Solution Structure of Phage λ Half-Operator DNA by Use of NMR, Restrained Molecular Dynamics, and NOE-Based Refinement†

James D. Baleja,‡ Richard T. Pon,§ and Brian D. Sykes*,‡

Department of Biochemistry and Medical Research Council of Canada Group in Protein Structure and Function, University of Alberta, Edmonton, Alberta T6G 2H7, Canada, and Department of Medical Biochemistry, University of Calgary, Calgary, Alberta T2N 4N1, Canada

Received November 7, 1989; Revised Manuscript Received January 17, 1990

ABSTRACT: The structure of a DNA decamer comprising the left half of the O_R3 operator from bacteriophage λ is determined in solution by using nuclear magnetic resonance spectroscopy and restrained molecular mechanics calculations. Nuclear magnetic resonance assignments for nonexchangeable protons are obtained by two-dimensional correlated and nuclear Overhauser effect (NOE) spectroscopies. Exchangeable proton resonances are assigned by one-dimensional NOE experiments. Coupling constant measurements from one- and two-dimensional experiments are used to determine approximate dihedral angles within the deoxyribose ring. Distances between protons are estimated by extrapolating distances derived from the time-dependent NOE intensities to initial mixing times. The sets of dihedral angles and distances form a basis for structure determination by restrained molecular dynamics. Separate runs start from classical A and from B DNA and converge to essentially identical structures (atomic root mean square difference of 0.8 Å). The structures are improved by NOE-based refinement in which observed NOE intensities are compared to those calculated by using a full matrix analysis procedure. Final NOE residual (R) factors were less than 0.19. The resultant structures are generally B type in character, but display local sequence-dependent variations in dihedral angles and in the spatial arrangement of adjacent base pairs. Although the entire structure exhibits a small bend, the central core of the half-operator, which comprises the sequence-specific recognition site for cro repressor, is straight.

The sequence-specific recognition of DNA by proteins is an important component in the control of cellular processes. Our working model of protein–DNA interaction is the cro repressor–O_R3 operator system from bacteriophage λ [for reviews, see Brennan and Matthews (1989a), Pabo and Sauer (1984), Takeda et al. (1983), and Ptashne et al. (1980)] in which the helix–turn–helix DNA-binding motif of the protein recognizes a specific sequence of DNA. The cro repressor is one of a family of structurally related DNA-binding proteins that include the lac repressor (Kaptein et al., 1985), the repressor from phage 434 (Wolberger et al., 1988) and from phage λ (Jordan & Pabo, 1988), and the catabolite gene activator protein (CAP) from *Escherichia coli* (McKay & Steitz, 1981). These proteins bind as dimers to 2-fold symmetric operator DNA sites, with the recognition helix of the DNA-binding motif of each monomer fitting into successive major grooves of B-type DNA. The DNA is directly read via network of hydrogen bonds and other contacts between amino acid side chains and the exposed functional groups in the major groove of DNA. This would suggest that the sequence diversity of DNA need not generate different DNA structures, but merely provide different patterns of hydrogen bonding in the protein–DNA recognition process.

However, the recent crystallographic determination of the related trp repressor complexed with its operator DNA shows no direct sequence-specific contacts between protein and DNA, suggesting that the trp repressor is recognized indirectly through sequence-specific changes in the geometry of the phosphate backbone to allow the formation of a tight protein–DNA complex (Otwinowski et al., 1988). In addition, although the X-ray structure of the repressor–operator complex from phage 434 shows no direct protein contact at the center of the operator (Aggarwal et al., 1988), substitution of these base pairs changes the binding affinity of the protein for its DNA (Koudelka et al., 1987, 1988). From modeling studies with B DNA, and from extensive biochemical data, the same situation is likely to exist with the phage λ cro repressor protein. There are no direct contacts between protein and the center of the O_R3 operator, but substitution of the central base pair results in reduced affinity of protein for DNA (Ohlendorf et al. 1982, Takeda et al., 1989). These results suggest a sequence-dependent change in conformation of the operator to accommodate for binding to the cro dimer (Kirpichnikov et al., 1984) or, more generally, that DNA does not merely play a passive role in protein–DNA recognition but instead, by altering its conformation from regular B DNA, is able to preform an optimal binding site for its cognate protein or, alternatively, is able to adopt a lower energy conformation in the protein–DNA complex.

In this report, we present the structure of part of the O_R3 operator DNA in the absence of protein as a basis for comparison to the structure of DNA within the protein–DNA complex. We have chosen to study a chemically synthesized 10 base-pair DNA fragment using two-dimensional NMR[1]

† This work was supported by the Medical Research Council of Canada and the Alberta Heritage Foundation for Medical Research.

* Address correspondence to this author.

‡ University of Alberta.

§ University of Calgary.

[1] Abbreviations: NMR, nuclear magnetic resonance spectroscopy; NOE, nuclear Overhauser effect; NOESY, two-dimensional NOE spectroscopy; DQF-COSY, double quantum filtered two-dimensional correlation spectroscopy; J, NMR spin–spin coupling constant; MD, molecular dynamics; rms, root mean square; EDTA, ethylenediaminetetraacetic acid.

0006-2960/90/0429-4828$02.50/0

techniques. The sequence comprises the left most 10 base pairs (L10) of the 21 base-pair operator and also contains the consensus cro repressor binding site (Kim et al., 1987). This half-operator binds to cro repressor at a ratio of 2 to 1, with nearly the same binding strength as the full operator (unpublished results), confirming that the complex is specific, as is observed with the lac repressor headpiece:half-operator DNA system (Scheek et al., 1983b). First, nonexchangeable proton resonances are assigned in a sequential manner by a combination of two-dimensional NOE (NOESY) and through-bond correlated (COSY) spectroscopies. Assignments are extended to base imino and hydrogen-bonded amino protons by one-dimensional NOE experiments in H_2O. A set of allowed dihedral angles is derived from a qualitative interpretation of COSY intensities and cross-peak patterns (Chary et al., 1988). From time-dependent NOE cross-peak intensities, another set of approximate interproton distances is obtained (Baleja et al., 1990). These two sets of data are then used as the basis for a structure determination by restrained molecular mechanics calculations (Kaptein et al., 1985; Nilsson et al., 1986). Convergence is achieved (rms deviation of 0.8 Å) starting from two quite different initial structures, namely, classical A- and B-type DNA (rms deviation of 5.4 Å).

The distance between two spins is often estimated by assuming inverse proportionality to the sixth root of the NOE cross-peak intensity. Distances so derived are only approximate since the cross-peak intensity due to direct cross-relaxation between spins i and j is modified by additional cross-relaxation with any spin k, especially if spin k exists such that $r_{ik} < r_{ij}$ or $r_{jk} < r_{ij}$. However, NOE cross-peak intensities may be predicted from the structures produced by dynamics calculations and compared directly to the observed intensities, eliminating approximate distance calculation (Keeper & James, 1984; Suzuki et al., 1986; Gupta et al., 1988, Boelens et al., 1989). The structures resulting from restrained molecular dynamics calculations are refined in an iterative manner so as to minimize the difference between the two sets of NOEs (Baleja et al., 1990). The final structures have NOE R factors of less than 0.19, consistent with the observed experimental data. The structure is of the B-type DNA family and exhibits sequence-dependent variations in conformational parameters. The full O_R3 operator is derived from the half-operator structure. Although some subtle differences from classical B DNA are displayed in the pattern of proposed protein contacts in the cro repressor binding site, it is basically consistent with the most recent cro repressor–DNA models (Takeda et al., 1989).

Experimental Procedures

Sample Preparation. The two deoxyoligonucleotides, d5′(TCTATCACCG) and d5′(CGGTGATAGA), were prepared on an Applied Biosystems DNA synthesizer using a large-scale synthesis procedure (Lee et al., 1988). Coupling yields for the syntheses averaged 95–97%. After deprotection with thiophenol and ammonium hydroxide, synthesis products were purified by chromatography on NACS-20 resin using a 0.25–0.45 M NaCl gradient in 12 mM NaOH (Lee et al., 1988). Pooled fractions were neutralized with acetic acid and desalted on a Sephadex G-25 column. Electrophoresis on 24% polyacrylamide/7 M urea gels and rechromatography on NACS-20 confirmed product homogeneity. An aliquot of each strand was digested to 5′-nucleotides for 1 h with 10 μL of ca. 0.1% snake venom phosphodiesterase in 20 mM Tris, 1 mM $MgCl_2$, pH 9 buffer. The concentration of each strand was determined by the absorbance of nucleotide mixture using the following molar absorbancies at 260 nm: dC, 7900; pdC, 6400; dA, 14 700; pdA, 15 300; dG, 11 800; pdG, 11 700; dT, 9600; pdT, 7600. Heating equimolar amounts of strands to 85 °C in 140 μL of a 60 mM KCl, 20 mM KPO_4, 50 μM EDTA, pH 7 buffer, and then allowing the solution to cool to room temperature over several hours, ensured strand annealing to form the duplex.

$$5'(T_1\,pC_2\,pT_3\,pA_4\,pT_5\,pC_6\,pA_7\,pC_8\,pC_9\,pG_{10})3' \quad \text{(strand 1)}$$

$$3'(A_{20}pG_{19}pA_{18}pT_{17}pA_{16}pG_{15}pT_{14}pG_{13}pG_{12}pC_{11})5' \quad \text{(strand 2)}$$

Solutions were passed over the Na^+ form of Chelex-100 to remove paramagnetic metal ions before lyophilization. Samples were dissolved and relyophilized repeatedly in increasing grades of D_2O and finally taken up in 0.65 mL of 99.997% D_2O. Final concentrations were 4 mM duplex, 0.2 M KCl, 10 mM K_2HPO_4, 10 mM KH_2PO_4, 20 μM EDTA, pH* 7.5 (direct meter reading).

