Solution Structure of the *Eco*RI DNA Sequence: Refinement of NMR-Derived Distance Geometry Structures by NOESY Spectrum Back-Calculations†

Willy Nerdal,‡,§ Dennis R. Hare,‖ and Brian R. Reid*,‡,⊥

Departments of Chemistry and Biochemistry, University of Washington, Seattle, Washington 98195

Received August 1, 1988; Revised Manuscript Received July 28, 1989

ABSTRACT: The solution structure of the self-complementary DNA duplex $[d(CGCGAATTCGCG)]_2$, which contains the *Eco*RI restriction site sequence GAATTC at the center, has been studied by two-dimensional nuclear magnetic resonance spectroscopy. Time-dependent nuclear Overhauser effect spectra were used to obtain the initial cross-relaxation rates between 155 pairs of protons. These initial cross-relaxation rates were converted into interproton distances and entered into a distance (bounds) matrix. A distance geometry algorithm (DSPACE) was used to create embedded starting structures and to refine these structures until they showed good agreement with the distance matrix; symmetry constraints were included in the refinement procedure, making the two strands in the refined distance geometry structures virtually identical and significantly improving the agreement with the distance matrix. The NOESY spectrum for one of these distance geometry structures was then calculated from the explicit coordinates by numerically integrating all the *z*-magnetization transfer pathways among neighboring protons within a specified radius. Distances in this distance geometry structure that did not agree with the experimental NOESY time course were then adjusted accordingly. This process was iterated until a good agreement between calculated and experimental NOESY spectra was reached. The final structure, which generates good agreement with the experimental NOESY spectrum, displays kinks at the C3–G4 base step and at the A6–T7 base step that appear to be similar to those reported for the *Eco*RI restriction site DNA bound to its endonuclease. The solution structure is not the same as the crystal structure of this DNA duplex.

The *Eco*RI restriction site DNA dodecamer, $[d(CGCGAATTCGCG)]_2$, has attracted much interest due to the fact that this sequence is one of the few DNA duplexes to crystallize in the B–DNA form (Wing et al., 1980; Dickerson & Drew, 1981). The single-crystal X-ray structure demonstrated a marked influence of base sequence on DNA helix structure (Dickerson & Drew, 1981; Dickerson, 1983). More recently the structure of the *Eco*RI endonuclease complexed to this DNA has been solved (McClarin et al., 1986). The specific binding of *Eco*RI endonuclease to its cognate hexanucleotide sequence GAATTC supports the notion that unique structural variations within this sequence are responsible for its recognition by the endonuclease.

Crystallization of DNA necessarily takes place under mildly dehydrating solution conditions; these conditions are reported to promote B–A or B–Z transitions in DNA (Feigon et al., 1985; Coll et al., 1986). Under dehydrating conditions, the B to A transition is the predominant structure alteration for adenine/thymine-rich DNA sequences, whereas alternating guanine–cytosine sequences undergo a B–Z-form transition under similar conditions (Pohl & Jovin, 1972; Coll et al., 1986). DNA sequences containing nonalternating guanine–cytosine regions such as GCC, exist as A-form duplexes when water availability is limited (Wang et al., 1982). This effect has been rationalized as the result of more efficient hydration by water bridging between the closer phosphate groups in the A-DNA and Z-DNA conformations in crystallization solutions containing 20–30% methylpentanediol (and similar solvents) in which the availability of *free* water is less than physiological (Saenger et al., 1986). The stem of a GC hairpin structure was shown to be a B-DNA structure in solution (Hare & Reid, 1986) but was found to be converted to Z-DNA by crystallization from 40% 2-propanol (Chattopadhyaya et al., 1988). In light of the potential structure-perturbing effects of crystallization solvents, it is important to determine DNA structure under the normal aqueous solution conditions in which it is recognized and bound by ligands; this can, in theory, be accomplished by NMR.[1]

The determination of biopolymer structure in solution relies on distance measurements. The interproton distances to be used in generating a solution structure are obtained by measuring the cross-relaxation rate between pairs of protons, which depends on the distance between those protons. If the distances are correct and the number of measured distances is large enough to avoid underdetermination, the structure can be calculated by distance geometry (DG) embedding methods; for recent reviews see Reid (1987) and Patel et al. (1987).

Assignments of all nonexchangeable protons except the 5′, 5″ protons in the $[d(CGCGAATTCGCG)]_2$ dodecamer have previously been reported by Hare et al. (1983). Although those studies were carried out at a lower temperature than the work reported here, apart from minor temperature variations in chemical shift, the spectrum remains essentially unchanged at 30 °C, and we have corroborated the assignments reported

† Supported in part by NIH Program Project Grant GM 32681 to B.R.R. and SBIR Grant GM 35620 to D.R.H.

‡ Department of Chemistry.

§ On leave from the Chemistry Department, University of Bergen, Bergen, Norway.

‖ Hare Research, Woodinville, WA 98072.

⊥ Department of Biochemistry.

[1] Abbreviations: 2D NMR, two-dimensional NMR spectroscopy; COSY, two-dimensional correlated NMR spectroscopy; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser effect spectroscopy; DG, distance geometry; DM, deviations between distance matrices; RMSD, root-mean-square coordinate deviations; RMD, restrained molecular dynamics; BCK, back-calculation of NOESY spectra.

0006-2960/89/0428-10008$01.50/0

earlier. The exchangeable imino protons of this duplex have been assigned previously by Patel et al. (1983).

In this work we report the determination of the solution structure of the $[d(CGCGAATTCGCG)]_2$ duplex. Interproton distance estimates based on the two-spin approximation were calculated from the initial rate of cross-relaxation between pairs of assigned protons in a time-dependent NOESY data set. A family of closely related distance geometry (DG) structures was calculated from these distances as described previously (Hare & Reid, 1986; Nerdal et al., 1988), and the NOESY spectrum corresponding to the coordinates of one of these structures was then calculated and compared to the experimental NOESY data to reveal distance errors caused by the two-spin approximation. Crosspeaks that did not match were identified, and the corresponding proton pairs were adjusted by increasing or decreasing the distance separating those two protons in the structure, as described by Banks et al. (1989). The NOESY spectrum of the adjusted, reannealed, structure was then recalculated and compared with the experimental NOESY spectrum again. This process was iterated until the two spectra matched almost perfectly, indicating that a structure very close to the true solution structure had been generated.

To compare the solution structure of $[d(CGCGAATTCGCG)]_2$ with the crystal structure of this dodecamer (Wing et al., 1980; Dickerson & Drew, 1981), the NOESY spectra corresponding to the crystal coordinates (with added protons) were also calculated. Since the two strands 1–12 and 13–24 are not identical in the crystal structure despite the twofold sequence symmetry, the NOESY spectra were calculated for each of the two strands separately.

Materials and Methods

The DNA dodecamer $[d(CGCGAATTCGCG)]_2$ was synthesized on a 10-μmol scale by using solid-phase phosphoramidite techniques as described previously (Hare et al., 1983). The crude DNA was purified by chromatography on Sephadex G-25, as described by Kintanar et al. (1987), and lyophilized to dryness. The purified lyophilized duplex was dissolved in 0.4 mL of buffer containing 10 mM sodium phosphate at pH 7.0 and repeatedly lyophilized to dryness, first from the aqueous buffer and then from D_2O solution. Finally 0.4 mL of 99.996% D_2O was added, and the solution was transferred to a 5-mm NMR tube. The NMR samples contained 26–28 mg of DNA, leading to a sample concentration of ca. 61 mg/mL, or 8.5–9 mM in duplex.

For the studies in H_2O solution, the 0.4-mL sample contained 90% H_2O and 10% D_2O. One-dimensional NOE spectra of the imino protons were acquired on a Bruker WM-500 spectrometer; the time-dependent NOESY experiments in D_2O were acquired on a home-built 500-MHz spectrometer (Gladden and Drobny, unpublished data). The sample temperature was regulated at 30 °C for all experiments in D_2O. The 1D NOE spectra of the exchangeable protons in H_2O were carried out at 18 °C to minimize the effects of solvent exchange. Eight NOESY spectra in D_2O, with mixing times of 0.001, 0.03, 0.06, 0.09, 0.12, 0.20, 0.50, and 1.2 s, were all collected within a single 6-day period without removing the sample from the spectrometer or changing any frequencies or gain settings. The pure absorption NOESY spectra were collected, by use of phase-sensitive methods (States et al., 1982), into 1024 complex points in t_2 and into 400 points in t_1. For each t_1 value, 32 scans were collected with a relaxation delay of 2.0 s between transients. All experiments in D_2O were carried out without suppressing the water resonance. The H_2O 1D NOE difference spectra of the exchangeable protons were acquired with irradiation times of 0.05, 0.10, 0.20, and 0.30 s, with the carrier frequency centered at 11.7 ppm and a Redfield 21412 observation pulse of 276 μs offset 3426 Hz from the water resonance (Redfield, 1978; Hare & Reid, 1982; Chou et al., 1984).

Following collection, the experimental data were processed on a DEC microVAX II computer using the FTNMR program developed by D.R.H. (Hare Research, Inc.). The phase-sensitive time-dependent NOESY spectra to be used for initial rate distance calculations were not apodized in the t_2 dimension, but were zero-filled to 2048 points and slightly (2 Hz) line-broadened to increase the signal-to-noise ratio. The first t_1 experiment was multiplied by 0.5 to suppress t_1 ridges in the spectra (Otting et al., 1985). The t_1 data were zero-filled to 2048 points and processed with a sine-squared 90° phase-shifted function; this light apodization function does not distort the signal intensities, and the resulting peaks have no truncation effects. The volume integral of each resolved crosspeak corresponding to each detectable proton–proton interaction was then measured from the NOESY data at 30, 60, 90, 120, and 200 ms. The integration algorithm (written by D.R.H.) calculates peak volumes by summating all points in the transformed data matrix within an elliptical cylinder encompassing the individual crosspeak; the size of the ellipse encompassing the cross peak to be integrated is set by the operator. Each unknown distance (r_{ij}) can be estimated relative to a fixed distance (r_{kl}) from the initial cross-relaxation rates R_{ij} and R_{kl} according to

$$R_{ij}/R_{kl} = (r_{kl})^6/(r_{ij})^6$$

if the two pairs of protons have the same correlation time, τ_c, and the product $\omega\tau_c \gg 1$. The latter criteria are satisfied in 12-bp DNA duplexes (Reid et al., 1989). Distances involving methyl groups were scaled to the thymine H6–CH_3 cross-relaxation rate by using a methyl pseudoatom located at the center of the triangle formed by the three protons (Tropp & Redfield, 1981). We used a methyl pseudoatom–H6 distance of 2.9 Å on the basis of thymine crystal coordinates. Because the pseudoatom simplification is not completely valid, and because X-ray data have a tendency to underestimate C–H bond lengths (Frey et al., 1973), the distances calculated between methyl groups and other protons were entered into the distance matrix with quite wide constraints varying from ±0.45 to ±1.15 Å depending on the distance involved and the quality of the particular initial rate data. A cytosine H5–H6 distance of 2.45 Å was used as a reference distance for all other NOEs between nonexchangeable protons. Although Clore and Gronenborn (1984) have claimed a 3-fold shorter correlation time for sugar protons compared to base protons, their conclusion is based on attributing a spin-diffusion effect to local motion, and we have recently demonstrated that sugar protons and base protons actually have the same correlation time (Reid et al., 1989).

