INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

ANALYSIS OF CW-EPR SPECTRA

AND

THE INTERNAL DYNAMICS OF DNA

by

Annabelle Wey Reese

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

1996

Approved by Bruce H Robinson
Chairperson of Supervisory Committee

Program Authorized
to Offer Degree Chemistry

Date December 13, 1996

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature Annabell Wylie

Date December 13, 1996

University of Washington

Abstract

ANALYSIS OF CW-EPR SPECTRA

AND

THE INTERNAL DYNAMICS OF DNA

by Annabelle Wey Reese

Chairperson of the Supervisory Committee: Professor Bruce H. Robinson
Department of Chemistry

A new lineshape analysis program is developed to accurately extract the relaxation rate R_2 from CW-EPR spectra in the fast motion limit, taking into account instrumental broadening, saturation from R_{1e}, and ^{13}C isotope effects. The broadening effects of spin label concentration and oxygen concentration on R_2 are quantitated and a deoxygenation protocol developed.

The internal dynamics of short (100 base pairs or less) DNA molecules spin labeled with two different spin-labeled base pairs are examined. The acetylene-linked C derivative is shown useful for studies of internal dynamics, thus permitting examination of both CG and AT base pairs. The quinolone T derivative, Q, is used to quantitate the length and position dependence of internal DNA dynamics, consistent with the weakly bending rod model's prediction with only a simplistic modification. The dynamic flexural persistence length and rms amplitude of oscillation are estimated. Q is further used to systematically examine the dependence of internal DNA dynamics on base sequence, and the dinucleotide and trinucleotide repeats are ranked for flexibility.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to thank Professors Bruce H. Robinson and Colin Mailer for their invaluable support and guidance during this project. I also thank Professor J. Michael Schurr, who developed the theory for DNA dynamics and has been a constant source of encouragement.

Eric J. Hustedt and Andreas Spaltenstein pioneered this effort through their groundbreaking work on T* spin labeled DNA. Eric also supplied the simulation program for simulating DNA spectra.

Special thanks go to Todd R. Miller, who synthesized the Q spin label, and Stephen C. Alley, who incorporated it into well over 100 DNA sequences. Thuc Nguyen made C* and Paula L. Fischhaber incorporated it into DNA.

I also thank my husband Steven A. Reese for supporting me throughout this endeavor.

CHAPTER 1: INTRODUCTION

SECTION 1.1. INTRODUCTION TO ELECTRON PARAMAGNETIC RESONANCE[1]

Electron paramagnetic resonance (EPR) spectroscopy is a form of magnetic resonance, similar to nuclear magnetic resonance (NMR), but using unpaired electrons instead of nuclei. Electrons in a magnetic field will align either with or against the magnetic field, with populations governed by the Boltzman distribution. Transitions between energy levels are induced by an electric field oscillating at frequency ν and are observed as energy absorption. The difference between energy levels is $\Delta E = g\beta H$, where g is the electronic g-factor ($g \approx 2$), β is the Bohr magneton, and H is the magnetic field, defined as the z-axis of the laboratory frame. Absorption occurs when $g\beta H = h\nu$, where h is Planck's constant and $h\nu$ is the energy of the oscillating electric field. NMR holds the magnetic field H constant while varying frequency ν; EPR holds the frequency ν fixed while varying the field H. EPR spectra are generally simpler than the analogous ^1H NMR spectra, because EPR detects only the unpaired electrons, which are usually few in number.

EPR experiments fall into two categories, continuous-wave- (CW-) EPR and time-domain- (TD-) EPR. Only CW-EPR experiments were used for this work. CW-EPR experiments investigate the steady-state status of the sample and are sensitive to motion on the order of nanoseconds to microseconds. The CW-EPR spectra reflect both molecular motion, characterized by correlation time τ, and magnetic interactions of the electron with the magnetic field, nearby nuclei, and other electrons.

Three magnetic interactions affect the spectral lineshape. First, the interaction between the electron spin $\bar{S}$ and the external magnetic field H (the Zeeman interaction) is responsible for the initial electronic absorption signal and is described by the g tensor in the Hamiltonian, $\bar{S}\cdot g\cdot H$. Second, the hyperfine interaction between the nearby nucleus (spin $\bar{I}$) and the electron splits the initial signal into $2\bar{I}+1$ manifolds; this coupling is described by the A tensor in the Hamiltonian, $\bar{I}\cdot A\cdot\bar{S}$. Spectra of samples containing naturally abundant ^{14}N ($\bar{I}$ = 1) have 3 manifolds; spectra of isotopically substituted ^{15}N ($\bar{I}$ = ½) samples have but 2 manifolds. Finally, the dipolar, or spin-spin, coupling of electron to electron may influence the lineshape as well, if the electrons are close enough to interact.

The Hamilitonian is then $\mathcal{H} = \bar{S}\cdot g\cdot H + \bar{I}\cdot A\cdot\bar{S}$ and the equation of motion is

$$\dot{\chi} = -i[\mathcal{H},\chi] - \Gamma(\chi-\chi_0) - R(\chi-\chi_0)$$

where Γ describes motion as defined by the diffusion coefficients D_{para} and D_{perp} for a rigid body, and R describes relaxation processes such as the spin-spin relaxation or the spin-lattice relaxation. Following standard procedure, g and A are assumed to be collinear.

Changes in motion are expressed through changes in the CW-EPR lineshape. As the correlation time τ increases, the molecular motion slows, and the spectra become broader and more complicated. This is the case for solids. In the fast motion limit, for example in solutions, the anisotropies of the A and g tensors are averaged away and the spectra become simpler.

As a consequence of the instrumentation, CW-EPR spectra look like derivative spectra, in contrast to most other spectroscopies, which yield absorption spectra. See Figure 1.1.

SECTION 1.2. SPIN LABELS

EPR monitors the motion of unpaired electrons within molecules. When the molecule of interest has no unpaired electrons, they may be introduced by attaching a chemical substituent, which contains the unpaired electron, to the molecule. The chemical substituent, called a spin label, must be sensitive to and report changes in its environment, without causing changes in the environment. Ideally the spin label is rigidly coupled to the molecule so that motion reported by the spin label reflects motion of the molecule alone and not of the spin label itself.[2]

Nitroxide free radicals have been a popular family of spin labels. The nitroxyl group is stabilized by the four adjacent methyl groups, as shown in the chemical representation of a nitroxide given in Figure 1.2. These relatively stable free radicals are readily detected with EPR. The orientation of the molecular frame is chosen such that the N-O bond defines the x axis and the z axis is defined in the direction of the nitrogen atom's p orbitals.

The nitroxide spin label has two resonance structures, as shown in Figure 1.3. As solvent polarity increases, the zwitterionic resonance structure with the unpaired spin density on the nitrogen nucleus is favored.

SECTION 1.3. OVERVIEW OF PROJECTS

In the following chapters, several projects will be described. Chapter 2 describes a new lineshape analysis program that was developed to accurately extract the relaxation rate R_2 from CW-EPR spectra in the fast motion limit, taking account of instrumental broadening, saturation from R_{1e}, and ^{13}C isotope effects. The broadening effects of spin label concentration and oxygen concentration on R_2 were quantitated and a deoxygenation protocol developed. Chapters 3 through 6 discuss the internal dynamics of DNA molecules spin labeled with two different spin-labeled base pairs.

NOTES TO CHAPTER 1

1. Nordio, P.L. in *Spin Labeling: Theory and Applications*, ed., L.J. Berliner, pp. 5-52. New York: Academic Press (1976).

2. Berliner, L.J. in *Spin Labeling: Theory and Applications*, ed. L.J. Berliner, pp. 1-4. New York: Academic Press (1976).

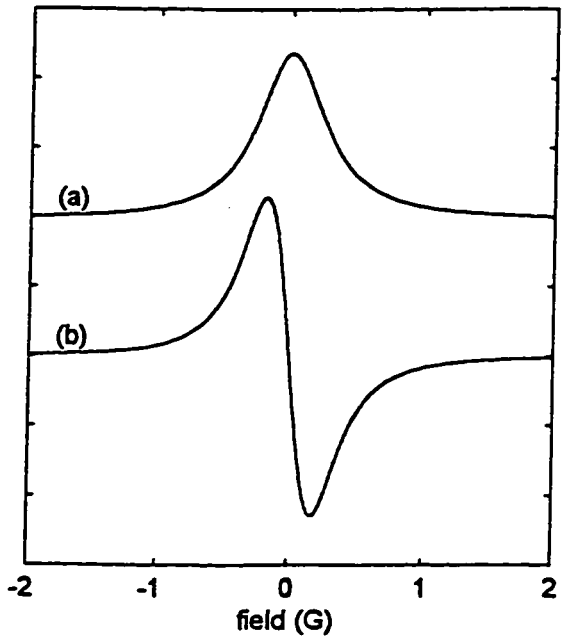


Figure 1.1 Absorption and derivative spectra. CW-EPR spectra look like derivative spectra (b), in contrast to most other spectroscopies, which yield absorption spectra (a).

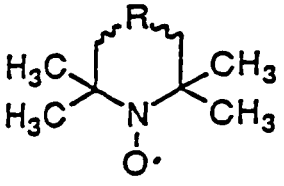


Figure 1.2 Chemical representation of a nitroxide. The N-O bond defines the x-axis; the z-axis points out of the page toward the reader.

$$N^{+}\text{---}O^{-} \text{ <---> } N\text{----}O\bullet$$

Figure 1.3. Resonance structures of a nitroxide.

CHAPTER 2: ANALYSIS OF CW-EPR LINEWIDTHS

SECTION 2.1. INTRODUCTION

Interactions of the spin label (which contains the unpaired electron necessary for EPR) with (paramagnetic) oxygen broaden the CW-EPR spectral linewidth, complicating analysis of the CW-EPR lineshape.[1-7] Therefore, elimination of, or at least understanding, the oxygen broadening effect was sought. In the process of developing a deoxygenation protocol, however, it was realized that the effect of oxygen is small enough that an accurate linewidth analysis protocol is required. Instrumental effects (observer power[8,9] and Zeeman modulation amplitude,[10] for example) additionally broaden the spectral line, as do interactions of the spin label with its neighbors (more of the same spin label)[4-6] and correlation time,[11] which can be varied via solution viscosity. Careful study of the oxygen effect requires that these broadening effects be understood as well. Thus, the goals of this project were to develop a deoxygenation protocol, to understand the oxygen broadening effect, to understand instrumental broadening effects, to understand the spin label concentration effect, and to understand the correlation time effect, as controlled through solution viscosity. To achieve these goals, CW-EPR spectra were collected (varying oxygen concentration, power, Zeeman modulation amplitude, spin label concentration, and solution viscosity, one at a time) and a new lineshape analysis routine was developed.

SECTION 2.2. BACKGROUND

The spin-spin relaxation rate of the electron, R_2, is a fundamental quantity which can, in principle, be directly extracted from the EPR linewidth, in the limit of no broadening influences.[11] However, experimentally, interactions of the spin label with others like it and with oxygen broaden the linewidth.[1-7] R_2 is the parameter that characterizes the spin label- and oxygen-broadened linewidth.

The linewidth is also broadened inhomogeneously by interaction of the spin label with nearby protons. In the fast-motion limit, the inhomogeneously broadened CW-EPR lineshape can be described by the Voigt function, which is the convolution of a Lorentzian line of full width at half height $2R_2$ and a Gaussian line characterized by width σ. The Lorentzian absorption line has the form $y = \frac{T_2}{\pi}\frac{1}{1+T_2^2(x-x_0)^2}$, where the line is centered about $x = x_0$. The full width at half height (in Hertz, Hz) is $\frac{2}{T_2} = 2R_2$, and the peak-to-peak width (in Hz) of the first derivative of this Lorentzian line is $\frac{2}{\sqrt{3}T_2} = \frac{2R_2}{\sqrt{3}}$. See Figure 2.1. (Widths in Hz can be converted to widths in Gauss, G, by the gyromagnetic ratio for the electron, $\gamma = 17.6\times10^6$ rad Hz/G.) The

inhomogeneous broadening is represented by the Gaussian line, of the form $y = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-x_0)^2}{2\sigma^2}\right)$, centered about $x = x_0$. 95% of the curve is contained within the range $\pm 2\sigma^2$. The convolution is $y = L*G = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(y)L(x-y)dy$, where L is the Lorentzian line and G is the Gaussian line, and so the fully convolved line is then $y = \frac{T_2}{2\pi^2\sigma} \int_{-\infty}^{\infty} \frac{\exp\left(\frac{-(y-y_0)^2}{2\sigma^2}\right)}{1+T_2^2(x-x_0-(y-y_0))^2} dy$. R_2 can be directly extracted from the CW-EPR spectral lines by deconvolving the Lorentzian and Gaussian lines. Because rigorous convolution is time consuming, the Voigt function has been approximated by the sum of the Gaussian and Lorentzian functions.[6,12] More recently, a direct convolution of the two functions using a fast fourier transform algorithm has been reported.[13] The approach described herein uses a numerical convolution.

Instrumental effects such as Zeeman modulation amplitude and frequency[10] and observer power[8,9] further broaden the EPR linewidth. As power increases, the linewidth is broadened:

(1) $$L^2 = L_0^2 + \frac{4}{3} h_1^2 \left(\frac{R_{2e}}{R_{1e}}\right)$$

(in the absence of modulation frequency and amplitude), where $L = \frac{2}{\sqrt{3}} \frac{R_2}{\gamma}$ is the peak-to-peak (ptp) width in Gauss, $L_0 = \frac{2}{\sqrt{3}} \frac{R_{2e}}{\gamma}$ is the intrinsic ptp width in the limit of no power, $h_1 \propto \sqrt{\text{power}}$, and R_{1e} and R_{2e} are the intrinsic spin-lattice and spin-spin relaxation rates of the electron. Rewriting equation 1 in terms of R,

(2) $$R_2^2 = R_{2e}^2 + h_1^2 \gamma^2 \left(\frac{R_{2e}}{R_{1e}}\right)$$ for R_{1e} and R_{2e} in Hertz, or

$$R_2^2 = R_{2e}^2 + h_1^2 \left(\frac{R_{2e}}{R_{1e}}\right)$$ for R_{1e} and R_{2e} in Gauss.

A plot of R_2^2 vs. h_1^2 yields intercept R_{2e}^2 and slope R_{2e}/R_{1e}.

Signal height also varies as a function of the incident observer power:[8,9]

(3) $$\Delta Y = \frac{c\sqrt{power}}{\left(1+\frac{power}{P}\right)^{\varepsilon}}$$

where ΔY is the ptp signal height, c is a constant, P is the product $R_{1e} \times R_{2e}$, and ε is a parameter that accounts for the lineshape (ε is 3/2 for a homogeneous first-derivative absorption line). By using $P = R_{1e} \times R_{2e}$ in conjunction with the ratio R_{2e}/R_{1e} from the line broadening study (equations 1 and 2), the intrinsic R_{1e} and R_{2e} can be calculated. The challenge is thus to explicitly include all these broadening effects so that R_2, and ultimately R_{2e}, can be accurately extracted from analysis of the EPR lineshape.

CTPO, 3-carbamoyl-2,2,5,5-tetramethylpyrrolin-1-yloxy, has been used extensively for EPR oxymetry; the EPR spectrum of CTPO changes dramatically as a function of oxygen concentration.[1-7] In 1990, Halpern et al. developed 4-hydro-3-carbamoyl-2,2,5,5-tetraper-deuteromethylpyrrolin-1-yloxy-d12 (mHCTPO).[5] This deuterated derivative of CTPO, with only one hydrogen remaining on the 5-membered ring, was developed to enhance the sensitivity of CTPO for spin-label oxymetry by simplifying the EPR spectra. The fully perdeuterated version, dCTPO, also exists.[5] Structures for CTPO, mHCTPO, and dCTPO are given in Figure 2.2.

Linear CW-EPR spectra of ^{14}N and ^{15}N CTPO have 3 and 2 manifolds, respectively, due to the hyperfine coupling between the nitroxide's N (of spin I, I = 1 for ^{14}N and $\frac{1}{2}$ for ^{15}N) and the unpaired electron. The number of manifolds is governed by 2*I+1.

SECTION 2.3. EXPERIMENTS

A. SAMPLES

Solutions of 5 mM ^{14}N dCTPO and ^{15}N dCTPO were made from powdered dCTPO[14] and nanopure water. To probe the dependence of R_2 on viscosity, a 5 mM ^{15}N dCTPO solution was also made using nominally pure glycerol and then diluted to generate stock solutions of varying % glyercol content. Serial dilution of the stock solutions yielded a range of solutions of different concentrations

Several different approaches were used to develop a deoxygenation protocol. Neither flowing N_2 nor argon gas past TPX[15] tubes containing the sample fully deoxygenated the sample. Bubbled samples sealed with red sira wax,[16] Critoseal,[17] and Torrseal[18] all leaked within a few hours, so that the samples were partially re-oxygenated. The protocol described below had the best success rate of deoxygenating the sample and keeping it deoxygenated.

Samples were deoxygenated with purified, solvent-saturated argon gas.[19] Argon flowed through an oxygen scavenger[20] and then a round-bottomed flask of solvent, sealed with a

rubber septum, to saturate the gas and reduce the amount of sample evaporation. The gas was bubbled through the solvent for at least 20 hours to purge the solvent completely, and then re-purged for 1 hour immediately before bubbling through the sample. Typically 1 ml of sample, in a culture tube sealed with a rubber septum, was bubbled for 3 hours. Either the solvent flask and the sample tube were cooled with ice baths or the glovebag environment (which contained the solvent flask and sample tube) was cooled with dry ice, again to reduce evaporation. The capillary in which the sample was later collected was placed in the gas outlet needle for purging.

Air samples were either not bubbled at all or bubbled with air that had passed through a dehydrator[21] and then bubbled through solvent for 50 hours. However, spectra of the samples that were air-bubbled had narrow linewidths, suggesting partial deoxygenation, perhaps through contamination of the dehydrating compound. Consequently only spectra of non-bubbled samples were used.

Samples were transferred to quartz capillaries of inner diameter 0.4 mm or 0.6 mm[22] via capillary action and sealed with red sira wax[16] at both ends. A syringe[23] was used to transfer oxygenated and deoxygenated samples into capillaries sealed at one end.[24] The open ends of the capillaries were temporarily sealed with red wax, then flame sealed within 10 minutes. "Infinite length" sample lengths (so-called because the sample length exceeded the length of exposure to microwaves) were > 1.5 cm for samples used in the Loop Gap Resonator[25] (LGR, described below) and > 4 cm for samples in the standard TE_{102} cavity. However, for the power studies, "point" samples, only 2 mm to 3 mm in length, were used to reduce magnetic field inhomogeneity effects. For samples studied in the TE_{102} cavity, the capillary was placed inside a quartz EPR tube, then inside the cavity. Tape was rolled around the capillaries, about 1 cm from the end without sample, to provide stability within the EPR tube and to keep the capillary standing upright.

B. INSTRUMENTATION

The 9 GHz, X-band homebuilt spectrometer has been described in detail elsewhere.[8,26] The klystron provides monochromatic radiation of fixed frequency; this radiation is carried through wave guides to the sample, which is held between the poles of the magnet. Helmholtz coils (controlled by a hall probe) vary the field of the magnet. The sample absorbs the radiation under resonance conditions, and the resulting signal is removed from the Zeeman (carrier) modulation frequency by an Ithaco Dynatrak-3 lock-in analyzer, which detects both the in-phase and out-of-phase (quadrature) signals. The signals are digitized by a Data Translation DT 2781 serial interface and stored on a DEC PDP 11/23 microcomputer.

The sample is placed in either a standard TE_{102} cavity or a Loop Gap Resonator (LGR). The LGR offers more efficient power (mW) to h_1 (G) conversion, with $\alpha_{LGR} = 4.5\ G/\sqrt{Watt}$,

while $\alpha_{TE102} = 1.5\ G/\sqrt{Watt}$:[8]

(4) $$h_1 = \alpha\sqrt{Power}$$ with h_1 in Gauss (G) and Power in Watts (W).

Consequently, heating of the sample is minimized and better signal-to-noise is achieved for any given power. Also, the area of exposure to the microwaves is much smaller, so sample length is reduced from ~4 cm to ~2 cm, which reduces the field inhomogeneity at the sample. Finally, the capillary goes directly into the LGR, so the EPR sample tube (required for the TE_{102} cavity) is no longer needed.

EPR spectra are derivative spectra, in contrast to, for example, NMR spectra, as a result of the instrumentation. Displaying the data in derivative form also enhances sensitivity to and resolution of the turning points and zero-cross-overs, which are the critical parts of the EPR spectrum. See Figure 2.1.

C. EXPERIMENTS

Linear CW-EPR experiments were run on the samples described above, using the PDP homebuilt spectrometer. For the concentration and viscosity studies, spectra were collected under the following conditions: 10 kHz modulation frequency, 0.050 G modulation amplitude, 128 G sweep range, 1024 points, 0.125 s response time, observer power 0.1 mW = 0.045 G h_1, 4 scans, temperature approximately 20°C. To examine the effect of Zeeman modulation amplitude, all conditions except modulation amplitude were as listed above, and modulation amplitude was varied. Similarly, to examine the effect of observer power, all conditions except power were as given above and power itself varied.

D. SIMULATIONS

The spectra were simulated to extract the best-fit parameters. The fitting routine is written in MATLAB[27] and simulates the EPR spectrum from a set of input parameters, optimizing these parameters to minimize the deviation of the simulation from the experimental data. The simulation produces both absorption and dispersion spectra for both in-phase and out-of- phase signals (with respect to the Zeeman modulation) at any harmonic and for any Zeeman modulation amplitude and frequency. The following parameters are included in the simulation:

- spectral line crossovers and Lorentzian linewidths
- Gaussian σ, an approximation to account for unresolved hyperfine splitting from nearby protons or deuterons
- absorption / dispersion phase rotation
- Zeeman modulation frequency, Zeeman modulation amplitude, and Zeeman phase
- spin-lattice relaxation rate R_{1e}
- observer power

- ^{13}C splitting

The fitting routine constructs the simulation by first generating Lorentzian lines of the given linewidths at the given positions, taking into account the instrumental parameters and R_{1e}. Each Lorentzian line is convolved with the ^{13}C splitting pattern. The lines are superimposed to create the complete spectrum. The complete signal is convolved with the Gaussian, then rotated to account for the absorption / dispersion mix and to minimize the quadrature signal. This signal is then compared to the experimental spectrum and the parameters optimized for best fit.

The standard nonlinear least squares method is used for the minimization. In addition, the (arbitrary) scale and baseline offset of the simulation are removed to reduce the number of non-linear parameters included in the minimization.[28] To do this, the variance, σ^2, is defined in terms of the model, scale (s), and baseline (b), where Y_i is the data point at the i^{th} position in the spectrum, and $\hat{Y}_i$ is the corresponding simulated point:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - s\hat{Y}_i - b\right)^2$$

Then σ^2 is minimized. By setting $\frac{d\sigma^2}{ds}$ and $\frac{d\sigma^2}{db}$ to zero, the optimal scale and baseline can be determined. If the variance of the data and the simulation are defined as

$$\sigma_y{}^2 = \left\langle (Y - \langle Y \rangle)^2 \right\rangle \text{ and } \sigma_{\hat{y}}{}^2 = \left\langle (\hat{Y} - \langle \hat{Y} \rangle)^2 \right\rangle$$

then the standard error is proportional to the correlation coefficient and the variance of the data:

$$\sigma^2 = (1 - R^2)\sigma_y{}^2$$

where R is the mean-corrected correlation coefficient:

(5)
$$R = \frac{\left\langle (Y - \langle Y \rangle)\right\rangle\left\langle (\hat{Y} - \langle \hat{Y} \rangle)\right\rangle}{\sigma_y \sigma_{\hat{y}}}$$

R is dimensionless and independent of scale and baseline. Because σ_y is a function of the data alone, it has no effect on the optimization of the simulation. Therefore, R accurately reflects the agreement between the experimental data and the simulation, and when R is maximized, σ is minimized. In the analysis program, the quantity $(1-R^2)$ is minimized with a Simplex type search (using the MATLAB routine fmins). The code is given in Appendix A.

To check the accuracy of the fitting program, a series of power study experiments was simulated and analyzed. When the power incident on the sample is increased, the signal height decreases and signal width increases per equations 2 and 3 above. From plots of signal height

and width vs. power, R_{1e} and R_{2e} are calculable. Synthetic data were generated including the effects of Zeeman modulation frequency, Zeeman modulation amplitude, R_{1e}, and power; the Lorentzian-only version of the data was subsequently convolved with a Gaussian (σ = 0.11 G) to produce the Lorentzian-Gaussian version of the data. From both versions the ptp heights and widths were measured. Estimates of R_2 were calculated from the relationship $\text{ptp width} = \frac{2R_2}{\sqrt{3}}$ and used to generate the appropriate plots; R_{1e} and R_{2e} were then calculated and compared to the true values of R_{1e} and R_{2e} used to create the synthetic data.

Results from this series of control experiments are listed in Table 2.1. R_{1e} and R_{2e}, when calculated from the Lorentzian-Gaussian convolved version of the data, are 11% to 33% greater than the R_{1e} and R_{2e} values used to create the data; R_{1e} and R_{2e} calculated from the non-convolved data are within 2% of the values used in creating the data. Clearly the correct approach is to use the R_2 from the non-convolved data. Evidently Gaussian convolution broadens the line such that the correct R_{1e} and R_{2e} are not retrieved. This result confirms the fitting algorithm: the Lorentzian linewidth R_2 is extracted first, then convolved with the Gaussian when simulating spectra.

In addition, this series of control experiments gives a measure of the intrinsic error in the calculations of R_{1e} and R_{2e}: without any noise in the synthetic data, the error in R_{1e} and R_{2e} is approximately 1% except for one case. The exceptionally large error in that case is believed to reflect the relatively low resolution of points in the synthetic spectra compounding the innate difficulty in calculating a small value. The analogous power study simulation using greater resolution of points yields less error; using the $R_{1e} \times R_{2e}$ value determined from this study with the slope R_{2e}/R_{1e} determined from line broadening from the lower-resolution study results in correct R_{1e} and R_{2e} values with smaller errors.

These simulations also show that high power data is needed to correctly calculate R_{1e} and R_{2e} (data not shown). In practical terms, this means that the LGR must be used for power studies, as the TE_{102} cavity's lower efficiency of conversion of power to h_1 makes high power data not experimentally achievable. Finally, to accurately calculate R_{1e} and R_{2e}, using the line broadening analysis alone yields larger errors than using both the line broadening (R_{2e}/R_{1e}) and the rollover ($R_{2e} \times R_{1e}$) analyses.

In review, then, CW-EPR experiments were run on samples of dCTPO, in the presence and absence of oxygen, varying spin label concentration, % glycerol, and instrumental parameters such as modulation amplitude, modulation frequency, and observer power. The

experimental spectra were simulated using the fitting program described above and R_2 was extracted. Typically, Gaussian σ, Zeeman modulation frequency and amplitude, spin-lattice relaxation rate R_{1e}, observer power h_1, and ^{13}C hyperfine splitting were kept fixed, while the spectral crossover positions and linewidths and the absorption/dispersion phase were optimized. R_2 was then analyzed as a function of modulation amplitude, spin label concentration, and oxygen concentration to elucidate these relationships; also, R_{1e} and R_{2e} were calculated.

SECTION 2.4. RESULTS & DISCUSSION

R^2, the square of the correlation coefficient R defined above, ranged from 0.97 (for a few cases) to greater than 0.99 (for most cases).

Preliminary simulations optimized the Gaussian σ to 0.1088 ± 0.0055 G; thereafter a single value of σ = 0.11 G was used for all the simulations to reduce the number of variables. This value of σ is in excellent agreement with the calculation of Bales et al. for perfectly deuterated ^{14}N mHCTPO, σ = 0.11 G, although the difference in compounds (^{14}N mHCTPO vs. ^{14}N dCTPO and ^{15}N dCTPO) must be noted.[29]

Spin label concentrations were calculated from the double integrals of the (noise-free) simulations without absorption / dispersion phase rotation. These concentrations were scaled relative to the double integral of a 1.0 mM ^{14}N CTPO sample to determine nominal concentrations.

A. INSTRUMENTAL EFFECTS

The effect of Zeeman modulation amplitude on R_2 is given in Table 2.2. Nominally 0.1 mM dCPTO samples were studied at different modulation amplitudes. The concentrations of all the samples were assumed to be the same in the analysis, although the double integrals indicated some variability in the sample concentrations. Excellent fits showed that the simulation routine adequately modeled the effect of modulation amplitude on the spectral line, even when the linewidth was overbroadened by modulation amplitudes much wider than the intrinsic linewidth, as shown in Figure 2.3. Beyond achieving agreement with experimental data at high modulation amplitude, the simulation program also consistently extracted the same lorentzian linewidth over a wide range of modulation amplitude. Without including the effect of modulation amplitude in the simulation, the apparent R_2 increased tremendously and nonlinearly with linearly increasing modulation amplitude. When modulation amplitude was explicitly included in the simulation, the R_2 remained constant, as shown in the relatively flat slope of the R_2 vs. modulation amplitude line. See Figure 2.4.

The effect of Zeeman modulation frequency on R_2 was not studied in detail; however, within the range of the study (10 kHz to 100 kHz), the effect was minimal. Modulation frequency

was explicitly included in simulations.

For the Loop Gap Resonator, the absorption / dispersion mixture required up to 20° phase rotation, whereas the phase correction for the TE_{102} cavity was less than 5°.

Power studies showed the effect of power on R_2 and allowed calculations of R_{2e} and R_{1e} as well, from purely CW-EPR techniques. Power studies on the ^{14}N isotope of dCTPO failed completely, as insufficient high-power data was collected. The results given in Figure 2.5 and Table 2.3 are based entirely on the ^{15}N dCTPO studies. Combining the information gathered from both the line broadening and rollover analyses, R_{1e} and R_{2e} were estimated respectively at 0.1075 ± 0.0018 G and 0.2392 ± 0.0040 G in air and 0.0707 ± 0.0018 G and 0.1272 ± 0.0033 G in argon, at T ≈ 20°C. Earlier Time-Domain EPR experiments in this laboratory predicted $R_{2e}/R_{1e} \approx 2$ in air at T ≈ 20°C;[30] these CW-EPR experiments yielded $R_{2e}/R_{1e} \approx 2.2$ (in air) and 1.8 (in argon).

B. Effect of spin label concentration

R_2 depended linearly upon sample concentration, for both ^{14}N and ^{15}N samples, as shown in Table 2.4 and Figure 2.6. The earliest studies used a small range of concentrations (0.1 mM to 0.5 mM) which resulted in nonsensical values of broadening per mM of spin label. A later study using a wider range of concentrations (0.5 mM to 5 mM) showed the dependence of R_2 on ^{15}N dCTPO (0% glycerol) concentration in air to be 132.0 ± 1.3 mG / mM. The analogous experiments with a wide concentration range were not performed on the ^{14}N spin label. Literature values for the effect of spin label concentration on R_2 are 145 ± 3 mG / mM[6] and 157 mG / mM[4] broadening, which are 9% and 16% greater than the value determined here. Perhaps the discrepancy is due to a difference in isotope: the literature values were derived from studies of ^{14}N mHCTPO, whereas this study examined ^{15}N dCTPO. Why isotopes should make this difference is unclear as yet.

C. Effect of viscosity

The viscosity experiments were intended to probe the dependence of R_2 on τ. A range of spin label concentrations in different % glycerol solvents were studied. The results, shown in Table 2.5 and 2.6 and Figures 2.7 and 2.8, suggest that R_2 depends on τ^{-1}. τ may be calculated using the difference of the linewidths of the two ^{15}N manifolds. This difference, in radians / second, is B in the expression $R_{2e}(m) = A(m)+mB(m)+m^2C(m)$, where m is $\pm\frac{1}{2}$ for the ^{15}N manifolds and A, B, and C are defined per Haas et al.[9] B is related to the correlation time τ through this expression, which came originally from Goldman et al.[31] and was later modified by

Hwang et al.,[32] before Haas et al. refined it further.[9] τ may also be calculated from the viscosity η of the solvent, using the Stokes-Einstein equation[33]

(6) $$\tau = \frac{4}{3}\pi r^3 \frac{\eta}{k_b T},$$

with $r = 3.0\times10^{-8}$ cm = 3Å, η = viscosity in poise[34], $k_b = 1.38\times10^{-16}$ erg/K, and T = 293 K.

τ ought not change with spin label concentration. However, for the samples in 70% glycerol, the linewidths, and the difference in linewidths, did change dramatically with concentration (data not shown), so that τ changed as the spin label concentration changed. Furthermore, disagreement between the two methods of calculating τ showed that the nominal % glycerol was up to 15% inaccurate. The 70% glycerol solution used to dilute the 70% glycerol, 5 mM dCTPO stock solution is believed to have been contaminated with water. This contamination was propagated through the serial dilution. Consequently, the experiments were compromised.

In a crude attempt to salvage the 70% glycerol data, the correlation time τ was calculated from the difference in linewidths for each concentration of the serial dilution. These correlation times were used to calculate the expected linewidths for each concentration, using the simple ratio $\frac{\text{linewidth}_{\text{calculated}}}{\text{linewidth}_{\text{stock}}} = \frac{\tau_{\text{calculated}}}{\tau_{\text{stock}}}$, where stock referred to the 5 mM ^{15}N dCTPO in nominally 70% glycerol solution. The newly calculated linewidths were used to calculate the effect of spin label concentration in 70% glycerol, and then this effect was compared to the corresponding effect of spin label concentration in 0%, 15%, and 40% glycerol.

The broadening caused by spin label concentration (slopes of R_2 vs. spin label concentration plots) in 0%, 15%, 40%, and 70% glycerol was plotted against $\frac{\omega\tau}{1+(\omega\tau)^{p+1}}$, where $\omega = 2\pi\nu$ (ν is the frequency at which these experiments were run, 9 GHz = 9×10^9 Hz), τ is the correlation time, as calculated from the difference in linewidths, and p is the dependence of R_2 on τ (p = 0.5 indicates that $R_2 \propto \tau^{-.5}$, p = 1.0 indicates that $R_2 \propto \tau^{-1}$, and p = 2 indicates that $R_2 \propto \tau^{-2}$). Neither the original nor the recalculated spin label concentration-caused broadening for the 70% glycerol dataset are correctly described by this analysis. However, the recalculated broadening is at least less than the broadening measured in the 40% glycerol dataset, which is qualitatively correct.

Neglecting the 70% glycerol datapoints, R_2 appears to depend inversely on τ ($R_2 \propto \tau^{-1}$) according to the least-squares best-fit lines, since the broadening caused by spin label concentration ought to be 0 when the spin label has a correlation time of τ = 0 s. Furthermore, if

p = 0.8, the broadening vs. $\frac{\omega\tau}{1+(\omega\tau)^{p+1}}$ line does go through the origin, although there is no rationale for why p should be 0.8. Nevertheless, p = 0.8 confirms qualitatively at least the p = 1 dependence described above. More rigorous analysis awaits the collection of a new dataset, using clean, uncontaminated glycerol in the serial dilutions.

However, even once "clean" results are determined, the broadening in 70% glycerol may still not be accurately described by this analysis, due to a "glycerol slippage" phenomenon described by Haas.[33] In higher % glycerol solutions, the correlation time τ was smaller than expected from the Stokes-Einstein equation (equation 6). A smaller τ, assuming $\omega\tau>1$ as has been the case heretofore, would result in a larger $\frac{\omega\tau}{1+(\omega\tau)^{p+1}}$, which would move the broadening in 70% glycerol closer to the line described by the 0%, 15%, and 40% glycerol data. Approximately a 5-fold decrease in correlation time, from 5.82×10^{-10} s to 1.35×10^{-10}, would pull the broadening into line. An approach to making this correction is described by Haas.[33]

D. Effect of Oxygen

The effect of O_2 was shown in the original spin label concentration studies, the modulation amplitude effect experiments, and the power studies, and is summarized in Table 2.7. From the original spin label concentration studies, the presence of O_2 caused 131.1 + 2.2 mG (^{14}N) and 130.9 ± 1.4 mG (^{15}N) broadening at the 0 mM SL limit. However, these values are highly suspect, since they came from analyses of data over a severely limited range of concentrations. In the modulation amplitude experiments, the presence of O_2 caused 113.1 ± 1.5 mG (^{14}N) and 116.1 ± 0.9 mG (^{15}N) broadening. Oxygen caused 112.0 ± 3.5 mG broadening in R_{2e} for ^{15}N from the power studies. The concentration of O_2 in water exposed to atmospheric air is 0.270 mM,[35,36] so the O_2 effect translates to 113.1 ± 1.5 mG / 0.270 mM = 418.9 ± 5.6 mG / mM O_2, for example. Neglecting the original spin label concentration study results, the average oxygen effect is 421.2 ± 6.2 mG / mM. Although the effect of O_2 has been calculated here from a line of only 2 points, the values compare acceptably with the literature (^{14}N) values: 476 ± 11 mG / mM[6] and 501 mG / mM[4]. Again the value determined here is 9% and 16% less that the literature values. Although the deoxygenation may have been incomplete, the discrepancy may again have been due to differences in isotope.

Section 2.5. Conclusions

A deoxygenation protocol was developed for deoxygenating the spin label dCTPO by bubbling solutions with cleaned argon. The effect of oxygen concentration on R_2 was within 16%

of the literature values.

A new analysis routine for CW-EPR lineshape simulation was developed. The simulation explicitly includes instrumental broadening effects (observer power, phase, Zeeman modulation), saturation effects requiring R_{1e}, and ^{13}C effects, in addition to the inhomogeneous broadening approximated by the Gaussian convolution. R_2 is accurately extracted under a variety of spectrometer conditions. Consequently, R_2 can be accurately determined even when the spectrum is instrumentally broadened. In practice, this result means that the spectrum can be deliberately overbroadened, for example to improve the signal to noise ratio, without compromising the final results.

The effect of spin label concentration on R_2 was also within 16% of the literature values. Although the viscosity effect experiments were compromised, R_2 appeared to be proportional to τ^{-1}. Additionally, power studies (a purely CW-EPR technique) resulted in estimates of R_{1e} and R_{2e} that compared favorably with estimates from TD-EPR.

This work suffered from an incomplete experiment set. Spin label concentration studies of deoxygenated ^{15}N dCTPO and of both oxygenated and deoxygenated ^{14}N dCTPO, using the wider range of concentrations (0.5 mM to 5 mM), should replace the studies using the smaller range of concentrations (0.1 mM to 0.5 mM). These studies should yield better estimates of the effect of both spin label and oxygen concentration. Power studies on both oxygenated and deoxygenated ^{14}N dCTPO are also in order and should be done using the LGR. These studies should give another estimate of R_{1e} and R_{2e} and the oxygen effect. Complete comparisons with literature results would necessitate repeating all the experiments with mHCTPO and CTPO as well, using both ^{14}N and ^{15}N isotopes. Halpern has already generously donated ^{15}N mHCTPO for that purpose. In addition, the viscosity effect experiments should be repeated with clean glycerol, to yield a better set of data from which to determine the dependence of R_{2e} on τ.

Halpern has also suggested incorporating oxygen-sensitive methylene blue, which changes from colorless to blue in the presence of oxygen, into the dCTPO sample, as a measure of the effectiveness of the deoxygenation protocol.[37] Additional Time Domain experiments on the CW-EPR samples would help to make the direct comparison of R_{1e} (CW) and R_{1e} (TD).

A major application of this deoxygenation work is in deoxygenating DNA samples for EPR study. However, bubbling may be physically hard on DNA molecules, so another protocol may need to be developed. For example, blowing on the sample, but not through it, may be more acceptable for biological samples.

Finally, the analysis routine used for this work accounted for inhomogeneous broadening by convolving the Lorentzian line with a Gaussian line. This convolution is a convenient

approximation, but it remains only an approximation. Explicitly including the individual splittings from the nearby methyl group proton and deuteron spins should yield a more accurate result. A new analysis routine has been developed to do exactly this. Preliminary results suggest that that the deuteration of dCTPO complicates the spectra rather than simplifying them, and that distinguishing between ^{14}N and ^{15}N isotopes is significant. Consequently, differences between literature values and those determined here may be at least partially attributed to isotopic differences: most literature work used [H,^{14}N] CTPO, but this work used mostly [D,^{15}N] CTPO.

NOTES TO CHAPTER 2

1. Popp, C.A., & Hyde, J.S. *J. Mag. Res.* **43**, 249-258 (1981).

2. Lai, C.S., Hopwood, L.E., Hyde, J.S., & Lukiewicz, S. *Proc. Natl. Acad. Sci. USA* **79**, 1166-1170 (1982).

3. Hyde, J.S. & Subczynski, W.K. *J. Mag. Res.* **56**, 125-130 (1984).

4. Bowman, M.K., Michalski, T.J., Peric, M., Halpern, H.J. *Pure and Applied Chemistry* **62**, 271-274 (1990).

5. Halpern, H.J., Peric, M., Nguyen, T., Spencer, D.P., Teicher, B.A., Lin, Y., & Bowman, M.K. *J. Mag. Res.* **90**, 40-51 (1990).

6. Halpern, H.J., Peric, M., Yu, C., Bales, B.L. *J. Mag. Res.* **A103**, 13-22 (1993).

7. Smirnova, T.I., Smirnov, A.I., Clarkson, R.B, Belford, R.L. *Mag. Res. in Medicine* **33**, 801-810 (1995).

8. Mailer, C., Haas, D.A., Hustedt, E.J., Gladden, J.G., & Robinson, B.H. *J. Mag. Res.* **91**, 475 - 496 (1991).

9. Haas, D.A., Mailer, C., & Robinson, B.H. *Biophys. J.* **64**, 594-604 (1993).

10. Robinson, B.H. *J. Chem. Phys.* **78**, 2268-2273 (1983).

11. Nordio, P.L. in *Spin Labeling: Theory and Applications*, ed. L.J. Berliner, pp 5-52. New York: Academic Press (1976).

12. Peric, M. & Halpern, H.J. *J. Mag. Res.* **A109**, 198-202 (1994).

13. Smirnov, A.I. & Belford, R.L. *J. Mag. Res.* **113**, 65-73 (1995).

14. C/D/N Isotopes, Vaudreuil, Quebec, Canada.

15. TPX is a methylpentene polymer used in deoxygenation protocols.[1] Samples of TPX were generously given by Boedeker Plastics, Inc. (TX) and Westlake Plastics Company (PA), and were machined by Mr. John Heutink in the Machine Shop of the Department of Chemistry, University of Washington.

16. Red sira wax was produced by BDH Isotope Co, England, and was a generous gift from Dr. Al Beth.

17. Critoseal, #8889-215008, Monoject Scientific, St. louis, MO.

18. Torr Seal Low Vapor Pressure Resin, Varian, Lexington MA.

19. The apparatus design was based upon that in the laboratories of Dr. D. Michael Heinekey, Department of Chemistry, University of Washington. The glassware was blown by Mr. Robert Morley in the Glassblowing Shop of the Department of Physics, University of Washington.

20. The oxygen scavenger (catalyst R3-11, Chemical Dynamics Corporation) was a generous gift from Dr. D. Michael Heinekey.

21. Drierite, anhydrous Calcium Sulfate, #13005, W.A. Hammond Drierite Company, LTD, Xenia, OH.

22. Clear fused quartz round capillary tubes, 0.6 mm ID x 0.84 mm OD x 100 mm long and 0.4 mm ID x 0.55 mm OD x 100 mm long, were purchased from Vitrocom, Inc., Mountain Lake, NJ.

23. A 10 μl Hamilton syringe with a 6", 33 gauge stainless steel needle was custom made by Hamilton Company, Reno, NV.

24. Capillaries were sealed at one end by Mr. Bob Moreley in the Glassblowing Shop of the Department of Physics, University of Washington.

25. Loop Gap Resonator, Medical Advances, Milwaukee Wisconson, Model #XP-0201.

26. Mailer, C., Danielson, J.D.S., & Robinson, B.H. *Rev. Sci. Instr.* **56**, 1917-1025 (1985).

27. MATLAB Software, The MathWorks, Inc., Natick, MA.

28. Hustedt, E.J., Spaltenstein, A., Kirchner, J.J., Hopkins, P.B., & Robinson, B.H. *Biochemistry* **32**, 1774-1787 (1993).