NMR Spectroscopy. All NMR spectra were obtained on a Varian VXR500 NMR spectrometer with an operating frequency of 500 MHz for protons. Phase-sensitive two-dimensional spectra in D_2O were taken with 2K complex points with appropriate phase cycling for quadrature detection and for eliminating axial peaks. Streaking along t_1 was reduced by multiplying the first domain time point by a factor optimized near 0.5 (Otting et al., 1986). The residual HOD signal was suppressed by presaturation. Spectra were normally taken at 20 °C, although spectra at 30 °C were useful to confirm assignments at the lower temperature. In particular, double quantum filtered COSY spectra at 30 °C had better signal-to-noise ratios, and small chemical shift changes, but were otherwise identical to 20 °C spectra. Free induction decays were weighted in each dimension to effect slight resolution enhancement and to avoid truncation effects.

NOESY spectra were collected by the hypercomplex method (States et al., 1982). Average mixing times of 50, 100, 150, and 250 ms were used with a random delay of ±10 ms incorporated to suppress zero quantum coherence. A total of 256 t_1 increments were taken, and for each t_1 value, 64 scans with an acquisition time of 0.215 s were signal averaged. The delay time between scans was 2.1 s. Spectra were zero-filled in each dimension so that final transformed spectra were 1K by 1K real data points. NOE intensities were quantified by determining the volume integral of each cross-peak. Nominally empty areas perpendicularly adjacent to each cross-peak were examined for base-line correction.

Through-bond correlations were observed with DQF-COSY spectra (Piantini et al., 1982). For each t_1 value, 64 scans with an acquisition time of 0.26 s were signal averaged with a delay time of 2.0 s between scans. A total of 512 t_1 increments were taken and final spectra were 2K by 1K data points.

One-dimensional spectra were taken in 85% H_2O/15% D_2O at 10 °C with 1-1 binomial suppression of the H_2O peak (Hore, 1983). For difference NOE spectra (Gronenborn & Clore, 1985), the decoupler power was chosen (γB_2 = 8 Hz) to completely saturate the peak during the mixing period while still retaining specificity. Mixing times between 100 and 500 ms were used. A total of 2048 transients with a preceding delay time of 2.5 s and an acquisition time of 0.5 s were summed into 13K data points.

Restrained Molecular Dynamics. Energy minimization and molecular dynamics (MD) calculations were carried out with the GROMOS program (van Gunsteren et al., 1985; de Vlieg et al., 1986) and force field, which consisted of the usual terms for bonds, bond angles, sinusoidal dihedral torsion, nonbonded interactions (van der Waals and electrostatics), and harmonic terms to maintain proper planar or tetrahedral geometries, and

to which two extra terms representing distance and dihedral restraints were added. The distance restraint square-well potential, E_{DIS}, was given by

$$E_{DIS} = 0.5CDIS[r_{ij}^{U} - r_{ij}]^2 \quad \text{if } r_{ij} > r_{ij}^{U}$$
$$= 0.5CDIS[r_{ij}^{L} - r_{ij}]^2 \quad \text{if } r_{ij} < r_{ij}^{L} \qquad (1)$$

where r_{ij}^{U} and r_{ij}^{L} were the upper and lower estimates of the distance between protons i and j, respectively, r_{ij} was the calculated distance, and CDIS was a force constant. The effective dihedral angle restraint, E_{DIH}, was represented by

$$E_{DIH} = 0.5CDLR[\phi_k^{U} - \phi_k]^2 \quad \text{if} \phi_k > \phi_k^{U}$$
$$= 0.5CDLR[\phi_k^{L} - \phi_k]^2 \quad \text{if } \phi_k < \phi_k^{L} \qquad (2)$$

where ϕ_k^{U} and ϕ_k^{L} were upper and lower allowed limits of the torsion angle, ϕ_k was the calculated angle, and CDLR was the force constant. The distance and dihedral angle experimental data sets were based on observations from the two-dimensional NMR spectra (see below).

Starting models were first subjected to 200 steps of steepest descents energy minimization. During the first 10 ps of each MD simulation, values of the distance restraint force constant were increased from 500 to 10000 kJ mol^{-1} nm^{-2} and the dihedral restraint force constant from 5 to 50 kJ mol^{-1} rad^{-2}. Velocities were reinitialized (taken from a Maxwellian distribution at 300 K) with every increase in the force constants (approximately every 1.5 ps). Newton's equations of motion were integrated with a time step of 2 fs, with all bond lengths kept rigid by the SHAKE algorithm (Ryckaert et al., 1977). The molecule was weakly coupled to a temperature bath with a reference temperature of 300 K with a coupling constant of 0.1 ps (Berendsen et al., 1984). A cutoff radius of 10 Å was applied for nonbonded interactions and the nonbonded pair list was updated every 20 fs. Molecular dynamics runs with the highest values of CDIS and CDLR were continued to 20 ps in total, and coordinates were averaged over the last 5 ps. Averaged molecular dynamics structures were then subjected to 200 steps of energy minimization to correct distortions in the structure caused by the averaging procdure.

We performed a total of four molecular dynamics runs: (1), with a starting model of L10 in a classical A-type DNA conformation (Arnott & Hukins, 1972), a set of distance restraints, and a set of dihedral restraints; (2), with a starting model in an average B-DNA configuration, and the experimental restraints; (3), with a starting A-DNA model and the sets of distance and dihedral angle restraints that correspond to the same atoms for the experimental restraints, but with values taken from the B-DNA model; and (4), with a starting B-DNA model with A-DNA restraints.

Structure Refinement. Molecular dynamics calculations result in structures that satisfy the two sets of experimental restraints—the approximate distance set and the dihedral angle set. Distance restraints based on the NOE intensities are most often inaccurate because of spin-diffusion effects. In addition, the Watson–Crick base-pairing distance restraints prevent any substantial deviations for the two bases of a base pair from lying in a single plane and may affect certain conformational parameters, such as propeller twist.

Structures were therefore subjected to further refinement, which was directly based on the NOE intensities (without distance restraints), by replacing the effective potential for distance restraints by an E_{NOE} potential for NOE restraints (Baleja & Sykes, 1988; Baleja et al., 1990):

$$E_{NOE} = 0.5CNOE\sum[NOE_{obs} - NOE_{calc}]^2 \qquad (3)$$

with the forces due to this pseudo energy potential calculated as

$$F_{x_i} = -\frac{\partial E_{NOE}}{\partial_{x_i}} = CNOE\sum[NOE_{obs} - NOE_{calc}]\frac{\partial NOE_{calc}}{\partial_{x_i}} \qquad (4)$$

Derivatives of the NOE with respect to a change in each Cartesian coordinate of proton i were evaluated numerically by changing the coordinate slightly (by 0.01 Å) and recalculating the NOE. Evaluation of the function was repeated for each proton in the molecule. Forces from the NOE potential were evaluated at the first step and were subsequently evaluated every 10 steps during the molecular mechanics run. The NOE "energy" gradients included both direct forces, where two protons, i and j, were pushed closer together if the calculated NOE_{ij} was too weak or stretched apart if too strong, and indirect forces that resulted from the effect of moving proton i on NOE_{jk} (Baleja et al., 1990). The pseudo energy forces were calculated for NOE intensities at 150 and 250 ms, which was a compromise between less signal-to-noise at shorter mixing times and greater spin diffusion effects at long times. A total of 294 NOE intensities at 150 ms and 309 NOE intensities at 250 ms were used for structure refinement. The force field here included a description of nonexchangeable protons. They were given no van der Waals radii, as the carbons they were attached to retained the united atom approach for the nonbonded interaction calculation. Terms for bond length (1.08 Å) and bond angles and improper dihedral angles were added to give proper stereochemistry. NOE intensities between nonexchangeable protons were calculated by assuming homonuclear dipolar relaxation for a macromolecule tumbling isotropically in solution (Solomon, 1955; Abragam, 1961; Baleja et al., 1990).

To take into account differential motion during refinement, an empirical correlation time reduction factor is assigned to each proton of a proton pair:

$$\tau_{ij} = \tau_c(S_iS_j) \qquad (5)$$

The product, S_{ij}^2, is related to the order parameter, S^2, of Lipari and Szabo (1982) and can vary between 0 and 1. A value of 1 indicates that the correlation time of the interproton vector is the same as the overall tumbling time of the macromolecule; a value of zero indicates complete motional freedom. Internal motions in DNA (Hogan & Jardetzky, 1980; Keepers & James, 1982; Clore & Gronenborn, 1984) are indicated here with correlation time reduction factors of less than unity, which shorten correlation times for more mobile protons. Values of 0.65, 0.85, and 0.9 were used empirically to reflect the increased motion of all thymine methyl groups, sugar 2′ and 2″ methylene protons, and the 5′ and 3′ terminal residues, respectively, with an overall correlation time of 3.8 ns (Baleja et al., 1990).

Structure refinement began with energy minimization using a dihedral force constant of 50 kJ mol^{-1} rad^{-2} and an NOE force constant of 1000 $(\Delta NOE)^{-2}$ kJ mol^{-1} and continued until the total potential energy changed by less than 0.001 kJ per step (about 50 steps). A further 100 steps of minimization with CNOE set to 2000 $(\Delta NOE)^{-2}$ kJ mol^{-1} completed the refinement procedure.[2] Helical parameters of the final structures were analyzed with the programs AHELIX, BROLL, and CYLIN (Fratini et al., 1982; Dickerson et al., 1985).

[2] Larger NOE force constants cause unacceptable energies or distortions in the proton stereochemistry.

