Determining local conformational variations in DNA. NMR structures of the DNA duplexes d(CHCCTAATCG) and d(CGTCACGCGC) generated using back calculation of the NOE spectra, a distan...

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · SEPTEMBER 1990

Impact Factor: 4.33 · DOI: 10.1016/0022-2836(90)90288-W · Source: PubMed

J. Mol. Biol. (1990) **214**, 711–736

Determining Local Conformational Variations in DNA

Nuclear Magnetic Resonance Structures of the DNA Duplexes d(CGCCTAATCG) and d(CGTCACGCGC) Generated Using Back-calculation of the Nuclear Overhauser Effect Spectra, a Distance Geometry Algorithm and Constrained Molecular Dynamics

William J. Metzler[1], Chuan Wang[2], Douglas B. Kitchen[2] Ronald M. Levy[2] and Arthur Pardi[1]†

[1]*Department of Chemistry and Biochemistry*
University of Colorado at Boulder
Boulder, CO 80309-0215, U.S.A.

[2]*Department of Chemistry*
Rutgers, The State University of New Jersey
New Brunswick, NJ 08903, U.S.A.

(Received 19 January 1990; accepted 9 April 1990)

Two-dimensional nuclear magnetic resonance (n.m.r.) spectroscopy and a variety of computational techniques have been used to generate three-dimensional structures of the two DNA duplexes d(CGCCTAATCG) and d(CGTCACGCGC). The central six base-pairs in these two decamers contain all ten dinucleotide pairs in DNA and thus, represent a model system for investigating how the local structure of DNA varies with base sequence. Resonance assignments were made for the non-exchangeable base protons and most of the C-1′–C-4′ sugar protons in both decamers. Three-dimensional structures were generated using a distance geometry algorithm and these initial structures were refined by optimizing the fit of back-calculated spectra against the experimental two-dimensional nuclear Overhauser effect (NOE) spectra. This back-calculation procedure consists of calculating NOE cross relaxation rates for a given structure by solution of the Bloch equations, and directly accounts for spin diffusion effects. Use of this refinement procedure eliminates some assumptions that have been invoked when generating structures of DNA oligomers from n.m.r. data. Constrained energy minimization and constrained quenched molecular dynamics calculations were also performed on both decamers to help generate energetically favorable structures consistent with the experimental data. Analysis of the local conformational parameters of helical twist, helical rise, propeller twist, displacement and the α, β, γ, ε and ζ backbone torsion angles in these structures shows that these parameters span a large range of values relative to the X-ray data of nucleic acids. However, the glycosidic and pseudorotation angles are quite well defined in these structures. The implications that these results have for determination of local structural variations of DNA in solution, such as those predicted by Callidine's rules, are discussed. Our results differ significantly from some previous studies on determining local conformations of nucleic acids and comparisons with these studies are made.

1. Introduction

Single-crystal X-ray studies of DNA oligomers have shown that the local conformation of the DNA varies with the base sequence (Wing *et al.*, 1980; Drew *et al.*, 1981; Wang *et al.*, 1981; McCall *et al.*, 1986; Saenger, 1984; Dickerson *et al.*, 1985; Dickerson, 1988; Hunter *et al.*, 1989). Calladine (1982) proposed a correlation between these local structural variations and the base sequence, with these variations arising from minimization of purine–purine clash between adjacent base-pairs. Recent studies have shown that the local conformation can vary significantly for different crystalline forms of the same oligomer (Sakked *et al.*, 1989; Jain

† Author to whom all correspondence should be addressed.

0022-2836/90/150711-26 $03.00/0

& Sundarlingam, 1989), where the molecular packing and hydration are proposed to effect the local structure of the DNA duplex in the crystal. Since DNA shows such variations in X-ray studies, it is important to complement these crystal structures with structural studies of DNAs in solution. Presently available two-dimensional (2D†) nuclear magnetic resonance (n.m.r.) spectroscopic techniques have made it possible to determine the three-dimensional structure of biopolymers in solution (for a review, see Wüthrich, 1986). The independent structure determination of the X-ray crystal (Pflugrath *et al.*, 1986) and solution (Kline *et al.*, 1986, 1988; Billeter *et al.*, 1989) structures of an α-amylase inhibitor protein has proven the ability of n.m.r. spectroscopy in determining the overall three-dimensional structures of small proteins. However, analogous studies for DNA duplexes are less easily interpreted since, in general, the questions concerning DNA structural variations within a given helical family involve much smaller atomic displacements, as compared to the kind of structural problems that have to date been addressed by n.m.r. studies of proteins. It is important to realize that many of the sequence-dependent structural variations observed in DNA (Dickerson *et al.*, 1985; Dickerson, 1988; Hunter *et al.*, 1989) involve quite subtle changes in structure. Several groups have performed simulation studies to try and address how precisely DNA conformational parameters can be determined (Nilsson *et al.*, 1986; Pardi *et al.*, 1988; Gronenborn & Clore, 1989), and it is clear that the precision with which the structures can be determined depends upon the type and quality of the n.m.r. distance data.

Recent studies using full relaxation matrix calculations (Borgias & James, 1988; Boelens *et al.*, 1989) or analysis of the Bloch equations (Banks *et al.*, 1989; Madrid *et al.*, 1989) have shown that neglecting spin diffusion effects (Kalk & Berendsen, 1976) in 2D nuclear Overhauser effect (NOE) data can lead to an inaccurate set of distance measurements. The use of full relaxation matrix or Bloch equation analysis allows refinement of structural models against the original experimental NOE data and thus leads to more accurate and precise distance constraints (Borgias & James, 1988; Boelens *et al.*, 1989). Here, we have generated structures of the double-stranded DNA decamers d(CGCCTAATCG) and d(CGTCACGCGC) using an iterative refinement procedure consisting of a distance geometry (DG) algorithm combined with back-calculation of the 2D NOE spectra by solution of the Bloch equations (Banks *et al.*, 1989) and constrained molecular dynamics calculations. We have made relatively conservative estimates for the initial precision of the input distance constraints and these distance constraints have only been modified in order to minimize differences between the volumes of the experimental and calculated 2D NOE cross peaks. This iterative refinement procedure was performed on several sets of structures: a structure starting from standard *A*-form DNA, a structure starting from *B*-form DNA, and five structures generated by a DG algorithm. The local conformational parameters were then analyzed and our results are compared with previous n.m.r. studies of nucleic acid duplexes in solution.

† Abbreviations used: 2D, two-dimensional; n.m.r., nuclear magnetic resonance; NOE, nuclear Overhauser effect; DG, distance geometry; t_1, evolution time; t_2, detection time; FID, free induction decay(s); DQF-COSY, double-quantum filtered correlated spectroscopy; TOCSY, total correlated spectroscopy; r.m.s., root-mean-square; p.p.m., parts per million.

2. Materials and Methods

(a) *n.m.r. sample preparation*

The 4 DNA decamers d(CGTCACGCGC), d(GCGCGTGACG), d(CGCCTAATCG) and d(CGATTAGGCG) were purchased from Pharmacia and were used without further purification. The 1st 2 strands were combined to form decamer 1, the 2nd 2 strands to form decamer 2, and the following numbering scheme was used:

```
                 1 2 3 4 5 6 7 8 9 10
Decamer 1  5' d (C-G-C-C-T-A-A-T-C-G)
               d (G-C-G-G-A-T-T-A-G-C) 5'
                20 19 18 17 16 15 14 13 12 11

                 1 2 3 4 5 6 7 8 9 10
Decamer 2  5' d (C-G-T-C-A-C-G-C-G-C)
               d (G-C-A-G-T-G-C-G-C-G) 5'
                20 19 18 17 16 15 14 13 12 11
```

The individual single strands for each DNA decamer were mixed in equimolar amounts with the concentrations estimated using extinction coefficients calculated from mono- and dinucleotides assuming only nearest-neighbor interactions (Fasman, 1979). The final concentrations for the n.m.r. samples were 2·5 mM in double-stranded DNA for both decamers. The samples were prepared in aqueous buffer containing 0·1 M-NaCl, 10 mM-sodium phosphate, 0·1 mM-sodium EDTA (pH 7·0). The n.m.r. samples were prepared by 2 cycles of lyophilization followed by dissolution of the sample in 99·8% (v/v) ^2H_2O, a final lyophilization and dissolution in 99·996% ^2H_2O, before transfer to 5 mm n.m.r. tubes.

(b) *2D n.m.r. experiments*

The n.m.r. experiments were carried out at 500 MHz on a Varian VXR-500S spectrometer with a sample temperature of 27(±0·1)°C. The 2D n.m.r. data were recorded in the phase-sensitive mode with quadrature detection in both dimensions employing the hypercomplex method (Mueller & Ernst, 1979; States *et al.*, 1982). Low-power continuous-wave irradiation was applied to the residual water signal during the 1·2 to 1·5 s recycle delay. The 2D data were collected in the following form: 2048 complex data points in t_2, 350 to 400 complex free induction decays (FID)s in t_1, 48 to 96 transients for each FID and with spectral widths in both dimensions of 6000 Hz. The double-quantum filtered correlated spectroscopy (DQF-COSY; Piantini *et al.*, 1982; Rance *et al.*, 1983), 2D NOE (Jeener *et al.*, 1979; States *et al.*, 1982) and total correlated spectroscopy (TOCSY) experiments (Braunschweiler & Ernst, 1983; Bax & Davis, 1985), were acquired by standard methods. The TOCSY experiments employed an 8·1 kHz spin-lock pulse of 60 to 70 ms, consisting of two

3 ms trim pulses surrounding an MLEV-16 composite pulse sequence (Bax & Davis, 1985). For decamer 1, 2D NOE spectra were acquired with mixing times of 1, 30, 60, 90, 120 and 200 ms and for decamer 2, the spectra were acquired with mixing times of 1, 30, 60, 90 and 120 ms. The n.m.r. data were transferred to a Sun 4/260 computer and processed with the FTNMR or FELIX programs (Hare Research, Inc.). The first complex points in both the t_1 and t_2 dimensions were appropriately scaled to eliminate ridge artifacts (Otting *et al.*, 1986). All 2D spectra were Fourier transformed into a 2048×2048 matrix by zero-filling, with both time domains in each data set multiplied by a 0 to 90° phase-shifted sine bell window function before Fourier transformation.

(c) *Determination of initial distance constraints*

For the iterative back-calculation procedure 2 types of distance constraints were generated for the 1st round of refinement. For large (corresponding to distances of the order of 3·7 Å (1 Å = 0·1 nm) or less) well-resolved cross peaks, distances were estimated from the build-up rate of the NOE using the 2 spin approximation and assuming a single correlation time for all protons. Volume integrals were measured from the NOE data at all of the 5 or 6 mixing times for each decamer. The following equation was used to estimate distances: $r_{ij} = r_c(\sigma_c/\sigma_{ij})^{1/6}$, where r_{ij} is an unknown distance, r_c, is a known calibration distance; σ_{ij} and σ_c, are build-up rates for the unknown and calibration distances, respectively, which were estimated from the initial slope of the NOE intensities as a function of the fixing time. The distance between the thymine C-6 proton and C-5 methyl group (2·90 Å) was used to calibrate distances involving a methyl group, and the distance between the cytosine C-5 and C-6 protons (2·45 Å) was used to calibrate all other proton–proton distances. The bounds for these initial distance constraints were set with an arbitrarily chosen precision of ±0·5 Å to reduce the potential for obtaining inconsistencies between constraints. Although an even more conservative constraint range could have been used, this range of ±0·5 Å was used as a compromise between the time required for the refinement and potential errors introduced by using too precisely defined distances. As will be seen, this precision for the distance constraints still leads to a wide range of structures consistent with the input data. For small cross peaks (where signal-to-noise limitations prevented useful analysis of the NOE build-up), for partially (or totally) overlapping cross peaks and for cross peaks whose volumes could not be accurately measured due to baseline variation in that particular spectral region, no attempt was made to quantify the NOE. Instead, the initial lower bound for the distance was set to the van der Waals' radii (1·8 Å) and the initial upper bound set to 5·0 Å. No distance constraints were initially added for completely overlapping cross peaks.

(d) *DG calculations*

The DG calculations (Crippen, 1981) were performed with the program DSPACE (Hare Research, Inc.) in a manner similar to that previously described (Hare *et al.*, 1986; Patel *et al.*, 1987; Nerdal *et al.*, 1988). In addition to the n.m.r.-derived distance constraints, the base-pairs were assumed to have standard Watson–Crick geometries with the imino N-H . . . N and the amino N-H . . . O hydrogen bonding distance constraints of 1·8 to 1·9 Å. All the structural information was entered in a distance bounds matrix that was then smoothed by application of triangle inequalities (Crippen, 1981). In addition to the distance information, the deoxyribose sugars were constrained to be in the D-stereoisomer during the refinement.

(e) *Iterative refinement using back-calculation of the 2D NOE spectra*

For the iterative back-calculation refinement procedure the 2D NOE spectrum for a given structure was calculated using the FORTRAN programs BKCALC and GNOE kindly provided by Dr Dennis R. Hare (Hare Research, Inc.). The procedure for back-calculation has been described (Banks *et al.*, 1989), and involves numerical integration of the Bloch equations, where the effects of spin diffusion are accounted for explicitly. These calculations require values for the cross relaxation scaling factor and the "Z leakage" rate. The cross relaxation scaling factor corresponds to a correlation time parameter and it was assumed to be constant for the whole molecule. The Z leakage rate accounts for the loss of all Z magnetization during the mixing time for an individual spin, except for that arising from NOE cross relaxation, and thus corresponds to a T_1 type parameter. The Z leakage rate was also assumed to be constant for the whole molecule. Banks *et al.* (1989) give a more extensive discussion of these parameters, and also describe the method that was used here for estimating values for these parameters.

For each decamer, 3 sets of structures were used as starting points for structure refinement: standard *A*-form DNA, standard *B*-form DNA and 5 structures previously generated with the program DSPACE. These 5 "initial DG structures" were generated with a precisely defined set of distance constraints (±0·2 Å or ±0·3 Å). These distances were generated using a 2-spin approximation and then fitting to the initial build-up rate of the NOE for 2D NOE spectra with mixing times of 60 and 100 ms for decamer 1 and 75 and 100 ms for decamer 2. The procedure used is similar to that previously described (Hare *et al.*, 1986; Nerdal *et al.*, 1988; Pardi *et al.*, 1988), and the full details are given by Wang (1989). We would like to emphasize that our choice of starting structures is rather arbitrary. The primary criterion for the selection of starting structures is that they should lead to refined structures that sample all of the conformational space consistent with the experimental data. Thus, the fact that our initial DG structures were generated with high precision (±0·2 Å to 0·3 Å) is not relevant and, as discussed below, the 7 defined structures are all consistent with our experimental data while still having quite large conformational variations.