29. Bales, B.L., Blum, R.A., Mareno, D., Peric, M., & Halpern, H.J. *J. Mag. Res.* **98**, 299-307 (1992).

30. Robinson, B.H., Haas, D.A., & Mailer, C. *Science* **363**, 490-493 (1994); Robinson, B.H., Reese, A.W., & Mailer, C., 18th International EPR Symposium, Denver CO (1995); Mailer, C., unpublished results.

31. Goldman, A., Bruno, G.V., Polnaszek, C.F., & Freed, J.H. *J. Chem. Phys.* **56**, 716-735 (1972).

32. Hwang, J.S., Mason, R., Hwang, L.P. & Freed, J.H. *J. Phys. Chem.* **79**, 489-511 (1975).

33. Haas, D.A. Ph.D. Thesis, University of Washington, 1993.

34. *Handbook of Chemistry*, ed. N.A. Lange, p. 1669. New York: McGraw-Hill Book Company (1961).

35. *Solubility Data Series, Volume 7. Oxygen and Ozone*, ed. R. Battino, p. 2. New York: Pergamon Press (1981)

36. *CRC Handbook of Chemistry and Physics*, ed. R.C. Weast, p. F-79. Cleveland, Ohio: CRC Press (1975).

37. Murphy, J.G., Smith, T.W., & Marsh, J.D. *J. Mol. Cell Cardiol.* **19**, 271-279 (1987).

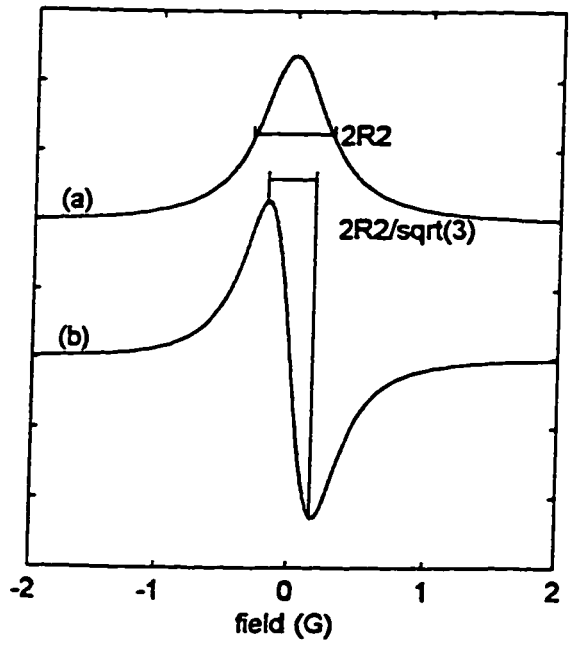
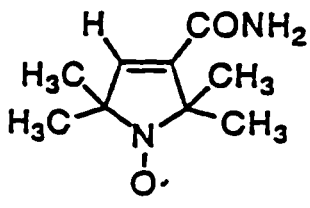
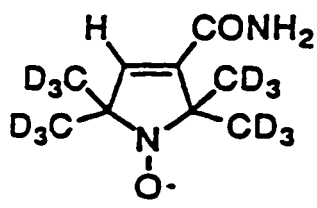


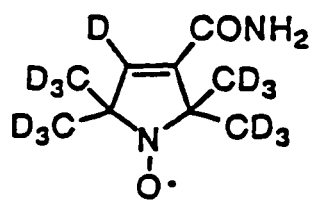
Figure 2.1. Absorption and derivative Lorentzian lines. The full width at half height on the absorption line is $2R_2$; the peak-to-peak width of the derivative line is $\frac{2R_2}{\sqrt{3}}$.



CTPO



mHCTPO



dCTPO

Figure 2.2. Structures of CTPO, mHCTPO, and dCTPO.

Table 2.1: Results of power simulations. The simulated Lorentzian line was analyzed both with ("with convolution") and without ("no convolution") the Gaussian convolution of 0.11 G. Estimates of R_2 were calculated from the measured peak-to-peak widths and used to generate the plots described by equation 2. The intercept of the R_2^2 vs. h_i^2 line is R_{2e}^2 (in G^2), from which R_{2e} (G) may be calculated. The slope of the line is R_{2e}/R_{1e}(G/G). Peak-to-peak heights were used with equation 3 and the product $P = R_{1e} \times R_{2e}$ (G^2) was extracted. R_{1e} (G) and R_{2e} (G) were then calculated from the quotient R_{2e}/R_{1e} and the product $R_{1e} \times R_{2e}$. "Resolution" designates the number of points used over a given scan range in the simulation. The true R_{1e} (G) and R_{2e} (G) that were used to create the synthetic data are also listed under "parameters."

	no convolution	with convolution	parameters
intercept = R_{2e}^2	0.0563 ± 0.0069	0.1179 ± 0.0045	128 points / 4G
R_{2e} from intercept	0.2372 ± 0.0290	0.3434 ± 0.0132	resolution
R_{2e}/R_{1e}	1.2596 ± 0.0132	1.2952 ± 0.0087	
$R_{1e} \times R_{2e}$	0.0374 ± 0.0002	0.0468 ± 0.0007	
R_{1e}	0.1723 ± 0.0019	0.1901 ± 0.0033	0.1700
R_{2e}	0.2170 ± 0.0033	0.2462 ± 0.0046	0.2200
intercept = R_{2e}^2	0.0494 ± 0.0004	0.1143 ± 0.0013	1024 points / 4G
R_{2e} from intercept	0.2222 ± 0.0019	0.3380 ± 0.0038	resolution
R_{2e} / R_{1e}	1.2970 ± 0.0008	1.2980 ± 0.0025	
$R_{1e} \times R_{2e}$	0.0376 ± 0.0001	0.0467 ± 0.0008	
R_{1e}	0.1703 ± 0.0001	0.1897 ± 0.0031	0.1700
R_{2e}	0.2208 ± 0.0002	0.2462 ± 0.0040	0.2200
intercept = R_{2e}^2	0.0083 ± 0.0027	0.0640 ± 0.0070	128 points / 4G
R_{2e} from intercept	0.0910 ± 0.0296	0.2529 ± 0.0278	resolution
R_{2e} / R_{1e}	1.2721 ± 0.0051	1.2543 ± 0.0133	
$R_{1e} \times R_{2e}$	0.0044 ± 0.0283	0.0080 ± 0.0006	
R_{1e}	0.0588 ± 0.3782	0.0799 ± 0.0061	0.0600
R_{2e}	0.0748 ± 0.4812	0.1002 ± 0.0077	0.0750
intercept = R_{2e}^2	0.0064 ± 0.0005	0.0624 ± 0.0020	1024 points / 4G
R_{2e} from intercept	0.0801 ± 0.0063	0.2498 ± 0.0080	resolution
R_{2e} / R_{1e}	1.2506 ± 0.0009	1.2666 ± 0.0038	
$R_{1e} \times R_{2e}$	0.0047 ± 0.0000	0.0081 ± 0.0002	
R_{1e}	0.0613 ± 0.0003	0.0800 ± 0.0020	0.0600
R_{2e}	0.0767 ± 0.0004	0.1013 ± 0.0253	0.0750

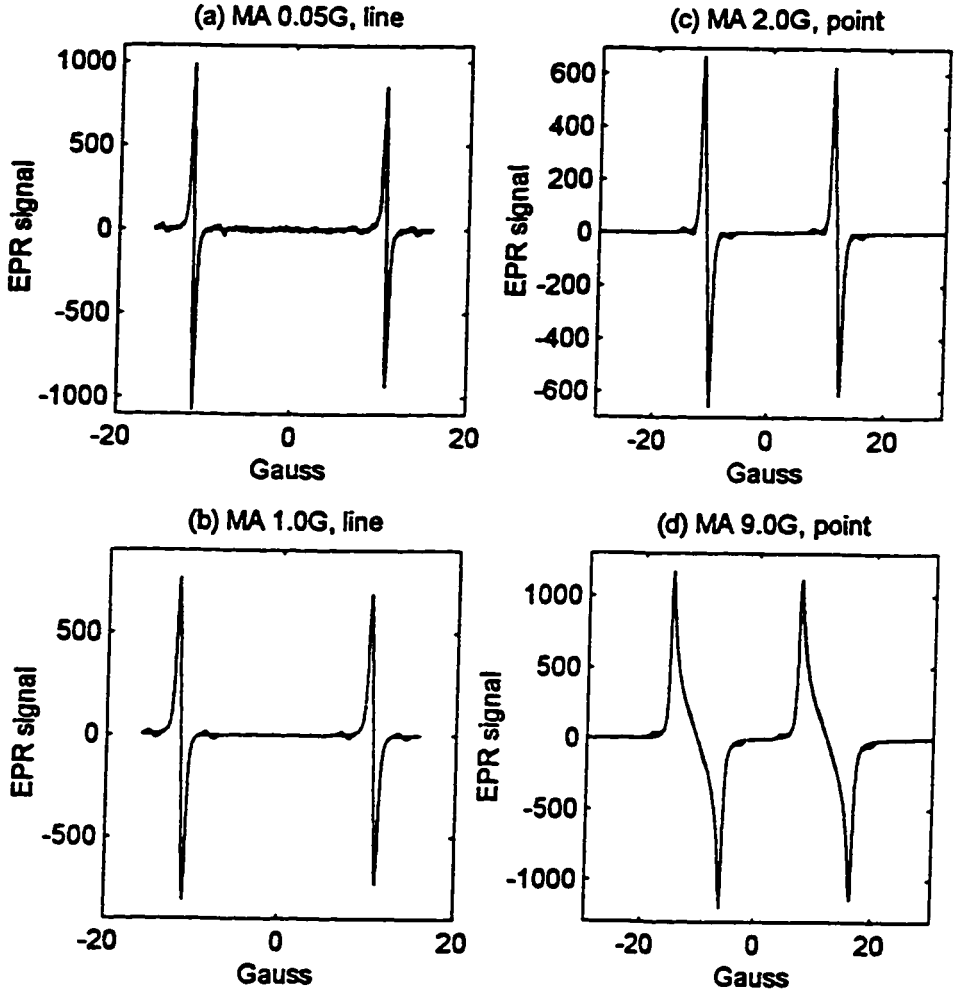


Figure 2.3. CW-EPR spectra of ^{15}N dCTPO, in air, are shown under conditions of varying Zeeman modulation amplitude (MA) and dCTPO concentration. In each case, the spectral simulation (dotted line) is overlaid on the data (solid line). Line samples were of "infinite length" as described in the text; point samples were 2 - 3 mm long. R^2 is the square of the correlation coefficient. Point samples were used for the higher modulation amplitude experiments, because the line samples suffered from the distribution of the modulation amplitude over the sample's active volume. Excellent fits of simulation to data were achieved, even when modulation amplitude was much larger than the intrinsic linewidth. (a) line sample, 0.06 mM ^{15}N dCTPO, MA = 0.05 G, R^2 = 0.9937. (b) line sample, 0.06 mM ^{15}N dCTPO, MA = 1.00 G, R^2 = 0.9981. (c) point sample, 0.20 mM ^{15}N dCTPO, MA = 2.0 G, R^2 = 0.9980. (d) point sample, 0.20 mM ^{15}N dCTPO, MA = 9.0 G, R^2 = 0.9987.

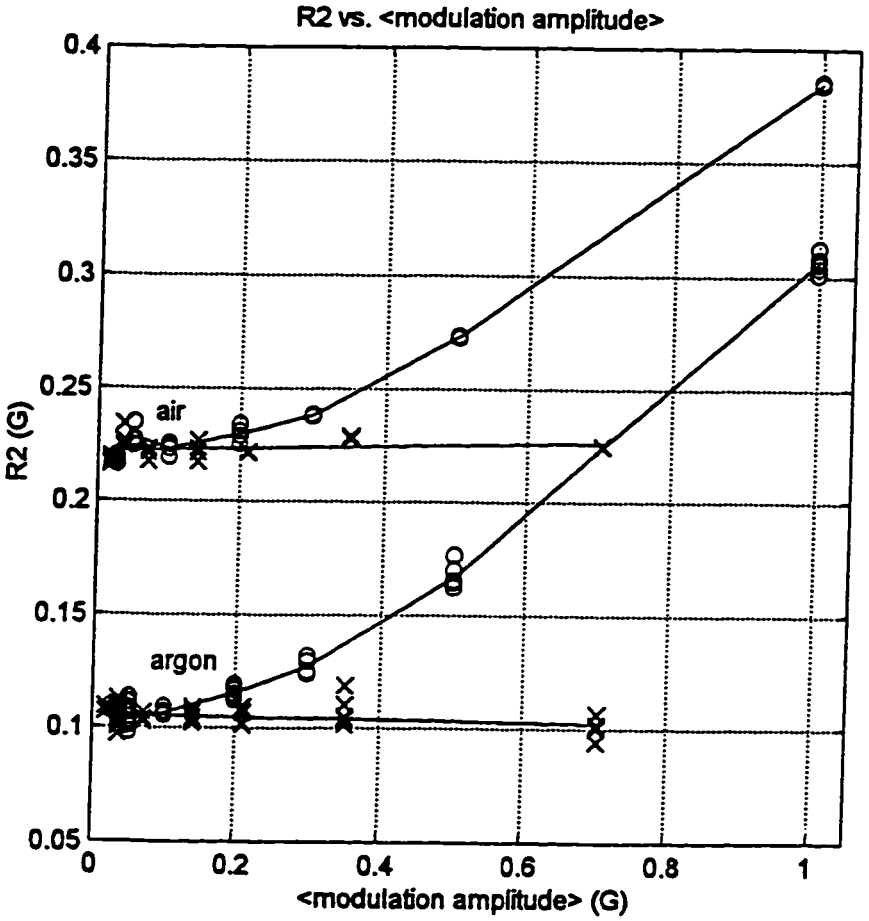


Figure 2.4. R_2 as a function of Zeeman modulation amplitude, both neglecting (o) and explicitly including (x) the effect of the modulation amplitude in the analysis. Because the experiments were run on line samples and there is a distribution of modulation amplitudes, R_2 is plotted against the average modulation amplitude. In this case, $\langle h_m \rangle = 0.705\ h_m$. This type of approximation is valid to $h_m \leq 8 * R_2$.

Table 2.2. Broadening of R_2 caused by the Zeeman modulation amplitude, when modulation amplitude is explicitly included in the analysis. Intercepts represent R_2 (in G) in the limit of no modulation amplitude; slopes give the amount of broadening of R_2 (in G) caused by increased modulation amplitude (in G). σ is the standard deviation of the least-squares best-fit line described by the intercept and slope.

		air	argon	R_2(air)-R_2(argon)
^{14}N	intercept	0.2416 ± 0.0026	0.1285 ± 0.0010	0.1131
	slope	0.0037 ± 0.0169	-0.0024 ± 0.0064	
	σ	0.0088	0.0032	
^{15}N	intercept	0.2220 ± 0.0011	0.1059 ± 0.0007	0.1161
	slope	0.0058 ± 0.0034	-0.0051 ± 0.0021	
	σ	0.0040	0.0037	

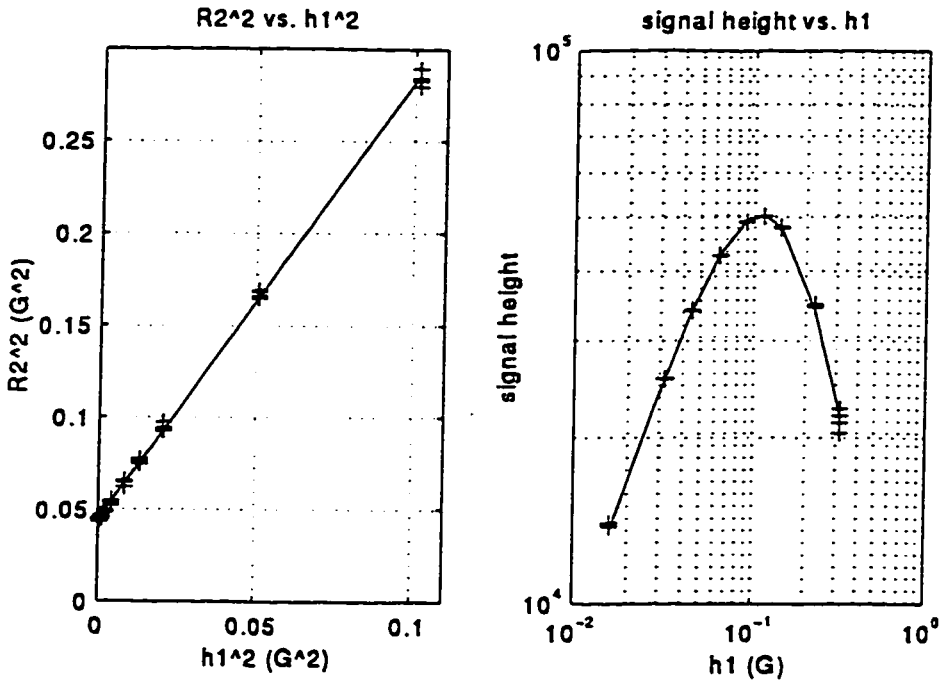


Figure 2.5. Representative plots of R_2 as a function of observer power. (a) linewidth broadening curve, based on equation 2. (b) rollover curve, based on equation 3.

Table 2.3. R_{1e} and R_{2e} calculated from power studies on ^{15}N dCTPO (insufficient high-power data was collected for ^{14}N dCTPO). "LW" represents the linewidth broadening analysis summarized in equation 2, $R_2^2 = R_{2e}^2 + h_1^2\left(\frac{R_{2e}}{R_{1e}}\right)$. $\frac{R_{2e}}{R_{1e}}$ is the slope determined from the R_2^2 vs. h_1^2 line; σ is the standard deviation of the best-fit line to the data. "Roll" represents the rollover analysis given in equation 3. ε accounts for the lineshape (ε = 3/2 for a homogeneous first-derivative absorption line). $P=R_{1e}\times R_{2e}$ is extracted from the rollover curve. R^2 is the correlation coefficient (R = 1 is a perfect fit). R_{1e} and R_{2e} (given in units of G) are calculated from both the quotient $\frac{R_{2e}}{R_{1e}}$ and the product $R_{1e}\times R_{2e}$. $\Delta R_{2e} = R_{2e,air} - R_{2e,argon}$.

^{15}N		air	air	argon	ΔR_{2e}
LW	$\frac{R_{2e}}{R_{1e}}$	2.0985 ± 0.0178	2.3720 ± 0.0102	1.7981 ± 0.0359	
	σ	0.0030	0.0020	0.0070	
roll	ε	1.5804	1.5163	1.2172	
	$R_{1e}\times R_{2e}$	0.0276 ± 0.0003	0.0238 ± 0.0003	0.0090 ± 0.0001	
	$1-R^2$	0.0002	0.0001	0.0004	
LW + roll	R_{1e}	0.1147 ± 0.0014	0.1002 ± 0.0012	0.0707 ± 0.0018	
	R_{2e}	0.2407 ± 0.0030	0.2376 ± 0.0027	0.1272 ± 0.0033	
	mean(R_{1e})	0.1075 ± 0.0018			
	mean(R_{2e})	0.2392 ± 0.0040			0.1120 ± 0.0035

Table 2.4. The effect of spin label concentration on R_2. R_{2e} is the intercept of the R_2 vs. spin label concentration ([SL]) line and represents R_2 extrapolated for a solution of 0 mM spin label. The slope of the R_2 vs. [SL] line gives the broadening caused by spin label concentration. R_{2e} is given in Gauss; the slope is given in G / mM spin label. σ is the standard deviation of the least-squares best fit line described by the intercept R_{2e} and slope. The second column of air data for ^{15}N dCTPO comes from the dataset using a wider (0.5 mM - 5 mM) range of spin label concentrations; the other data are derived from the narrower (0.1 mM - 0.5 mM) range of spin label concentrations. $\Delta R_{2e} = R_{2e,air} - R_{2e,argon}$.

^{14}N	air	air	argon	ΔR_{2e}
R_{2e}	0.2507 ± 0.0015		0.1196 ± 0.0019	0.1311 ± 0.0022
slope	0.1179 ± 0.0059		0.1020 ± 0.0079	
σ	0.0110		0.0049	
^{15}N				
R_{2e}	0.2305 ± 0.0015	0.1857 ± 0.0024	0.0996 ± 0.0008	0.1309 ± 0.0014
slope	0.0380 ± 0.0049	0.1320 ± 0.0013	0.0447 ± 0.0028	
σ	0.0068	0.0062	0.0048	

Table 2.5. The effect of spin label concentration on R_2 in solutions of varying viscosity. The estimates of τ (in seconds) are calculated from the linewidth difference between the two manifolds of the ^{15}N spectrum and from the nominal viscosity. R_{2e}, slope, and σ are defined as for Table 2.4. The second entry for the 70% glycerol case is derived from a re-calculation as described in the text.

% glycerol	τ (Δ linewidth)	τ (viscosity)		
0	3.00×10^{-11}	2.81×10^{-11}	R_{2e}	0.1857 ± 0.0024
			slope	0.1320 ± 0.0013
			σ	0.0062
15	4.70×10^{-11}	4.24×10^{-11}	R_{2e}	0.1489 ± 0.0027
			slope	0.1066 ± 0.0015
			σ	0.0025
40	8.95×10^{-11}	1.05×10^{-10}	R_{2e}	0.1471 ± 0.0010
			slope	0.0722 ± 0.0008
			σ	0.0031
70	5.82×10^{-10}	6.42×10^{-10}	R_{2e}	0.3026 ± 0.0030
			slope	0.0838 ± 0.0022
			σ	0.0102
70	5.82×10^{-10}	6.42×10^{-10}	R_{2e}	0.3726 ± 0.0016
			slope	0.0629 ± 0.0013
			σ	0.0026

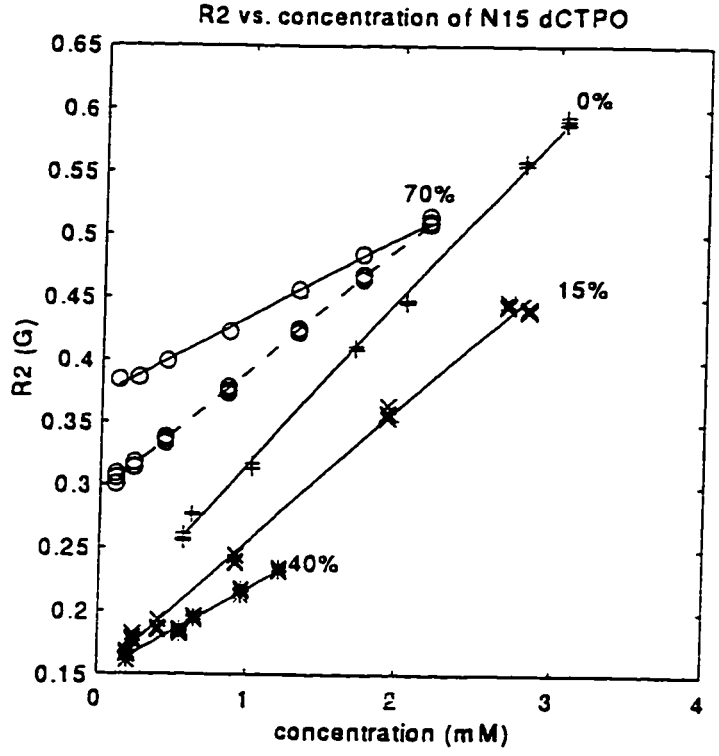


Figure 2.6. R_2 (in G) as a function of spin label concentration (in mM) in 0% (+), 15% (x), 40% (*), and 70% (o) glycerol solutions. Least-squares best-fit lines are superimposed on the data; intercepts, slopes, and σ for each fit are listed in Table 2.5. The dotted line passes through the original 70% glycerol data; the solid line passes through the recalculated 70% glycerol data.

Table 2.6. Intercepts, slopes, and σ of least-squares best-fit lines (shown in Figure 2.7) to the 0%, 15%, and 40% glycerol data, for m = 0.5, 0.8, 1.0, and 2.0.

m	intercept	slope	σ
0.5	-0.1171 0.0408	0.4631 ± 0.0851	0.0077
0.8	0.0008 ± 0.0041	0.2778 ± 0.0107	0.0016
1.0	0.0272 ± 0.0015	0.2430 ± 0.0044	0.0008
2.0	0.0680 ± 0.0098	0.2396 ± 0.0547	0.0094

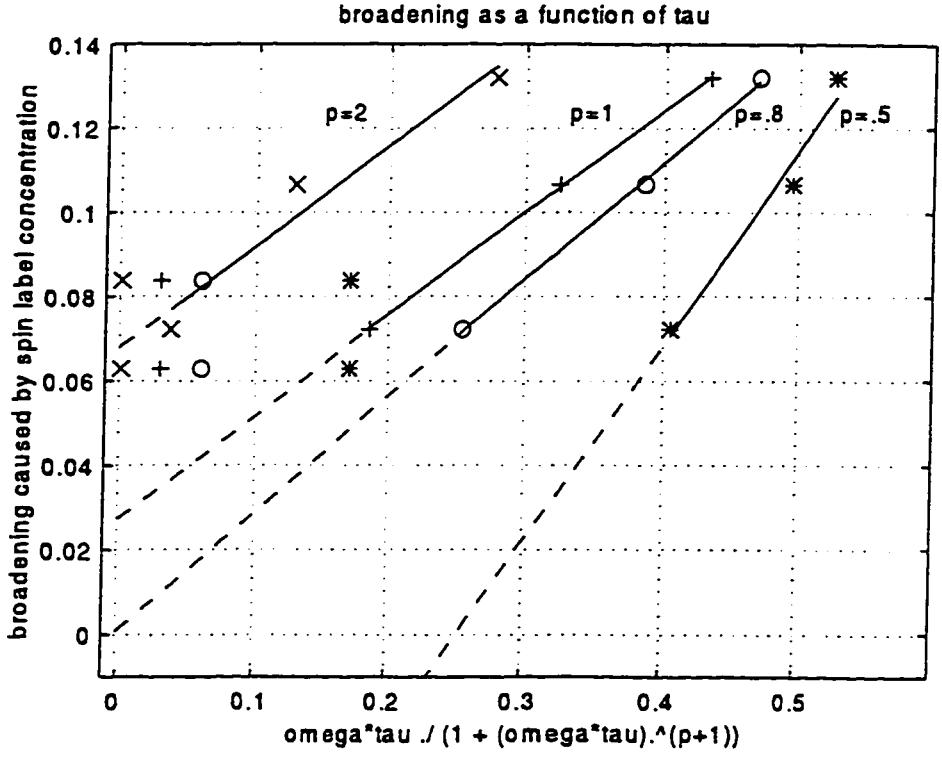


Figure 2.7. R_2 broadening caused by spin label concentration (in G / mM) as a function of $\frac{\omega\tau}{1+(\omega\tau)^{p+1}}$, where $\omega = 2\pi\nu$ (ν is the frequency at which these experiments were run, 9 GHz = 9×10^9 Hz), τ is the correlation time (in seconds), calculated from the difference in linewidths, and p is the dependence of R_2 on τ, $R_2 \propto \tau^{-p}$. p = 0.5 (*), p = 0.8 (o), p = 1.0, and p = 2 (x) are depicted. Least-squares best-fit lines to the 0%, 15%, and 40% glycerol data are superimposed on the data; intercepts, slopes and σ for each fit are given in Table 2.6.

Table 2.7. The effect of oxygen concentration on R_{2e}. Results are summarized from Table 2.2 (broadening caused by Zeeman modulation amplitude [MA]), Table 2.3 (results of power studies [power]), and Table 2.4 (broadening caused by spin label concentration [SL]). Because values from Table 2.4 are derived from analyses of data over severely limited ranges of spin label concentration, the values are highly suspect and are neglected in calculating the mean effect of oxygen concentration on R_{2e}. $\Delta R_{2e} = R_{2e,air} - R_{2e,argon}$ is given in mG and shows the difference in R_{2e} for samples in the presence and absence of oxygen. The concentration of O_2 in the aerated samples is 0.270 mM;[35,36] the broadening caused per mM oxygen is $\frac{R_{2e,air} - R_{2e,argon}}{0.270\ mM}$ and is given in mG / mM.

	analysis	ΔR_{2e}	broadening
^{14}N	[SL]	(131.1 ± 2.2)	(485.6 ± 8.2)
	MA	113.1 ± 1.5	418.9 ± 5.6
^{15}N	[SL]	(130.9 ± 1.4)	(484.8 ± 5.2)
	MA	116.1 ± 1.0	430.0 ± 3.7
	power	112.0 ± 3.5	414.8 ± 13.0
mean neglecting [SL] results			421.2 ± 6.2

CHAPTER 3: DNA DYNAMICS

SECTION 3.1. INTRODUCTION

Understanding the dynamics of solution state DNA is fundamental to understanding the functionality of DNA.[1] Knowledge of how DNA moves by itself may provide insight into how DNA replicates and how DNA interacts with other molecules such as proteins and drugs. Because of the interest in this topic, many different techniques have been used to investigate this question, including fluorescence polarization anisotropy, optical spectroscopy, NMR, and EPR, with considerable differences in the interpretation of the resulting data.[2-5] One issue of debate has been the effect of sequence on dynamics. Two particular bones of contention have been the estimate for the dynamic flexural persistence length of DNA and the estimate of the magnitude of DNA's internal motions.

SECTION 3.2. EPR ON DNA

There are a number of advantages in using EPR spectroscopy to study DNA dynamics. EPR is well suited for the investigation of molecular motion in the nanosecond to microsecond time domains. The tumbling and internal motions of short (<100 base pairs) duplex DNA molecules fall within this time window. In linear CW-EPR experiments, changes in correlation time are reflected by changes in the spectral lineshapes and line positions, as shown in Figures 3.1 and 3.2. As the correlation time increases, the lineshapes broaden and the spectral width increases. In addition, EPR spectroscopy is very sensitive and relatively small amounts of sample are required, for example, compared to NMR: EPR uses nanomoles of DNA, whereas the corresponding NMR experiment uses hundreds of nanomoles of DNA. Another advantage is that EPR experiments yield relatively simple spectra: the spectrum is generally that of the single unpaired electron and its interactions with nearby nuclei (usually only one). The corresponding ^{1}H NMR experiment would yield a much more complex spectrum, from the many NMR-detectable nuclei present in the DNA sample, all of which interact and produce signals. (NMR spectra on isotopically-labeled (^{2}H, ^{13}C, or ^{31}P, for example) DNA samples would be considerably simplified, similar to the EPR spectra.)

Natural DNA contains no unpaired electrons and therefore is EPR silent. To examine DNA with EPR requires that the DNA molecule be made EPR-active. Either intercalating or covalently attaching ("spin-labeling") a chemical substituent that contains a single unpaired electron (the spin label) to the DNA meets this requirement. The spin label must perturb the structure and function of the DNA molecule as little as possible and should also have minimal independent motion. Any such motion should, ideally, be easily distinguished from that of the DNA molecule and should not compete with or obscure the motion of the DNA molecule.[6] Although spin-labeling the DNA molecule does constitute an unnatural modification of a base pair and consequently the entire molecule, the technique also renders site-specific labeling

possible. Knowing exactly where on the molecule the spin label sits means that effects can be studied site-specifically, rather than averaged over the length of the entire molecule, as is required for intercalated or otherwise randomly labeled samples.

SECTION 3.3. HISTORY OF SPIN LABELS

In the early 1980s, acridine spin labels were intercalated into DNA over the length of the entire molecule.[7] By 1984, Bobst et al. had attached a spin label to a specific site on DNA, using a nitroxide free radical on the end of a long flexible tether (6 σ bonds).[8] Although the length of the tether practically guarantees that the spin label does not interfere with the DNA motion, the independent spin label motion compromises the interpretation of the data, as the rapid motion of the spin label itself obscures the more subtle and slower DNA dynamics.[5,9] By 1995, Bobst et al. had reported another spin label with a tether of only 2 σ bonds, which more faithfully reports DNA dynamics.[10,11]

However, by 1988, Spaltenstein et al. had designed a different class of spin label, robust enough to be used in automated DNA synthesis.[12] This spin label, T*, is a nitroxide attached by an acetylene tether to a modified thymine (deoxyuridine). T* pairs to A as does unmodified T, and is shown in Figure 3.3. Although T* does not interfere with the structure[9,12] and is highly sensitive to DNA dynamics,[5] it has two drawbacks. First, the spin label is not completely rigidly attached to the DNA base, so the spin label may spin about the σ bond of the acetylene tether. Second, depyrimidination may occur after exposure to high temperatures (6 hours at 90°C), causing the modified base to disintegrate.[13] Nevertheless, much of the ground-breaking work in studying DNA internal dynamics using EPR was done with T*, and it is upon these studies that the current work is based.

A new spin label, Q, was designed to address the two drawbacks of T* and is shown in Figure 3.3. Coupling the nitroxide with a rigid 6-membered aromatic ring to the modified base eliminates rotation about the tether. Replacing the nitrogen atom with carbon at the sugar-base connection (creating a "C-nucleoside" instead of an "N-nucleoside") removes the threat of depyrimidination. Solomon et al. demonstrated that the precursor to this new spin label pairs best with the commercially available 2-Aminopurine, P, rather than A.[14] Synthesis of Q was completed in 1995.[15]

Q was also incorporated into synthetic DNA in 1995.[16,17] Incorporation was confirmed by quantitation of coupling yields (released trityl cation), the length of the strand as estimated by electrophoretic mobility (denaturing polyacrylamide gel electrophoresis (d-PAGE)), significant UV absorption for purified single strands at 330 nm where dQ absorbs light but unlabeled DNA does not, and the return of dQ in HPLC analysis following digestion of the purified single strands with a phosphodiesterase/phosphotase mixture.[16,17]

Native polyacrylamide gel electrophoresis (PAGE) demonstrated that the spin-labeled DNA strands matched preferentially to the P-containing complement strands, consistent with design. UV-monitored thermal denaturation measurements further investigated the thermal stability of the QP base pair. The melting temperature T_m denotes the temperature at which 50% of the DNA duplexes have dissociated to single strands. T_m for an 11-mer containing the QP base pair (5'-d(CGA GGQ CCA GC)·5'-d(GC TGG PCC TCG)) was 45°C, 7°C lower than T_m (52°C) for the analogous unlabeled 11-mer (5'-d(CGA GGT CCA GC)·5'-d(GC TGG ACC TCG)). The circular dichroism (CD) spectrum of both 11-mers possessed negative and positive molar ellipticities at ca. 250 and 280 nm, respectively, characteristic of right-handed double helices. However, the molar ellipticity per reside from 225 to 300 nm was more negative for the 11-mer containing QP than for the unlabeled 11-mer. Because the impact of the chromophore in Q on the CD spectrum is unknown, no further structural conclusions were drawn.[16,17] 1D and 2D NMR experiments and assignments further characterized the spin-labeled DNA as being overall in the B-form.[18]

SECTION 3.4. DNA DYNAMICS

A brief description of the classes of DNA motion and of what EPR can and cannot measure follows. During the last 20 years, Schurr and co-workers have developed the "weakly bending rod" model of the modes of motion of DNA.[3,19] The DNA molecule is considered to be a flexible rod-shaped object, with the long axis defined as z. As sketched in Figure 3.4, the rod, similar to a string of tuna-fish cans, is comprised of base pairs that exhibit local cylindrical symmetry ($\Delta x^2=\Delta y^2$). The base pairs are connected by bonds behaving as Hookean springs, which provide constraints against which the molecule may bend or twist.

The DNA molecule's motions come from three distinct sources, schematically shown in Figure 3.5: (1) global motion, (2) internal collective motion, and (3) internal local motion. Global motion, also called global rotation or uniform rotation, represents Brownian motion of the entire molecule as it tumbles through space as a rigid object with no internal flexibility. The motion is independent of spin label position but does depend on DNA length: the longer the molecule, the more slowly the molecule moves. Global rotation can be classified as motion parallel ("axial") and perpendicular ("end over end" or "head over heels" tumbling) to the z axis.

Internal motions refer to motion within the DNA molecule. Internal collective motion is motion at one point of the molecule as a consequence of motions from the rest of the molecule. The motion depends on both position and length and is classified as flexural motion (about x or y), like bending the elbow, and torsional motion (about z), like wringing a dishcloth. Internal local motion refers to the aggregate motion of the spin label and the base pair to which it is attached, as well as the motion of the spin label and the base pair alone, independently from the rest of the

DNA. Internal local motion is independent of DNA length.

EPR spectra reflect the dynamics of spin-labeled DNA. At 9 GHz, the frequency at which all experiments were run, the EPR spectra are relatively insensitive to the g tensor anisotropy and therefore to motion about the z axis of the nitroxide. Furthermore, because the spin labels T* and Q are oriented approximately 20° off the principal axis (z) of the DNA molecule,[5,18] the EPR spectra of T* and Q-labeled DNA are again relatively insensitive to motion about the z axis but highly sensitive to motion perpendicular to z, namely x-y motion.

SECTION 3.5. GENERAL EXPERIMENTAL METHODS

Duplex DNA containing the new spin label Q was synthesized by the automated phosphoramidite method, except for incorporation of the dQ residue, which was conducted manually to conserve the phosphoramidite precursor to Q.[17] Both the protonated ^{14}N and deuterated ^{15}N isotopes of Q were used. DNA was suspended in PNE buffer (10 mM phosphate buffer (pH 7.0), 115 mM Na^+, 0.1 mM EDTA) or in 50% (weight /volume, w/v) sucrose in PNE buffer. Only non-self-complementary sequences were studied. For the length and position dependence studies, a typical mixture of labeled (Q-containing) to partner (P-containing) DNA strand was 1:1.5; the sequence dependence mixtures were 1:1.

Samples of DNA were transferred to quartz capillary tubes[20] via capillary action and sealed at the bottom with red sira wax.[21] Typical samples were 10 mg/ml; for a 50-mer this represents approximately 0.6 mM DNA.[22] At 1 nitroxide per DNA and assuming 100% labeling, this translates to 0.6 mM NO· for each 50-mer sample.

Experiments were run on two different X-band (9 GHz) spectrometers, the homebuilt PDP spectrometer (described in Chapter 2) and the Bruker EMX commercial spectrometer.[23] A Varian TE_{102} cavity and a Medical Advances Loop Gap Resonator (LGR, also described in Chapter 2) were used on the homebuilt PDP spectrometer; only a TE_{102} cavity was used on the Bruker spectrometer.

Spectra were collected on the homebuilt PDP spectrometer under the following conditions: 10 kHz modulation frequency, 1.0 G modulation amplitude, 128 G sweep range, 1024 points, 0.125 s response time, 0.1mW = 0.045 G power, 5 - 10 scans per spectrum. The temperature was controlled by nitrogen gas flowing through the cavity or LGR via a dry ice / acetone bath and was regulated to within 2°C.

Spectra were collected on the Bruker spectrometer under very similar conditions, with the following changes: 80 G sweep range, 81.92 ms time constant, 163.96 ms conversion time, 2 mW = 0.063G power, 10 scans. Temperature was controlled by nitrogen gas flowing through the cavity via a liquid nitrogen dewar and regulated to within 0.1°C.

The EPR spectra of the spin-labeled DNAs are typical of objects tumbling in the

nanoseconds to hundreds of nanoseconds time scale. As a result, the spectra look very different from the simple, 2 or 3 sharp line spectra of rapidly-tumbling CTPO seen in Chapter 2. Although there still are 2 or 3 main manifolds (^{15}N and ^{14}N respectively), the spectral features are broadened and the spectral width is increased. The ^{14}N spectra are more sensitive to the x and y components of the g and A tensors; the ^{15}N spectra, to the z component. Consequently, pooling ^{14}N and ^{15}N data improved the level of sensitivity. The middle line of the ^{14}N spectra shows more hyperfine features but is also more difficult to simulate, because of overlap; as a result the majority of the analysis was done with DNAs labeled with ^{15}N Q, although ^{14}N Q-labeled DNAs were also studied.

SECTION 3.6. SIMULATIONS

The EPR spectra of spin-labeled DNA reflect both the DNA dynamics described above and the interaction of the spin label with its magnetic environment. The experimental spectra were simulated, first to obtain the rigid limit g and A tensors characterizing the magnetic interactions and then to investigate the internal dynamics.

The simulation program is based upon the Stochastic Liouville Equation and has been described in detail elsewhere.[5] Input parameters for the program are

1. the magnetic tensors g (g_{xx}, g_{yy}, g_{zz}) and A (A_{xx}, A_{yy}, A_{zz}), which are assumed to be coincident,
2. the intrinsic Lorentzian (homogeneous) linewidth r, and
3. a post-calculation Gaussian convolution of width σ, representing (inhomogeneous) broadening caused by nearby nuclear spins other than the nitroxide's nitrogen, such as non-equivalent protons or deuterons.

To simulate dynamics, the additional input parameters are

4. the global dynamics characterized by the correlation times τ_{perp} (perpendicular to the z axis) and τ_{para} (parallel to the z axis), where $\tau_a = \frac{1}{6D_a}$, a = perp or para, and D_a are the rotational diffusion coefficients described below,
5. the tilt angle θ_{tilt}, reflecting the orientation of the coincident g and A tensors of the spin label relative to the axially symmetric rotational diffusion tensors of the DNA duplex, and
6. the parameter $\langle\beta^2\rangle$ as described below.

The rigid limit (no dynamics) g and A tensors were extracted from the spectra of the ^{14}N- and ^{15}N-labeled 50-mers in 50% w/v sucrose at 0°C. The sequences are described in detail in Chapter 4 and are the left-, middle-, and right-labeled 50-mers. In 50% w/v sucrose at 0°C, the global dynamics (τ_{perp} = 7150 ns, τ_{para} = 640 ns, from Table 3.2) are so slow as to be

invisible on the EPR time scale. Only the magnetic interactions (characterized by the tensors) remained. The 6 tensors (g_{xx}, g_{yy}, g_{zz}; A_{xx}, A_{yy}, A_{zz}) were optimized separately for the ^{14}N and ^{15}N spectra sets (3 DNA samples for each isotope), using 0.5 G Gaussian broadening, and allowing the homogeneous linewidths to float also. The A tensor for ^{15}N, A(^{15}N), was then calculated with A(^{15}N) * 0.7131 = A(^{14}N), where 0.7131 is the ratio of the gyromagnetic ratios for the two isotopes. The g(^{15}N) tensor was taken to be the same as the g(^{14}N) tensor. The calculated tensors were applied to the ^{15}N data to see how well the calculated tensors fit. The analogous procedure was used to fit the ^{14}N data with the calculated ^{14}N tensors. The compromise set of tensors found to fit both set of spectra is given in Table 3.1. Overlays of data and simulations are shown in Figure 3.6.

Global dynamics were calculated, as a function of DNA molecule length (N base pairs), temperature, and viscosity, per Tirado and de la Torre[24] assuming radius R = 12.0 Å and length L = 3.4N Å.[5] For a right circular cylinder of length L and radius R, the rotational diffusion coefficients are

(1) $$Dperp=\frac{3k_bT(\ln(p)+\delta_{perp})}{\pi\eta L^3} \qquad Dpara=\frac{k_bT}{(1+\delta_{para})A\pi\eta LR^2}$$

where $p=\frac{L}{2R}$, A = 3.841, k_b = 1.38x10^{-16} ergs/K is the Boltzman constant, T is the absolute temperature (K), and η is the solvent viscosity in poise. Following the parametrization of Hustedt et al.,[5]

(2) $$\delta_{perp}=-0.662+\frac{0.913}{p}-\frac{0.042}{p^2} \qquad \delta_{para}=\frac{0.689}{p}-\frac{0.203}{p^2}$$

The viscosity of the buffer alone was assumed to be that of pure water and the viscosity of the sucrose solutions was calculated according to Barber.[25] Correlation times were calculated using $\tau_a=\frac{1}{6D_a}$, where a = perp or para, and are given in Table 3.2.