Using the above procedures for distance measurement from initial cross-relaxation rates in high-quality NOESY spectra, we estimate a measurement precision within 0.1–0.2 Å for distances of 2–3 Å (where reliable crosspeak volumes can be determined at 30–60 ms) and no more than 0.2–0.3 Å for distances of 3–4 Å (where the earliest reliable crosspeak intensities require about 100 ms to develop). Error estimates were determined by plotting the best slopes through the points to generate the estimated distance together with lines corresponding to ±0.1, ±0.2, and ±0.3 Å; the first two or three time points were always comfortably within the quoted delimiters. These error estimates in the calculated distances were entered into the corresponding upper and lower bounds of the distance

matrix. Distances involving imino protons are inherently less accurate than those between nonexchangeable protons for several reasons. Imino proton magnetization can relax via solvent exchange (although this is reduced to less than 1 s^{-1} at 18 °C), and there is no suitable reference fixed distance involving the imino proton. Regarding the latter problem, an intrabase pair thymine imino to adenine H2 distance of 2.9 Å was used as a reference for scaling interbase pair imino–imino and imino–H2 cross-relaxation rates. This value was chosen based on the average ring NH to adenine H2 distance in a normally hydrogen-bonded classical AT pair. Because of these potential sources of error, distances involving imino protons were given quite wide constraints in the distance matrix, ranging from ±0.5 to ±0.7 Å, i.e., 1.0–1.4 Å between the upper and lower bounds, depending on the particular case and quality of the rate data.

Distance Geometry Algorithm

The essential concepts involved in the distance geometry algorithm DSPACE have been described earlier (Hare & Reid, 1986; Hare et al., 1986a,b; Nerdal et al., 1988; Banks et al., 1989), and only a brief summary of the algorithm will be given here. All distances implicit in the primary structure, as well as those obtained from the experimental data, were assembled and stored in a bounds matrix. Each undefined lower bound was set to the sum of the van der Waals radii of the two atoms involved, and the corresponding undefined upper bounds were set to an arbitrarily large value (1000 Å in this case). The bounds matrix now contained all the availabile distance information and was next subjected to several smoothing procedures (Nerdal et al., 1988). Finally, the entire bounds matrix was smoothed by using the triangle and inverse triangle inequalities, as described by Crippen and Havel (Crippen, 1981; Havel & Wüthrich, 1985).

At this stage, several different distance matrices were generated by using random numbers between the upper and lower bounds for each matrix element to produce several symmetrized trial matrices, each of which was converted to a metric matrix and embedded in 3-space (Havel et al., 1979; Crippen, 1981). The resulting structures are all distorted to varying degrees since the trial distances define structures that generally do not fit into 3-space. Refinement of the trial structures was carried out by using a combination of conjugate gradient refinement and an annealing algorithm, which uses the deviation from the distance bounds as a pseudopotential (Nerdal et al., 1988). Since the NMR spectrum indicates perfect twofold symmetry for the duplex in solution, a symmetry penalty was also included in the penalty function at later stages of the refinement (the two DNA strands are not identical in the initial embedded structure); with the added penalty, the strands become virtually identical in the refined structure. Refinement of the molecule with the additional symmetry constraints greatly reduces the conformational space available to the structure and significantly improves the agreement between the refined structure and the bounds matrix.

The refinement of each trial structure required 1–2 h on a Convex C1 computer or ca. 30 h on a VAX 11/780. This refinement procedure was applied to several embedded structures, creating a family of DG structures, each of which satisfied the distance constraints, but which exhibited small variations among themselves.

Validity of the Distance Geometry Algorithm. To test the ability of the DG algorithm to determine a known DNA structure, given the interproton distances for that structure, the following test was carried out. All proton–proton distances within 5 Å in an idealized B-form DNA (Arnott & Hukins, 1972) having the sequence $[d(CGCGAATTCGCG)]_2$ were entered into a hypothetical distance matrix; in the first test, exact interproton distances, without error bounds, were used. This distance matrix was then subjected to smoothing, embedding, and refinement using the DSPACE algorithm; the refinement included symmetry constraints, as described earlier (distance geometry algorithm).

The resulting structure generated by the DSPACE algorithm is shown in Figure 1 (top) superimposed upon the actual structure from which the distances were obtained; the agreement is excellent (RMSD of 0.51 Å per residue). Thus, the algorithm is indeed capable of recreating the correct structure when given all exact interproton distances below 5 Å.

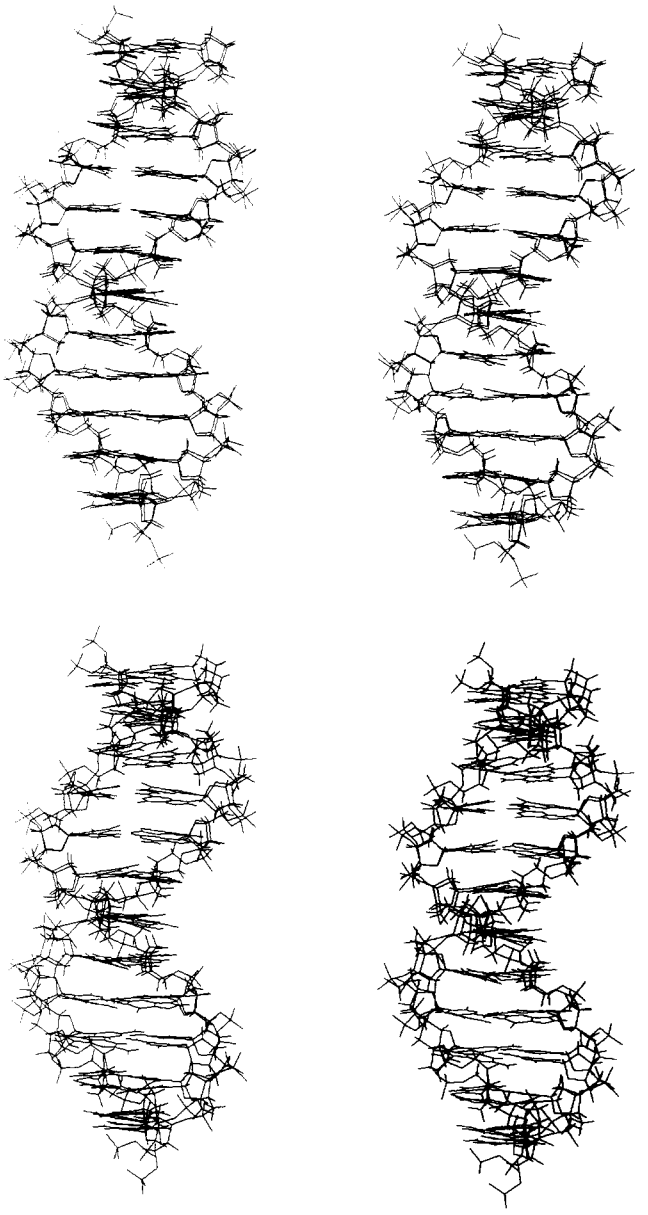

FIGURE 1: Stereoviews of the *Eco*RI dodecamer structure calculated from the proton–proton distances in idealized B-DNA geometry, superimposed on the idealized structure from which the distances were taken. (Top) Exact proton–proton distances for all proton–proton pairs within 5 Å of each other were used in the distance matrix. (Bottom) Only the 155 proton–proton distances corresponding to those measured in the experimental data set were used in the distance matrix and were given the uncertainty bounds of the experimental measurements. The RMSD is 0.51 Å for the top pair and 1.38 Å for the bottom pair.

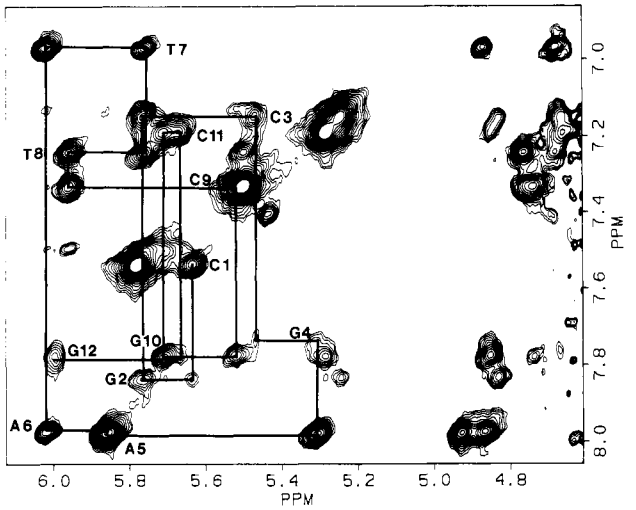

FIGURE 2: Contour plot of the H8/H6–1′H/3′H region in the NOESY spectrum of the $[d(CGCGAATTCGCG)]_2$ dodecamer (200-ms mixing time). The intraresidue H8/H6–1′H crosspeaks are labeled with the residue number. The sequential connectivity, $(n + 1)$-H8/H6–(n)1′H–(n)H8/H6–$(n - 1)$1′H, etc., is connected by a solid line.

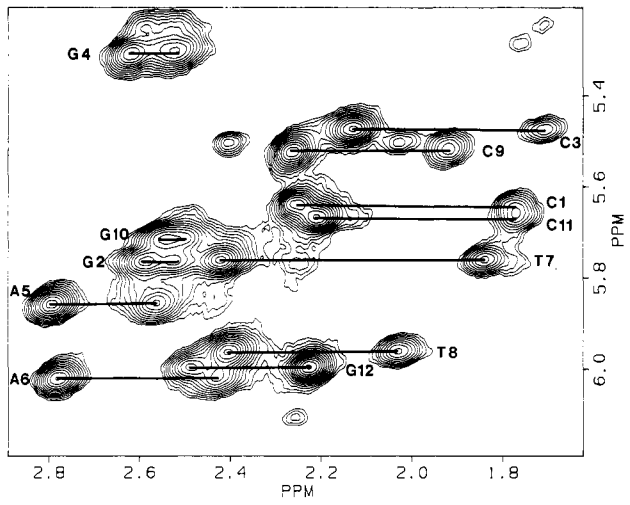

FIGURE 3: Contour plot of the sugar 1′H–2′H,2″H region in the NOESY spectrum of the $[d(CGCGAATTCGCG)]_2$ dodecamer (200-ms mixing time). The residues are labeled at the position of their 1′H chemical shift. For each residue, a horizontal bar connects the 1′H–2″H crosspeak (downfield) to the 1′H–2′H crosspeak (upfield).