Table I: Proton Chemical Shift Assignments for L10[a]

residue	H6/H8	H2/H5/CH_3	1′	2′	2″	3′	4′	5′, 5″[b]	NH[e]
T_1	7.53	1.67	6.08	2.20	2.52	4.72	4.14	3.73, 3.73	*f*
C_2	7.74	5.80	6.04	2.22	2.52	4.83	4.23	4.10, 4.08[c]	8.46
T_3	7.47	1.71	5.63	2.23	2.53	4.89	4.18	4.11, 4.14[c]	13.69
A_4	8.38	7.42	6.29	2.70	2.96	5.03	4.45	4.15, 4.18	
T_5	7.17	1.39	5.86	2.00	2.41	4.83	4.17	4.15, 4.30	13.50
C_6	7.49	5.61	5.50	2.03	2.37	4.82	4.11	*d*, *d*	8.41
A_7	8.26	7.68	6.17	2.70	2.86	5.00	4.40	4.04, 4.12	
C_8	7.27	5.28	5.81	1.93	2.32	4.76	4.14	4.23, 4.09	8.11
C_9	7.42	5.60	5.57	1.95	2.29	4.79	4.07	4.04[c], 4.02[c]	8.64
G_{10}	7.93		6.15	2.61	2.36	4.67	4.17	4.07,[c] 4.01[c]	13.10
C_{11}	7.56	5.84	5.71	1.83	2.35	4.67	4.05	3.69, 3.69	
G_{12}	7.92		5.54	2.69	2.74	4.97	4.31	4.07, 3.96	13.15
G_{13}	7.79		6.00	2.58	2.77	4.97	4.43	4.16, 4.18	12.82
T_{14}	7.12	1.39	5.76	1.98	2.39	4.86	4.16	4.22, 4.10[c]	13.60
G_{15}	7.88		5.60	2.67	2.77	5.00	4.35	*d*, *d*	12.44
A_{16}	8.17	7.68	6.18	2.58	2.88	4.99	4.44	4.23, 4.25[c]	
T_{17}	7.11	1.37	5.51	1.86	2.20	4.80	4.08	4.22, 4.15[c]	13.47
A_{18}	8.10	7.35	5.86	2.57	2.72	4.98	4.34	4.00, 4.08	
G_{19}	7.67		5.50	2.44	2.57	4.92	4.29	4.11,[c] *d*	12.59
A_{20}	8.07	7.87	6.27	2.53	2.39	4.63	4.22	4.18, 4.09	

[a] Chemical shifts (ppm) are relative to disodium 2,2-dimethyl-2-silapentane-5-sulfonate. Nonexchangeable proton assignments are at 20 °C and exchangeable proton assignments are at 7.7 °C. [b] 5′ and 5″ protons are not assigned stereospecifically. [c] Tentative assignment. [d] Unassigned 5′ and 5″ resonances are between 4.05 and 4.25 ppm. [e] Assignments are for the H1 imino proton of guanine, the H3 imino proton of thymine, or the hydrogen-bonded amino proton (on N4) of cytosine. [f] No unambiguous resonance was observed for T_1 because of spectral overlap, rapid exchange with H_2O, and poor signal intensity.

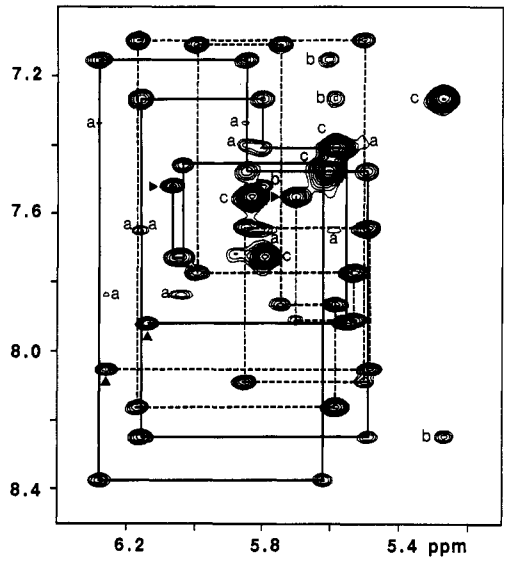

FIGURE 1: Assignment of base and 1′ protons of L10 DNA. The mixing time of the 500-MHz NOESY experiment was 250 ms. Sequential connectivities between H6/H8 and 1′ protons are shown by the solid line for the strand 1 (residues 1–10) and by the dashed line for strand 2 (residues 11–20). (a) Cross-peaks to adenine H2 protons. (b) NOE cross-peaks between the H5 base proton of cytosine and the H6/H8 proton of the preceeding base. (c) H5–H6 cytosine cross-peaks. Resonance assignments are given in Table I.

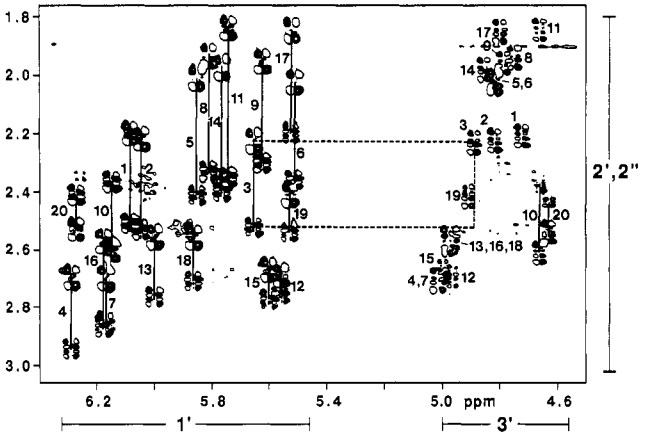

FIGURE 2: Assignment of 2′, 2″, and 3′ protons. The DQF-COSY ^1H NMR spectrum was taken at 30 °C. Cross-peaks between 1′ ↔ 2′ and 1′ ↔ 2″ protons are indicated by residue number. Except for the 3′ terminal, 2′ protons resonate upfield to the 2″ proton of the same residue. The dashed line illustrates the extension of the assignment from 2′, 2″ protons to the 3′ proton for residue T_3. As observed with other residues, the 2″ ↔ 3′ cross-peak is absent. 2′ ↔ 3′ cross-peaks are noted by residue number.

RESULTS

Resonance Assignment. A prerequisite for the determination of a solution structure by NOE measurements is the assignment of a resonance to a specific proton of the macromolecule. Resonance lines in nucleic acids can be identified by a combination of COSY and NOESY experiments, which measure respectively, coherent and incoherent transfer of magnetization between individual nuclei. A qualitative interpretation of the NOESY spectra reveals that the L10 decamer is right-handed (Cohen, 1987) and resonances were therefore assigned by following the procedures developed for right-handed helices (Feigon et al., 1982; Scheek et al., 1983; Hare et al., 1983; Chazin et al., 1986). Regions of the NOESY and DQF-COSY spectra used for assignment are shown in Figures 1 and 2, and Figures S2–S4 of the supplementary material. Exchangeable protons were assigned by one-dimensional difference NOE spectroscopy (Gronenborn & Clore, 1985), as shown in Figure S-5.

Distance Determination. Distances between protons are obtained by a distance extrapolation procedure. Distances are derived at each mixing time, plotted against the mixing time, and extrapolated back to zero mixing time as a first-order correction for spin diffusion effects:

$$r_{ij} = \lim_{\tau_m \to 0} r_{ij}(\tau_m) = \lim_{\tau_m \to 0} \left[\frac{\mathrm{NOE}_{\mathrm{ref}}(\tau_m) r_{\mathrm{ref}}(\tau_m)^6}{\mathrm{NOE}_{ij}(\tau_m)} \right]^{1/6} \quad (6)$$

This method yields the same quality of distances as that obtained by fitting the NOE buildup curve with a polynomial in τ_m and determining the slope at $\tau_m = 0$, but offers advantages in being self-correcting for changes in instrument gain between mixing time experiments and allowing a more direct visualization of spin diffusion effects in derived distances

(Baleja et al., 1990; Majumdar & Hosur, 1989; Nilges et al., 1987). For base and 1′ sugar protons, the cytosine H5–H6 reference distance of 2.46 Å is used. Remaining distances between nonexchangeable protons are determined by using the average $NOE_{ref}r_{ref}^6$ products for the C2′ methylene 2′ ↔ 2″ proton pair (1.76 Å) and the 1′ ↔ 2″ pair (2.3 ± 0.1 Å for all sugar puckers). Upper and lower bounds on these distances are estimated from the distance extrapolation curve. For all distances greater than 3.5 Å and involving both an aromatic base and either of the C2′ methylene protons, upper bounds are increased by 0.2 Å to account for systematic distance underdetermination for this arrangement of protons (Baleja et al., 1990).

NOE intensities are calibrated for exchangeable protons by assuming an average 3.4-Å distance for imino–imino NOEs between sequential residues, 2.5 Å for the guanine H1 ↔ cytosine H4a (hydrogen-bonded amino proton) interstrand proton pair, and 3.0 Å for the thymine H3 ↔ adenine H2 interstrand proton pair. Increased upper distance estimates and decreased lower distance estimates were used for the distances involving exchangeable protons because of measurement at a lower temperature, and loss of magnetization due to exchange with H_2O.

For overlapping peaks, lower bounds for the distances, corresponding to the proton pairs involved, could be estimated by using all of the observed NOE intensity for each pair. The set of distance restraints comprised 322 upper distance bounds and 330 lower distance bounds, determined from the NOE intensities for input into molecular dynamics calculations. NOE measurements and distances are given in the experimental restraints table of the supplementary material. Little residue to residue variation is seen at short mixing times for interproton vectors of fixed length, in either distances or NOE intensities, indicating only a small amount of increased motion for terminal residues on the nanosecond time scale. At longer mixing times, however, NOE intensities are dissimilar for terminal residues. Much of this effect can be modeled by calculations taking spin diffusion into account, indicating that the lower spin density at the ends of the molecule results in less spin diffusion and the NOE intensity for short covalently bound distances is therefore lower.

