VOLUME 115, NUMBER 4
FEBRUARY 24, 1993

JOURNAL
OF THE
AMERICAN
CHEMICAL
SOCIETY

Base Inclinations in Natural and Synthetic DNAs

Ping-Jung Chou and W. Curtis Johnson, Jr.*

Contribution from the Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331. *Received January 22, 1992*

Abstract: A sophisticated algorithm is developed to analyze flow linear dichroism data on nucleic acids for individual base inclinations. Measured absorption and linear dichroism data for synthetic AT and GC polymers and natural DNAs are analyzed. The reliability of the algorithm is tested on data for the synthetic polymers, and the results are similar to earlier, more straightforward analyses. For the first time, specific base inclinations are derived for all four bases individually from the linear dichroism data for natural deoxyribonucleic acids. For B-form DNA in aqueous solution at moderate salt the inclinations from perpendicular are as follows: d(A) = 16.1 ± 0.5; d(T) = 25.0 ± 0.9; d(G) = 18.0 ± 0.6; d(C) = 25.1 ± 0.8 deg. Our results indicate that the bases in synthetic and natural DNAs are not perpendicular to the helix axis, not even in the B form.

Introduction

Watson and Crick depicted their helical structure for DNA with 10 base pairs per turn with the bases perpendicular to the helix axis. This was consistent with Wilkins' X-ray patterns for fibers of DNA at high humidity, the B form. Although the data in diffraction patterns from fibers are limited, subsequent model building indicated a 10-fold repeat with bases perpendicular to the helix axis for the B form.[1] However, DNA is known to be polymorphic,[2-9] with the particular structure sensitive to sequence, cation type, temperature, and solvent (or, in the case of fibers and crystals, the humidity). A structural model built on X-ray diffraction, however, may depend on packing forces, and not actually exist in solution where DNA molecules are relatively free.

Linear dichroism (LD) is a method for determining the inclination angle of a given kind of base in a DNA molecule in solution.[10-28] It is based on the facts that (1) each kind of base

(1) Langridge, R.; Marvin, D. A.; Seeds, W. E.; Wilson, H. R.; Hooper, C. W.; Wilkins, M. H. F.; Hamilton, L. D. *J. Mol. Biol.* **1960**, *2*, 38–61.
(2) Fuller, W.; Wilkins, M. H. F. *J. Mol. Biol.* **1965**, *12*, 60–80.
(3) Arnott, S.; Hukins, H. W. L. *J. Mol. Biol.* **1972**, *149*, 761–786.
(4) Bram, S.; Tougard, P. *Nature (London) New Biol.* **1972**, *239*, 128–131.
(5) Girod, J. C.; Johnson, W. C., Jr.; Huntington, S. K.; Maestre, M. F. *Biochemistry* **1973**, *12*, 5092–5096.
(6) Ivanov, V. I.; Minchenkova, L. E.; Schyolkina, A. K.; Poletayev, A. I. *Biopolymers* **1973**, *12*, 89–110.
(7) Pohl, F. M. *Nature* **1976**, *260*, 365–366.
(8) Leslie, A. G. W.; Arnott, S.; Chandrasekaran, R.; Ratliff, R. L. *J. Mol. Biol.* **1980**, *143*, 49–72.
(9) Shakked, Z.; Kennard, O. In *Structural Biology*; McPherson, A., Jurnak, F., Ed.; Wiley: New York, 1983.
(10) Cavalieri, L. F.; Rosenberg, B. H.; Rosoff, M. *J. Am. Chem. Soc.* **1956**, *78*, 5235–5238.
(11) Gray, D. M.; Rubenstein, I. *Biopolymers* **1968**, *6*, 1605–1631.
(12) Wada, A. *Appl. Spectrosc. Rev.* **1972**, *6*, 1–30.
(13) Ding, D.; Rill, R.; van Holde, K. E. *Biopolymers* **1972**, *11*, 2109–2124.
(14) Hofrichter, J.; Eaton, W. A. *Annu. Rev. Biophys. Bioeng.* **1976**, *5*, 511–560.
(15) Nordén, B. *Appl. Spectrosc. Rev.* **1978**, *14*, 157–248.
(16) Hogan, M.; Dattagupta, N.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 195–199.
(17) Rizzo, V.; Schellman, J. *Biopolymers* **1981**, *20*, 2143–2163.
(18) Charney, E.; Yamaoka, K. *Biochemistry* **1982**, *21*, 834–842.
(19) Lee, C.; Charney, E. *J. Mol. Biol.* **1982**, *161*, 289–303.
(20) Diekmann, S.; Hillen, W.; Jung, M.; Wells, R. D.; Pörschke, D. *Biophys. Chem.* **1982**, *15*, 157–167.
(21) Matsuda, K.; Yamaoka, K. *Bull. Chem. Soc. Jpn.* **1982**, *55*, 1727–1733.
(22) Matsuoka, Y.; Nordén, B. *Biopolymers* **1982**, *21*, 2433–2452.
(23) Matsuoka, Y.; Nordén, B. *Biopolymers* **1983**, *22*, 1731–1746.
(24) Edmondson, S. P.; Johnson, W. C., Jr. *Biochemistry* **1985**, *24*, 4802–4806.
(25) Charney, E.; Chen, H. H.; Henry, E. R.; Rau, D. C. *Biopolymers* **1986**, *25*, 885–904.
(26) Charney, E.; Chen, H. H. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 1546–1549.
(27) Charney, E. *Q. Rev. Biophys.* **1988**, *21*, 1–60.

0002-7863/93/1515-1205$04.00/0

has different π–π^* transitions with dipole moments of known direction in the base plane; (2) the long DNA molecules can be aligned so that, at least on average, the helical axis lies in the direction of alignment; and (3) the anisotropic absorption of transition dipoles in a base can be expressed as a function of the base inclination angle from perpendicular to the helical axis.

DNA molecules are generally aligned either in films or fibers by the shear forces of flow or by their special polyelectrolyte properties in an orienting electric field. Of course, complete alignment is impossible. Base inclinations are deduced in the case of flow LD by modeling the shear forces in the flow cell, extrapolating to infinite shear, or making use of the variation in LD as a function of wavelength. Base inclinations are usually deduced in the case of electric dichroism by making measurements at various fields and extrapolating to infinite field. The orientation problem may be further complicated by the possible existence of tertiary superstructures, which would prevent complete alignment of the helix axis in the direction of alignment even at infinite shear or infinite field. Recent recognition that bent DNA does exist, typified by the kinetoplast fragments, means that tertiary superstructures deserve serious consideration. Detailed reviews have been written covering these points.[12,14,15,27]

In an LD measurement the absorption is measured parallel and perpendicular to the direction of alignment at one or more wavelengths, and the data are conveniently expressed as the reduced dichroism given by

$$L(\lambda) = [A_{\parallel}(\lambda) - A_{\perp}(\lambda)]/A(\lambda) = \mathrm{LD}(\lambda)/A(\lambda)$$

where $A(\lambda)$ is the normal isotropic absorption at wavelength λ. If the base planes in B-form DNA are nearly perpendicular to the helix axis, then for complete alignment in the absence of complicating factors, $L(\lambda)$ will be -1.5 for the in-plane π–π^* transitions, regardless of the wavelength and corresponding transition dipole directions.

Most electric dichroism work since 1978 has utilized samples of homogeneous length and reduced dichroism at the absorption maximum of 260 nm extrapolated to infinite field.[16,18-20] Measurements have been made on different DNA lengths, with the idea that it should be easier to obtain complete alignment for short lengths of DNA without exterior complications. However, considering all of the data together, it is clear that the shorter the DNA length the *lower* the magnitude of the negative L(260 nm). At one extreme Lee and Charney[19] obtained -1.41 for a DNA length of 9200 base pairs, while Hogan et al.[16] obtained -1.11 for a DNA length of 154 base pairs at the other extreme. Hogan et al.[16] interpreted their data in terms of a base inclination from perpendicular of about 17°. In contrast, Dieckmann et al.[20] and Lee and Charney[19] noted that a bent tertiary structure in the DNA would rationalize the values for L(260 nm) as a function of DNA length; as the DNA length increases it is presumed that the DNA becomes increasingly straight in the orienting electric field. This data would still be consistent with the bases perpendicular to the helix axis if (1) extrapolations to infinite field are not correct or (2) the DNA has a tertiary superstructure so that complete alignment is impossible. Rau and Charney[29] have questioned the extrapolation to infinite field and have provided a model for the orientation of the DNA as a function of field that explains the observed data. When everything is taken into consideration, Charney et al.[25] believe that L(260 nm) = -1.41 for the long DNA molecules is consistent with the Watson–Crick structure and its average base tilts of about 10°.

Flow LD measurements also give a negative reduced dichroism for B-form DNA.[10-12,17,22-24] The data are independent of wavelength between 280 and 250 nm, suggesting that the bases are perpendicular to the helix axis. The reduced dichroism is less negative in the 250–220-nm region, and this change in L has been presumed to be due to out-of-plane n–π^* transitions.[11] In general, workers have interpreted their flow LD data as being consistent with the Watson–Crick model. Our laboratory has extended the LD measurements of nucleic acids into the vacuum UV region to 175 nm.[24,30-33] Our data over this extended range show a reduced dichroism that varies with wavelength for natural B-form DNA,[24,31] indicating that the bases are not perpendicular to the helix axis.