The tilt angle θ_{tilt}, defined by Hustedt et al.,[5] describes the orientation of the g and A tensors of the spin label relative to the rotational diffusion tensors of the DNA molecule. The A tensor is treated below; the g tensors may be treated equivalently by replacing A with g. The A tensor at any non-diagonal orientation, A', may be written using an Euler angle rotation, $A'=\boldsymbol{R}\cdot A_d\cdot\boldsymbol{R}^T$, where A_d is the A tensor in its diagonal orientation and $\boldsymbol{R}=\boldsymbol{R}_\psi\cdot\boldsymbol{R}_\theta\cdot\boldsymbol{R}_\beta$, where

$$(3)\qquad R_\psi = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix},\ R_\theta = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix},\ \text{and } R_\beta = \begin{pmatrix} \cos\beta & \sin\beta & 0 \\ -\sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

See Figure 3.7. We average over β, described below. ψ interchanges A_{xx} and A_{yy}, but for the cylindrical symmetry used here to describe DNA, $A_{xx} = A_{yy}$, so the spectrum is independent of ψ and in fact ψ is fixed at 0° in the simulation program. Following the treatment of T* by Hustedt et al.,[5,22] θ was assumed to be 20° but was not optimized. Subsequent 2D NMR experiments and initial structure determination confirmed this assumption. More recent refinement of the structure estimates $\theta \approx 26°$.[18] Recent optimization of θ, following the procedure described by Hustedt et al.,[5,22] shows $\theta \approx 15°$-$20°$. Consequently, although using an optimized θ would have been better, the initial assumption of $\theta = 20°$ was acceptable.

The internal dynamics are characterized here by a single parameter $\langle\beta^2\rangle$, which pre-averages the tensors, assuming that the internal dynamics are moving rapidly. Rapid motion is defined as $\Delta A \cdot \tau < 1$, where ΔA is the amount by which the tensors are narrowed by the motion and τ is the correlation time of the motion.[26] The parameter $\langle\beta^2\rangle$ represents the mean squared amplitude of the DNA motion away from its equilibrium position, and as such, the "flexibility" of the molecule. $\langle\beta^2\rangle$ describes how much the spectrum must be narrowed to best simulate the spectral width without changing the line shapes. $\langle\beta^2\rangle$ was defined by Hustedt et al:[5]

$$(4)\qquad \begin{aligned} \bar{A}_{xx} &= A_{xx} + \tfrac{1}{2}\langle\beta^2\rangle(A_{zz} - A_{perp}) & \bar{g}_{xx} &= g_{xx} + \tfrac{1}{2}\langle\beta^2\rangle(g_{zz} - g_{perp}) \\ \bar{A}_{yy} &= A_{yy} + \tfrac{1}{2}\langle\beta^2\rangle(A_{zz} - A_{perp}) & \bar{g}_{yy} &= g_{yy} + \tfrac{1}{2}\langle\beta^2\rangle(g_{zz} - g_{perp}) \\ \bar{A}_{zz} &= A_{zz} - \langle\beta^2\rangle(A_{zz} - A_{perp}) & \bar{g}_{zz} &= g_{zz} - \langle\beta^2\rangle(g_{zz} - g_{perp}) \end{aligned}$$

$$\text{where}\quad A_{perp} = \tfrac{1}{2}(A_{xx} + A_{yy})$$
$$g_{perp} = \tfrac{1}{2}(g_{xx} + g_{yy})$$

As explained by Hustedt et al., these expressions were defined for *limited* amplitude motion ($\beta <$ ca. 20°) and are applicable to both 1-dimensional and 2-dimensional dynamics.[5]

The EPR spectra of Q-labeled DNA in 0% sucrose were simulated to determine the internal dynamics, using the rigid limit g and A tensors (given in Table 3.1), and the global dynamics as calculated from de la Torre (given in Table 3.2). θ_{tilt} was fixed at 20°, the Lorentzian linewidth was fixed at r = 0.4 G, and a σ = 0.5 G Gaussian convolution was included. Only $\langle\beta^2\rangle$ was varied. The simulation routine minimized χ^2 to optimize $\langle\beta^2\rangle$; $\langle\beta^2\rangle$ was then manually refined to achieve the best agreement between the spectral widths of the experimental data and the simulation, ignoring agreement of the inner parts of the spectra. Error in β, σ_β,

represents the point at which agreement between data and simulations using β and $\beta \pm \sigma_\beta$ was equivalent and is approximately 0.003 radians. Given $\sigma_\beta = 0.003$, σ_{β^2} was calculated with Gauss's rule: $\sigma_{\beta^2}^2 = \left(\frac{\partial(\beta^2)}{\partial\beta}\right)^2 \sigma_\beta{}^2 = (2\beta)^2\sigma_\beta^2 = 4\beta^2\sigma_\beta^2 = 4\beta^2(0.003)^2$, so that $\sigma_{\beta^2} = \sqrt{4\beta^2(0.003)^2}$.

In using the simulation protocol described above, the following assumptions were made:

1. The spectra of DNA in 50% w/v sucrose and at 0°C were assumed to be entirely uninfluenced by DNA dynamics, and thus true rigid limit spectra from which to extract the rigid limit g and A tensors. Spectra of DNA attached to DEAE Sephadex in PNE closely resembled those of DNA in 50% sucrose in PNE, suggesting that the two samples were equally immobile, even though the mechanisms of restraint were different. On DEAE the DNA was attached to a matrix; in sucrose the viscosity was high. However, neither sample of DNA is truly immobile. Following Hustedt et al., the 50% sucrose, 0°C condition was the approximation chosen for the rigid limit case, to slow the spin label dynamics and avoid charge density problems, while not significantly changing the environment of the nitroxide.[5,22] To estimate the amount of residual motion of the 50-mers in 50% sucrose at 0°C, $\langle\beta^2\rangle_{theo}$ was calculated, using the weighted $\langle\eta^2\rangle$ model discussed in detail in Chapter 4. Fixing these calculated, non-zero $\langle\beta^2\rangle_{theo}$ in the simulations, the g and A tensors were optimized to estimate the "true" rigid limit g and A tensors, given in Table 3.3. Although the "standard" values for A_{xx} and A_{yy} used for all the simulations are about 5% low compared to the "true" values, A_{zz} is within 1% of the "true" value.
2. The de la Torre values for the global dynamics were assumed to be valid for this work. Tirado and de la Torre gave no results for $p < 2$, but Hustedt et al. extended their results to the 12 bp duplex case, $p = 1.7$,[5] and Eimar et al. further extended the results to the 8 bp duplex case ($p = 1.1$).[5] The shortest DNA duplex used in this study was 14 bp ($p = 2$).
3. The tilt angle θ_{tilt} of 20° determined for the T* spin label[5] was assumed for the new spin label Q, as discussed above. Subsequent 2-D NMR experiments and EPR spectral optimization of θ_{tilt} have verified this assumption to within ± 6°.[15]

The validity of assumptions 4 and 5 will be discussed in detail in Chapter 4.

4. The effect of torsional dynamics on the EPR spectra of Q-labeled DNA duplexes was assumed to be negligible and consequently was entirely neglected in the simulation.
5. The internal dynamics were assumed to be purely in the fast limit, so that $\langle\beta^2\rangle$ preaveraged

the g and A tensors.

SECTION 3.7. INITIAL EPR RESULTS

As with T*, the EPR spectra of a Q-labeled single strand, short (14-mer) duplex, and long (self-assembling concatamer) duplex are easily distinguishable.[27] See Figure 3.8.

Comparison of spectra of Q- and T*-spin-labeled DNA in 50% w/v sucrose in PNE at 0°C reveals far narrower EPR spectral lines for Q-spin-labeled DNA. A representative example is shown in Figure 3.9. The narrower lines result from highly constrained dynamics, as would arise from Q being rigidly attached to the DNA molecule. The broader lines of the T* spectrum reflect the fundamental limitation of T*, the independent motion from the nitroxide spinning about the sigma bond of the acetylene tether of T*.

The narrowness of the spectral lines of Q-labeled DNA exposed previously unseen spin-spin splitting. Attempts to increase or decrease the size of the features included repurifying the DNA, changing pH to 8.0, decreasing salt concentration, diluting the DNA with solvent (PNE buffer, 50% sucrose w/v in PNE, DEAE in PNE) or other unlabeled DNA, and varying temperatures and length. None of these attempts significantly changed the size of the features, within experimental error (overlays of spectra showed no distinguishable differences in the width of the splitting or the height of the satellite peaks). The splitting corresponded to that of ^{13}C, with satellite peak heights on the order of 8% of the main peak. Typically ^{13}C splitting results in satellite peak heights of ca. 2%, from the 2 ^{13}C atoms immediately adjacent to the nitroxide, in the α positions. The amplified ^{13}C effect here suggests coupling of the nitroxide with the ^{13}C atoms in both the α and β positions but was not explored further. These features were less pronounced in the spectra of 50-mers in 50% sucrose at 0°C and did not influence tensor selection.

SECTION 3.8. OVERVIEW OF FOLLOWING CHAPTERS

In Chapter 4, the internal dynamics of DNA as a function of length, spin label position, and temperature will be discussed. Chapter 5 will discuss the beginning of a study on the sequence dependence of DNA dynamics. In Chapter 6, the synthesis, incorporation, and characterization of a new spin label, C*, and its use in studying DNA dynamics, will be described.

NOTES TO CHAPTER 3.

1. Pecora, R. *Science* **251**, 893-897 (1991).

2. Levy, G.C., Craik, D.J., Kumar, A., & London, R.E. *Biopolymers* **22**, 2703-2726 (1983); Eimar, W., Williamson, J.R., Boxer, S.G., & Pecora, R. *Biochemistry* **29**, 799-811 (1990); Collini, M., Chirico, G., & Baldini, G. *J. Chem. Phys* **104**, 6058-6065 (1996), and others.

3. Schurr, J.M., Fujimoto, B.S., Wu, P., & Song, L. Fluorescence studies of nucleic acids, in *Topics in Fluorescence Spectroscopy, Vol 3. Biological Applications*, ed. J.R. Lakowicz, pp. 137-229. New York: Plenum Press (1992); Allison, S.A. & Schurr, J.M. *Chem. Phys.* **41**, 35-59 (1979); Wilcoxon, J. & Schurr, J.M. *Biopolymers* **22**, 2273-2321 (1983); Wu, P., Fujimoto, B.S., & Schurr, J.M. *Biopolymers* **26**, 1463-1488 (1987); Song, L., Allision, S.A., & Schurr, J.M. *Biopolymers* **29**, 1773-1791 (1990); Fujimoto, B.S. & Schurr, *Nature* **344**, 175-178 (1990); Shibata, J.H., Fujimoto, B.S., & Schurr, J.M. *Biopolymers* **24**, 1909-1930 (1985).

4. Bobst, E.V., Keyes, R.S., Cao, Y.Y., & Bobst, A.M. *Biochemistry* **35**, 9309-9313 (1996) and references therein.

5. Hustedt, E.J., Spaltenstein, A., Kirchner, J.J., Hopkins, P.B., & Robinson, B.H. *Biochemistry* **32**, 1774-1787 (1993).

6. Berliner, L.J. in *Spin Labeling, Theory and Applications*, ed. L.J. Berliner, pp. 1-4. New York: Academic Press (1976).

7. Robinson, B.H., Lerman, L.S., Beth, A.H., Frisch, H.L., Dalton, L.R., & Auer, C. *J. Mol. Biol.* **139**, 19-44 (1980).

8. Bobst, A.M., Kau, S.-C., Toppin, R.C., Ireland, J. C., & Thomas, I.E. *J. Mol. Biol.* **173**, 63-74 (1984).

9. Hustedt, E.J., Spaltenstein, A., Kirchner, J.J., Hopkins, P.B., & Robinson, B.H. *Biochemistry* **34**, 4369-4375 (1995).

10. Keyes, R.S. & Bobst, A.M. *Biochemistry* **34**, 9265-9276 (1995).

11. Robinson, B.H., Mailer, C., & Drobny, G.P. submitted to *Annual Review of Biophysical and Biomolecular Structures* (1996).

12. Spaltenstein, A., Robinson, B.H., & Hopkins, P.B. *J. Am. Chem. Soc.* **110**, 1299-1301 (1988).

13. Spaltenstein, A., Robinson, B.H., & Hopkins, P.B. *Biochemistry* **28**, 9484-9495 (1989).

14. Solomon, M.S. Ph.D. Thesis, University of Washington, 1992; Solomon, M.S.; & Hopkins, P.B. *J. Org. Chem.* **58**, 2232-2243 (1993).

15. Miller, T.R. & Hopkins, P.B. *Bioorg. Med. Chem. Lett.* **4**, 981-986 (1994); Miller, T.R. Ph.D. Thesis, University of Washington (1995).

16. Miller, T.R., Alley, S.C., Reese, A.W., Solomon, M.S., McCallister, W.V., Mailer, C., Robinson, B.H., & Hopkins, P.B. *J. Am. Chem. Soc.* **117**, 9377-9378 (1995).

17. Alley, S.C. Ph.D. Thesis, University of Washington (1996). After detritylation of the residue 3' to the intended position for Q, the automated cycle was interrupted. The controlled pore glass

(CPG) column was removed from the synthesizer and attached to two syringes, one filled with 100 ul of 0.1 M tetrazole in anhydrous CH_3CN and the other with 100 ul of 130 mM solution of phosphoramidite, also in anhydrous CH_3CN. The solutions were passed back and forth approximately 100 times through the column for 2.5 minutes. The liquid was discarded and the column returned to the synthesizer to finish DNA synthesis under full automation. DNA was cleaved from the resin and deprotoected in concentrated aqueous ammonia at 55 C for 16 h and then dried in a Savant Speed Vac concentrator.

18. Huang, H., Miller, T.R., Duncan, J.A., Alley, S.C., Zhu, L., Reid, B.R., & Hopkins, P.B., in preparation.

19. Robinson, B.H. & Drobny, G.P. *Annual Review of Biophysical and Biomolecular Structures* **24**, 523-549 (1995).

20. VitroCom, Inc., Mountain Lakes NJ, clear fused quartz round capillary tubing, 0.6 mm ID x 0.84 mm OD x 100 mm long.

21. BDH Chemical supplies, gift from Al Beth.

22. Hustedt, E.J. Ph.D. Thesis, University of Washington, 1989:

$$\frac{2\ OD * (50\ \mu g / OD)}{10\ \mu l} = \frac{100\ \mu g}{10\ \mu l} = 10\ \mu g / \mu l = 10\ mg / ml$$

$$\frac{2\ OD * (50\ \mu g / OD)}{(330\ (\mu g / \mu mol) / bp * 50bp)} = 0.006\ \mu mol \qquad \frac{0.006\ \mu mol}{10\ \mu l} = 0.0006\ \mu M = 0.6\ mM\ DNA$$

23. Bruker Instruments, Billerica MA.

24. Tirado, M.M. & de la Torre, J.G. *J. Chem. Phys.* **73**, 1986-1993 (1980).

25. Barber, E.J. *National Cancer Inst. Monogr.* **21**, 219-239 (1966).

26. Nordio, P.L. in *Spin Labeling, Theory and Applications*, ed. L.J. Berliner, pp. 5-52. New York: Academic Press (1976).

27. Spaltenstein, A., Robinson, B.H., & Hopkins, P.B. *J. Am. Chem. Soc.* **111**, 2303-2305 (1989).

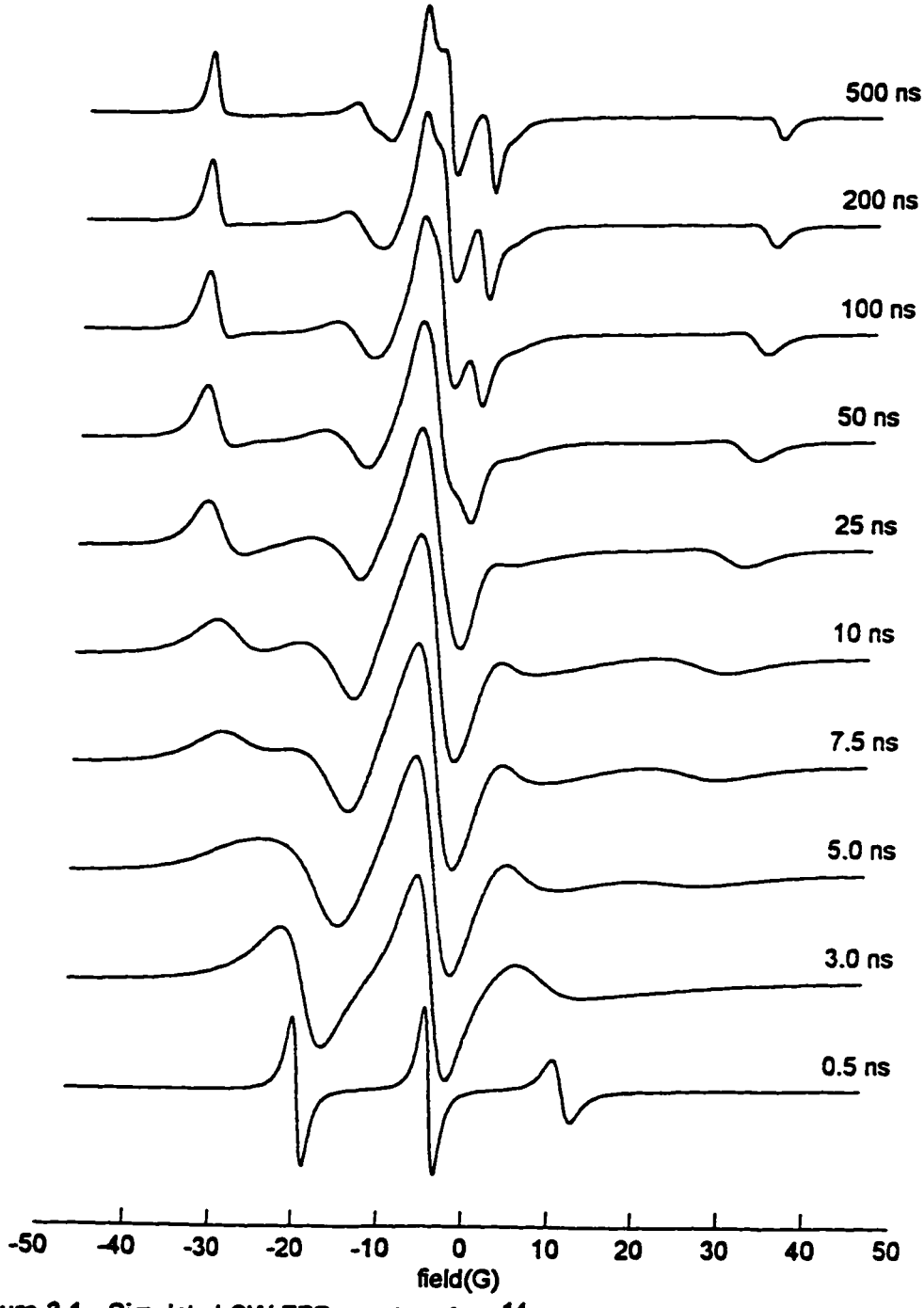


Figure 3.1. Simulated CW-EPR spectra of an ^{14}N nitroxide assuming the spin label undergoes isotropic rotational diffusion. The isotropic correlation times, τ_{rot}, are given on the right side of each simulated spectrum. Parameters used in the simulation are g_{xx} = 2.0079, g_{yy} = 2.0060, g_{zz} = 2.0026; A_{xx} = 6.18 G, A_{yy} = 6.05 G, A_{zz} = 34.23 G; θ = 20°, r = 0.6 G, σ = 0 G. The values of τ_{rot} are 500 ns, 200 ns, 100 ns, 50 ns, 25 ns, 10 ns, 7.5 ns, 5.0 ns, 3.0 ns, and 0.5 ns from top to bottom. Note that the lineshapes broaden and spectral width increases as τ increases.

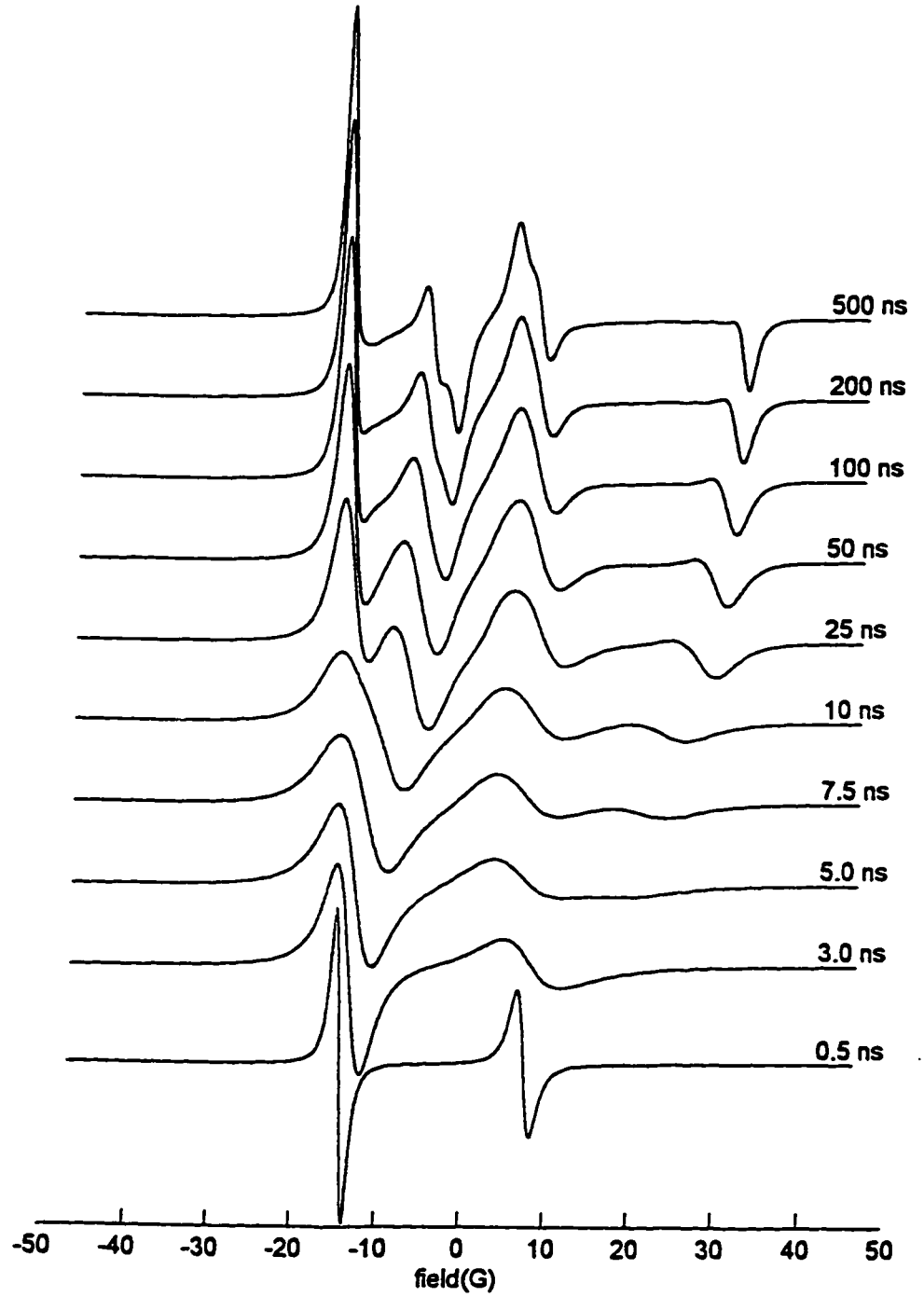


Figure 3.2. Simulated CW-EPR spectra of an ^{15}N nitroxide assuming the spin label undergoes isotropic rotational diffusion. The isotropic correlation times, τ_{rot}, are given on the right side of each simulated spectrum. Parameters used in the simulation are g_{xx} = 2.0079, g_{yy} = 2.0060, g_{zz} = 2.0026; A_{xx} = 6.18 G, A_{yy} = 6.05 G, A_{zz} = 34.23 G; θ = 20°, r = 0.6 G, σ = 0 G. The values of τ_{rot} are 500 ns, 200 ns, 100 ns, 50 ns, 25 ns, 10 ns, 7.5 ns, 5.0 ns, 3.0 ns, and 0.5 ns from top to bottom. Note that the lineshapes broaden and spectral width increases as τ increases.

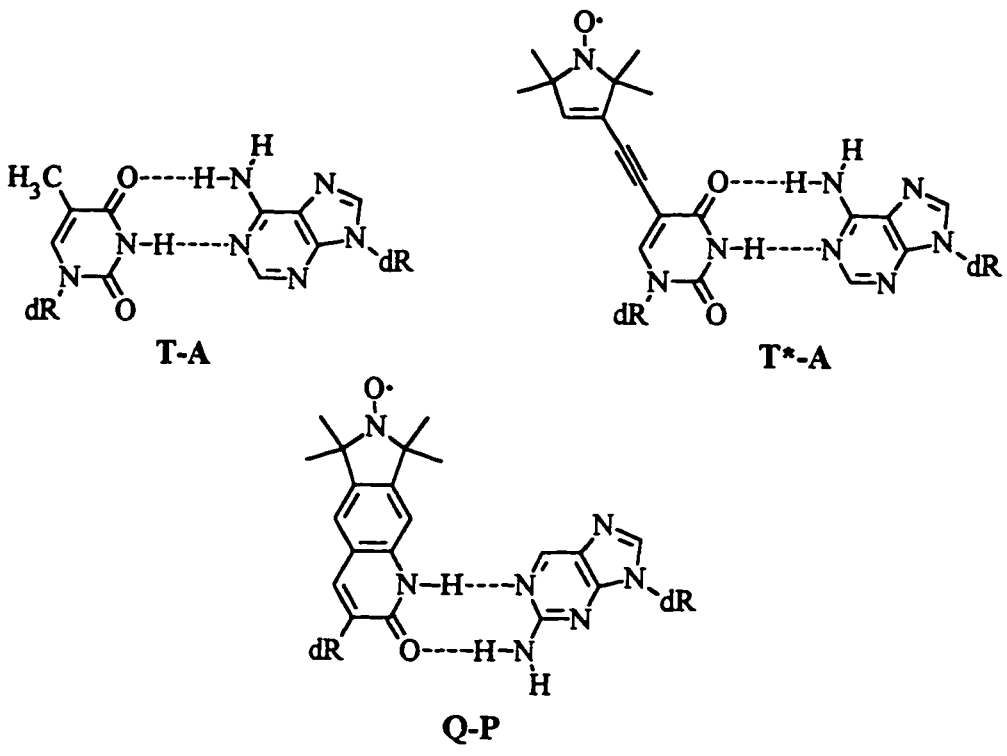


Figure 3.3. Structures of basepairs AT, AT*, and PQ.

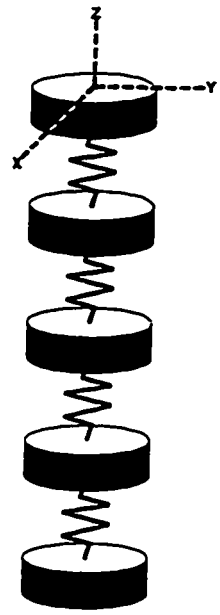


Figure 3.4. Schematic representation of the DNA molecule with the x, y, and z axes. Figure by Jerry King / 3 Kings Design.

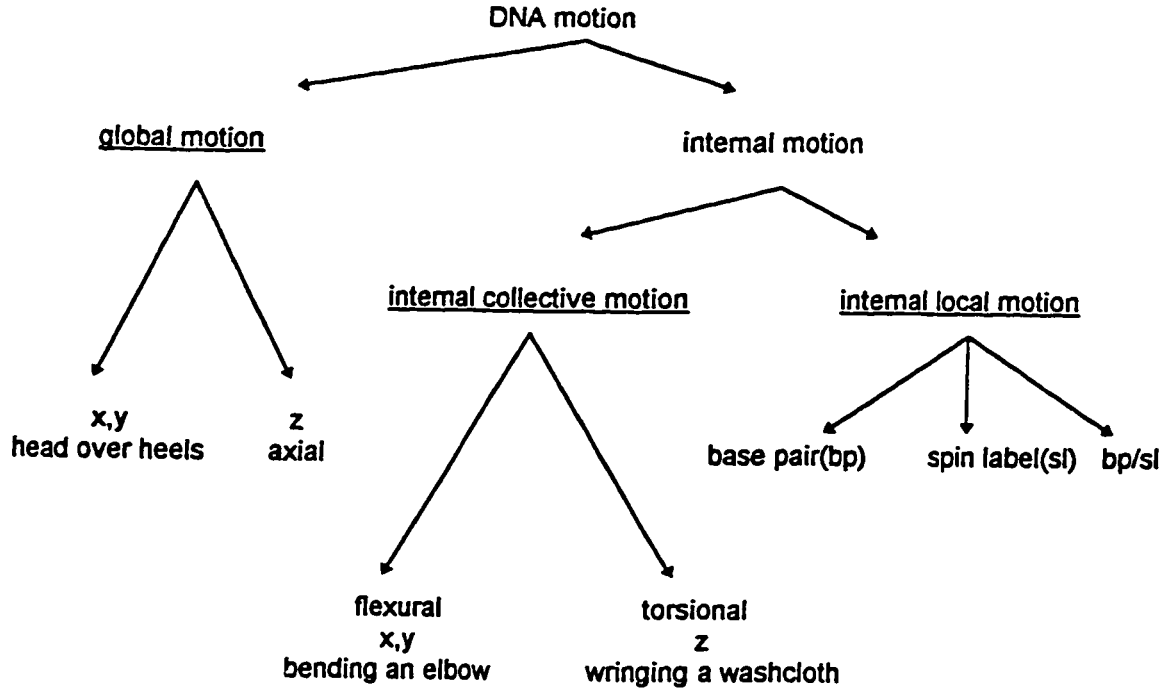


Figure 3.5. Schematic representation of the categories of DNA dynamics.

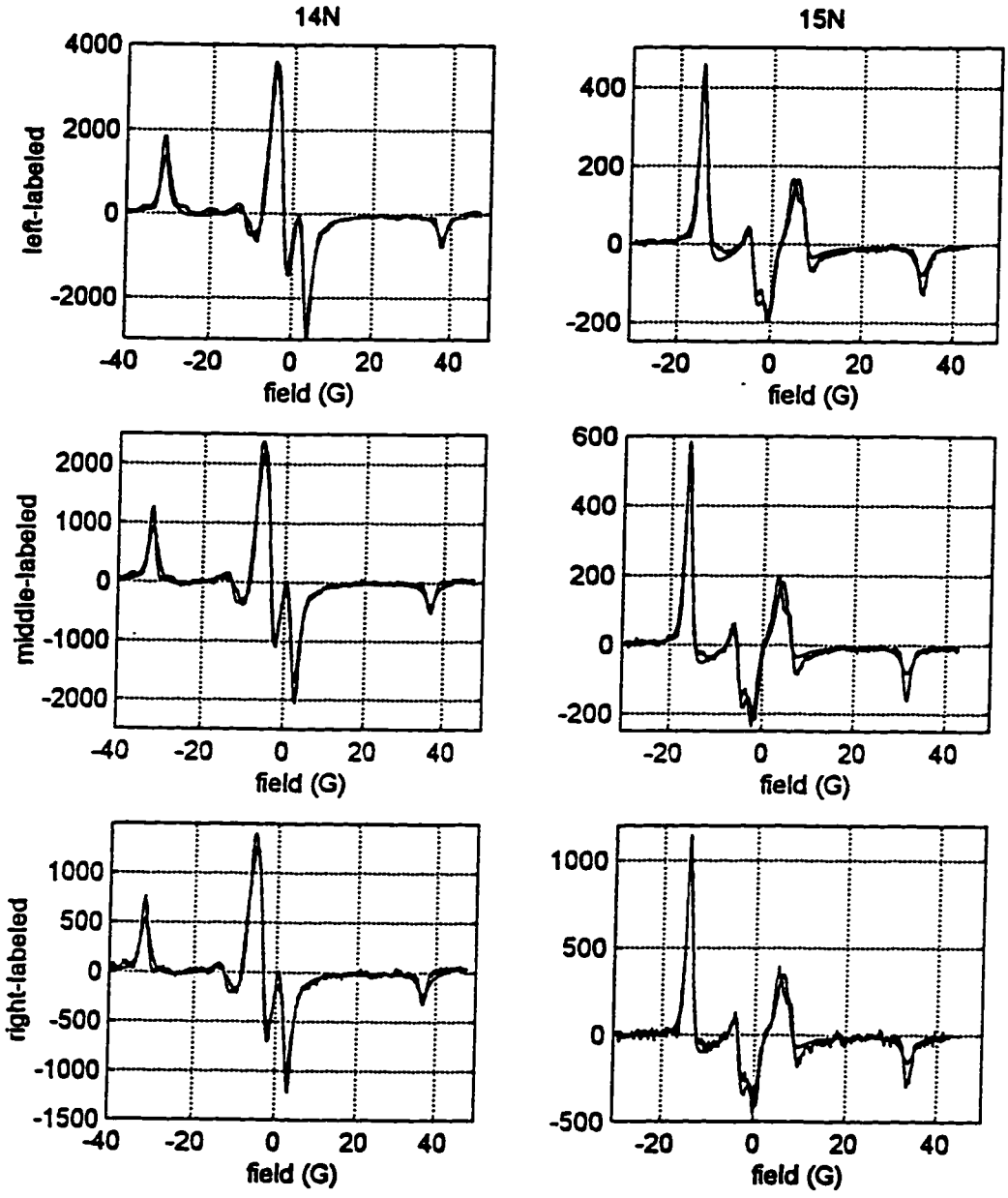


Figure 3.6. CW-EPR spectra for the ^{14}N and ^{15}N 50-mers, labeled at positions 13, 26, and 39, in 50% w/v sucrose at 0°C, with least squares best fit simulations overlaid. Specific sequences were (left) 5'-GGT GCG ACC TCG NAT CGT GCT CCT CAT CTT CGT GTC CTA ATC TAT CGT AT, (middle) 5'-TAA TCT ATC GTA TGG TGC GAC CTC GNA TCG TGC TCC TCA TCT TCG TGT CC, (right) 5'-CAT CTT CGT GTC CTA ATC TAT CGT ATG GTG CGA CCT CGN ATC GTG CTC CT. Simulations held σ = 0.5 G fixed and varied the Lorentzian linewidth (r) while optimizing the g and A tensors. A single set of tensors was determined for these spectra; values are given in Table 3.1. Lorentzian linewidth values are (^{14}N center line; l) 1.26 G; (^{14}N center line; m) 1.22 G; (^{14}N center line; r) 1.22 G; (^{15}N both lines; l) 0.75 G, 1.47 G; (^{15}N both lines; m) 0.54 G, 1.38 G; (^{15}N both lines; r) 0.50 G, 1.31 G. R^2, the correlation coefficients, were (^{14}N; l,m,r) 0.9679, 0.9595, 0.9534; (^{15}N; l,m,r), 0.9581, 0.9420, 0.9412.

Table 3.1. Least squares best fit tensors determined from spectra of 50-mers in 50% w/v sucrose in PNE at 0°C.

	xx	yy	zz
g	2.0084	2.0068	2.0034
A (^{14}N)	6.58	4.98	34.23
A (^{15}N)	9.22	6.98	48.00

Table 3.2. Rotational correlation times (in nanoseconds) for DNA molecules of radius R = 12.0Å and length 3.4*N Å, where N is the number of base pairs, as a function of temperature (T,°C) and % sucrose (viscosities in centipoise calculated per Barber). Correlation times are calculated with $\tau_a = \frac{1}{6D_a}$, where a = perp or para, and D_{perp} and D_{para} are calculated per Tirado and de la Torre.

0% w/v sucrose

	0°C		10°C		20°C		30°C		40°C	
N	τ_{perp}	τ_{para}	τ_{perp}	τ_{para}	τ_{perp}	τ_{para}	τ_{perp}	τ_{para}	τ_{perp}	τ_{para}
11	12.32	6.95	8.73	4.92	6.47	3.65	4.97	2.81	3.94	2.22
12	14.32	7.45	10.14	5.28	7.52	3.91	5.78	3.01	4.58	2.38
14	18.82	8.44	13.34	5.98	9.89	4.43	7.60	3.41	6.02	2.70
15	21.35	8.92	15.13	6.32	11.22	4.69	8.62	3.60	6.83	2.85
20	37.21	11.33	26.37	8.03	19.55	5.95	15.03	4.57	11.90	3.62
22	45.18	12.28	32.02	8.70	23.74	6.45	18.24	4.96	14.45	3.93
28	75.37	15.11	53.41	10.71	39.59	7.94	30.43	6.10	24.10	4.83
30	87.68	16.05	62.13	11.38	46.06	8.43	35.41	6.48	28.04	5.13
50	284.4	25.42	201.5	18.01	149.4	13.35	114.8	10.26	90.92	8.13
60	441.4	30.08	312.8	21.32	231.9	15.80	178.3	12.15	141.2	9.62
100	1578	48.7	1118	34.5	828.8	25.59	637.1	19.67	504.4	15.58
η	1.78		1.31		1.00		0.80		0.65	

50% w/v sucrose

	0°C		10°C		20°C		30°C		40°C	
N	τ_{perp}	τ_{para}	τ_{perp}	τ_{para}	τ_{perp}	τ_{para}	τ_{perp}	τ_{para}	τ_{perp}	τ_{para}
11	309.6	174.7	168.1	94.80	99.44	56.09	63.02	35.55	42.22	23.82
12	359.8	187.3	195.3	101.6	115.6	60.13	73.24	38.11	49.07	25.53
14	473.0	212.0	256.7	115.1	151.9	68.09	96.27	43.16	64.50	28.91
15	536.7	224.3	291.3	121.7	172.4	72.03	109.2	45.65	73.18	30.58
20	935.3	284.7	507.6	154.5	300.4	91.43	190.4	57.94	127.5	38.82
22	1136	308.6	616.4	167.5	364.7	99.11	231.2	62.81	154.9	42.08
28	1894	379.9	1028	206.2	608.4	122.0	385.6	77.32	258.3	51.80
30	2204	403.5	1196	219.0	707.8	129.6	448.6	82.13	300.5	55.03
50	7148	638.9	3879	346.7	2295	205.2	1455	130.0	974.7	87.12
60	11096	756	6022	410.4	3563	242.8	2258	153.9	1513	103.1
100	39655	1224	21522	665	12735	393	8071	249.2	5408	167.0
η	44.7		25.2		15.4		10.1		7.0	

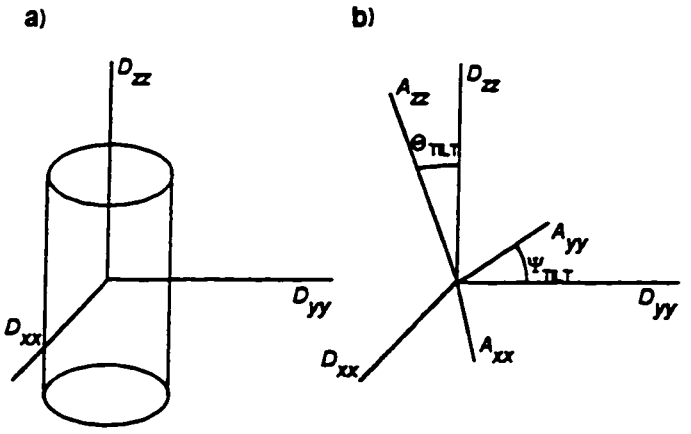


Figure 3.7. Orientation of DNA and spin label coordinate systems. a) Orientation of the rotational diffusion tensor ($D_{perp} = D_{xx} = D_{yy}$; $D_{para} = D_{zz}$) with respect to the DNA cylinder. b) Orientation of the spin label coordinate system with respect to the rotational diffusion tensor, as defined by angles θ_{tilt} and ψ_{tilt}.

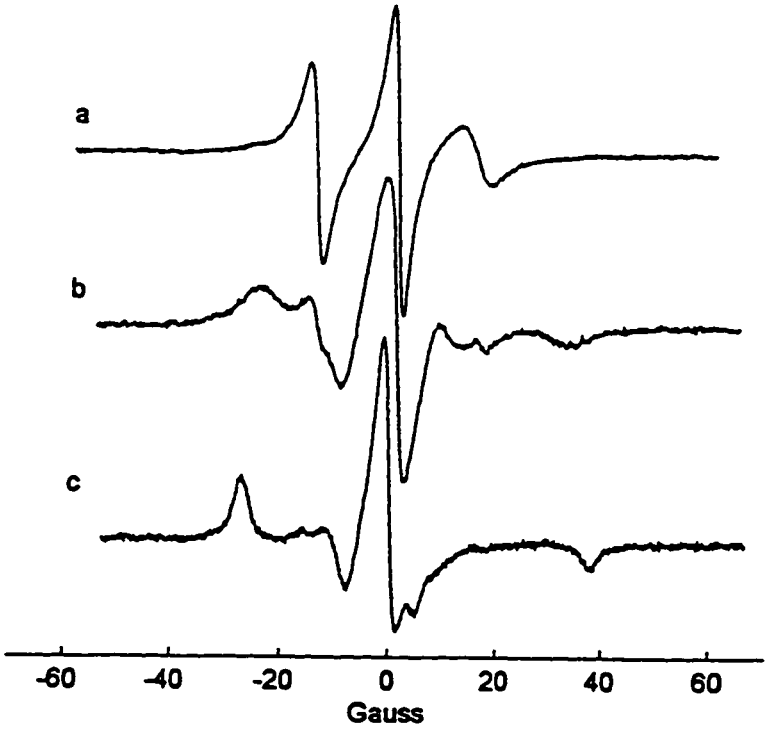


Figure 3.8. CW-EPR spectra of a) a single strand 5'-d(GCT TAA GCT QCG CG), b) a 14-mer duplex 5'-d(GCC TAC CTG QGA CG)·5'-d(CG TCP CAT GTA GGC), and c) a self-assembling concatemer 5'-d(GTG PCA TGQ CAC GCG CGC), in PNE at 20°C.

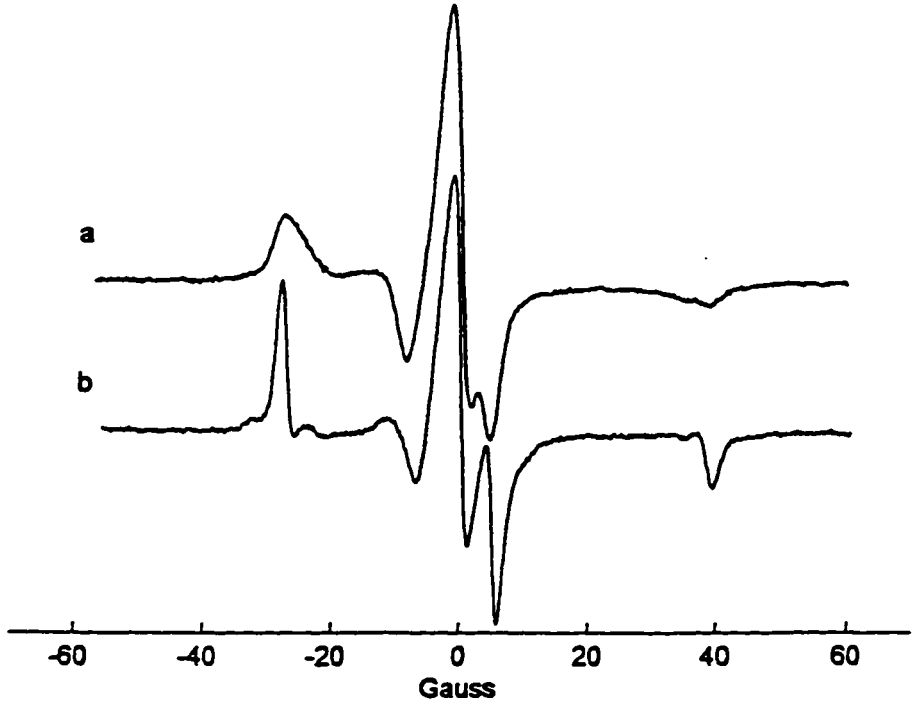


Figure 3.9. CW-EPR spectra of a) 5'-d(GCC TAC ATG T*GA CG)·5'-d(CG TCA CAT GTA GGC) and b) 5'-d(GCC TAC ATG QGA CG)·5'-d(CG TCP CAT GTA GGC) in 50% w/v sucrose in PNE at 0°C.