Glycosidic Dihedral Angles. The geometry of the five-member sugar ring of DNA can be described by five torsion angles v_0–v_4. Because of ring closure, the values of v_0–v_4 are interrelated:

$$v_n = v_{max} \cos [P + 144(n - 2)] \quad n = 0, 4 \qquad (7)$$

where v_{max} is the maximum amplitude of the sugar ring pucker, and P is the sugar pseudorotational angle (Altona & Sundaralingam, 1972). The magnitude of three-bond coupling constants between protons is dependent on the intervening dihedral angle and therefore reflects the pseudorotational angle that specifies the conformation of the sugar ring (Hosur et al., 1986).

Excepting 3′ termini, 2″ ↔ 3′ correlations are absent in the COSY spectrum (Figure 2). This indicates that the sugar pseudorotational angle is between 100 and 250° for all non-3′ terminal sugar rings (Hosur et al., 1986). Furthermore, the separation of the outer lines in cross-peaks involving the 1′ proton is greater than 14 Hz in the 1′ frequency dimension. This indicates the sum of coupling constants to the 1′ proton for all residues to be greater than 14 Hz, which further limits the pseudorotation angle to lie between 100 and 200° (Chary et al., 1988; Rinkel & Altona, 1987). The observation of 3′ ↔ 4′ correlations (Figure S-4) for all residues narrows the pseudorotation angle to be between 105 and 175°. Allowed ranges of glycosidic dihedral angles are obtained for each nucleotide (Table S-1) from the coupling constant data by assuming a v_{max} of 35° (Rinkel & Altona, 1987) and by using eq 7.

Right-Handed DNA Helix Restraints. To preserve the right-handed character of the DNA during the molecular dynamics calculations, it was sometimes necessary to constrain backbone dihedral angles to be in a broad allowed region of torsional angle space (Gronenborn & Clore, 1989). The allowed angles (α, −90 to −30°; β, <−145 and >135°; γ, 30–90°; ϵ, <−60 and >140°; ζ, <−45 and >150°) are derived from a table of conformation angles found in the different DNA types (Suzuki et al., 1986) and from consideration of individual variations in right-handed helices from X-ray crystallographic studies (Dickerson et al., 1985). These right-handed helix restraints would cause no violations for any of the average A-, B-, alternating B-, C-, D-, or wrinkled D-DNA forms, nor for any individual angles found in the best studied single-crystal X-ray structures of B DNA (Dickerson & Drew, 1981; Privé et al., 1987), except for two β angles (of residues G_3 and C_{12}) in the B-DNA form of a phosphothiolate analogue of DNA (Cruse et al., 1986).

Base pairs were kept Watson–Crick hydrogen bonded by distance restraints between bases. These were as follows for all base pairs: $r(C_{1'}–C_{1'}$, across the base pair) = 10.87 ± 0.20 Å. For A·T base pairs: $r(A_{N6}–T_{O4})$ = 2.8 ± 0.1 Å, $r(A_{HN}–T_{O4})$ = 1.7 ± 0.1 Å, $r(A_{N1}–T_{H3})$ = 1.7 ± 0.1 Å, $r(A_{N1}–T_{N3})$ = 2.8 ± 0.1 Å. For G·C pairs: $r(G_{O6}–C_{HN})$ = 1.6 ± 0.1 Å, $r(G_{O6}–C_{N4})$ = 2.7 ± 0.1 Å, $r(G_{N1}–C_{N3})$ = 2.8 ± 0.1 Å, $r(G_{H1}–C_{N3})$ = 1.7 ± 0.1 Å, $r(G_{N2}–C_{O2})$ = 2.8 ± 0.1 Å, $r(G_{HN}–C_{O2})$ = 1.7 ± 0.1 Å (Arnott & Hukins, 1972). These gave 120 distance restraints in addition to the 652 noted above.

After approximately 10 ps of molecular dynamics, the major groove had a tendency to collapse on two occasions—when starting from B-type DNA with experimental restraints, and when starting from A-type DNA with B-DNA restraints. A structure formed with a lower van der Waals energy (by 100 kJ), but also with an increased total distance restraint energy (by 30 kJ). The distance restraint force constants could be increased in order to dominate the van der Waals contribution, although an increased force on inaccurate distance restraints would likely cause distortions in the structure. Alternatively, we could have further experimented with the electrostatic and van der Waals interactions, for example, by adding counterions to the phosphates. Instead we have added a few "generous" distance restraints between C1′ atoms on opposite sides of the major groove (Table S1). For example, the distance between T_5 C1′ and G_{19} C1′ is 12.4 and 9.7 Å in A and in B DNA, respectively. During the MD run, these atoms are required to be greater than 8 Å apart. No violations occur with either the right-handed DNA dihedral angle set or with the major groove distance restraints in any of the structures presented in this paper.

Molecular Dynamics Calculations. Several small alterations to the GROMOS force field were made in order to be more consistent with the nucleic acid force field of the CHARMM molecular mechanics program (Nilsson & Karplus, 1986). The normal van der Waals radius on united methylene carbons was reduced from 2.22 to 2.10 Å to avoid steric clashes between C2′ and one of the oxygens on the 3′-phosphate. Methine carbons were given van der Waals radii of 2.05 Å. Corresponding 1–4 van der Waals interactions were left unaltered. The effect of solvent was approximated by a $1/\epsilon r$ screening function where r was the separation of the charged groups in angstroms (Brooks et al., 1983) and ϵ was equal to

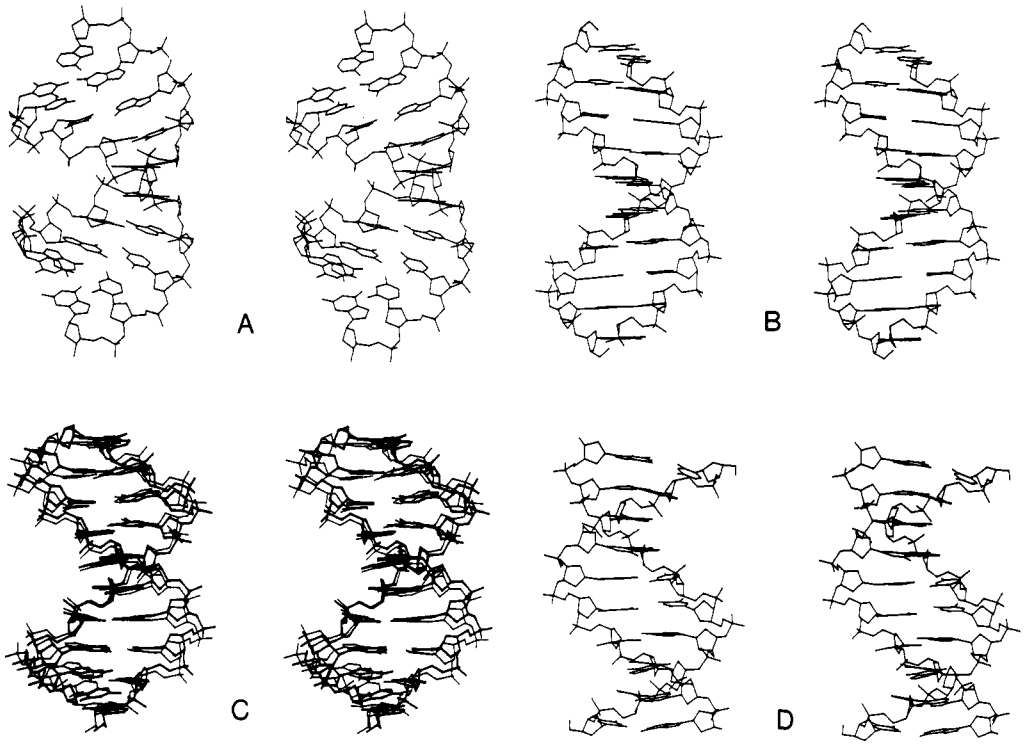

FIGURE 3: Structural models for L10 DNA. (A) Initial model structure in a classical A-DNA conformation. (B) Initial model structure in a B-DNA geometry. (C) Two superimposed final refined structures, R-A and R-B, resulting from restrained molecular dynamics and NOE-based refinement. R-A is shown in bold. (D) The average of structures R-A and R-B. Structure R_{av} has been rotated by 90° to provide a view into the major groove. Protons have been omitted to preserve clarity in the diagrams.

Table II: Atomic rms Differences (Å) between L10 Structures[a]

	A DNA	B DNA	R-A	R-B	BAI	ABI
A DNA		5.43	4.82	4.14	1.49	6.31
B DNA	4.69		1.30	1.82	5.01	1.82
R-A	4.28	1.08		0.80	4.39	2.29
R-B	3.66	1.55	0.74		3.77	2.88
BAI	1.30	4.49	4.06	3.48		5.66
ABI	5.54	1.38	1.88	2.44	5.17	

[a] A DNA and B DNA are starting structures with regular A and B geometries, respectively. R-A and R-B are refined structures after distance-restrained molecular dynamics and NOE-based refinement. BAI is the MD structure resulting from applying A-DNA distance restraints to a B-DNA model. ABI is the MD structure resulting from applying B-DNA distance restraints to an A-DNA model. Numbers above the diagonal indicate the rms atomic deviation between two structures. Numbers below the diagonal show the comparison for the middle 8 base pairs.

4 (Weiner et al., 1984; Cuniasse et al., 1989). The net charge on each phosphate group was reduced to −0.32e (Tidor et al., 1983; Nilsson et al., 1986).

The molecular dynamics protocol and force field are tested by following the method of Nilsson (Nilsson et al., 1986). A set of distance restraints and dihedral angles (between the same atoms for which there are corresponding experimental measurements and with the same error estimates) are taken from classical B DNA and are applied to a MD simulation starting from A DNA and vice versa. In both cases, the expected transition occurs with the A to B structure (ABI) having an rms deviation of 1.8 Å from B DNA, and the B to A structure (BAI) has an rms deviation from A DNA of 1.5 Å (Table II). Note that although these distances are taken from idealized structures, they also are from canonically straight DNA with no adjustments made for sequence-dependent variations.