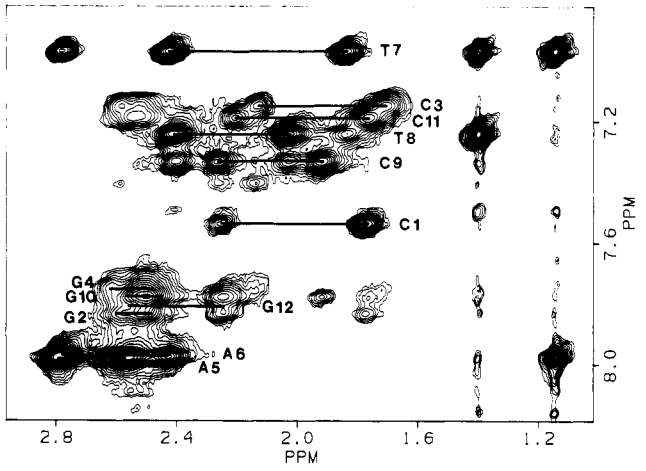

FIGURE 4: Contour plot of the H8/H6–2′H,2″H region in the NOESY spectrum of $[d(CGCGAATTCGCG)]_2$ (200-ms mixing time). The residues are labeled at the position of the H8/H6 chemical shift.

In reality one can never experimentally measure *all* the interproton distances shorter than 5 Å, and furthermore the experimentally determined subset of distances contains experimental uncertainties. A second, more realistic, test of the capability of the DSPACE algorithm to produce the correct DNA structure was also carried out by using *only* the 155 proton–proton distances that were actually measured experimentally in the NOESY spectrum. Furthermore, the same uncertainty (upper and lower bounds) used in the experimental data was applied to each distance. This less-determined distance matrix was also subjected to smoothing, embedding, and refinement (including symmetry constraints) by the DSPACE algorithm. The results are shown in Figure 1 (bottom). The structure generated by the algorithm is now a less perfect representation of the actual structure from which the distances were obtained (RMSD of 1.38 Å per residue) but is nevertheless a good approximation of the original structure and does not contain any gross systematic distortions. This test experiment demonstrates conclusively that the algorithm generates reasonable approximations of the true structure when supplied with only a subset of experimentally amenable, but correct, distances containing typical experimental uncertainty; it does not shed any light on whether the experimentally measured distances are in fact correct.

Calculating the NOESY Spectra. The method for back-calculating the NOESY spectra from any of the generated distance geometry (DG) structures using the BKCALC algorithm and further refinement of the DG structure by changing incorrect proton–proton distances until the calculated NOESY spectrum matches the experimental data, have been previously described in some detail by Banks et al. (1989). These calculations were carried out on a Silicon Graphics IRIS 4D/20 personal computer.

RESULTS

Assignment of the nonexchangeable protons in the *Eco*RI restriction sequence $[d(CGCGAATTCGCG)]_2$ DNA has been carried out previously (Hare et al., 1983). Contour plots of the base proton–1′H/3′H region, the 1′H–2′H,2″H region, and the base proton–2′H,2″H region in the 200-ms NOESY spectrum of the present data set are shown in Figures 2–4, respectively. The present data set was collected at a higher temperature than the data set reported earlier (Hare et al., 1983), and these contour plots serve to relate the two data sets and to aid in the identification of the crosspeaks shown in subsequent figures. Stacked plots of these three crosspeak regions are shown in Figures 5 and 7–9.

The NOESY crosspeaks were integrated as described under Materials and Methods, and distances were calculated by measuring the initial growth rates of the various crosspeaks and scaling them to the growth rate of the appropriate reference distance crosspeak. All crosspeaks involving methyl resonances were scaled to the thymine methyl–H6 crosspeak growth rate, and imino proton distances were scaled to the growth rate of the thymine imino–adenine H2 NOE peak. All other nonexchangeable crosspeak growth rates were scaled to the cytosine H5–H6 distance of 2.45 Å. The crosspeak buildup as a function of mixing time for three different regions of the NOESY spectrum is shown in Figure 5 from 30 to 120 ms. The experimental data for the same crosspeak regions at 200 ms is shown at the top of Figure 7. It is important to point out that although several crosspeaks partially overlap at time points when they are close to maximum intensity (120–200 ms), many of these same crosspeaks can be individually integrated at earlier time points when their intensities are

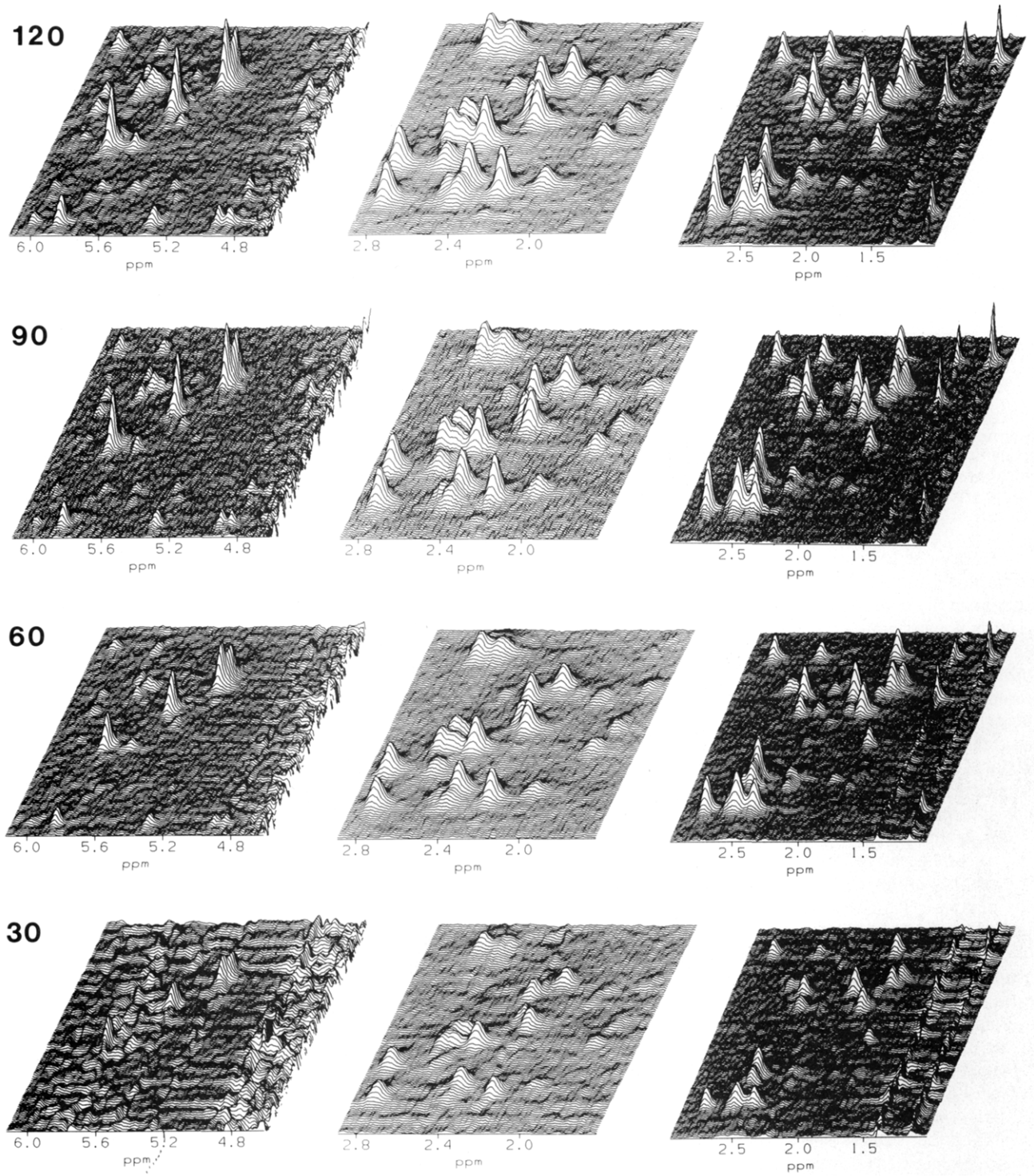

FIGURE 5: Stacked plots of the time dependence of the development of the NOESY spectrum for the three regions shown in Figures 2–4. Increasingly longer mixing times in the NOESY experiment are shown from bottom to top at 30, 60, 90, and 120 ms. The H8/H6–1′H/3′H region is on the left; the 1′H–2′H,2″H region is in the middle; the H8/H6–2′H,2″H region is on the right. Stacked plots of the same regions at 200-ms mixing time are shown at the top of Figure 7.

smaller, and initial rates can still be determined for such peaks. The crosspeak volumes were then plotted as a function of mixing time for each resolved NOESY crosspeak. Figure 6 displays the time dependence of the crosspeak volumes for four proton pairs with interproton distance separations ranging from 2.3 to 3.7 Å (determined from the initial slope). The T7–H6/T7–2′H crosspeak corresponding to a distance of 2.3 Å is obviously in the nonlinear regime by a mixing time of 90–120 ms, while the T7–H6/T7–1′H crosspeak corresponding to a distance of 3.7 Å does not appear above the noise to a sufficient level to be reliably integrated until mixing times of 90–120 ms. In the latter case the crosspeak continues to be in the linear buildup regime out to 200 ms. When the effects of spin diffusion become severe, as at 500 ms, the crosspeak volume is no longer a true reflection of the interproton distance separation for that proton pair (Kalk & Berendsen, 1976), and by 500–1200 ms all peaks tend to achieve approximately equal volume, regardless of the proton–proton distance separation they represent, due to the effects of spin diffusion and magnetization leakage.