It is not straightforward to relate either electric or flow LD data to base inclinations; the measurement depends not only on base inclinations but also on the angle that the dipole for each transition makes with the axis around which the base is inclining. However, with this extended data we were able to compare the relative values of the reduced dichroism for the 260- and 200-nm π–π^* regions to obtain a minimum average base inclination from perpendicular of about 15° for standard B-form DNA.[24,31] We do not attempt to model our flow or extrapolate our data to infinite alignment. The beauty of extending the LD data to shorter wavelengths is that absolute measurements are not necessary, and base inclinations can be determined from the wavelength dependence (overall spectral shape) of the data. DNA tertiary structure, such as a superhelical coil or simple bending, affects the LD as a multiplicative factor, which affects the values at infinite field or flow but which does not affect the wavelength dependence of the data.[12,14,15,27]

We have also measured the LD of simple repeating double-stranded AT and GC polynucleotides from 320 to 175 nm.[32,33] This data can be decomposed into the individual absorption bands, and since the transition dipole directions are known, it has been analyzed for inclinations and axes of inclination for the various bases. The reduced dichroism for these double-stranded polynucleotides varies with wavelength, indicating that the base planes are not perpendicular to the helix axis. Many workers believe that loss of negative reduced dichroism around 230 nm is due to an n–π^* transition with an out-of-plane transition dipole. We analyzed our data without the 245–212-nm spectral region, and the wavelength dependence of the data still predicted significant inclinations for the bases. Furthermore, the 230-nm feature in the reduced dichroism was found to be due to the angle that the π–π^* transition dipoles made with the inclination axis in this region, and the existence of out-of-plane n–π^* transitions need not be postulated to explain the measurements. If the minimum magnitude of the reduced dichroism for B-form DNA at 223 nm is compared with the maximum magnitude at 260 nm, a minimum average base inclination of about 19° is derived for natural DNA in the standard B form.[24]

Here we develop a sophisticated algorithm in order to analyze the LD data of natural nucleic acids as a function of wavelength for individual base inclinations and axes of inclination. With an algorithm that relies so heavily on the computer, it is important to be sure that the results are not an artifact generated by the computer. Thus we use the method to reanalyze the data for the synthetic AT and GC polynucleotides, which were analyzed in a more straightforward way in the original publications.[32,33] The results of this sophisticated algorithm are in reasonable agreement with the results of the original, simpler analyses. Furthermore, the inclinations and axes of inclination that we derive for the individual bases in B-form DNA predict an L(260 nm) of -1.40 for perfect alignment of the DNA helix axis along the direction of orientation. This agrees with the values obtained by extrapolating electric dichroism data to infinite field for monodispersed samples of long DNAs[19,20] and supports the argument that large electric fields should overwhelm configurational and thermal bending for long DNAs.[27,29] The fact that L(260 nm) = -1.40 at perfect orientation can correspond to significant base inclinations demonstrates that it is important to take into account the relative

(28) Flemming, J.; Pohle, W.; Weller, K. *Int. J. Biol. Macromol.* **1988**, *10*, 248–254.
(29) Rau, D. C.; Charney, E. *Biophys. Chem.* **1983**, *17*, 35–50.
(30) Causley, G. C.; Johnson, W. C., Jr. *Biopolymers* **1982**, *21*, 1763–1780.
(31) Dougherty, A. M.; Causley, G. C.; Johnson, W. C., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 2193–2195.
(32) Edmondson, S. P.; Johnson, W. C., Jr. *Biopolymers* **1985**, *24*, 825–841.
(33) Edmondson, S. P.; Johnson, W. C., Jr. *Biopolymers* **1986**, *25*, 2335–2348.

orientation of transition dipoles to the axes around which the bases incline when interpreting LD data.

Flemming et al.[28] have used infrared LD to investigate the base inclinationof A- and B-form DNA in oriented films. They find inclinations from perpendicular of 28–30° for the A form and 18–30° for the B form, in agreement with our work. Theoretical calculations support large base inclinations in DNA.[34,35] In particular, Sarai et al.[35] find that the origin of the B-form double helix can be attributed in large part to the atomic charge pattern in the base pairs. That is, the base pairs alone have a strong tendency to form a helical structure independent of the backbone. Further, propeller twisting is found to enhance the electrostatic interaction by positioning favored atom pairs closer together. One might expect that, in aqueous solution where the DNA is free of the packing effects found in crystals and fibers, bases may be freer to assume larger propeller twists with the concomitant larger base inclination in order to maximize favorable base–base interactions. Ansevin and Wang[36] have proposed a new model for the Z-form with a fair base inclination. Edmondson used the molecular mechanical program AMBER[37] to investigate the potential energy of conformations consistent with his LD results for poly[d(A)–d(T)].[38] He found that the large 50° propeller twist maximized intrastrand base-stacking interactions and that the total potential energy was comparable to that calculated for X-ray diffraction models of DNA. Large propeller twists do not really preclude hydrogen bonding, because hydrogen bonds are not very directional.

Here we reanalyze our published LD data using a more sophisticated algorithm and recently determined transition dipole directions. Large inclinations are confirmed for the bases in synthetic polymers, and specific inclinations are determined for the first time for the bases in natural DNA.

Methods

Algorithm. Let $y(\lambda_i)$, $i = 1, ..., m$, a tabulated function of wavelength λ, be a measured spectrum and assume that it can be approximated as $Y(\lambda_i)$ by N components:

$$Y(\lambda_i) = \sum_{j=1}^{N} f(\lambda_i, \mathbf{p}_j)$$

Each component $f(\lambda_i,\mathbf{p}_j)$ is an analytical function, and $\mathbf{p}_j$ is a vector of parameters. Decomposition of the spectrum into its N components determines all $\mathbf{p}_j$'s such that $Y(\lambda_i)$ can best approximate $y(\lambda_i)$ for $i = 1, ..., m$.

The sum of squares error (residual) will be used here to measure the goodness of the approximation. Let $\mathbf{x}$, a vector of size n, represent all $\mathbf{p}_j$'s, and let $\mathbf{F}(\mathbf{x})$ be the vector of m residuals:

$$\mathbf{F}(\mathbf{x}) = \mathbf{y} - \mathbf{Y}(\mathbf{x})$$

in which $\mathbf{y}$ and $\mathbf{Y}$ are the vector forms of $y(\lambda_i)$ and $Y(\lambda_i)$, respectively. The problem of decomposition is now equivalent to determining an $\mathbf{x}$ that minimizes the sum of squares error, $\|\mathbf{F}(\mathbf{x})\|_2^2$. Because $\mathbf{F}$ is not a linear function of $\mathbf{x}$ in our case, we will be using a nonlinear least-squares fitting to solve this problem.

On the basis of preliminary studies, the method we chose is the derivative-free Levenberg–Marquardt algorithm,[39] abbreviated as DFLM. From an initial guess of $\mathbf{x}$, DFLM iteratively generates a sequence of approximations toward the minimum for the sum of squares error of $\mathbf{F}$ by

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [C_k\mathbf{D} + \mathbf{J}(\mathbf{x}_k)^T\mathbf{J}(\mathbf{x}_k)]^{-1}\mathbf{J}(\mathbf{x}_k)^T\mathbf{F}(\mathbf{x}_k)$$

at the kth iteration. In this expression C_k is the Levenberg–Marquardt coefficient (a positive real number), $\mathbf{J}(\mathbf{x}_k)$ is the $m \times n$ numerical Jacobian matrix evaluated at $\mathbf{x}_k$, and $\mathbf{D}$ is a diagonal matrix with entries equivalent to the diagonal of $\mathbf{J}(\mathbf{x}_k)^T\mathbf{J}(\mathbf{x}_k)$. The ith row and jth column of the Jacobian matrix at each iteraction is calculated by

$$\frac{\mathbf{F}_i(\mathbf{x} + h\mathbf{u}_j) - \mathbf{F}_i(\mathbf{x})}{h}$$

in which $\mathbf{u}_j$ is the jth unit vector corresponding to each variable and h is a small real number used to approximate the first partial derivative of $\mathbf{F}_i$ with respect to $\mathbf{x}_j$.

The Levenberg–Marquardt coefficient is systematically updated according to the results of the previous iteration. This allows the behavior of DFLM to switch smoothly between the Gauss–Newton and steepest descent algorithms, and it is this flexibility that enables DFLM to locate the global minimum within a multidimensional space much faster than, say, Powell's conjugate gradient algorithm[40] used in our preliminary studies.

