_2–CH system the methylene protons could exhibit sums-of-couplings of 25–26 Hz when β is 0° ± 15°, in a P–O–CH_2–furanose sugar ring β angles of 0° ± 30° are sterically impossible due to van der Waals reasons (steric clashes between the phosphate and the sugar

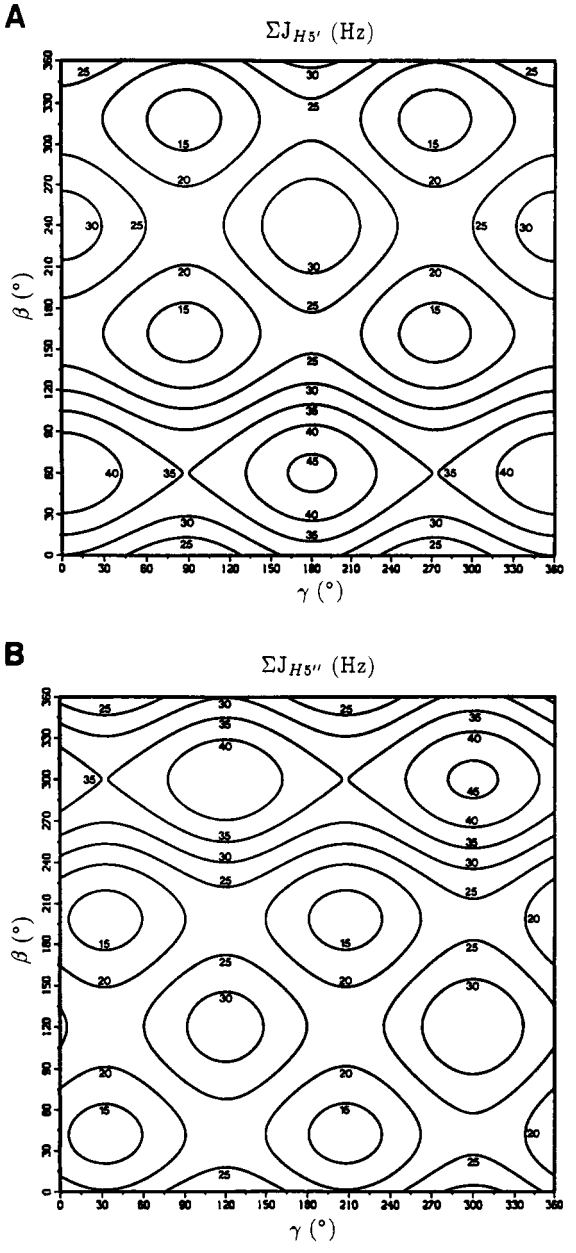

FIGURE 7: Contour maps of the sum of J coupling constants for (A) H5′ and (B) H5″ as a function of the γ and β angles. The sum of J couplings includes $J_{H5'-H5''}$, $J_{H5'/H5''-H4'}$, and $J_{H5'/H5''-P}$. A value of 12 Hz was used for the geminal J coupling between H5′ and H5″.

ring) and can be rejected. Thus, β can readily be constrained to the range 105°–255° merely by observing the H5′ and H5″ line widths (without knowing which is which). Although β can be further constrained by superimposing the already determined 20°–100° range for γ (vide supra), in the present study only the wider generic 180° ± 75° β constraints were used, after transformation to distances, in the distance bounds matrix. The H5′–P and H5″–P distances corresponding to these angle bounds are 2.65–3.40 Å. The β angle can thus be partially constrained using ordinary NOESY data alone, without any other structural assumptions. The experimentally determined β angle range encompasses the P–O–C5′–C4′ torsion angles found in crystallographic data (Holbrook et al., 1978) and from quantum chemical predictions (Saran et al., 1973).