The back-calculation refinement proceeded in 2 stages using a procedure similar to that of Banks *et al.* (1989). First, stack plots were compared to look for deviations between calculated and experimental spectra. The advantage of using stack plots for performing this analysis (as compared with contour plots), is that one is able to eliminate visually any problems resulting from baseline offset between 2 spectra and one can spot easily where ridges in the spectrum may affect the true intensity of a cross peak. The stack plots of the back-calculated 2D NOE spectra for each decamer were compared with the experimental (120 ms or 200 ms) 2D NOE spectra to judge how well the model structure fit the experimental data. Because of the r^{-6} dependence of the NOE, a small error in the distance between 2 protons in the model structure can manifest itself as a large change in the intensity of the cross peak for these 2 protons. Thus, extremely intense or missing cross peaks were observed

commonly in the back-calculated 2D NOE spectra of the starting structures in the 1st round of refinement. All regions of the 2D NOE spectra for which the protons were assigned were analyzed, but the cross peaks involving methyl protons were not used in the iterative refinement procedure due to the known difference in correlation times for methyl groups. Distance constraints were manually modified if there were significant differences in the intensities of a cross peak in the experimental and back-calculated spectra. Additional constraints were added for assigned protons where cross peaks were observed in the back-calculated but not in the experimental spectra, and these are referred to as non-NOE constraints. Refinement was considered complete when no significant change in the net difference (determined visually) in intensity between the back-calculated and experimental spectra was observed for several rounds of refinement. A total of 10 iterations were performed on each structure. It is important to note that our distance constraints were modified only when required to improve the fit between the calculated and the experimental spectra. This leads to the widest possible range between the lower and upper bounds of the distance constraints, while still fitting the experimental data. In no case was the range of a distance constraint ever reduced to less than 0·4 Å. Tighter distance constraints were not used, even if there was a perfect fit between the experimental and the calculated cross peak volumes, because in the presence of spin diffusion there is more than 1 way of changing the size of a cross peak. For example, it is possible to change the volume of a cross peak by changing the distance between the 2 protons involved in the cross peak, or by changing the distance between a 3rd proton and 1 of the protons in the cross peaks. Such spin diffusion, or 3-spin, effects can lead to a variety of structures consistent with the volume of an individual cross peak. Therefore, a perfect fit for a cross peak volume *does not imply* a perfectly defined distance between the protons in this cross peak.

The 2nd stage of each round of iterative refinement involved modification of our model structures to fit the new set of distance constraints. This was performed in the program DSPACE, where a new bounds-matrix was generated and then used to calculate the penalty function during refinement. Structures were "heated" using a simulated annealing-type algorithm (Hare *et al.*, 1986; Nerdal *et al.*, 1988) to allow them to sample conformational space consistent with the new distance constraints and then "cooled" with damped annealing. A final round of conjugate gradient minimization brought the structures into stable local minima. The new model structures generated with this procedure were then subjected to back-calculation to see how well they fit the experimental data.

Separate back-calculation and refinement procedures were performed on each decamer for the *A*, *B* and 1 of the initial 5 DG structures. The remaining 4 DG structures were "carried along" in the refinement procedure using the new set of distance constraints generated from back-calculation of the 5th structure. The final 3 sets of structures will be referred to as refined *A*, refined *B* and refined DG structures, and correspond to the 10 rounds of iteratively refined structures starting from the standard *A*, standard *B* and initial DG structures, respectively.

(f) *Energy minimization and molecular dynamics calculations*

Constrained energy minimization was performed on the refined *A*, refined *B* and 5 refined DG structures for each DNA decamer. These calculations were performed with the program IMPACT (Bassolino *et al.*, 1988), using the all-atom force-field parameters described by Weiner *et al.* (1986) and employing the energy potential functions of bond stretching, bond angle bending, hindered rotation about a single bond, van der Waals' interactions, electrostatic interactions, hydrogen bonding and distance constraints of the form described by Bassolino *et al.* (1988). For each structure, 500 steps of conjugate gradient minimization were performed. A distance-dependent dielectric constant was used for the electrostatic term, so no counter ions were included in the calculation. The non-bonded list was updated every 10 steps with a non-bonded cutoff of 7·5 Å. The NOE distance constraints for the constrained energy minimizations consisted of constraints between the same pairs of protons as used in the back-calculation refinement, except that exact distances for each pair of protons were measured in each of the refined structures and were entered with a precision of ±0·1 Å. As will be discussed below, the reason for using these distance constraints was to ensure that the final energy-minimized structures fit well to the experimental NOE data. The form of the potential used for the distance constraints has been described by Bassolino *et al.* (1988), and a relative weighting factor of 10 was used for the distance constraints (for comparison, the weighting factor for a bond stretching term for a C–C single bond was around 200).

Constrained quenched molecular dynamics was performed on the energy-minimized structure in order to search for additional minima in the energy surface. The potentials used in the molecular dynamics calculations were the same as those used in the energy minimizations. The structures were heated to 2500 K for 0·5 ps (0·5 fs/step) of constrained dynamics using both the NOE and hydrogen bonding distance constraints, then cooled to 100 K for 1·0 ps of constrained dynamics with the NOE and hydrogen bonding distance constraints. A final 500 steps of constrained minimizations were performed with no hydrogen bonding distance constraints, but instead included the standard hydrogen bonding potential discussed above.

(g) *Calculation of conformational parameters*

The conformational parameters were calculated by the programs AHELIX (written by J. Rosenberg), and slightly modified versions of BROLL, CYLIN and DTORAN (written by R. E. Dickerson). The definitions for the helical twist, helical rise, propeller twist, tilt, roll, displacement (also known as dislocation; Saenger, 1984), backbone torsion angles and pseudorotation angle have been described (Frantini *et al.*, 1982; Dickerson, 1983; Saenger, 1984). Since the sign for the displacement, roll, and possibly tilt, generated by these programs varied with the absolute orientation of the molecule, we will only report and analyze the magnitude of these parameters here. Some of the conformational parameters depend upon the definition of the helix axis of the DNA and the best fit helical axis was calculated by the method of Rosenberg *et al.* (1976), using the co-ordinates of the C-1′ atoms in each decamer. The helical twist corresponds in Dickerson's nomenclature, to the global twist and is calculated from the C-1′–C-1′ vectors of successive base-pairs. To calculate displacement, the long axes of the base-pairs (through the C-6–C-8 atoms) are projected into a normal plane of the best-fit helical axis. The displacement is calculated as the distance of the projected

C-6–C-8 vector from the best-fit axis. The definitions of tilt and roll correspond, in Dickerson's nomenclature, to θ-roll and θ-tilt, and are calculated as the base-plane through both bases in the pair and are defined relative to the neighboring base-pair. It is important to realize that some of the helical parameters are defined relative to the helical axis and, therefore, depend on how this best-fit axis is determined for a curved helix. The pseudorotation angle was defined and calculated as described by Altona & Sundaralingam (1972).

3. Results

(a) *Proton resonance assignments*

Resonance assignment of the proton spectra of the DNA decamers were made by standard methods (Feigon *et al.*, 1983; Hare *et al.*, 1983; Scheek *et al.*, 1984; Weiss *et al.*, 1984; Wüthrich, 1986), and involved first the application of DQF-COSY and TOCSY spectra to group protons into individual spin systems on the aromatic bases or sugar rings. Intraresidue NOEs were used to link an aromatic base with its sugar, and then interresidue sugar proton–base proton NOEs were used to make sequential assignments (Wüthrich, 1986). In a right-handed double helix a C-6 or C-8 base proton generally shows NOEs to the C-1′, C-2′ and C-2″ protons of its own sugar ring as well as to the C-1′, C-2′ and/or C-2″ protons on the 5′ neighboring sugar (Wüthrich, 1986). Figure 1 shows part of a NOE connectivity diagram that was used to make sequential assignments for the C-6 or C-8 and C-1′ protons in decamer 2. Similar connectivities were made for the C-6 or C-8 and C-2′, C-2″ and C-3′ protons. This procedure led to the assignment of all the non-exchangeable base protons (except the adenosine C-2 base protons), all the C-1′, C-2′, C-2″, C-3′ protons and most of the C-4′ protons of the sugar rings in both decamers. The C-2′ and C-2″ protons were differentiated by the different sizes for their C-1′ proton intraresidue NOEs and for their $^3J_{H1'-H1'}$ coupling constants (data not shown; Wang, 1989). The adenosine C-2 protons were assigned, along with the exchangeable imino protons, by standard methods (data not shown; Wüthrich, 1986) with details given by Wang, (1989). The resonance assignments for the proton spectra of decamer 1 and decamer 2 are given in Tables 1 and 2 respectively.

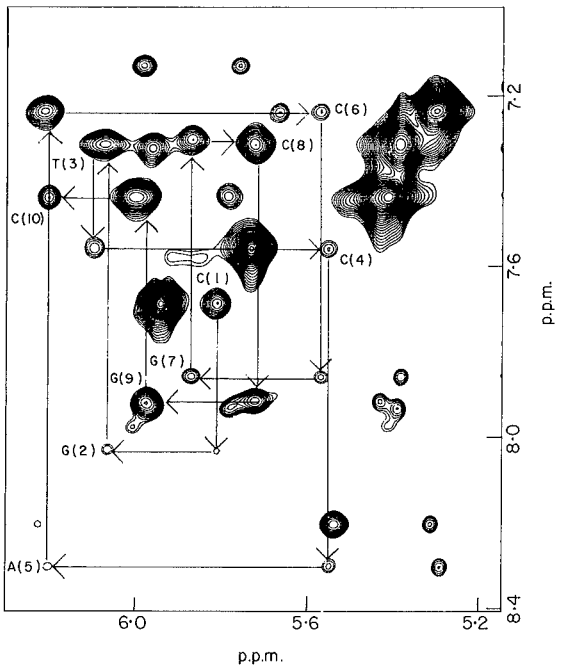

Figure 1. Expansion of the H-1′/H-6, H-8 region of the 2D NOE spectrum for decamer 2. An example of the NOE connectivities used in making sequential assignments (see the text) is shown for residues 1 through 10. The intra-residue base-proton to C-1′-proton NOEs for these residues are labeled in the Figure.

(b) *Iterative refinement of the decamers using back-calculation of the 2D NOE spectra, minimization and simulated annealing*

Refinement of the standard *A*-form, standard *B*-form, and the initial DG structures were carried out by comparison of stacked plots of the experimental and back-calculated 2D NOE spectra as described in Materials and Methods. Figure 2(a) and (b) show stacked plots for the H-1′/H-3′–H-8/H-6/H-2 region of the experimental 200 (or 120) millisecond 2D NOE spectra, and the calculated spectra for one of the initial and refined DG structures for decamers 1 and 2, respectively. Since three separate sets of back-calculations were performed for each decamer, Figures 3 and 4 show analogous plots of this region of the spectrum for both decamers for the *A* and *B*-structure refinements, respectively.

The number and precision of the distance constraints increased in each round during the ten rounds of iterative refinement. Table 3 lists changes in the distance constraints for the DG structures of both decamers from the initial to the final round of refinement. A complete list of the distance constraints, cross peak intensities and co-ordinates for the fully refined DNA structures are available from the authors upon request.

(c) *Energy minimization and molecular dynamics of the refined DG structures*

The refined *A*, refined *B* and refined DG structures for decamers 1 and 2 were subjected to constrained energy minimization to improve the energetics and covalent geometry of the structures. Table 4 gives the potential energies (in kcal/mol: 1 cal = 4·184 J) for the constrained energy minimized structures of both decamers. The conformations of the phosphate groups are very disordered in the refined DG structures and this situation did not significantly change upon constrained minimization (see below). Thus, we performed constrained

Table 1

Chemical shifts, δ, for the assigned proton resonances of decamer 1

Residue	H-6/H-8	H-1′	H-2′	H-2″	H-3′	H-4′	H-5/CH_3/H-2
C(1)	7·66	5·80	2·01	2·45	4·74	4·10	5·94
G(2)	8·00	5·96	2·77	2·85	5·03	(4·40)†	
C(3)	7·45	6·06	2·25	2·55	4·86	4·30	5·41
C(4)	7·56	5·89	2·10	2·50	4·78	4·18	5·59
T(5)	7·40	5·67	2·12	2·48	4·90	4·16	1·71
A(6)	8·31	5·97	2·68	2·78	5·06	4·42	6·80
A(7)	8·17	6·18	2·57	2·91	4·99	4·47	7·62
T(8)	7·15	5·91	2·00	2·44	4·84	4·17	1·34
C(9)	7·49	5·76	2·02	2·41	4·86	4·15	5·66
G(10)	7·97	6·20	2·40	2·64	4·71	4·21	
C(11)	7·63	5·76	2·39	1·91	4·73	4·08	5·92
G(12)	7·99	5·61	2·77	2·86	5·04	4·37	
A(13)	8·29	6·34	2·76	3·01	5·09	4·54	7·88
T(14)	7·19	5·95	2·02	2·58	4·87	4·23	1·41
T(15)	7·38	5·77	2·14	2·48	4·92	4·16	1·67
A(16)	8·19	6·02	2·70	2·91	5·07	4·41	7·29
G(17)	7·63	5·58	2·55	2·66	4·99	4·38	
G(18)	7·69	5·91	2·49	2·70	4·97	4·39	
C(19)	7·33	5·77	1·95	2·36	4·86	4·16	5·35
G(20)	7·97	6·20	2·40	2·64	4·71	4·21	

Chemical shifts are listed in parts per million (p.p.m.) at 27 °C, pH 7·0. The chemical shifts are relative to the residual water signal that resonates at 4·75 p.p.m. and have errors of ±0·01 p.p.m.

† The parentheses indicate a tentative assignment.

quenched dynamics calculations on the constrained energy-minimized *A*, *B* and DG structures (for details, see Materials and Methods), to test if improved sampling of the energy-distance penalty surface would lead to more regular phosphate conformations. The potential energies of the final constrained quenched dynamics structures for the three sets of calculations for both decamers are also given in Table 4.

(d) *Structural analysis of the decamer structures: r.m.s. deviations and conformational parameters*

r.m.s. deviations were calculated for all atoms in the central six base-pairs of both decamers. Table 5 lists the average r.m.s. distance deviations for the refined, constrained energy-minimized and constrained quenched dynamics *A*, *B* and DG structures for decamers 1 and 2. It is interesting to note

Table 2

Chemical shifts, δ, for the assigned proton resonances of decamer 2

Residue	H-6/H-8	H-1′	H-2′	H-2″	H-3′	H-4′	H-5/CH_3/H-2
C(1)	7·68	5·80	2·08	2·48	4·73	4·11	5·93
G(2)	8·02	6·05	2·72	2·85	5·02	4·41	
T(3)	7·31	6·08	2·17	2·54	4·91	4·27	1·45
C(4)	7·56	5·54	2·09	2·41	4·86	4·15	5·72
A(5)	8·30	6·19	2·73	2·88	5·04	4·43	7·68
C(6)	7·23	5·56	1·92	2·31	4·81	4·13	5·28
G(7)	7·85	5·86	2·62	2·71	4·98	4·35	
C(8)	7·30	5·71	1·97	2·35	4·84	4·17	5·37
G(9)	7·91	5·96	2·64	2·77	4·99		
C(10)	7·44	6·19	2·20	2·20	4·50	4·07	5·42
G(11)	7·97	5·99	2·61	2·81	4·88	4·27	
C(12)	7·43	5·77	2·13	2·46	4·90	4·23	5·40
G(13)	7·92	5·95	2·58	2·63	5·02	4·39	
C(14)	7·32	5·70	2·05	2·42	4·83	4·21	5·37
G(15)	7·91	5·97	2·62	2·78	4·99		
T(16)	7·12	5·74	1·93	2·35	4·86	4·17	1·48
G(17)	7·91	5·52	2·71	2·77	5·01	4·35	
A(18)	8·19	6·22	2·65	2·88	5·04	4·13	7·87
C(19)	7·24	5·65	1·84	2·29	4·81	4·13	5·30
G(20)	7·89	6·15	2·59	2·37	4·67	4·19	

Chemical shifts are listed in parts per million (p.p.m.) at 27 °C, pH 7·0. The chemical shifts are relative to the residual water signal that resonates at 4·75 p.p.m. and have errors of ±0·01 p.p.m.

Table 3
Classification of NOE constraints for decamers 1 and 2

Distance range†	Number of NOE constraints: Initial DG structures	Refined DG structures
A. *Decamer 1*		
=0·4	0	148
0·5–0·7	0	24
0·7–1·0	164	10
1·0–2·0	0	43
2·0–3·0	0	19
>3·0	73	2
Non-NOE‡	0	47
Total	237	293
B. *Decamer 2*		
=0·4	0	132
0·4–0·7	0	41
0·7–1·0	135	19
1·0–2·0	0	52
2·0–3·0	0	6
>3·0	80	2
Non-NOE‡	0	38
Total	215	290

† The distance range is the difference between the upper and lower bounds in Å.