The diagonal elements of the $n \times n$ matrix $[\mathbf{J}(\mathbf{x}_k)^T\mathbf{J}(\mathbf{x}_k)]^{-1}$ are the variances of the elements of $\mathbf{x}_k$ at the kth iteration if $\mathbf{F}$ is a linear function of $\mathbf{x}$ and measurement errors are normally distributed.[41] However, we do not know our error distribution, and our function is not linear. Although strictly speaking our diagonal elements are not the variances, they will be related to the variances, and the difference between the diagonal elements for two consecutive iterations will still tell us whether $\mathbf{x}_k$ is more stable than $\mathbf{x}_{k-1}$. Of course one can take the sum of squares error in fitting a spectrum to a minimum, but there is error in the data that is being fit so exactly. Instead we monitor the stability of $\mathbf{x}_k$ through the diagonal elements of $[\mathbf{J}(\mathbf{x}_k)^T\mathbf{J}(\mathbf{x}_k)]^{-1}$ and stop fitting when $\mathbf{x}_k$ is stable.

Fitting Monomer Absorption Spectra. To decompose a monomer absorption spectrum into its constituent bands, we must first choose an analytical function that can best described the shape for each absorption band. Gaussian or Lorentzian functions are most often used in the decomposition of UV or IR spectra.[28,32] However, the shape of a UV absorption band is generally asymmetric, and this is well represented by the log-normal function.[42,43] With four parameters (band center μ, an integrated intensity ζ, width at half-height σ, and skewness ρ), the log-normal function for a single band as a function of wavelength is

$$A(\lambda) = \zeta \exp\left\{-\tfrac{1}{2}\left[\frac{\ln (G/R)}{Z} - Z\right]^2\right\} \Big/ \sqrt{2\pi}ZG \qquad \text{if } G > 0$$

$$A(\lambda) = 0 \qquad \text{if } G \le 0$$

in which $G = \mu + R - \lambda$, $R = 2\sigma\rho/(\rho^2 - 1)$, and $Z = \ln \rho/(2 \ln 2)^{1/2}$. In some cases the skewness increased unreasonably to fit imperfect data perfectly. We limited ρ to $1.0 \le \rho \le 1.5$, and this limit barely affected the fit.

Thus, if a spectrum is to be decomposed into N individual bands, $4N$ variables would have to be determined, and the fitted spectrum (as opposing to measured spectrum) is

$$A_{\text{base}}(\lambda) = \sum_{i=1}^{N} A(\lambda, \mu_i, \zeta_i, \sigma_i, \rho_i)$$

Since we know from other work how many bands exist within the measured spectrum for each monomer,[44-48] we know the value

(34) Levitt, M. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 640–644.
(35) Sarai, A.; Jazur, J.; Nussinov, R.; Jernigan, R. L. *Biochemistry* **1988**, *27*, 8498–8502.
(36) Ansevin, A. T.; Wang, A. H. *Nucleic Acids Res.* **1990**, *18*, 6119–6126.
(37) Weiner, P. K.; Kollman, P. A. *J. Comput. Chem.* **1981**, *2*, 287–303.
(38) Edmondson, S. P. *Biopolymers* **1987**, *26*, 1941–1956.
(39) Brown, K. M.; Dennis, J. E., Jr. *Numer. Math.* **1972**, *18*, 289–297.
(40) Brent, R. P. In *Algorithms for Minimization Without Derivatives*; Thomas J. Watson Research Center: Yorktown Heights, NY, 1973.
(41) Alper, J. S.; Gelb, R. I. *J. Phys. Chem.* **1990**, *94*, 4747–4751.
(42) Siano, D. B.; Metzler, D. E. *J. Chem. Phys.* **1969**, *51*, 1856–1861.
(43) Siano, D. B. *J. Chem. Educ.* **1972**, *49*, 755–757.
(44) Clark, L. B. *J. Am. Chem. Soc.* **1977**, *99*, 3934–3938.
(45) Clark, L. B. *J. Phys. Chem.* **1989**, *93*, 5345–5347.
(46) Clark, L. B. *J. Phys. Chem.* **1990**, *94*, 2873–2879.
(47) Zaloudek, F.; Novros, J. S.; Clark, L. B. *J. Am. Chem. Soc.* **1985**, *107*, 7344–7351.
(48) Novros, J. S.; Clark, L. B. *J. Phys. Chem.* **1986**, *90*, 5666–5668.

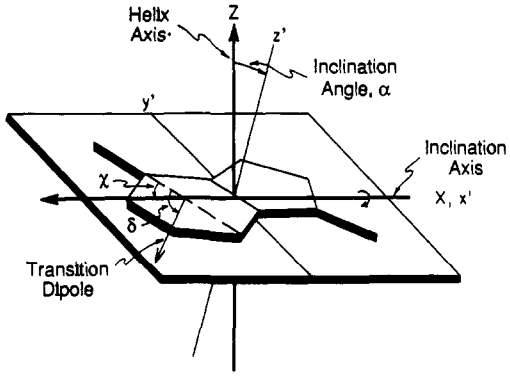

Figure 1. Diagram showing adenine inclining around the x axis by an inclination angle α. The angle between the reference $N_3 \rightarrow C_6$ and the inclination axis is χ, and the angle between $N_3 \rightarrow C_6$ and the transition dipole is δ.

of N for each base, which corresponds to the smallest number of bands necessary to give a satisfactory fit to the absorption spectrum.

To begin the decomposition, initial values for position μ, intensity ζ, and width σ are taken from previous work.[32,33,44-48] Skewness ρ is arbitrarily assigned the value 1.2. Fitting to the monomer spectra by the DFLM algorithm is quite straightforward and the results are stable.

Fitting Polymer Absorption and LD Spectra. The parameters determined by fitting the monomer absorption spectra are the initial guesses for simultaneously fitting the absorption and LD spectra for each type of polymer using the DFLM algorithm. The relation between measured isotropic absorption and LD for a transition dipole i of base j is given by[15,32,33]

$$LD_{ij}(\lambda) = A_{ij}(\lambda)3S[3 \sin^2 \alpha_j \sin^2 (\chi_j - \delta_{ij}) - 1]/2 \quad (1)$$

in which δ_{ij} is the angle between transition dipole i and the vector $N \rightarrow C_6$ if base j is a purine or $N_1 \rightarrow C_4$ if j is a pyrimidine, α_j is the inclination of base j from perpendicular to the helix axis (the result of both twist and tilt), χ_j is the angle between the in-plane axis (perpendicular to the helix axis) around which the base inclines and the vector to which δ_{ij} references, and S is the factor that makes up for imperfect orientation in the flow. The signs of χ, δ, and α follow the right-handed Cartesian coordinate system, and the angles are illustrated in Figure 1.

Since our polymers in these studies contain more than one base, the absorption measured and LD spectra are as follows:

$$A_{poly}(\lambda) = \sum_{j=1}^{M}\sum_{i=1}^{N_j} A_{ij}(\lambda)$$

$$LD_{poly}(\lambda) = \sum_{j=1}^{M}\sum_{i=1}^{N_j} LD_{ij}(\lambda)$$

in which N_j is the number of transitions for the jth base and M is the number of bases. We are analyzing the wavelength dependence of the data, so that imperfect orientation, including the effects of tertiary superstructures, does not affect our analysis.[12,14,15,27]

The objective is to determine the parameters for all bands and the α and χ angles for the bases in the polymer through the DFLM algorithm as described above, simultaneously fitting the absorption and LD spectra. We reiterate that, due to the large number of variables and the different scales of measurement errors for the absorption and LD spectra, our chosen fit is at the unique point along the minimization path not too far from the global minimum of residuals, at which most variables have the smallest variance. We monitor all variances after each iteration and choose the bottom of the multidimensional valley of variances as our end point.

The transition dipole directions, δ_{ij}, associated with transition i of base j, must be known to fit LD spectra, and these are taken from Clark and co-workers.[44-48] Initial values for parameters for the absorption and LD bands are those from our fitting of monomers (Table I). Initial α and χ angles for the synthetic polymers are from earlier work[24,32,33] and for DNA are from our results for the synthetic polymers.

Table I. Decomposition of Monomer Absorption Spectra

monomer	μ (nm)	$\zeta \times 10^{-3}$	σ (nm)	ρ	δ (deg)
dAMP	266.4	162.7	11.2	1.20	83[a]
	255.0	319.1	13.9	1.33	25[a]
	206.6	467.0	10.5	1.21	-45[a]
	195.3	78.7	6.1	1.38	15[a]
	184.9	282.7	7.6	1.29	72[a]
	173.6	60.2	4.5	1.00	-45[a]
TMP	265.1	363.0	18.0	1.25	-9[b]
	204.7	409.5	19.7	1.50	-53[b]
	176.6	190.7	5.8	1.42	-26[b]
dGMP	274.5	288.7	16.7	1.50	-4[c]
	248.5	309.5	13.9	1.10	-75[c]
	198.8	471.2	11.6	1.03	-71[c]
	183.2	449.4	11.6	1.50	41[c]
dCMP	269.0	301.2	15.3	1.12	6[d]
	228.1	319.2	19.8	1.31	-35[d]
	211.6	86.8	7.1	1.00	76[d]
	196.5	403.1	9.9	1.43	86[d]
	170.1	94.0	12.4	1.03	0[d]

[a] References 45 and 46. [b] Reference 48. [c] Reference 44. [d] Reference 47.