α and ζ Angles. Since the α (O3′–P–O5′–C5′) and ζ (C3′–O3′–P–O5′) angles do not involve any protons, it is impossible to directly constrain these angles using ^{1}H NMR spectroscopy since there are no relevant J couplings or direct NOEs. One approach has been attempted using the chemical shift of ^{31}P (Roongta et al., 1990; Powers et al., 1990), but the relationship between chemical shift and backbone torsion angles is not straightforward and phosphorus chemical shifts are influenced by parameters other than the adjacent torsion angles. Dickerson and Drew (1981) reported a linear correlation between ζ and ϵ in the crystal structure of a DNA dodecamer; Gorenstein also reported a ϵ–ζ correlation in a DNA solution structure determined by NMR and energy minimization (Nikonowitz & Gorenstein, 1990) although DNA energy potentials (especially electrostatics) are notoriously unreliable. Nevertheless, the ζ angle may well be limited by the ϵ angle. In the NOESY spectra, inspection of the H1′ (5.2–6.2 ppm) to H4′/H5′/H5″ (3.4–4.6 ppm) region shows that the shortest experimental distance for any H1′–H5′/H5″ distance is 3.0 Å; note that no explicit assignments are required for this conclusion. Thus, we can safely conclude that the (n)H1′–(n + 1)H5′/H5″ distance is ≥3.0 Å. Including the (n)H5′/H5″ and (n – 1)H5′/H5″, this generates a further 68 additional lower bound constraints. The (n)-H1′–(n + 1)H5″ distance is closely related to both the backbone and sugar conformations. In this context, we note that β is already partially limited by line width/coupling constraints, γ and δ are restricted by both NOE and coupling constraints, and ϵ is constrained by the limits on the proton–phosphorus coupling. Thus, the major function of the experimentally determined (n)H1′ to (n + 1)H5′/H5″ lower bound is to indirectly restrict the remaining torsion angles α and ζ, given the C–O–P–O–C–C bond lengths and bond angles. It is important to emphasize that none of the individual backbone distance constraints, whether derived from couplings or NOEs, are precise single-valued solutions, they consist of conservative partially restricted zones or ranges. However, the combination of these ranges of allowed values work synergistically to produce quite highly restricted overall allowed structures that satisfy all the observed parameters.

Distance Geometry Structure—Testing the Additional Constraints. In order to detrmine the importance and influence of these backbone constraints, canonical A-form and B-form DNA structures were refined by using a simulated annealing algorithm combined with conjugate gradient refinement against a distance bounds matrix containing the experimentally determined distance bounds either with or without the additional backbone constraints described above. The initial distance constraints include intrasugar H–H, intraresidue base–H1′/H2′, interresidue base–H1′/H2′/H2″, and base–base proton distances. The distance constraints for the backbone linker were (i) intraresidue H6/H8–H5′/H5″, H2′/H2″–H5′/H5″ lower bounds, and H4′–H5′/H5″ distance bounds for γ, (ii) (n)H3′–(n + 1)P distance bounds for ϵ, (iii) P–H5′/H5″ distance bounds for β, and (iv) interresidue (n)H1′–(n + 1)H5′/H5″ lower bounds for the combination of all backbone angles including α and ζ. In parallel, a random-embedded DG structure was also refined against the same distance bounds matrix containing all the backbone constraints. From each starting structure, five refined structures were generated to produce close-to-global-minimum structures; the structures derived from A-DNA, B-DNA, and DG-DNA with the lowest penalty were then chosen for comparison. Figure 8A shows the superimposed structures of [d-(GCCGTTAACGGC)]$_2$ refined from both canonical A-form and B-form DNAs without backbone angle constraints; Figure 8B shows the corresponding A-derived and B-derived structures refined with backbone constraints. To measure the similarity between structures, coordinate root-mean-square deviations

Table I: Coordinate Root-Mean-Square Deviations per Atom (Å) between Structures Determined by Distance Geometry[a]

	B[b]	A[b]	RB[c]	RA[c]	DG[c]
B[b]		1.56	0.93	0.76	1.13
A[b]	1.01		1.49	1.43	1.93
RB[c]	0.77	1.21		0.50	1.02
RA[c]	0.62	1.03	0.30		0.74
DG[c]	0.71	1.21	0.37	0.30	

[a] RMSD values for all 12 base pairs are above the diagonal and those for the middle 8 base pairs are below the diagonal. [b] The structures refined from ideal B-form and A-form DNA structures without additional backbone constraints. [c] The structures refined from ideal B- and A-form DNA and random-embedded structures with additional backbone distance constraints.