The two starting structures (A- and B-type DNA models) for the MD runs with experimental restraints are depicted in the top half of Figure 3. Application of molecular dynamics produces structures MD-A and MD-B. They have an rms atomic deviation of less than 0.8 Å, which is comparable to the fluctuations during the averaging over the last 5 ps of the MD run. This indicates that molecular dynamics has sampled conformational space well, and that both MD runs have converged essentially to the same structure (Scalfi Happ et al., 1988). Both structures agree with the experimental dihedral and distance restraints equally well, as indicated by the pseudo potential energies listed in Table III.

Table III: Potential Energies (kJ/mol) for L10 Structures[a]

total[a]	electrostatic	van der Waals	bond	angle	improper dihedral	torsion	distance restraint	dihedral restraint	NOE restraint	
A DNA_m	−858	−325	−1855	26	220	40	705	240	92	
B DNA_m	−981	−321	−1787	28	243	18	800	37	0	
MD-A	−1155	−328	−1934	24	209	45	752	73	4	
MD-B	−1161	−330	−1924	26	209	46	740	70	4	
R-A	−1168	−321	−1985	15	223	136	790		1	124
R-B	−1166	−322	−1970	15	229	139	775		1	125
R_{av}	−1181	−319	−1995	14	224	132	786		1	125

[a] Energies are given for initial A- and B-DNA models after energy minimization with distance restraints, with CDIS equal to 500 kJ mol^{-1} nm^{-2} and CDIH at 5 kJ mol^{-1} rad^{-2}. For the remaining structures, CDIS and CDIH are 10000 kJ mol^{-1} nm^{-2} and 50 kJ mol^{-1} rad^{-2}. MD structures A and B result from energy minimization of the average structures from the MD runs starting with the initial A- and B-DNA models. For refined (R) structures, the tabulations for total energy exclude the contributions from nonexchangeable protons.

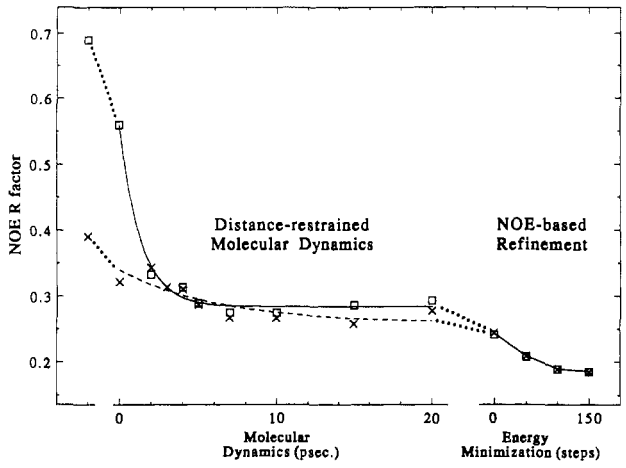

FIGURE 4: Improvement in the NOE *R* factor as structure determination proceeds. Initial A-DNA models (□) and B-DNA models (×) are first subjected to energy minimization with approximate distance restraints. Twenty picoseconds of distance-restrained molecular dynamics reduces the *R* factor to near 0.27. Averaging over the last 5 ps (second dotted line) results in structures with *R* factors of 0.24. Structures are refined by comparing observed NOE intensities to NOEs calculated by using a full matrix analysis procedure and minimizing the difference between the two sets of NOEs.

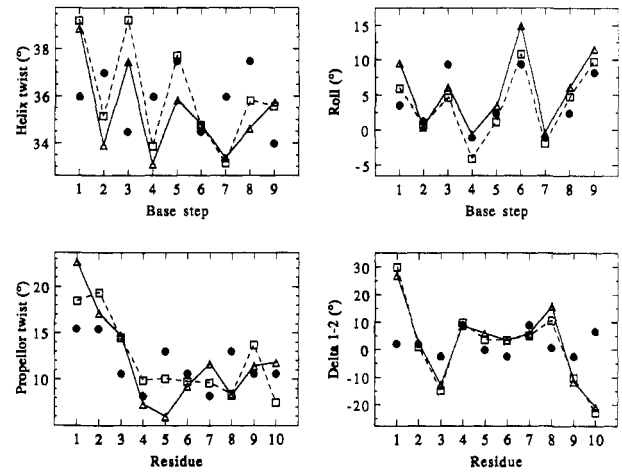

FIGURE 5: Variation in base roll, global helical twist, propeller twist, and δ_{1-2}, the difference of the $C_4'-C_3'$ torsion angle at each end of a base pair, for the refined structures R-A (□) and R-B (△). Solid circles indicate the predictions based on Calladine's (1982) rules. Precise descriptions for the conformational parameters are given by Dickerson et al. (1985).

Structure Refinement and Assignment. Each of these structures is then refined separately, replacing the distance restraints with NOE restraints (Baleja et al., 1990). The refined structures, R-A and R-B, again have similar energies and have an atomic deviation of less than 0.8 Å. These two refined structures are fit to each other (over all atoms) in Figure 3C. The bases superimpose best, having an rms deviation of 0.6 Å. The ribose sugars have rms deviations of 0.86 Å, and the phosphate backbone, 1.05 Å. The less precise determination of the backbone may be due to increased motion (Hogan & Jardetzky, 1980), although there is less data giving direct information on the backbone conformation. Both structures are B type in character and have rms deviations approximately 1.5 Å from canonical B-type DNA. Excluding the terminal base pairs, they exhibit 10.3 ± 0.2 base pairs per turn of the helix, with an average twist angle of 35.3° and an average spacing between base pairs of 3.4 Å. The structures bend slightly into the major groove, with the R-B bending somewhat more and being more A-DNA-like, despite starting from B-type DNA (Table II). If either half of the molecules are fit to each other (top or bottom 5 base pairs), the rms deviation for the superimposed atoms is 0.65 Å, and the rms deviation for the remainder is 1.3 Å.

The convergence of the DNA structures from the widely different starting structures indicates that the final refined structures, R-A and R-B, provide reasonable representations of the structure of the half-operator in solution and therefore are discussed in some detail below. The refinement procedure results in little change from the average MD structure (less than 0.1-Å rms deviation). This is consistent with structures that are energetically near a global minimum. Long restrained molecular dynamics simulations, which sample much more conformational space than energy minimization, could not be undertaken because of the computer time required (Baleja et al., 1990).

The NOE residual factor *R* is used to monitor the fit of the NOEs calculated from a structure to observed NOE intensities (Lefèvre et al., 1987; Gupta et al., 1988; Baleja & Sykes, 1988):

$$R = \frac{\sum|\mathrm{NOE}_{obs} - \mathrm{NOE}_{calc}|}{\sum \mathrm{NOE}_{obs}} \qquad (8)$$

where the summation runs over the number of observables. To keep an analogy with the standard crystallographic *R* factor, weighting factors based on the standard deviation in the NOE intensity are not included. Figure 4 shows the improvement of the NOE *R* factor as molecular structure determination proceeds. The first reduction in the *R* factor arises from energy minimization of initial models A and B. During molecular dynamics calculations, the *R* factor steadily decreases for both runs to a value near 0.27. Up to 20 ps, the comparison between observed and calculated NOE intensities is not iterated upon directly, but here the structure is attempting to meet distance restraints inferred from the NOE intensities. Structures obtained by averaging over the last 5 ps of the molecular dynamics calculations produce structures MD-A and MD-B, each with NOE *R* factors of about 0.24. Final refinement is based on the comparison between observed NOEs and NOE intensities calculated by using a full matrix analysis procedure and the differential motion model. After energy minimization the NOE *R* factor is less than 0.19 for both structures R-A and R-B. We have shown (Baleja et al., 1990) that such an *R* factor indicates that the structure is consistent with the NOE data within reasonable experimental error.

Structural Features. As suggested by the overall shape of the molecules and the similar energies and NOE *R* factors, convergence to the same structure is also seen in details of the spatial relationships of the base pairs. Four of these conformational parameters, helical twist, roll, propeller twist, and difference in δ torsion angles across the base pair (Figure 5), are often predicted from a set of rules based on the crystal structures of the B-DNA dodecamer and other A-type DNA molecules [for a dicussion, see Dickerson (1983) and Dickerson et al. (1985)]. Calladine's rules (1982) predict only the roll angle well, which is the rotation that the base pair makes with respect to its adjacent base pair about its long axis perpendicular to the overall helix axis. The generally poor correlations for the other parameters have been observed in a number of studies, suggesting that the Calladine steric clash model is insufficient to account for all local helix perturbations (Dickerson et al., 1985; Scalfi Happ et al., 1988; Nilges et al., 1987; Lefèvre et al., 1987; Privé et al., 1987). For example, the twist angle is anticorrelated about base step 3–4. The interstrand purine–purine clash is avoided at this pyrimidine–purine step instead by base-pair roll, by a large negative