Uncertainties in Transition Dipole Directions. The measured directions of the transition dipoles are assumed to be correct and unchanged for all polymers and DNAs studied. However, as mentioned in the reports of dipole direction measurements, there are uncertainties in these directions. To determine how the uncertainties affect our results, we repeated each fitting 100 times with the transition dipole directions randomly varied within ±10°. The average value from the 100 runs for each variable (parameters for each band and α,χ angles of each base) is our reported value, and the standard deviation is for the 100 runs.

Validation of α and χ Angles. A given base pair will have quantities that vary with α and χ angles, such as hydrogen-bond distance and angle, distance between purine C_8 and pyrimidine C_6, distance between the two C_1' atoms and propeller twist (the dihedral angle between base planes). By constructing base pairs from our α and χ angles, calculating these base-pair parameters, and comparing with published parameters, we can determine whether α and χ angles derived this way are reasonable. Another reason for this validation has to do with the sign of α. Because positive and negative α angles of the same magnitude give the same LD spectrum, we investigated the four possible base pairings with the signs of the angles as +/+, +/−, −/+, and −/− for each base pair.

With atomic coordinates for the four bases taken from Arnott,[49] construction of a base pair begins by placing the bases in a plane (assigned to be the xy plane) perpendicular to the direction of light polarization (assigned to be the z axis). Each base plane is rotated about the x axis for α deg. Because LD contains only information about a base instead of a base pair, we are free to move the two bases in space and rotate around the z axis as long as we keep the angle between each base plane and the xy plane constant. With minimal effort the two or three hydrogen-bond distances can be adjusted to an acceptable value, and then other base-pair parameters are calculated.

Results and Discussion

Decomposition of Monomer Absorption Spectra. Absorption spectra of dAMP, TMP, dGMP, and dCMP are decomposed into 6, 3, 4, and 5 bands, respectively, as shown in Figures 2–5. The position, intensity, width, and skewness of each band are listed in Table I. Obviously, the parameters for the 173.6-nm band of dAMP and the 170.1-nm band of dCMP are neither well determined nor particularly relevant. However, the red end of a shorter wavelength band is necessary in this region to realistically

(49) Arnott, S. *Prog. Biophys. Mol. Biol.* **1970**, *21*, 265–319.

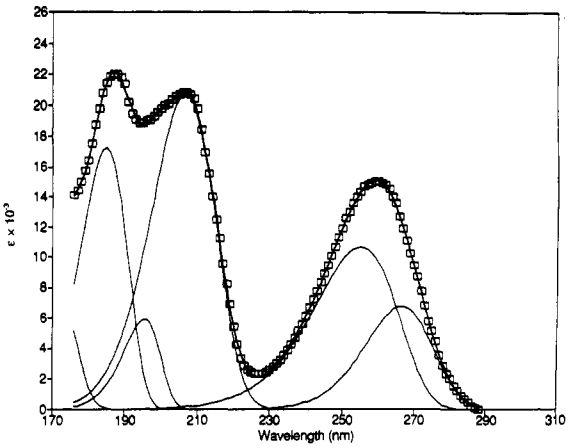

Figure 2. Decomposition of dAMP absorption spectrum: □ is measured, — is fitted.

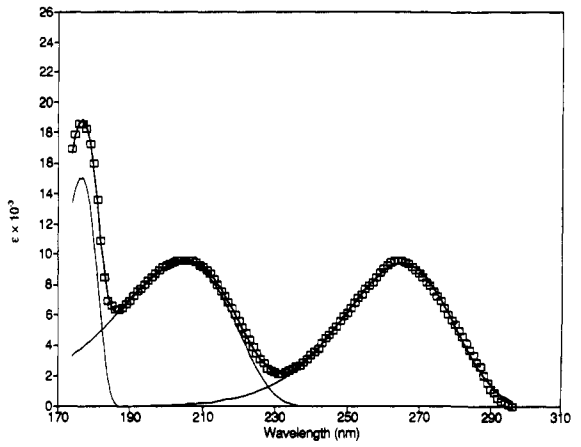

Figure 3. Decomposition of TMP absorption spectrum: □ is measured, — is fitted.

fit the data. Also listed are the corresponding transition dipole directions, which are vital for successful decomposition of an LD spectrum.

During preliminary studies only four bands were used to decompose the dAMP spectrum, as with work previously done in our laboratory.[32] The result is fine except for the region between 190 and 210 nm, and according to Clark[46] a minor band is also in this region, which we now include in our fit (Figure 2). Fitting of the TMP spectrum is relatively easy because its three components are well separated (Figure 3), but these bands are not Gaussian and show that the log-normal function with its skewness parameter is more suitable to approximate electronic absorption bands. The major components of the dGMP and dCMP spectra can be distinguished as peaks or shoulders (Figures 4 and 5).

Decomposition of the Absorption and LD Spectra for the Synthetic Polymers. Using the decomposition of the monomer absorption spectra as the initial guess, a polymer absorption spectrum could be decomposed, and the resulting μ, ζ, σ, and ρ parameters were used to decompose its LD spectrum with the α's and χ's as the variables. The problem with this two-step procedure is that the fits to absorption and LD are correlated. We also tried fitting both spectra at the same time with all of the variables, and the error in fitting the LD spectrum scaled to reflect the fact that the intensity of the LD spectrum is much smaller than that of the absorption spectrum. As the total sum of squares error for the fitting is minimized, the sum of squares error for the LD spectrum is nearly synchronous with that of absorption spectrum (Figure 6), and we can let the fitting proceed until the global minimum for the sum of squares error is reached. The error in the fit will be less than the error in the measurements, and this unrealistic fitting results in some unrealistic band parameters. For example, a band position may move to 400 or 100 nm, a band intensity may become zero, or a band width may decrease to 0.1

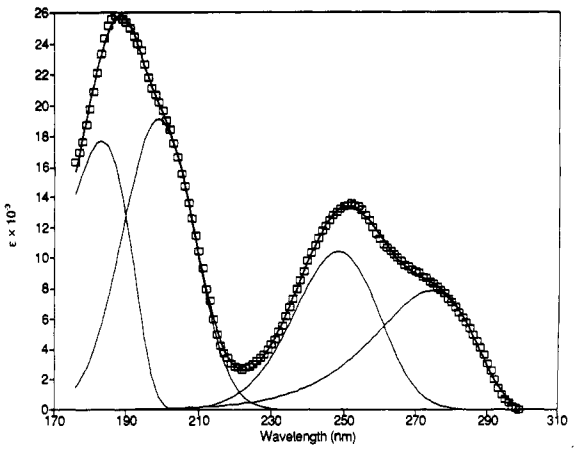

Figure 4. Decomposition of dGMP absorption spectrum: □ is measured, — is fitted.

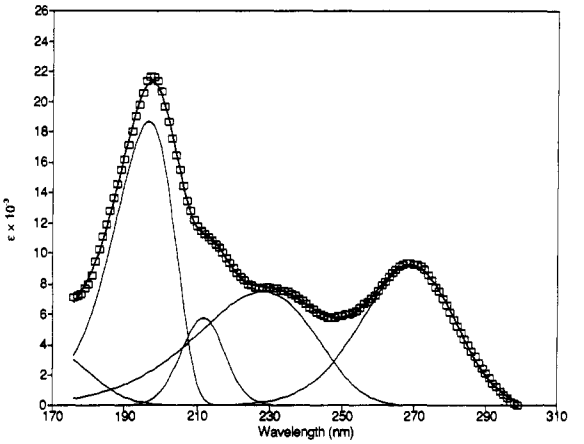

Figure 5. Decomposition of dCMP absorption spectrum: □ is measured, — is fitted.

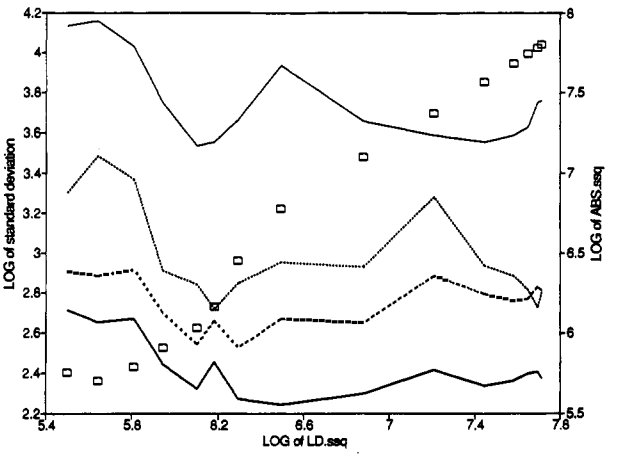

Figure 6. Sum of squares error for absorption (ABS.ssq) versus LD (LD.ssq): correlated in this algorithm, □. LD.ssq decreases as ABS.ssq decreases, here for A-form DNA. To avoid overfitting the data we looked for a stable solution with small variances in the variables. The log of the standard deviation for the inclination of A (– –), T (—), G (—), and C (···) is minimized for all four bases, as for the other variables, around LD.ssq of 6.0–6.4.

nm. There is error in the measurements, so we must stop the fitting before it is overdone. Thus, we choose to stop when the variables become stable, as described in the Methods section. The advantage of this procedure is 2-fold: (1) the point along the minimization path is easily identified; and (2) the weight assigned to scale fitting errors of the LD spectrum has little effect on the fitted results, so that we can weigh both absorption and LD spectra equally.