All the resulting distances calculated from the initial rate two-spin approximation, as well as their uncertainty estimates,

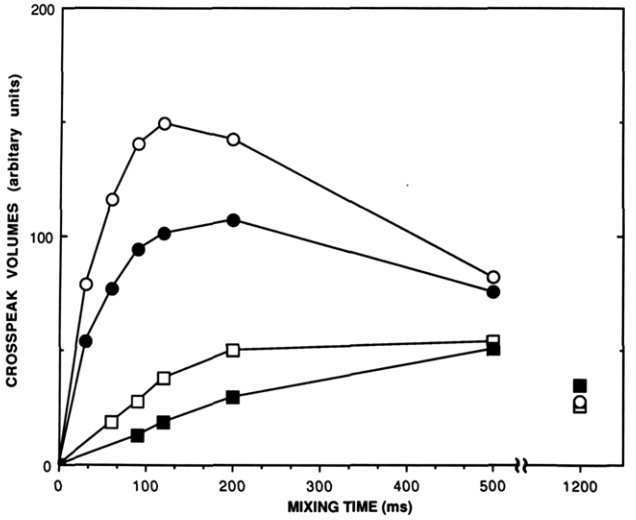

FIGURE 6: Time dependence of the intensity (volume integrals) for four representative NOESY crosspeaks: (open circles) T7–H6/T7–2′H; (filled circles) T7–H6/A6–2″H; (open squares) T7–H6/A6–1′H; (filled squares) T7–H6/T7–1′H. Peak volumes were integrated as described under Materials and Methods. See the text for details.

were next entered into the upper and lower bounds of the distance matrix. The distance matrix was then smoothed, symmetrized, embedded in 3-space, and refined as described under Materials and Methods. Embedding several independently symmetrized matrices created a family of unique unbiased starting structures for subsequent refinement. Each embedded starting structure was then refined against the distance matrix to minimize deviations.

The refined structures generated with distances calculated from the two-spin approximation were all very similar; comparing pairs of structures resulted in root-mean-square coordinate deviations (designated RMSD) of ca. 2.0 Å per residue. At the same time all the refined structures had low distance deviations from the distance matrix (designated DM). A ball and stick model stereoview of a refined DG structure is shown in Figure 10. The structures differ significantly from "normal DNA", and we initially suspected that this might be due either to the algorithm itself or to inherent underdetermination in the distance data set. However, the structures are all quite similar among themselves, arguing against underdetermination; furthermore, when only the distances corresponding to the experimentally measured distances are taken directly from idealized B-DNA fiber coordinates (together with the experimental upper and lower bounds values), the resulting distance matrix embeds and refines to regenerate only the starting structure from which the distances were taken, thus validating the algorithm (see Materials and Methods). This

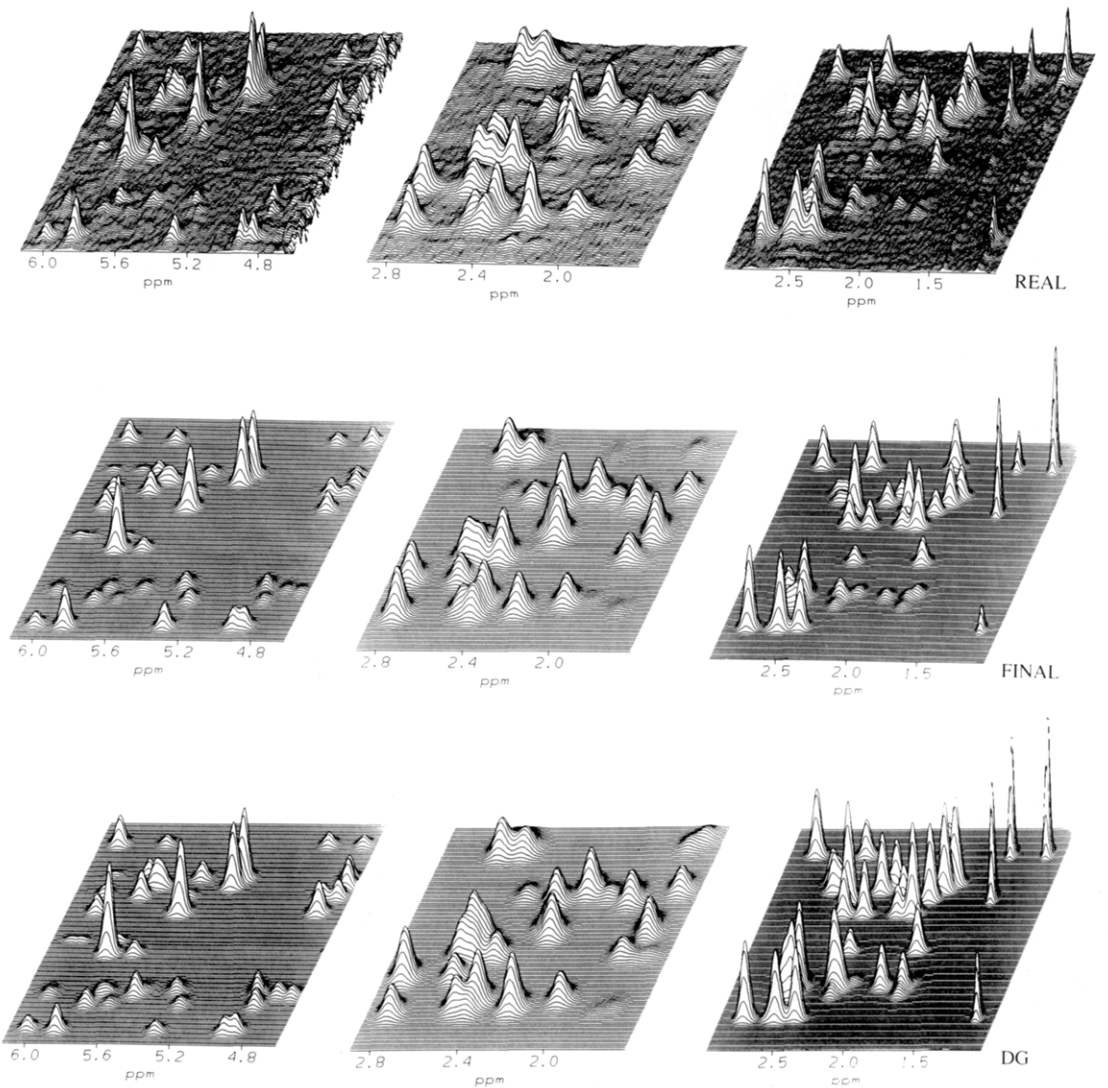

FIGURE 7: Stacked plots of the NOESY spectra at 200-ms mixing time for (top, middle, and bottom) real (experimental), final refined structure, and DG (two-spin approximated) structure. The three regions shown are the same as those of Figures 2–4: H8/H6–1′H, 3′H (left), 1′H–2′H,2″H (middle), and H8/H6–2′H,2″H (right).

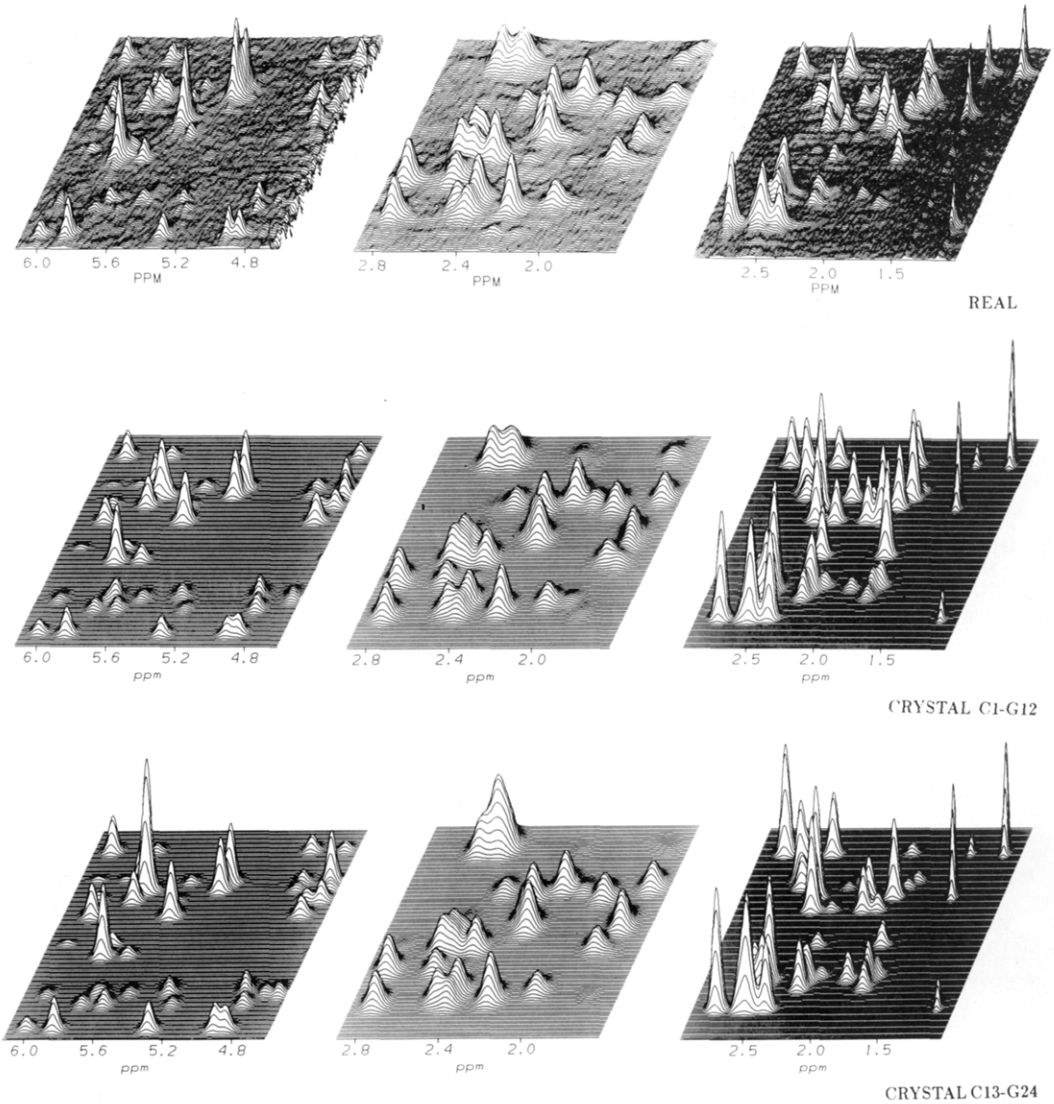

FIGURE 8: Stacked plots of NOESY spectra at 200-ms mixing time for (top, middle, and bottom) real (experimental), crystal C1–G12, strand one in the crystal structure, and crystal C13–G24, strand two in the crystal structure. The three regions shown are the same as those of Figures 2–4: H8/H6–1′H (left), 1′H–2′H,2″H (middle), and H8/H6–2′H,2″H (right).

strongly suggests that the "abnormal" structures generated from the experimental distances are in fact valid or at least a good representation of the experimental interproton distance estimates.