Table IV: Dihedral Angles for L10

residue	χ	α	β	γ	δ	ε	ζ	P
T_1	−126.6			62.1	136.9	−174.5	−86.7	155.3
C_2	−107.4	−72.4	187.1	55.8	124.1	−180.9	−88.9	135.1
T_3	−115.0	−65.7	174.9	57.6	112.4	−177.6	−96.8	117.7
A_4	−105.9	−65.9	172.1	62.5	120.9	−185.5	−93.1	129.1
T_5	−115.9	−62.3	172.1	61.5	122.1	−177.8	−98.7	129.7
C_6	−111.1	−67.3	175.1	61.5	121.5	−175.0	−91.8	126.0
A_7	−111.2	−72.4	174.7	57.7	117.1	−181.5	−99.3	159.5
C_8	−109.6	−68.5	177.4	60.5	131.3	−184.3	−89.5	143.8
C_9	−116.0	−63.6	176.6	59.4	116.4	−172.2	−90.5	122.0
G_{10}	−116.8	−80.7	179.3	50.6	103.2			96.9
C_{11}	−115.6			67.4	124.9	−173.5	−91.0	134.0
G_{12}	−106.5	−74.7	182.7	55.9	127.1	−178.8	−102.3	138.1
G_{13}	−116.4	−71.5	182.8	55.6	118.1	−178.9	−84.0	125.6
T_{14}	−113.1	−63.9	170.7	57.1	111.7	−174.7	−91.5	116.1
G_{15}	−108.3	−67.0	174.9	55.7	117.0	−185.4	−92.1	125.7
A_{16}	−112.5	−64.0	178.2	56.7	118.8	−176.9	−92.1	125.8
T_{17}	−116.9	−62.7	169.7	59.1	111.5	−173.4	−104.1	116.4
A_{18}	−102.5	−61.1	166.3	64.5	126.1	−177.6	−91.7	136.1
G_{19}	−106.7	−67.2	171.1	58.7	122.3	−178.1	−102.7	130.0
A_{20}	−100.8	−71.0	166.7	61.2	108.3			134.7
av	−113.0	−64.9	175.1	59.0	119.6	−178.1	−89.5	128.2
B DNA	−98.0	−46.0	−147.0	36.0	156.0	−195.0	−96.0	192.0
A DNA	−154.0	−87.0	−152.0	45.0	83.0	−182.0	−47.0	14.0
dodec[a]	−117.0	−63.0	171.0	54.0	123.0	−169.0	−108.0	

[a] Average dihedral angles for the DNA dodecamer, CGCGAATTCGCG, (Dickerson & Drew, 1981). Numerical values were not reported for the sugar pucker, which ranged from C2′ endo (P = 162 ● 18°) through C1′ exo (P = 126 ± 18°) to O1′ endo (P = 90 ± 18°), with one residue, G24, C3′ endo (P = 18 ± 18°).

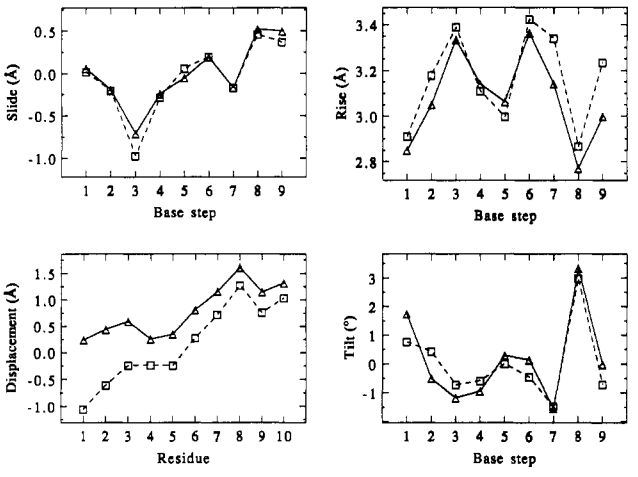

FIGURE 6: Variation in rise, displacement, slide, and tilt angles for the refined structures R-A (□) and R-B (△).

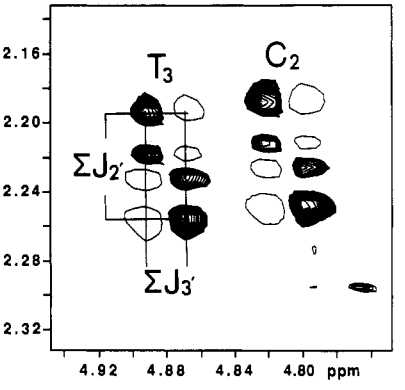

FIGURE 7: Expansion of 2′ ↔ 3′ DQF-COSY cross-peaks for residues C_2 and T_3. $\sum J_{2'}$ and $\sum J_{3'}$ indicate the outer separations of the multiplet lines (approximately 30 and 12 Hz, respectively).

slide (the distance between midpoints of the base pairs when viewed down the helix axis), and by increasing the rise between base pairs slightly (Figure 6). The twist angle, propeller twist angles, and the difference in δ torsion angles across the base pair show less dependence on sequence than is observed in crystal structures, where external forces can result in the perturbation of the DNA structure. For example, although the precision of crystallographic determinations cannot be disputed when compared to other methods, DNA oligomers of mixed sequence have a marked tendency to crystallize in a A-type DNA form, whereas the B-type DNA is clearly preferred in solution (Dickerson et al., 1985).

The distance from the C6–C8 vector of the base pair to the helix axis, the displacement, averages +4 Å for A DNA and −0.2 Å for B DNA. Although certainly not large in magnitude here (Figure 6), slightly more positive values for R-B reflect parts of its A-DNA-type character. The small displacements and tilt angles (a tilt angle being the angle that the normals of the adjacent mean base-pair planes make with each other) indicate that local twist axes are nearly the same and are coincident with the global twist axis. This is less true for the terminal base pairs. A reduction in the network of experimental restraints at the ends of the duplex allows the structure to be more at the mercy of any inadequacies in the molecular force field description. More importantly, chain-termination effects cause a discontinuity in the electrostatic and van der Waals interactions of the peripheral base pairs, likely causing some structural distortions (Grütter et al., 1988).

It is clear that R-A and R-B are essentially equivalent structures. Therefore we have averaged and energy minimized these two structures to form a structure R_{av}, which has, of course, the same R factor as R-A and R-B (0.185). This structure is shown in Figure 3D and has been rotated by 90° from that of Figure 3C to present the view from a different perspective. Torsion angles for this structure are shown in Table IV. Angles for individual structures, R-A and R-B, varied by less than 3° from the values listed here, except for the 5′ terminal γ angles (±5°). The angles are near the averages observed in the crystal structures of B-DNA sequences (Dickerson et al., 1985; Cruse et al., 1986; Dickerson & Drew, 1981; Privé et al., 1987). Less diversity is displayed for each type of dihedral angle, which may be due to averaging

in solution. Note that the initial models chosen are derived from fiber diffraction data and, although adequate as initial models, do not necessarily have realistic dihedral angles. The fact that both determinations of L10 in solution by NMR arrive at a reasonable conformation gives credence to the methodology employed.

However, the ϵ and ζ angles appear to be slightly different from the average angles from single-crystal structures. This is, in turn, an average between the B_I trans–gauche$^-$ conformation (for the ϵ and ζ torsion angles, respectively) and the less common B_{II} gauche$^-$–trans conformation seen in the crystallographic determinations. The average B_I ϵ angle observed in crystal structures of B DNA is near −173°, and the ζ angle is about −85°, which are the angles observed for this decamer in solution. It is not surprising that the B_I conformation should be the one exclusively preferred in solution, as the B_{II} conformation tends to lock the preceding sugar ring in a C2′ endo conformation and restricts mobility (Privé et al., 1987; Dickerson & Drew, 1981). The B_{II} conformation appears to be an artifact of crystal-packing forces (Dickerson et al., 1987). Note also that β angles following a B_I phosphate conformation average 176° in B-DNA crystal structures, which is the average observed here in solution.

Sugar rings are centered on a C1′ exo range of pseudorotational angle, except the 3′-terminal residues 10 and 20, which are O1′ endo (Altona & Sundaralingam, 1972). No attempt was made to interpret our results for conformational averaging between C2′ endo and C3′ endo, or any other two possible sugar ring conformations (Zhou et al., 1988). Elegant work with coupling constant data (Hosur et al., 1988) indicates little conformational averaging, and that a single pseudorotation angle intermediate between C2′ and C3′ endo is adequate to explain the experimental data for each residue.

The base-stacking arrangement is typical for B-type DNA (Figure S-6). A greater overlap is found for purine–pyrimidine steps (steps 4–5 and 7–8) than for pyrimidine–purine (steps 3–4, 6–7, and 9–10), as is observed in the crystal structure of a DNA dodecamer (Dickerson & Drew, 1981), and as predicted by fiber diffraction studies (Arnott et al., 1969). The stacking arrangement appears symmetrical across the helix axis, as expected for B DNA, and as seen in the dodecamer.

Discussion

There are a number of limitations on the accuracy of the structure determination presented here. Especially in the absence of a chosen molecular mechanics force field, it is still insufficiently understood to what extent the few hundred NOE intensities (or the derived short distances) and dihedral angles define, including a description of the dynamics, the three-dimensional structure of oligonucleotides (van de Ven & Hilbers, 1988). For example, although the structures of the L10 decamer are experimentally determined to have phosphates in a B_I conformation, both starting models also have B_I phosphates, and there are no experimental data available that unambiguously distinguish the two conformations, B_I and B_{II}, from one another.

Several lines of evidence however do support a B_I conformation. The positioning of the nucleotide units relative to each other, in the presence of even an approximate force field, is likely sufficient to localize the backbone torsion angles (Nilges et al., 1987). In crystal structures, the B_I conformation is more common than the B_{II} conformation, especially in regions of the DNA not affected by intermolecular hydrogen bonds or crystal contacts. This suggests that the B_I configuration is lower in energy and is more stable. Some B_{II} conformations are observed in B DNA that have ϵ angles near −120°. This would correspond to a coupling constant between the 3′ proton and the nearest phosphate greater than 10 Hz (Dickerson et al., 1985; Hosur et al., 1988). Figure 7 features an expansion of a typical nonterminal 2′ ↔ 3′ cross-peak, with outer separations of the multiplet lines indicated in the 2′ frequency dimension by $\sum J_{2'}$ and in the 3′ dimension by $\sum J_{3'}$. These outer separations reflect the sum of coupling constants to the 2′-proton ($\sum J_{2'} = J_{2'3'} + J_{2'2''} + J_{1'2'}$) and to the 3′ proton ($\sum J_{3'} = J_{2'3'} + J_{2''3'} + J_{3'4'} + J_{3'P}$), with the active coupling, $J_{2'3'}$, giving the antiphase characteristic to the cross-peak pattern. $\sum J_{3'}$ and $\sum J_{2'}$ are approximately 12 and 30 Hz, respectively. For all sugar geometries, $J_{2'3'}$ is never less than 6 Hz (Hosur et al., 1988). From the absence of correlations in Figure 2, $J_{2''3'}$ is very near zero. However, all 3′–4′ correlations are seen (Figure S-4), implying a $J_{3'4'}$ of at least 2 Hz. Therefore, for all nonterminal residues, the 3′ ↔ P coupling constant is small (<5 Hz). From Figure 23 of the review by Hosur et al. (1988), the near 180° ϵ angles obtained here for a B_I conformation are consistent with this coupling constant being quite small.