Table II: Backbone Torsion Angles (deg) of Structures Determined by Distance Geometry without Backbone Distance Constraints[a]

residue	ϵ	ζ	α	β	γ	δ
C2	53 (1)	56 (4)	263 (6)	182 (1)	27 (6)	117 (9)
C3	116 (80)	341 (87)	218 (2)	208 (1)	45 (6)	108 (8)
G4	180 (6)	272 (5)	274 (51)	196 (13)	30 (11)	124 (4)
T5	191 (6)	262 (4)	310 (2)	180 (11)	34 (4)	120 (6)
T6	115 (88)	356 (112)	312 (4)	177 (1)	32 (0)	112 (5)
A7	181 (4)	265 (7)	294 (31)	195 (13)	5 (37)	130 (11)
A8	184 (6)	267 (4)	349 (53)	172 (16)	353 (41)	130 (10)
C9	124 (93)	338 (109)	310 (1)	175 (2)	31 (6)	107 (9)
G10	199 (4)	258 (1)	273 (69)	193 (12)	21 (16)	139 (4)
G11	196 (0)	264 (0)	320 (1)	185 (1)	357 (7)	132 (1)

[a] The results are means followed by their RMS deviations in parentheses. Structures were determined from canonical A- and B-form structure without backbone distance constraints.

Table III: Backbone Torsion Angles (deg) of Structures Determined by Distance Geometry with Backbone Distance Constraints[a]

residue	ϵ	ζ	α	β	γ	δ
C2	169 (11)	279 (9)	243 (19)	181 (6)	54 (21)	116 (4)
C3	168 (8)	286 (6)	269 (11)	200 (10)	54 (3)	110 (3)
G4	149 (3)	298 (2)	282 (4)	182 (2)	55 (4)	113 (2)
T5	184 (1)	270 (1)	237 (5)	213 (6)	68 (2)	110 (2)
T6	176 (1)	281 (1)	288 (4)	178 (3)	56 (1)	109 (1)
A7	180 (1)	277 (0)	284 (1)	183 (1)	58 (1)	116 (1)
A8	172 (6)	281 (6)	276 (2)	179 (2)	60 (2)	115 (1)
C9	173 (3)	288 (3)	267 (8)	182 (6)	64 (1)	108 (2)
G10	196 (1)	261 (1)	266 (3)	198 (3)	40 (5)	134 (3)
G11	197 (1)	262 (1)	282 (3)	187 (2)	33 (1)	124 (1)

[a] The results are means followed by their RMS deviations. Structures were determined from random-embedded DG and canonical A- and B-form structures with additional backbone distance constraints described in the text.

(RMSD) per atom are shown in Table I for all pairwise combinations of the structures including the refined DG-derived structure. The structures refined with the backbone distance constraints (0.3–0.4 Å RMSDs for the middle 8 base pairs) are better converged than those without backbone distance constraints (RMSD ca. 1.0 Å). Table II lists the backbone angles of the two structures refined from ideal A- and B-form DNA without backbone constraints, and Table III lists the backbone angles of the three structures refined with backbone constraints from random-embedded DG-, ideal A-, and B-form DNA. The backbone angles of the structures refined without backbone constraints have quite large variations in ϵ, ζ, α, and γ. The ϵ values include some "tucked-in" phosphate conformations (ϵ = 50°–120°), and γ includes some eclipsed conformations (γ = 0° ± 10°), which are at variance with the NMR coupling and H5′/H5″ NOE data. In contrast, the structures refined with the additional backbone constraints have a quite narrow range of variation for each specific backbone angle. Furthermore, the overall range for any particular torsion angle throughout the molecule is now much less variable; the ϵ angle range is 145°–200°, the ζ angles are