Reassured that the algorithm was not leading us into wrong structures and that the number of experimentally measured distances was sufficient to constrain the structure into a relatively small region of conformational space, we next calculated the NOESY spectrum for one of the DG structures generated from experimental distances. In calculating the theoretical spectrum, each proton of a proton pair corresponding to a crosspeak in the NOESY spectrum is surrounded by a sphere of 5-Å radius, and all interproton distances within these two spheres are then measured and stored. This process is repeated for each proton pair in the NOESY spectrum. The collection of subsystems is then allowed to cross-relax in the presence of magnetization leakage to the lattice. The leakage rate is determined experimentally as the difference between magnetization lost from the autopeak and magnetization appearing in the crosspeaks, and the observable magnetization at all spins is calculated numerically according to r^{-6} and updated every 0.1 ms, as described by Banks et al. (1989). While the choice of 5-Å spheres appears arbitrary, we find empirically that when the sphere surrounding the cross-relaxing proton pairs is increased from 5 to 10 Å, a difference of only 1–2% is observed in the crosspeak intensities out to 200 ms.

The accuracy of the DG structure was then evaluated by comparing the calculated crosspeak intensities for the DG structure to those found in the experimental data (Figure 7, bottom and top). For the large majority of crosspeaks, the agreement with experiment is quite good, indicating that in most respects the DG structure is a reasonable approximation of the actual structure. However, some of the crosspeaks in the calculated NOESY do not match the experimental data; in many cases these correspond to distances that could not be measured experimentally due to spectral overlap—see, for example, the base H8/H6–2′H, 2″H region in Figure 7 (right) and in Figure 4. The crosspeaks between the H8 protons of G2, G4, G10, and G12 and their own 2′,2″ protons as well as the 2′,2″ protons of the preceding residue amount to a total of 16 possible distance constraints. However, the spectral overlap in this region permits only 5 of these 16 crosspeaks to be measured reliably. This part of the calculated NOESY spectrum (Figure 7, bottom right) has several crosspeaks that are more intense than in the experimental spectrum; this is largely caused by the lack of measured input distance constraints in this overlapped region. Furthermore, many of the resolved H8/H6–2′H/2″H proton pairs tended to cluster near

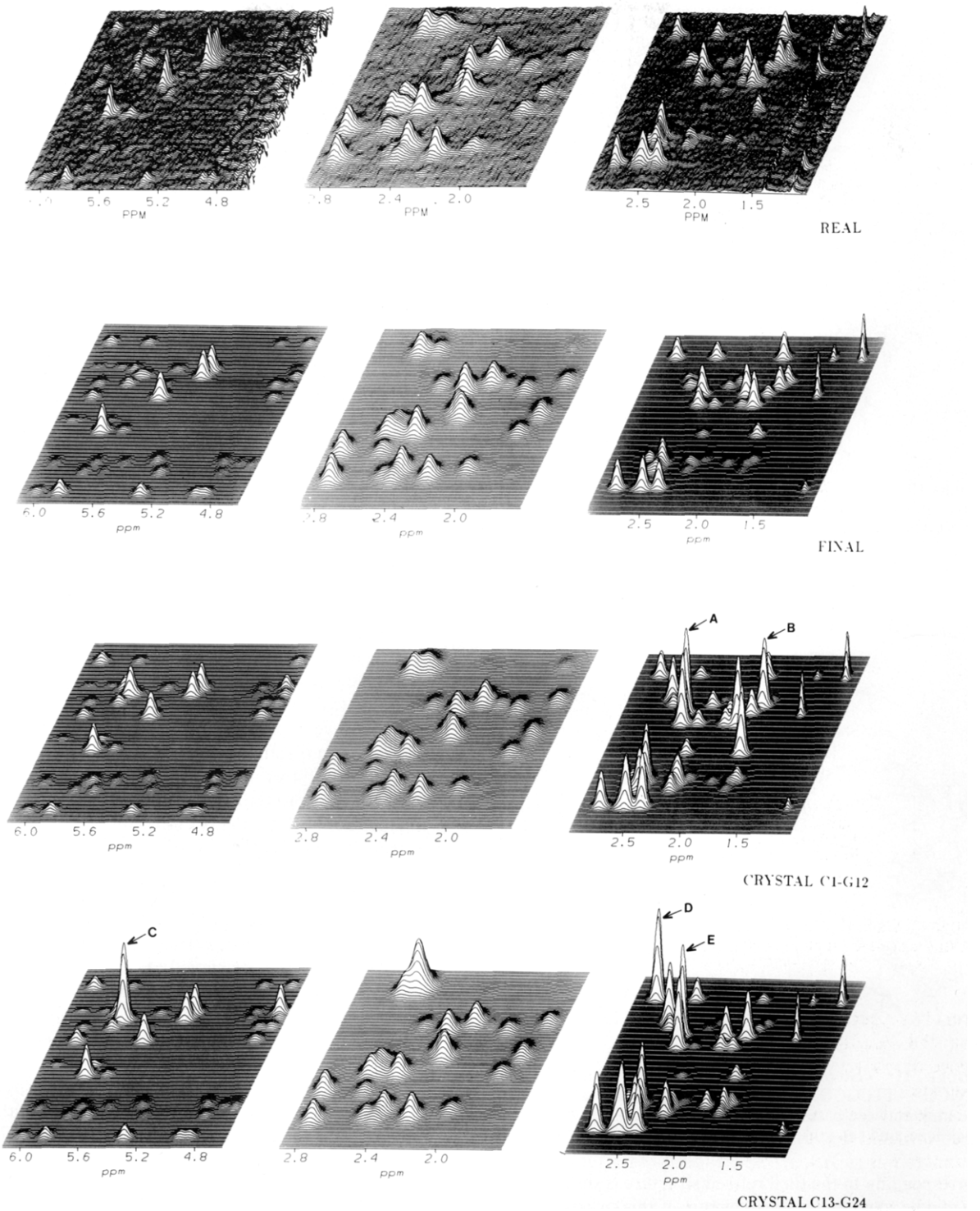

FIGURE 9: Stacked plots of the NOESY spectra at 60-ms mixing time for (from the top down) real (experimental), final refined structure, crystal C1–G12, strand one in the crystal structure, and crystal C13–G24, strand two in the crystal structure. The three regions shown are the same as those of Figures 2–4: H8/H6–1′H,3′H (left), 1′H–2′H,2″H (middle), and H8/H6–2′H,2″H (right). Arrows A–E indicate five of the crosspeaks in the calculated NOESY spectra from the crystal structure that deviate markedly from the real (experimental) data. Arrow A indicates the overlapping T8–H6/T8–2″H and T8–H6/T7–2″H crosspeak (strand one), and arrow E indicates the corresponding overlapping T20–H6/T20–2″H and T20–H6/T19–2″H crosspeak (strand two). Arrow B points to the H6/2′H crosspeak in residue C11. Arrow C points to the overlapping C23–H6/1′H and C23–H6/G22–1′H crosspeak. Arrow D points to the T19–H6/A18–2″H crosspeak. See the text for details.

the lower bounds of the distance matrix. For example, the T7–H6/A6–2″H rate shown in Figure 6 was given an initial distance constraint of 2.3–2.7 Å. However, in the DG structure shown in Figure 10 and simulated in Figure 7, this distance was found to be 2.24 Å, i.e., just outside the lower bound. Further refinement was next carried out by adjusting the offending distances in the bounds matrix of the particular DG structure under consideration—typically by 0.1–0.5 Å. The DG structure was then rerefined against this modified bounds matrix by using the DSPACE algorithm, and the NOESY spectrum of the resulting structure was again back-calculated from the new coordinates, as described previously (Banks et al., 1989). This procedure was repeated several times to finally produce a structure that generated a calculated NOESY spectrum in good agreement with the real data at all mixing times. The NOESY spectrum at 200 ms

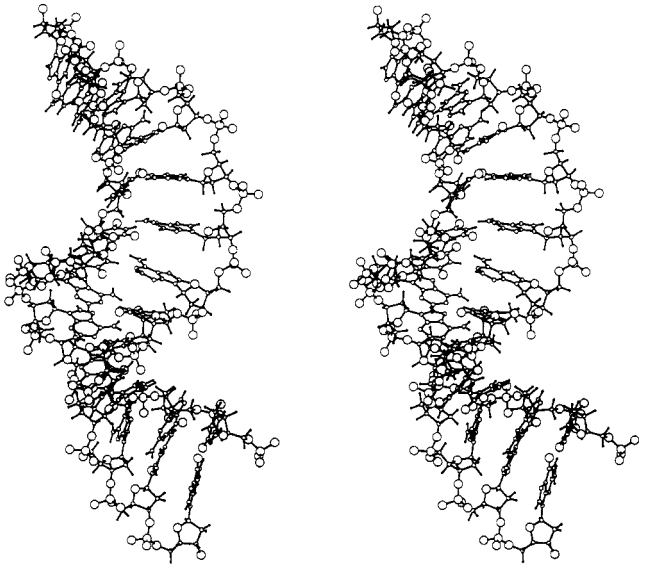

FIGURE 10: Stereoview model of a distance geometry (DG) structure after refinement against the distance matrix, but before refinement by NOESY back-calculation.

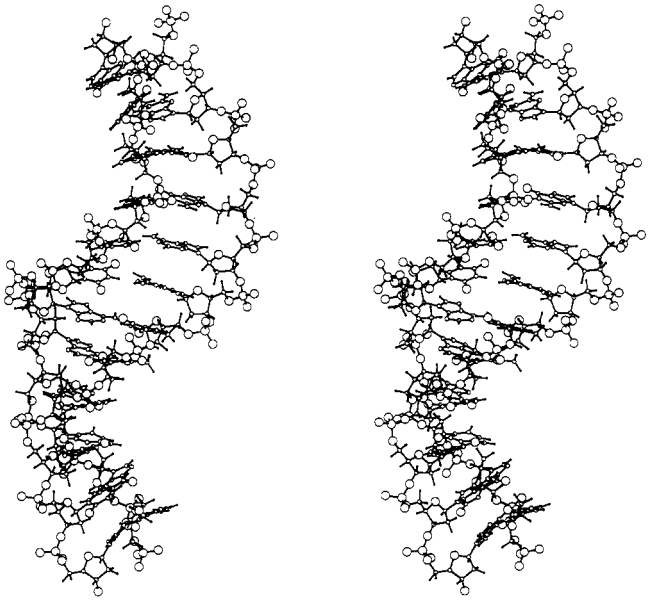

FIGURE 11: Stereoview of the final structure of the [d-(CGCGAATTCGCG)]$_2$ dodecamer refined by back-calculation. Note the kink between bases A6 and T7, which can best be seen in the backbone, and the kink between bases C3 and G4.