‡ Additional distance constraints required to fit the calculated NOE spectra to the experimental data (see the text).

Table 5
Average r.m.s. deviations (Å) for decamers 1 and 2

Dynamics	Refined	Constrained energy-minimized	Constrained quenched
A. *Decamer 1*			
DG–DG	1·69	1·54	1·88
A–DG	3·35	3·38	3·19
B–DG	1·59	1·52	1·79
A–*B*	2·75	2·90	3·23
A-form†–DG	3·92	3·76	3·60
B-form†–DG	2·87	2·60	2·36
B. *Decamer 2*			
DG–DG	1·96	1·78	1·65
A–DG	2·36	2·10	2·26
B–DG	1·70	1·43	1·51
A–*B*	1·90	1·77	1·64
A-form†–DG	4·26	4·10	3·98
B-form†–DG	2·65	2·46	2·58

The r.m.s. deviations are calculated for all atoms in the central 6 base-pairs.

† These are the standard *A* or *B*-form structures for the decamers.

that while the r.m.s. deviations indicate that the refined *B* structure is very similar to all five refined DG structures for both decamers, the refined *A* structures have much larger r.m.s. deviations from either the refined DG or refined *B* structures in both decamers (approx. 3·0 Å and 2·3Å in decamers 1 and 2, respectively).

The following local structural parameters were analyzed for the central six base-pairs in both DNA decamers: helical rise, helical twist, roll, tilt, propeller twist, displacement, glycosidic angle, α, β, γ, δ, ε, ζ torsion angles and the pseudorotation angle. Table 6 lists the average, standard deviation, and total range of values for each of these parameters (except for the α, β, γ, ε and ζ torsion angles) in the initial, refined, constrained energy minimization and quenched dynamics structures of decamer 1. The corresponding values of these parameters for decamer 2 are given in Table 7. The variations in these parameters in both decamers will be discussed below.

4. Discussion

In X-ray crystallography, structure refinement proceeds by minimizing the deviation of the calculated structure factors for a given structural model from the experimental structure factors, and usually incorporates information on the covalent geometry (Hendrickson & Konnert, 1981). The procedure for refining the solution structure of a molecule against the original 2D n.m.r. data is not as straightforward, since one needs to know the correlation times for individual proton pairs in order to calcu-

Table 4
Relative energies (in kcal/mol) for decamers 1 and 2

DNA structure	Refined	Constrained energy minimized	Constrained quenched dynamics
A. *Decamer 1*			
A	$7{\cdot}5\times10^3$	−513 (7·6)†	−541 (9·8)
B	$7{\cdot}6\times10^3$	−529 (11)	−521 (16)
DG‡	$7{\cdot}6\pm0{\cdot}1\times10^3$	−540±89 (8·2±0·8)	−567±55 (13±1·7)
B. *Decamer 2*			
A	$7{\cdot}2\times10^3$	−610 (12)	−566 (13)
B	$7{\cdot}2\times10^3$	−596 (9·1)	−589 (13)
DG‡	$7{\cdot}3\pm1{\cdot}3\times10^3$	−638±27 (12±1·7)	−620±17 (15±1·7)

1 cal = 4·184 J.

† Values in parentheses are the relative energies of the distance constraint violations.

‡ Average values for 5 DG structures±standard deviation.

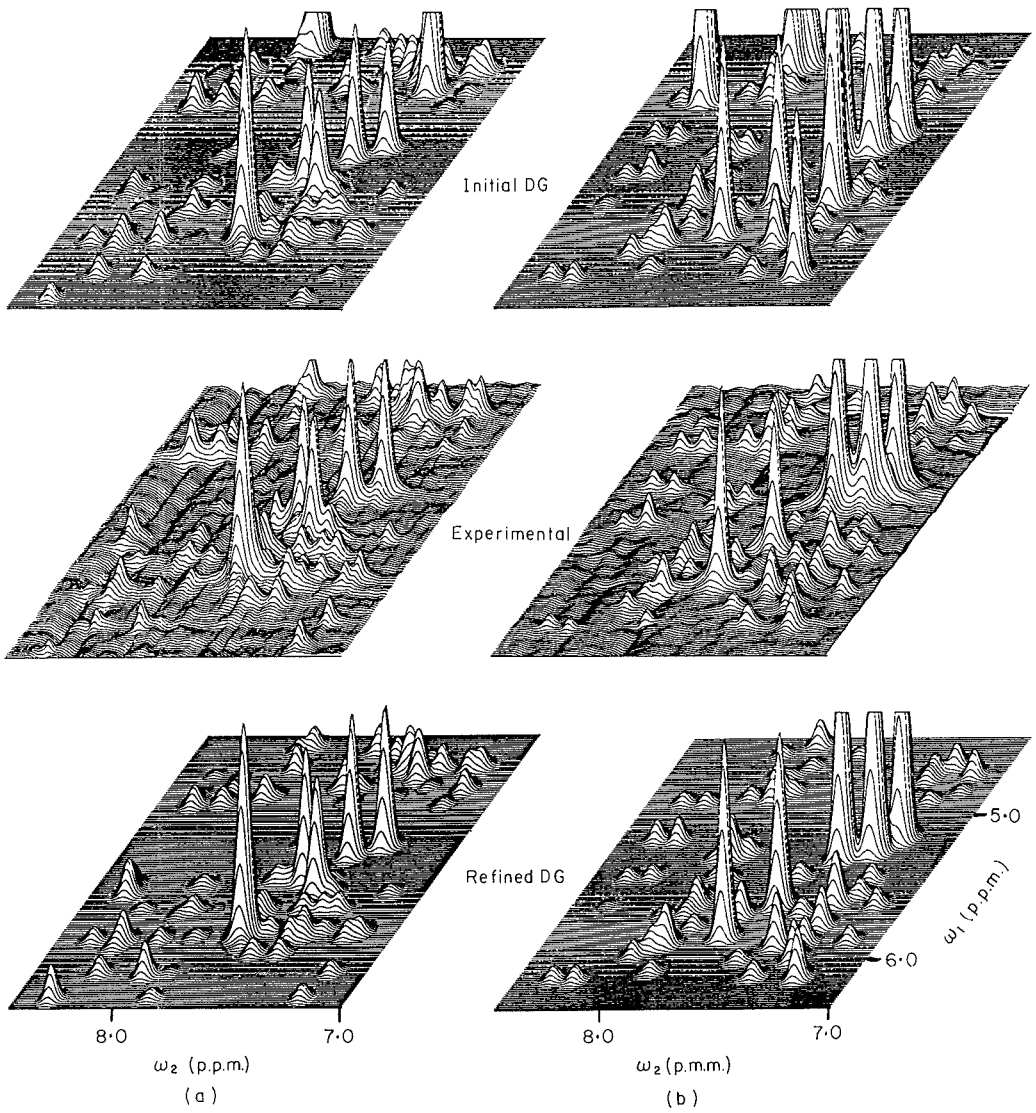

Figure 2. Stack plots of the H-6/H-8 to H-1′/H-3′ region for 1 of the 5 refined DG structures of (a) decamer 1, and (b) decamer 2. The top spectra are 1 of the calculated initial DG structures, the middle spectra are the experimental NOE data (200 ms for decamer 1 and 120 ms for decamer 2), and the bottom spectra are the calculated spectra for the final refined DG structures.

late accurate distances from the measured NOEs. However, structure refinement methods have been developed for n.m.r. where, by assuming a single correlation time for the whole molecule, it is possible to calculate all proton–proton NOEs. Using such a back-calculation (Banks *et al.*, 1989; Madrid *et al.*, 1989) or full-relaxation matrix analysis (Borgias & James, 1988; Boelens *et al.*, 1989) it is possible to account directly for spin diffusion effects (Kalk & Berendsen, 1976) in the molecule. Simulations by Borgias & James (1988) have shown that for short DNA oligomers (such as those used here), the errors introduced by the assumption of a single correlation time are generally much less than those produced by neglecting spin diffusion effects. Thus, by using a procedure that refines the structure against the original data we have much more confidence in the accuracy and precision of the final structures.

One goal of this study was to search for sequence-specific structural variations in DNA double helices, such as those observed in X-ray crystal structures (Wing *et al.*, 1980; Drew *et al.*, 1981; Wang *et al.*, 1981; McCall *et al.*, 1986; Saenger, 1984; Dickerson *et al.*, 1985; Dickerson, 1988; Hunter *et al.*, 1989) and predicted by Calladine's rules (Calladine, 1982; Dickerson, 1983). The studies performed here were aimed at understanding how accurately and precisely the solution structures of two DNA decamers could be determined by refining the structure against the original NOE data, and by inclusion of energy potential functions in the structure generation process. The central six base-pairs of the two decamers studied here contain all ten possible nearest-neighbor interactions in DNA and thus serve as a model for local structural variations with base sequence.

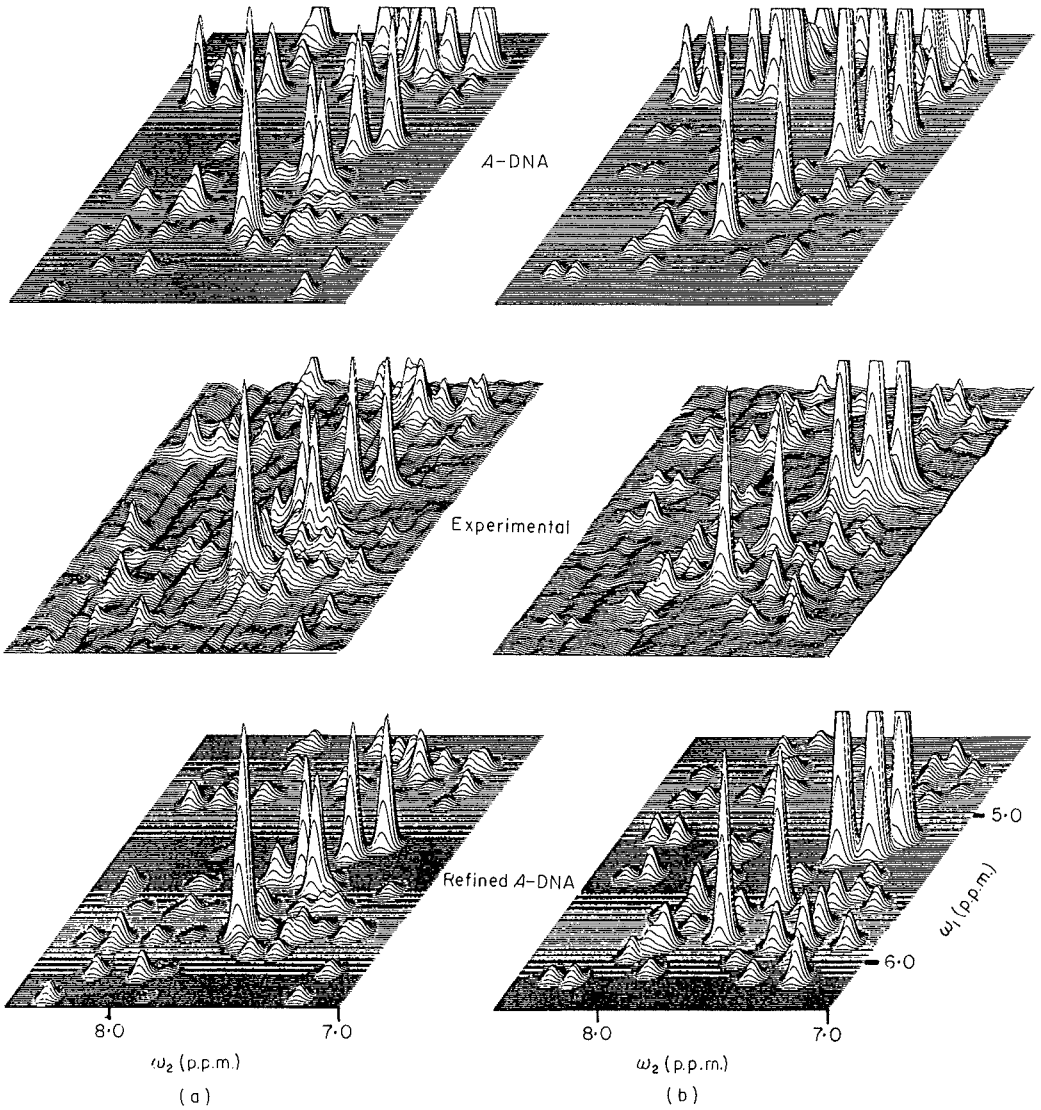

Figure 3. Stack plots of the H-6/H-8 to H-1′/H-3′ region for the refined *A*-DNA structure of (a) decamer 1, and (b) decamer 2. The top spectra are the calculated standard *A*-form structures, the middle spectra are the experimental NOE data (200 ms for decamer 1 and 120 ms for decamer 2), and the bottom spectra are the calculated spectra for the final refined *A* structures.

(a) *Structure refinement using back-calculation of the 2D NOE spectra*

The initial DG structures of both decamers were refined against the original 2D NOE data by the back-calculation procedure described in Materials and Methods. It is apparent from comparison of the calculated spectra with the experimental spectra (Figs 2 to 4) that none of the starting structures represents a close fit to the experimental data. However, iterative refinement of the structures yields calculated spectra that fit quite well to the n.m.r. data. The refinement was judged to have converged when the total deviations from one round of refinement to the next was essentially unchanged. At this stage the majority of the significant deviations were from cross peaks for protons on the ends of the decamers (data not shown). A possible explanation of these deviations is that the assumption of a single static structure in solution is not valid for the terminal base-pairs. This is consistent with the known fraying of the ends of DNA oligomers (Pardi *et al.*, 1981; Bloomfield *et al.*, 1974).