Table 3.3. Estimates of the "true" rigid limit g and A tensors, calculated by including the theoretical $\langle\beta^2\rangle$ = 0.0164, 0.0110, and 0.0169 representing internal dynamics remaining at T = 0°C for the left- (l), middle- (m), and right- (r) labeled 50-mers in 50% w/v sucrose. Simulations held σ = 0.5 G fixed and varied the Lorentzian linewidth (r) while optimizing the g and A tensors. Lorentzian linewidth values (for standard tensor set) are (^{14}N center line; l) 1.26 G; (^{14}N center line; m) 1.22 G; (^{14}N center line; r) 1.22 G; (^{15}N low field line; l) 0.75 G; (^{15}N low field line; m) 0.54 G; (^{15}N low field line; r) 0.50 G. Lorentzian linewidth values (for "true" tensor set) are (^{14}N center line; l) 1.19 G; (^{14}N center line; m) 1.15 G; (^{14}N center line; r) 1.16 G; (^{15}N low field line; l) 0.73 G; (^{15}N low field line; m) 0.50 G; (^{15}N low field line; r) 0.46 G. R^2 is the correlation coefficient between data and simulation.

	g	A, ^{14}N	A, ^{15}N	$\frac{A(true)}{A(standard)}$	$\langle\beta^2\rangle$ (fixed)	R^2, ^{14}N	R^2, ^{15}N	
standard	2.0084	6.58	9.22		0	0.9679	0.9581	l
	2.0068	4.98	6.98		0	0.9595	0.9420	m
	2.0034	34.23	48.00		0	0.9534	0.9412	r
true	2.0087	6.33	8.87	0.96	0.0164	0.9476	0.9430	l
	2.0067	4.72	6.62	0.95	0.0110	0.9439	0.9265	m
	2.0033	34.74	48.71	1.01	0.0169	0.9368	0.9296	r

CHAPTER 4: LENGTH AND PROBE POSITION DEPENDENCE OF DNA DYNAMICS

SECTION 4.1. INTRODUCTION

The weakly bending rod model of DNA dynamics predicts that the internal motion of DNA depends upon the length of the DNA molecule.[1] In particular, the model predicts that the amplitude of the internal collective dynamics of DNA should be greater at the ends than at the middle of the DNA molecule. Furthermore, as the DNA molecule length increases, the amplitude of internal collective dynamics should also increase.[2] In the work described here, the accuracy of these predictions is confirmed. The dynamic flexural persistence length, P_{df}, is estimated and the temperature dependence of P_{df} is examined.

SECTION 4.2. EXPERIMENTS

Four synthetic DNA duplexes based on a single sequence were designed to constitute a "self-incorporating" series of 14, 28, 50, and 100 base pairs, such that the 28-mer contained the 14-mer in its center, the 50-mer contained the 28-mer in its center, and the 100-mer contained the 50-mer in its center, as illustrated in Figure 4.1. This series carried the spin label Q in the middle position and is explicitly listed in Table 4.1.

Each middle-labeled length sequence was cyclically permuted to form the left- (1/4 from the left, 5', end) and right-labeled (1/4 from the right, 3', end) cases, illustrated in Figure 4.1.[3] Location of the spin label in each position and length sequence is given in Table 4.2. The cyclic permutation maintained the immediately local sequence as well as overall base composition while investigating different spin label positions. However, because the left- and right-labeled sequences were cyclic permutations of the middle-labeled sequence, the left- and right-labeled sequences were not fully consistent with the "self-incorporating" model mentioned above. Had the left- and right-labeled sequences been made fully consistent with the "self-incorporating" model, then the overall base composition within each length set would have varied for the left-, middle-, and right-labeled samples. For example, the left-, middle-, and right-labeled 14-mers would have had different overall base composition.

The sequence 5'd(TAC GAA CCT GAG CTA TTT CCC GGC TTA ATC TAT CGT ATG GTG CGA CCT CGQ ATC GTG CTC CTC ATC TTC GTG TCC GGA AAG GTG GTG CTC ACT GGA GCA A) with its (P-containing) complement was chosen as the basic template. The sequence came from the Drosophila Melanogaster TATA-box binding protein TFIID gene, positions 1062 to 1161, where the A at position 1112 was replaced with Q, the spin label; Q was paired against P, 2-Aminopurine. The sequence was chosen to have approximately a 1:1 ratio of GC and AT base pairs, and also to avoid known oddities such as runs of a single base. The majority of the DNA samples used the [^{15}N,D] isotope of Q; some DNAs used the [^{14}N,H] isotope of Q, and all were pooled for the analysis.

In addition, one other permutation of the basic template sequence was used, with the

[^{15}N,D] spin label in position 6 of a 50-mer, approximately 10% of the way from the left, 5', end. The sequence was 5'd(CCT CGQ ATC GTG CTC CTC ATC TTC GTG TCC GGA AAG GTG GTG CTC ACT GG), with its complement.

CW-EPR experiments using the PDP homebuilt spectrometer described earlier were performed on the DNA samples at T = 0°, 10°, 20°, 30°, and 40°C, in PNE buffer (10 mM phosphate buffer (pH 7.0), 115 mM Na^+, 0.1 mM EDTA) and also in 50% w/v sucrose in PNE. Experimental conditions were given in Chapter 3. The 50-mer sample with Q in position 6 was studied only at T = 30°C.

As described in Chapter 3, tensors were extracted from the spectra of the six 50-mers (left-, middle-, and right-labeled, ^{14}N and ^{15}N) in 50% w/v sucrose in PNE buffer at 0°C. (Overlays of simulation on data are shown in Figure 3.6.) These tensors (listed in Table 3.1) were then used in the simulations of all the length- and position-dependence study DNA molecules (0% sucrose, T = 0°, 10°, 20°, 30°, and 40°C). The global dynamics used for these simulations are given in Table 3.2. $\langle\beta^2\rangle$, the single parameter that measures the narrowing of the spectrum representing the internal dynamics of the DNA molecule, is listed in Table 4.3.

As with the T* spin-labeled DNA study, for the short 14-mer duplexes in buffer at temperatures T ≥ 10°C, the global dynamics collapsed the lineshape such that the spectral width $2A_{zz}$ could not be measured. The simulations for these spectra were relatively insensitive to the value of $\langle\beta^2\rangle$. Consequently these values were not used in the analysis of $\langle\beta^2\rangle$ as a function of length and position.[4]

SECTION 4.3. RESULTS AND DISCUSSION

A. ANALYSIS OF $\langle\beta^2\rangle$ AS A FUNCTION OF LENGTH AND POSITION

Following Hustedt et al.,[4] we consider the total rms amplitude of internal motion (measured), $\langle\beta^2\rangle_{total}$, to be the sum of the internal collective motion, $\langle\eta^2\rangle_{total}$, and the internal independent motion, $\langle\beta_0{}^2\rangle_{total}$:

$$\langle\beta^2\rangle_{total} = \langle\eta^2\rangle_{total} + \langle\beta_0{}^2\rangle_{total} \tag{1}$$

We further take all the internal collective dynamics to be internal collective bending dynamics, neglecting torsional dynamics. To determine $\langle\eta^2\rangle_{total}$, the internal collective bending motion for DNA with N base pairs, Song, Allison, and Schurr examined the potential bending energy, assuming nearest-neighbor interactions:[5]

$$U = \frac{1}{2}\kappa \sum_{i\geq 1}^{N-1} \left((\eta_{i+1} - \eta_i)^2 + (\xi_{0i+1} - \xi_{0i})^2 \right), \tag{2}$$

where κ is the bending force constant, η represents bending in the yz plane, and ξ represents

bending in the xz plane. Because motions in the xz and yz planes are uncoupled and effectively the same, only one set of dynamics need to be considered in detail, for example the yz motions:

$$
\begin{aligned}
U' &= \tfrac{1}{2}\kappa\Sigma\left((\eta_{i+1}-\eta_i)^2\right) \\
&= \tfrac{1}{2}\kappa\left((\eta_2-\eta_1)^2+(\eta_3-\eta_2)^2+(\eta_4-\eta_3)^2+\ldots\right) \\
&= \tfrac{1}{2}\kappa\left(\eta_2^2-2\eta_1\eta_2+\eta_1^2+\eta_3^2-2\eta_2\eta_3+\eta_2^2+\eta_4^2-2\eta_3\eta_4+\eta_3^2+\ldots\right) \\
&= \tfrac{1}{2}\kappa\left(\eta_1^2-2\eta_1\eta_2+\eta_3^2+(\eta_2^2+\eta_2^2)-2\eta_2\eta_3+(\eta_3^2+\eta_3^2)-2\eta_3\eta_4+\eta_4^2+\ldots\right) \\
&= \tfrac{1}{2}\kappa\left(\eta_1^2-\eta_1\eta_2-\eta_2\eta_1+2\eta_2^2-\eta_2\eta_3-\eta_3\eta_2+2\eta_3^2-\eta_3\eta_4-\eta_4\eta_3+\ldots\right) \\
&= \tfrac{1}{2}\kappa\hat{\eta}^T\boldsymbol{A}\hat{\eta}
\end{aligned}
\tag{3}
$$

where $\hat{\eta} = \begin{pmatrix}\eta_1\\ \eta_2\\ \vdots\\ \eta_N\end{pmatrix}$ and $\boldsymbol{A} = \begin{vmatrix} 1 & -1 & 0 & \cdots & & 0 \\ 1 & 2 & -1 & 0 & & 0 \\ 0 & -1 & 2 & -1 & & 0 \\ \vdots & & & & & \vdots \\ 0 & \cdots & & -1 & 2 & -1 \\ 0 & \cdots & & & -1 & 1 \end{vmatrix}$.

Diagonalizing the matrix $\boldsymbol{A}$ yields the eigenvalues Λ and the matrix of eigenvectors $\boldsymbol{Q}$: $\boldsymbol{Q}\cdot\Lambda\cdot\boldsymbol{Q}^T = \boldsymbol{A}$. Wu et al. showed that the internal collective bending dynamics of DNA, $\langle\eta^2\rangle_{total}$, can be calculated using Λ and $\boldsymbol{Q}$:[7]

$$
\langle\eta^2\rangle_{total} = 2\langle\eta_i^2\rangle = 2\sum_{i\geq 2}^{N} Q_{il}^2 d_l^2 , \tag{4}
$$

where $d_l^2 = \left(\frac{h}{P_{df}\Lambda_l}\right)$, $\Lambda_l = 4\sin^2\left(\frac{(l-1)\pi}{2N}\right)$, and $Q_{il}^2 = \frac{2}{N}\cos^2\left(\frac{(i-\frac{1}{2})(l-1)\pi}{N}\right)$. d_l^2 represents the mean-squared amplitude of the l^{th} bending mode and Q_{il}^2 shows the contribution of the l^{th} mode to the motion of the i^{th} base pair. These modes of internal bending motion are analogous to the modes of a violin string. The l = 0 and l = 1 modes represent global translation and rotation, respectively, and are not included in the calculation of $\langle\eta^2\rangle$. The l = 2 mode looks like a horseshoe, and the l = 3 mode looks sinusoidal. The motion as measured at position i then is the sum of the contributions of the l = 2 mode, l = 3 mode, and so on. The analytical solution[7] to equation 4 is

$$
\langle\eta^2\rangle_{total} = 2\langle\eta^2\rangle = 2\Sigma Q_{il}^2 d_l^2 = \left(\frac{h}{6P_{df}}\right)\times N\left(1+3\left(\frac{2i-N-1}{N}\right)^2\right) = \left(\frac{h}{6P_{df}}\right)\times f(N,i) \tag{5}
$$

where h = 3.4 Å is the rise per base pair, P_{df} is the dynamic flexural persistence length (a measure of stiffness: larger P_{df} implies a stiffer piece of DNA), i is the position under study, and $f(N,i) = N\left(1+3\left(\frac{2i-N-1}{N}\right)^2\right)$.

In addition, Wilcoxon and Schurr suggested that P_{df} follows one of two models: either the bending angle between two adjacent base pairs obeys a square-well potential, so P_{df} is independent of temperature T, or, the bending angle obeys a harmonic potential, so P_{df} is inversely proportional to T:[6]

$$P_{df}(T) = P_{df}(T = 293K) \times f(T)$$

(6)
$$f(T) = 1 \quad \text{for square well}$$
$$f(T) = \left(\frac{293K}{T}\right) \quad \text{for harmonic well}$$

Furthermore, the internal independent motion $\langle\beta_0^2\rangle$ is independent of length but does depend on temperature. If, following Hustedt et al., [4] a power law dependence on T is assumed,

(7)
$$\langle\beta_0^2\rangle = \langle\beta_0^2(T = 293K)\rangle \times \left(\frac{T}{293K}\right)^q$$

Consequently,

$$\begin{aligned}\langle\beta^2\rangle_{total} &= \langle\eta^2\rangle_{total} + \langle\beta_0^2\rangle_{total} \\ &= \left(\frac{h}{6P_{df}}\right)\left(\frac{1}{f(T)}\right) \times N\left(1+3\left(\frac{2i-N-1}{N}\right)^2\right) + \langle\beta_0^2(T = 293K)\rangle \times \left(\frac{T}{293K}\right)^q \\ &= \left(\frac{h}{6P_{df}}\right)\left(\frac{1}{f(T)}\right) \times f(N,i) + \langle\beta_0^2(T = 293K)\rangle \times \left(\frac{T}{293K}\right)^q\end{aligned} \quad (8)$$

where $f(N,i) = N\left(1+3\left(\frac{2i-N-1}{N}\right)^2\right)$.

This is of the form y = mx + b, so a plot of $\langle\beta^2\rangle_{total}$ vs. f(N,i) yields the slope, $m = \left(\frac{h}{6P_{df}}\right)\left(\frac{1}{f(T)}\right)$ and the intercept, $b = \langle\beta_0^2(T = 293K)\rangle \times \left(\frac{T}{293K}\right)^q$. Data from the left-, middle-, and right-labeled DNA sequences should then all lie on one line, the persistence length P_{df} may be calculated from the slope, and the length-independent $\langle\beta_0^2\rangle$ is calculable from the intercept. If $\langle\beta^2\rangle_{total}$ is plotted against N instead, then the slope is $m = \left(\frac{h}{6P_{df}}\right)\left(\frac{1}{f(T)}\right) \times N\left(1+3\left(\frac{2i-N-1}{N}\right)^2\right)$. The left-

and right-labeled DNA sequences correspond to cases of i = N/4 and i = 3N/4; the middle-labeled case is of i = N/2. Consequently, the slopes of $\langle\beta^2\rangle_{total}$ vs. N for the end-labeled DNA should be roughly twice as steep as the slope of $\langle\beta^2\rangle_{total}$ vs. N for the middle-labeled DNA.

$\langle\beta^2\rangle$ from all the Q-spin-labeled DNAs was analyzed as a function of length and position using this analysis, with q = 7.1, per Hustedt et al..[4] Results are given in Table 4.4. As shown in Figure 4.2, the calculated $\langle\eta^2\rangle$ + $\langle\beta_0^2\rangle$ line does not pass through much of the data. An additional attempt was made to fit $\langle\beta^2\rangle$ as a function of length and position without requiring the power-law dependence (equation 7), so that $\langle\beta_0^2\rangle$ was not a simple function of temperature. Although at T=0°C the fit of (calculated) $\langle\eta^2\rangle$ + $\langle\beta_0^2\rangle$ to (experimental) data $\langle\beta^2\rangle$ is acceptable (Figure 4.3b), the left- and right-labeled DNA data roll over by N = 100, more clearly shown when $\langle\beta^2\rangle$ is plotted against N, as in Figure 4.3a. The steady increase in $\langle\beta^2\rangle$ as N increases (seen for N = 14, 28, and 50 bp) diminishes at N = 100 for the left- and right-labeled cases, but not for the middle-labeled case. Furthermore, the middle-labeled DNA data falls along a much steeper line than that determined from the analysis of the left-, middle-, and right-labeled DNA spectra at T = 0, 10, 20, 30, and 40°C. A line twice as steep would pass through the end-labeled DNA $\langle\beta^2\rangle$ IF a rollover at large N is permitted for the end-labeled cases.

This puzzle may be explained by considering the time constants of the modes of internal dynamics, that is, how fast these modes move. The time constants, τ_l, are inversely proportional to the Λ_l, which are used in the calculation of $\langle\eta^2\rangle$.[5,7] Furthermore, the τ_l have been shown to increase with increasing length of DNA sequences, as shown in Figure 4.4.[5] In the analysis of Hustedt et al., the internal dynamics of spin-labeled DNA were treated as being entirely in the fast motion limit: rapid internal dynamics resulted in a motionally averaged, effective set of A and g tensors, and the time constants of all the modes were used to calculate $\langle\eta_i^2\rangle$. However, if the time constants were not in the fast motion limit, then including the time constants of all the modes in the motional averaging of the tensors is no longer completely correct.

To reduce the contribution of the longer modes to the calculated $\langle\eta_i^2\rangle$, a simple weighting scheme was used. This weighting was of the form $\langle\eta_i^2\rangle_{weighted} = 2 \sum_{l=modes} Q_{il}^2 d_l^2 w_{i,l}$, where $w_{i,l} = e^{-(2m\tau_l \Delta A_{i,l})}$, m is a constant weighting premultiplier, τ_l is the time constant for mode l, and $\Delta A_{i,l}$ reflects how much the A tensor is narrowed at position i because of the l^{th}

motional mode. The time constants τ_l were calculated from the eigenvalues Λ_l of the matrix ***A***, $\tau_l = \frac{\gamma}{\kappa\Lambda_l}$, where $\gamma = 3.03\times10^{-21}$ erg s is the friction factor, κ is the force constant, calculated from the relationship[5] $\frac{P_{df}}{h} = \frac{\kappa}{k_b T}$, and Λ_l is the l^{th} eigenvalue described above. This differs from the protocol of Song et al.[5] but was considered a reasonable first approximation. $\Delta A_{i,l}$ was defined from Griffith and Jost as follows:[8]

$$
\begin{aligned}
A_{i,l} &= A_{zz}\cos^2\theta + A_{perp}\sin^2\theta \\
\Delta A_{i,l} &= A(\Delta) - A(0) \\
&= A_{zz}\cos^2\Delta + A_{perp}\sin^2\Delta - A_{zz}\cos^2 0 - A_{perp}\sin^2 0 \\
&= A_{zz}\cos^2\Delta + A_{perp}\sin^2\Delta - A_{zz} \\
&= A_{zz}(1-\cos^2\Delta) + A_{perp}\sin^2\Delta \\
&= A_{zz}(\sin^2\Delta) + A_{perp}\sin^2\Delta \\
&= (A_{zz} + A_{perp})\sin^2\Delta \\
&= (A_{zz} + A_{perp})\sin^2(2Q_{il}^2 d_l^2)
\end{aligned}
\qquad (9)
$$

When $(-m2\tau_l\Delta A_{i,l}) \approx 0$, for example if τ_l is small, representing fast motion, then $e^{-(2m\tau_l\Delta A_{i,l})} \approx 1$, and the l^{th} mode fully contributes to the calculated $\langle\eta_i^2\rangle$. When $(-m2\tau_l\Delta A_{i,l})$ is large, for example if τ_l is large, representing slower motion, then $e^{-(2m\tau_l\Delta A_{i,l})}$ becomes small, and the l^{th} motional mode's contribution is diminished. The weighting transforms the $\langle\beta^2\rangle$ vs. f(N,i) curve from a straight line at all lengths to a straight line that rolls over at large N, shown in Figure 4.5.

In addition, because the center merely translates in the l = 2 (horseshoe) mode of motion, the l = 2 mode does not contribute to the internal collective rotational motion measured at the center of the DNA sequence. In contrast, the l = 2 mode contributes greatly to the internal rotational motion measured at the ends of the sequence. In the (sinusoidal) l = 3 mode, the center point rotates dramatically. This mode makes the greatest contribution to the internal motion observed at the center of the DNA sequence. Consequently, if 10 ns is taken to represent "slow" dynamics, $\langle\beta^2\rangle \propto \langle\eta^2\rangle$ should be linear as a function of length until the l = 3 mode begins to contribute for the middle-labeled DNAs (ca. 70 base pairs from Figure 4.4) and until the l = 2 mode begins to contribute for the end-labeled DNAs (ca. 110 base pairs). In Figure 4.5, the curved ($\langle\eta^2\rangle_{weighted}$) lines deviate noticeably from the straight ($\langle\eta^2\rangle = \langle\eta^2\rangle_{not\ weighted}$) lines at approximately 70 and 110 base pairs as well, for the end- and middle-labeled cases,

respectively.

The difference between $\langle \eta_i^2 \rangle$ (calculated) and $\langle \beta_i^2 \rangle$ (experimental data) was minimized. The weighting premultiplier m and persistence length P_{df} yielding the best fits were nearly the same for the entire set of data; a single weighting premultiplier (m = 4) was finally used to reduce the number of variables. Using the ^{14}N A tensors ($A(^{15}N) \times 0.7131 = A(^{14}N)$) in the weighting resulted in larger $\langle \eta_i^2 \rangle$ and therefore $\langle \eta_i^2 \rangle + \langle \beta_0^2 \rangle$, but experimentally, $\langle \beta_i^2(^{14}N) \rangle < \langle \beta_i^2(^{15}N) \rangle$. This discrepancy shows that the simplistic weighting scheme needs additional refinement. All analysis of $\langle \beta^2 \rangle$ as a function of length, position, and temperature used the ^{15}N A tensors in the weighting scheme. However, that this simplistic weighting scheme was effective at all shows for the first time that the time constants of the motional modes are distinguishable through this technique, and that the calculated time constants scale correctly with mode index, despite not following Song et al. exactly. Results are shown in Tables 4.5 and 4.6 and in Figure 4.6. These are consistent with the predictions of the weakly bending rod theory, specifically that $\langle \beta_i^2 \rangle$ depends on length and position per equation 5.

Figure 4.6a shows $\langle \beta_i^2 \rangle$ as a function of length N. $\langle \beta_i^2 \rangle$ for the left- and right-labeled DNA series clearly belongs to a different, steeper, curve than $\langle \beta_i^2 \rangle$ for the middle-labeled DNA series. The solid lines depict the theoretical values $\langle \eta_i^2 \rangle + \langle \beta_0^2 \rangle$ calculated from the weakly bent rod theory and show reasonable agreement to the data. The slopes of the left- and right-labeled DNA $\langle \beta_i^2 \rangle$ are approximately two times greater than the slope of the middle-labeled DNA $\langle \beta_i^2 \rangle$; theory predicts that the ratio of slopes should be 1.78.

Figure 4.6b shows $\langle \beta_i^2 \rangle$ as a function of length and position, $f(N,i) = N\left(1+3\left(\frac{2i-N-1}{N}\right)^2\right)$. The data fall on a curve that is linear at low N, as the original theory specifies, and then rolls over using the simplistic weighting model described above, based on the time constants of the motional modes.

From Tables 4.5 and 4.6, the best-fit persistence length at 20°C is $P_{df}(20°) = 1250$ Å ± 75 Å (6%). $\langle \beta_0^2 \ (T=20°C) \rangle = 0.0082$, so the rms amplitude of flexibility at 20°C is $\beta_0 = \sqrt{\langle \beta_0^2 (T = 20°C) \rangle} * \frac{180}{\pi} = 5.2° \pm 0.2°$. P_{df} is slightly less than the current best estimate of dynamic persistence length from Schurr et al., 1500 Å.[9] The rms amplitude of flexibility at 20°C is also slightly less than the value of 7-10° obtained by Wilcoxon and Schurr from dynamic light scattering experiments.[10]

The estimates are approximately half the values estimated by Hustedt et al. (P_{df} = 2500 ± 340 Å , β_0 = 10.3° ± 0.2°) from their studies of DNA dynamics using T*.[4] The Hustedt et al. study used only middle-labeled DNA and did not weight the contributions of the time constants of the motional modes. The analogous analysis on only the middle-Q-labeled DNA yields P_{df} = 1400 Å. A comparison of the (middle-)T*- and middle-Q-labeled DNA $\langle\beta^2\rangle$ at T = 20°C is shown in Figure 4.7. When left-, middle-, and right-Q-labeled DNA data are analyzed, with the simplistic weighting described above, P_{df} = 1250 Å. Adding left- and right-labeled DNAs lowered the P_{df}, because the diminished contributions from the time constants of the motional modes appeared sooner on the end-labeled DNAs as compared to the middle-labeled DNAs. However, that our estimate of P_{df} is approximately half as large as that of Hustedt et al. is primarily due to the more rigidly attached spin label, shown in the contrast of the intercepts in Figure 4.7, $\langle\beta_0^2(Q)\rangle < \langle\beta_0^2(T^*)\rangle$. Table 4.7 presents the total root-mean-square amplitude of length-independent motion, β_0, as a function of temperature for Q- and T*-labeled DNAs. This motion represents the (length-independent) motion of the spin label itself, the base pair, and the spin label / base pair aggregate, and is calculated by $\beta_0 = \sqrt{\langle\beta_0^2\rangle} \times \frac{180°}{\pi}$. In every case, $\beta_0(Q) < \beta_0(T^*)$. Since the individual base pair motion should be the same for both labels, Q appears to have less independent spin label motion that T*, per design.

Although $\langle\beta_6^2(T=30°C)\rangle$ = 0.0854 for the 50-mer labeled in position 6 was not included in the analysis to determine P_{df}, the calculated $\langle\eta_6^2\rangle + \langle\beta_0^2\rangle$ (using the results from Table 4.5) and experimental $\langle\beta_6^2\rangle$ agreed to within 20%. Using $\langle\beta_0^2(T=30°C)\rangle$ = 0.0138, P_{df} from this single piece of data was estimated as ≈ 1000 Å, 20% smaller than P_{df} = 1250 Å from the other DNAs. Using P_{df} = 1250 Å, at T=30°C, $\langle\beta_6^2\rangle - \langle\eta_6^2\rangle = \langle\beta_0^2\rangle \approx 0.02$, which translates to 8.1°, 20% larger than 6.7°, calculated from $\langle\beta_0^2\rangle$ = 0.0138. This discrepancy again suggests that the simplistic weighting scheme used here needs refinement; nevertheless, for a first approximation, it yielded satisfactory results.

B. EFFECT OF TORSIONAL DYNAMICS

The effect of torsional dynamics on the EPR spectra of Q-labeled DNA duplexes was assumed to be negligible and consequently was entirely ignored in the simulations, as Hustedt et al. also did.[4] All experiments were done at 9 GHz frequency (X-band). At this frequency, EPR spectra are relatively insensitive to the g anisotropy and therefore to motion about the z-axis of the nitroxide. At higher microwave frequencies, such as 35 GHz (Q-band), the influence of the

nonaxial g tensor is enhanced, increasing sensitivity to rotation about the z-axis. In such a case, the torsional dynamics would have to be included.[4]

Schurr et al. have developed equations for the torsional dynamics directly analogous to those describing flexural (bending) dynamics:[1,2,7]

$$(10) \qquad \langle \varphi^2 \rangle_{total} = 2\langle \varphi_i^2 \rangle = 2\sum_{l\geq 2}^{N} Q_{il}^2 d_l^{t2} = \left(\frac{k_b T}{6\alpha}\right) \times N\left(1+3\left(\frac{2i-N-1}{N}\right)^2\right)$$

where $d_l^{t2} = \frac{k_b T}{\alpha \Lambda_l}$, k_b = 1.38x10^{-16} erg/K is Boltzman's constant, T is absolute temperature in Kelvin, α is the torsional force constant, and Λ_l and Q_{il}^2 are defined as for $\langle \eta^2 \rangle_{total}$. Using the relationship[5] $\frac{P_{df}}{h} = \frac{\kappa}{k_b T}$, the expression for $\langle \eta^2 \rangle_{total}$ may be rewritten as

$$(11) \qquad \langle \eta^2 \rangle_{total} = 2\langle \eta_i^2 \rangle = 2\sum_{l\geq 2}^{N} Q_{il}^2 d_l^{f2} = \left(\frac{k_b T}{6\kappa}\right) \times N\left(1+3\left(\frac{2i-N-1}{N}\right)^2\right),$$

so that $\langle \eta_i^2 \rangle$ and $\langle \varphi_i^2 \rangle$ are then related by their force constants:

$$(12) \qquad \frac{\langle \varphi_i^2 \rangle}{\langle \eta_i^2 \rangle} = \frac{\kappa}{\alpha}.$$

For P_{df} = 1250 Å = 1250x10^{-8} cm, k_b = 1.38x10^{-16} erg/K, T = 293 K, h = 3.4 Å = 3.4 x10^{-8} cm, κ = 1.5x10^{-11} erg is calculated. Using α = 4x10^{-12} erg,[1,2] $\langle \varphi_i^2 \rangle \approx 3.75 \langle \eta_i^2 \rangle$.

Just as the motional modes are used in the calculation of $\langle \eta_i^2 \rangle$, the modes are also related to $\langle \varphi_i^2 \rangle$: $\tau_l^t = \frac{\gamma^t}{\alpha \Lambda_l}$, where γ^t = 6x10^{-23} erg s is the torsional friction factor. Just as for the calculation of $\langle \eta_i^2 \rangle$, the time constant for the l = 2 motional mode does not contribute to the internal torsional dynamics measured at the center of a DNA duplex but does contribute at the ends. The time constants for the first 4 modes for both torsional dynamics (τ_l^t) and flexural dynamics ($\tau_l = \tau_l^f$), using γ^t = 6x10^{-23} erg s and α = 4x10^{-12} erg,[1,2] are shown as a function of length N in Figure 4.8. The calculated $\langle \eta_i^2 \rangle$ and $\langle \varphi_i^2 \rangle$ (no weighting, all motional modes contributing) are shown as a function of length N in Figure 4.9. In agreement with the prediction, $\langle \varphi_i^2 \rangle \approx 3.72 \langle \eta_i^2 \rangle$.

The effect of torsional dynamics on the EPR spectrum was next estimated as follows:

let the spin label vector be $v_{original} = \begin{pmatrix} \eta \\ \eta \\ \varphi \end{pmatrix}$ (recall that η describes motion perpendicular to the z axis and φ describes motion parallel to the z axis). Assume that the spin label vector rotates by θ off the z axis, purely around the y axis. The new orientation of the spin label vector is

$$(13) \qquad v_{new} = R_\theta v_{original} = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} \eta \\ \eta \\ \varphi \end{pmatrix} = \begin{pmatrix} \cos\theta \cdot \eta - \sin\theta \cdot \varphi \\ \eta \\ \sin\theta \cdot \eta + \cos\theta \cdot \varphi \end{pmatrix}.$$

Because motion about the z axis is not seen at X-band, the average orientation is then

$$\begin{aligned} \langle v_{new}{}^T \cdot v_{new} \rangle &= \frac{1}{2}\left\langle (\cos\theta \cdot \eta - \sin\theta \cdot \varphi \quad \eta) \cdot \begin{pmatrix} \cos\theta \cdot \eta - \sin\theta \cdot \varphi \\ \eta \end{pmatrix} \right\rangle \\ &= \frac{1}{2}\left(\cos^2\theta \cdot \langle \eta^2 \rangle + \sin^2\theta \cdot \langle \varphi^2 \rangle - 2\cos\theta\sin\theta\langle \eta\varphi \rangle + \langle \eta^2 \rangle\right) \\ &= \frac{1}{2}\left(\cos^2\theta \cdot \langle \eta^2 \rangle + \sin^2\theta \cdot \langle \varphi^2 \rangle + \langle \eta^2 \rangle\right) \\ &= \frac{1}{2}\left(\cos^2\theta \cdot \langle \eta^2 \rangle + \sin^2\theta \cdot \frac{\kappa}{\alpha}\langle \eta^2 \rangle + \langle \eta^2 \rangle\right) \\ &= \frac{1}{2}\langle \eta^2 \rangle\left(1 + \cos^2\theta + \sin^2\theta \cdot \frac{\kappa}{\alpha}\right) \end{aligned} \qquad (14)$$

For $\theta = 0°$, the $\langle \varphi^2 \rangle$ contribution is exactly 0; for $\theta = 20°$, the $\langle \varphi^2 \rangle$ contribution is approximately 20% of $\langle v_{new}{}^T \cdot v_{new} \rangle$. Consequently the neglect of the contribution from torsional dynamics to the internal dynamics of DNA constitutes an error of approximately 20% for $\theta = 20°$.

C. ANALYSIS OF $\langle \beta^2 \rangle$ AS A FUNCTION OF TEMPERATURE ALONE

As described above, Wilcoxon and Schurr suggested that the persistence length P_{df} is either independent of temperature T or inversely proportional to T, depending on whether the bending dynamics between two adjacent base pairs are governed by a square-well potential or a harmonic potential.[6] Hustedt et al.[4] were unable to distinguish between these cases, as were Wilcoxon and Schurr.[6] In the analysis of $\langle \beta^2 \rangle$ as a function of length and position above, the unweighted analysis showed a very slight preference for the harmonic model; the weighted analysis showed an equally slight preference for the square well model. Neither preference was statistically significant. Consequently, $\langle \eta^2 \rangle$, which is inversely proportional to P_{df} (equation 5), was directly analyzed as a function of temperature.

Within a position and length series, $\langle \beta^2 \rangle$ demonstrates a clear dependence on T: with increasing temperature, the spectra narrow and $\langle \beta^2 \rangle$ increases, implying an increase in the internal dynamics. See Table 4.3. However, $\langle \eta^2 \rangle = \langle \beta^2 \rangle - \langle \beta_0{}^2 \rangle$ is the quantity of interest.

Because the ^{14}N dataset was incomplete, only the ^{15}N dataset was analyzed.

Originally, $\langle\beta^2\rangle$ for the 50-mers and the 14-mers, $\langle\beta(50)^2\rangle$ and $\langle\beta(14)^2\rangle$, was plotted as a function of T. As shown in Figure 4.10, the plot of $\langle\beta(14)^2\rangle$ vs. T is widely scattered and curves upwards, even crossing the analogous line for $\langle\beta(50)^2\rangle$ vs. T. This unexpected result has several possible explanations. First, the lineshape for the 14-mers collapses at $T \geq 10°C$, going from a relatively broad lineshape to one of sharper lines. Because of this collapse, the simulations for these spectra were relatively insensitive to $\langle\beta^2\rangle$. Second, the dynamics for such short pieces of DNA are dominated by global motion, and the values for $\langle\beta(14)^2\rangle$ are small, as predicted by the weakly bending rod model ($\langle\beta_i^2\rangle \propto \langle\eta_i^2\rangle$, and $\langle\eta_i^2\rangle \propto$ length (equation 5)). Therefore the true (low) values for $\langle\beta(14)^2\rangle$ may be obscured by noise. Third, according to the UV-monitored thermal denaturation curves of an 11-mer containing the natural AT base pair and the analogous sequence containing the labeled QP base pair, up to 30% of the molecules had begun to melt at the highest temperature used in this study, 40°C.[10] (see Figure 4.11) The more rapid motion of the single-stranded DNA would be reflected in larger $\langle\beta(14)^2\rangle$ values. Fourth, the Lorentzian linewidth r should decrease as temperature T increases, since $r = \Delta A^2\tau$, where $\Delta A = \langle\beta^2\rangle\Delta A_0$, and correlation time τ decreases as T increases. Since r was kept fixed for the entire set of simulations, the unexpectedly large $\langle\beta(14)^2\rangle$ values may reflect compensation for the decreased τ with increased T, rather than increased internal dynamics. Oxygen broadening of the linewidth may have countered this effect, however.

Because of the uncertainties in $\langle\beta(14)^2\rangle$, $\langle\beta_0^2\rangle$ (from Tables 4.5 and 4.6 of the length and position dependence studies) was used. In contrast to $\langle\beta(14)^2\rangle$, $\langle\beta_0^2\rangle$ does not curve upwards as sharply (Figure 4.10). The line for $\langle\beta_i(50)^2\rangle - \langle\beta_0^2\rangle = \langle\eta_i^2\rangle_{calc}$ as a function of T can be written as $\langle\eta_i^2\rangle$ = slope × T + intercept. If slope = 0, then $\langle\eta_i^2\rangle$ and P_{df} are independent of T, consistent with the square well model. If intercept = 0, then $\langle\eta_i^2\rangle \propto T$, so P_{df} is inversely proportional to T, consistent with the harmonic potential, and the slope should be approximately $\langle\eta_i^2(20C{=}293K)\rangle/293 \times T \approx 1.13\times10^{-4}$ rad^2/K.

When plotted against T in Kelvin (Figure 4.12), the linear least squares lines for the difference $\langle\beta_i(50)^2\rangle - \langle\beta_0^2\rangle = \langle\eta_i^2\rangle_{calc}$ show modest positive slopes of $2.2\times10^{-4} \pm 1.0\times10^{-4}$ rad^2/K (left- and right-labeled, pooled) and $2.8\times10^{-4} \pm 1.3\times10^{-4}$ rad^2/K (middle-labeled) with intercepts of -0.023 ± 0.030 rad^2 (LR) and -0.055 ± 0.040 rad^2 (M). When the left- and right-labeled cases are separated, the slopes become $1.2\times10^{-4} \pm 0.8\times10^{-4}$ (L) and $3.2\times10^{-4} \pm 1.9\times10^{-4}$ (R) rad^2/K,

with intercepts 0.008 ± 0.024 rad^2 (L) and -0.053 ± 0.057 rad^2 (R). Although the intercepts are not 0 and have large errors, the intercepts come from extrapolating from a limited data range (40 degrees out of a total range of 0K to 273+40C = 323K). When the fit is constrained to be a proportionality, the slopes are $1.5\times10^{-4} \pm 0.2\times10^{-4}$ rad^2/K (LR) and $0.9\times10^{-4} \pm 0.1\times10^{-4}$ rad^2/K (M). Slopes, intercepts, and measure of fit (σ) are given in Table 4.8.

Despite the sizeable uncertainties on the slopes and intercepts of $\langle\eta^2\rangle$ vs. T, the slopes are positive. Furthermore, the linear fit is best, followed by the proportional fit, and then the no-slope fit, so that $\langle\eta_i^2\rangle$ is then proportional to temperature T. Equation 5 above shows that $\langle\eta_i^2\rangle$ is inversely proportional to P_{df}, $\langle\eta_i^2\rangle \propto \frac{h}{6P_{df}} \times f(N,i)$. Thus, the proportionality of $\langle\eta_i^2\rangle$ to T implies that, at least for sequences in this study, the persistence length P_{df} is proportional to 1/T, favoring the harmonic potential model slightly.

SECTION 4.4. CONCLUSIONS

The spin label Q was used to study the length, position, and temperature dependence of internal DNA dynamics. Results are consistent with the weakly bending rod model of Schurr and co-workers. In particular, as length increases, the spectra narrow, which was interpreted as an increase in the internal dynamics. The increase goes linearly to a point, after which the curve begins to roll over. The rollover of theory was achieved by using a simplistic weighting scheme that diminishes the contributions of the slower motional modes to $\langle\eta_i^2\rangle$ for large N. As the spin label position is moved toward the ends of the DNA molecule, the spectra also narrow, again interpreted as an increase in internal dynamics. The magnitude of the increase in internal dynamics agrees with the prediction of the weakly bending rod model:

$$\langle\beta_i^2\rangle = \langle\eta_i^2\rangle + \langle\beta_0^2\rangle \propto N\left(1+\left(\frac{3(2i-N-1)}{N}\right)^2\right).$$

As temperature increases, $\langle\beta^2\rangle$, the measure of internal dynamics, also increases. This slight increase suggests that, for the sequences studied here, the dynamic flexural persistence length, P_{df}, depends inversely on temperature. A harmonic potential then governs the bending angle between two adjacent base pairs.

The dynamic flexural persistence length at 20°C is estimated as 1250 ± 75 Å, and the rms amplitude of oscillation is estimated as 5.2° ± 0.2° at 20°C. These values are in line with Schurr's current best estimates of P_{df} (1500 Å) and rms amplitude (7° - 10°). Comparison with the previous T* work shows that the P_{df} is approximately half as long and the rms amplitude is approximately half as large. This factor of approximately 2 reflects primarily the rigidity of Q as

a probe relative to T*; T* has more independent spin label motion from the spinning about the acetylene tether.

Sources of error include choice of tensors, measurement of $\langle\beta_i^2\rangle$, neglect of $\langle\varphi_i^2\rangle$, and the analysis model itself. As discussed in Chapter 3, the 50% sucrose at 0°C condition was chosen as the best compromise between slowing the dynamics of the spin label and significantly changing the environment of the nitroxide. The "rigid limit" spectra are not truly in the rigid limit, so there is error in the tensor values. This error is, however, constant through the entire data set. There is an error of approximately 0.003 in β. As discussed above, the torsional dynamics $\langle\varphi_i^2\rangle$ were completely neglected in the analysis of the spectra. This is a systematic (and not constant) error throughout the entire data set. Furthermore, that the interior of the spectra did not fit completely indicates that the analysis model in the simulation program is not entirely correct. A post-calculation Gaussian convolution is used to compensate for broadening by the nearby non-equivalent nuclei, but it also is not exact. Finally, the weighting scheme described here used a simplistic approach to a complex problem and could be refined to reduce the discrepancies found in the analysis of $\langle\beta_i^2\rangle$ for the 50-mer labeled at position 6 and to deal with the change in $\langle\beta_i^2\rangle$ as a function of nitrogen isotope.

This series of experiments could be improved upon and supplemented by a number of additional experiments. Examining 70-mers, 90-mers and 130-mers would test the accuracy of the predictions given here. The theoretical curve in Figure 4.5 suggests that the curve for $\langle\beta_i^2\rangle$ as a function of length N for the left- and right-labeled DNA molecules peaks around 90 base pairs and drops through a local minimum around 130 base pairs. $\langle\beta_i^2\rangle$ as a function of length N for the middle-labeled DNA is predicted to begin flattening at 130 base pairs. Additional ^{14}N isotope spin labeled DNA molecules would yield a more complete comparison between isotopes. Synthesizing the left- and right-labeled DNA molecules of lengths 14, 28, 50, and 100 base pairs such that they were consistent with the "self-incorporating" model (illustrated in Figure 4.1) would resolve questions about the longer-range sequence effects on the spin label Q by maintaining the identical sequences for the end-labeled cases as length increases. Deoxygenating the samples would narrow the spectral lines by removing the oxygen broadening, allowing a more precise fitting of simulation to data. Finally, an exact replication of the sequences studied by Hustedt et al. would permit a complete comparison between the spin labels T* and Q.

Given the success here with the length, position, and temperature dependence of DNA dynamics, the spin label appears to be adequate for the job of examining the sequence dependence of DNA dynamics as well. This will be discussed in the next chapter.

NOTES TO CHAPTER 4

1. Schurr, J.M., Fujimoto, B.S., Wu, P., & Song, L. Fluorescence studies of nucleic acids, in *Topics in Fluorescence Spectroscopy, Vol 3. Biological Applications*, ed. J.R. Lakowicz, pp. 137-229. New York: Plenum Press (1992).

2. Allison, S.A. & Schurr, J.M. *Chem. Phys.* **41**, 35-59 (1979).

3. Cyclic permutation means figuratively taking the sequence, joining the ends, and making a circle, then cutting the circle in three different places to create the three different position members of each length sequence.

4. Hustedt, E.J. Ph.D. Thesis, University of Washington, 1989;, Hustedt, E.J., Spaltenstein, A., Kirchner, J.J., Hopkins, P.B., & Robinson, B.H. *Biochemistry* **32**, 1774-1787 (1993).

5. Wu, P., Fujimoto, B.S., & Schurr, J.M. *Biopolymers* **26**, 1463-1488 (1987); Song, L., Allision, S.A., & Schurr, J.M. *Biopolymers* **29**, 1773-1791 (1990).

6. Wilcoxon, J. & Schurr, J.M. *Biopolymers* **22**, 2273-2321 (1983)

7. Robinson, B.H. & Drobny, G.P. *Annual Review of Biophysical and Biomolecular Structures* **24**, 523-549 (1995).

8. Griffith, O.H. & Jost, P.C. in *Spin Labeling: Theory and Applications*, ed. L.J. Berliner, pp. 454-523. New York: Academic Press (1976).

9. Fujimoto, B.S. & Schurr, *Nature* **344**, 175-178 (1990).

10. Miller, T.R., Alley, S.C., Reese, A.W., Solomon, M.S., McCallister, W.V., Mailer, C., Robinson, B.H., & Hopkins, P.B. *J. Am. Chem. Soc.* **117**, 9377-9378 (1995); Alley, S.C. Ph.D. Thesis, University of Washington, 1996.

	length series,"self-incorporating" model	position series (14,28,50,100-mers)
14-mer:	----Q----	---Q--------
28-mer:	-----(14-mer)-----	------Q------
50-mer:	---------(28-mer)---------	--------Q---
100-mer:	----------(50-mer)----------	

Figure 4.1. Schematics of duplex DNA sequences used in the length and position dependence studies, illustrating the "self-incorporating" model and cyclic permutation. Each Q-labeled strand was matched with its P-containing complement strand.

Table 4.1. Sequences for middle-labeled DNA; each duplex used the sequence given here with its P-containing complement strand. Q represents the Q spin label.