corresponding to the final refined structure is shown in Figure 7 (middle row), and a stereo picture of this structure is shown in Figure 11. The extent to which the back-calculation refinement changed the DG structure can be seen in Figure 12, where the DG and final structures are superimposed. The important point to be made is that, unlike the DG structure, the final structure now satisfies the time dependence of all the local multiple-spin relaxation networks and explicitly incorporates spin diffusion. The corollary of this is that the DG structure, which does not, must be satisfying some incorrect distances in the original bounds matrix; i.e., some of the initial rate two-spin approximated distances must be wrong due to second-order spin diffusion effects even at 30 ms. Separate test experiments on averaged DNA structures in this laboratory indicate that base–base and base–3′H are the most seriously underestimated distances from an initial rate two-spin approximation (K. M. Banks, personal communication). It

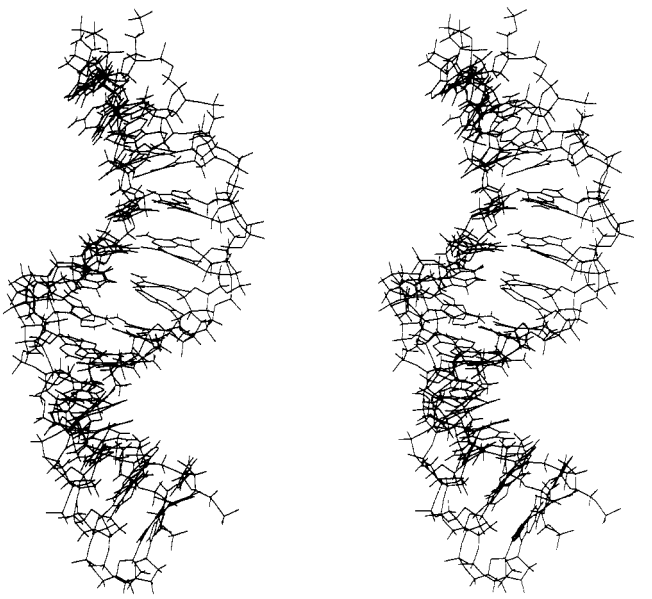

FIGURE 12: Stereoview of the distance geometry (DG) structure shown in Figure 10 superimposed on the final structure shown in Figure 11. The average RMS deviation between the two structures is 2.596 Å per residue.

is worth noting that the T7–H6/T7–2′H and T7–H6/A6–2″H distances (initially measured as 2.1–2.5 and 2.3–2.7 Å by the two-spin approximation) changed from values of 2.54 and 2.24 Å in the DG structure to 2.32 and 2.44 Å in the final structure; i.e., the two-spin estimates of 2.3 ± 0.2 and 2.5 ± 0.2 Å were not only very precise but also accurate for these types of interproton distances.

At this point we naturally wondered how similar the solution structure was to the published crystal structure of this dodecamer duplex, and the back-calculation algorithm gave us a means to address this question. Protons were added to the X-ray coordinates, and the NOESY spectra at 200 ms corresponding to the two strands in the crystal structure are shown in Figure 8, with the first strand 1–12 in the middle and the second strand 13–24 at the bottom. The actual experimental spectrum (real) is displayed at the top of Figure 8 for ease of comparison. As was the case for the DG structure, the H8/H6–2′H/2″H region (right column) shows the greatest deviations from the experimental NOESY spectrum. Furthermore, the back-calculated NOESY spectra for the two individual strands of the crystal structure also show quite large variations between themselves. The fact that the crystal structure is not symmetrical, whereas the NMR resonances of G2 and G14, C3 and G15, etc., are exactly superimposed, indicating perfect symmetry in solution, already precludes the crystal structure being the same as the solution structure. The mixing time of 200 ms is somewhat late to qualify as an initial rate, and to more closely compare proton–proton distances, the NOESY spectra at 60 ms are also shown in Figure 9, together with the final and real data. At 60 ms the effects of spin diffusion are much less severe than at 200 ms, and the crosspeak intensities more accurately reflect the corresponding proton–proton distances. The final structure generates a NOESY spectrum in good agreement with the experimental NOESY spectrum at 60 ms, but the two crystal structure strands generate NOESY spectra with marked deviations from the experimental spectrum. The arrows A–E in Figure 9 indicate some of the more aberrant crosspeaks that differ from the experimental data. Peak A contains the two overlapping crosspeaks of proton pairs T8–H6/T8–2″H and T8–H6/T7–2″H in strand one, while peak E is the corresponding crosspeak

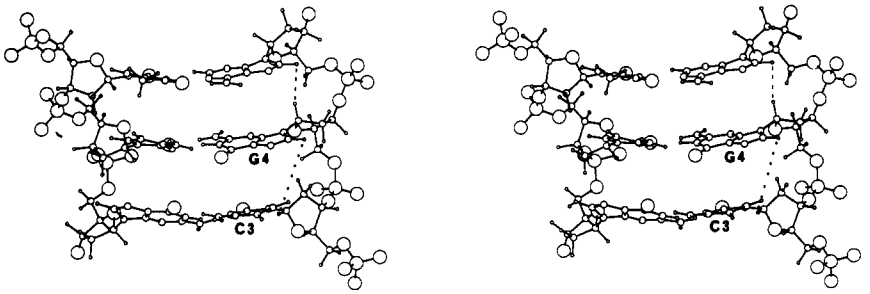

FIGURE 13: Expanded stereoview of base pairs 3, 4, and 5. Note the difference in sugar ring orientation of residues G4 and C3. The dashed line is drawn between the G4–1′H and A5–H8 protons, the shortest H8/H6(n)–1′H(n – 1) distance (2.9 Å) in the molecule; the corresponding crosspeak is located in the lower central region of the H8/H6–1′H region in Figure 9. The dotted line is drawn between the G4–H8 and C3–1′H protons, the longest H8/H6(n)–1′H(n – 1) distance (5.67 Å) in the duplex; the corresponding crosspeak is located in the lower central region of Figure 2. See the text for details.

in strand two of the crystal structure. In the final refined structure the two proton–proton distances T8–H6/T8–2″H and T8–H6/T7–2″H are 3.70 and 2.40 Å, respectively, whereas they are 4.05 and 1.84 Å in strand one and 3.72 and 1.84 Å in strand two in the crystal structure. Peak B is the single C11–H6/C11–2′H crosspeak in strand one in the crystal structure. The C11–H6/C11–2′H distance is 2.43 Å in the final structure and 1.91 Å in strand one of the crystal structure, whereas the corresponding C23–H6/C23–2′H distance in strand two of the crystal is 3.26 Å, producing a rather weak crosspeak in the bottom simulation. Peak C is the crosspeak containing the two overlapping C23–H6/C23–1′H and C23–H6/G22–1′H crosspeaks in strand two of the crystal structure. These two proton–proton distances in the symmetric final structure are 3.59 and 3.37 Å, respectively, and 3.68 and 2.03 Å in strand two in the crystal structure. The corresponding C11–H6/C11–1′H and C11–H6/G10–1′H distances in strand one in the crystal structure are 3.75 and 2.38 Å, respectively. Peak D is the single T19–H6/A18–2″H crosspeak in strand two of the crystal structure. This proton–proton distance in the symmetric final structure is 2.44 Å but is 1.84 Å in strand two of the crystal structure. The corresponding T7–H6/A6–2″H in strand one in the crystal structure is 2.35 Å.

Examination of the final refined structure in Figure 11 reveals a number of bends and kinks in the sequence. It appears that there is a kink between residues A6 and T7 involving slide and some opening into the major groove. Between base pair C3–G10 and base pair G4–C9 there is also a kink opening into the minor groove caused by roll between these base pairs; an expanded view of base pairs 3–5 is shown in Figure 13. Figure 14 shows stereoviews of the stacking overlap in the six base-pair steps, C1–G2, G2–C3, C3–G4, G4–A5, A5–A6, and A6–T7, in the symmetric final structure, and Table I lists the sugar pseudorotation angles and sugar conformations, as well as the glycosyl torsion angles for all but the less accurately determined terminal residues C1 and G12. The helical parameters of the refined duplex were calculated with the routine NEWHELIX (kindly provided by R. E. Dickerson) using the Cambridge Convention (1989) and are presented in Table II.

Table I: Glycosyl and Sugar Torsion Angles for the Nonterminal Residues in the Final Structure of the [d(CGCGAATTCGCG)]$_2$ Dodecamer[a]

residue	glycosyl χ (deg)	pseudorotation angle P (deg)	sugar conformation
G2	307	251	$^{4}_{0}$T
C3	249	93	^{0}E
G4	318	135	$_{1}$E–$^{2}_{1}$T
A5	268	106	$^{0}_{1}$T
A6	241	112	$^{0}_{1}$T
T7	229	139	$^{2}_{1}$T
T8	257	144	$^{2}_{1}$T
C9	283	199	$_{3}$E
G10	281	91	^{0}E
C11	219	103	$^{0}_{1}$T

[a] The glycosyl torsion angles are listed according to the Cambridge Convention (1989).

Table II. Helical Parameters for the Final Structure[a]

1 12
C-G-C-G-A-A-T-T-C-G-C-G
G-C-G-C-T-T-A-A-G-C-G-C
12 1

base pairs	twist (deg)	slide (Å)	roll (deg)	tilt (deg)
G2–C11/C3–G10	14.6	−0.79	29.2	−3.7
C3–G10/G4–C9	5.8	1.15	13.5	−1.5
G4–C9/A5–T8	23.3	−0.45	2.4	0.6
A5–T8/A6–T7	26.7	−2.12	6.7	0.3
A6–T7/T7–A6	28.1	−3.01	−9.1	0.0

base pair	propeller twist (deg)
G2–C11	−25.0
C3–G10	−20.1
G4–C9	−10.3
A5–T8	−11.5
A6–T7	−31.1

[a] The terminal bases C1 and G12 are not included due to terminal fraying effects and few NMR-determined distances. Helical parameters are listed according to the Cambridge Convention (1989) and calculated by the routine NEWHELIX (kindly provided by R. E. Dickerson). Twist and slide are calculated from the C1′–C1′ vectors. Roll, propeller twist, and tilt are calculated from the best plane through both bases.

Discussion

In the present study of the *Eco*RI dodecamer duplex we have shown that distance geometry combined with NMR-derived distances generates a closely related family of structures for this DNA sequence in solution, with only minor variations between structures. Compared to our previous study of a Pribnow promoter DNA sequence (Nerdal et al., 1988), the present study yielded a larger number of distance constraints (12–13 per residue) and did not suffer the larger variation in distance geometry structures due to underdetermination in the earlier study.