Table II. Decomposition of Poly[d(A)–d(T)] Absorption and LD Spectra

base	μ (nm)	$\zeta \times 10^{-3}$	σ (nm)	ρ	α (deg)	χ (deg)
d(A)	276.9 ± 1.0	26.1 ± 4.6	10.1 ± 0.5	1.20 ± 0.11	23.2 ± 0.8	−28.4 ± 3.7
	255.2 ± 0.8	94.8 ± 2.2	12.4 ± 0.3	1.23 ± 0.09		
	207.9 ± 0.2	124.7 ± 5.5	11.7 ± 0.3	1.01 ± 0.00		
	196.9 ± 0.2	52.8 ± 2.8	6.7 ± 0.1	1.02 ± 0.01		
	185.4 ± 0.2	148.2 ± 4.9	7.9 ± 0.2	1.24 ± 0.02		
	172.5 ± 0.2	40.5 ± 6.2	4.1 ± 0.3	1.10 ± 0.06		
d(T)	268.0 ± 2.3	113.6 ± 2.5	17.9 ± 1.4	1.32 ± 0.14	42.1 ± 2.5	21.1 ± 3.2
	203.7 ± 1.3	137.4 ± 9.9	21.8 ± 1.3	1.19 ± 0.09		
	177.5 ± 0.2	142.3 ± 5.0	5.8 ± 0.1	1.08 ± 0.06		

Table III. Decomposition of Poly[d(AT)–d(AT)] Absorption and LD Spectra

base	μ (nm)	$\zeta \times 10^{-3}$	σ (nm)	ρ	α (deg)	χ (deg)
d(A)	271.3 ± 0.4	49.6 ± 0.5	13.6 ± 0.1	1.19 ± 0.03	18.6 ± 0.6	−16.1 ± 3.4
	256.2 ± 0.2	89.4 ± 0.6	14.1 ± 0.2	1.26 ± 0.05		
	206.5 ± 0.1	161.6 ± 1.4	13.0 ± 0.1	1.05 ± 0.02		
	195.7 ± 0.0	39.6 ± 0.6	6.2 ± 0.1	1.06 ± 0.01		
	185.5 ± 0.0	112.7 ± 0.3	7.4 ± 0.0	1.07 ± 0.01		
	174.3 ± 0.2	28.3 ± 2.0	3.8 ± 0.1	1.07 ± 0.06		
d(T)	268.5 ± 0.1	114.5 ± 1.0	16.9 ± 0.1	1.20 ± 0.01	34.8 ± 2.0	18.7 ± 3.2
	203.7 ± 0.7	159.6 ± 3.8	23.9 ± 0.4	1.41 ± 0.02		
	177.2 ± 0.0	78.7 ± 0.9	5.5 ± 0.1	1.13 ± 0.00		

Table IV. Decomposition of B-Form Poly[d(G)–d(C)] Absorption and LD Spectra

base	μ (nm)	$\zeta \times 10^{-3}$	σ (nm)	ρ	α (deg)	χ (deg)
d(G)	279.7 ± 0.1	86.2 ± 0.8	16.5 ± 0.2	1.50 ± 0.00	20.1 ± 0.6	116.8 ± 3.5
	248.3 ± 0.1	135.7 ± 0.6	13.5 ± 0.1	1.00 ± 0.00		
	196.4 ± 0.2	145.7 ± 1.0	12.0 ± 0.2	1.01 ± 0.01		
	179.8 ± 0.0	234.4 ± 1.5	10.1 ± 0.1	1.39 ± 0.01		
d(C)	263.6 ± 0.1	98.4 ± 0.4	15.1 ± 0.2	1.15 ± 0.01	33.8 ± 1.0	189.8 ± 3.8
	221.8 ± 0.3	98.4 ± 0.7	17.8 ± 0.2	1.25 ± 0.02		
	211.0 ± 0.2	40.7 ± 1.7	9.2 ± 0.2	1.02 ± 0.01		
	193.4 ± 0.1	124.7 ± 1.4	10.1 ± 0.2	1.01 ± 0.00		
	182.6 ± 0.2	105.5 ± 1.0	11.2 ± 0.1	1.00 ± 0.00		

Table V. Decomposition of B-Form Poly[d(GC)–d(GC)] Absorption and LD Spectra

base	μ (nm)	$\zeta \times 10^{-3}$	σ (nm)	ρ	α (deg)	χ (deg)
d(G)	279.7 ± 0.1	81.4 ± 0.7	16.8 ± 0.1	1.46 ± 0.01	21.4 ± 0.5	130.7 ± 2.8
	248.7 ± 0.1	130.1 ± 0.4	14.1 ± 0.1	1.02 ± 0.01		
	196.3 ± 0.2	140.7 ± 1.1	12.4 ± 0.1	1.01 ± 0.01		
	180.0 ± 0.0	232.7 ± 0.7	10.1 ± 0.0	1.35 ± 0.01		
d(C)	263.7 ± 0.1	96.1 ± 0.4	14.7 ± 0.1	1.10 ± 0.01	34.0 ± 0.7	184.0 ± 3.2
	221.4 ± 0.2	92.9 ± 1.3	17.6 ± 0.3	1.39 ± 0.01		
	209.9 ± 0.5	35.1 ± 1.6	9.9 ± 0.3	1.04 ± 0.03		
	193.1 ± 0.1	124.1 ± 1.5	10.2 ± 0.1	1.10 ± 0.02		
	182.7 ± 0.1	103.1 ± 1.2	10.7 ± 0.1	1.09 ± 0.02		

Figure 6 shows the standard deviation in α of the four bases for A-form DNA. We see that a stable solution with low standard deviation is achieved roughly when the log of the sum of squares error for the absorption (ABS.ssq) and LD (LD.ssq) is 6.0–6.4. The exact choice does not affect the results significantly, and the method is stable. Further fitting leads to a larger standard deviation and instability, as Figure 6 shows. Note that ABS.ssq and LD.ssq are not perfectly correlated.

The results of decomposing the absorption and LD spectra for B-form poly[d(A)–d(T)] are shown in Figure 7 and listed in Table II. For adenine, the first band of d(A) shifts toward longer wavelengths by +10.5 nm from that of dAMP, while the positions of all other bands remain about the same. The second band of d(T) is the only one in our studies having a positive LD (remember that the resultant LD spectrum for a nucleic acid is negative everywhere). Numerically, the sign of an LD band depends on α and $\chi - \delta$, as can be seen from eq 1. The larger both angles are, the more likely that a transition will have a positive LD band. Since α and χ for the d(T) base from the best fit are 42.1° and 21.1° (Table II) and δ for the second band is −53° (Table I), the expression within the brackets in eq 1 is positive.

Table III lists the results for B-form poly[d(AT)–d(AT)] (decomposition not shown). Only the first band of d(A) and the second band of d(T) are significantly different from their counterparts in dAMP and TMP, respectively. If compared with the results for poly[d(A)–d(T)] (Table II), we find that the third band of d(T) and all but the first band of d(A) are about the same for both polymers. The inclinations of d(A) and d(T) are somewhat smaller than those of poly[d(A)–d(T)].

Decomposition of B-form poly[d(G)–d(C)] (Figure 8) and poly[d(GC)–d(GC)] (not shown) spectra gives very similar results in band parameters and α,χ angles (Tables IV and V), but some band parameters deviate from those of dGMP and dCMP. The first band of d(G) shifts −5.2 nm with respect to that of dGMP, and all of the first four bands of d(C) shift −5.3, −6.7, −1.7, and −3.4 nm, respectively, from those of dCMP.

Table VI lists the results for decomposition (not shown) of the Z-form poly[d(GC)–d(GC)] spectra. Each band for both d(G) and d(C) resembles the corresponding one in B-form poly[d(G)–d(C)] and poly[d(GC)–d(GC)] (Tables IV and V), except for a 3.9-nm difference in the position of the first band for d(G).