Since the iterative refinement procedure involved separate distance constraints and back-calculation on only three of the seven structures (see Materials and Methods), it is important to see how well all five fully refined DG structures fit to the original 2D NOE data. Figures 5 and 6 show stacked plots of the aromatic proton to C-2′ and C-2″ proton region of the experimental 2D NOE spectra for the refined *A*, refined *B* and five refined DG structures of decamers 1 and 2, respectively. Although there are some minor differences in cross peak volumes among the seven back-calculated spectra, these plots show that all seven structures give quite good fits to the

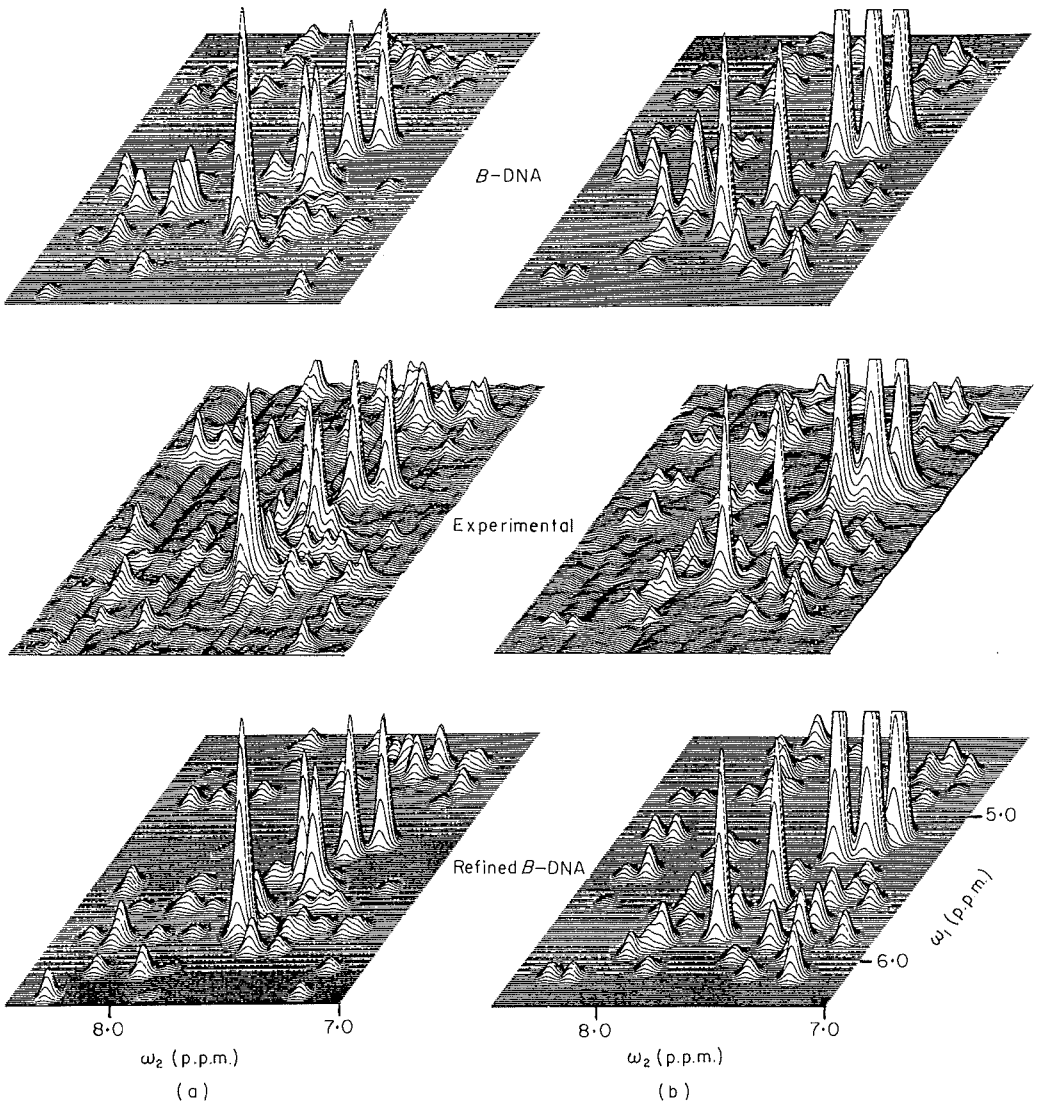

Figure 4. Stack plots of the H-6/H-8 to H-1′/H-3′ region for the refined *B* structure of (a) decamer 1, and (b) decamer 2. The top spectra are the calculated standard *B*-form structures, the middle spectra are the experimental NOE data (200 ms for decamer 1 and 120 ms for decamer 2), and the bottom spectra are the calculated spectra for the final refined *B* structures.

experimental data. As mentioned above, the terminal residues show much larger conformational variations than the interior residues, and the fits of the calculated and experimental spectra were often poorer for terminal residues. Thus, for the vast majority of the non-terminal cross peaks the variation of volumes among the seven structures is quite small. This becomes particularly obvious when one compares the variation in volumes obtained here with what is obtained if the standard structure refinement procedure were employed. For example, the average variation in volumes for cross peaks in the spectral region shown in Figures 5 and 6 (the refined DG, *A* and *B* structures) is only 17% and 10%, respectively. This is in contrast to what is obtained in the initial starting structures (which were generated by more standard methods where a high precision was used for the input distance constraints and spin diffusion was neglected), where the average volume variations are 69% and 51%, respectively, for the same spectral regions. This shows that fits for the refined structures are much better than for the starting structures. Similar improvements in fits are found for the base proton–H-1′, H-3′ proton (Figs 2 to 4) and other regions of the 2D NOE spectrum (data not shown).

To give some perspective to the magnitude of the observed variations in volumes among the seven back-calculated structures, we will consider volume variations of DNA structures generated using standard methods, that is, using a two-spin approximation and neglecting spin diffusion. Distance constraints in the standard methods are generally entered with a precision of ±0·2 to 0·4 Å for distances less than 3·0 Å. (See below for a discussion of the validity of the use of highly precise

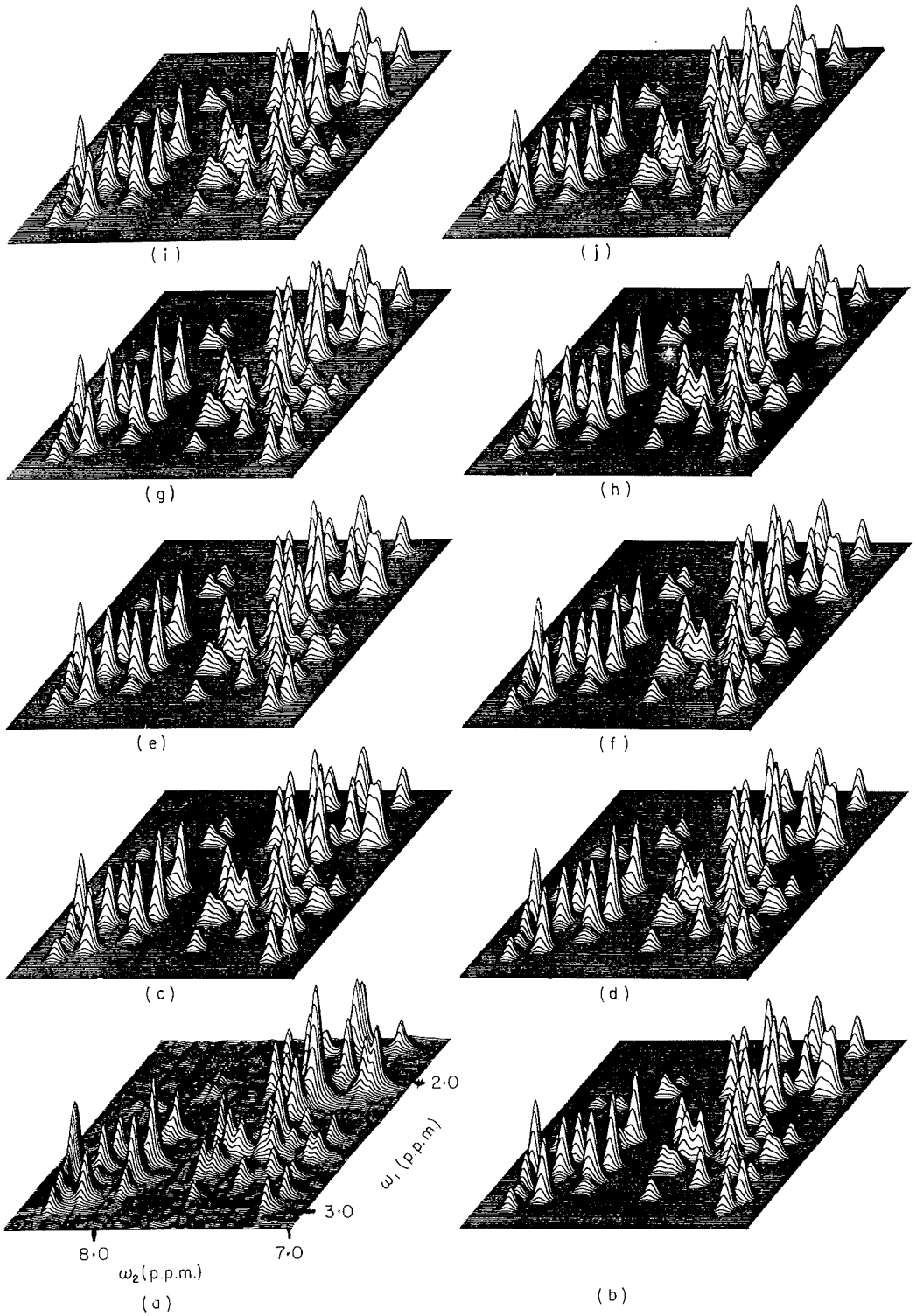

Figure 5. Stack plots of the H-6/H-8 to H-2′/H-H2″ region for the structures of decamer 1: (a) the experimental NOE data (200 ms); (b) to (f) the calculated spectra for the final refined DG structures 1 to 5; (g) the calculated spectrum for the final refined *A* structure; (h) the calculated spectrum for the final refined *B* structure; (i) the calculated spectrum for the constrained minimized structure of DG 1; (j) the calculated spectrum of the constrained dynamics structure of DG 1.

constraints.) Thus, a set of structures can be generated for which individual proton–proton distances vary by as much as 0·8 Å while still *perfectly* satisfying the input distance constraints. A 0·5 Å range in a distance of this magnitude ($\leq$3·0 Å) would correspond to a difference in volume of greater than 300%. Thus, with the standard methods, structures which perfectly satisfy the distance constraints could have very poor fits to the actual experimental data; that is they would have very large volume

Table 6

Conformational parameters for decamer 1

		Initial structures		Refined structures		Constrained energy minimized structures		Constrained quenched dynamics structures	
	Residues†	Average‡	Range§	Average‖	Range§	Average‖	Range§	Average‖	Range§
Helical rise (Å)	C(3)–C(4)	3·6±0·3	0·6	4·0±0·5	1·6	3·7±0·6	1·6	3·9±0·5	1·3
	C(4)–T(5)	3·7±0·3	0·8	4·2±0·4	1·1	4·1±0·4	1·1	3·9±0·4	1·2
	T(5)–A(6)	3·2±0·3	0·7	3·8±0·3	1·0	4·0±0·4	1·1	4·0±0·6	1·9
	A(6)–A(7)	3·2±0·4	0·8	3·2±0·2	0·7	3·3±0·4	1·0	3·2±0·4	1·1
	A(7)–T(8)	3·6±0·3	0·6	4·2±0·2	0·7	4·1±0·2	0·6	4·0±0·2	0·5
	Average¶	0·3	0·7	0·3	1·0	0·4	1·1	0·4	1·2
Helical twist (°)	C(3)–C(4)	30±1·3	3	39±2·8	9	39±4·0	11	37±4·4	12
	C(4)–T(5)	31±2·3	6	29±4·6	13	31±4·6	14	31±2·6	8
	T(5)–A(6)	34±5·0	14	26±7·6	22	26±6·6	19	28±7·1	21
	A(6)–A(7)	35±2·5	6	32±3·3	9	33±1·5	4	34±2·3	7
	A(7)–T(8)	30±2·7	7	20±5·4	18	22±5·7	17	31±7·0	19
	Average¶	2·8	7	4·7	14	4·5	13	4·7	13
Roll (°)	C(3)–C(4)	6±2	6	26±19	52	25±18	54	24±22	67
	C(4)–T(5)	5±3	8	14±10	30	6±7	19	4±3	10
	T(5)–A(6)	18±8	21	30±15	38	17±12	31	16±13	32
	A(6)–A(7)	2±2	6	4±2	6	4±3	7	6±3	8
	A(7)–T(8)	8±4	11	30±8	22	21±7	20	17±8	21
	Average¶	4	10	11	30	9	26	10	28
Tilt (°)	C(3)–C(4)	2·2±0·4	1·0	3·4±2·1	6·2	3·8±2·2	5·7	2·7±4·1	12
	C(4)–T(5)	2·5±1·3	3·4	5·0±3·1	8·4	3·2±1·8	5·3	2·5±1·4	4
	T(5)–A(6)	2·2±1·8	4·6	9·0±4·6	10·1	2·2±1·7	4·4	5·1±5·1	13
	A(6)–A(7)	2·1±1·6	3·6	3·0±2·6	7·8	3·1±1·8	5·0	3·8±2·9	8
	A(7)–T(8)	0·6±0·4	0·9	3·4±3·0	7·8	2·7±1·4	3·4	3·7±1·8	5
	Average¶	1·1	2·7	3·1	7·9	1·8	4·8	3·1	8
Displacement (Å)	C(3)	2·7±1·3	3·2	2·1±1·3	3·0	1·9±1·4	3·6	2·3±1·1	2·8
	C(4)	1·3±0·5	1·3	1·7±1·3	3·9	1·9±1·3	4·0	2·0±1·4	3·8
	T(5)	0·7±0·6	1·5	2·1±1·4	4·2	2·1±1·5	3·6	1·8±1·3	3·8
	A(6)	0·7±0·8	1·5	2·1±1·3	3·8	2·1±1·4	3·8	1·8±1·0	3·1
	A(7)	0·6±0·6	1·3	1·6±1·1	2·7	1·9±0·9	2·4	1·8±1·0	2·8
	T(8)	0·9±0·6	1·6	2·5±1·3	3·7	2·9±1·0	2·6	2·3±1·6	3·7
	Average¶	0·7	1·8	1·3	3·5	1·3	3·3	1·2	3·3
Propeller twist (°)	C(3)	9±2	6	24±6	16	19±7	19	19±9	25
	C(4)	10±4	10	39±10	27	30±9	26	22±11	33
	T(5)	13±8	21	35±4	11	26±7	19	25±9	27
	A(6)	17±2	5	19±6	19	15±4	11	23±9	26
	A(7)	23±7	17	17±8	21	18±7	23	26±10	27
	T(8)	16±6	13	25±8	21	18±2	5	20±7	17
	Average¶	5	12	7	19	6	17	9	26
δ Torsion angle (°)	C(3)	136±6	16	116±12	34	125±14	44	127±22	73
	C(4)	142±5	13	153±5	12	128±19	53	137±12	40
	T(5)	103±29	68	140±6	17	147±19	45	119±16	46
	A(6)	135±19	43	140±26	67	141±12	32	147±6	16
	A(7)	108±16	37	157±6	15	155±10	28	152±9	23
	T(8)	153±22	54	95±10	27	85±12	37	97±8	27
	A(13)	135±15	36	162±9	23	164±4	14	152±14	45
	T(14)	123±21	48	81±5	12	91±14	37	101±29	81
	T(15)	120±12	29	78±3	11	105±10	30	121±14	40
	A(16)	136±12	30	113±24	55	117±14	42	135±16	49
	G(17)	147±10	26	156±9	24	144±13	43	146±15	46
	G(18)	148±7	16	137±26	70	136±27	86	124±25	60
	Average¶	14	35	12	31	14	41	16	45
Glycosidic angle (°)	C(3)	−107±5	12	−137±11	26	−130±5	13	−128±13	38
	C(4)	−59±1	3	−142±5	13	−129±10	25	−133±6	17
	T(5)	−106±25	66	−103±14	43	−106±10	27	−125±4	12
	A(6)	−110±11	26	−93±26	69	−88±19	56	−97±18	44
	A(7)	−111±4	11	−107±5	13	−94±5	18	−118±10	32
	T(8)	−106±10	26	−103±22	55	−102±10	24	−113±6	16

Table 6 *(continued)*

		Initial structures		Refined structures		Constrained energy minimized structures		Constrained quenched dynamics structures	
	Residues†	Average‡	Range§	Average‖	Range§	Average‖	Range§	Average‖	Range§
	A(13)	−122±3	8	−66±11	29	−70±9	22	−86±13	32
	T(14)	−127±6	16	−80±11	33	−93±8	21	−107±13	34
	T(15)	−121±10	23	−109±4	12	−114±5	14	−132±8	20
	A(16)	−89±15	40	−116±18	41	−105±13	33	−121±20	56
	G(17)	−79±11	26	−103±25	75	−101±18	50	−119±23	64
	G(18)	−121±24	56	−76±15	39	−87±15	43	−110±23	66
	Average¶	10	26	14	37	10	29	13	36
Pseudorotation (°)	C(3)	145±7	18	140±26	73	137±18	53	144±34	110
	C(4)	152±5	12	181±5	14	149±30	73	149±13	33
	T(5)	105±38	85	183±8	25	196±26	85	125±20	51
	A(6)	149±29	62	157±28	70	142±14	35	156±11	29
	A(7)	116±23	55	170±6	15	163±13	39	163±12	28
	T(8)	167±26	67	104±13	37	96±12	38	107±6	22
	A(13)	148±20	49	211±12	35	177±15	40	167±19	57
	T(14)	134±26	62	88±7	19	88±7	24	111±40	118
	T(15)	135±25	63	91±4	12	105±6	17	126±12	33
	A(16)	149±15	41	132±33	74	125±21	66	145±18	59
	G(17)	165±15	35	176±12	33	158±17	47	156±14	41
	G(18)	157±15	38	242±61	201	191±65	185	143±33	90
	Average¶	20	49	18	51	20	58	19	56

† Only 1 residue of a base-pair is listed. Thus, C(3) represents the base-pair C(3)–G(18); C4 represents the base-pair C(4)–G(17), etc.
‡ The average of the 5 DG structures are listed with their standard deviation.
§ The range is defined as the difference between the highest and lowest values of a given parameter in the set of 5 or 7 structures.
‖ The average of the 5 DG structures and the *A* and *B*-structures are listed with their standard deviation.
¶ The average of the standard deviation or range for the 6 base-pairs is given for each set of structures.

deviations. Furthermore, it is common to find distances that violate the input distance constraint range by as much as 0·5 Å (Wüthrich, 1986). Such a violation in the distance constraints could then lead to volume differences of greater than 650%. Therefore, it should be clear that the variation in the volumes of the spectra shown here are very small relative to those obtained by more standard methods. However, as will be discussed below, even with these good fits to the experimental data, our structures still show quite large variations in many local conformational parameters.