Despite the ability to constrain the family of distance geometry structures to a much more restricted region of conformational space, the convergence of these structures is merely a measure of the extent of determination in the number of distances and does not constitute proof that the correct structure lies within this narrow range of structures. The distance geometry structures simply satisfy the experimentally

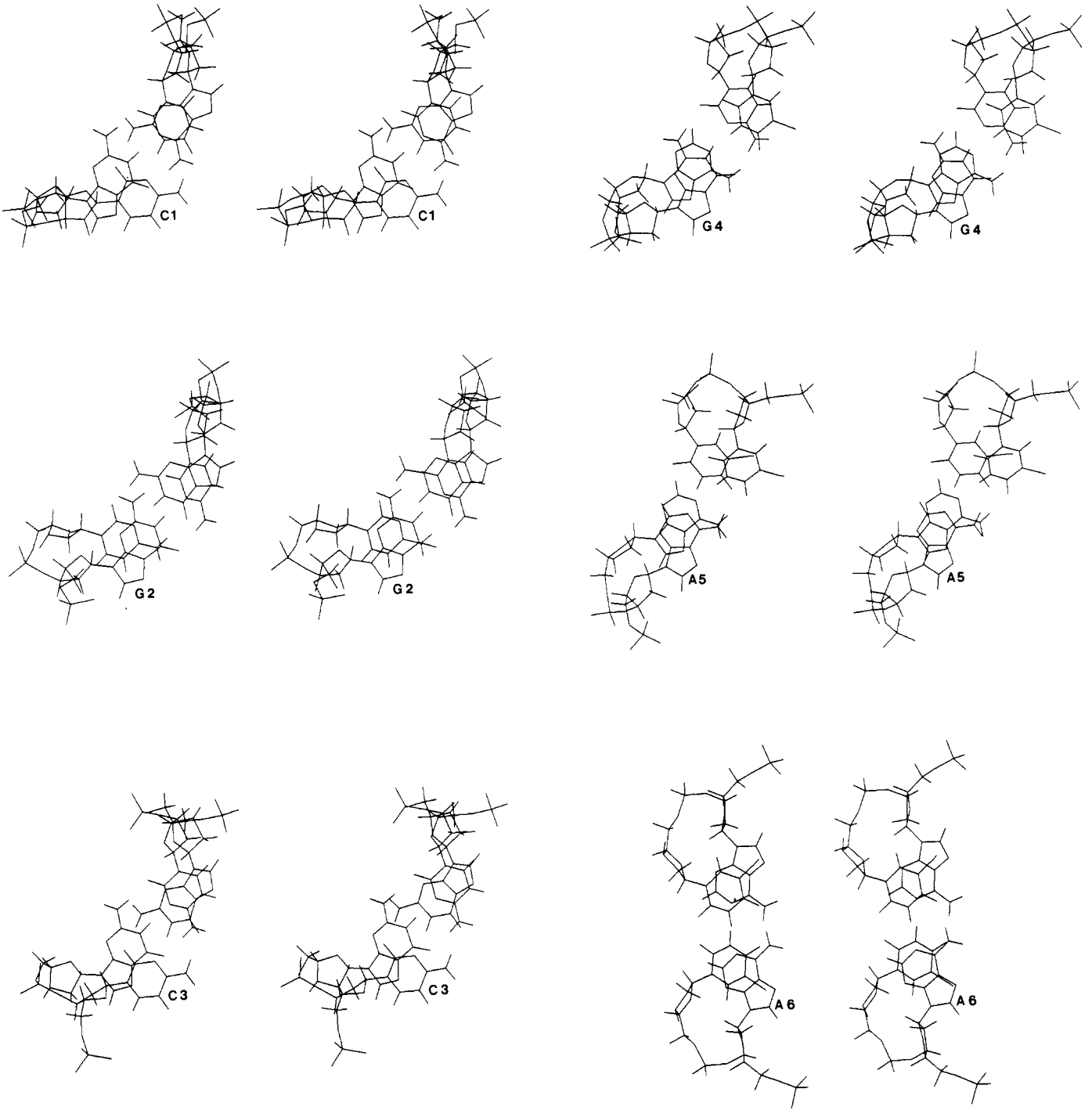

FIGURE 14: Stereoviews of the individual base-pair stacking overlap in the symmetric final refined $[d(CGCGAATTCGCG)]_2$ dodecamer, starting with the C1–G2 step at the upper left corner. Viewed "wide eyed" the C1 residue appears on top of the G2 residue, in the G2–C3 step the G2 residue appears on top of the C3 residue, etc. Note the broadening of the major groove at the A5–A6 step and at the A6–T7 step.

measured distance bounds, and it is important to remember that these distances are calculated from a two-spin approximation. Although the error in the two-spin-approximated distance estimates are progressively reduced at shorter mixing times, for reasons of spectrometer sensitivity it is not experimentally feasible to reduce the mixing time below 20–30 ms yet still integrate the very small off-diagonal peaks reliably at sample concentrations of 5–10 mM. Thus, there is no guarantee that all of the 30-ms "initial rate" distance estimates are free of second-order spin-diffusion contributions that lead to incorrect, underestimated distances. If some of the distances being satisfied are incorrect, the resulting structure obviously cannot be correct; this objection applies equally well to distance geometry structures and to structures generated by molecular dynamics refinement of classical DNA structures using NMR-derived distance constraints.

Back-calculation of the NOESY spectrum constitutes a much more rigorous test of whether the distance geometry structure, or any other proposed structure, is the actual structure in solution. By numerically integrating all the higher order indirect pathways of magnetization transfer within a given local environment, spin diffusion is converted from an undesirable complication into a useful diagnostic tool. Back-calculation of the NOESY spectrum of any of the individual DG structures reveals that some of the initial rate distance estimates are indeed in error; i.e., as a class the base to 3′H crosspeaks for the DG structures are uniformly too intense compared to the experimental data, indicating that

these distances are underestimated in the two-spin approximation—see Figure 7. These distances are in the 3.5–4-Å range, and their crosspeaks do not appear above the noise level until 60–90 ms (see Figure 5), at which time second-order spin-diffusion pathways have begun to contribute and the true initial rate is no longer being measured. An additional source of deviation from the experimental NOESY data comes from overlapped crosspeaks that are entered into the distance matrix with very wide bounds; even after reduction by triangulation, the bounds range remains moderately wide, and any value within this wide range eliminates the penalty during refinement.

Although the DG structures are a highly converged set of very similar structures, back-calculation reveals that they are not the correct structures but merely a set of consistently wrong structures that satisfy a particular subset of consistently incorrect distance estimates. However, in addition to this important function, back-calculation provides the opportunity to correct the errors in distance estimates caused by spin diffusion. Increasing the distance value in the distance matrix between proton pairs whose crosspeaks are too large in the back-calculation and refining the structure against this new bounds matrix by simulated annealing and conjugate gradient refinement will generate a structure closer to the true structure. Several cycles of this procedure produce the final structure, the coordinates of which reproduce the experimental NOESY spectrum (and its time dependence) almost perfectly and certainly far better than does the DG structure—see Figure 7. We believe that ab initio calculation of a DG structure that satisfies the NMR-derived two-spin distance approximations, followed by back-calculation refinement of this unbiased model (which is incorrect but reasonably close to the real structure), represents the closest one can come to attaining an experimentally derived solution structure for DNA.

At this point it is perhaps worth discussing the strengths and weaknesses of this approach and comparing it to the alternative approach of molecular dynamics restrained by NMR-derived distance measurements (Nilges et al., 1987a,b; Clore et al., 1988). In both methods the experimentally measured distance constraints are derived from the two-spin approximation and suffer from exactly the same spin-diffusion-induced errors in distance estimates. Thus, the RMD structures satisfy some incorrect distances while also satisfying some energy potentials, some of which, e.g., electrostatic repulsion and solvation, are of dubious accuracy. In the absence of back-calculation refinement, DG structures and RMD structures will introduce rather similar kinds of structural errors, and neither can be expected to produce the true solution structure.

A further weakness common to both approaches is the absence of assignments or distance constraints for the 5′ and 5″ protons of the DNA backbone. In the present DG–BKC approach this leads to a wide range of backbone torsion angles α, β, and γ for the unconstrained backbone linkages between residues. The CH_2OPO linker is allowed to bridge adjacent sugars in any of a variety of ways, provided bond lengths, bond angles, and van der Waals distances are not compromised. In contrast, the RMD approach uses an idealized average B-DNA starting model. In the absence of any experimental distance constraints from the backbone, these torsional angles for the backbone linker remain largely unperturbed. While the resulting model is perhaps more esthetic, and is certainly more regular, it is important to realize that there is no experimental evidence to support this backbone conformation; neither is there any reason to believe that it is the actual conformation in solution. Opinion about backbone conformation has gradually changed from a fairly rigid structure about which base stacking and sugar conformation adjust accordingly to a rather flexible structure that adjusts like a shock absorber according to base stacking (Calladine & Drew, 1984). In our refinement the 5′H and 5″H are included in the back-calculation, and thus their indirect effects are incorporated in the simulation, but with significantly less precision than direct distance constraints. The uncertainty concerning the backbone linker conformation is not unique to NMR structure determination; even in single-crystal X-ray structures the backbone linker is susceptible to perturbation by crystal packing forces, and this is the least well determined part of the molecule in crystallographic studies (Dickerson et al., 1988).

Turning to the structural characteristics of the molecule in solution, the refined final structure that reproduces the experimental NOESY spectra is shown in Figures 11 and 13. Among several peculiarities, one of the more noteworthy features is an apparent kink or discontinuity between the third and fourth base pairs from either end. This base-pair step has an extremely low helical twist angle (Figure 14 and Table II) combined with a moderately high positive roll angle that opens the kink into the minor groove (Figure 11 and Table II). This results in a rather poor stack of G4 on C3 (see Figure 14). Figure 13 shows an expanded view of this portion of the molecule, and it is important to emphasize that the somewhat peculiar spatial orientation of the A5 and G4 H8 protons to the 1′H, 2′H, and 2″H of C3, G4, and A5 is directly supported, indeed even required, by the raw experimental data. The rather close proximity of A5–H8 to G4–1′H is indicated by a dashed line in Figure 13, while the much greater distance between G4–H8 and C3–1′H is indicated with a dotted line. Historically, the widely different NOE intensities of these adjacent H8–(n − 1)1′H steps in our earlier qualitative studies on this molecule already indicated sequence-dependent deviations from idealized B-DNA (Hare et al., 1983). The close proximity of A5–H8 to G4–1′H is forces upon the structure by the observation that this is the largest single H8–1′H crosspeak in the entire molecule—see the intense interresidue peak at 5.31 ppm in Figure 2 and also Figures 5, 7, and 9. The growth rate of this resolved single peak corresponds to 3.1–3.2 Å in the two-spin distance estimation and refines to 2.9 Å in the final structure; it is by far the shortest base-to-1′H distance in the entire structure. Conversely, the G4–H8–C3–1′H crosspeak is barely detectable at 200 ms (see Figures 2 and 7) or even at 300 ms (Hare et al., 1983). This distance was scored as >4 Å in the initial distance matrix and refined to a value of 5.67 Å in the final structure. From Figure 13, the 2′H and 2″H of C3 are perfectly placed to serve as an intervening spin conduit for magnetization transfer between G4–H8 and C3–1′H, and the presence of this second-order pathway presumably explains the time required before this crosspeak barely appears out of the noise; spin diffusion through the intervening C3–2′H and –2″H after an appropriate induction delay is the major reason why these protons separated by more than 5 Å even develop a crosspeak above the noise level at all.