Two remarks can be made regarding the decompositions of the three d(G)–d(C) polymers. First, the fourth band of dGMP and the fifth band of dCMP almost exchange their positions after fittings to polymers, from 183 and 170 nm for the monomers to 177 and 183 nm for the polymers. Numerically, it means either that the path between the two points of the multidimensional space is clear-cut (which may also mean the exact answer is actually

Table VI. Decomposition of Z-Form Poly[d(GC)-d(GC)] Absorption and LD Spectra

base	μ (nm)	$\zeta \times 10^{-3}$	σ (nm)	ρ	α (deg)	χ (deg)
d(G)	283.6 ± 0.1	96.0 ± 1.3	16.2 ± 0.2	1.50 ± 0.00	27.1 ± 1.1	137.6 ± 3.6
	249.3 ± 0.2	126.8 ± 1.9	14.7 ± 0.2	1.01 ± 0.01		
	197.9 ± 0.6	143.3 ± 3.0	12.6 ± 0.2	1.04 ± 0.04		
	177.3 ± 0.5	222.1 ± 1.8	11.4 ± 0.2	1.05 ± 0.02		
d(C)	265.4 ± 0.3	97.1 ± 1.3	15.4 ± 0.4	1.04 ± 0.02	32.1 ± 1.7	201.5 ± 2.8
	217.9 ± 0.4	93.6 ± 2.6	15.9 ± 0.3	1.19 ± 0.06		
	206.4 ± 0.4	32.7 ± 1.8	6.4 ± 0.3	1.03 ± 0.02		
	193.4 ± 0.3	115.8 ± 2.1	9.6 ± 0.4	1.39 ± 0.07		
	184.5 ± 0.2	94.8 ± 3.4	7.2 ± 0.2	1.02 ± 0.02		

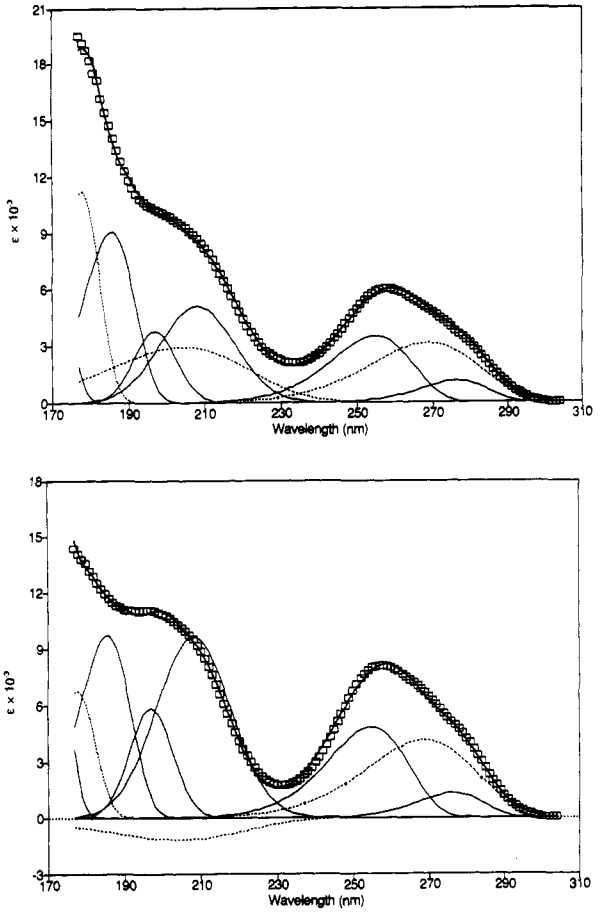

Figure 7. Decomposition of poly[d(A)-d(T)] absorption spectrum (a) and negative of the normalized LD spectrum (b): □ is measured, — is fitted.

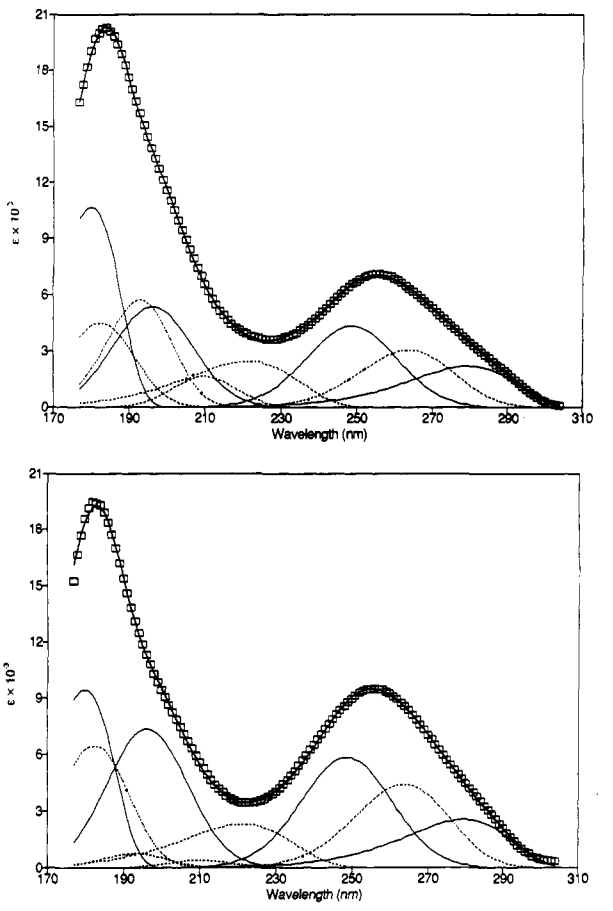

Figure 8. Decomposition of poly[d(GC)-d(GC)] absorption spectrum (a) and negative of the normalized LD spectrum (b): □ is measured, — is fitted.

there) or that the DFLM algorithm is smart enough to swell the two bands, exchange their positions slowly, and then shrink the two bands (this is what actually happened) to reach the point of minimum variances. Second, we get almost the same results for d(C) for the three polymers, including all of its bands and α,χ angles. This may indicate that cytosine is less sensitive to its environment or that it faces the same surroundings in the three polymers.

Decomposition of DNA Absorption and LD Spectra. Natural DNA is typically studied in three different forms in solution. In aqueous solution with moderate salt (here 0.01 M Na^+ as phosphate buffer, pH 7) DNA exhibits a well-known conservative circular dichroism (CD) spectrum with a maximum at 275 nm and a minimum at 248 nm.[50] This B-form DNA has 10.4 bp/turn,[51,52] and we denote it as 10.4B-DNA. At high concentrations of salt (here 5.5 M NH_4F), in 95% methanol, or when wrapped around histone cores, the 275-nm band of B-form DNA collapses, and this form has 10.2 bp/turn.[53] We denote this form as 10.2B-DNA. In 80% ethanol,[5,6] or here 80% 2,2,2-trifluoroethanol,[24] DNA has the nonconservative CD typical of the A form. The LD has been measured for all three forms,[31] and we analyze these LD spectra here for the first time.

Decomposition of DNA absorption and LD spectra presents another set of problems. First, the computer time required for each iteration is more than 4 times longer than that in fitting two-base polymers. Second, the step size between two iterations must be small enough so that the DFLM algorithm can find the path leading to the point of minimum variances and stay there through several iterations. Third, as the step size gets smaller, round-off errors become more significant in computing the Jacobian matrix and matrix inversion, resulting in meaningless variances.

We overcome these problems to obtain the results listed in Tables VII–IX for decomposition of absorption and LD spectra (not shown). Differences among the three DNAs for each band are generally small. Inclinations for the B forms are about 15° for the purines and 26° for the pyrimidines. As expected d(A),

(50) Brahms, J.; Mommaerts, W. F. H. M. *J. Mol. Biol.* **1964**, *10*, 73–88.
(51) Wang, J. C. *Cold Spring Harbor Symp. Quant. Biol.* **1978**, *43*, 29–34.
(52) Wang, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 200–203.
(53) Baase, W. A.; Johnson, W. C., Jr. *Nucleic Acids Res.* **1979**, *6*, 797–814.

Table VII. Decomposition of 10.4B-DNA Absorption and LD Spectra

base	μ (nm)	$\zeta \times 10^{-3}$	σ (nm)	ρ	α (deg)	χ (deg)
d(A)	271.4 ± 0.0	24.6 ± 0.1	14.9 ± 0.1	1.21 ± 0.00	16.1 ± 0.5	46.5 ± 4.7
	255.2 ± 0.1	43.4 ± 0.2	13.3 ± 0.0	1.27 ± 0.01		
	206.4 ± 0.0	73.2 ± 0.1	12.8 ± 0.0	1.05 ± 0.00		
	195.5 ± 0.0	17.5 ± 0.2	6.4 ± 0.0	1.01 ± 0.00		
	185.5 ± 0.0	50.8 ± 0.1	7.4 ± 0.0	1.09 ± 0.01		
	174.4 ± 0.0	13.9 ± 0.1	3.5 ± 0.0	1.13 ± 0.00		
d(T)	268.5 ± 0.0	54.7 ± 0.2	17.4 ± 0.1	1.19 ± 0.01	25.0 ± 0.9	1.8 ± 3.3
	204.2 ± 0.1	70.6 ± 0.1	23.9 ± 0.1	1.41 ± 0.00		
	177.1 ± 0.0	35.1 ± 0.1	5.6 ± 0.0	1.15 ± 0.00		
d(G)	279.5 ± 0.1	40.5 ± 0.1	15.9 ± 0.0	1.50 ± 0.00	18.0 ± 0.6	114.8 ± 8.6
	248.9 ± 0.0	61.7 ± 0.2	13.2 ± 0.0	1.04 ± 0.01		
	196.3 ± 0.1	63.9 ± 0.2	12.4 ± 0.1	1.01 ± 0.01		
	180.0 ± 0.0	105.2 ± 0.1	10.2 ± 0.0	1.35 ± 0.01		
d(C)	263.2 ± 0.0	45.8 ± 0.1	15.2 ± 0.0	1.05 ± 0.00	25.1 ± 0.8	215.8 ± 3.0
	222.0 ± 0.1	42.2 ± 0.3	18.1 ± 0.1	1.20 ± 0.04		
	209.7 ± 0.1	15.4 ± 0.1	9.8 ± 0.1	1.01 ± 0.01		
	193.2 ± 0.0	55.6 ± 0.2	10.2 ± 0.0	1.08 ± 0.00		
	182.5 ± 0.0	46.6 ± 0.1	10.9 ± 0.0	1.11 ± 0.00		