Figure 7(a) and (b) show stereo views of the superposition of the central six base-pairs of the refined *A*, refined *B* and five refined DG structures of decamers 1 and 2, respectively. These structures were superimposed as the best fit of the C-8, C-6 and C-1′ atoms of the central six base-pairs. It is clear that there is a significant amount of conformational variation among these structures. The variation is much larger than was observed in structures generated from simulated n.m.r. data (Pardi *et al.*, 1988). Tables 6 and 7 show the conformational parameters for the refined DG structures of decamer 1 and 2, respectively. The average standard deviation and average range for almost all of these parameters increase in going from the initial DG structures to the refined DG structures (despite the fact that the variation in the volumes decreases). For example, in decamer 1 the average standard deviation and range of helical twist in the five DG structures increases by 68 and 100%, respectively, in going from the initial DG structures to the refined DG structures. This result might seem surprising at first given the fact that the iterative refinement generates an improved fit of the calculated and experimental cross peak volumes. However, it is important to realize that the distance constraints for the initial DG structures were generated assuming a simple two-spin approximation and then entered assuming that the distance could be defined very precisely (±0·2 or ±0·3 Å; Wang, 1989). Thus, the larger variations in the conformational parameters for the refined structures simply reflect the fact that for the refined DG structures we have directly accounted for spin diffusion effects and have used a more conservative set of distance constraints than was employed in the initial DG structures.

(b) *The role of the distance constraints in the refinement process*

The standard method for generating three-dimensional structures of biomolecules is to estimate proton–proton distance constraints from experimental NOE data and then use these constraints as input for calculations that generate structures consistent with these data (Wüthrich, 1986). In this method, the residual violations between the

Table 7
Conformational parameters for decamer 2

		Initial structures		Refined structures		Constrained energy minimized structures		Constrained quenched dynamics structures	
	Residues†	Average‡	Range§	Average‖	Range§	Average‖	Range§	Average‖	Range§
Helical rise (Å)	T(3)–C(4)	3·5±0·1	0·3	3·6±0·6	1·6	3·5±0·5	1·6	3·7±0·5	1·2
	C(4)–A(5)	3·4±0·5	1·2	3·7±0·4	1·1	3·6±0·3	0·8	3·9±0·3	1·2
	A(5)–C(6)	2·4±0·4	1·0	3·5±0·4	1·2	3·5±0·4	1·2	3·9±0·3	0·9
	C(6)–G(7)	3·5±0·2	0·4	3·3±0·2	0·5	3·5±0·2	0·5	3·7±0·2	0·9
	G(7)–C(8)	3·1±0·6	1·5	3·3±0·3	0·7	3·5±0·2	0·6	3·6±0·4	1·0
	Average¶	0·4	0·9	0·4	1·0	0·3	0·9	0·3	1·0
Helical twist (°)	T(3)–C(4)	22±6	14	34±5	13	35±5	16	38±4	13
	C(4)–A(5)	32±4	10	38±4	13	37±4	10	36±5	15
	A(5)–C(6)	24±2	5	32±4	11	31±4	12	29±4	11
	C(6)–G(7)	44±3	7	39±5	13	38±4	10	35±6	17
	G(7)–C(8)	32±7	17	27±4	10	27±3	7	26±3	9
	Average¶	4	11	4	12	4	11	4	13
Roll (°)	T(3)–C(4)	16±2	6	6±8	23	7±8	24	5±3	9
	C(4)–A(5)	14±6	14	7±5	15	6±3	6	9±4	12
	A(5)–C(6)	2±1	3	3±2	5	4±2	4	3±3	8
	C(6)–G(7)	8±3	7	16±14	40	11±10	28	15±12	32
	G(7)–C(8)	8±6	18	3±12	7	2±2	5	3±3	8
	Average¶	4	10	6	18	5	13	5	14
Tilt (°)	T(3)–C(4)	2·7±2·4	6·3	4·7±3·5	11·0	4·1±2·7	7·4	6·8±2·2	6·2
	C(4)–A(5)	3·9±1·5	3·8	3·2±2·9	8·8	2·4±1·4	3·8	3·0±2·7	8·2
	A(5)–C(6)	1·0±0·6	1·5	3·5±2·8	7·9	2·1±1·2	3·7	1·5±1·6	4·3
	C(6)–G(7)	3·3±2·2	6·0	5·1±3·3	9·3	1·6±1·8	5·1	3·0±1·8	4·2
	G(7)–C(8)	2·7±2·2	4·8	2·9±1·1	3·3	1·4±0·8	2·8	1·3±0·6	1·4
	Average¶	1·8	4·5	2·7	8·1	1·6	4·5	1·8	4·9
Displacement (Å)	T(3)	1·4±0·7	1·8	1·1±1·1	3·1	1·4±0·8	2·5	1·7±1·0	3·0
	C(4)	1·1±1·0	2·4	1·1±0·6	1·7	1·1±0·6	2·0	1·3±0·6	1·5
	A(5)	0·8±0·3	0·9	1·6±0·8	2·6	1·6±0·9	2·6	1·7±0·8	2·3
	C(6)	1·0±0·5	1·2	1·4±0·8	2·2	1·3±0·8	2·5	1·3±0·9	2·6
	G(7)	0·5±0·5	1·2	0·8±0·6	1·5	0·9±0·7	2·1	0·7±0·5	1·2
	C(8)	1·3±1·2	2·5	2·1±0·7	1·9	1·9±1·1	3·1	0·9±0·9	2·6
	Average¶	0·7	1·7	0·8	2·2	0·8	2·5	0·8	2·2
Propeller twist (°)	T(3)	27±9	26	22±11	35	24±7	17	25±11	29
	C(4)	17±5	14	21±9	29	21±7	21	19±7	23
	A(5)	21±7	19	25±11	28	14±6	17	15±5	16
	C(6)	16±5	14	22±5	15	12±6	18	17±5	14
	G(7)	16±4	9	20±7	17	11±5	14	12±5	13
	C(8)	16±7	17	18±6	18	11±3	8	10±8	22
	Average¶	6	16	8	24	6	16	7	19
δ Torsion angle (°)	T(3)	153±5	13	161±42	42	152±14	42	136±14	38
	C(4)	87±5	10	88±7	22	94±23	64	115±14	38
	A(5)	123±19	52	126±12	39	142±7	24	134±7	20
	C(6)	115±6	14	113±33	73	128±12	34	125±12	32
	G(7)	118±2	6	144±11	32	144±14	35	137±12	31
	C(8)	112±7	15	112±20	57	136±25	65	127±19	53
	G(13)	118±17	40	119±23	72	117+30	92	110±11	31
	C(14)	104±5	13	120±34	79	135±11	30	132±15	43
	G(15)	105±9	22	135±20	61	141±12	32	133±11	31
	T(16)	102±2	6	116±21	49	108±23	56	117±8	21
	G(17)	142±12	29	152±10	29	158±11	33	153±4	11
	A(18)	95±10	28	161±10	31	144±13	37	148±12	34
	Average¶	8	21	18	49	16	45	12	32
Glycosidic angle (°)	T(3)	−67±23	54	−84±16	46	−93±12	39	−113±10	25
	C(4)	−84±7	18	−89±21	60	−97±14	42	−108±12	38
	A(5)	−100±6	14	−95±14	41	−96±7	24	−106±10	29
	C(6)	−154±7	16	−141±8	24	−134±13	34	−135±8	27
	G(7)	−112±12	29	−118±8	21	−108±11	33	−108±8	26
	C(8)	−137±6	16	−130±11	28	−114±13	45	−117±14	39

Table 7 *(continued)*

		Initial structures		Refined structures		Constrained energy minimized structures		Constrained quenched dynamics structures	
	Residues†	Average‡	Range§	Average‖	Range§	Average‖	Range§	Average‖	Range§
	G(13)	−102±16	37	−110±25	60	−106±10	26	−113±7	19
	C(14)	−131±11	29	−153±12	32	−131±7	18	−133±13	39
	G(15)	−107±9	22	−143±16	49	−114±15	44	−114±11	30
	T(16)	−113±5	13	−160±9	28	−140±9	24	−140±11	34
	G(17)	−82±33	75	−97±25	62	−87±17	48	−99±11	32
	A(18)	−99±16	44	−79±7	19	−90±8	22	−98±5	13
	Average¶	13	31	14	39	11	33	10	29
Pseudorotation angle (°)	T(3)	192±25	62	185±25	82	167±16	43	152±14	38
	C(4)	83±7	16	105±8	25	105±21	60	122±10	24
	A(5)	137±31	83	143±10	33	140±6	15	136±6	15
	C(6)	118±8	20	134±34	78	131±20	60	128±10	27
	G(7)	126±5	13	160±12	34	151±9	29	140±7	18
	C(8)	114±10	25	127±20	58	139±19	55	133±12	33
	G(13)	127±26	63	142±39	113	131±39	99	118±9	22
	C(14)	106±8	18	134±28	67	138±17	44	132±13	37
	G(15)	108±15	36	146±16	46	139±9	26	137±13	39
	T(16)	107±2	6	137±33	71	113±20	47	123±10	25
	G(17)	180±31	76	174±13	36	177±13	31	168±9	26
	A(18)	97±15	41	188±7	19	166±15	39	171±16	45
	Average¶	15	38	20	55	17	46	11	29

† Only 1 residue of a base-pair is listed. Thus, T(3) represents the base-pair T(3)–A(18); C4 represents the base-pair C(4)–G(17), etc.
‡ The average of the 5 DG structures are listed with their standard deviation.
§ The range is defined as the difference between the highest and lowest values of a given parameter in the set of 5 or 7 structures.
‖ The average of the 5 DG structures and the *A* and *B* structures are listed with their standard deviation.
¶ The average of the standard deviation or range for the 6 base-pairs is given for each set of structures.

distances in a given structure and input distances are used to judge the goodness of fit. In the back-calculation procedure employed here, the optimization involves fitting the cross peak *volumes* of the experimental and calculated 2D NOE spectra. Thus, the distance constraints in this procedure only serve as a driving force for optimizing the fit of the experimental and calculated spectra. Because it is the volume that is being fit, and not the distance constraints, it is not relevant whether a particular distance constraint is satisfied as long as the calculated cross peak volume agrees well with the experimental cross peak volume. For this reason a comparison of the distance violations in the various generated structures, as is commonly done in the standard structure generation process, is not meaningful. Instead, comparisons should be made between the experimental and calculated spectra as shown in Figures 6 and 7. As will be discussed below, a potential improvement in the procedure used here would be to perform the refinement directly against the volumes (instead of distance constraints), and preliminary results of such a procedure have been reported (Boelens *et al.*, 1989).

For the constrained energy minimization and quenched molecular dynamics simulations performed here, the distance constraints play a more standard role. As discussed in Materials and Methods, these calculations employed optimization by fitting to a different set of input distance constraints for each structure, where the distances were measured from each of the refined *A*, refined *B* and five refined DG structures. This procedure was used so that the interproton distances in these structures, which give rise to good fits to the experimental data, changed as little as possible, thus ensuring that the constrained energy minimization and quenched dynamics generated structures also have good fits to the experimental data. As seen for decamer 1, in Figure 5(g) and (h) (and for decamer 2 in Fig. 6(g) and (h)), the constrained energy-minimized and quenched dynamics structures still fit well to the experimental data, even though they were refined by optimization of distance constraints.

(c) *Relative energies of the decamer structures*

Table 4 gives the total potential energies (in kcal/mol) for the refined, constrained energy-minimized and constrained quenched dynamics structures for the *A*, *B* and DG refinements for both decamers. The refined structures have large total potential energies ($>10^3$ kcal/mol), with essentially all of this energy arising from the repulsive term in the Lennard–Jones potential. This effect is caused by slightly different templates (primarily bond lengths and bond angles) in the DSPACE and IMPACT programs, which leads to large energies for the r^{-12} term in the Lennard–Jones potential. This high energy is eliminated in less than 50 cycles of mini-

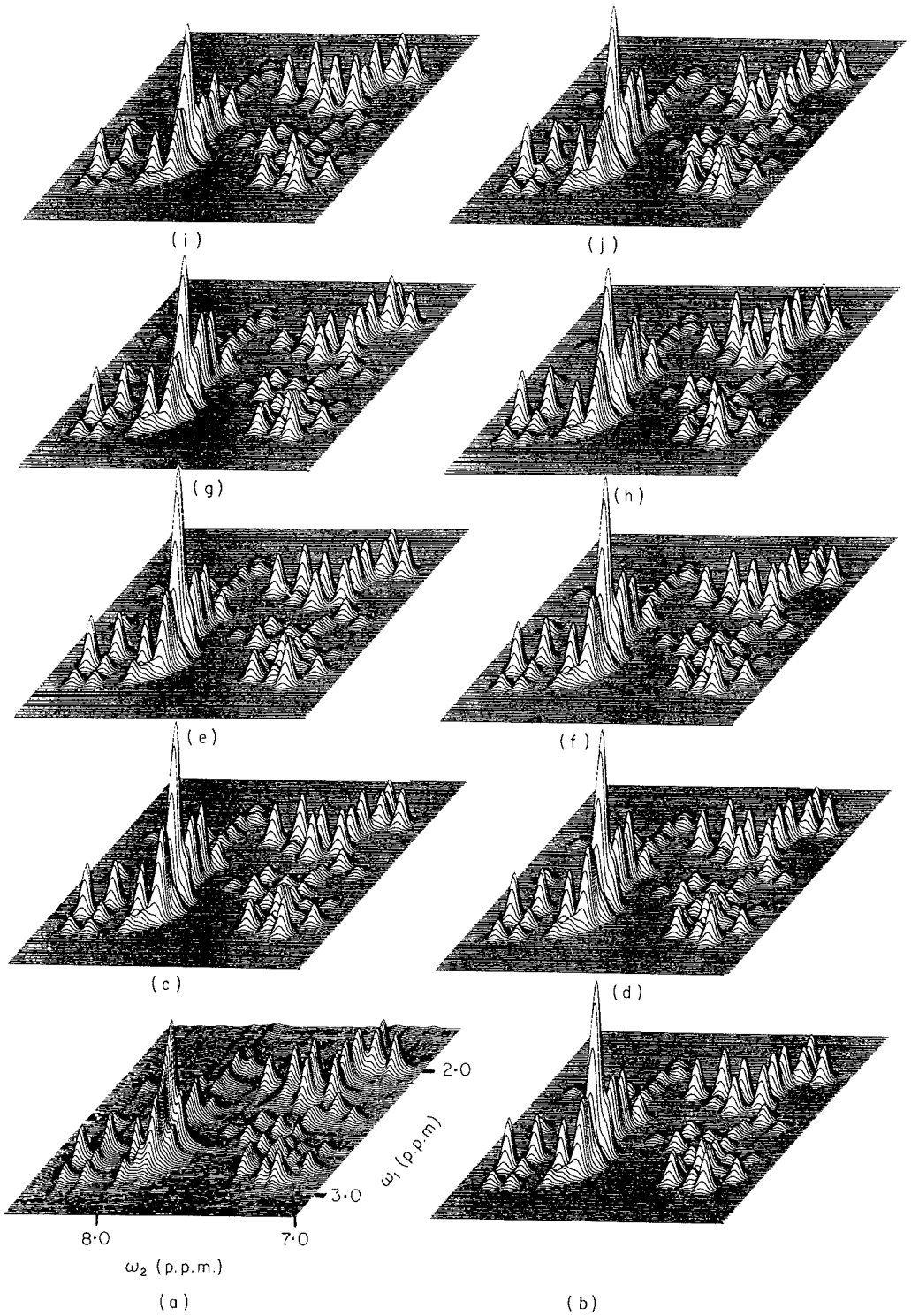

Figure 6. Stack plots of the H-6/H-8 to H-2′/H-H2″ region for the structures of decamer 2: (a) the experimental NOE data (120 ms); (b) to (f) the calculated spectra for the final refined DG structures 1 to 5; (g) the calculated spectrum for the final refined *A* structure; (h) the calculated spectrum for the final refined *B* structure; (i) the calculated spectrum for the constrained minimized structure of DG 1; (j) the calculated spectrum of the constrained dynamics structure of DG 1.

mization, by which point the total energy from the Lennard–Jones term is negative. Table 4 also gives energies corresponding to the distance constraint violations for these different structures indicating that the distance constraints are well fitted in all sets of structures.