Table IV: Backbone Torsion Angles (deg) of Structures Determined by DG with Only NOE-Related Backbone Distance Constraints[a]

residue	ϵ	ζ	α	β	γ	δ
C2	54 (3)	46 (16)	263 (7)	175 (4)	39 (12)	105 (4)
C3	114 (82)	343 (88)	199 (0)	212 (13)	65 (1)	100 (3)
G4	54 (0)	63 (0)	248 (57)	188 (13)	60 (10)	113 (0)
T5	183 (3)	268 (2)	184 (1)	194 (4)	67 (7)	108 (2)
T6	182 (3)	277 (1)	291 (8)	173 (9)	59 (6)	114 (3)
A7	121 (88)	349 (105)	279 (4)	189 (3)	49 (3)	115 (7)
A8	55 (2)	64 (6)	234 (59)	180 (1)	65 (10)	109 (1)
C9	57 (0)	49 (2)	188 (0)	188 (8)	61 (2)	94 (0)
G10	195 (0)	260 (0)	208 (2)	197 (6)	56 (4)	130 (6)
G11	201 (2)	281 (18)	282 (4)	185 (2)	35 (1)	127 (6)

[a] The results are means followed by RMS deviations. Structures were determined by distance geometry (DG) with only the NOE-related backbone distance constraints starting from canonical A- and B-form structures described in the text.

260°–300°, the α angle range is 230°–290°, the β angles are 170°–220°, the γ angles are 40°–70°, and the δ angles are 100°–140°. All backbone angles in the structures refined with backbone angle constraints were found to agree well with the initial distance bounds. It is important to emphasize that these results were derived purely from experimental NMR data and distance geometry without any additional input such as assumptions from X-ray crystallographic data or energy minimization. However, the resulting backbone angle ranges agree quite well with X-ray crystal data (Holbrook et al., 1978; Prive et al., 1991). Interestingly, while the backbone conformation can theoretically be B_I (ϵ = 180°, ζ = 270°) or B_{II} (ϵ = 270°, ζ = 180°) purely on the basis of the NMR-derived H3′–P J coupling constants, all ϵ angles were found to adopt the B_I conformation. This may well be related to the indirect effect of H1′–(n + 1)H5″ distance constraints on the ϵ torsion angle.

To study the significance of only the J coupling-derived backbone information, the A-form and the B-form canonical DNA structures were refined against the distance matrix with only the "NOE-derived" backbone constraints, i.e., H1′–H5′/H5″, base–H5′/H5″, and H2′/H2″–H5′/H5″ distance bounds. These refined structures were much less well constrained, containing ϵ values of 50°–200°, ζ values of 40°–290°, α values of 180°–300°, β values of 170°–220°, γ values of 30°–70°, and δ values of 90°–130° (Table IV), indicating insufficient distance constraints to limit ϵ, ζ, and α. Thus, the H3′–P and H5′/H5″–P distance constraints on ϵ and β (derived from torsional constraints via maximal H–P coupling) are obviously required to fully determine the backbone conformation.

In the data presented above, we have emphasized the methods required to successfully constrain polynucleotide backbone conformation solely from proton NMR data. These methods were implemented and tested on the DNA duplex [d(GCCGTTAACGGC)]$_2$ and in the present paper we have focused on the efficacy of the methods rather than on the structure that emerges. The detailed analysis of the [d(GCCGTTAACGGC)]$_2$ structure, including local and global helical parameters, will be presented in a separate paper.

Summary

In DNA or RNA duplexes, the 6-bond backbone C3′–O3′–P–O5′–C5′–C4′–C3′ contains only four protons. This seriously limits the ability to define the backbone conformation by ^{1}H–^{1}H distance geometry (DG) methods. The problem is further compounded by the inability to assign the crowded H5′ and H5″ backbone protons and invariably leads to backbone-undetermined DG structures. In the present study we have implemented reliable methods to determine β, γ, and ϵ (and indirectly α and ζ) angles by lower bound NOE dis-