Table VIII. Decomposition of 10.2B-DNA Absorption and LD Spectra

base	μ (nm)	$\zeta \times 10^{-3}$	σ (nm)	ρ	α (deg)	χ (deg)
d(A)	271.7 ± 0.1	23.9 ± 0.2	14.4 ± 0.1	1.23 ± 0.01	14.9 ± 0.6	96.6 ± 3.7
	255.1 ± 0.0	42.6 ± 0.1	13.2 ± 0.1	1.18 ± 0.01		
	206.7 ± 0.1	68.8 ± 0.2	13.1 ± 0.1	1.01 ± 0.01		
	194.9 ± 0.0	16.4 ± 0.1	6.7 ± 0.1	1.01 ± 0.00		
	185.4 ± 0.1	45.8 ± 0.2	8.1 ± 0.1	1.01 ± 0.01		
	174.1 ± 0.1	11.7 ± 0.5	3.6 ± 0.1	1.15 ± 0.04		
d(T)	268.3 ± 0.2	53.6 ± 0.4	17.8 ± 0.2	1.09 ± 0.01	28.1 ± 1.3	31.9 ± 3.0
	204.7 ± 0.4	66.1 ± 0.5	23.3 ± 0.2	1.41 ± 0.03		
	177.1 ± 0.0	34.5 ± 0.6	5.2 ± 0.1	1.12 ± 0.01		
d(G)	279.5 ± 0.1	38.9 ± 0.2	16.0 ± 0.1	1.50 ± 0.00	13.9 ± 1.7	142.5 ± 4.2
	248.8 ± 0.1	58.6 ± 0.2	13.4 ± 0.1	1.08 ± 0.01		
	196.5 ± 0.1	59.5 ± 0.1	12.2 ± 0.1	1.01 ± 0.01		
	179.7 ± 0.0	96.1 ± 0.2	10.5 ± 0.0	1.27 ± 0.00		
d(C)	263.5 ± 0.1	43.7 ± 0.2	15.4 ± 0.1	1.07 ± 0.01	27.7 ± 0.7	201.2 ± 2.5
	221.5 ± 0.1	40.4 ± 0.1	17.3 ± 0.1	1.18 ± 0.02		
	210.2 ± 0.2	14.9 ± 0.1	11.9 ± 0.3	1.09 ± 0.03		
	193.9 ± 0.1	51.2 ± 0.3	10.6 ± 0.1	1.01 ± 0.01		
	182.2 ± 0.1	42.1 ± 0.3	11.5 ± 0.1	1.02 ± 0.01		

Table IX. Decomposition of A-Form DNA Absorption and LD Spectra

base	μ (nm)	$\zeta \times 10^{-3}$	σ (nm)	ρ	α (deg)	χ (deg)
d(A)	270.3 ± 0.2	25.8 ± 0.7	16.9 ± 0.7	1.13 ± 0.01	27.8 ± 1.0	7.0 ± 1.3
	256.1 ± 0.1	44.7 ± 0.4	14.1 ± 0.1	1.34 ± 0.01		
	206.9 ± 0.1	76.1 ± 0.2	12.5 ± 0.1	1.00 ± 0.00		
	196.0 ± 0.1	18.4 ± 0.4	6.3 ± 0.1	1.01 ± 0.01		
	185.4 ± 0.0	50.0 ± 0.3	7.4 ± 0.0	1.13 ± 0.01		
	174.0 ± 0.0	11.3 ± 0.2	3.5 ± 0.0	1.18 ± 0.01		
d(T)	268.6 ± 0.1	56.8 ± 0.6	17.0 ± 0.1	1.16 ± 0.01	34.7 ± 0.9	−5.4 ± 1.4
	206.6 ± 0.1	75.9 ± 0.3	23.2 ± 0.1	1.21 ± 0.01		
	176.9 ± 0.0	34.3 ± 0.3	5.5 ± 0.0	1.23 ± 0.01		
d(G)	280.0 ± 0.1	42.3 ± 0.4	16.3 ± 0.1	1.50 ± 0.00	14.3 ± 1.0	95.3 ± 6.7
	248.9 ± 0.1	61.7 ± 0.4	14.6 ± 0.1	1.06 ± 0.01		
	196.8 ± 0.1	65.2 ± 0.5	12.0 ± 0.1	1.04 ± 0.02		
	179.9 ± 0.1	103.7 ± 0.4	10.3 ± 0.0	1.36 ± 0.01		
d(C)	263.4 ± 0.1	47.7 ± 0.2	14.7 ± 0.0	1.06 ± 0.01	35.2 ± 0.5	216.1 ± 1.4
	219.9 ± 0.4	49.6 ± 0.6	14.7 ± 0.3	1.12 ± 0.05		
	209.9 ± 0.2	18.6 ± 0.2	8.4 ± 0.1	1.01 ± 0.01		
	193.5 ± 0.1	54.6 ± 0.5	11.2 ± 0.2	1.01 ± 0.01		
	182.4 ± 0.1	45.1 ± 0.2	11.3 ± 0.1	1.07 ± 0.01		

d(T), and d(C) have larger inclinations in the A form, but our results indicate that d(G) is unchanged from the B form.

Comparison to Previous Results. Angles α for the five synthetic polymers are similar to those calculated previously in our laboratory from the same data.[32,33] Differences in the inclination axis, χ, are not surprising as this parameter has been recognized as unstable. The new inclination angles, α, result in the same message: base pairs are inclined, even in the B form.

One factor that is responsible for any differences between this analysis and previous analyses is the different optimization algorithm. Although the advantage of the simplex method used previously is that one can tell local ssq (sum of squares error) minima from the global minimum "by running the program several times",[32] the scale of the problem actually turns the advantage into a disadvantage, because "several times" could be infinite to assure that the global minimum of ssq is found. Two other disadvantages in using the simplex algorithm are that there is no correlation term defined for any two variables and that the algorithm uses only ssq, and not individual squared errors. The DFLM algorithm used in this study has none of these drawbacks.

Furthermore, transition dipole directions are different, especially for the base adenine, which also has a different number of transitions. A skewness parameter is added to define the shape of an absorption band. Previous calculations aimed to fit one reduced

Table X. A/T Base-Pair Parameters

		hydrogen bond						
		$A.N_6$–$T.O_4$		$A.N_1$–$T.N_3$				
	sign of α for A/T pair	length (Å)	angle (deg)	length (Å)	angle (deg)	propeller twist (deg)	$A.C_1'$–$T.C_1'$ (Å)	$A.C_8$–$T.C_6$ (Å)
poly[d(A)–d(T)]	+/+	2.80	102	3.00	94	53	10.84	9.67
	+/–	2.80	112	3.00	123	40	10.18	8.99
	–/+	2.80	110	2.91	121	40	10.44	9.22
	–/–	2.80	105	3.00	95	53	10.73	9.47
poly[d(AT)–d(AT)]	+/+	2.81	110	3.00	103	40	10.82	9.75
	+/–	2.80	114	3.00	120	38	10.38	9.26
	–/+	2.84	111	2.90	121	38	10.43	9.30
	–/–	2.80	113	3.00	103	40	10.74	9.81
10.4B-DNA	+/+	2.92	123	3.00	121	10	10.82	9.88
	+/–	3.00	109	2.95	121	41	10.29	9.37
	–/+	3.00	108	2.91	119	41	10.44	9.46
	–/–	2.87	124	3.00	121	10	10.78	9.86
10.2B-DNA	+/+	2.94	122	3.00	122	13	10.84	9.86
	+/–	3.00	107	2.81	115	43	10.71	9.72
	–/+	3.00	106	2.92	110	43	10.91	9.76
	–/–	2.85	123	3.00	120	13	10.80	9.84
A-form DNA	+/+	3.00	114	2.80	120	25	10.57	9.74
	+/–	2.80	109	3.00	123	58	9.27	8.39
	–/+	2.80	107	2.97	123	57	9.66	8.61
	–/–	3.00	115	2.80	121	25	10.48	9.76