(d) *Conformational analysis of the constrained energy minimized structures for the decamers*

Figure 8 shows a stereo view of all ten base-pairs of one of the constrained energy minimized DG structures of decamer 2. Throughout this study

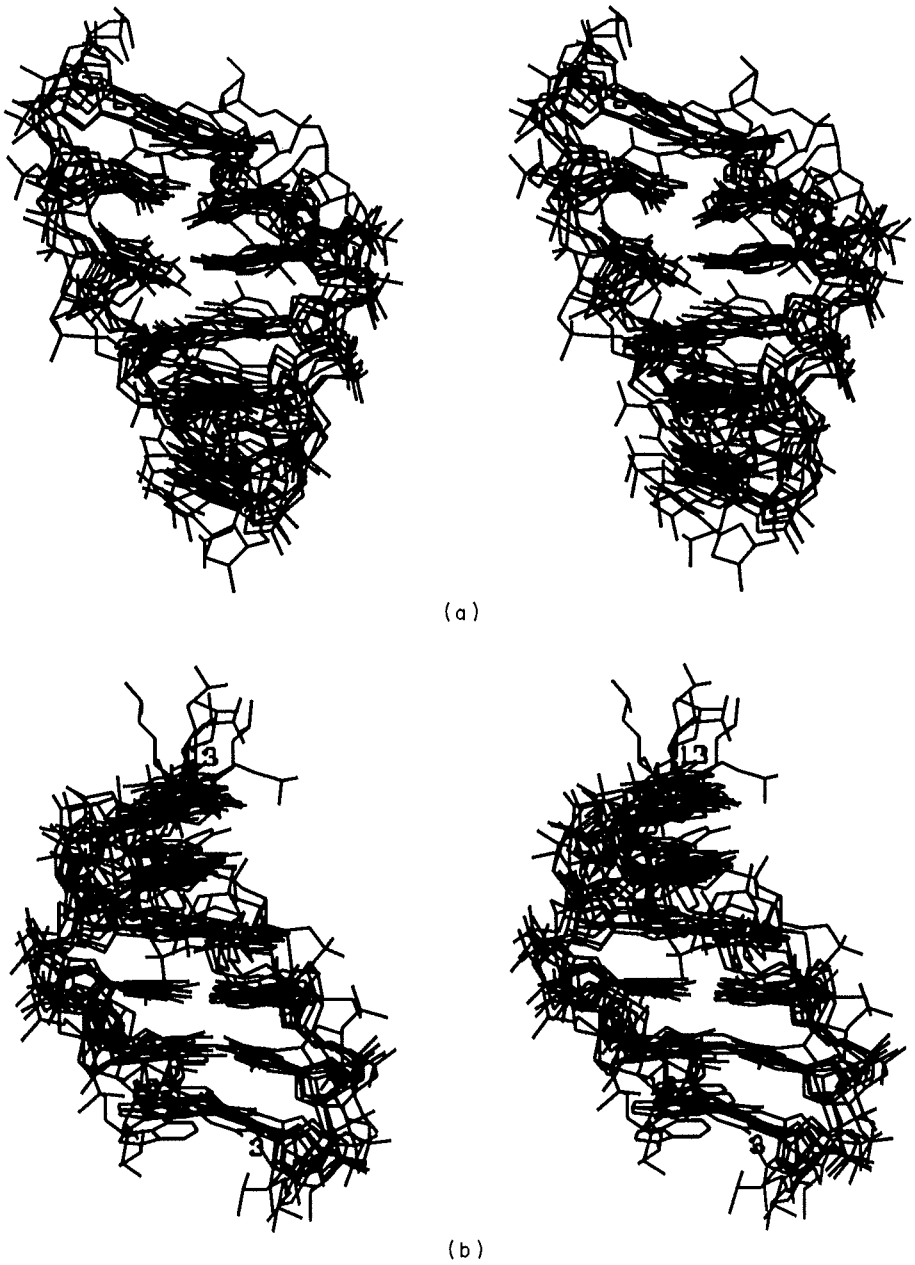

Figure 7. Stereo view of the superposition of the central 6 base-pairs of the constrained energy-minimized *A*, *B* and 5 refined DG structures for (a) decamer 1, and (b) decamer 2. Superposition was performed using only C-1′, C-6 and C-8 atoms of the central 6 base-pairs.

it was observed that the two terminal base-pairs are subject to larger conformational variations than the internal base-pairs, and one sees quite abnormal geometries for the end base-pairs in Figure 8. Our previous simulation studies using an optimal n.m.r. data set also indicated that the terminal base-pairs could not be precisely and accurately defined (Pardi *et al.*, 1988), and so here we only analyzed the central six base-pairs in both decamers.

Table 6 gives complete data on the conformational parameters for the various structures of decamer 1, but the trends in these data are more readily seen in Figure 9(a) to (h), which shows plots of the variation with base sequence for helical rise, helical twist, magnitude of the roll, magnitude of the displacement, propeller twist, δ torsion angle, glycosidic angle and pseudorotation angle, respectively, for the average of the constrained energy-minimized *A*, *B*, and five DG structures of decamer 1. In addition the plots also show the range of values observed in these seven structures, as well as the total range for each parameter that has been observed in X-ray studies of double-stranded DNA (Saenger, 1984). We have chosen to analyze the ranges for these structures, rather than the standard deviations since we are searching for *all* conformations consistent with the experimental data and there is no reason to assume that the ensemble of

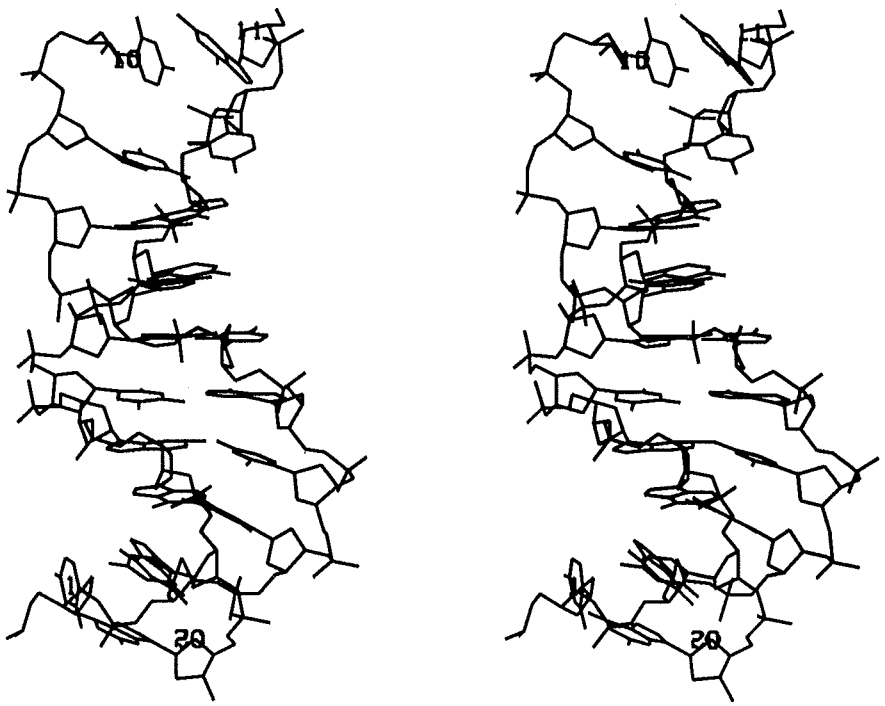

Figure 8. Stereo view of a constrained energy-minimized DG structures of decamer 2, showing an example of the abnormal geometry for the end base-pairs and variable phosphate conformations that are consistent with the experimental spectra.

structures consistent with the data is a Gaussian distribution of the conformational parameters. Thus, the standard deviation may not necessarily reflect the true distribution of the conformational parameters. However, this parameter is still given in Table 6 so that these results can be compared with previous studies.

Figure 9 shows that the range of helical rise, helical twist and propeller twist that are consistent with back-calculation of the 2D NOE data and energetics of decamer 1, on average span approximately the *whole range of values* for these parameters that have been observed in right-handed DNA double-strands (given by the broken horizontal lines in the Fig.). Thus, it is clear that using the data and methods discussed here, such as refinement against the original NOE data, these parameters cannot be defined precisely for this decamer. Figure 9 and Table 6 show that the displacement and the δ torsion angle, can only be defined to, on average, no better than 50% of the range of values observed from X-ray studies of double-stranded DNAs, whereas the glycosidic angle and pseudorotation angle are quite well defined in these structures relative to the observed range in X-ray studies. It is important to note that we have somewhat arbitrarily defined the "relative" precision with which a given parameter can be defined, by taking the ratio of the average deviation for each parameter in the refined structures to the total range of values for this parameter that have been observed in X-ray studies of double-stranded DNAs (for a more extensive discussion, see Pardi *et al.*, 1988). Thus, two parameters that are defined with the same numerical precision can have distinctly different classifications as to how well they are defined. For example, if both the pseudorotation angle and δ torsion angle have numerical precisions of approximately 45°, their relative precisions are 15% *versus* 56%, respectively, due to a range of 300° for the pseudorotation angle but only 80° for the δ torsion angle.

Similar trends to those discussed above for decamer 1 are observed for decamer 2, as shown in Table 7 and Figure 10. Again this is most easily seen in Figure 10, where the range of helical twist, helical rise and propeller twist in constrained energy-minimized DG structures also span a large percentage of the range of values observed in double-stranded DNA.

(e) *Constrained quenched dynamics structures*

As seen in Figure 8, and from analysis of the α, β, γ, ε and ζ backbone torsion angles (data not shown), the conformation of the phosphate groups are not well determined in these decamers. This is expected from the results of computer simulations of DNA using a set of distance constraints that mimic an optimal n.m.r. data set (Pardi *et al.*, 1988). In both the structures generated here and in the simulations, the phosphate backbone appears to exist in local "pseudomirror image" geometries. The next step in the structure refinement procedure was to perform constrained quenched dynamics calculations on both decamers to help search for the lowest

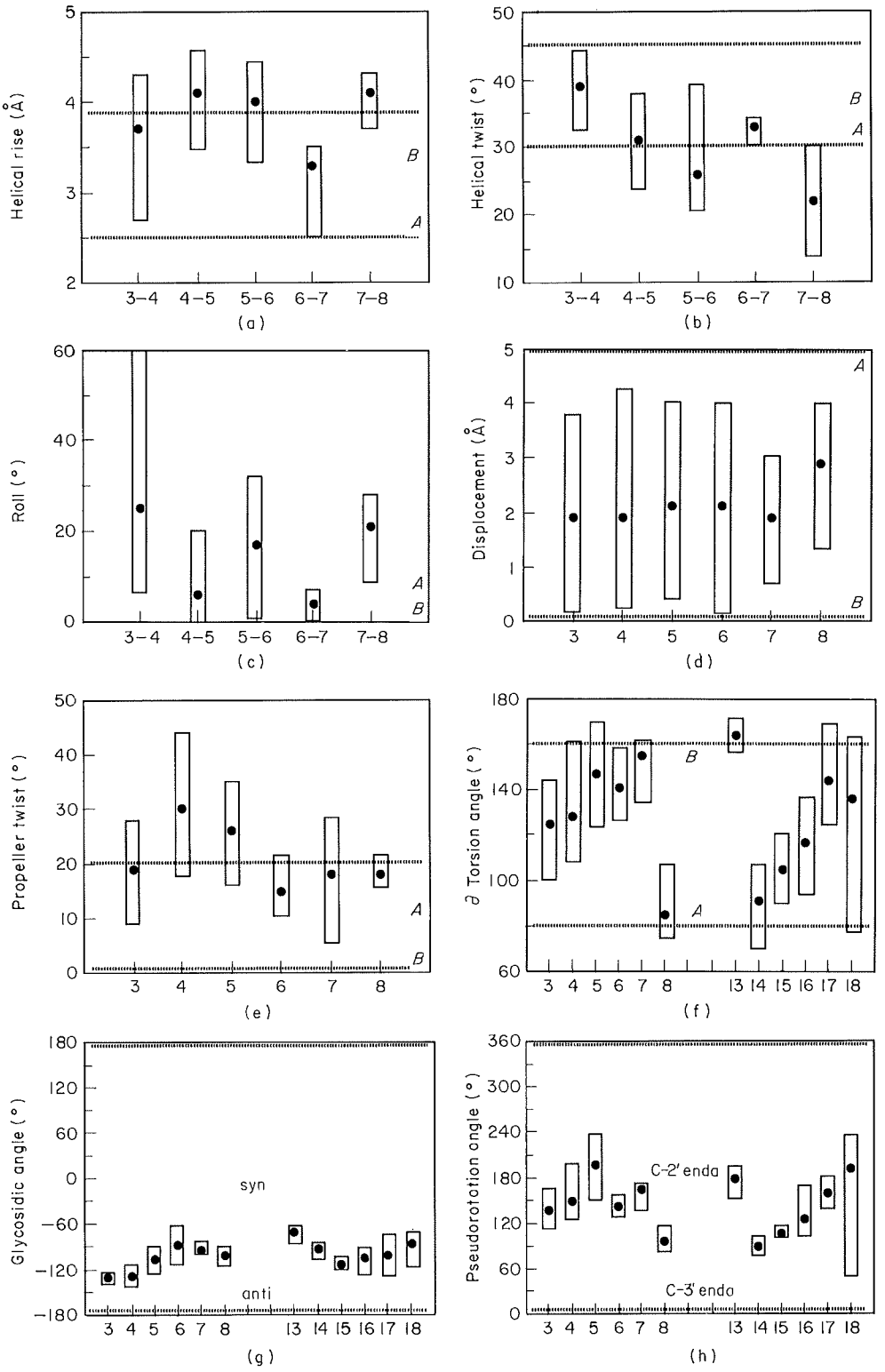

Figure 9. Plots of the conformational parameters for constrained energy-minimized structures of decamer 1: (a) helix rise; (b) helix twist; (c) roll (magnitude); (d) displacement (magnitude); (e) propeller twist; (f) δ torsion angle; (g) glycosidic angle; (h) pseudorotation angle. The broken horizontal lines indicate the range of values observed for a given parameter in X-ray data of right-handed DNA (see the text; Pardi *et al.*, 1988). The bars shows the range observed for these parameters in the 5 constrained energy-minimized DG, the constrained energy-minimized *A* and constrained energy-minimized *B* structures. The filled circle indicates the average value of the parameter for these 7 structures.