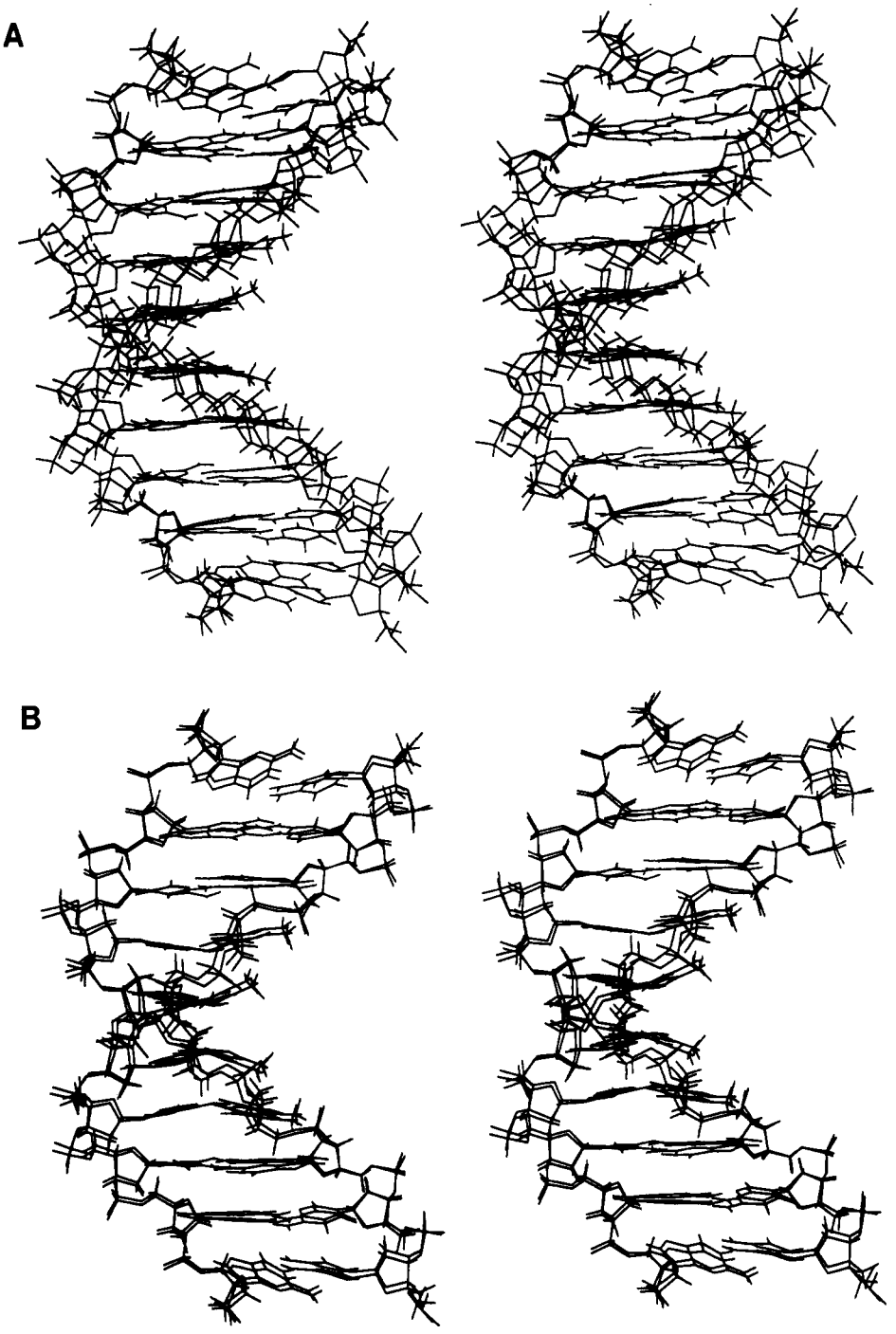

FIGURE 8: Stereoview (cross-eye) of superimposed structures for the internal 10-base-pair segment of $[d(GCCGTTAACGGC)]_2$ refined from canonical B-DNA and A-DNA using the DSPACE distance geometry algorithm either (A) without or (B) with additional back-bone constraints. Note the greatly improved backbone convergence in B.