The MD calculation could start with a model including B_{II} phosphates, or more robust methods in conformational sampling could be employed, such as dynamical simulated annealing [e.g., Holak et al. (1988)]. The accuracy of the method described here appears to be between approximately 0.8 (from the rms deviation of structures R-A and R-B) and 1.8 Å (from the interconversion between the A- and B-DNA fiber diffraction models). The 1.8-Å limit is potentially an overestimate, which would be improved by choosing better models that take sequence-dependent variations into account. The incorporation of NOE-based refinement circumvents inaccuracies in distances caused by neglecting indirect cross-relaxation pathways for the NOE between two protons and therefore one of the most serious problems in structure determination by NMR. Often at long mixing times, NOE cross-peaks can be observed between protons even 6 and 7 Å apart, but this information is lost since it cannot be directly related to a distance. The NOE refinement method indicates a way to obtain more parameters that cover a greater range of conformational space, and to better determine structures in solution by NMR.

Our initial motion model assumes that all interproton vectors tumble isotropically and are of invariant length. Therefore, besides some anisotropic behavior of motion, the NOE between protons not fixed by covalent geometry can be modified by distance fluctuations. The observed cross-peak is an approximate r^6 weighted average, and is therefore dominated by the close approach of the atoms. A more sophisticated model could be used (Lipari & Szabo, 1982; Rowan et al., 1974; Keepers & James, 1984; Noggle & Shirmer, 1971), but the incorporation of model-free empirical parameters here is sufficient with respect to the precision of the NOE data experimentally collected.

Although it is preferable to use NOEs directly in a long restrained MD run, thereby performing a better search of conformational space and eliminating the step of making rather approximate distance determinations, the computational time would be cumbersome. Moreover, it is unclear whether the structure produced by using the NOE directly in MD runs would be essentially more correct than that produced by using approximate distances, followed by refinement with NOEs, given any inadequacies in the motional model used here (Baleja et al., 1990).

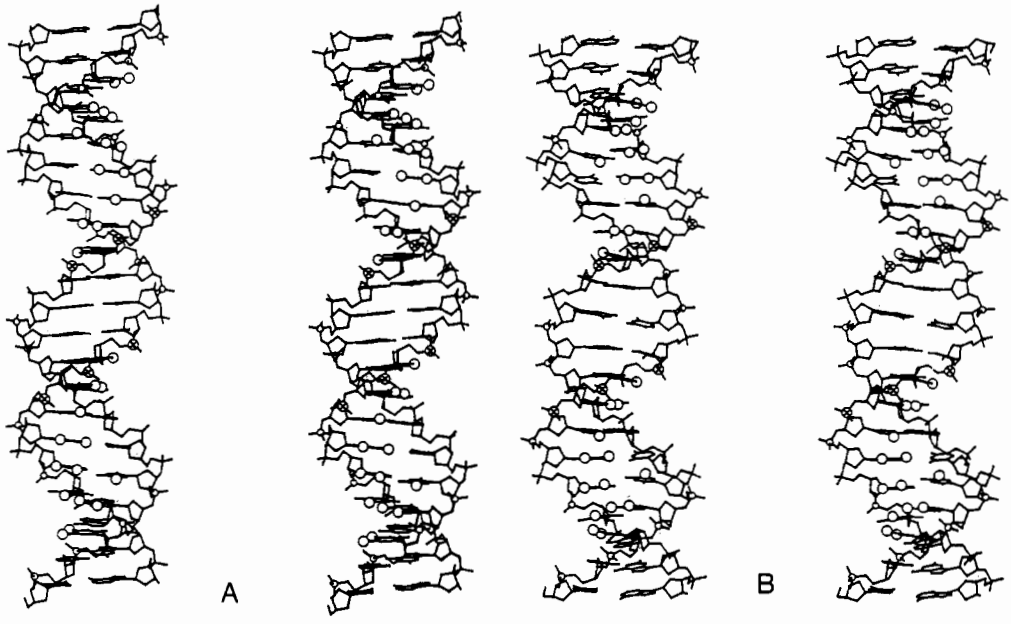

FIGURE 8: Models for the 21 base-pair O_R3 operator DNA. (A) B-type DNA. (B) The operator constructed by using the solution structure of the half-operator (L10) determined by use of NMR, restrained molecular dynamics, and NOE-based refinement. Spheres in the major groove indicate the sequence-specific contacts made by the cro repressor protein. Ethylation of the phosphates denoted by cross-hatched spheres prevent cro repressor binding. Other phosphate atoms proposed to be involved in nonspecific protein contact are indicated by smaller spheres.

Figure 8 shows two structures for the full O_R3 operator DNA. On the left is the classical B-DNA conformation first used to model the interaction between cro repressor and its operator DNA (Ohlendorf et al., 1982). Spheres in the major groove represent all atoms to which the protein is proposed to make sequence-specific contact (Takeda et al., 1989). Ethylation of the phosphates represented by cross-hatched spheres prevents cro repressor binding (Ptashne et al., 1980). Other proposed nonspecific contacts to phosphates are shown as small spheres. In Figure 8B, a consensus operator is built with two half-operators, using the conformation of the decamer determined above. The L10 sequence does not include the middle base pair of the full operator. The central G·C base pair is constructed here by using average dihedral and twist angles and rise values between base pairs. The similarity in the protein contact site between the models is striking and suggests that the monomeric unit of cro repressor recognizes DNA by the pattern of hydrogen bonds and van der Waals contacts presented by the DNA, and not by any gross morphological feature.

This study does not preclude any improvements in the model such as adjustments in the positioning of the two halves of the operator, for example, by bending at the center. The significance of the difference between L10 and classical B DNA awaits further improvements in the NMR methodology, and confimation of the structure by other means, such as X-ray crystallography. A preliminary crystallographic determination of the cro dimer–17 base-pair O_R3 complex (Brennan & Matthews, 1989b) indicates that, at the center of the operator, the DNA is overwound, and the minor groove is somewhat compressed.[3] A detailed comparison of the DNA structure in the cro repressor sequence-specific site between free and protein-bound DNA is expected and will be possible in the very near future.

[3] The cro monomers retain nearly the same structure as observed in the crystal structure of the free protein (Anderson et al., 1981), although there are substantial changes in the relative positions of monomers in the cro dimer.

ACKNOWLEDGMENTS

We thank Dr. R. Dickerson for the helix analysis and Dr. W. van Gunsteren for the GROMOS biomolecular simulation programs.

SUPPLEMENTARY MATERIAL AVAILABLE

One table listing the experimental restraints used in the structure determination and six figures showing spectra used for resonance assignments and a structure illustrating base stacking (15 pages). Ordering information is given on any current masthead page. A listing of the coordinates of the final refined structure of L10, R_{av}, is available from the authors.

REFERENCES

Abragam, A. (1961) *The Principles of Nuclear Magnetism*, Oxford University Press, London.

Aggarwal, A. K., Rodgers, D., Drottar, M., Ptashne, M., & Harrison, S. C. (1988) *Science (Washington, D.C.) 242*, 899–907.

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

Anderson, W. F., Ohlendorf, D. H., Takeda, Y., & Matthews, B. W. (1981) *Nature (London) 290*, 754–758.

Arnott, S., & Hukins, D. W. L. (1972) *Biophys. Biochem. Res. Commun. 47*, 1504–1509.

Arnott, S., Dover, S. D., & Wonacott, A. J. (1969) *Acta Crystallogr., Sect. B 25*, 2192–2206.

Baleja, J. D., & Sykes, B. D. (1988) *Biophys. J. 53*, 80a.

Baleja, J. D., Moult, J., & Sykes, B. D. (1990) *J. Magn. Reson.* (in press).

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984) *J. Chem. Phys. 81*, 3684–3690.

Bodenhausen, G., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 1304–1309.

Boelens, R., Koning, T. M. G., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1989) *J. Magn. Reson. 82*, 290–308.

Brennan, R. G., & Matthews, B. W. (1989a) *J. Biol. Chem. 264*, 1903–1906.

Brennan, R. G., & Matthews, B. W. (1989b) *Trends Biochem. Sci. 14*, 286–290.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.

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

Chary, K. V. R., Hosur, R. V., Govil, G., Chen, C., & Miles, H. T. (1988) *Biochemistry 27*, 3858–3867.

Chazin, W. J., Wüthrich, K., Hyberts, S., Rance, M., Denny, W. A., & Leupin, W. (1986) *J. Mol. Biol. 190*, 439–453.

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

Cohen, J. S. (1987) *Trends Biochem. Sci. 12*, 133–135.

Cruse, W. B. T., Salisbury, S. A., Brown, T., Cosstick, R., Eckstein, F., & Kennard, O. (1986) *J. Mol. Biol. 192*, 891–905.

Cuniasse, Ph., Sowers, L. C., Eritja, R., Kaplan, B., Goodman, M. F., Cognet, J. A. H., Le Bret, M., Guschlbauer, W., & Fazakerley, G. V. (1989) *Biochemistry 28*, 2018–2026.

de Vlieg, J., Boelens, R., Scheek, R. M., Kaptein, R., & van Gunsteren, W. F. (1986) *Isr. J. Chem. 27*, 181–188.