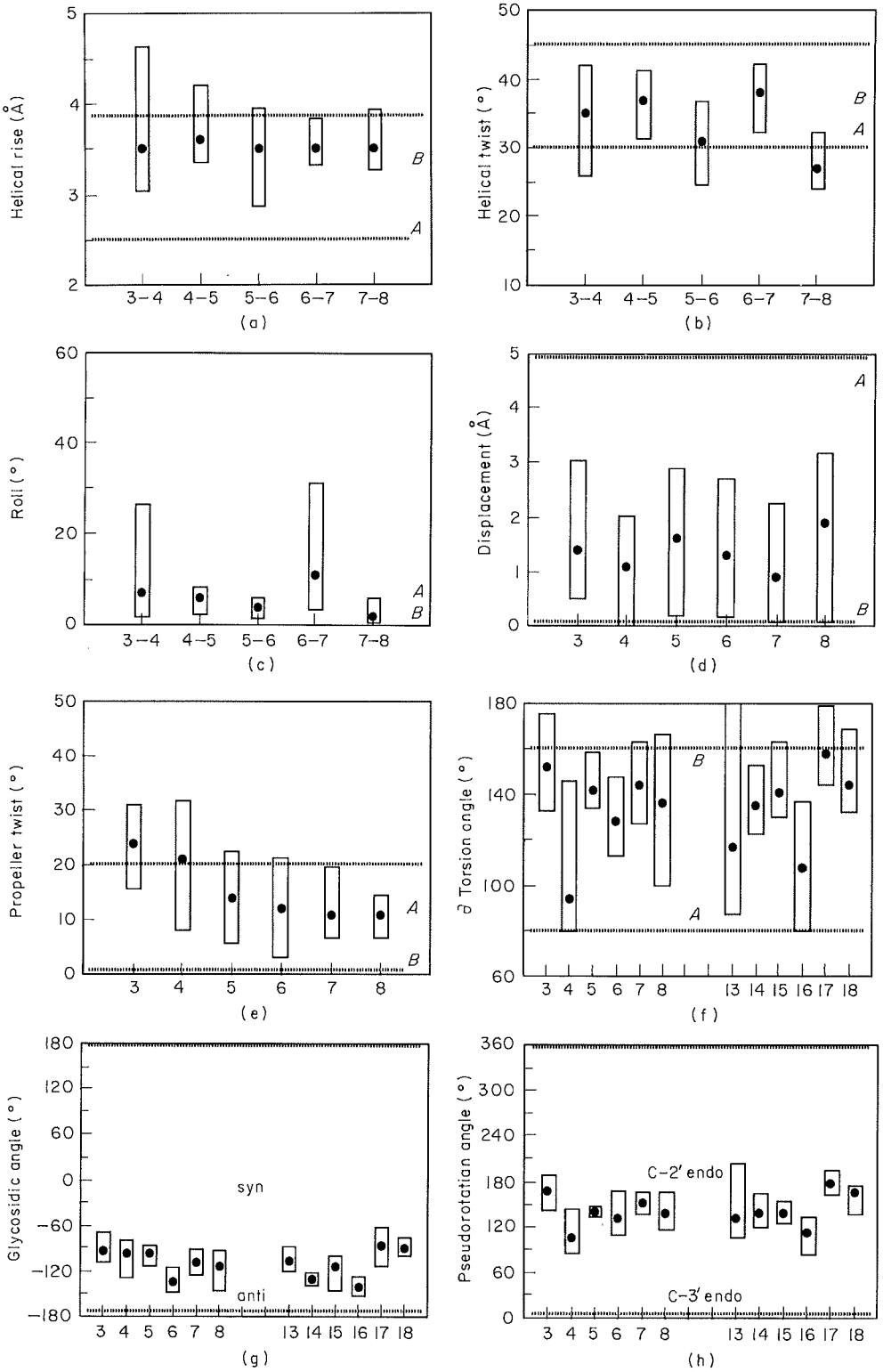

Figure 10. Plots of the conformational parameters for constrained energy-minimized structures of decamer 2: (a) helix rise; (b) helix twist; (c) roll (magnitude); (d) displacement (magnitude); (e) propeller twist; (f) δ torsion angle; (g) glycosidic angle; (h) pseudorotation angle. The broken horizontal lines indicate the range of values observed for a given parameter in X-ray data of right-handed DNA (see the text, Pardi *et al.*, 1988). The bars show the range observed for these parameters in the 5 constrained energy-minimized DG, the constrained energy minimized *A* and constrained energy minimized *B* structures. The filled circles indicates the average value of the parameter of these 7 structures.

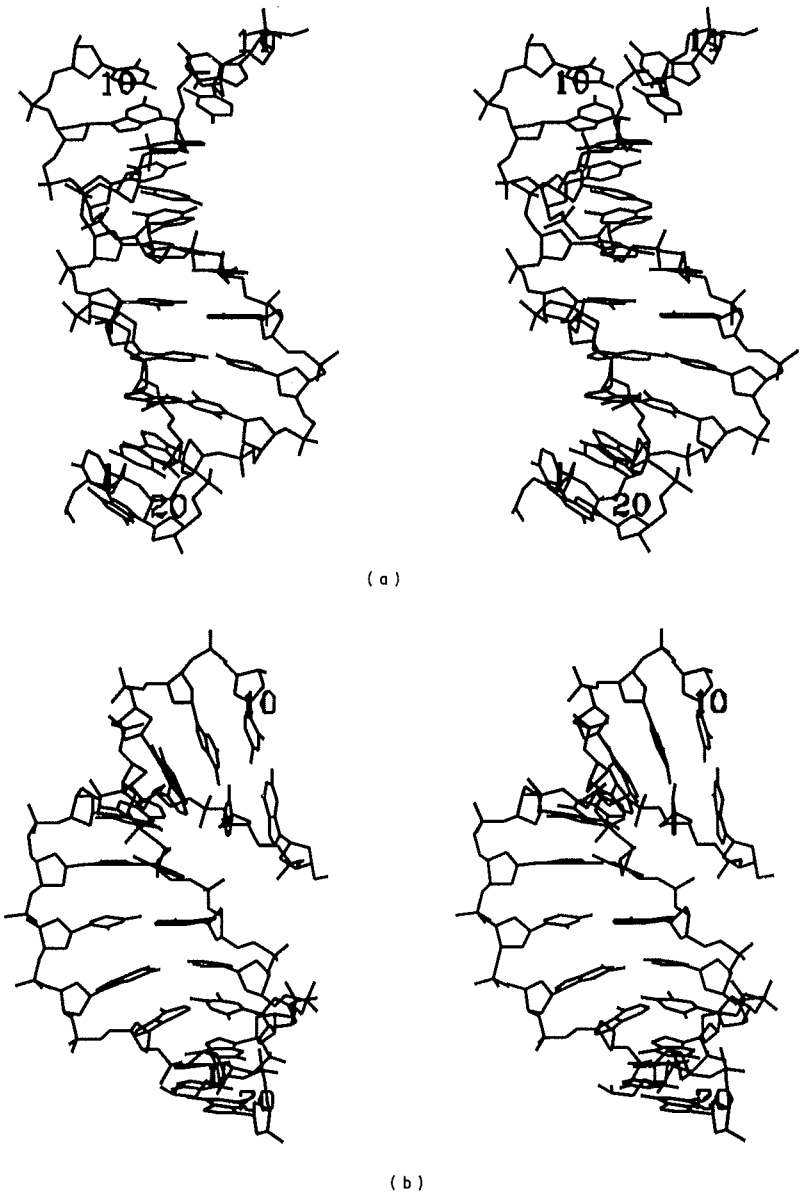

Figure 11. (a) Stereo view of the constrained energy-minimized structure of decamer 1 that was used as input for the long constrained quenched dynamics calculations. This structure corresponds to the constrained energy-minimized DG 1 structure. (b) Stereo view of this structure generated by the long constrained quenched dynamics calculation (see the text).

energy phosphate conformation. If a particular phosphate conformation is energetically much more favorable, then the observed disorder could be eliminated based on energetic criteria.

Constrained quenched dynamics calculations were performed as described in Materials and Methods and, as shown in Table 5, the average r.m.s. deviations among the seven structures for each decamer are very similar to those for the constrained energy-minimized structures. In addition, the energies for the dynamics structures are quite similar to the constrained minimized structures (see Table 4). Analysis of the range and standard deviations of the conformational parameters (Tables 6 and 7) also indicates that the constrained quenched dynamics procedure does not significantly improve the precision with which the local conformation of DNA can be determined from n.m.r. distance data. Since these calculations were performed to try to improve the precision of the phosphate backbone torsion angles, analysis of the seven constrained quenched dynamics structures for both decamer 1 and decamer 2 shows that these calculations do not lead to better determined conformations for the phosphate groups (data not shown).

(f) *Additional constrained molecular dynamics calculations*

The constrained quenched dynamics calculations were performed to search for energetically favorable structures consistent with input distance

constraints. This method has been used (Nilsson *et al.*, 1986; Happ *et al.*, 1988; Clore *et al.*, 1988) to generate structures of DNA oligomers where the dynamics calculations were performed for much longer times (approx. 15 to 20 ps) compared to the 1·5 picoseconds used here. In order to test how the length of the dynamics calculations affect the generated structures, we performed a series of constrained quenched dynamics calculations on decamer 1 and 2 where the molecules were "heated" to 1500 K for eight picoseconds (with a step size of 0·5 fs), and then progressively cooled for 2·5 picoseconds each to 1000 K, 500 K and 100 K. The 100 K structures were then subjected to 500 cycles of constrained energy minimization. The initial and final structures from this "long" constrained dynamics calculation for decamer 1 are given in Figure 11(a) and (b), respectively. The structure generated in the long quenched dynamics structure has a number of features similar to *A*-form geometry, including a large displacement (on average 3·7 Å) and a small helical rise (approx. 2·9 Å). The energy of this structure is 25% lower than the starting structure and the NOE violations are also slightly reduced. In addition, this structure still fits well to the experimental NOE data (data not shown), but has a significant r.m.s. deviation (approx. 4·0 Å) from the energy-minimized or constrained quenched dynamics structures of decamer 2. Additional calculations with simulated data (data not shown) indicate that it is the energy potentials (most likely the electrostatic potential) and not the distance information that is driving the structures to this "*A*-like" geometry, and we are presently investigating the source of the forces that are driving the structures to this alternate conformation (W.J.M., R.M.L & A.P., unpublished results). However, for the purposes of the discussion here these results clearly show that longer constrained quenched dynamics can generate a non-*B*-type structure that is *still consistent* with the experimental data. These preliminary data indicate that the results presented here may actually *underestimate* the range of structures that are consistent with NOE data.

(g) *Limits on determining local structural variations of DNA in solution by n.m.r. data*

We have performed computer simulations to determine how accurately and precisely local structural parameters can be determined from a set of proton–proton distances that mimic an optimal n.m.r. data set (i.e. a set of perfectly accurate and highly precise (±0·2 Å) distance constraints; Pardi *et al.*, 1988). In this simulation, it was shown that while the α, β, γ, ε and ζ backbone torsion angles cannot be determined precisely from an optimal set of n.m.r. distance constraints, many other conformational parameters could be defined precisely and accurately in the simulations. Unfortunately, it is difficult to determine directly the accuracy and precision with which the experimental distance constraints can be measured, so here, we generated a set of structures by refining against the original 2D NOE data, starting with a set of rather conservatively (relative to previous studies; Nilsson *et al.*, 1986; Happ *et al.*, 1988; Clore *et al.*, 1988) defined distance constraints. The only reason that the value and/or precision of a particular distance constraint was modified was to improve the fit between the calculated and experimental 2D NOE data. Therefore, by using this iterative back-calculation refinement procedure we tried to avoid bias introduced by making an arbitrary assumption about the precision with which distance constraints can be measured. The next step was to include energy terms in the refinement procedure to produce energetically favorable structures that were consistent with the original n.m.r. data. We found that using n.m.r. distance data and the presently available computational techniques, the conformational parameters of helical twist, helical rise, propeller twist, roll, the α, β, γ, ε and ζ torsions angles and probably the tilt, δ torsion angle and displacement, cannot be determined precisely enough to justify comparisons of local structural variations with base sequence. In many cases, a parameter can only be defined with a precision that is similar to the whole range of values that has ever been observed for this parameter in X-ray structures of right-handed DNA duplexes (ranging from *A* to *D*-form DNAs; Saenger, 1984).

The structural analysis performed here, as well as, essentially, all n.m.r. structure determinations of DNA duplexes of which we are aware, assume that the DNA oligomer exists in a single static conformation in solution. However, it is more likely that the molecule actually exists in multiple conformations in solution, with rapid interconversion among the various conformations. Borgias & James (1988) give an excellent discussion of the problems associated with determining the structure of a DNA duplex where the individual sugar residues are in rapid equilibrium between S and N sugar puckers. As these authors point out, attempts to fit the NOE data to a population of exchanging conformations would only serve to add many more parameters to an already underdetermined system, and therefore would lead to a decrease in precision with which any given parameter could be defined. The end result of such an analysis would be that the DNA structures would be even more poorly defined than those given here.

(h) *Comparison of these decamer structures with previously determined n.m.r. solution structures of DNA oligomers*

A number of workers have used 2D NOE data and computational techniques to generate detailed solution structures of DNA duplexes (for reviews, see Reid, 1987; Patel *et al.*, 1987; Van der Ven & Hilbers, 1988). In some of these studies, the conclusions of how well local structures of DNA can be determined are quite different from those made here

and we will comment on possible explanations for these differences.

Reid and co-workers have generated solution structures of a DNA oligomer containing the *Bcl*I restriction site (Banks *et al.*, 1989), and one containing the *Eco*RI restriction site (Nerdal *et al.*, 1989) using a procedure very similar to that performed here. In their studies a number of initial structures (8 for the *Bcl*I restriction site) were generated by the DSPACE DG algorithm, but *only one* of these structures for each oligomer was then refined by minimizing deviations between the experimental NOEs and NOEs calculated from the model structure. These studies showed that use of the back-calculation refinement procedure enables one to generate a model structure which fits quite well to the NOE build-up rates. However, the fact that a particular model structure fits the experimental data well does not mean that this structure is unique, since it is possible that there are many different structures which fit well to the experimental data. Therefore, we chose to perform the back-calculation refinement procedure on all seven of our initial starting structures. The results here show that there are very large local structure variations in a set of structures all of which have calculated NOE spectra that are equally consistent with the experimental data for a single mixing time. Further studies will need to be done to see if refinement of a set of structures by fitting to the build-up of the NOE will lead to more precisely defined local conformation in DNA duplexes.