tances to *unassigned* H5′/H5″ resonances combined with ^{1}H line widths or conservative use of $\sum J$ measurements. The γ angle is constrained by lower bound NOE distances from H2′ and from H6/H8 to *any* H5′/H5″ in the 3.8–4.6 ppm region as well as by $\sum J_{H4'}$. The ϵ angle is partially constrained by the H3′ line width and/or further constrained by $\sum J_{H3'}$ (E. COSY), i.e., subtracting the minimum possible $\sum J_{H3'-H}$ from the observed $\sum J_{H3'}$ (E. COSY) to arrive at the maximum possible $J_{H3'-P}$. The β angle is partially constrained via H5′–P and H5″–P distance bounds from H5′–P and H5″–P maximum J couplings derived from observed H5′ and H5″ line widths. The α and ζ angles are indirectly constrained by lower bounds on the observed (n)H1′ to $(n + 1)$H5′/H5″ NOEs combined with the prior partial constraints on β, γ, and ϵ. These added constraints greatly assist refinement of structures and lead to quite well-determined backbone conformations. Coordinate RMSD values per atom between structures refined from random-embedded DG structures, as well as ideal A- and B-form DNA, were less than 0.4 Å for the central 8 base pairs indicating good convergence. All individual backbone angles for the central 8 base pairs were constrained to within 10°.

Further refinement of these structures using back-calculation methods is in progress and will be reported in a future paper. To speed up these structural studies, a semiautomatic back-calculation approach has been developed. This approach, as well as the variation in helical parameters of the back-calculated structures, will also be discussed in a future paper.

ACKNOWLEDGMENTS

We thank Drs. John Orban and Peter Flynn for informative

discussions and Drs. Peter Flynn and Jeff Davis for critical reading. We also thank Susan Ribeiro and Julie Miller for DNA synthesis and Dr. Dennis Hare (and Hare Research) for providing the excellent software programs FTNMR, FELIX, and DSPACE.

Registry No. d(GCCGTTAACGGC), 138166-98-0.