The short A5–H8–G4–1′H distance is produced by an atypical G4 sugar orientation (Figure 13) combined with a high glycosyl torsion angle (Table I). This pattern of a long G4–H8–C3–1′H distance followed by a short A5–H8–G4–1′H distance may explain the highly underwound low twist angle at the C3–G4 step (see Figure 14 and Table II). While it is difficult to sort out cause and effect, the net result of this combination of deviations from classical B-DNA geometry is

a kink at the 3–4 step, caused by underwinding plus a modest amount of positive roll and a small amount of positive slide that opens to the minor groove.

The second notable feature of the refined structure in Figure 11 is the relatively broad minor groove at the peripheral GC regions of the duplex and the marked narrowing of the minor groove at the central four AT pairs, together with another kink at the A6–T7 step on the twofold axis. This second kink is caused by a high propeller twist (negative propeller twist in the new Cambridge convention) in the two central base pairs and a reasonably high negative roll value combined with negative helical slide (see Figures 11 and 14 and Table II); the slide is also obvious from the cross-strand NOE from A6–H2 to the 1′H of T20 (T8)—see the unlabeled T8–1′H crosspeak at 5.95 ppm in Figure 2. The net result of this negative roll kink is an opening toward the minor groove at the center of the molecule. Table II lists the twist, slide, roll, tilt, and propeller twist values for the various base-pair steps determined with the program NEWHELIX (kindly provided by R. E. Dickerson); only the internal five steps of the symmetrical duplex are listed since the terminal base pairs are subject to end-fraying artifacts and are not well constrained by experimental distances.

From a broader, less-detailed perspective, the morphological features of the duplex are dominated by the two kinks discussed above, and an overall bending of the helix axis. Our observation of two types of kink in the duplex is qualitatively similar to the observation of kinks at the 3–4 and 6–7 steps observed in the intermediate resolution single-crystal X-ray structure of this same duplex complexed to its cognate *Eco*RI endonuclease (McClarin et al., 1986). The 6–7 kink has been designated Neo-kink I and the 3–4 kink Neo-kink II by McClarin et al. (1986) with the prefix neo implying that the kinking is induced by the nuclease protein. The assumption that the enzyme caused these kinks is due to the fact that initially they were not seen in the crystal structure of this DNA sequence alone (Wing et al., 1980; Dickerson & Drew, 1981). The ability of crystallization solvents to induce structural distortion or rearrangement in DNA is well-known (Saenger et al., 1986). In fact, a DNA hairpin that we showed to have a B-DNA stem with a stacked loop in solution (Hare & Reid, 1986) was later found to contain a Z-DNA stem and an unstacked loop when crystallized from 40% 2-propanol (Chattopadhyaya et al., 1988). In light of these findings, we think it equally likely that the kinks are not induced by the nuclease but preexist in solution and are perturbed or attenuated by the solvent during crystallization of the naked DNA. In fact, reanalysis of DNA crystal structure reveals minor structural variations, albeit of much lesser extent, at the same 3–4 and 6–7 steps that have been termed "proto neo-kinks" (R. E. Dickerson, personal communication). Our current hypothesis is that these kinks preexist in the DNA in solution as a necessary consequence of this particular sequence and are in fact the structural basis of sequence-dependent DNA recognition. As part of this hypothesis we propose that, in the X-ray studies of the complex, the "natural" structure of the DNA was initially recognized and bound in true aqueous solution and the tight association with the protein preserved this structure and prevented the distortion by subsequent addition of crystallization agents. At the moment this remains a hypothesis based on qualitative similarities; we await the release of coordinates for the kinked DNA in the nuclease complex for more quantitative comparisons. We note in passing that, although without a firm structural interpretation, the presence of backbone discontinuities in aqueous solution at the 3–4 and 6–7 steps in this same DNA sequence was observed by ^{31}P NMR several years ago (Ott & Eckstein, 1985).

In addition to the kinks, the other gross morphological feature of our refined structure is an overall bending of the duplex, which is caused mostly by the 3–4 kinks. DNA bending, as monitored by retardation on polyacrylamide gel electrophoresis, has been documented for A_3T_3 tracts (Hagerman, 1984) and for A_5T tracts (Wu & Crothers, 1984), but GA_2T_2C sequences have not formally been a member of the growing class of bent DNAs. While attempting to prepare oligo-d(GCGAATTCGC)$_n$ duplexes for analysis, we noted that repeating decamers containing the GAATTC sequence were recently analyzed by polyacrylamide gel electrophoresis (Diekmann & McLaughlin, 1988) and were indeed found to be "bent", i.e., retarded. The extent of retardation was found to be approximately halfway between that of A_3T_3 and "normal" DNA. Thus, there is independent evidence and precedent that the structure of the GAATTC *Eco*RI restriction sequence is in fact somewhat bent. As a note of caution, however, we emphasize that, at the current level of development of NMR-derived solution structure determination of DNA, the structure in Figure 11 should not be used as a reference standard to quantitatively evaluate the precise bending angle for the overall duplex. The NMR distance method extends only to ca. 4.5 Å, and while we have confidence in the details of local geometry, the precise orientation of CG3 with respect to AT6 is necessarily third-hand and subject to some level of error propagation. Pardi et al. (1988) have recently published an insightful simulation analysis of the extent to which the various DNA helical parameters can be accurately determined by distance geometry alone, in the absence of back-calculation; helical rise, propeller twist, and dislocation are among the inherently least well-determined parameters in this test experiment.

ACKNOWLEDGMENTS

W.N. gratefully acknowledges the Royal Norwegian Council for Scientific and Industrial Research (NTNF), Oslo, Norway, for a fellowship. D.R.H. gratefully acknowledges support from NIH SBIR Grant GM 35620, which provided partial funding of the computer equipment utilized for most of the distance geometry calculations in this project. We also thank N. Susan Ribeiro for synthesizing the DNA dodecamer and Mary Coventry for her assistance in preparation of the manuscript.

Registry No. d(CGCGAATTCGCG), 77889-82-8; restriction endonuclease *Eco*RI, 80498-17-5.

REFERENCES

Arnott, S., & Hukins, D. W. (1972) *Biochem. Biophys. Res. Commun. 47*, 1504–1510.

Banks, K. M., Hare, D. R., & Reid, B. R. (1989) *Biochemistry 28*, 6996–7010.

Calladine, C. R., & Drew, H. R. (1984) *J. Mol. Biol. 178*, 773–782.

Cambridge Convention (1989) *EMBO J. 8*, 1–4.

Chattopadhyaya, R., Ikuta, S., Grzeskowiak, K., & Dickerson, R. E. (1988) *Nature (London) 334*, 175–179.

Chou, S.-H., Wemmer, D. E., Hare, D. R., & Reid, B. R. (1984) *Biochemistry 23*, 2257–2262.

Clore, G. M., & Gronenborn, A. M. (1984) *FEBS Lett. 172*, 219–225.

Clore, G. M., Oschkinat, H., McLaughlin, L. W., Benseler, F., Happ, C. S., Happ, E., & Gronenborn, A. M. (1988) *Biochemistry 27*, 4185–4197.

Coll, M., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1986) *J. Biomol. Struct. Dyn. 4*, 157–172.

Crippen, G. M. (1981) *Distance Geometry and Conformational Calculations*, Research Studies Press/Wiley, Chichester, U.K.

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

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

Dickerson, R. E., Goodsell, D. S., Kopka, M. L., & Pjura, P. E. (1987) *J. Biomol. Struct. Dyn. 5*, 557–579.

Diekmann, S., & McLaughlin, L. W. (1988) *J. Mol. Biol. 202*, 823–834.

Feigon, J., Wang, A. H.-J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1985) *Science 230*, 82–84.

Frey, M. N., Koetzle, T. F., Lehmann, M. S., & Hamilton, W. C. (1973) *J. Chem. Phys. 59*, 915–924.

Hagermann, P. J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4632–4636.

Hare, D. R., & Reid, B. R. (1982) *Biochemistry 21*, 5129–5135.

Hare, D. R., & Reid, B. R. (1986) *Biochemistry 25*, 5341–5350.

Hare, D. R., Wemmer, D. E., Chou, S.-H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319–336.

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., Crippen, G. M., & Kuntz, I. D. (1979) *Biopolymers 18*, 73–81.

Kalk, A., & Berendsen, H. J. C. (1976) *J. Magn. Reson. 24*, 343–351.

Kintanar, A., Klevit, R. E., & Reid, B. R. (1987) *Nucleic Acids Res. 15*, 5845–5862.

McClarin, J. A., Frederick, C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J., & Rosenberg, J. M. (1986) *Science 234*, 1526–1541.

Nerdal, W., Hare, D. R., & Reid, B. R. (1988) *J. Mol. Biol. 201*, 717–739.

Nilges, M., Clore, G. M., Gronenborn, A. M., Brunger, A. T., Karplus, M., & Nilsson, L. (1987a) *Biochemistry 26*, 3718–3733.

Nilges, M., Clore, G. M., Gronenborn, A. M., Piel, N., & McLaughlin, L. W. (1987b) *Biochemistry 26*, 3734–3744.

Ott, J., & Eckstein, F. (1985) *Biochemistry 24*, 2530–2535.

Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1985) *J. Magn. Reson. 66*, 187–193.

Pardi, A., Hare, D. R., & Wang, C. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8785–8789.

Patel, D. J., Ikuta, S., Kozlowski, S., & Itakura, K. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2184–2188.

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

Pohl, F. M., & Jovin, T. M. (1972) *J. Mol. Biol. 67*, 375–396.

Redfield, A. G. (1978) *Methods Enzymol. 49*, 253–270.

Reid, B. R. (1987) *Q. Rev. Biophys. 20*, 1–34.

Reid, B. R., Banks, K., Flynn, P., & Nerdal, W. (1989) *Biochemistry* (preceding paper in this issue).

Saenger, W., Hunter, W. N., & Kennard, O. (1986) *Nature (London) 324*, 385–388.

States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.

Tropp, J., & Redfield, A. G. (1981) *Biochemistry 20*, 2133–2140.

Wang, A. H.-J., Fujii, S., van Boom, J. H., & Rich, A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 3968–3972.

Wemmer, D. E., Chou, S.-H., & Reid, B. R. (1984) *J. Mol. Biol. 180*, 41–60.

Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature (London) 287*, 755–758.

Wu, H.-M., & Crothers, D. M. (1984) *Nature (London) 320*, 509–513.