Clore, Gronenborn and co-workers have performed a number of n.m.r. structural studies on DNA duplexes, and come to quite different conclusions as to how precisely and accurately local structural parameters can be determined from NOE distance data (Happ *et al.*, 1988; Clore *et al.*, 1988, and references cited therein). These different conclusions appear to arise from different assumptions made by their group and ours in the procedures used for generating structures. The different assumptions include the following. (1) We have refined our structures against the experimental data using a procedure that accounts directly for spin diffusion, whereas their analysis assumed that spin diffusion does not affect the measured distances at the mixing times that they have used. (2) They have assumed a quite high precision ($-0{\cdot}2$ to $+0{\cdot}3$ Å for distances <3 Å and $-0{\cdot}3$ to $+0{\cdot}4$ Å for distances between 3 and 5 Å) for their input distance constraints, whereas we have initially used relatively much more conservative distance constraints (see Materials and Methods) and only modified the constraints when required to fit the experimental and back-calculated data. (3) In some of their recent studies they have included torsion angle constraints to help generate better defined structures.

We will now address possible effects of these different assumptions. Several studies have shown that neglecting spin diffusion effects for molecules of the size studied here can lead to an inaccurate set of distance constraints (Borgias & James, 1988; Boelens *et al.*, 1989; Madrid *et al.*, 1989). This is the reason why we adopted the back-calculation refinement procedure that accounts directly for spin diffusion effects. We have started from a relatively conservative set of distance constraints and, as discussed previously, during the refinement we fit to the experimental 2D NOE spectra, where the distance constraints serve only as a driving force during the refinement. Gronenborn & Clore (1989) generated structures of a DNA duplex by varying the precision with which the input distance constraints were defined. Their results show that the variation of the structures is affected directly by the precision of the input distance constraints. Thus, the differences between the work presented here and that of Clore and co-workers (Happ *et al.*, 1988; Clore *et al.*, 1988 and references cited therein), concerning how well the local conformation can be determined, is at least partially due to the effects of spin diffusion and different assumptions for the precision of the input distance constraints.

Another important difference in the work here and that of Clore and co-workers concerns their assumption of backbone torsion angle constraints. For example, in a recent study they have used the measured $^3J_{1'2'}$ coupling constants for the sugar protons in a DNA duplex to constrain the δ torsion angle to a range of 110 to 170° (Clore *et al.*, 1988). Although the addition of such coupling constant data is potentially valuable for improving the quality of nucleic acid structures, as seen in Figures 9 and 10 and Tables 6 and 7, the δ torsion angle is already generally defined within this range by the refinement procedure performed here using NOE distance data. Thus, the addition of this constraint will not significantly change the results presented here. Probably one of the crucial differences between these two studies arise because in some recent studies (Gronenborn & Clore, 1989), but apparently not in earlier ones (Nilsson *et al.*, 1986; Happ *et al.*, 1988), Clore and co-workers have constrained the α, β, γ, ε and ζ torsion angles to the rotamer conformations observed in standard *A* and *B*-form geometries. We have chosen not to make such an assumption since there is no *experimental evidence* to constrain these angles in solution. In fact, there is direct evidence from X-ray crystal studies that the phosphate backbone is variable (Sakked *et al.*, 1989; Dickerson, 1988; Jain & Sundarlingam, 1989). For example, in the structure of a *B*-DNA dodecamer, Dickerson and co-workers found that the phsophate backbone existed in both a $B_{\rm I}$ or $B_{\rm II}$ conformation, which arose from different rotamer conformations for the ε and ζ torsion angles (Dickerson *et al.*, 1985; Dickerson, 1988), thus showing that there is no "standard" conformation for *B*-DNA. In addition, Westhof *et al.* (1986) observed that in a molecular dynamics refinement of a *Z*-DNA X-ray data set, the phosphate groups were rather disordered. Thus, even in the limited number of crystal structures of DNA duplexes presently available, there is variation in the rotamer

populations for the phosphate backbone and an assumption constraining the backbone torsion angles to a "standard" conformation is unwarranted. This assumption could certainly lead to some of the difference between the conclusions of Clore and co-workers and those presented here. We believe that this assumption can give quite misleading results since it will force the DNA duplex to have a regular backbone geometry even when there is no direct experimental evidence to support this conclusion. However, it may be possible to restrict the conformations of some backbone torsion angles by experimental measurements of homonuclear and heteronuclear coupling constants.

5. Conclusions

There has been some controversy as to how precisely and accurately the local conformational parameters of a nucleic acid duplex in solution can be determined from n.m.r. data. Different orkers have come to quite different conclusions as to whether sequence-dependent local structural variations, such as those predicted by Callidine's rules, can be observed in solution for nucleic acids (Clore *et al.*, 1988; Happ *et al.*, 1988). The results presented here indicate that these differences are a direct result of different assumptions made in the structure generation procedure. For example, Gronenborn & Clore (1989) state that using appropriate mixing times one can ignore spin diffusion effects and obtain accurate distances from NOE measurements with a precision of aproximately 10% for distances up to approximately 4 Å (which then leads to input distance constraints with precision of $\pm 0{\cdot}2$ to 0·4 Å). Although it is theoretically possible to ignore spin diffusion by obtaining spectra at very short mixing times, it is extremely difficult to prove that spin diffusion is not contributing to the NOE cross peak intensities for any practically useful mixing time. In fact, Reid *et al.* (1989) presented data that indicated that there may be significant spin diffusion effects even at mixing times of 30 milliseconds. In addition a number of groups have shown that neglecting spin diffusion effects can lead to an inaccurate set of distance constraints for practically useful mixing times (Borgias & James, 1988; Boelens *et al.*, 1989; Madrid *et al.*, 1989; Clore & Gronenborn, 1989). An alternate procedure to obtaining data at shorter mixing times would be to eliminate the need for ignoring the effects of spin diffusion. Therefore, here we have used an improved structure generation process that directly accounts for spin diffusion by employing a back-calculation procedure involving fitting to the experimental 2D NOE spectra. The use of a back-calculation procedure, such as that performed here, will help to eliminate the controversy over the precision with which the input distance constraints can be defined from n.m.r. data because one can directly fit to the experimental cross peak volumes and need not fit to distance constraints. By combining the back-calculation refinement with constrained energy-minimized and constrained quenched molecular dynamics calculations, we have generated a set of seven energetically favorable structures consistent with the experimental 2D NOE data for two different DNA decamers. Analysis of these structures shows that the majority of the local conformational parameters, such as helical twist, helical rise, propeller twist and backbone torsion angles cannot be defined precisely, using the n.m.r. data and computational procedure described here. Thus, from our data it is not possible to determine if sequence-dependent local structural variations, such as those predicted by Callidine's rules, exist for DNA duplexes in solution. However, it is clear that the procedure described here can be used to determine more qualitative structural information such as syn or anti glycosidic angles, S or N sugar puckers, formation of Watson–Crick base-pairs etc. These data can be extremely useful in structure determination of nucleic acids such as identification of hairpins (Van den Ven & Hilbers, 1988), triple-standard structures (Rajagopal & Feigon, 1989) or unusual conformations such as those found in telomeric DNA (Henderson *et al.*, 1987).

Although the procedure described here is an improvement over previous methods of structure generation because it refines the structures against the original experimental data and directly accounts for spin diffusion effects, there are a number of additional improvements which should lead to better defined structures. A direct fitting to the experimental cross peak volumes would be desirable. Kaptein and co-workers have described such an improvement involving refinement against the NOE relaxation matrix (Boelens *et al.*, 1989). It will be interesting to see how the precision of the structures changes using this procedure, as well as other procedures being developed which account for spin diffusion. Another important improvement would be the addition of homonuclear and heteronuclear coupling constants for constraining the phosphate backbone torsion angles in nucleic acids. ^{1}H-^{31}P coupling constant data would be helpful in defining these torsion angles. The $^{3}J_{C\text{-}5'H\text{-}P}$, $^{3}J_{C\text{-}5''H\text{-}P}$ and $^{3}J_{C\text{-}3'H\text{-}P}$ coupling constants give information on the β and ε torsion angles through Karplus-type relations. In addition, the $^{3}J_{C\text{-}5'H\text{-}C\text{-}4'H}$ and $^{3}J_{C\text{-}5''H\text{-}C\text{-}4'H}$ homonuclear coupling constants give information on the γ torsion angle. Unfortunately, there are no analogous coupling constants that can be used to give information on the α and ζ torsion angles in nucleic acids. However, it is clear that the addition of coupling constant data will help constrain the backbone torsion angles in DNA, which should then lead to better defined structures.

Thus, there are a number of avenues which can be pursued to try to obtain better defined local conformations in nucleic acid duplexes. However, until structures are generated with these additional data, and appropriate simulations and calibration studies have been performed, care should be taken in inter-

preting many of the subtle local conformational parameters of nucleic acid structures produced from n.m.r. data.

We thank Drs D. R. Hare, I. Tinoco, Jr and J. R. Williamson for useful discussions and Dr D. R. Hare for providing us with the BKCALC, DSPACE, FELIX and FTNMR programs. This work was supported in part by grants from the Searle Scholars Program of the Chicago Community Trust (85-C110), and NIH GM35807 to A.P., NSRA GM13075 to W.J.M. and GM30580 to R.M.L. The 500 MHz n.m.r. spectrometer was purchased with partial support from NIH grant RR03283.

References

Altona, C. & Sundaralingam, M. (1972). *J. Amer. Chem. Soc.* **94**, 8205–8212.

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

Bassolino, D. A., Hirata, F., Kitchen, D., Kominos, D., Pardi, A. & Levy, R. M. (1988). *Int. J. Supercomput. Applic.* **2**, 41–61.

Bax, A. & Davis, D. G. (1985) *J. Magn. Reson.* **65**, 355–360.

Billeter, M., Kline, A. D., Braun, W., Huber, R. & Wüthrich, K. (1989). *J. Mol. Biol.* **206**, 1677–1687.

Bloomfield, V. A., Crothers, D. M. & Tinoco, I., Jr (1974). *Physical Chemistry of Nucleic Acids*, Harper & Row, New York.

Boelens, R., Scheek, R. M., Dijkstra, K. & Kaptein, R. (1985). *J. Magn. Reson.* **62**, 378–388.

Boelens, R., Koning, 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. (1988). *J. Magn. Reson.* **79**, 493–512.

Braunschweiler, L. & Ernst, R. R. (1983). *J. Magn. Reson.* **53**, 521–528.

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

Clore, G. M. & Gronenborn, A. M. (1989). *J. Magn. Reson.* **84**, 398–409.

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

Crippen, G. M. (1981). *Distance Geometry and Conformation Calculations*, John Wiley, New York.

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

Dickerson, R. E. (1988). In *Unusual DNA Structures* (Wells, R. D. & Harvey, S. C., eds), pp. 287–293, Springer Verlag, New York.

Dickerson, R. E., Kopka, M. L. & Pjura, P. (1985). In *Biological Macromolecules and Assemblies* (Jurnak, F. A. & McPherson, A., eds), vol. 2, pp. 37–126, John Wiley, New York.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 2179–2183.

Fasman, G. (1979). *Handbook of Biochemistry, Selected Data for Molecular Biology*, 3rd edit. pp. 586, CRC Press, Cleveland.

Feigon, J., Leupin, W., Denny, W. A. & Kearns, D. R. (1983). *Biochemistry*, **22**, 5943–5941.

Frantini, A. V., Kopha, L., Drew, H. R. & Dickerson, R. E. (1982). *J. Biol. Chem.* **257**, 14686–14707.

Gronenborn, A. & Clore, M. (1989). *Biochemistry*, **28**, 5978–5984.

Happ, C. S., Happ, E., Nilges, M., Gronenborn, A. M. & Clore, G. M. (1988). *Biochemistry*, **27**, 1735–1743.

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

Hare, D. R., Sharpiro, L. & Patel, D. J. (1986). *Biochemistry*, **25**, 7455–7456.

Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., Jr & Blackburn, E. H. (1987). *Cell*, **51**, 899–908.

Hendrickson, W. A. & Konnert, J. H. (1981). In *Biomolecular Structure, Conformation, Function and Evolution* (Srinivasan, R., Subramanian, E. & Yatindra, N., eds), vol. 1, pp. 43–57, Pergamon, Oxford.

Hunter, W. N., D'Estaintot, L. & Kennard, O. (1989). *Biochemistry*, **28**, 2444–2451.

Jain, S. & Sundarlingam, M. J. (1989). *J. Biol. Chem.* **264**, 12780–12784.

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

Kalk, A., & Berendsen, H. J. C. (1976). *J. Magn. Reson.* **24**, 343–366.

Kline, A. D., Braun, W. & Wüthrich, K. (1986). *J. Mol. Biol.* **189**, 377–382.

Kline, A. D., Braun, W. & Wüthrich, K. (1988). *J. Mol. Biol.* **204**, 675-724.

Madrid, M., Mace, J. E. & Jardetzky, O. J. (1989). *J. Magn. Reson.* **48**, 267–278.

McCall, M., Brown, T., Hunter, W. N. & Kennard, O. (1986). *Nature (London)*, **322**, 661–664.

Mueller, L. & Ernst, R. R. (1979). *Mol. Phys.* **38**, 963–992.

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

Nerdal, W., Hare, D. R. & Reid, B. (1989). *Biochemistry*, **28**, 10008–10021.

Nilsson, L., Clore, G. M., Gronenborn, A. M., Brünger, A. T. & Karplus, M. (1986). *J. Mol. Biol.* **188**, 455–475.

Otting, G., Widmer, H., Wanger, G. & Wüthrich, K. (1986). *J. Magn. Reson.* **66**, 187–193.

Pardi, A., Martin, F. H. & Tinoco, I., Jr (1981). *Biochemistry*, **20**, 3986–3996.

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

Patel, D. J., Sharpiro, L. & Hare, D. (1987). *Quart. Rev. Biophys. Chem.* **16**, 423–454.

Pflugrath, J. M., Weigand, G., Huber, R. & Vertesy, L. (1986). *J. Mol. Biol.* **189**, 383–386.

Piantini, U., Sorensen, O. W. & Ernst, R. R. (1982). *J. Amer. Chem. Soc.* **104**, 6800–6801.

Rajagopal, P. & Feigon, J. (1989). *Nature (London)*, **339**, 637–640.

Rance, M., Sorensen, O. W., Bodenhausen, G., Wanger, G., Ernst, R. R. & Wüthrich, K. (1983). *Biochem. Biophys Res. Commun.* **117**, 479–485.

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

Reid, B. R., Banks, K., Flynn, P. & Nerdal, W. (1989). *Biochemistry*, **28**, 10001–10007.

Rosenberg, J. M., Seeman, N. C., Day, R. O. & Rich, A. (1976). *Biochem. Biophys. Res. Commun.* **69**, 979–987.

Saenger, W. (1984). *Principles of Nucleic Acid Structure*, Springer-Verlag, Berlin.

Sakked, Z., Guerstein-Guzikevich, G., Eisenstein, M., Frolow, F. & Rabinovich, D. (1989). *Nature (London)*, **342**, 456–460.

Scheek, R. M., Boelens, R., Russo, N., van Boom, J. H. & Kaptein, R. (1984). *Biochemistry*, **23**, 1371–1376.

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

Van den Ven, J. M. & Hilbers, C. W. (1988). *Eur. J. Biochem.* **178**, 1–38.

Wang, C. (1989). Ph.D. thesis, Rutgers University.

Wang, A. H. J., Quigley, G. J., Kolpak, F. J., van der Marel, G., van Boom, J. H. & Rich, A. (1981). *Science*, **211**, 171–176.
Weiss, M. A., Patel, D. J., Sauer, R. T. & Karplus, M. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 130–134.
Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986). *J. Comput. Chem.* **7**, 230–252.
Westhoff, E., Chevrier, B., Gallion, S. L., Weiner, P. K. & Levy, R. M. (1986). *J. Mol. Biol.* **190**, 699–712.
Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1980). *Nature (London)*, **287**, 755–758.
Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.

Edited by R. Huber