REFERENCES

Altona, C. (1982) *Recl. Trav. Chim. Pays Bas 101*, 413–433.
Altona, C., & Sundaralingam, M. (1972) *J. Am. Chem. Soc. 94*, 8205–8212.
Arnott, S., & Hukins, D. W. L. (1972a) *Biochem. Biophys. Res. Commun. 47*, 1504–1509.
Arnott, S., & Hukins, D. W. L. (1972b) *J. Mol. Biol. 81*, 93–105.
Baleja, J. D., Pon, R. T., & Sykes, B. D. (1990) *Biochemistry 29*, 4828–4839.
Banks, K. M., Hare, D. R., & Reid, B. R. (1989) *Biochemistry 28*, 6996–7010.
Boelens, R., Koning, T. M. G., & Kaptein, R. (1988) *J. Mol. Struct. 173*, 299–311.
Boelens, R., Koning, T. M. G., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1989) *J. Magn. Reson. 82*, 290–308.
Borgias, B. A., & James, T. L. (1990) *J. Magn. Reson. 87*, 475–487.
Celda, B., Widmer, H., Leupin, W., Chazin, W. J., Denny, W. A., & Wüthrich, K. (1989) *Biochemistry 28*, 1462–1471.
Chary, K. V. R., & Modi, S. (1988) *FEBS Lett. 233*, 319–324.
Chary, K. V. R., Modi, S., Hosur, R. V., Govil, G., Chen, C.-Q., & Miles, H. T. (1989) *Biochemistry 28*, 5240–5245.
Clore, G. M., & Gronenborn, A. M. (1989) *Crit. Rev. Biochem. Mol. Biol. 24*, 479–564.
Crippen, G. M. (1981) *Distance Geometry and Conformational Calculations*, Research Studies/Wiley, Chichester.
Davies, D. B. (1978) *Prog. Nucl. Magn. Reson. Spectrosc. 12*, 135–225.
Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761–786.
Drobny, G., Pines, A., Sinton, S., Weitkamp, D. P., & Wemmer, D. E. (1979) *Faraday Symp. Chem. Soc. 13*, 49–55.
Gochin, M., & James, T. L. (1990) *Biochemistry 29*, 11171–11180.
Griesinger, C., Sorensen, O. W., & Ernst, R. R. (1986) *J. Chem. Phys. 85*, 6837–6852.
Griesinger, C., Sorensen, O. W., & Ernst, R. R. (1987) *J. Magn. Reson. 75*, 474–492.
Gronenborn, A. M., Clore, G. M., & Kimber, B. J. (1984) *Biochem. J. 221*, 723–736.
Haasnoot, C. A. G., de Leeuw, F. A. A. M., & Altona, C. (1980) *Tetrahedron 36*, 2783–2792.
Hall, L. D., & Malcolm, R. B. (1972a) *Can. J. Chem. 50*, 2092–2101.
Hall, L. D., & Malcolm, R. B. (1972b) *Can. J. Chem. 50*, 2102–2110.
Hall, L. D., & Malcolm, R. B. (1972c) *Can. J. Chem. 50*, 2111–2118.
Hare, D. R., & Reid, B. R. (1986) *Biochemistry 25*, 5341–5350.
Hare, D. R., Shapiro, L., & Patel, D. J. (1986a) *Biochemistry 25*, 7445–7456.
Hare, D. R., Shapiro, L., & Patel, D. J. (1986b) *Biochemistry 25*, 7456–7464.
Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 21*, 5129–5135.
Havel, T. F., Kuntz, I. D., & Crippen, G. M. (1983) *Bull. Math. Biol. 45*, 665–720.
Holbrook, S. R., Sussman, J. L., Warrant, R. W., & Kim, S.-H. (1978) *J. Mol. Biol. 123*, 631–660.
Hosur, R. V., Ravikumar, M., Chary, K. V. R., Sheth, A., Govil, G., Zu-Kun, T., & Miles, H. T. (1986) *FEBS Lett. 205*, 71–76.
Hosur, R. V., Govil, G., & Miles, H. T. (1988) *Magn. Reson. Chem. 26*, 927–944.
Kintanar, A., Klevit, R. E., & Reid, B. R. (1987) *Nucleic Acids Res. 15*, 5845–5862.
Kirkpatrick, S., Gelatt, C. D., Jr., & Vecchi, M. P. (1983) *Science 220*, 671–680.
Lankhorst, P. P., Haasnoot, C. A. G., Erkelens, C., & Altona, C. (1984) *J. Biomol. Struct. Dyn. 1*, 1387–1405.
Mellema, I.-R., Pieters, J. M. L., van der Marel, G. A., Van Boom, J. H., Haasnoot, C. A., & Altona, C. (1984) *Eur. J. Biochem. 143*, 285–301.
Marian, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953) *J. Chem. Phys. 21*, 1087–1092.
Nerdal, W., Hare, D. R., & Reid, B. R. (1988) *J. Mol. Biol. 201*, 717–739.
Nerdal, W., Hare, D. R., & Reid, B. R. (1989) *Biochemistry 28*, 10008–10021.
Nikonowicz, E. P., & Gorenstein, D. G. (1990) *Biochemistry 29*, 8845–8858.
Nikonowicz, E. P., Meadows, R. P., & Gorenstein, D. G. (1990) *Biochemistry 29*, 4193–4204.
Nilges, M., Clore, G. M., Gronenborn, A. M., Brünger, A. T., Karplus, M., & Nilsson (1987a) *Biochemistry 26*, 3718–3733.
Nilges, M., Clore, G. M., Gronenborn, A. M., Piel, N., & McLaughlin, L. W. (1987b) *Biochemistry 26*, 3734–3744.
Nilsson, L., Clore, G. M., Gronenborn, A. M., Brunger, A. T., & Karplus, M. (1986) *J. Mol. Biol. 188*, 455–475.
Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1985) *J. Magn. Reson. 66*, 187–193.
Powers, R., Jones, C. R., & Gorenstein, D. G. (1990) *J. Biomol. Struct. Dyn. 8*, 253–294.
Prive, G. G., Yanagi, K., & Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 177–199.
Reid, B. R. (1987) *Q. Rev. Biophys. 20*, 1–34.
Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621–649.
Roongta, V. A., Jones, C. R., & Gorenstein, D. G. (1990) *Biochemistry 29*, 5245–5258.
Saran, A., Pullman, B., & Perahia, D. (1973) *Biochim. Biophys. Acta 287*, 211–231.
Sklenar, V., & Bax, A. (1987) *J. Am. Chem. Soc. 109*, 7525–7526.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
van de Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1–38.
Westhof, E., & Sundaralingam, M. (1980) *J. Am. Chem. Soc. 102*, 1493–1500.
Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, pp 203–255, John Wiley & Sons, New York.