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

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

Dickerson, R. E., Kopka, M. L., & Pjura, P. (1985) in *Biological Macromolecules and Assemblies*, Vol. 2, Nucleic Acids and Interactive Proteins (Jurnak, F. A., & McPherson, A., Eds.) pp 37–126, 490–493, Wiley-Interscience, New York.

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

Feigon, J., Wright, J., Leupin, W., Denny, W. A., & Kearns, D. R. (1982) *J. Am. Chem. Soc. 104*, 5540–5541.

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

Gronenborn, A. M., & Clore, G. M. (1985) *Prog. Nucl. Magn. Reson. Spectrosc. 17*, 1–33.

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

Grütter, R., Otting, G., Wüthrich, K., & Leupin, W. (1988) *Eur. Biophys. J. 16*, 279–286.

Gupta, G., Sarma, M. H., & Sarma, R. H. (1988) *Biochemistry 27*, 7909–7919.

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

Hogan, M. E., & Jardetzky, O. (1980) *Biochemistry 19*, 3460–3468.

Holak, T. A., Engström, Å., Kraulis, P. J., Lindeberg, G., Bennich, H., Jones, T. A., Gronenborn, A. M., & Clore, G. M. (1988) *Biochemistry 27*, 7620–7629.

Hore, P. J. (1983) *J. Magn. Reson. 55*, 283–300.

Hosur, R. V., Ravikumar, M., Chary, K. V. R., Sheth, A., Govil, G., Zu-kun, T., & Miles, H. T. (1986) *FEBS Lett. 205*, 71–76.

Hosur, R. V., Govil, G., & Miles, H. T. (1988) *Magn. Reson. Chem. 26*, 927–944.

Jordan, S. R., & Pabo, C. O. (1988) *Science (Washington, D.C.) 242*, 893–899.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179–182.

Keepers, J. W., & James, T. L. (1982) *J. Am. Chem. Soc. 104*, 929–939.

Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404–426.

Kim, J. G., Takeda, Y., Matthews, B. W., & Anderson, W. F. (1987) *J. Mol. Biol. 196*, 149–158.

Kirpichnikov, M. P., Hahn, K. D., Buck, F., Rüterjans, H., Chernov, B. K., Kurochkin, A. V., Skryabin, K. G., & Bayev, A. A. (1984) *Nucleic Acids Res. 12*, 3551–3561.

Koudelka, G. B., Harrison, S. C., & Ptashne, M. (1987) *Nature (London) 326*, 886–888.

Koudelka, G. B., Harbury, P., Harrison, S. C., & Ptashne, M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4633–4637.

Lee, M., Chang, D.-K., Hartley, J. A., Pon, R. T., Krowicki, K., & Lown, J. W. (1988) *Biochemistry 27*, 445–455.

Lefèvre, J.-F., Lane, A. N., & Jardetzky, O. (1987) *Biochemistry 26*, 5076–5090.

Lipari, G., & Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546–4559.

Majumdar, A., & Hosur, R. V. (1989) *Biophys. Biochem. Res. Commun. 159*, 886–892.

McKay, D. B., & Steitz, T. A. (1981) *Nature (London) 290*, 744–749.

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

Nilsson, L., & Karplus, M. (1986) *J. Comput. Chem. 7*, 591–616.

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

Noggle, J. H., & Schirmer, R. E. (1971) *The Nuclear Overhauser Effect*, Academic Press, New York.

Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., & Matthews, B. W. (1982) *Nature (London) 298*, 718–723.

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

Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature (London) 335*, 321–329.

Pabo, C. O., & Sauer, R. T. (1984) *Annu. Rev. Biochem. 53*, 293–321.

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

Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., & Dickerson, R. E. (1987) *Science (Washington, D.C.) 238*, 498–504.

Ptashne, M., Jeffrey, A., Johnson, A. D., Maurer, R., Meyer, B. J., Pabo, C. O., Roberts, T. M., & Sauer, R. T. (1980) *Cell 19*, 1–11.

Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621–649.

Rowan, R., McCammon, J. A., & Sykes, B. D. (1974) *J. Am. Chem. Soc. 96*, 4773–4780.

Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327–337.

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

Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., & van Boom, J. H. (1983a) *J. Am. Chem. Soc. 105*, 2914–2916.

Scheek, R. M., Zuiderweg, E. R. P., Klappe, K. J. M., van Boom, J. H., Kaptein, R., Rüterjans, H., & Beyreuther, K. (1983b) *Biochemistry 22*, 228–235.

Solomon, I. (1955) *Phys. Rev. 99*, 559–565.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Suzuki, E., Pattabiraman, N., Zon, G., & James, T. L. (1986) *Biochemistry 25*, 6854–6865.
Takeda, Y., Ohlendorf, D. H., Anderson, W. F., & Matthews, B. W. (1983) *Science (Washington, D.C.) 221*, 1020–1026.
Takeda, Y., Sarai, A., & Rivera, V. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 439–443.
Tidor, B., Irikura, K., Brooks, B. R., & Karplus, M. (1983) *J. Biomol. Struct. Dyn. 1*, 231–252.
van de Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1–38.
van Gunsteren, W. F., Boelens, R., Kaptein, R., Scheek, R. M., & Zuiderweg, E. R. P. (1985) in *Molecular Dynamics and Protein Structure* (Hermans, J., Ed.) pp 92–99, Polycrystal Bookservice, Western Springs, IL.
Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr., & Weiner, P. (1984) *J. Am. Chem. Soc. 106*, 765–784.
Wolberger, C., Dong, Y., Ptashne, M., & Harrison, S. C. (1988) *Nature (London) 335*, 789–795.
Zhou, N., Monogaran, S., Zon, G., & James, T. L. (1988) *Biochemistry 27*, 6013–6020.

A Circular Dichroic Study of Helical Structure in Flagellar Dynein†

Gabor Mocz* and I. R. Gibbons
Pacific Biomedical Research Center, University of Hawaii, Honolulu, Hawaii 96822
Received October 6, 1989; Revised Manuscript Received January 18, 1990

ABSTRACT: The circular dichroic spectra of outer arm dynein from sea urchin sperm flagella, of its separated α and β heavy-chain complexes, and of the two major fragments produced by tryptic digestion of the β heavy chain have been measured over the range 190–240 nm. Although the spectra show significant individuality, in all cases they qualitatively resemble those of typical globular proteins with mixed regions of α-helix and β-sheet (α/β-type structure) or with separate α-helix- and β-sheet-rich regions ($\alpha+\beta$-type structure). Quantitative analyses of the spectra by both constrained and unconstrained least-squares curve-fitting procedures indicate that the intact dynein contains ~26% α-helix. The separated β heavy-chain complex and its ATPase-containing amino-terminal domain (fragment A) both have spectra resembling that of intact dynein, and they appear to contain 32% and 23% α-helix, respectively. The carboxy-terminal domain of the β heavy chain (fragment B) and the separated α heavy chain have significantly different spectra; however, they each appear to contain 26–36% α-helix. These data suggest that dynein does not contain an extensive α-helical domain, such as is found in the carboxy-terminal rod region of the other motor proteins myosin and kinesin.

Dynein ATPases play a central role in the motility of cilia and sperm flagella, as well as in some forms of cytoplasmic microtubule-based motility (Gibbons, 1981, 1988). In the axonemes of cilia and flagella, dynein constitutes the inner and outer arms on the doublet microtubules where its function is to generate localized sliding between the doublets by coupling the chemical cycle of ATP binding and hydrolysis with cyclic detachment and reattachment to successive sites along the adjacent doublets. Cytoplasmic dyneins are believed to play a role in vesicle transport in nerve axons (Schnapp & Reese, 1989; Schroer et al., 1989) and may also be involved in mitosis (Pratt, 1984).

Electron microscopy has demonstrated that soluble outer arm dynein has a structure consisting of two or three globular heads, each ~15 nm in diameter, that are connected to a common base by a flexible stem 25–30 nm long (Johnson & Wall, 1983; Goodenough & Heuser, 1984). The examination of the dynein arms in situ on the axonemes has shown that their fixed structural attachment to the A subfiber of each doublet microtubule involves the basal region of their flexible stem and that the heads form the ATP-sensitive site of interaction with the B subfiber of the adjacent doublet microtubule. The cytoplasmic dynein from brain has a similar two-headed electron microscopic structure (Vallee et al., 1988).

Analysis of polypeptide composition has indicated that axonemal dyneins are complex proteins with 9–10 subunits of 3 distinct categories of size. The outer arm dynein from sea urchin sperm flagella is comprised of one α and one β heavy chain of ~470 kDa each, a set of three intermediate-size chains of 76–122 kDa, and at least four light chains of 15–25 kDa (Gibbons et al., 1976; Mocz et al., 1988). The outer arm dyneins of *Chlamydomonas* flagella and *Tetrahymena* cilia contain three heavy-chain subunits (Johnson, 1985; Piperno & Luck, 1979). In all cases, the number of distinct heavy chains is equal to the number of globular heads visible by electron microscopy, which strongly suggests that each globular head/stem unit corresponds to a single heavy-chain polypeptide, with each of these units having a similar tadpole-shaped structure. Direct evidence to support this hypothesis has been obtained by electron microscopy of the separated α and β heavy-chain subunits of outer arm dynein from sea urchin sperm flagella (Sale et al., 1985). The separated β heavy-chain complex is sufficient to produce microtubule motility in an in vitro motility assay system (Sale & Fox, 1988; Vale et al., 1989).

Following the initial observations of Ogawa and Mohri (1975), recent work from this laboratory has shown that

† Supported by Grant GM30401 from the National Institute of General Medical Sciences.
* Address correspondence to this author at Kewalo Laboratory, 41 Ahui St., Honolulu, HI 96813.

0006-2960/90/0429-4839$02.50/0