N	sequence
14	5'-d(GAC CTC CQA TCG TG)
28	5'-d(ATG GTG C-[14-mer]-CTC CTC A
50	5'-d(TAA TCT ATC GT-[28-mer]-TCT TCG TGT CC
100	5'-d(TAC GAA CCT GAG CTA TTT CCC GGC T-[50-mer]-GGA AAG GTG GTG CTC ACT GGA GCA A

Table 4.2. Position (i) of the spin label Q in the duplex DNA molecules of length N base pairs.

N (base pairs)	left	middle	right
14	4	8	12
28	8	15	22
50	13	26	39
100	26	51	76

Table 4.3. $\langle \beta_i^2 \rangle$ as a function of length (N base pairs), temperature (T), position (i), and nitrogen isotope (^{14}N or ^{15}N).

N	T(°C)	isotope	left $\langle \beta_i^2 \rangle$	middle $\langle \beta_i^2 \rangle$	right $\langle \beta_i^2 \rangle$
14	0	^{15}N	0.0069	0.0000	0.0100
14	10	^{15}N	0.0156	0.0121	0.0256
14	20	^{15}N	0.0324	0.0132	0.0328
14	30	^{15}N	0.0529	0.0400	0.0576
14	40	^{15}N	0.0784	0.0676	0.0900
14	0	^{14}N	0.0000	0.0000	0.0081
14	10	^{14}N	0.0100	0.0001	0.0225
14	20	^{14}N	0.0121	0.0000	0.0256
14	30	^{14}N	0.0335	0.0361	0.0400
14	40	^{14}N	0.0361	0.0207	0.0625
28	0	^{15}N	0.0156	0.0061	0.0196
28	10	^{15}N	0.0231	0.0081	0.0289
28	20	^{15}N	0.0328	0.0169	0.0306
28	30	^{15}N	0.0462	0.0289	0.0462
28	40	^{15}N	0.0506	0.0400	0.0529
28	0	^{14}N	0.0064	0.0000	
28	10	^{14}N			
28	20	^{14}N	0.0380	0.0144	
28	30	^{14}N			0.0324
28	40	^{14}N			
50	0	^{15}N	0.0361	0.0121	0.0256
50	10	^{15}N	0.0400	0.0256	0.0380
50	20	^{15}N	0.0529	0.0400	0.0552
50	30	^{15}N	0.0600	0.0441	0.0625
50	40	^{15}N	0.0729	0.0576	0.0702
50	0	^{14}N	0.0266	0.0156	0.0199
50	10	^{14}N	0.0361		
50	20	^{14}N	0.0400		
50	30	^{14}N	0.0462	0.0324	0.0361
50	40	^{14}N	0.0552		
100	0	^{15}N	0.0342	0.0324	0.0361
100	10	^{15}N	0.0484	0.0441	0.0441
100	20	^{15}N	0.0552	0.0441	0.0529
100	30	^{15}N	0.0576	0.0552	0.0600
100	40	^{15}N	0.0756	0.0784	0.0784

Table 4.4. Results from analysis of $\langle\beta_i^2\rangle$ as a function of length and position, following Hustedt et al. Analysis with the power law dependence of the intercept (b) on temperature (T) is denoted b(T); analysis without the power law dependence is b(free). σ is the square root of the variance.

	square well			
	$P_{df}(20°)$ (Å)	$\langle\beta_0^2\rangle(20°)$, radians	$\beta_0(20°)$, degrees	σ
b(T)	3280 ± 330	0.025 ± 0.002	9.0° ± 0.9°	0.101
b(free)	2560 ± 210	5.6×10^{-4} ± 0.002	1.4° ± 0.1°	0.084

	harmonic			
	P_{df} (20°) (Å)	$\langle\beta_0^2\rangle(20°)$, radians	$\beta_0(20°)$, degrees	σ
b(T)	3010 ± 300	0.023 ± 0.002	8.8° ± 0.9°	0.100
b(free)	2540 ± 210	0.001 ± 0.002	2.1° ± 0.2°	0.084

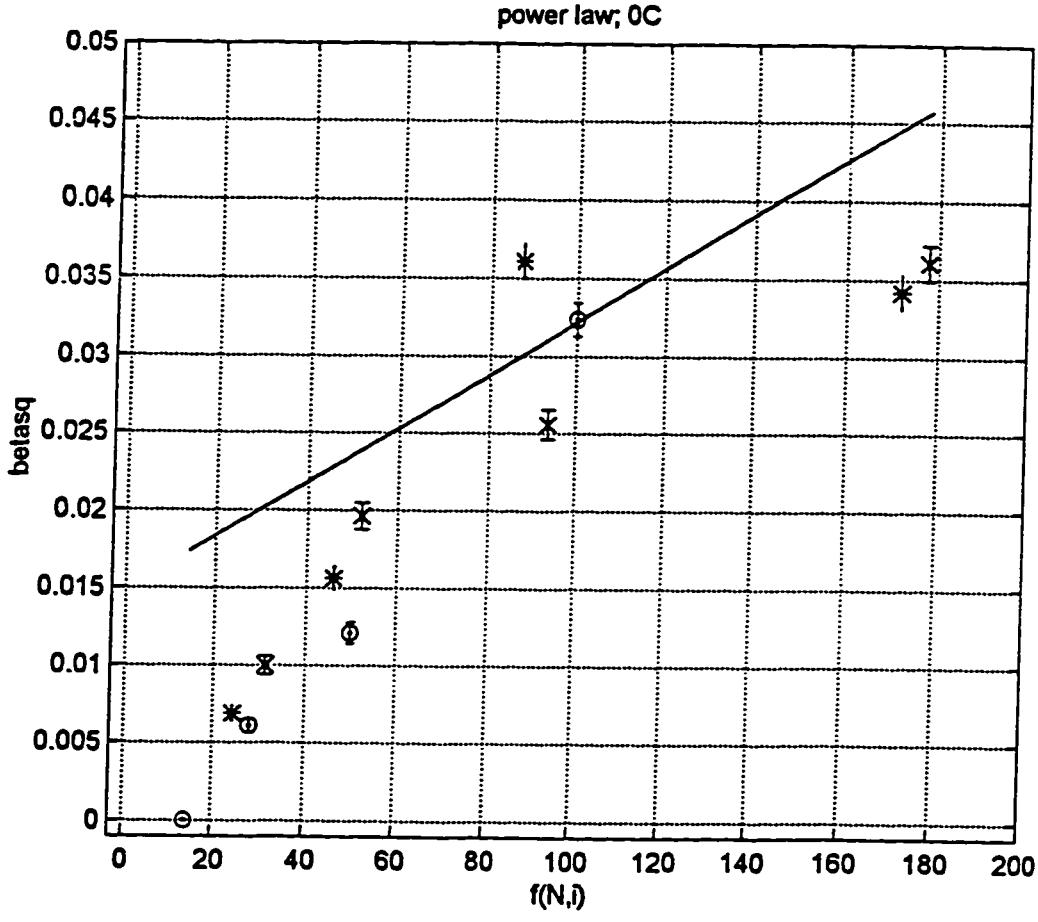


Figure 4.2. $\langle\beta_i^2\rangle$ as a function of length (N base pairs) and position (i), per Hustedt et al.[4] at T = 0°C, requiring the power law dependence of the intercept, b, on temperature T (equation 7). $f(N,i) = N\left(1+3\left(\frac{2i-N-1}{N}\right)^2\right)$. Calculated $\langle\eta_i^2\rangle+\langle\beta_0^2\rangle$ (solid line), ^{15}N left-labeled DNA (*), ^{15}N middle-labeled DNA (o), and ^{15}N right-labeled DNA (x) are shown. Parameters determined from this analysis are given in Table 4.4.

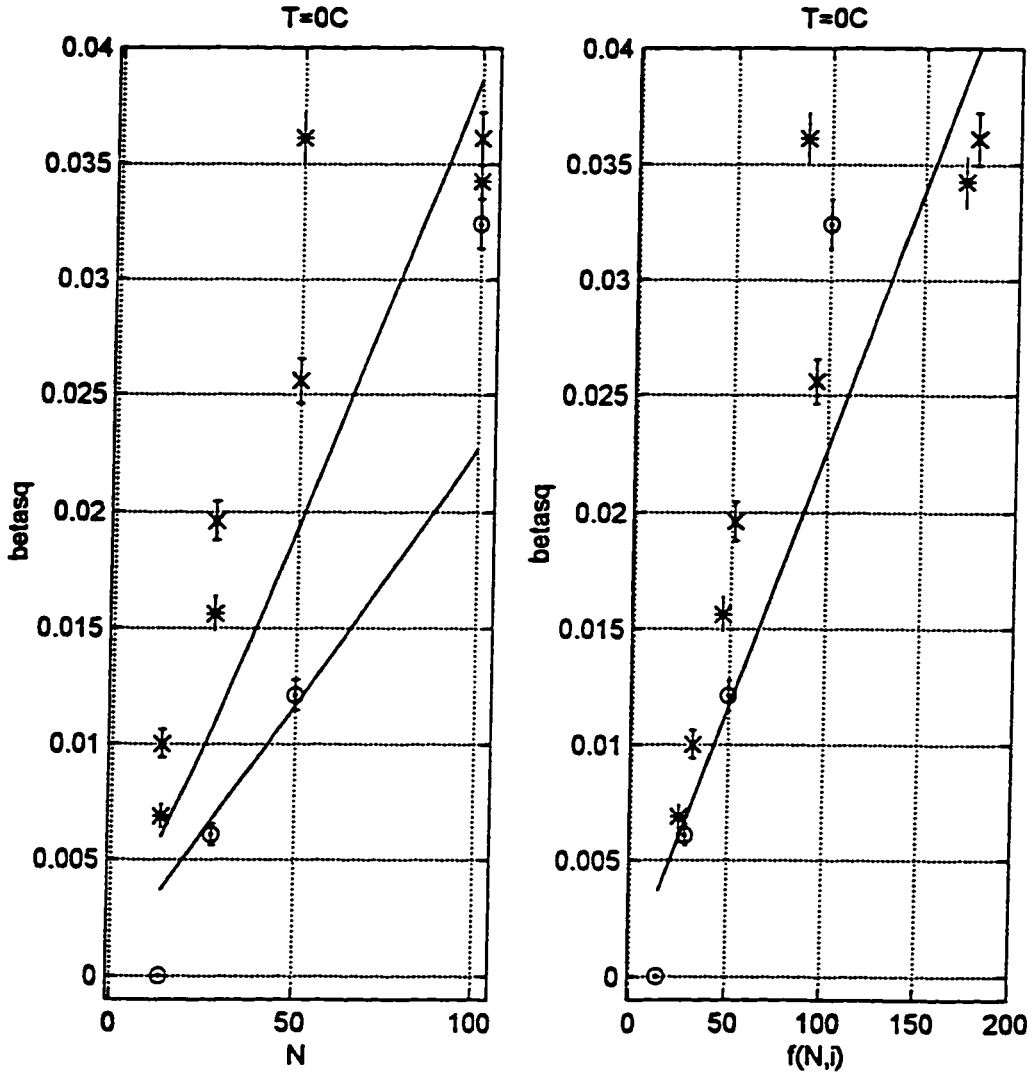


Figure 4.3. $\langle\beta_i^2\rangle$ as a function of length (N base pairs) and position (i), per Hustedt et al. [4] at T = 0°C, without the power law dependence of the intercept, b, on temperature T. $f(N,i) = N\left(1+3\left(\frac{2i-N-1}{N}\right)^2\right)$. a) $\langle\beta_i^2\rangle$ vs. N. b) $\langle\beta_i^2\rangle$ vs. f(N,i). Calculated $\langle\eta_i^2\rangle+\langle\beta_0^2\rangle$ (solid lines), ^{15}N left-labeled DNA (*), ^{15}N middle-labeled DNA (o), and ^{15}N right-labeled DNA (x) are shown. Parameters determined from this analysis are given in Table 4.4.

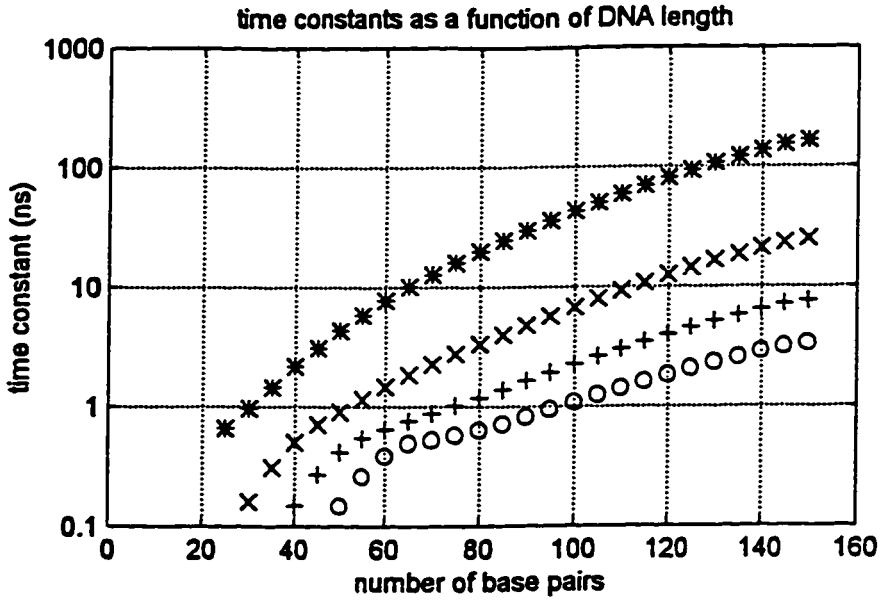


Figure 4.4. Time constants of motional modes of internal dynamics as a function of length, using P_{df} = 1250 Å, h = 3.4 Å, $\gamma = 3.03 \times 10^{-21}$ erg s, $\kappa = 1.5 \times 10^{-11}$ erg, T = 20°C, and viscosity = 0.01 poise. Modes shown are l = 2 (*), l = 3 (x), l = 4 (+), and l = 5 (o).

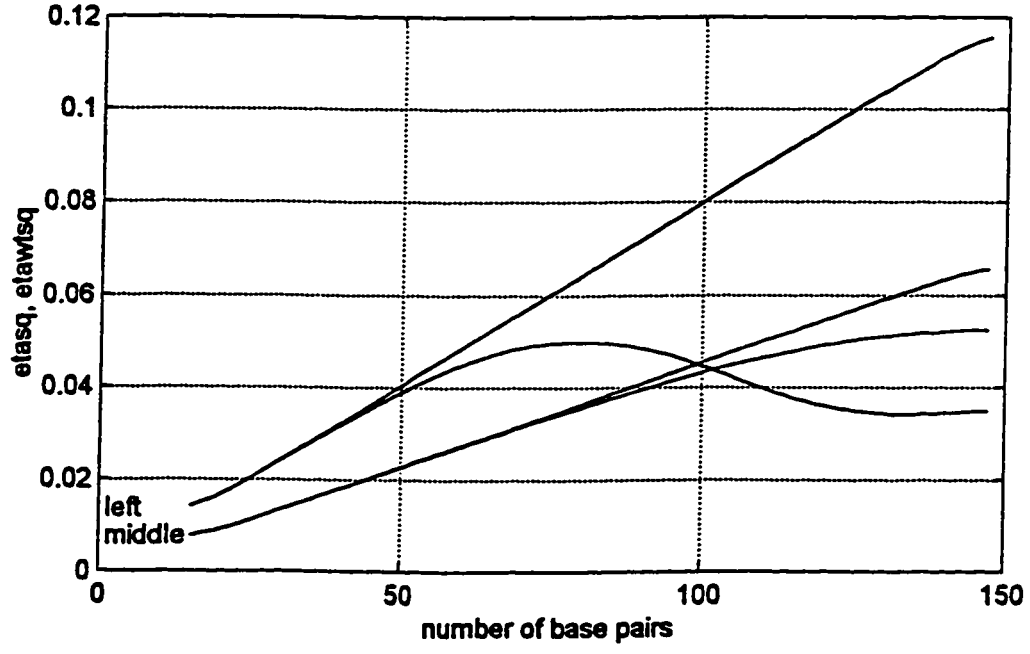


Figure 4.5. Effect of weighting the contributions of the l motional modes to $\langle \eta_i^2 \rangle$, using P = 1250 Å, h = 3.4 Å, A_{xx} = 9.223 G, A_{yy} = 6.982 G, and A_{zz} = 48.002 G. $\langle \eta_i^2 \rangle$ (etasq) vs. length (N base pairs), no weighting, is plotted with solid lines. $\langle \eta_i^2 \rangle_{weighted}$ (etawtsq) vs. length is plotted with dotted lines. Upper line shows $\langle \eta_i^2 \rangle$ and $\langle \eta_i^2 \rangle_{weighted}$ when DNA (length N base pairs) is labeled at position N/10 from the left (5') end; lower line shows $\langle \eta_i^2 \rangle$ and $\langle \eta_i^2 \rangle_{weighted}$ when DNA is labeled at position N/2 from the left (5') end.

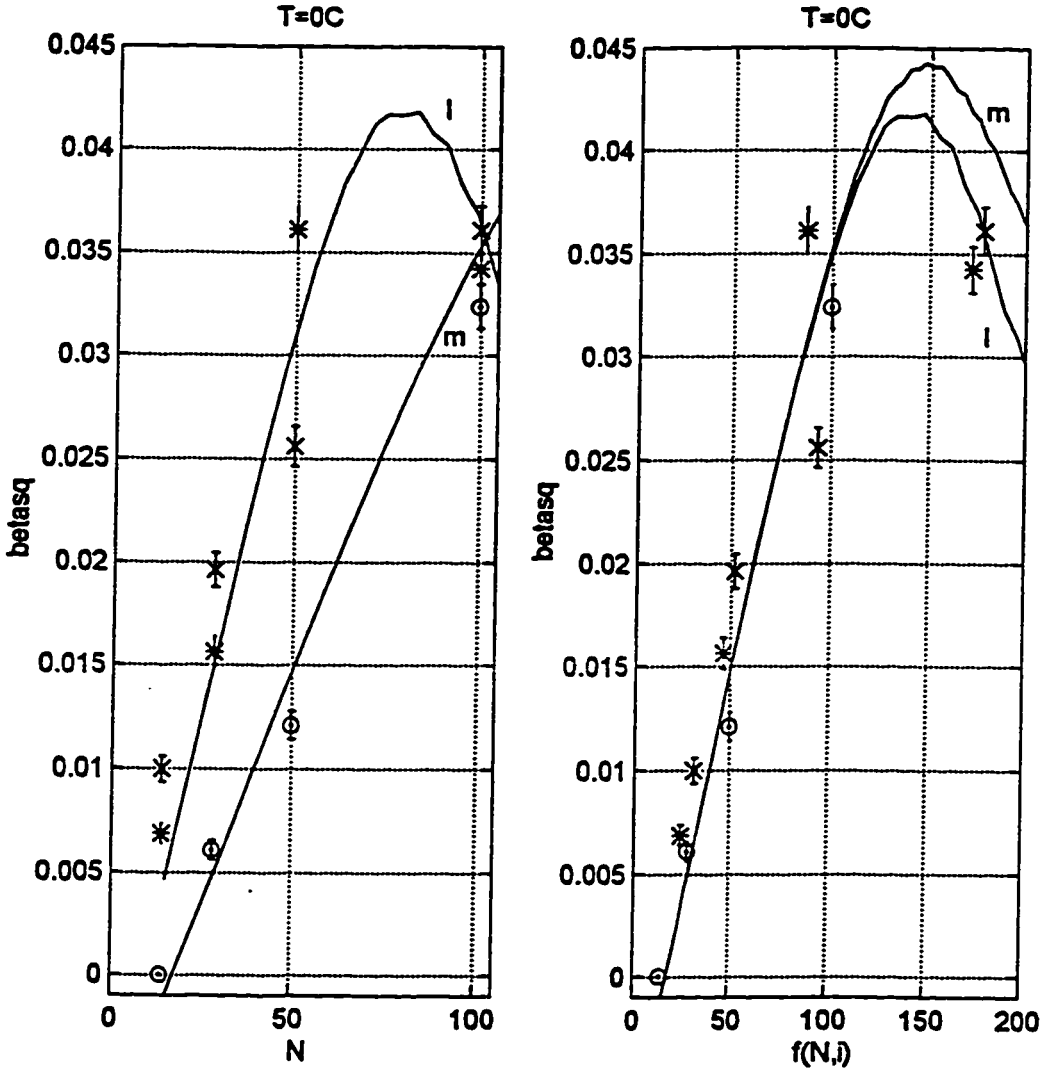


Figure 4.6. $\langle\beta_i^2\rangle$ as a function of length (N base pairs) and position (i), using data from all positions and lengths, at T = 0°C. a) $\langle\beta_i^2\rangle$ vs. N b) $\langle\beta_i^2\rangle$ vs. f(N,i). Each plot shows $\langle\eta_i^2\rangle+\langle\beta_0^2\rangle$ calculated using P_{df} = 1250 Å (solid lines, l = left, m = middle), left-labeled DNA (*), middle-labeled DNA (o), and right-labeled DNA (x). $f(N,i) = N\left(1+3\left(\frac{2i-N-1}{N}\right)^2\right)$.

Table 4.5. Results from weighted analysis of $\langle\beta_i^2\rangle$ as a function of length and position, using the square well potential model, in which P_{df} is independent of temperature T. σ is the square root of the variance.

T (°C)	# points	σ	P_{df}(20°C), Å	$\sigma(P_{df})$, Å	$\langle\beta_0^2\rangle$, rad^2	β_0, deg	$\sigma(\beta_0)$, deg
0	18	0.00088	1500	96	-0.0061	0	0.03
10	10	0.00093	1250	35	0.0006	1.4	0.03
20	11	0.00163	1250	54	0.0082	5.2	0.2
30	10	0.00259	1250	74	0.0138	6.7	0.4
40	10	0.00161	1250	31	0.0299	9.9	0.3

sum(σ) 0.0076

Table 4.6. Results from weighted analysis of $\langle\beta_i^2\rangle$ as a function of length and position, using the harmonic potential model, in which P_{df} is inversely dependent on T. σ is the square root of the variance.

T (°C)	# points	σ	P_{df} (20°C), Å	$\sigma(P_{df})$, Å	$\langle\beta_0^2\rangle$, rad^2	β_0, deg	$\sigma(\beta_0)$, deg
0	18	0.00086	1500	93	-0.0049	0	0.1
10	10	0.00093	1250	35	0.0015	2.2	0.1
20	11	0.00163	1250	54	0.0082	5.2	0.2
30	10	0.00264	1250	75	0.0131	6.6	0.4
40	10	0.00166	1250	33	0.0283	9.6	0.3

sum(σ) 0.0077

Table 4.7. Total rms amplitude of the length independent motion, β_0, as a function of temperature T from the T* and Q spin-labeled DNA studies. The model for T* takes the length-independent motion to be the isotropic two dimensional local motion of the labeled base pair.

Temp (°C)	Q, square well	Q, harmonic	T*
0	(0 ± 0.03)	(0 ± 0.1)	7.9 ± 0.2
10	1.4 ± 0.03	2.2 ± 0.1	
20	5.2 ± 0.2	5.2 ± 0.2	10.3 ± 0.2
30	6.7 ± 0.4	6.6 ± 0.4	
40	9.9 ± 0.3	9.6 ± 0.3	12.9 ± 0.2

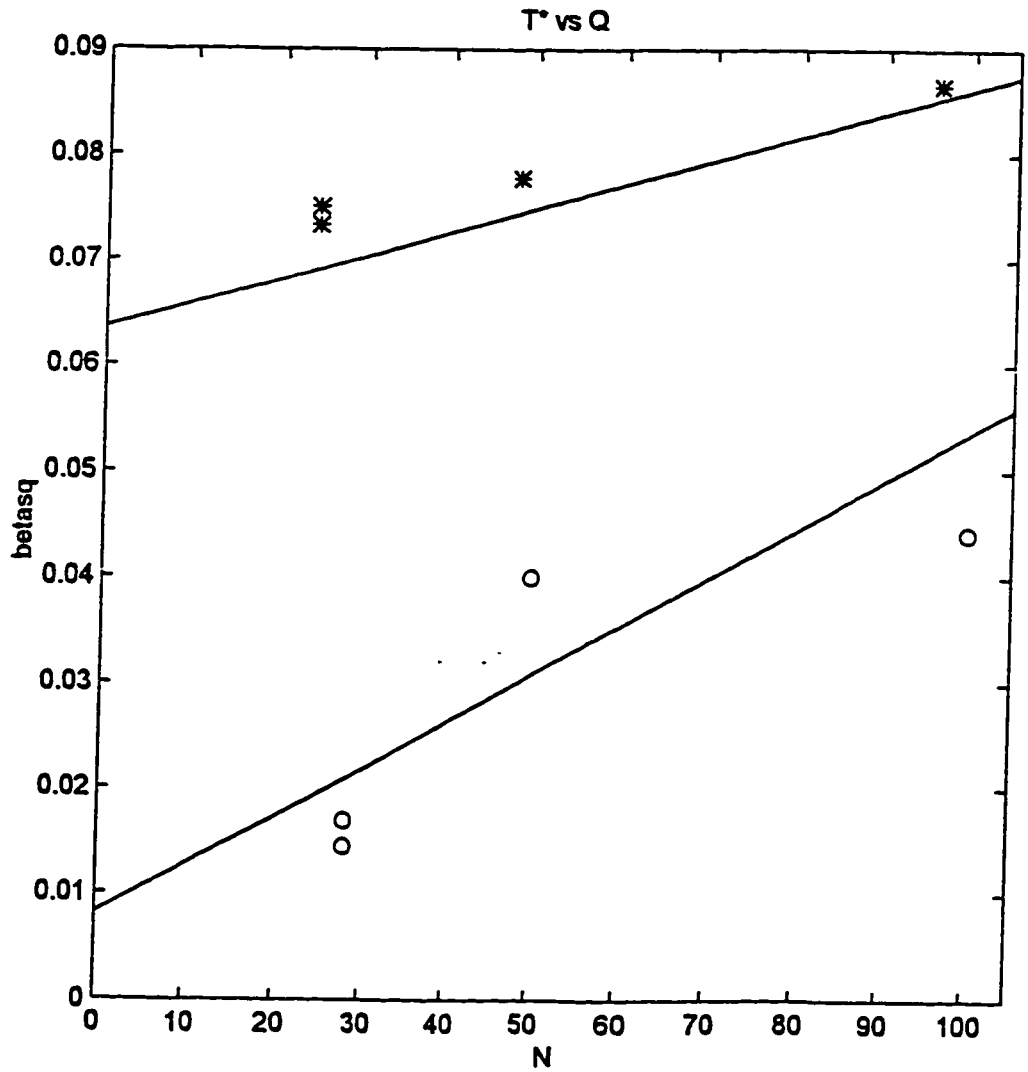


Figure 4.7. Comparison of $\langle\beta_i^2\rangle$ for T*- (*) and middle-Q- (o) spin-labeled DNA as a function of length (N base pairs), at T = 20°C. Solid lines show calculated $\langle\eta_i^2\rangle+\langle\beta_0^2\rangle$ curves with P_{df} = 2500 Å (T*) and 1250 Å (Q).

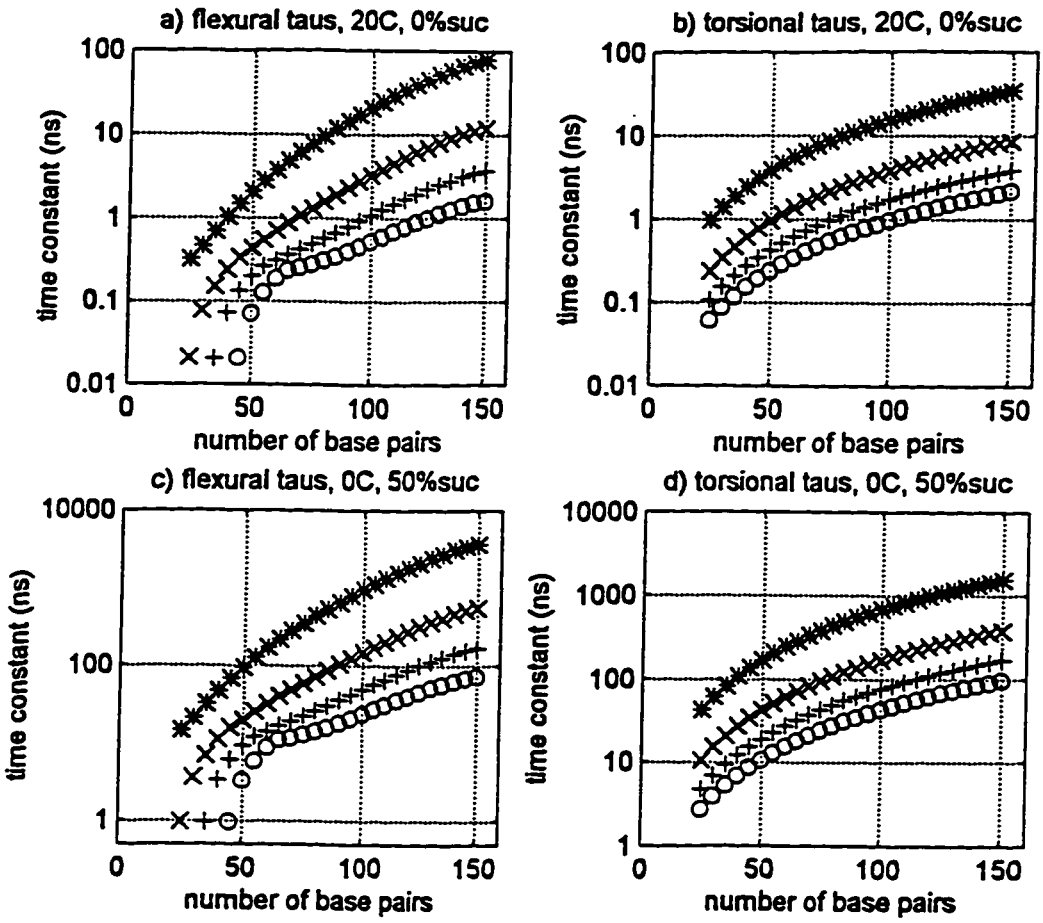


Figure 4.8. Time constants of both torsional (τ_l^t) and flexural (τ_l^f) motional modes of internal dynamics as a function of length (N base pairs), using P_{df} = 1250 Å, h = 3.4 Å. a) τ_l^f, T = 20°C, 0% sucrose, b) τ_l^t, T = 20°C, 0% sucrose, c) τ_l^f, T = 0C, 50% sucrose, d) τ_l^t, T = 0°C, 50% sucrose.

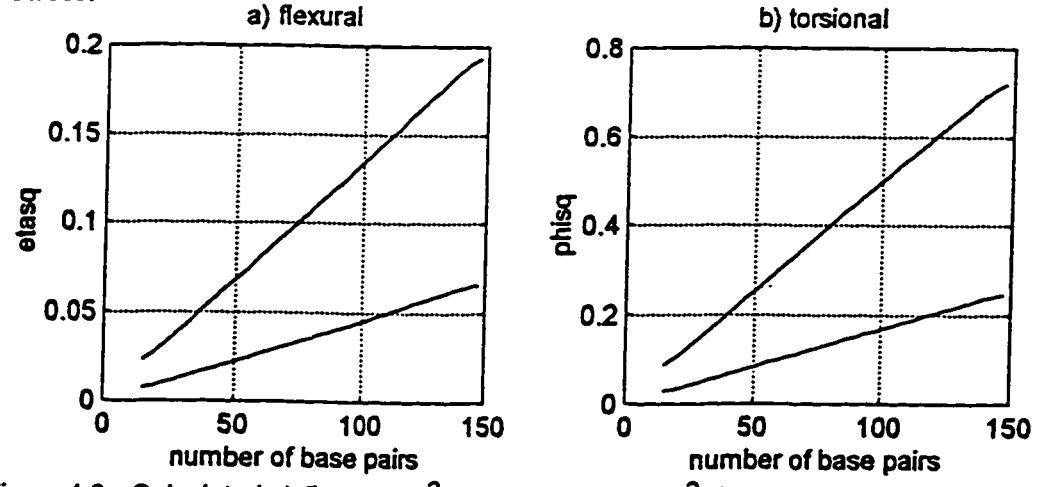


Figure 4.9. Calculated a) flexural $\langle\beta_l^2\rangle$ and b) torsional $\langle\varphi_l^2\rangle$ internal dynamics with no weighting (all motional modes contributing) as a function of length (N base pairs), using P_{df} = 1250 Å, h = 3.4 Å, T = 20°C, and viscosity = 0.01 poise.

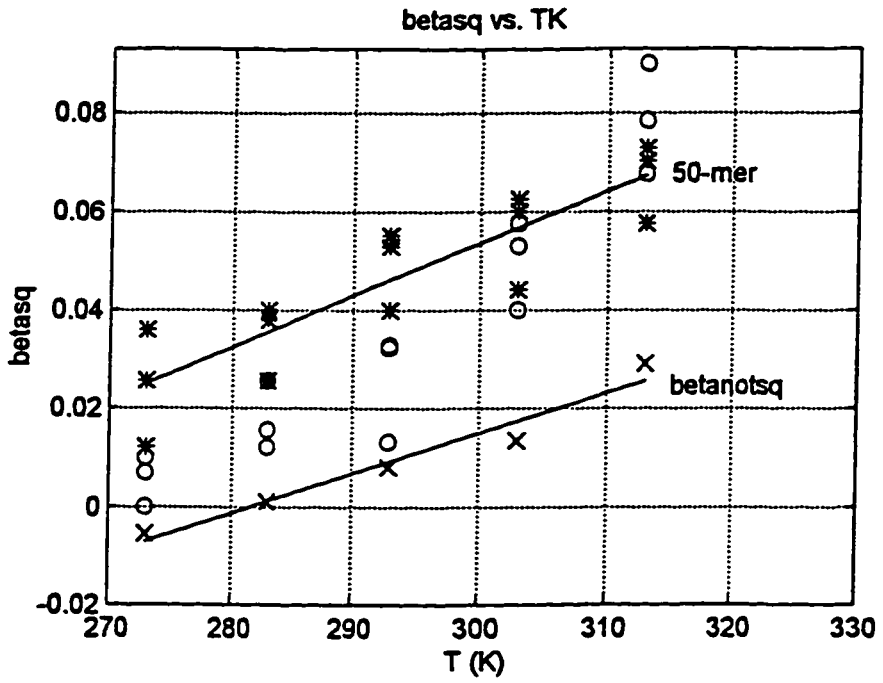


Figure 4.10. $\langle\beta_0^2\rangle$ (x), $\langle\beta_i^2(14)\rangle$ (o), and $\langle\beta_i^2(50)\rangle$ (*) as a function of temperature T. Solid lines are linear least squares fits through the data.

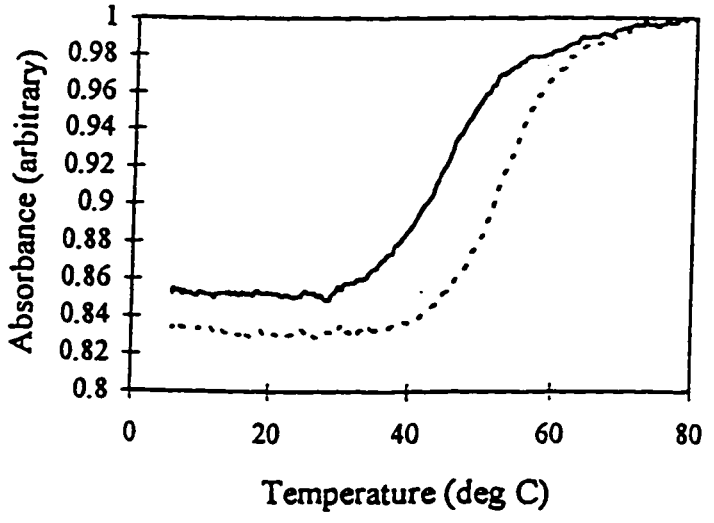


Figure 4.11. UV-monitored thermal denaturation curves of 5'-d(CGAGGTCCAGC)·5'-d(GCTGGACCTCG) (dashed line) and of 5'-d(CGAGGQCCAGC)·5'-d(GCTGGPCCTCG) (solid line) in PNE buffer.

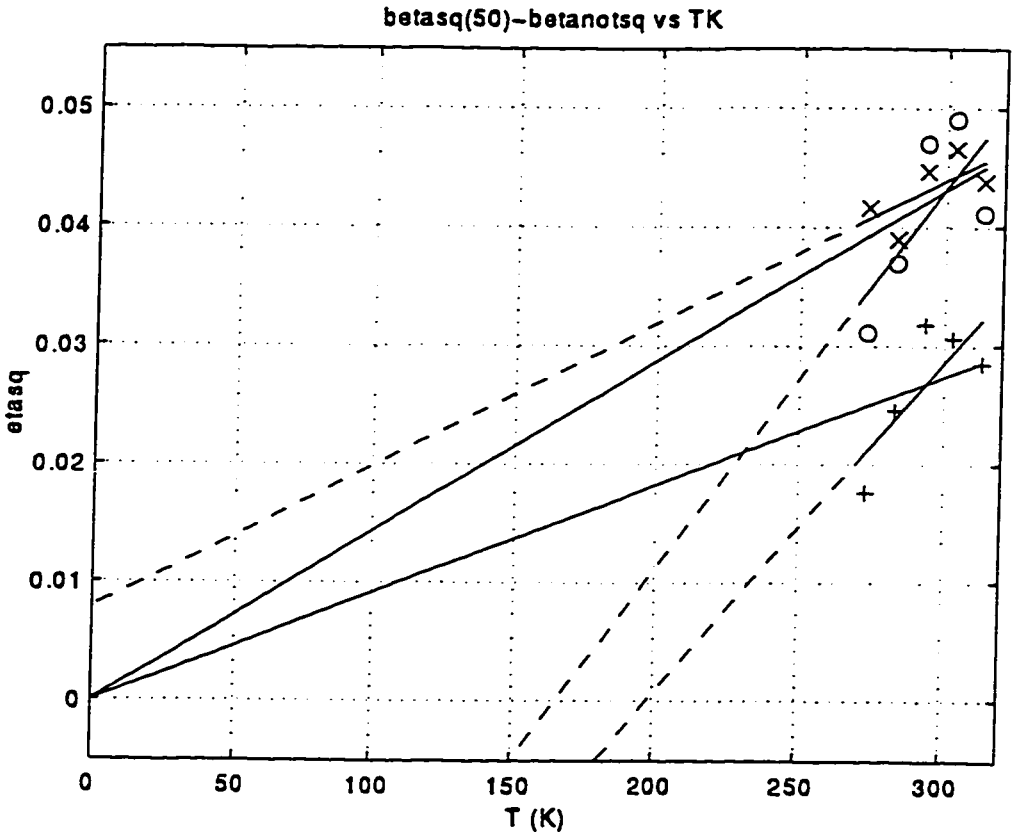


Figure 4.12. $\langle\beta_i^2(50)\rangle - \langle\beta_0^2\rangle = \langle\eta_i^2\rangle$ as a function of temperature T. $\langle\eta_i^2\rangle$ from left-labeled (L,x), middle-labeled (M,+), and right-labeled (R,o) cases are shown. Dashed lines are linear least-squares fits to the data; solid lines depict the proportional case. Slopes, intercepts, and measures of fit σ are given in Table 4.8.

Table 4.8. Results from analysis of $\langle\beta_i^2(50)\rangle - \langle\beta_0^2\rangle = \langle\eta_i^2\rangle$ as a function of temperature T. $\langle\eta_i^2\rangle$, LR, used the $\langle\beta_i^2(50)\rangle$ data from both left- and right-labeled DNA. Slope is given as $m\times10^{-4}$, for example $1.5 \times10^{-4} \pm 1.3 \times10^{-4}$ for $\langle\eta_i^2\rangle$ pooled. Proportional case had intercept (0,0); no slope case had slope = 0.

$\langle\eta_i^2\rangle$	# points	intercept	slope ($\times10^{-4}$)	σ
linear				
LR	10	-0.023 ± 0.030	2.2 ± 1.0	0.0040
L	5	0.008 ± 0.024	1.2 ± 0.8	0.0020
R	5	-0.053 ± 0.057	3.2 ± 1.9	0.0047
M	5	-0.055 ± 0.040	2.8 ± 1.3	0.0033
proportional				
LR	10	–	1.5 ± 0.2	0.0042
M	5	–	0.9 ± 0.1	0.0042
no slope				
LR	10	0.042 ± 0.002	–	0.0051
M	5	0.027 ± 0.002	–	0.0039

CHAPTER 5: SEQUENCE DEPENDENCE OF DNA DYNAMICS

SECTION 5.1. INTRODUCTION

Certain DNA sequences are favored in interactions which require bending or folding of the DNA molecule, such as protein-DNA interactions.[1] Examples include the CAP half-site and the TATA-binding box, which are believed to form dramatic bends to accommodate the protein.[2] Perhaps these sequences are more flexible or at least are more easily manipulated to the appropriate configuration. The effect of sequence upon the internal dynamics of DNA molecules has been a topic of much speculation.[3] The beginning of a systematic study of DNA dynamics as a function of base sequence is described here, with the development of a metric by which to measure the effect of sequence, followed by a ranking of the flexibilities of the dinucleotide and trinucleotide repeats.

SECTION 5.2. EXPERIMENTS

A 50-mer segment from the middle-labeled duplex 100-mer studied in Chapter 4 was used as the control sequence and is shown in Figure 5.1. This sequence was systematically varied to examine the sequence dependence of internal DNA dynamics. These duplex DNA sequences contained three regions, as shown in Figure 5.1: the label region, the test region, and the tail region. The label region, usually 11 base pairs long, contained the Q spin label, always in position 6 of the 50-mer. Bases in the test region were varied to examine different sequences. Sequence "units," such as dinucleotide repeats, were repeated to amplify the effects. The test region was usually 15 base pairs long but varied in composition and sometimes length, the difference in length being supplied with the control sequence bases in the tail region.

Samples were made using the automated phosphoramidite method and Klenow filling.[4] The Klenow fragment is a piece of DNA polymerase I and contains a 5' to 3' polymerase activity as well as a 3' to 5' exonuclease activity. Given a template strand, a primer, and nucleotide triphosphates, the Klenow fragment will fill the gap and make double stranded DNA. In this case, the automated phosphoramidite method was used to synthesize a 50-mer containing P in position 45 and the complementary 11-mer containing Q in position 6. These two DNA strands comprise the template and primer, which the Klenow fragment used to create the duplex 50-mer sequence. All the sequence dependence samples were generated using the same primer (with Q in position 6) and only the complementary (P-containing) strand was separately synthesized for each sample, thus conserving the spin label Q. DNA was resuspended in PNE buffer (10 mM phosphate buffer, pH 7.0, 115 mM Na^+, 0.1 mM EDTA). Typically samples were 10 mg/ml, or 0.6 mM in DNA.[5] The majority of samples used the [^{15}N,D] isotope of Q.

The sequences examined are listed in Table 5.1. The control sequence, called NT, is shown in Figure 5.1. Inserts (15 base pairs long) of each of the 6 unique dinucleotide repeats (AA = $(AA)_7A$, AC, AG, AT, CC, CG) and the 10 unique trinucleotide repeats (CAG = $(CAG)_5$,

CGG, AAT, AAG, AAC, ACC, ACG, ACT, AGG, ATC) replaced bases 12 through 26 of the control sequence to examine the effect of the sequence repeats. Additionally, a randomized 15-mer insert (XX) was examined, in principle simultaneously examining $4^{15} \approx 1.07 \times 10^9$ possible sequences. For the single-blind samples (MN, YZ), S. Alley made the samples knowing what the sequences were, but A. Reese collected and analyzed EPR spectra knowing only that the sequences were dinucleotide repeats. A 15-mer insert of the TG repeat (TG), the other side of the AC dinucleotide repeat, was tested to determine whether the technique was sensitive to which side of the base pair (AC or TG) was on the spin-labeled strand. Length and proximity of the flexible region were also tested: an 8-mer insert (SA, short AT) and a 30-mer insert (LA, long AT) of the AT dinucleotide repeat, each starting at position 12, were studied, as were 15-mer inserts of the AT dinucleotide repeat in positions 12 through 26 (AT), 24 through 38 (AU), and 36 through 50 (AV). The 15-mer inserts of TAX and ATX (using random bases at positions X) were tested to determine whether the AT and TA base pairs could be distinguished. The TATA box, TATA TAAA (TATA), and the CAP half-site, TGTG ACAT (CAP), were also examined. Finally, the effects of phasing and structural effects on dynamics were studied with two two-full-turn dinucleotide repeats $(AA)_{10}$ (MA, middle AA) and $(AT)_{10}$ (MT, middle AT) and the 15-mers $(AA)_7A$ (AO) and $(AT)_7A$ (TO) shifted downstream by 5 base pairs (positions 17 through 31 instead of 12 through 26). Representative sequences are explicitly listed in Table 5.2.