Table XI. G/C Base-Pair Parameters

		hydrogen bond								
		$G.O_6$–$C.N_4$		$G.N_1$–$C.N_3$		$G.N_2$–$C.O_2$				
	sign of α for G/C pair	length (Å)	angle (deg)	length (Å)	angle (deg)	length (Å)	angle (deg)	propeller twist (deg)	$G.C_1'$–$C.C_1'$ (Å)	$G.C_8$–$C.C_6$ (Å)
B-form poly[d(G)–d(C)]	+/+	3.00	112	2.80	113	3.37	109	49	10.55	9.70
	+/–	2.80	122	3.00	116	2.83	126	27	10.61	9.80
	–/+	2.80	118	3.00	114	3.13	124	25	10.74	9.81
	–/–	3.00	114	2.80	115	3.10	113	48	10.42	9.71
B-form poly[d(GC)–d(GC)]	+/+	3.00	111	2.80	112	3.30	108	45	10.44	9.70
	+/–	2.80	119	3.00	115	2.80	124	34	10.48	9.67
	–/+	2.80	113	2.81	115	2.80	124	34	10.28	9.53
	–/–	3.00	113	2.80	115	3.03	111	44	10.29	9.70
Z-form poly[d(GC)–d(GC)]	+/+	3.00	106	2.80	108	3.70	98	52	10.71	9.74
	+/–	2.80	120	2.98	113	2.80	123	31	10.53	9.73
	–/+	2.80	115	2.82	114	2.80	123	31	10.35	9.61
	–/–	3.00	107	2.80	110	3.54	102	52	10.63	9.75
10.4B-DNA	+/+	3.00	108	2.80	110	3.31	106	43	10.71	9.77
	+/–	2.80	119	2.93	113	2.80	123	11	10.67	9.89
	–/+	2.80	118	2.92	112	2.91	121	10	10.71	9.89
	–/–	3.00	108	2.80	111	3.17	110	43	10.66	9.77
10.2B-DNA	+/+	3.00	112	2.80	113	3.07	110	35	10.47	9.80
	+/–	2.80	121	2.95	114	2.80	124	26	10.61	9.82
	–/+	2.80	117	2.82	114	2.80	123	26	10.46	9.72
	–/–	2.91	114	2.80	114	3.00	113	36	10.48	9.79
A-form DNA	+/+	3.00	112	2.80	115	3.48	109	49	10.84	9.76
	+/–	2.80	123	2.84	119	2.80	125	21	10.58	9.86
	–/+	2.80	122	2.83	118	2.91	122	21	10.62	9.86
	–/–	3.00	111	2.80	115	3.32	113	49	10.80	9.76

LD spectrum, while this work fits absorption and LD spectra simultaneously. Previously, the position of a band for a given base was fixed for all polymers containing that base, but the position is variable now. Because every change we made departing from the previous study is an improvement, the current results should be more reliable and stable.

Repeated Fittings with Randomized Transition Dipole Directions. Two types of error can affect the results of spectral decomposition: the error in measuring absorption and LD spectra and the error in determining transition dipole directions. The effects of the first type of error were minimized by using stability of the band parameters as the criterion for fitting the data. The effects of the second type of error can be studied through the Monte Carlo method. The direction of a transition dipole in a given base may not necessarily be the same for the monomer and a polymer containing the base. Thus we repeated each of the fittings 100 times with each transition dipole direction randomly perturbed by a value sampled from an uniform distribution in the range ±10°. Average and standard deviations are calculated from the 100 independently fitted results for each variable, and these are the results presented in Tables II–IX. The α and χ angles show no dependence on either ABS.ssq or LD.ssq, indicating that they are very stable around our chosen solution and are fairly insensitive to the ±10° variation in the transition dipole directions used to obtain these α and χ values.

In Tables II–IX relatively large standard deviations for band parameters often occur at bands sitting near the ends of a spectrum and for the second band of d(T). Standard deviations of χ angles in each table are always greater than those of α angles, especially for the three DNAs. The difference in stability between fitted α and χ angles was observed in earlier studies, but its physical meaning is not clear.

Building a Base-Pair Model. Table X lists parameters for the four base pairs built from α and χ angles determined for d(A) and d(T) in each of the two synthetic polymers and three DNAs. When building a base pair, positions and orientations of the two paired bases are adjusted so that the two hydrogen-bond lengths are between 2.80 and 3.00 Å and hydrogen-bond angles

Table XII. Cross-Pair Hydrogen Bond of Poly[d(A)-d(T)]

	A1.N_6-T2.O_4			A1.N_6-T2.O_4	
sign of α	length (Å)	angle (deg)	sign of α	length (Å)	angle (deg)
+/+	5.15	90	-/+	5.42	105
+/-	1.63	101	-/-	2.87	118

($A.C_6$-$A.N_6$-$T.O_4$ and $T.C_4$-$T.N_3$-$A.N_1$) are close to 120°. For each resulting base pair the propeller twist, the distance between the two C_1' atoms, and the distance between $A.C_8$ and $T.C_6$ are determined. Because the imposed restrictions are not very tight in this procedure, for each base pair we can actually derive a number of possible conformations, each with slightly different values of the parameters. Thus the results presented in Table X are not unique, nor the best, but simply possible.

One interesting feature regarding the uncertainty in the sign of the α angles is that only poly[d(A)-d(T)] has a propeller twist in +/-, or -/+ smaller than that in +/+ or -/-. It is independent of how the base pair is built because of the large α angle for d(T). Another feature is that all four possible propeller twist angles for 10.4B-DNA are about the same as those for 10.2B-DNA, although the χ angles of d(A) and d(T) for 10.4B-DNA are significantly different from those for 10.2B-DNA. Equally striking is the fact that all parameters for the +/+ and -/- variations of 10.4B-DNA are virtually the same as their counterparts in 10.2B-DNA. From the two hydrogen-bond angles, the +/+ and -/- pairs in both B-form DNAs are considered more acceptable than the +/- and -/+ pairs. The +/+ and -/- pairs of A-form DNA are also our choices, because their $A.C_1'$-$T.C_1'$ and $A.C_8$-$T.C_6$ distances are more realistic than those of the +/- and -/+ pairs. In general, conformations with a smaller propeller twist for the AT pair have more favorable base-pair parameters, and these are found in the +/+ and -/- pairs for all but poly[d(A)-d(T)].

The GC base pairs were built for the three d(G)-d(C) polymers and three DNAs, and the parameters are listed in Table XI. Among the four sign possibilities for each base pair, there are always two (+/+ and -/-) that have hydrogen-bond lengths beyond the 2.8-3.0 Å range, primarily due to their large propeller twists. Since the amino group of d(G) rotates within the closed base pair,[54] we allowed the hydrogen-bond length of $G.N_2$-$C.O_2$ to be a larger value and adjusted the other two within the range of 2.8-3.0 Å. In all cases the smaller propeller twist is also accompanied by better hydrogen-bond angles ($C.C_4$-$C.N_4$-$G.O_6$, $G.C_6$-$G.N_1$-$C.N_3$, and $G.C_2$-$G.N_2$-$C.O_2$). The $A.C_1'$-$T.C_1'$ and $A.C_8$-$T.C_6$ distances appear to be irregular for all four pairs and, thus, are of no help in determining which one conformation is reasonable.

One Step Further for Poly[d(A)-d(T)]. Even though base pairs can be built from the calculated α and χ angles for poly[d(A)-d(T)], the α angle of 42.1° for d(T) is rather large. Supporting evidence comes from the structure of poly[d(A)-d(T)] in the B form, which has a large propeller twist so that an extra hydrogen bond forms between $A.N_6$ of one base pair and $T.O_4$ of the next pair.[55] To verify the existence of this cross-pair hydrogen bond in a B-form structure, we need to define some parameters. In our coordinate system a standard B-form helix of DNA will have its helical axis at $x = 0.86$ and $y = 2.40$ Å.[56] The rise, dz, is 3.38 Å, and the rotational angle along the helical axis between two base pairs is +36°. Adding the twist and tilt from this work and then generating a second base pair through rotation defines a new helix axis displaced $dx = +0.0$ and $dy = +0.3$ Å.

Now, for each of the four possible AT pairs, namely, +/+, +/-, -/+, and -/-, we put a second pair (A2-T2) on top of the first one (A1-T1) according to the above parameters and calculate the length (A1.N_6-T2.O_4) and angle (A1.C_6-A1.N_6-T2.O_4) of this special hydrogen bond. The results listed in Table XII show that there indeed exists a hydrogen bond of length 2.87 Å and angle 118° between the A1.C_6 and T2.O_4 atoms if the paired sign of the α angles is -/-, and the bond length is actually the shortest distance among any two atoms between bases of A1-A2, A1-T2, A2-T1, and T1-T2. For +/+, +/-, and -/+ pairs, the hydrogen bond can barely form, and some atomic contact distances are too small to be acceptable. However, as has been stressed earlier, we have more than enough degrees of freedom to determine a possible structure, and the results presented here should not be taken as unique. In the case of finding this cross-pair hydrogen bond, we tried only the smallest and most reasonable dx and dy that can give results satisfying our conditions, and the -/- pair appeared to be the first one of choice.

Thus, the large α angle for d(T), as well as the large propeller twist angle between d(A) and d(T), is possible, and the overall picture may be considered as the actual conformation of poly[d(A)-d(T)] in solution.

Acknowledgment. This research was supported by PHS Grant No. GM43133 from the Institute of General Medical Sciences.

(54) Williams, L. D.; Williams, N. G.; Shaw, B. R. *J. Am. Chem. Soc.* **1990**, *112*, 829-833.

(55) Bhattacharyya, D.; Bansal, M. *J. Biomol. Struct. Dyn.* **1990**, *8*, 539-572.

(56) Takusagawa, F. *J. Biomol. Struct. Dyn.* **1990**, *7*, 795-809.