CW-EPR experiments were run on both the homebuilt PDP and the commercial Bruker spectrometers at 30° ± 2°C (PDP) and 30.0° ± 0.1°C (Bruker); spectrometers and experimental conditions were described in Chapter 3. The experiments were run at 30°C so the differences in internal dynamics as a result of sequence would be dramatic.

Specifically, a rigid n-mer tumbling in the absence of internal dynamics appears to tumble about its center at rates determined by the rotational diffusion coefficients D ($\text{rate} = \frac{1}{\tau}$, $\tau = \frac{1}{6D}$) for that n-mer. If however the n-mer is attached to a brick by a universal joint and must tumble about its end, then the n-mer appears to tumble at rates governed by length 2×n. That is, the n-mer appears to be twice as long as it really is. Consequently, if a 50-mer has a highly flexible region from positions 12 to 26, the motion should approximate that of a (2×11) 22-mer tumbling freely in solution. If instead the region from positions 12 to 26 is completely rigid, then the global dynamics of this 50-mer are simply those of a 50-mer.

At 30°C the (global dynamics) rotational diffusion coefficients, expressed as correlation times, are 18.24 ns (τ_{perp}) and 4.96 ns (τ_{para}) for a 22-mer and 114.83 ns (τ_{perp}) and 10.25 ns (τ_{para}) for a 50-mer (Table 3.2). This 6-fold difference in τ_{perp} is both dramatic and within the range of greatest sensitivity for the CW-EPR experiment.

The optimized tensors (given in Table 3.1), global dynamics as calculated from Tirado and de la Torre (given in Table 3.2), and the internal dynamics as measured with $\langle\beta^2\rangle$ were used to simulate the spectra from these CW-EPR experiments. θ_{tilt} = 20°, r = 0.4 G (Lorentzian linewidth), and σ = 0.5 G (Gaussian broadening) were used.

SECTION 5.3. CONTROLS AND THEIR RESULTS

To determine whether the differences between relatively rigid and flexible sequences could be measured at all, two potentially grossly distorted sequences and their controls were synthesized and spectra were collected and simulated. The first sequence was a 60-mer with a gap from position 16 to position 30 (named gap); the second was a 50-mer with 2 propyl linkers (S) inserted between positions 11 and 12 (floppy). The sequences are given in Table 5.3. The results are shown in Figure 5.2, in which the experimental spectra of the potentially distorted sequences are overlaid with experimental spectra of a 22-mer and a 50-mer for floppy and simulations of a 30-mer and a 60-mer for gap. The shorter DNA spectra are in better agreement with the spectra for both gap and floppy than the longer DNA spectra; the full length DNA spectra fit the controls well. In addition, simulations using the dynamics of 50- and 60-mers fit floppy and gap very poorly, but simulations using the dynamics of 22- and 30-mers gave much better agreement with the spectra for floppy and gap (Table 5.4). These results showed that gap and floppy were indeed more flexible, and that the distinction between the relatively rigid controls and the more flexible sequences could be seen. While the results for the gap and floppy sequences were dramatic, spectra for the sequence-dependence samples yielded much more subtle differences, as expected.

Another control involved the $(AA)_7A$ and $(AT)_7A$ (AA and AT) samples in 50% sucrose at 0°C. These conditions represent the rigid limit compromise between slowing the dynamics and changing the nitroxide's environment.[5] If the measured difference between relatively rigid (AA, $\langle\beta^2\rangle$ = 0.0820 at 30°C and 0% sucrose) and flexible (AT, $\langle\beta^2\rangle$ = 0.0934 at 30°C and 0% sucrose) sequences were truly representative of a difference in only dynamics, then, in the absence of dynamics, the spectral differences should disappear as well. Only the common, underlying magnetic interactions (characterized by tensors) would remain, and one set of tensors would then adequately fit the rigid limit spectra of DNA samples. Reasonable fits resulted from application of the standard tensor set (given in Table 3.1) to the spectra (R^2 = 0.9330 (AA), 0.9455 (AT)). Better fits resulted from simulating each spectrum individually to determine its optimal set of tensors, given in Table 5.5 (R^2 = 0.9808 (AA), 0.9789 (AT)). Although these sets of tensors were not identical, each tensor element was within 0.8% of the corresponding element in the other set. Furthermore, a single set of tensors that fit both spectra was found; agreement between simulations and spectra was as good as the agreement found for the individually

optimized sets of tensors. That $A_{xx} \approx A_{yy}$ (rather than $A_{xx} > A_{yy}$ as was the case for the standard tensor set) is attributed to the general insensitivity of the ^{15}N data to the inner (A_{xx}, A_{yy}) elements. Had a set of (^{14}N) samples been used as well, the distinction between A_{xx} and A_{yy} might have become clearer. Thus, in the absence of dynamics, these sequences (with different levels of dynamic flexibility) had the same underlying tensors. Unfortunately, this result did not extend to the control sequence in 50% sucrose at 0°C (NT, $\langle\beta^2\rangle = 0.0853$ at 30°C and 0% sucrose). This spectrum could not be pooled with the $(AA)_7A$ and $(AT)_7A$ spectra and fit with a common set of tensors as the NT spectrum was too narrow. The only features distinguishing this sample from the $(AA)_7A$ and $(AT)_7A$ samples were that the NT sample was 5 months newer and that the sequences were different. Perhaps the difference in age accounts partially for this result (aging phenomenon discussed below); however, since the $(AA)_7A$ did not show appreciable "aging," there may be some other explanation. Tensors used in this control series are given in Table 5.5.

Section 5.4. Results and Discussion

Over the 7 months of experiments, $\langle\beta^2\rangle$, the measure of internal dynamics, decreased for the heterogeneous dinucleotide repeats (AC, AG, AT, CG) and several other samples, as shown in Figure 5.3 and Table 5.6. The homogeneous dinucleotide repeats (AA, CC), the control sequence (NT), and the random sequence (XX) did not appear to "age" in this way. Ethanol precipitation of the DNA and resuspension in PNE buffer appeared to "reset the clock," such that the earlier, higher values of $\langle\beta^2\rangle$ were reachieved. This aging phenomenon is not understood, but appears to be reversible through ethanol precipitation and resuspension in buffer.

Table 5.7 gives $\langle\beta^2\rangle$ as a function of sequence, where $\langle\beta^2\rangle$ is the measure of the internal dynamics. Larger $\langle\beta^2\rangle$ indicates greater internal flexibility. The intrinsic error on β is approximately 0.003 radians. The errors in $\langle\beta^2\rangle$ given in Table 5.7 are the larger of the two following options: a) the standard deviation of the independent measurements of $\langle\beta^2\rangle$, and b) $\sqrt{4 \times \langle\beta^2\rangle \times (0.003)^2}$ (see Note 7). The dinucleotide and trinucleotide repeats are ranked by flexibility in Table 5.8. For $\theta_{tilt} = 20°$, AT was the most flexible of the dinucleotide repeats and AA the least flexible. Flexibilities for both the control sequence, NT, and the randomized sequence, XX, were in the middle of the flexibility range, neither exceptionally flexible nor rigid. The single-blind experiments, MN and YZ, which were CG and AC respectively, yielded results very similar to $\langle\beta^2\rangle$ for CG and AC. The TG repeat and the AC repeat gave similar results, suggesting that, at least by this technique, the DNA helix is isotropic with respect to the two

strands. As the (relatively flexible) AT repeat region was lengthened, from 8 (SA) to 15 (AT) to 30 (LA) base pairs, the flexibility as seen by the spin label at position 6 increased. As the AT repeat region was moved away from the spin label, from starting at position 12 (AT) to position 24 (AU) to position 36 (AV), the effect of the flexibility as reported by the spin label at position 6 diminished. Similar results for ATX and TAX indicate that AT and TA are indistinguishable at this level. The TATA box and the CAP half-site sequences were nearly as flexible as the equivalent (short) AT repeat, SA. The high degree of flexibility of the TATA box and CAP half-site sequences suggests that dynamic flexibility may be involved in protein-DNA interactions. Of the trinucleotide repeats, AAT was most flexible and CGG was least flexible.

Most of the trinucleotides yielded results consistent with one of two models. One model correlates the trinucleotide flexibility with the most flexible link of the chain, where the links represent the dinucleotide steps comprising the trinucleotide chain. For example, AAT contains the AA, AT, and TA dinucleotide steps. $\langle\beta^2(AAT)\rangle = 0.0922$ is close to $\langle\beta^2(AT)\rangle = \langle\beta^2(TA)\rangle = 0.0934$, the most flexible of the component dinucleotide repeats. The other model averages the flexibilities of the dinucleotide steps comprising the trinucleotide repeat to generate the trinucleotide flexibility. For example, CAG contains the CA, AG, and GC dinucleotide steps. $\langle\beta^2(CAG)\rangle = 0.0877$ is closer to the average of $\langle\beta^2\rangle$ for the three repeats (0.0866 (CA), 0.0899 (AG), 0.0842 (GC), average = 0.0869) than to $\langle\beta^2\rangle$ for GC, the most flexible component ($\langle\beta^2(GC)\rangle = 0.0842$). However, three of the ten trinucleotides were more flexible than either model predicts. Flexibilities for two of the trinucleotides, AAC and ACG, did not fit either model at all, and one, AGG, was on the edge. That seven of the ten trinucleotide repeats did have flexibilities correlated in some way with the component dinucleotide step flexibilities suggests that a dinucleotide step model is partially correct but needs refinement.

A simple model of a DNA molecule with two regions of flexibility characterized by force constants κ and κ' was developed to calculate the theoretical $\langle\beta_6^2\rangle_{theo} = \langle\eta_6^2\rangle + \langle\beta_0^2\rangle$. $\langle\eta_6^2\rangle$ and $\langle\eta_6^2\rangle_{weighted}$ were calculated according to the procedure described in detail in Chapter 4, using a modified ***A*** matrix (given in Note 8). Later a more sophisticated model was developed, in which the force constants for each dinucleotide step were used (the modified ***A*** matrix is given in Note 9). The remainder of the calculation is exactly the same as described in Chapter 4. Briefly, diagonalizing the matrix ***A*** yields the eigenvalues Λ and the matrix of eigenvectors ***Q*** ($\boldsymbol{Q}\cdot\Lambda\cdot\boldsymbol{Q}^T = \boldsymbol{A}$). $\langle\eta^2\rangle_{total} = 2\langle\eta_i^2\rangle = 2\sum_{i\geq 2}^{N} Q_{il}^2 d_l^2$ and $\langle\eta_i^2\rangle_{weighted} = 2\sum_{l=modes} Q_{il}^2 d_l^2 w_{il}$, where $d_l^2 = \left(\frac{h}{P_{df}\Lambda_l}\right)$, $\Lambda_l = 4\sin^2\left(\frac{(l-1)\pi}{2N}\right)$, and $Q_{il}^2 = \frac{2}{N}\cos^2\left(\frac{(i-\frac{1}{2})(l-1)\pi}{N}\right)$. P_{df} = 1250 Å is the

dynamic flexural persistence length (from Chapter 4) for the control (NT) sequence, h = 3.4 Å is the rise per base pair,[6] d_l^2 represents the mean-squared amplitude of the l^{th} bending mode and Q_{il}^2 shows the contribution of the l^{th} mode to the motion of the i^{th} base pair. The weighting is calculated with $w_{i,l} = e^{-(2m\tau_l \Delta A_{i,l})}$, where τ_l is the time constant for mode l ($\tau_l = \frac{\gamma}{\kappa \Lambda_l}$, where γ is the friction factor, κ is the force constant calculated by $\kappa = \frac{P_{df} k_b T}{h}$ (k_b = 1.38x10^{-16} erg / K is Boltzman's constant, T = 303 K is the absolute temperature), and Λ_l is the l^{th} eigenvalue), m is the weighting premultiplier, and $\Delta A_{i,l} = (A_{zz} + A_{perp}) \sin^2(2Q_{il}^2 d_l^2)$ reflects how much the A tensor is narrowed at position i because of the l^{th} motional mode. The l = 0 and l = 1 modes represent global translation and rotation, respectively, and are not included in the calculation of $\langle \eta^2 \rangle$.

The simple model (with 2 regions of flexibility) was used for three calculations. First, the relative values of κ'/κ required to produce the $\langle \beta_6^2 \rangle$ determined for the dinucleotide repeats were extracted. The persistence length P_{df}, κ, and $\langle \beta^2 \rangle$ are related:

(1) $$\frac{P_{df}}{h} = \frac{\kappa}{kbT} \propto \frac{1}{\langle \beta^2 \rangle}.$$

Based on this relationship, the value of $\langle \beta_6^2 \rangle$ was converted to a κ'/κ value and persistence length for each dinucleotide repeat. The persistence length thus varies as a function of sequence. The maximum root mean squared angle θ that each sequence can achieve was also calculated. Let the potential energy $U = \frac{1}{2}\kappa\theta^2$, so then the probability of being at that energy is

(2) $$P = \frac{\exp(-\frac{U}{k_b T})}{Z} = \frac{\exp(-\frac{\kappa\theta^2}{2k_b T})}{Z}, \text{ where } Z = \sqrt{\frac{2\pi k_b T}{\kappa}}.$$

Then the average mean squared angle is

(3) $$\langle \theta^2 \rangle = \int_{-\infty}^{\infty} \theta^2 P(\theta) d\theta = \frac{k_b T}{\kappa}$$

(note that P is in the form of a Gaussian, $G = \frac{\exp(-\theta^2 / 2\sigma^2)}{\sqrt{2\pi\sigma^2}}$, and $\int_{-\infty}^{\infty} \theta^2 G(\theta) d\theta = \sigma^2$)

and so the root mean squared angle (in one dimension) is

(4) $$\theta_{rms} = \sqrt{\langle \theta^2 \rangle} = \sqrt{\frac{k_b T}{\kappa}}$$

Results are given in Table 5.9.

Second, the model was used to calculate how $\langle\beta_6^2\rangle$ ($\langle\beta^2\rangle$ at position 6, where the spin label resides) should change as the region of flexibility changed in length and position. Sequences were based on the relatively flexible AT dinucleotide repeat (AT = $(AT)_7A$); sequences SA, AT, LA, AU, and AV are explicitly listed in Table 5.2. The calculated and experimental values of $\langle\beta_6^2\rangle$ are shown in Table 5.10. As the length of the test region increased in length (SA = 8 bp, AT = 15 bp, LA = 30 bp), both calculated and experimental values of $\langle\beta_6^2\rangle$ increased. The results were consistent with a single persistence length P_{df} and a single force constant ratio κ'/κ. As the region of flexibility moved further away from the spin label, the degree of flexibility reported by the spin label decreased as predicted also. (AT started at position 12, AU started at position 24, AV started at position 36.) The calculated $\langle\beta_6^2\rangle$ was within the experimental error of the measured $\langle\beta_6^2\rangle$ for each case.

Third, the model was used to predict how $\langle\beta_6^2\rangle$ should change for the A-tract sequences, AA ($(AA_7)A$), AO ($(AA_7)A$ shifted 5 bp downstream), and MA ($(AA)_{10}$). Sequences are explicitly listed in Table 5.2. For both AO and MA, the sequences were more flexible than expected: $\langle\beta^2\rangle_{calc} < \langle\beta^2\rangle_{expt}$, as shown in Table 5.11. A change in the net tilt angle, θ_{tilt}, between the spin label and the DNA, may explain this result. The dot product of two vectors $\vec{A}$ and $\vec{B}$ is defined as $\vec{A}\cdot\vec{B} = |A||B|\cos\theta$, where θ is the angle between the vectors. Here, $\theta_{tilt} = \cos^{-1}\left(\frac{\vec{v}_{sl}\cdot\vec{v}_{DNA}}{|v_{sl}||v_{DNA}|}\right)$. If the orientation of the DNA ($\vec{v}_{DNA}$) changes, then θ_{tilt} changes also. In particular, A_5 is believed to induce a 20° bend in the DNA molecule.[10] As more bends are introduced (AA vs. MA) or as the position of the bends changes (AA vs. AO), the overall orientation of the entire DNA molecule ($\vec{v}_{DNA}$) with respect to the spin label ($\vec{v}_{sl}$) changes and must be included in the simulation to accurately depict the internal DNA dynamics. See Figure 5.4.

A simplified calculation of the vector $\vec{v}_{DNA}$ follows. The orientation of each base pair in the DNA duplex (N = 50 bp) with respect to the base pair immediately before it is given by the assumption of 10 bp / turn = 10 bp / 360° or 36° per base pair.[11] That is, the i^{th} base pair is rotated 36° from the i-1^{st} base pair and is 1 base pair further along the DNA helix axis (z), such that a right-handed helix is formed. This can be written as $angle_i$ = [18 0 18], where the three numbers represent Euler angles.[12] For an A-tract base pair, assuming that the 20° bend for A_5 can be evenly divided into a 4° bend per A, the angles are assigned as [18 4 18]. The (3x50)

matrices of Euler angles for the sequences AA, AO, MA, and a non A-tract sequence are given schematically in Table 5.12. The coordinates of each base pair, in the coordinate system of the first base pair, were calculated by repeated application of the Euler angle rotations, starting from the N^{th} (= 50^{th}) base pair and working backwards to the first base pair.

Using these coordinates, the linear equations of the modified Kirkwood algorithm were solved and the modified Rotne-Prager-Yamakawa-Oseen tensor was calculated.[13] This calculation used the "mesoscopic model," in which subunits of diameter 31.8 Å (radius 15.6 Å) replace the individual base pairs (31.8 Å per subunit / 3.4 Å per base pair ≈ 9 base pairs per subunit). This 6x6 tensor is the 6x6 friction factor matrix and is related to the diffusion tensors by $D = \frac{k_b T}{ff}$. The upper left 3x3 submatrix of this 6x6 friction factor matrix gives the translational friction factors; the lower right 3x3 submatrix gives the rotational friction factors. Diagonalizing the rotational friction factor matrix yields the eigenvectors $\bar{v}'$ and the eigenvalues d. Rotation of the matrices of eigenvectors and eigenvalues by matrices R such that the eigenvalues were in the order $\begin{vmatrix} \tau_{perp} & 0 & 0 \\ 0 & \tau_{perp} & 0 \\ 0 & 0 & \tau_{para} \end{vmatrix}$ was required. The vectors $\bar{v}_{DNA}$ were calculated with

$$\bar{v}_{DNA} = \bar{v}' \times R \times \begin{vmatrix} 0 \\ 0 \\ 1 \end{vmatrix}.$$

$\bar{v}_{sl}$ was defined as [6×18=108 20 108]. If the DNA helix were unwound to resemble a railroad track, both the spin label and A-tract bend would point the same direction, because θ, the middle angle, was positive for both the spin label and the A-tract. The spin label is believed to point into the major groove;[14] therefore the A-tract is believed also to bend into the major groove. Although gel-mobility and cyclization evidence has indicated that the A-tract bends into the minor groove and x-ray crystallographic evidence has suggested no bend at all, recent experiments in the literature show that the dehydrating agents used in x-ray crystallography greatly reduce curvature and may influence the (non-) bending seen in x-ray crystallography.[15]

Using $\bar{v}_{sl}$= [108 20 108], the tilt angles θ_{tilt} were calculated: $\theta_{tilt} = \cos^{-1}\left(\frac{\bar{v}_{sl} \cdot \bar{v}_{DNA}}{|v_{sl}||v_{DNA}|}\right)$.

Results are given in Table 5.13. AA is bent from the helix axis by about 8°, coincident with the spin label, so the net $\theta_{tilt} \approx 12°$. The shifted sequence, AO, is bent about 6° in the opposite direction, so $\theta_{tilt} \approx 26°$. The lengthened A-tract sequence, MA, is nearly coincident with the z-axis (off by 3°), so $\theta_{tilt} \approx 17°$ (Figure 5.4). When recalculated with these modified θ_{tilt}, $\langle\beta^2(AA)\rangle$

$\approx \langle\beta^2(AO)\rangle \approx \langle\beta^2(MA)\rangle$: the underlying internal dynamics are the same for the three sequences, in agreement with the calculation (Table 5.11). This analysis merely indicates that including an A-tract bend decreases the discrepancy between experiment and calculation and is not intended as a rigorous examination of the A-tract bend.

The more sophisticated force constant model used the force constants for every dinucleotide step to calculate $\langle\beta^2\rangle$. This model was used to try to distinguish between AT and TA, AC and CA, AG and GA, and CG and GC and to generate a self-consistent set of force constant ratios κ'/κ for the dinucleotide steps. ($\langle\beta^2\rangle$ calculated with the modified θ_{tilt} for AA, MA, and AO was used, $\langle\beta^2(AA)\rangle = \langle\beta^2(MA)\rangle = \langle\beta^2(AO)\rangle = 0.0882$.) Several different "optimal" matrices were found, underscoring the uncertainty of this approach. These matrices, each with $\sigma^2 \approx 4\times10^{-6}$, are listed in Table 5.14. One matrix, fcsca2, was done by hand. The others were computer-optimized to minimize the difference between experimental and calculated $\langle\beta^2\rangle$, using different starting matrices and using varied numbers of data points. fconea ("force constants - ones - trial a") and fconeb started with matrices of all ones, using 9 and 12 sequences in the optimization. The sequences used in the optimization were AA, AC, AG, AT, CC, CG, AGC, ATC, and ACT, plus CCG, AAG, and ACC for the 12 sequence set. fdia ("force constants - dinucleotides - trial a") and fdib came from optimizing the dinucleotide-only matrix, fdi (see Table 5.8), using 12 and 6 (dinucleotides only) sequences. fdiba and fdibb were further refinements of fdib, using 9 and 12 sequences. Optimization of the force constants with 27 sequences (sequences from 12 sequence set plus NT, AAT, AAC, AGG, ACG, CAP, TATA, AO, TO, MA, MT, AU, AV, SA, and LA) followed similar trends but became unstable ($\frac{\kappa'}{\kappa}$(GC) $\geq$ 4).

Errorbars were calculated in two ways. First, for each matrix, the uncertainty of each element in the matrix was calculated. For each element, $\sigma^2(a_0)$ was calculated at $\frac{\kappa'}{\kappa} = a_0$, $a_0+\Delta$, and $a_0-\Delta$, holding all other elements fixed, where a_0 is the value of the element as given in the matrix and $\Delta = 0.1$. For example, for fconeb, σ^2 was calculated at $\frac{\kappa'}{\kappa}$(AT) = 0.7952, 0.8052, and 0.7852. The uncertainty in a_0 was $\sigma^2_{a_0} = \frac{\sigma^2(a_0)}{\left(\frac{2\sigma^2(a_0)-(\sigma^2(a_0-\Delta)+\sigma^2(a_0+\Delta))}{\Delta^2}\right)}$. A larger $\sigma^2(a_0)$ indicates greater uncertainty in a_0; a smaller $\sigma^2(a_0)$ indicates less uncertainty. These $\sigma^2(a_0)$ are reported for each element in each matrix in Table 5.14. For comparison, the errors were also calculated for a matrix of all ones, shown in Table 5.15. The larger $\sigma^2(a_0)$ for CT =

AG, GC, CA, and CG suggest an intrinsic uncertainty for these force constant ratios, perhaps due to uneven representation of the dinucleotide steps in this data set.

Second, the matrices fcsca2, fconea, fconeb, fdia, fdib, fdiba, and fdibb were pooled and the element-by-element averages and standard deviations were calculated (Table 5.16). These standard deviations give a sense of how broad a range of values was found (in these 7 matrices) for each κ'/κ element and are shown schematically in Figure 5.5. The TT = AA dinucleotide step had the least uncertainty; the GC step had the greatest uncertainty.

A consensus ranking of the flexibilities of the dinucleotide steps, using these 7 matrices, is given in Table 5.17. $\frac{\kappa'}{\kappa}$(GC) > $\frac{\kappa'}{\kappa}$(CG) (5 times out of 7), so GC is more rigid than CG. $\frac{\kappa'}{\kappa}$(CC) > $\frac{\kappa'}{\kappa}$(TT) (7/7): CC is more rigid that TT. Together, $\frac{\kappa'}{\kappa}$(AT) and $\frac{\kappa'}{\kappa}$(TA) < $\frac{\kappa'}{\kappa}$(GC) and $\frac{\kappa'}{\kappa}$(CG) (7/7). The nearly even split for AT vs. TA ($\frac{\kappa'}{\kappa}$(AT) > $\frac{\kappa'}{\kappa}$(TA), 4/7) is shown by the overlap of errorbars shown in Figure 5.5. The hand-done matrix shows $\frac{\kappa'}{\kappa}$(CA) > $\frac{\kappa'}{\kappa}$(AC) and $\frac{\kappa'}{\kappa}$(AG) > $\frac{\kappa'}{\kappa}$(GA). However, the consensus from the computer-optimized matrices is the opposite: $\frac{\kappa'}{\kappa}$(AC) > $\frac{\kappa'}{\kappa}$(CA) (5/6) and $\frac{\kappa'}{\kappa}$(GA) > $\frac{\kappa'}{\kappa}$(AG) (5/6), although the large errorbars (Figure 5.5) for these $\frac{\kappa'}{\kappa}$ ratios for these 4 steps do indicate considerable uncertainty in the ratios. Finally, base pairs with a leading G (GT = AC, GA, GC) were more rigid dinucleotide steps than their counterparts (CA = TG, AG, CG). This phenomenon is not understood and might be a reflection of DNA dynamics but again might be related to the representation of the dinucleotide steps in the sequences used for optimization. Perhaps a series of DNA sequences with a more even representation of the dinucleotide steps would be helpful.

Although general trends are observed from this approach, the uncertainties in the $\frac{\kappa'}{\kappa}$ ratios indicate that the dinucleotide step model is not the complete explanation for sequence dependence of internal DNA dynamics. Some kind of longer-range effect is also involved.

SECTION 5.5. CONCLUSIONS

DNA molecules spin labeled with Q were used to systematically study the dependence of internal DNA dynamics upon base sequence. The flexibilties of the dinucleotide and trinucleotide repeats were ranked. Seven of the ten trinucleotide repeats had flexibilities dependent in some way on the flexibilities of the component dinucleotide steps; the remaining three trinucleotide repeats all were more flexible than a dinucleotide step model would predict. The flexibilities for

the single-blind samples MN and YZ agreed with the flexibilities for their true sequences CG and AC. The similarity in flexibility of the TG and AC repeats suggested that the DNA molecule is isotropic with respect to the two chains comprising the helix. Moving the flexible region away from the spin label reduced the effect of the flexibility, and the results were satisfactorily if simplistically modeled using 1 region of flexibility κ' sandwiched between two regions of flexibility κ. The same model of local flexibility satisfactorily predicted the results of increasing the length of the flexible region: as the flexible region increased in length, the flexibility as seen by the spin label increased as well. The TATA box and CAP half-site sequences were nearly as flexible as the equivalent (short) AT repeat, providing evidence that internal collective bending dynamics may be involved in the presumed flexibility of these two particular sequences in DNA-protein interactions. Assuming and including the A_5-induced 20° bend revealed that neither introducing more bends nor phasing the bends changed the internal DNA dynamics; that is, the measure of internal dynamics was not compromised by structural changes.

This study represents merely the beginning of more detailed work on the dependence of internal DNA dynamics on base sequence. For example, Harrington has suggested that GGGCCC may induce a bend as did A_5.[16] This proposal could be tested by examining the lengthened and shifted GGGCCC repeats (analogous to the AA repeats). Furthermore, the TT repeat should be examined as well, as TT and AA are not the same if the A-tract induces a bend. Using T_{15} on the Q-labeled strand might show a bend in the opposite direction. Also, if the unique tetranucleotides were studied, pooling that information with the dinucleotide and trinucleotide data might permit distinguishing the trinucleotide steps (ACT vs. ATC for example) and determining whether the trinucleotides or the dinucleotides are the fundamental repeating step. This series of experiments should be complemented with computer simulation using a variation on the more sophisticated force constant model described above. Deoxygenation of the samples would remove one source of line-broadening (collisions with paramagnetic oxygen) and might make the spectral differences less subtle. Additional experiments of the sequences studied here in 50% sucrose at 0°C would either confirm or disprove the assumption that the sequences can be fit with 1 set of tensors in the "absence" of dynamics. Finally, the aging phenomenon could be tracked and studied more carefully, and, ideally, eliminated.

NOTES TO CHAPTER 5

1. (a) von Hippel, P.V. *Science* **263**, 769-770 (1994); (b) van der Vliet, P.C., Verrijzer, C.P. *Bioessays* **15**, 25-32 (1993).

2. Schultz, S.C., Shields, G.C., & Steitz, T.A. *Science* **253**, 1001-1007 (1991); Kim, Y., Geiger, J.H., Hanh, S., & Sigler, P.B. *Nature* **365**,512-520 (1993); Kim, J.L., Nicolav, D.B., & Burley, S.K. *Nature* **365**, 520-527 (1993).

3. Lyubchenko, Y.L., Shlyakhtenko, L.S., Appella, E., & Harrington, R.E. *Biochemistry* **32**, 4121-4127 (1993), and others.

4. Maniatis, T., Fritsch, E.F., & Sambrook, J., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory (1982); Alley, S.C. Ph.D. Thesis, University of Washington, 1996.

5. Hustedt, E.J. Ph.D. Thesis, University of Washington, 1989;, Hustedt, E.J., Spaltenstein, A., Kirchner, J.J., Hopkins, P.B., & Robinson, B.H. *Biochemistry* **32**, 1774-1787 (1993).

$$\frac{2OD \times 50\frac{\mu g}{OD}}{10\ \mu l} = \frac{100\ \mu g}{10\ \mu l} = 10\ \mu g/\mu l = 10\ mg/ml$$

$$\frac{2OD \times 50\frac{\mu g}{OD}}{330\frac{\mu g}{\mu mol \cdot bp} \times 50bp} = 0.006\mu mol \qquad \frac{0.006\mu mol}{10\mu l} = 0.6mM\ DNA.$$

6. Allison, S.A.. & Schurr, J.M. *Chem. Phys.* **41**, 35-59 (1979); Wu, P., Fujimoto, B.S., & Schurr, J.M. *Biopolymers* **26**, 1463-1488 (1987); Fujimoto, B.S. & Schurr, *Nature* **344**, 175-178 (1990); Song, L. & Schurr, J.M. *Biopolymers* **30**, 229-237 (1990).

7. Given $\sigma_\beta = 0.003$, σ_{β^2} may be calculated with Gauss's rule:

$$\sigma_{\beta^2}^2 = \left(\frac{\partial(\beta^2)}{\partial\beta}\right)^2 \sigma_\beta^2 = (2\beta)^2\sigma_\beta^2 = 4\beta^2\sigma_\beta^2 = 4\beta^2(0.003)^2 \text{ so that } \sigma_{\beta^2} = \sqrt{4\beta^2(0.003)^2}\ .$$

8. The matrix A used for the simplistic model of a DNA molecule with 2 regions of flexibility, characterized by force constants κ and κ':

$$A = \begin{pmatrix}
\frac{\kappa}{\kappa} & -\frac{\kappa}{\kappa} & 0 & \cdots & & & & & & & 0 \\
-\frac{\kappa}{\kappa} & 2\frac{\kappa}{\kappa} & \ddots & 0 & \cdots & & & & & & 0 \\
0 & -\frac{\kappa}{\kappa} & \ddots & -\frac{\kappa}{\kappa} & 0 & \cdots & & & & & \vdots \\
\vdots & 0 & \ddots & 2\frac{\kappa}{\kappa} & -\frac{\kappa}{\kappa} & 0 & \cdots & & & & \\
 & \vdots & 0 & -\frac{\kappa}{\kappa} & \frac{\kappa+\kappa'}{\kappa} & -\frac{\kappa'}{\kappa} & 0 & & & & \\
 & & & 0 & -\frac{\kappa'}{\kappa} & 2\frac{\kappa'}{\kappa} & \ddots & 0 & & & \\
 & & & & 0 & -\frac{\kappa'}{\kappa} & \ddots & -\frac{\kappa'}{\kappa} & 0 & & \\
 & & & & & 0 & \ddots & \frac{\kappa+\kappa'}{\kappa} & -\frac{\kappa}{\kappa} & 0 & \\
 & & & & & & 0 & -\frac{\kappa}{\kappa} & 2\frac{\kappa}{\kappa} & \ddots & 0 \\
 & & & & & & & 0 & -\frac{\kappa}{\kappa} & \ddots & -\frac{\kappa}{\kappa} \\
 & & & & & & & & 0 & \ddots & 2\frac{\kappa}{\kappa} \\
0 & \cdots & & & & & & & & 0 & -\frac{\kappa}{\kappa}
\end{pmatrix}$$

$$= \begin{pmatrix}
1 & -1 & 0 & \cdots & & & & & & & 0 \\
-1 & 2 & \ddots & 0 & \cdots & & & & & & 0 \\
0 & -1 & \ddots & -1 & 0 & \cdots & & & & & \vdots \\
\vdots & 0 & \ddots & 2 & -1 & 0 & \cdots & & & & \\
 & \vdots & 0 & -1 & \frac{\kappa+\kappa'}{\kappa} & -\frac{\kappa'}{\kappa} & 0 & & & & \\
 & & & 0 & -\frac{\kappa'}{\kappa} & 2\frac{\kappa'}{\kappa} & \ddots & 0 & & & \\
 & & & & 0 & -\frac{\kappa'}{\kappa} & \ddots & -\frac{\kappa'}{\kappa} & 0 & & \\
 & & & & & 0 & \ddots & \frac{\kappa+\kappa'}{\kappa} & -1 & 0 & \\
 & & & & & & 0 & -1 & 2 & \ddots & 0 \\
 & & & & & & & 0 & -1 & \ddots & -1 \\
 & & & & & & & & 0 & \ddots & 2 \\
0 & \cdots & & & & & & & & 0 & -1 \\
 & & & & & & & & & -1 & 1
\end{pmatrix}$$

9. The matrix A used for the more sophisticated model of a DNA molecule (length N base pairs) using the N force constants (for each dinucleotide step):

$$A = \begin{pmatrix} \frac{\kappa_1}{\kappa_1} & -\frac{\kappa_1}{\kappa_1} & 0 & \cdots & & & 0 \\ -\frac{\kappa_1}{\kappa_1} & \frac{\kappa_1+\kappa_2}{\kappa_1} & -\frac{\kappa_2}{\kappa_1} & \ddots & & & \vdots \\ 0 & -\frac{\kappa_2}{\kappa_1} & \frac{\kappa_2+\kappa_3}{\kappa_1} & -\frac{\kappa_3}{\kappa_1} & & & \\ \vdots & \ddots & -\frac{\kappa_3}{\kappa_1} & \frac{\kappa_3+\kappa_4}{\kappa_1} & \ddots & & \\ & & & -\frac{\kappa_4}{\kappa_1} & \ddots & -\frac{\kappa_{N-1}}{\kappa_1} & 0 \\ & & & & \ddots & \frac{\kappa_{N-1}+\kappa_N}{\kappa_1} & -\frac{\kappa_N}{\kappa_1} \\ 0 & \cdots & & & 0 & -\frac{\kappa_N}{\kappa_1} & \frac{\kappa_N}{\kappa_1} \end{pmatrix}$$

10. Koo, H.S., Woo, H.M., & Crothers, D.M. *Nature* **320**, 501-506 (1986).

11. A more accurate estimate is 10.5 ± 0.1 base pairs / turn for natural DNA and 10.1 ± 0.1 base pairs / turn for A tracts, but 10 base pairs / turn was used for simplicity. Saenger, W., *Principles of Nucleic Acid Structure*, in the Springer Advanced Texts in Chemistry series, series ed., Charles R. Cantor, p. 271. New York: Springer-Verlag (1984).

12. Goldstein, H. *Classical Mechanics*, pp 107-109. Reading MA: Addison-Wesley (1980). The Euler angles are defined as in Goldstein, with rotation of the axes by ψ about the z axis, followed by rotation by θ about the new y axis, and finally rotation β about the new z axis. The rotation matrices are $\boldsymbol{R} = \boldsymbol{R}_\psi \cdot \boldsymbol{R}_\theta \cdot \boldsymbol{R}_\beta$, where

$$\boldsymbol{R}_\psi = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \boldsymbol{R}_\theta = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}, \text{ and } \boldsymbol{R}_\beta = \begin{pmatrix} \cos\beta & \sin\beta & 0 \\ -\sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

13. Bernal, J.M.G. & de la Torre, J.G. *Biopolymers* **19**, 751-766 (1980); Gebe, J.A. Ph.D. Thesis, University of Washington, 1995.

14. Huang, H., Miller, T.R., Duncan, J.A., Alley, S.C., Zhu, L., Reid, B.R., & Hopkins, P.B., in preparation.

15. Sprous, D., Zacharias, W., Wood, Z.A., & Harvey, S.C. *Nucleic Acids Research* **23**, 1816-1821 (1995); Harvey, S.C., Diakic, M., Griffith, J., Harrington, R., Park, K., Sprous, D., & Zacharias, W. *J. Biomol. Struct. Dyn.* **13**, 301-307 (1996).

16. Diakic, M. & Harrington, R.E. *PNAS* **93**, 3847-3852 (1996).

Figure 5.1. Schematic of testing sequences, with control sequence listed as an example. The label region extends from position 1 to position 11; the test region extends from position 12 to 26, and the tail completes the sequence to 50 base pairs. The sequence was matched to its P-containing complement to form duplex DNA.

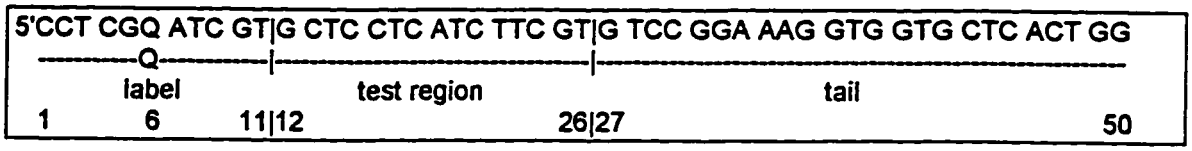


Table 5.1. Duplex DNA sequences studied.

sequence	name	category	start of test region	test region length
control	NT	control	12	15
$(AA)_7A$, $(AC)_7A$, $(AG)_7A$, $(AT)_7A$, $(CC)_7C$, $(CG)_7C$	AA, AC, AG AT, CC, CG	dinucleotide repeats	12	15
$(XX)_7X$	XX	random sequence	12	15
$(MN)_7M$, $(YZ)_7Y$	MN, YZ	blind samples	12	15
$(TG)_7T$	TG	inverse of AC	12	15
$(AT)_4$, $(AT)_{15}$	SA, LA	length	12	8,30
$(AT)_7A$, $(AT)_7A$	AU, AV	position	24,36	15
$(CAG)_5$, $(CGG)_5$, $(AAT)_5$, $(AAG)_5$, $(AAC)_5$, $(ACC)_5$, $(ACG)_5$, $(ACT)_5$, $(AGG)_5$, $(ATC)_5$	CAG, CGG, AAT, AAG, AAC, ACC, ACG, ACT, AGG, ATC	trinucleotide repeats	12	15
$(TAX)_5$, $(ATX)_5$	TAX, ATX	AT vs. TA	12	15
TATA TAAA	TATA	TATA box	12	8
TGTG ACAT	CAP	CAP half-site	12	8
$(AA)_{10}$, $(AT)_{10}$	MA, MT	2 full turns	12	20
$(AA)_7A$, $(AT)_7A$ shift 5	AO, TO	shifted 5 bp downstream	17	15

Table 5.2. Representative sequences, with the varied piece in **bold** type. Each sequence was matched with its P-containing complement to form duplex 50-mers.

name	sequence
NT	5'CCT CGQ ATC GT**G CTC CTC ATC TTC GTG** TCC GGA AAG GTG GTG CTC ACT GG
AC	5'CCT CGQ ATC GTA **CAC ACA CAC ACA ACG** TCC GGA AAG GTG GTG CTC ACT GG
TG	5'CCT CGQ ATC GTT **GTG TGT GTG TGT GTG** TCC GGA AAG GTG GTG CTC ACT GG
TATA	5'CCT CGQ ATC GTT **ATA TAA ATC** TTC GTG TCC GGA AAG GTG GTG CTC ACT GG
CAP	5'CCT CGQ ATC GTT **GTG ACA** TTC TTC GTG TCC GGA AAG GTG GTG CTC ACT GG
SA	5'CCT CGQ ATC GTA **TAT ATA** TTC TTC GTG TCC GGA AAG GTG GTG CTC ACT GG
AT	5'CCT CGQ ATC GTA **TAT ATA TAT ATA TAG** TCC GGA AAG GTG GTG CTC ACT GG
LA	5'CCT CGQ ATC GTA **TAT ATA TAT ATA TAT ATA TAT ATA TAT ATG** CTC ACT GG
AU	5'CCT CGQ ATC GTG CTC CTC ATC TTA **TAT ATA TAT ATA TAG** GTG CTC ACT GG
AV	5'CCT CGQ ATC GTG CTC CTC ATC TTC GTG TCC GGA AA**A TAT ATA TAT ATA TA**
AA	5'CCT CGQ ATC GTA **AAA AAA AAA AAA AAG** TCC GGA AAG GTG GTG CTC ACT GG
AO	5'CCT CGQ ATC GTG CTC **CAA AAA AAA AAA AAA AGA** AAG GTG GTG CTC ACT GG
MA	5'CCT CGQ ATC GTA **AAA AAA AAA AAA AAA AAA AGA** AAG GTG GTG CTC ACT GG

Table 5.3. Potentially grossly distorted and control sequences. S represents a propyl linker.

gap control (gac)	5'GAC CTC GQA TCG TGC TCC TCA TCT TCG TGT CCG GAA AGG TGG TGC - CTG GAG CPT AGC ACG AGG AGT AGA AGC ACA GGC CTT TCC ACC ACG -TCA CTG GAG CAA AGG AGT GAC CTC GTT TCC
gap	5'GAC CTC GQA TCG TGC CCG GAA AGG TGG TGC - CTG GAG CPT AGC ACG AGG AGT AGA AGC ACA GGC CTT TCC ACC ACG -TCA CTG GAG CAA AGG AGT GAC CTC GTT TCC
control	5'CCT CGQ ATC GT GC TCC TCA TCT TCG TGT CCG GAA AGG TGG TGC - GGA GCP TAG CA CG AGG AGT AGA AGC ACA GGC CTT TCC ACC ACG -TCA CTG G AGT GAC C
floppy	5'CCT CGQ ATC GTS SGC TCC TCA TCT TCG TGT CCG GAA AGG TGG TGC - GGA GCP TAG CAS SCG AGG AGT AGA AGC ACA GGC CTT TCC ACC ACG -TCA CTG G AGT GAC C

Table 5.4. Results of simulations for potentially grossly distorted and control sequences (sequences given in Table 5.3). Sequence length is N base pairs; N' = 2 × n, where n is the length of the DNA segment before the discontinuity, n = 15 for gap, 11 for floppy. Simulations used global dynamics for the N-mers and N'-mers (see Table 3.2), 0.4 G homogeneous linewidth, 0.5 G Gaussian broadening, and the standard tensor set given in Table 3.1. R^2 is the correlation coefficient between data and simulation; R = 1 represents a perfect fit. $\langle\beta^2\rangle$ measures the amplitude of internal dynamics.

	N	$\langle\beta^2\rangle$	R^2	N'	$\langle\beta^2\rangle$	R^2
gap control	60	0.0784 ± 0.0017	0.9579			
gap	60	0.1024 ± 0.0019	0.8836	30	0.0562 ± 0.0014	0.9709
control	50	0.0854 ± 0.0018	0.9841			
floppy	50	0.1444 ± 0.0023	0.8282	22	0.0784 ± 0.0017	0.9401

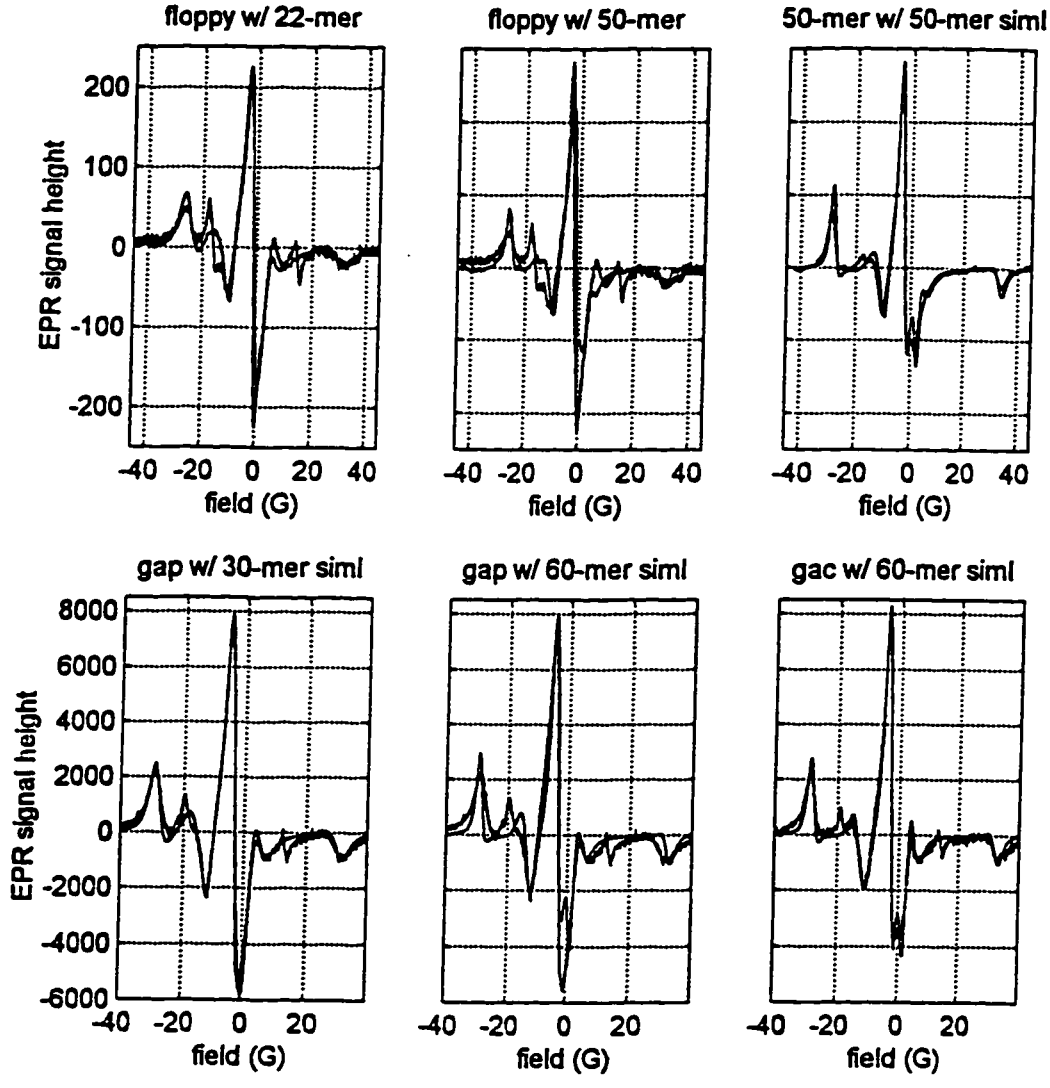


Figure 5.2. Experimental spectra of floppy and gap overlaid with simulations (for gap) and experimental (for floppy) spectra of shorter and full length DNAs.

Table 5.5. Tensors determined for $(AA)_7A$, $(AT)_7A$, and NT in 50% sucrose in PNE at 0°C. R^2 is the correlation coefficient between data and simulation; R = 1 represents a perfect fit.

	standard	AA alone	AT alone	AA, AT pooled	NT alone
gxx	2.0084	2.0084	2.0081	2.0081	2.0084
gyy	2.0068	2.0064	2.0061	2.0061	2.0064
gzz	2.0034	2.0029	2.0027	2.0026	2.0030
Axx	9.22	8.83	8.82	8.82	9.11
Ayy	6.98	8.64	8.64	8.64	8.91
Azz	48.00	48.03	47.99	48.01	47.69
R^2	0.9330(AA) 0.9455(AT)	0.9808	0.9789	0.9808(AA) 0.9789(AT)	0.9776

Table 5.6. $\langle\beta^2\rangle$ as a function of time to illustrate "aging" phenomenon. AA and AT were refreshed on 5/9; the remaining samples were refreshed on 5/21.

date	1/12	1/18	3/21	3/25	4/4	4/9	4/30	5/1
days	2	8	71	75	85	90	111	112
AA	0.0858	0.0841	0.0812	0.0824	0.0812	0.0807	0.0807	0.0807
AC	0.0888	0.0900	0.0870	0.0835	0.0841	0.0835	0.0807	-
AG	0.0918	0.0900	0.0870	0.0930	-	0.0870	0.0823	0.0823
AT	0.0955	0.0949	0.0918	0.0942	0.0900	0.0900	0.0756	0.0676
CC	0.0858	0.0841	0.0858	0.0864	-	0.0835	0.0824	-
CG	0.0876	0.0841	0.0829	0.0853	-	0.0812	-	-
NT	0.0858	0.0841	0.0858	0.0930	-	-	0.0835	-
XX	0.0882	0.0900	0.0870	0.0900	-	-	0.0864	-

date	5/2	5/10	5/22	5/29	5/31	6/12	6/24	7/12
days	113	121	133	140	142	154	166	170
AA	0.0804	0.0824	0.0824	0.0824	0.0818	0.0824	0.0824	0.0812
AC	-	0.0762	0.0870	0.0870	-	0.0882	-	-
AG	-	0.0740	0.0900	0.0900	-	0.0900	-	-
AT	0.0625	0.0942	0.0936	0.0936	0.0955	0.0942	-	-
CC	-	0.0841	0.0853	0.0858	-	0.0858	-	-
CG	-	0.0671	0.0841	0.0841	-	-	-	-
NT	-	-	0.0853	0.0858	0.0841	-	-	-
XX	-	-	0.0870	-	0.0841	-	-	-

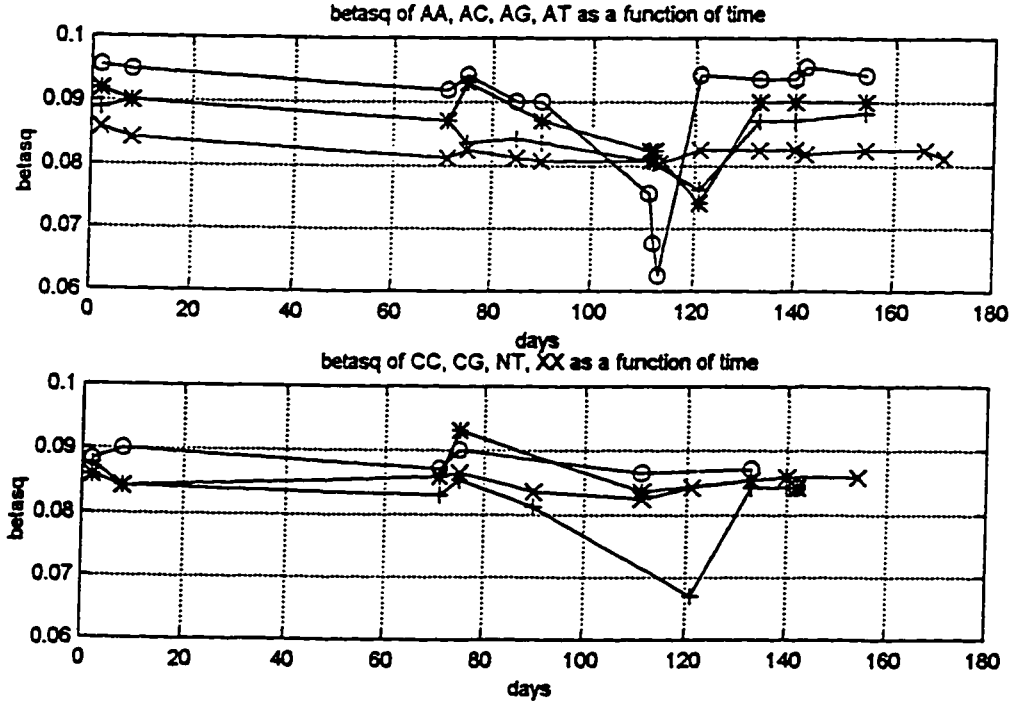


Figure 5.3. $\langle\beta^2\rangle$ as a function of time (days since the samples were made) to illustrate "aging" phenomenon. a) $\langle\beta^2\rangle$ for AA (x), AC (+), AG (*), and AT (o); b) $\langle\beta^2\rangle$ for CC (x), CG (+), NT (*), and XX (o). The AA and AT samples were refreshed on May 9 (day 120), and $\langle\beta^2\rangle$ for AT rose dramatically. The remaining samples were refreshed on May 21 (day 132).

Table 5.7. $\langle\beta^2\rangle$ as a function of sequence, using $\theta_{tilt} = 20°$. $\theta_{tilt} \neq 20°$ for values in *italics* (see Table 5.11). # points = number of independent measurements of $\langle\beta^2\rangle$ for each sequence.

name	# points	mean($\langle\beta^2\rangle$)	name	# points	mean($\langle\beta^2\rangle$)
AA	16	0.0820 ± 0.0017	CAG	3	0.0876 ± 0.0021
AA	*16*	*0.0882 ± 0.0018*	CGG	4	0.0856 ± 0.0018
AC	9	0.0866 ± 0.0024	TAX	3	0.0900 ± 0.0018
AG	8	0.0899 ± 0.0021	ATX	3	0.0912 ± 0.0018
AT	11	0.0934 ± 0.0020	TATA	3	0.0888 ± 0.0018
CC	10	0.0849 ± 0.0017	CAP	3	0.0882 ± 0.0018
CG	8	0.0842 ± 0.0018	AAT	4	0.0922 ± 0.0018
NT	9	0.0854 ± 0.0018	AAG	3	0.0882 ± 0.0018
XX	6	0.0881 ± 0.0018	ACC	2	0.0861 ± 0.0021
			AAC	3	0.0900 ± 0.0025
AU	3	0.0872 ± 0.0018	ACG	3	0.0916 ± 0.0018
AV	3	0.0853 ± 0.0018	ACT	2	0.0906 ± 0.0034
			AGG	2	0.0921 ± 0.0030
SA	2	0.0900 ± 0.0018	ATC	2	0.0885 ± 0.0021
LA	3	0.0949 ± 0.0018			
			MN (=CG)	3	0.0847 ± 0.0017
MA	4	0.0876 ± 0.0018	YZ (=AC)	3	0.0880 ± 0.0018
MA	*4*	*0.0882 ± 0.0019*			
MT	4	0.0961 ± 0.0019	AO	2	0.0921 ± 0.0018
			AO	2	*0.0882 ± 0.0023*
TG	3	0.0878 ± 0.0018	TO	2	0.0961 ± 0.0019

Table 5.8. Ranking of the dinucleotide and trinucleotide repeats using $\langle\beta^2\rangle$ calculated with $\theta_{tilt} = 20°$. $\langle\beta^2\rangle$ for AA using $\theta_{tilt} \neq 20°$ is reported in *italics* (see Table 5.11).

name	mean($\langle\beta^2\rangle$)	mean($\langle\beta^2\rangle$)	name
AA	0.0820		
CG	0.0842		
CC	0.0849		
		0.0856	CGG
		0.0861	ACC
AC	0.0866		
		0.0876	CAG
AA	*0.0882*	0.0882	AAG
		0.0885	ATC
AG	0.0899		
		0.0900	AAC
		0.0906	ACT
		0.0916	ACG
		0.0921	AGG
		0.0922	AAT
AT	0.0934		

Table 5.9. $\langle\beta^2\rangle$, force constant ratios κ'/κ, persistence length P_{df}, and maximum possible θ for dinucleotide repeats, assuming P_{df} = 1250 Å for κ'/κ=1. $\kappa = (P_{df} / h) \times k_b \times T = 1.537\times10^{-11}$ erg for h = 3.4 Å, $k_b = 1.38\times10^{-16}$ erg/K, and T = 303K. θ was calculated from equation 4, $\theta = \sqrt{\langle\theta^2\rangle} = \sqrt{\frac{k_b T}{\kappa}}$. Calculated values for AA using $\theta_{tilt} \neq 20°$ are reported in *italics* (see Table 5.11).

name	$\langle\beta^2\rangle$	κ'/κ	P_{df} (Å)	θ (rad)	θ (deg)
AA	0.0820	1.140	1430	0.0488	2.80
CG	0.0842	1.045	1310	0.0510	2.92
CC	0.0849	1.020	1280	0.0516	2.96
NT	0.0854	1.000	1250	0.0522	2.99
AC	0.0866	0.963	1200	0.0531	3.04
AA	*0.0882*	*0.908*	*1140*	*0.0547*	*3.14*
AG	0.0899	0.858	1070	0.0563	3.23
AT	0.0934	0.772	970	0.0594	3.40

Table 5.10. Experimental and calculated values of $\langle\beta^2\rangle$ for AT derivatives, using κ'/κ = 0.772.

name	$\langle\beta^2\rangle_{expt}$	$\langle\beta^2\rangle_{calc}$	$\langle\beta^2\rangle_{expt}$-$\langle\beta^2\rangle_{calc}$
SA	0.0900 ± 0.0018	0.0906	-0.0006
AT	0.0934 ± 0.0020	0.0934	0.0000
LA	0.0949 ± 0.0018	0.0967	-0.0018
AT	0.0934 ± 0.0020	0.0934	0.0000
AU	0.0872 ± 0.0018	0.0888	-0.0016
AV	0.0853 ± 0.0018	0.0858	-0.0005

Table 5.11. Experimental and calculated values of $\langle\beta^2\rangle$ for A-tract samples. $\Delta\langle\beta^2\rangle = \langle\beta^2\rangle_{expt} - \langle\beta^2\rangle_{calc}$.

	$\theta_{tilt} = 20°$, $\kappa'/\kappa = 1.140$			$\theta_{tilt} \neq 20°$, $\kappa'/\kappa = 0.908$			
name	β_{expt}	β_{calc}	$\Delta\langle\beta^2\rangle$	θ_{tilt}	β_{expt}	β_{calc}	$\Delta\langle\beta^2\rangle$
AA	0.0820 ± 0.0017	0.0820	0.0000	12.2°	0.0882 ± 0.0018	0.0882	0.0000
MA	0.0876 ± 0.0018	0.0812	0.0064	16.5°	0.0882 ± 0.0019	0.0888	-0.0006
AO	0.0921 ± 0.0018	0.0829	0.0092	26.3°	0.0882 ± 0.0023	0.0876	0.0006

Table 5.12. Positions of elements in (3x50) matrices of Euler angles used to calculate $\bar{v}_{DNA}$ for A-tract sequences.

name	18 18 18 18 ... 18 18 18 18 ... 18 18 18 ... 18 18 18 ψ 0 0 0 0 ... 0 4 4 4 ... 4 0 0 ... 0 0 0 θ 18 18 18 18 ... 18 18 18 18 ... 18 18 18 ... 18 18 18 β		
no bend	1	11	12 50
AA	1	11 12	26 27 50
MA	1	11 12	31 32 50
AO	1	16 17	31 32 50

Table 5.13. Vectors ($\bar{v}_{DNA}$), angles that the overall DNA molecule is bent from the z axis, and θ_{tilt} for A-tract sequences.

	$\bar{v}_{DNA} = [x\ y\ z]$	angle bent from z	θ_{tilt}
no bend	[0 0 1]	0°	20.0°
AA	[-0.144 -0.006 0.990]	+8°	12.2°
MA	[0.004 -0.001 0.998]	+3°	16.5°
AO	[0.112 0.008 0.994]	-6°	26.3°

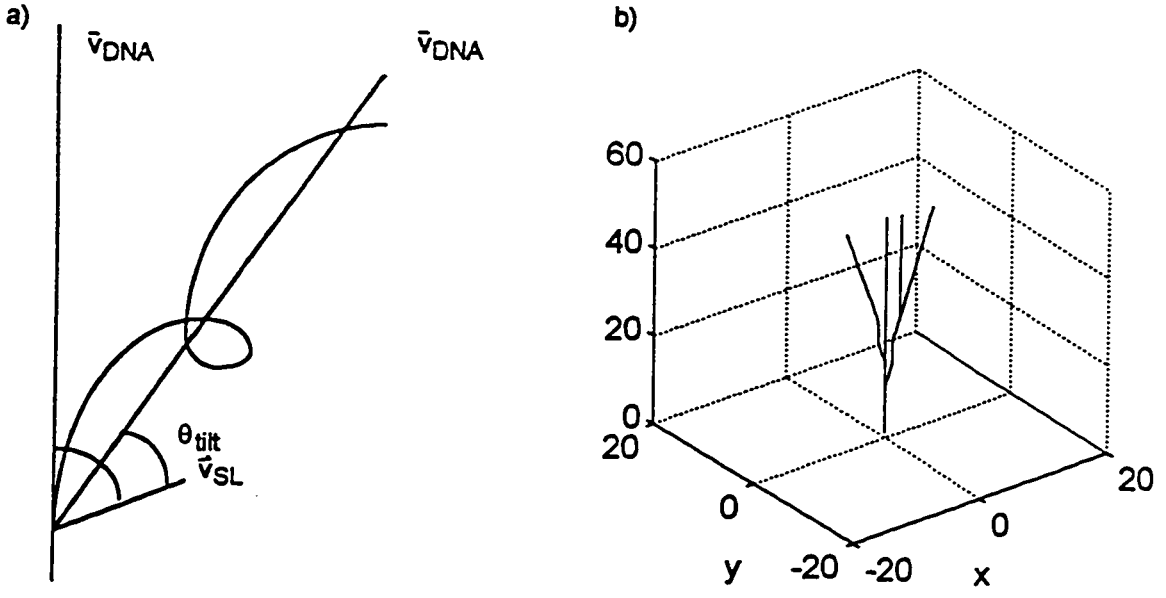


Figure 5.4. a) Schematic view of overall DNA orientation for $(AA)_7A$, $(AA)_7A$ shifted 5 bp, $(AA)_{10}$, and the control (no bend), assuming the A_5-induced 20° bend. b) Sketch of $\bar{v}_{DNA}$, $\bar{v}_{SL}$, and θ_{tilt}.

Table 5.14. Collection of force-constant matrices with $\sigma^2 = 4\times10^{-6}$. Elements of the matrices are the neighboring bases, not the base pairs, as follows:

AT	AA	AC	AG
TT	TA	TC	TG
GT	GA	GC	GG
CT	CA	CC	CG

Because of symmetry relationships (for example, AA = TT), only the lower diagonal matrix is used for calculations and optimization. The larger value for each comparison (AT vs. TA, GT vs. CA, CT vs. GA, and GC vs. CG) is in **bold** type.

fcsca2 $\sigma^2 = 4.05\times10^{-6}$

0.883 ± 0.102			
0.984 ± 0.116	0.793 ± 0.072		
1.037 ± 0.175	0.886 ± 0.108	**1.151 ± 0.502**	
0.948 ± 0.198	**1.118 ± 0.322**	1.118 ± 0.225	1.118 ± 0.256

fconea $\sigma^2 = 3.80\times10^{-6}$ 9 seq

0.821 ± 0.085			
0.977 ± 0.110	**0.858 ± 0.082**		
1.012 ± 0.158	**0.973 ± 0.125**	**1.137 ± 0.436**	
0.889 ± 0.167	**1.052 ± 0.257**	1.065 ± 0.189	1.056 ± 0.212

fconeb $\sigma^2 = 3.83\times10^{-6}$ 12 seq

0.795 ± 0.079			
0.978 ± 0.111	**0.874 ± 0.085**		
1.231 ± 0.235	**1.390 ± 0.258**	**1.746 ± 0.901**	
0.704 ± 0.101	0.861 ± 0.168	1.022 ± 0.170	0.738 ± 0.098

fdi $\sigma^2 = 2.77\times10^{-5}$

0.772 ± 0.158			
0.908 ± 0.214	0.772 ± 0.142		
0.963 ± 0.271	0.858 ± 0.195	1.045 ± 0.703	
0.858 ± 0.301	0.963 ± 0.399	1.020 ± 0.352	1.045 ± 0.371

fdia $\sigma^2 = 3.81\times10^{-6}$ 12 seq

0.827 ± 0.086			
0.980 ± 0.112	**0.840 ± 0.078**		
1.238 ± 0.236	**1.308 ± 0.227**	**1.630 ± 0.789**	
0.778 ± 0.108	0.861 ± 0.168	1.025 ± 0.170	0.762 ± 0.104

fdib $\sigma^2 = 4.19\times10^{-6}$ 6 seq

0.841 ± 0.093			
0.976 ± 0.116	0.835 ± 0.082		
1.041 ± 0.176	0.926 ± 0.118	1.071 ± 0.899	
0.927 ± 0.190	1.018 ± 0.249	1.066 ± 0.199	**1.119 ± 0.252**

fdiba $\sigma^2 = 4.21\times10^{-6}$ 9 seq

0.902 ± 0.109			
0.976 ± 0.116	0.779 ± 0.071		
1.072 ± 0.189	**0.998 ± 0.151**	1.047 ± 0.380	
0.869 ± 0.165	0.990 ± 0.240	1.045 ± 0.191	**1.097 ± 0.247**

fdibb $\sigma^2 = 3.96\times10^{-6}$ 12 seq

0.851 ± 0.094			
0.982 ± 0.115	0.819 ± 0.077		
1.187 ± 0.228	**1.207 ± 0203**	**1.423 ± 0.660**	
0.770 ± 0.126	0.896 ± 0.190	1.028 ± 0.178	0.840 ± 0.134

Table 5.15. Errorbars calculated for a matrix of all ones, $\sigma^2 = 2.57 \times 10^{-5}$

1.000 ± 0.073			
1.000 ± 0.359	1.000 ± 0.561		
1.000 ± 0.624	1.000 ± 0.524	1.000 ± 0.924	
1.000 ± 1.106	1.000± 0.886	1.000 ± 0.507	1.000 ± 0.771

Table 5.16. Element by element average and standard deviation of force constants from fcsca2, fconea, fconeb, fdia, fdib, fdiba, and fdibb.

0.846 ± 0.037			
0.977 ± 0.003	0.828 ± 0.034		
1.116 ± 0.098	1.098 ± 0.201	1.315 ± 0.282	
0.833 ± 0.098	0.971 ± 0.101	1.053 ± 0.034	0.961 ± 0.174

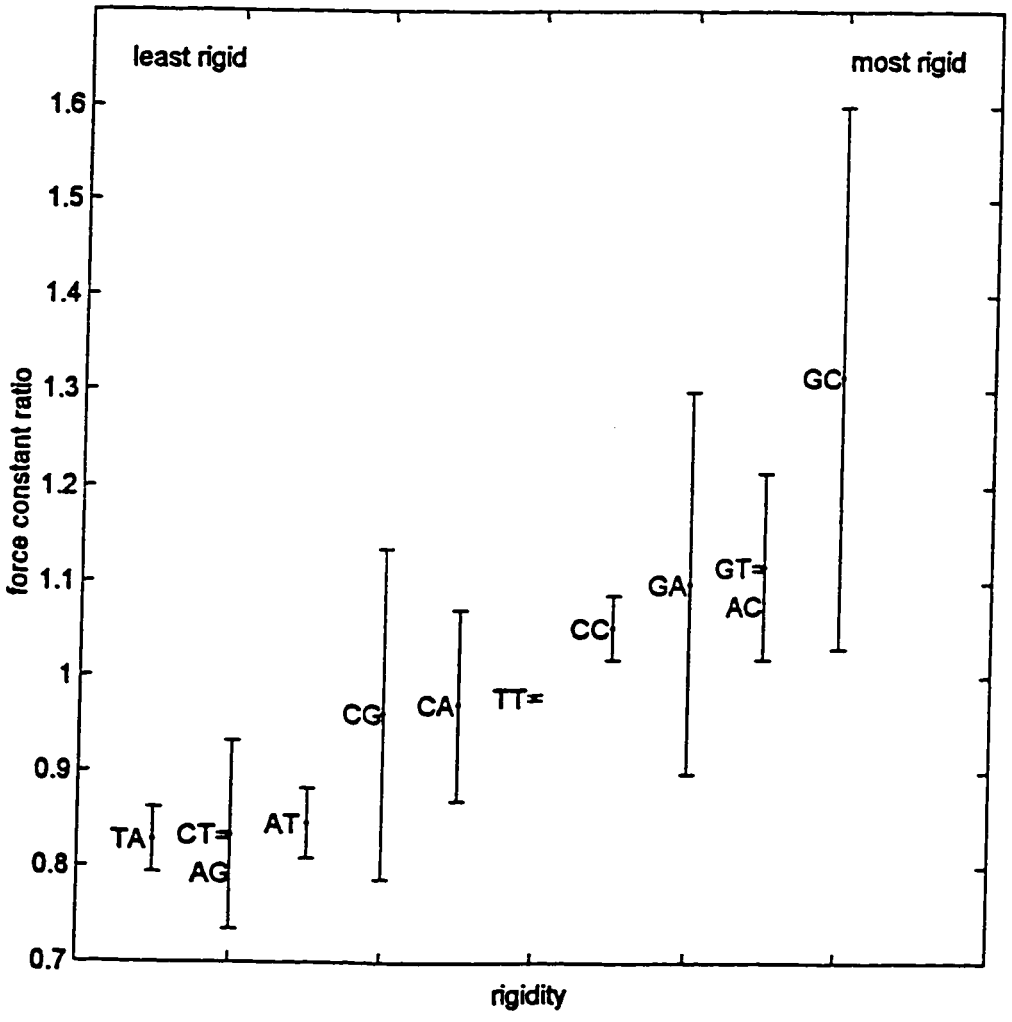


Figure 5.5. Element by element average and standard deviation of force constants from fcsca2, fconea, fconeb, fdia, fdib, fdiba, and fdibb. y axis represents force constant ratios, κ'/κ.

Table 5.17. Ranking of the dinucleotide steps.

name	κ'/κ	
GC	1.315 ± 0.282	most rigid
GT = AC	1.117 ± 0.098	
GA	1.098 ± 0.201	
CC	1.053 ± 0.034	
TT = AA	0.979 ± 0.003	
CA	0.971 ± 0.101	
CG	0.961 ± 0.174	
AT	0.846 ± 0.037	
CT = AG	0.833 ± 0.098	
TA	0.829 ± 0.034	most flexible

CHAPTER 6: C*, ANOTHER PROBE TO STUDY DNA DYNAMICS

SECTION 6.1. INTRODUCTION

During the last decade, two spin labels robust enough for automated DNA synthesis were designed, synthesized, and incorporated into DNA molecules for the purpose of studying DNA dynamics. Both these spin labels, T* and Q, were thymidine derivatives, and as such probed the AT base pair.[1,2] A third spin label, C*, has been developed to extend the site-specific spin-labeling technique to include the CG base pair as well.[3] This spin label, C*, is a spin-labeled analog of deoxycytidine, similar in structure to its predecessor T*, as shown in Figure 6.1. DNA molecules spin-labeled with C* were examined via EPR spectroscopy and compared with T*- and Q- spin-labeled DNA. The spectra indicate that C*, like its predecessors T* and Q, is useful for studying DNA dynamics.

SECTION 6.2. EXPERIMENTS

The spin label C* was synthesized in 9 steps and incorporated into synthetic DNA molecules by the automated phosphoramidite method, with manual coupling of the dC* residue, again to conserve spin label. Incorporation was confirmed by colorimetric quantitation of released trityl cation and the length of the strand as estimated from electrophoretic mobility using denaturing polyacrylamide gel electrophoresis (d-PAGE).[3] Furthermore, dC* was returned following exhaustive enzymatic hydrolysis of the phosphodiester backbone using a phosphodiesterase/phosphatase mixture (HPLC analysis).[3] Also, native polyacrylamide gel electrophoresis (PAGE) demonstrated that the spin-labeled DNA strands formed duplexes with the G-containing complement strands.[3]

DNA sequences are given in Table 6.1. C* was incorporated into the 11-mer 5'-d(TTT TTC* TTT TT) and also into a modified (self-complementary) Dickerson-like dodecamer, 5'-d(CGC GGA TC*C GCG). The non-self-complementary 12-mers 5'-d(GTC CC*C ACT CCC), 5'-d(GTC CC*G ACT CCC) and the T* analog to the second 12-mer were studied. The 14-mers 5'-d(GGA TAA TC*C GTG CC), 5'- d(GCC TAC ATG C*GA CG) and the T* and Q analogs to the second 14-mer were also examined. According to PAGE results, the 12-mer 5'-d(GTC CC*C ACT CCC) was also successfully extended to a 51-bp duplex using a variation of polymerase chain reaction (PCR), but EPR was not run on this sample.[4]

Samples were prepared for EPR by diluting 0.5 OD of the spin-labeled DNA in 10 μL PNE buffer (10 mM phosphate buffer (pH 7.0), 115 mM Na^+, 0.1 mM EDTA) or in 50% (weight /volume, w/v) sucrose in PNE buffer. Non-self-complementary DNA molecules were admixed with 1.5:1 molar excess of the unlabeled strand.

EPR spectra were collected on both the PDP spectrometer described in Chapter 2 using the Loop Gap Resonator (LGR) and the Bruker spectrometer described in Chapter 3 using the TE_{102} cavity. Conditions were as given in Chapter 3, except that 1 - 4 scans were taken per

spectrum on the PDP spectrometer. Spectra were collected at 0° and 20°C ± 2°C on the PDP an ± 0.1°C on the Bruker.

EPR spectra were simulated by the method described in Chapter 3. Spectra from samples in 50% sucrose at 0°C were simulated to find tensors using a fixed Gaussian convolution of 1 G. The sequences varied, so the magnetic environment varied and no one set of tensors was applicable to all the data. Sets of tensors used to simulate spectra are given in Table 6.2. Spectra from samples in 0% sucrose at 0° and 20°C were then simulated using the tensors in Table 6.2 and global dynamics given in Table 3.2. Internal dynamics were measured with $\langle\beta^2\rangle$ in the simulations. θ_{tilt} = 20°, r = 0.4 G (Lorentzian linewidth), and σ = 0.5 G (Gaussian broadening) were fixed.

SECTION 6.3. RESULTS AND DISCUSSION

The EPR spectra of a representative DNA sequence, 5'-d(GCC TAC ATG C*GA CG), are shown in Figure 6.2. The spectrum of the single-stranded DNA (Figure 6.2b) shows motion much slower than that observed for the monomeric nucleoside (Figure 6.2a), as previously observed for single-stranded DNA spin labeled with T* and Q.[1,2] When mixed with its complement, the motion of the DNA molecule slowed, as expected upon formation of the slower-tumbling duplex DNA (Figure 6.2c). The spectrum of the duplex DNA was consistent with uniform motion rotational correlation times τ_{perp} = 18.82 ns and τ_{para} = 8.44 ns (Table 3.2, 14-mer in 0% sucrose at 0°C). The other simulation parameters were g = 2.0076, 2.0069, 2.0041; A = 5.73 G, 10.71 G, 30.68 G; σ = 0.5 G; r = 0.4G; $\langle\beta^2\rangle$ = 0.152; R^2 = 0.9649. The EPR spectrum of the DNA duplex suspended in 50% w/v sucrose in PNE buffer at 0°C was simulated (Figure 6.2d, R^2 = 0.9743) to find the tensors used in simulating Figure 6.2c and resembles the spectrum of the analogous T*-spin-labeled DNA duplex characterized in parallel (Figure 6.2e with overlaid simulation, R^2 = 0.9576). Figure 6.2f shows the spectrum and simulation for the corresponding Q-spin-labeled DNA duplex, also in 50% w/v sucrose in PNE at 0°C (R^2 = 0.9673).

The discrepancies between the simulations and experimental spectra for the (short, 12 to 14 bp) C*-labeled DNA samples in 50% sucrose at 0°C may have been due to unexpected internal motion. The overall spectral width is narrower than that shown by the Q-labeled DNA, suggesting that the tensors may have been pre-averaged by internal dynamics, even at 0°C in 50% sucrose (τ_{perp} = 473 ns and τ_{para} = 212 ns). Using longer DNAs, such as 50-mers (as used for characterizing Q) might have been more effective (the global dynamics correlation times for a 50-mer in 50% sucrose at 0°C are τ_{perp} = 7150 ns and τ_{para} = 639 ns, more closely approximating the "no motion" assumption for DNA in 50% sucrose at 0°C implicit in determining the tensors this way). Nevertheless, even with longer DNAs, spectral line broadening would

have made simulations difficult.

The spectral line broadening is clearly seen by comparing Figures 6.2d and e to Figure 6.2f. The outer manifolds (at ca. -30 G and +35 G) are broader than single Lorentzian lineshapes convolved with Gaussians. The broad lineshapes suggest the presence of multiple components. This may be due to different orientations of the nitroxide moiety of T* and C*, since these spin labels are subject to sigma bond rotation about the acetylene tether. The two most likely orientations are "face up" and "face down," and in fact the spectra can be reasonably simulated as two-component spectra. One example is shown in Figure 6.3.

The values of $\langle\beta^2\rangle$ for the C*-labeled DNAs seem large relative to $\langle\beta^2\rangle$ found for nearly analogous Q-labeled DNAs. For example, $\langle\beta^2\rangle = 0.152$ for the DNA shown in Figure 6.2c, a 14-mer labeled in position 10 at 0°C, whereas $\langle\beta^2\rangle \approx 0.010$ for a 14-mer labeled in position 12 at 0°C (Table 4.3). Although the high $\langle\beta^2\rangle$ may be due in part to incorrect tensors, the difference between $\langle\beta_0^2\rangle$ for C* and T* and $\langle\beta_0^2\rangle$ for Q may also be partially responsible. From Chapter 4, $\langle\beta_i^2\rangle = \langle\eta_i^2\rangle + \langle\beta_0^2\rangle$. $\langle\beta_i^2\rangle$ measures the total internal dynamics; $\langle\eta_i^2\rangle$ measures the internal collective dynamics and is a function of h (rise per base pair), persistence length P_{df}, length N base pairs, and spin label position i; and $\langle\beta_0^2\rangle$ represents the internal local dynamics of independent spin label motion, base pair motion, and the motion of the base pair and spin label together but independent of the rest of the DNA molecule. For n-mers labeled in the same position i with the different spin labels C*, T*, and Q, $\langle\eta_i^2(C^*)\rangle = \langle\eta_i^2(T^*)\rangle = \langle\eta_i^2(Q)\rangle$. However, C*, like T* before it, can spin about the acetylene tether, thereby exhibiting spin label motion independent of the DNA motion, but Q cannot. Consequently, $\langle\beta_0^2\rangle$ for C* and T* > $\langle\beta_0^2\rangle$ for Q and so $\langle\beta_i^2\rangle$ for C* and T* > $\langle\beta_i^2\rangle$ for Q.

However, $\langle\beta_0^2\rangle$ for C* should be similar to that of T*, $\langle\beta_0^2(C^*)\rangle > \langle\beta_0^2(T^*)\rangle$, so that $\langle\beta_i^2(C^*)\rangle \approx \langle\beta_i^2(T^*)\rangle$, but instead, $\langle\beta_i^2(C^*)\rangle > \langle\beta_i^2(T^*)\rangle$ (0.152 > ca. 0.04).[1] Perhaps this discrepancy can be attributed to the tensor choice for C*. If the A_{zz} tensor had been chosen too wide, then $\langle\beta^2\rangle$ would have been forced artificially high to compensate. Using a longer (e.g., 50 bp) piece of DNA might have yielded a better tensor fit.

A comparison of the cost of synthesis for the three spin labels, T*, C*, and Q, reveals a substantial difference. The phosphoramidite precursor to Q involves a 22-step synthesis, whereas the precursors to T* and C* require only 8 and 9 steps in their respective syntheses. C* requires about 50% more time to synthesize, at 3 to 4 times the cost of the T* spin label, due to lower reaction yields. Hence T* is more easily prepared than C*, both of which are much more

easily prepared than Q. Furthermore, T* and C* are designed to base pair to naturally occurring nucleosides, A and G respectively, but Q requires (commercially available) 2-aminopurine (P) on the opposite strand. Consequently, DNA molecules studied with the QP base pair have 2 unnatural nucleosides, whereas those with the T*A or C*G base pair contain just one unnatural nucleoside. Because minimal perturbation of the DNA molecule is preferred for these DNA dynamics studies, T* and C* are superior to Q in this respect.

However Q was deliberately designed to have two strengths to offset its drawbacks: thermal stability and rigidity (seen in sharp spectral lineshapes). The three spin labels differ in stability both at and after exposure to high temperatures. As the temperature increases, the EPR lineshapes for the T* and C* labeled DNA molecules collapse more quickly to sharp lines than does the lineshape for the Q labeled DNA (Figure 6.4). At 40°C, spectra for both the T* and C* labeled DNA molecules represent motion on the order of subnanoseconds, while the Q-labeled DNA spectrum retains some of the slower motion character, representing motion on the order of 3 to 5 ns.

After 36 hours of exposure to 80°C, both T* and C*-containing duplexes show a fast component that resembles the EPR spectrum of a nucleoside monomer that had not been incorporated into DNA (Figure 6.5). The fast component may be explained by depyrimidination of the spin labeled residue, in which the unnatural base with its nitroxide moiety disintegrates, becoming a fast-tumbling small molecule which has the distinctive fast motion EPR signature. This fast component began to appear after 3 hours of exposure to 80°C and most likely was fully present well before 36 hours of exposure were completed. Q, being a C-nucleoside (the N at the base-sugar bond was replaced with a C, preventing depyrimidination), is inherently more stable to this type of thermal degradation.

As noted above, the T* and C* spectra contain broadened features suggestive of independent spin label rotation about the acetylene tether, in addition to the DNA motion itself. In contrast, spectra of the Q-labeled DNAs are characterized by narrower lines. The Q spin label is rigidly attached and has no opportunity to rotate, thus allowing a more rigorous analysis of the DNA motions because subtle changes can be observed, rather than being hidden under the broadened lines as would result with T* and C* labeled DNAs. This is particularly useful for examining the subtleties of sequence dependence of internal DNA dynamics discussed in Chapter 5.

Another way to compare the relative rigidities of these three spin labels is with the order parameter S.[6] S reflects how rigidly the spin label is attached to the DNA. When S = 1, the spin label is completely rigidly attached and there is no independent spin label motion at all; S = 0 indicates completely flexible attachment of the spin label to the DNA. Hustedt et al. related the

order parameter S to $\langle\beta_i^2\rangle$: $S \approx 1-\frac{3}{2}\langle\beta_i^2\rangle$ for $\beta < 20°$.[1] Using this relationship, the order parameter S was calculated for C*, T*, and Q and is listed in Table 6.3. C* is the most flexible of these spin labels, with S ranging from 0.76 to 0.90. T* is next, with S ranging from 0.80 to 0.95, and Q is most rigid, with S ranging from 0.90 to 1.00. Although such rigid attachment means that the spin labels report the DNA dynamics faithfully, the possibility exists that the spin labels themselves interfere with the local DNA dynamics. However, even if the local dynamics were influenced, the longer-range internal collective motions would still be faithfully reported.[7]

Bobst and coworkers have developed many different spin labels with tethers (from the nitroxide spin label to the DNA) of 2 to 11 σ bonds.[8] Although the length and freedom of motion of the tethers practically guarantee that the spin labels do not interfere with the DNA dynamics, the independent motions of the spin labels obscure the DNA dynamics.[1,7] Nevertheless, the most recent, a 2-σ-bond tethered spin label called DUMTA, does successfully report DNA dynamics.[9] The order parameter for DUMTA is approximately 0.35,[9] indicating more flexibility than C*, T*, and Q, but still reporting the same DNA dynamics as C*, T*, and Q.[7]

The spin labels T*, C*, and Q have different uses. The spin labels T* and C* are better for a quicker, lower-resolution screening to find oddities in DNA dynamics. Although these two spin labels yield EPR spectra compromised by the spin label motion, the spectra are sensitive to changes in DNA dynamics. The independent motion of the spin labels actually amplifies the changes in motion due to changes in structures. However, the T* and C* spin labels do not give unambiguous information about the structures themselves, due to the fundamental limitation that spin label motion and DNA motion cannot be separated for these spin labels. The Q spin label provides more direct dynamic information about the structures themselves, and would be best used once the sequences of interest were identified through the T* and C* screenings.

The difference in how the spin labels might be used is illustrated in a brief examination of flanking sequences. Two sequence contexts were examined: CC*G and CC*C. The spectra of the CC*G sequences in 0% sucrose at 0°C show internal motion approximately two times faster than the motion recorded in the spectra of the CC*C sequences (τ_{iso} = 14.2 ns vs. 8.0 ns, Figure 6.6). This trend was observed previously on the T* labeled DNA molecules.[5] C* clearly can show differences in internal motion and is an effective spin label to examine sequence context differences.

The spectra of the 12-mers containing the C*G context (in 50% sucrose at 0°C) were much more difficult to adequately simulate than the longer (14-mer) C*G context sequences, all the C*C context sequences, and T*- and Q-labeled DNAs. Entirely different tensor sets were required for these spectra, with much narrower A_{zz} (see Table 6.1). Although the spectra of

these 12-mers in 50% sucrose at 0°C may have been narrowed because of dramatic changes in the magnetic environment, the spectra may have also been compromised by internal dynamics (τ_{perp} = 360 ns and τ_{para} = 187 ns for a 12-mer in 50% sucrose at 0°C), possibly even compounded by the faster C*G motion described above.

The EPR spectra of one sequence yielded unexpectedly slow dynamics for the single stranded form of 5'-d(GGA TAA TC*C GTG CC) (Figure 6.7). The spectrum shows motion even slower than that of the duplexes shown in Figure 6.6. Preliminary polyacrylamide gel electrophoresis and circular dichroism experiments suggest that the single strand exists in some higher order structure. This unusual result suggests that the spin label C* may be useful for screening for anomalous DNA structures as well as DNA molecules exhibiting unusual dynamics.

SECTION 6.4. CONCLUSIONS

A new spin label, C* was synthesized, incorporated into DNA, and the DNA was characterized. EPR spectra of DNA samples labeled with C* indicate that the spin label does report DNA motion, albeit compromised with independent spin label motion.

The spin label T* is more accessible than C*, requiring an 8-step synthesis rather than a 9-step synthesis, at one quarter to one third the cost. Both T* and C*, however, are more accessible than Q, which requires a 22-step synthesis. The spin labels T* and C* both suffer from the fundamental limitation of sigma bond rotation of the nitroxide about the acetylene tether; Q was designed to address this limitation. Using either T* or C* also runs the risk of depyrimidination at or after exposure to high temperatures; Q is inherently more stable to such thermal degradation. In view of cost of synthesis, thermal stability, and narrowness of EPR spectra, the spin labels T* and C* are best suited for lower resolution, quick screening for oddities and the more expensive Q for more detailed investigation of the oddities. With this family of spin labels, both the AT and CG base pairs can be investigated site-specifically.

In this series of experiments to determine the utility of the C* spin label, the primary source of error lies in the simulations of the spectra. The poor fits indicate that the model is not accurate for the situation. However, this could be remedied in part at least by using longer DNA samples (for example, 50-mers) to determine tensors. With more accurate tensors, better simulations of the spectra would result. Optimization of θ_{tilt}, r (Lorentzian linewidth), and σ (width of the Gaussian convolution) would also improve the simulations.

This series of experiments with C* labeled DNA merely begins the characterization of C* and its application to the study of DNA dynamics. A number of experiments would further this work. Use of longer DNA samples would better determine the tensors for C*. EPR spectra of the PCR-extended DNA (from 12 to 51 base pairs) mentioned in the beginning of this chapter would corroborate the gel results that the DNA was indeed extended and determine whether the spin label survived the extension process. The anomalous sequence discussed above needs

further investigation regarding the odd higher order structure. To complete the thermal degradation comparison, sequences containing the three spin labels, T*, C*, and Q, should be exposed to 80°C for more closely connected time points (for example, 0.5, 1, 3, 5, 10, 15, 20, 25, 30, 35, and 40 hours) to determine after what point each spin label detaches from the DNA. Finally, a more systematic study of sequences labeled with C* is in order, as the flanking sequence study mentioned above clearly is just the beginning.

NOTES TO CHAPTER 6

1. Hustedt, E.J., Spaltenstein, A., Kirchner, J.J., Hopkins, P.B., & Robinson, B.H. *Biochemistry* **32**, 1774-1787 (1993).

2. Miller, T.R., Alley, S.C., Reese, A.W., Solomon, M.S., McCallister, W.V., Mailer, C., Robinson, B.H., & Hopkins, P.B. *J. Am. Chem. Soc.* **117**, 9377-9378 (1995).

3. Fischhaber, P.L., Reese, A.W., Nguyen, T., Kirchner, J.J., Hustedt, E.J., Robinson, B.H., & Hopkins, P.B. submitted to *Nucleosides and Nucleotides* (1996).

4. Fischhaber, P.L., unpublished results.

5. Kirchner, J.J., unpublished results.

6. Griffith, O.H. & Jost, P.C. in *Spin Labeling: Theory and Applications*, ed. L.J. Berliner, pp. 454-523. New York: Academic Press (1976)

7. Robinson, B.H., Mailer, C., & Drobny, G.P. submitted to *Annual Review of Biophysical and Biomolecular Structures* (1996).

8. Bobst, E.V., Keyes, R.S., Cao, Y.Y., & Bobst, A.M. *Biochemistry* **35**, 9309-9313 (1996) and references therein.

9. Keyes, R.S. & Bobst, A.M. *Biochemistry* **34**, 9265-9276 (1995).

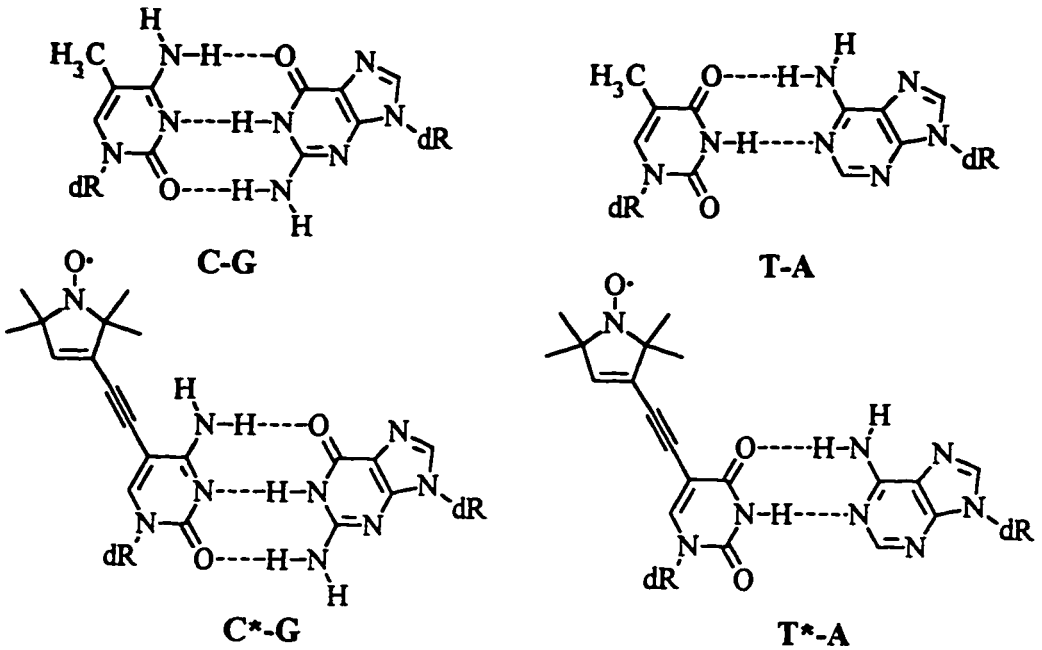


Figure 6.1. Structures of the spin labeled base pairs CG, C*G, TA, and T*A.

Table 6.1. C* spin labeled sequences of length N base pairs. Sequences listed were paired with their partner (G-containing) complementary strands to form duplexes.

N	sequence	name	
11	TTT TTC* TTT TT		
12	CGC GGA TC*C GCG	plf1	self-complementary
12	GTC CC*C ACT CCC	plf3	CC*C context
12	GTC CC*G ACT CCC	plf6	CC*G context
12	GTC CT*G ACT CCC	plf7,8	CT*G context, T* analog
14	GGA TAA TC*C GTG CC	plf4	TC*C context
14	GCC TAC ATG C*GA CG	plf2	GC*G context
14	GCC TAC ATG T*GA CG	plf9	GT*G context, T* analog
14	GCC TAC ATG QGA CG	plfq	GQG context, Q analog

Table 6.2. Tensor sets used in simulations of experimental spectra. "Standard" (std) tensor sets are the sets in Table 3.1 (Q) and from Hustedt et al.[1] (T*). The "data sets" column shows which data sets were analyzed with each tensor set.

spin label	data sets	g_{xx}	g_{yy}	g_{zz}	A_{xx}	A_{yy}	A_{zz}
Q (std)	plfq	2.0084	2.0068	2.0034	6.58	4.98	34.23
T* (std)		2.0076	2.0059	2.0029	6.27	10.10	32.21
T*	plf9	2.0070	2.0061	2.0036	5.22	10.85	31.13
C*	plf1,2,3,4,8	2.0076	2.0069	2.0041	5.73	10.71	30.68
C*	plf6	2.0084	2.0071	2.0058	5.77	13.04	26.54
C*	plf7	2.0071	2.0063	2.0040	5.63	11.55	28.37

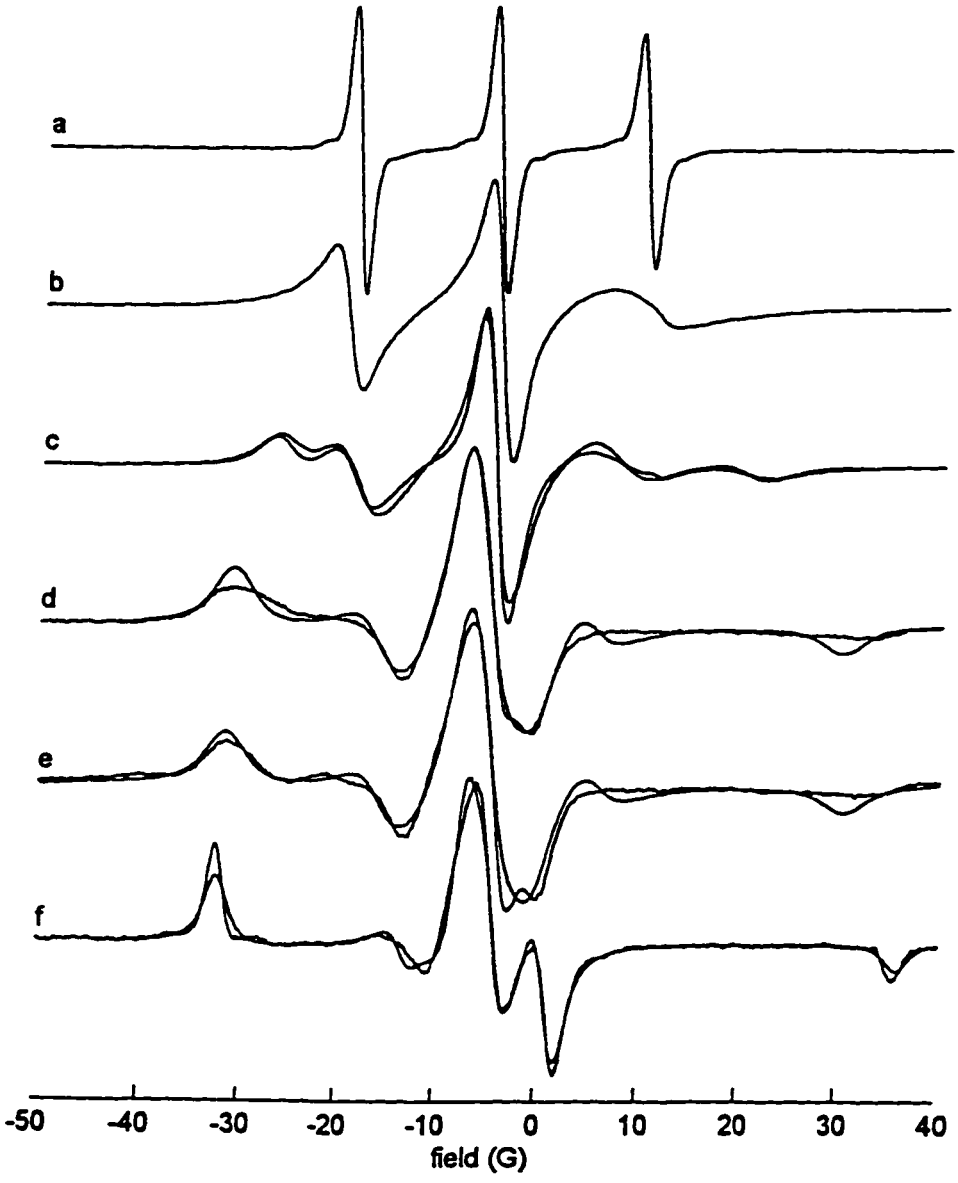


Figure 6.2. CW-EPR spectra of a) monomer C* (20°C); b) single-stranded 5'-d(GCC TAC ATG C*GA CG) in PNE buffer (0°C) (plf2); c) duplex 5'-d(GCC TAC ATG C*GA CG)·5'-d(CG TCG CAT GTA GGC) in PNE buffer (0°C) (plf2); d) 5'-d(GCC TAC ATG C*GA CG)·5'-d(CG TCG CAT GTA GGC) in 50% w/v sucrose in PNE buffer (0°C) (plf2); e) 5'-d(GCC TAC ATG T*GA CG)·5'-d(CG TCA CAT GTA GGC) in 50% w/v sucrose in PNE buffer (0°C) (plf9); f) 5'-d(GCC TAC ATG QGA CG)·5'-d(CG TCP CAT GTA GGC) in 50% w/v sucrose in PNE buffer (0°C) (plfq). Simulations are overlaid for c), d), e), and f). Tensors are listed in Table 6.2. For c), additional simulation parameters were rotational correlation times 18.82 ns (τ_{perp}) and 8.44 ns (τ_{para}), σ = 0.5 G; r = 0.4G, θ_{tilt} = 20°, $\langle\beta^2\rangle$ = 0.152 (R^2 = 0.9649). Simulations for d), e), and f) used σ = 1.0 G and r = 2.0 G, 2.0 G, and 1.2 G, respectively, with R^2 = 0.9743, 0.9576, and 0.9673.

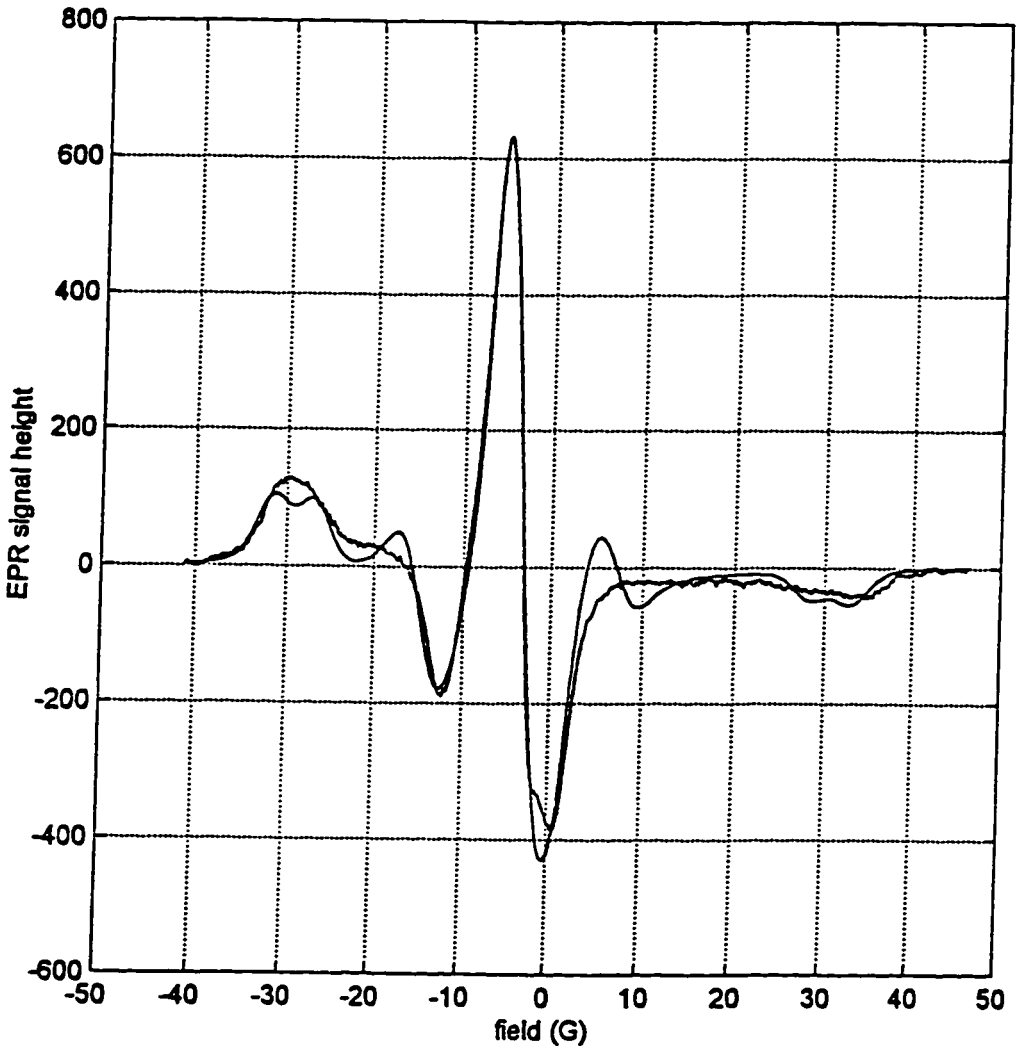


Figure 6.3 CW-EPR spectrum of duplex DNA 5'-d(GCC TAC ATG C*GA CG)·5'-d(CG TCG CAT GTA GGC) in 50% w/v sucrose in PNE buffer (0°C) with a 2-component simulation overlaid. Parameters were A_{xx} = 5.73 G, A_{yy} = 10.71 G, $A_{zz,1}$ = 27.84 G, $A_{zz,2}$ = 32.50 G, g_{xx} = 2.0076, g_{yy} = 2.0069, g_{zz} = 2.0041 (R^2 = 0.9677).

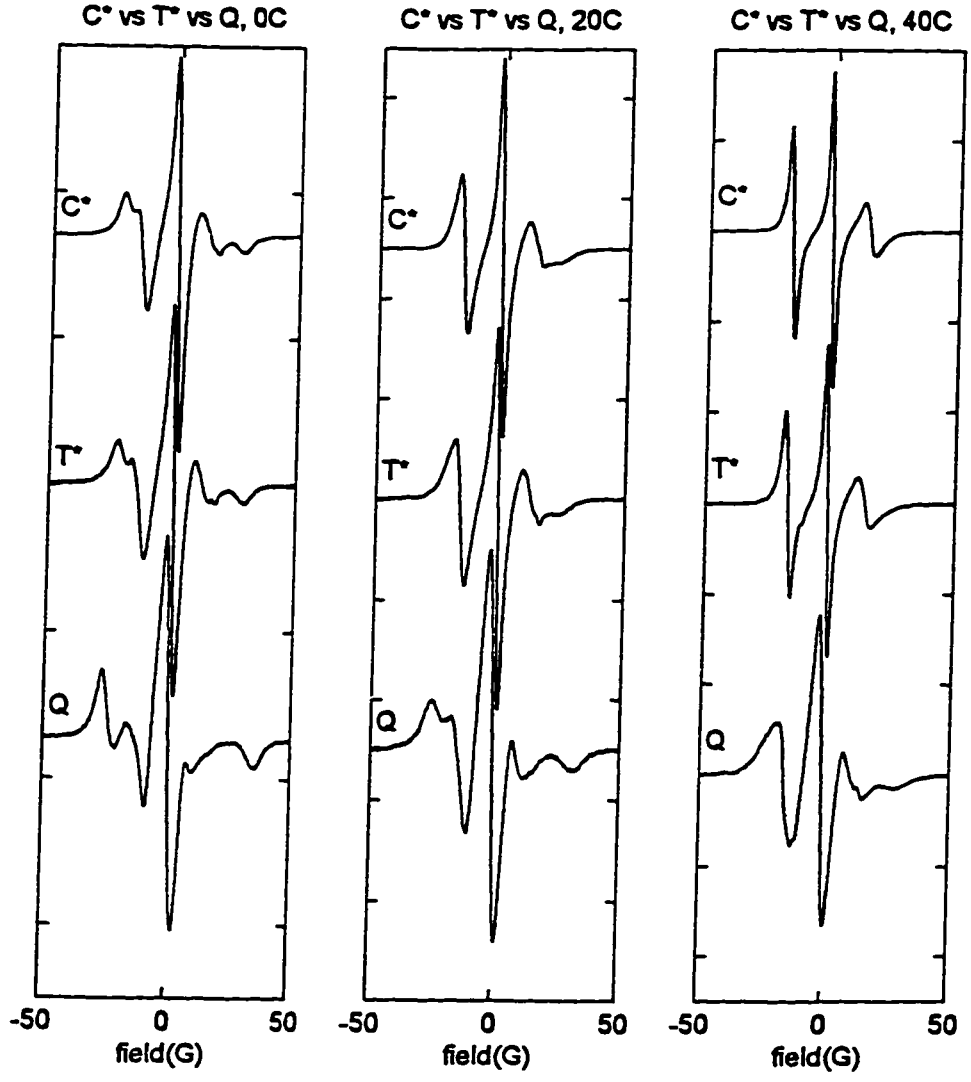


Figure 6.4. CW-EPR spectra of C*-, T*-, and Q-labeled duplex DNA 5'-d(GCC TAC ATG XGA CG)·5'-d(CG TCY CAT GTA GGC), where X and Y are (C*,G), (T*,A), and (Q,P), in PNE buffer at 0°, 20°, and 40°C.

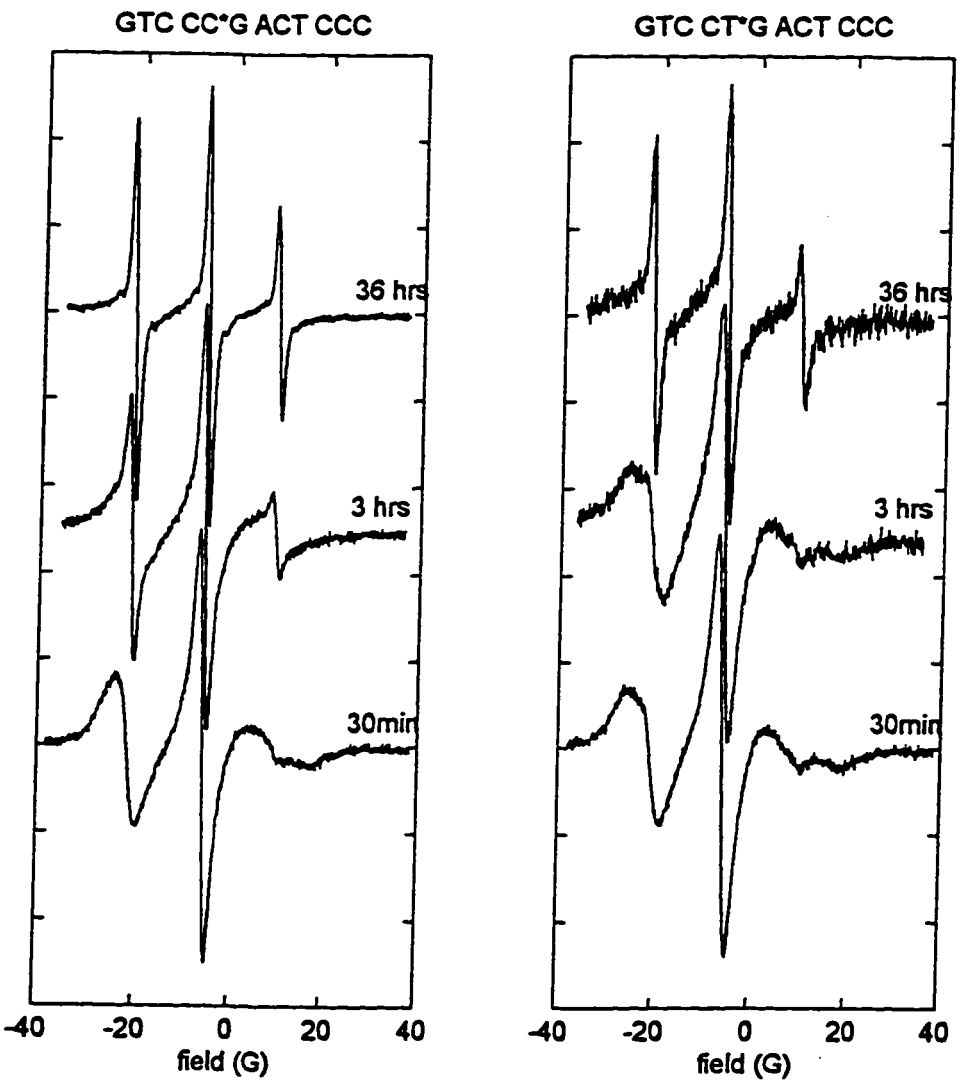


Figure 6.5. CW-EPR spectra of C*- and T*-labeled duplex DNA 5'-d(GTC CXG ACT CCC)·5'-d(GGG AGT CYG GAC), where X and Y are (C*,G) and (T*,A), in PNE buffer at 0°C after exposure to 80°C for 30 minutes, 3 hours, and 36 hours.

Table 6.3. Order parameter S and $\langle\beta^2\rangle$ for the spin labels DUMTA, C*, T*, and Q.

spin label	range of $\langle\beta^2\rangle$	S
DUMTA	n/a	0.35
C*	0.0625 - 0.1600	0.76 - 0.90
T*	0.0332 - 0.1225	0.80 - 0.95
Q	0.0000 - 0.0784	0.90 - 1.00

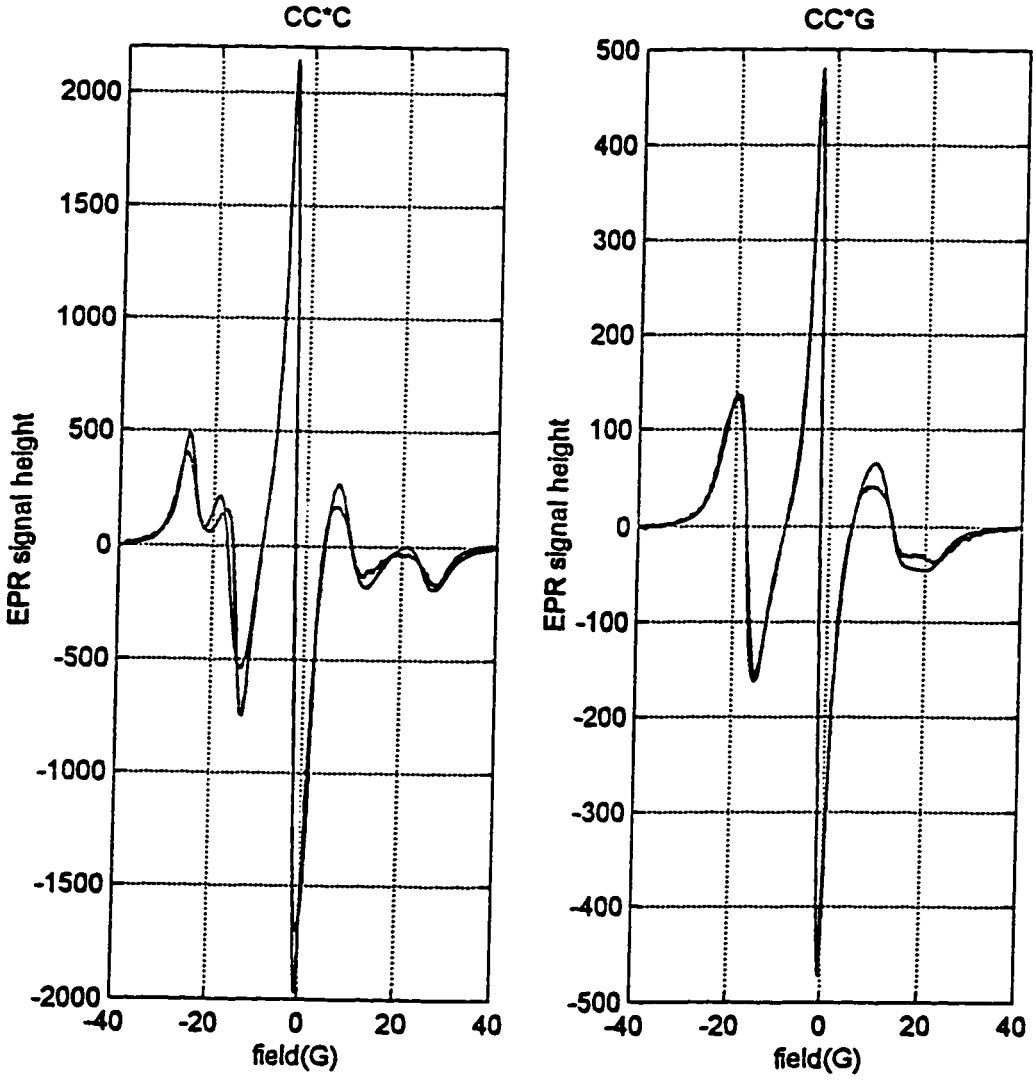


Figure 6.6. CW-EPR spectra of duplex DNA with CC*C and CC*G flanking sequences, a) 5'-d(GTC CC*C ACT CCC)·5'-d(GGG AGT GGG GAC) (plf3) and b) 5'-d(GTC CC*G ACT CCC)·5'-d(GGG AGT CGG GAC) (plf6) in PNE buffer (0°C) overlaid with simulations using isotropic correlation times 14.2 ns and 8.0 ns respectively. Tensors are given in Table 6.2. Additional simulation parameters were θ_{tilt} = 20°, r = 0.4 G, and σ = 0.5 G.

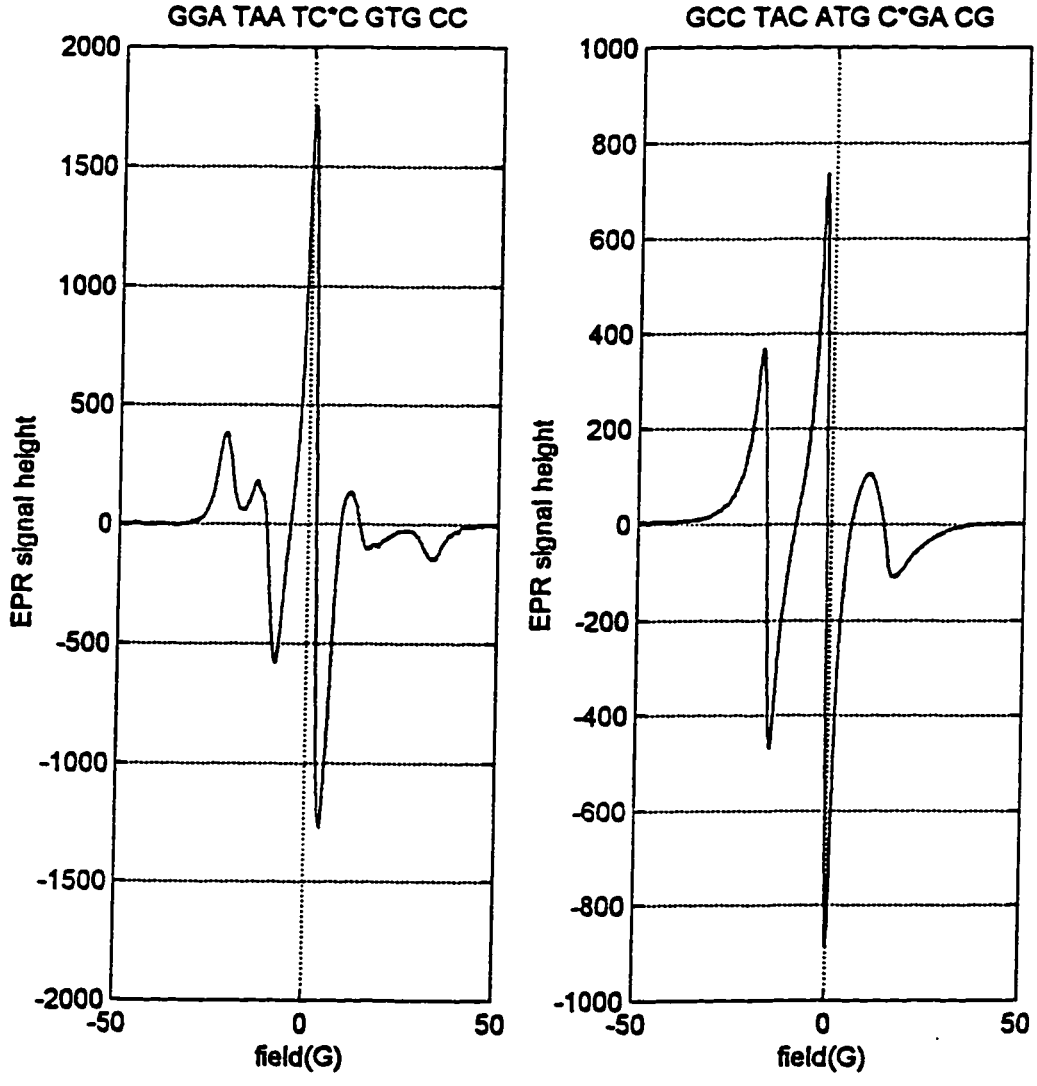


Figure 6.7. CW-EPR spectra of single stranded DNA a) 5'-d(GGA TAA TC*C GTG CC) and b) 5'-d(GCC TAC ATG C*GA CG) in PNE buffer (0°C) to illustrate the unusually slow dynamics of the first sequence.

CHAPTER 7: SUMMARY

The first project described in this work revolved about CW-EPR lineshape analysis for the purpose of extracting R_2 accurately from the spectral lineshape. A simulation routine was developed using the MATLAB software package. Rather than using analytical functions to calculate the lineshape, this routine does the direct numerical convolution to extract R_2. Furthermore, this analysis explicitly includes instrumental effects (for example, from observer power and Zeeman modulation amplitude and frequency), R_{1e} effects, and ^{13}C effects. Consequently, spectra may now be deliberately overmodulated to improve the signal-to-noise ratio without compromising the final results.

This analysis program was used to quantitate the broadening effects of spin label concentration and oxygen concentration upon R_2. Results were 9% to 16% smaller than the literature values for both the spin label and oxygen concentration effects. R_{1e} and R_{2e} were estimated from analyses of power studies (a purely CW-EPR technique); the ratio of R_{2e}/R_{1e} was in qualitative agreement with the ratio found from time-domain EPR studies. R_2 appears to be proportional to τ^{-1}, where τ is the correlation time, based on analyses of the spin label concentration effect in 0%, 15%, 40%, and 70% glycerol. However, as the data were compromised from water-contaminated glycerol, a new set of data should be collected using clean glycerol to properly elucidate the dependence of R_2 on τ.

Further refinement of each of these effects awaits the completion of a new analysis program which has been developed. This new program explicitly does the convolutions for the nearby protons and deuterons, instead of subsuming them under a broad Gaussian convolution. Preliminary results show that distinguishing between the ^{14}N and ^{15}N isotopes and the D (^2H) and H (^1H) isotopes is significant. This distinction may account, at least in part, for the differences between the spin label concentration and oxygen concentration effects found here and those in the literature: the work done here used predominantly ^{15}N dCTPO, whereas the literature work used ^{14}N mHCTPO. Results from this new analysis will be used to calculate R_{2e} from first principles.

The remaining projects described in this work examined the dynamics of DNA molecules, using two different spin labels. The acetylene-linked C derivative, C*, was successfully incorporated into synthetic DNA. As its predecessor, the acetylene-linked T derivative T*, C* can be used to study DNA dynamics, but the EPR spectra for both T*- and C*-spin labeled DNAs are broadened by the rotation of the nitroxide about the sigma bond of the acetylene. This independent spin label motion renders the data less useful, as analysis is compromised. Nevertheless, the availability of both T* and C* now permits examination of both the AT base pair and the CG base pair.

The quinolone T derivative, Q, was also successfully incorporated into synthetic DNA, but at much higher synthetic cost (22 synthesis steps rather than less than 10 steps for T* and C*). However, EPR spectra of DNAs labeled with Q show much finer resolution than those labeled with T* or C*. The narrower lines bespeak the rigidity with which Q is attached to the DNA and allow for much cleaner analysis of spectra. Q is believed to report only the DNA motion now, uncompromised by independent spin label motion.

The weakly bending rod model's prediction of the length and position dependence of DNA internal dynamics was confirmed with only a simplistic weighting modification. The internal dynamics appear to depend very slightly on temperature as well, implying that the dynamic flexural persistence length for this particular sequence family, P_{df}, depends (very slightly) inversely on temperature. P_{df} at 20°C was estimated as 1250 ± 75 Å, and the rms amplitude of oscillation (independent base pair motion) at 20°C was estimated as 5.2° ± 0.2°.

The narrow lines of CW-EPR spectra of Q-labeled DNA molecules made possible the systematic investigation of the dependence of internal DNA dynamics on base sequence. The flexibilities of the dinucleotide and trinucleotide repeats were ranked, and the effects of changing the "flexible region" length and position were quantitated. This work represents the beginning of more detailed work on sequence dependence of DNA dynamics.

BIBLIOGRAPHY

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

Alley, S.C. Ph.D. Thesis, University of Washington, in preparation.

Allison, S.A. & Schurr, J.M. *Chem. Phys.* **41**, 35-59 (1979).

Bales, B.L., Blum, R.A., Mareno, D., Peric, M., & Halpern, H.J. *J. Mag. Res.* **98**, 299-307 (1992).

Barber, E.J. *National Cancer Inst. Monogr.* **21**, 219-239 (1966).

Berliner, L.J. in *Spin Labeling: Theory and Applications*, ed. L.J. Berliner, pp 1-4. New York: Academic Press (1976).

Bernal, J.M.G. & de la Torre, J.G. *Biopolymers* **19**, 751-766 (1980)

Bobst, E.V., Keyes, R.S., Cao, Y.Y., & Bobst, A.M. *Biochemistry* **35**, 9309-9313 (1996).

Bobst, A.M., Kao, S.-C., Toppin, R.C., Ireland, J. C., & Thomas, I.E. *J. Mol. Biol.* **173**, 63-74 (1984).

Bowman, M.K., Michalski, T.J., Peric, M., Halpern, H.J. *Pure and Applied Chemistry* **62**, 271-274 (1990).

Carrington, A., & McLachlan, A.D. *Introduction to Magnetic Resonance.* New York: Harper and Row (1967).

Collini, M., Chirico, G., & Baldini, G. *J. Chem. Phys* **104**, 6058-6065 (1996).

CRC Handbook of Chemistry and Physics, ed. R.C. Weast, p. F-79. Cleveland, Ohio: CRC Press (1975).

Dlakic, M. & Harrington, R.E. *PNAS* **93**, 3847-3852 (1996).

Eimar, W., Williamson, J.R., Boxer, S.G., & Pecora, R. *Biochemistry* **29**, 799-811 (1990).

Edmonds, A.R. *Angular Momentum in Quantum Mechanics.* Princeton NJ: Princeton University Press (1974).

Fischhaber, P.L., Reese, A.W., Nguyen, T., Kirchner, J.J., Hustedt, E.J., Robinson, B.H., & Hopkins, P.B. submitted to *Nucleosides and Nucleotides*, 1996.

Fujimoto, B.S. & Schurr, *Nature* **344**, 175-178 (1990).

Gebe, J.A. Ph.D. Thesis, University of Washington, 1995.

Goldman, A., Bruno, G.V., Polnaszek, C.F., & Freed, J.H. *J. Chem. Phys.* **56**, 716-735 (1972).

Goldstein, H. *Classical Mechanics*, pp 107-109. Reading MA: Addison-Wesley (1980).

Griffith, O.H. & Jost, P.C. in *Spin Labeling: Theory and Applications*, ed. L.J. Berliner, pp. 454-523. New York: Academic Press (1976)

Haas, D.A., Mailer, C., & Robinson, B.H. *Biophys. J.* **64**, 594-604 (1993).

Haas, D.A. Ph.D. Thesis, University of Washington, 1993.

Halpern, H.J., Peric, M., Yu, C., Bales, B.L. *J. Mag. Res.* **A103**, 13-22 (1993).

Halpern, H.J., Peric, M., Nguyen, T., Spencer, D.P., Teicher, B.A., Lin, Y., & Bowman, M.K. *J. Mag. Res.* **90**, 40-51 (1990).

Handbook of Chemistry, ed. N.A. Lange, p. 1669. New York: McGraw-Hill Book Company (1961).

Harvey, S.C., Dlakic, M., Griffith, J., Harrington, R., Park, K., Sprous, D., & Zacharias, W. *J. Biomol. Struct. Dyn.* **13**, 301-307 (1996).

Huang, H., Miller, T.R., Duncan, J.A., Alley, S.C., Zhu, L., Reid, B.R., and Hopkins, P.B., in preparation.

Hustedt, E.J., Spaltenstein, A., Kirchner, J.J., Hopkins, P.B., & Robinson, B.H. *Biochemistry* **34**, 4369-4375 (1995).

Hustedt, E.J., Spaltenstein, A., Kirchner, J.J., Hopkins, P.B., & Robinson, B.H. *Biochemistry* **32**, 1774-1787 (1993).

Hustedt, E.J. Ph.D. Thesis, University of Washington, 1989.

Hwang, J.S., Mason, R.P., Hwang, L.P. & Freed, J.H. *J. Phys. Chem.* **79**, 489-511 (1975).

Hyde, J.S. & Subczynski, W.K. *J. Mag. Res.* **56**, 125-130 (1984).

Keyes, R.S. & Bobst, A.M. *Biochemistry* **34**, 9265-9276 (1995).

Kim, Y., Geiger, J.H., Hanh, S., & Sigler, P.B. *Nature* **365**,512-520 (1993).

Kim,J.L., Nicolav, D.B., & Burley, S.K. *Nature* **365**, 520-527 (1993).

Kivelson, D. *Journal of Chemical Physics* **33**, 1094-1106 (1960).

Koo, H.S., Woo, H.M., & Crothers, D.M. *Nature* **320**, 501-506 (1986).

Lai, C.S., Hopwood, L.E., Hyde, J.S., & Lukiewicz, S. *Proc. Natl. Acad. Sci. USA* **79**, 1166-1170 (1982).

Levy, G.C., Craik, D.J., Kumar, A., & London, R.E. *Biopolymers* **22**, 2703-2726 (1983).

Lyubchenko, Y.L., Shlyakhtenko, L.S., Appella, E., & Harrington, R.E. *Biochemistry* **32**, 4121-4127 (1993).

Mailer, C., Haas, D.A., Hustedt, E.J., Gladden, J.G., & Robinson, B.H. *J. Mag. Res.* **91**, 475 - 496 (1991).

Mailer, C., Danielson, J.D.S., & Robinson, B.H. *Rev. Sci. Instr.* **56**, 1917-1025 (1985).

Maniatis, T., Fritsch, E.F., & Sambrook, J., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory (1982).

Miller, T.R. Ph.D. Thesis, University of Washington, 1995.

Miller, T.R., Alley, S.C., Reese, A.W., Solomon, M.S., McCallister, W.V., Mailer, C., Robinson, B.H., & Hopkins, P.B. *J. Am. Chem. Soc.* **117**, 9377-9378 (1995).

Miller, T.R. & Hopkins, P.B. *Bioorg. Med. Chem. Lett.* **4**, 981-986 (1994);

Murphy, J.G., Smith, T.W., & Marsh, J.D. *J. Mol. Cell Cardiol.* **19**, 271-279 (1987).

Nordio, P.L. in *Spin Labeling: Theory and Applications*, ed. L.J. Berliner, pp 5-52. New York: Academic Press (1992).

Pecora, R. *Science* **251**, 893-897 (1991).

Percival, P.W., & Hyde, J. S. *J. Mag. Res.* **23**, 249-257 (1976).

Peric, M. & Halpern, H.J. *J. Mag. Res.* **A109**, 198-202 (1994).

Popp, C.A., & Hyde, J.S. *J. Mag. Res.* **43**, 249-258 (1981).

Redfield, A.G. *Phys. Rev.* **98**, 1787-1809 (1955).

Robinson, B.H., Mailer, C., & Drobny, G.P. submitted to *Annual Review of Biophysical and Biomolecular Structures* (1996).

Robinson, B.H. & Drobny, G.P. *Annual Review of Biophysical and Biomolecular Structures* **24**, 523-549 (1995).

Robinson, B.H., Reese, A.W., & Mailer, C., 18th International EPR Symposium, Denver CO (1995).

Robinson, B.H., Haas, D.A., Mailer, C. *Science* **363**, 490-493 (1994).

Robinson, B.H. *J. Chem. Phys.* **78**, 2268-2273 (1983).

Robinson, B.H., Lerman, L.S., Beth, A.H., Frisch, H.L., Dalton, L.R., & Auer, C. *J. Mol. Biol.* **139**, 19-44 (1980).

Rose, M.E. *Elementary Theory of Angular Momentum*. John Wiley, NY, 1957.

Saenger, W., *Principles of Nucleic Acid Structure*, in the Springer Advanced Texts in Chemistry series, series ed., Charles R. Cantor, p. 271. New York: Springer-Verlag (1984).

Schultz, S.C., Shields, G.C., & Steitz, T.A. *Science* **253**, 1001-1007 (1991).

Schurr, J.M., Fujimoto, B.S., Wu, P., & Song, L. Fluorescence studies of nucleic acids, in *Topics in Fluorescence Spectroscopy, Vol 3. Biological Applications*, ed. J.R. Lakowicz, pp. 137-229. New York: Plenum Press (1992).

Smimov, A.I. & Belford, R.L. *J. Mag. Res.* **113**, 65-73 (1995).

Smimova, T.I., Smimov, A.I., Clarkson, R.B, Belford, R.L. *Mag. Res. in Medicine* **33**, 801-810 (1995).

Solomon, M.S.; & Hopkins, P.B. *J. Org. Chem.* **58**, 2232-2243 (1993).

Solomon, M.S. Ph.D. Thesis, University of Washington, 1992.

Solubility Data Series, Volume 7. Oxygen and Ozone, ed. R. Battino, p. 2. New York: Pergamon Press (1981)

Song, L., Allision, S.A., & Schurr, J.M. *Biopolymers* **29**, 1773-1791 (1990).

Song, L. & Schurr, J.M. *Biopolymers* **30**, 229-237 (1990).

Spaltenstein, A., Robinson, B.H., & Hopkins, P.B. *Biochemistry* **28**, 9484-9495 (1989).

Spaltenstein, A., Robinson, B.H., & Hopkins, P.B. *J. Am. Chem. Soc.* **110**, 1299-1301 (1988);

Spaltenstein, A., Robinson, B.H., & Hopkins, P.B. *J. Am. Chem. Soc.* **111**, 2303-2305 (1989).

Sprous, D., Zacharias, W., Wood, Z.A., & Harvey, S.C. *Nucleic Acids Research* **23**, 1816-1821 (1995)

Tirado, M.M.; de la Torre, J.G. *J. Chem. Phys.* **73**, 1986-1993 (1980).

van der Vliet, P.C., Verrijzer, C.P. *Bioessays* **15**, 25-32 (1993).

von Hippel, P.H. *Science* **263**, 769-770 (1994).

Wilcoxon, J. & Schurr, J.M. *Biopolymers* **22**, 2273-2321 (1983).

Wu, P., Fujimoto, B.S., & Schurr, J.M. *Biopolymers* **26**, 1463-1488 (1987).

Appendix A. Spectra and Simulations

Simulations are shown overlaid on the experimental spectra for the left-, middle-, and right-labeled DNA sequences discussed in Chapter 4, at T = 0°, 10°, 20°, 30°, and 40°C.

Figure A.1. Left-labeled DNA, T = 0°C.

Figure A.2. Left-labeled DNA, T = 10°C.

Figure A.3. Left-labeled DNA, T = 20°C.

Figure A.4. Left-labeled DNA, T = 30°C.

Figure A.5. Left-labeled DNA, T = 40°C.

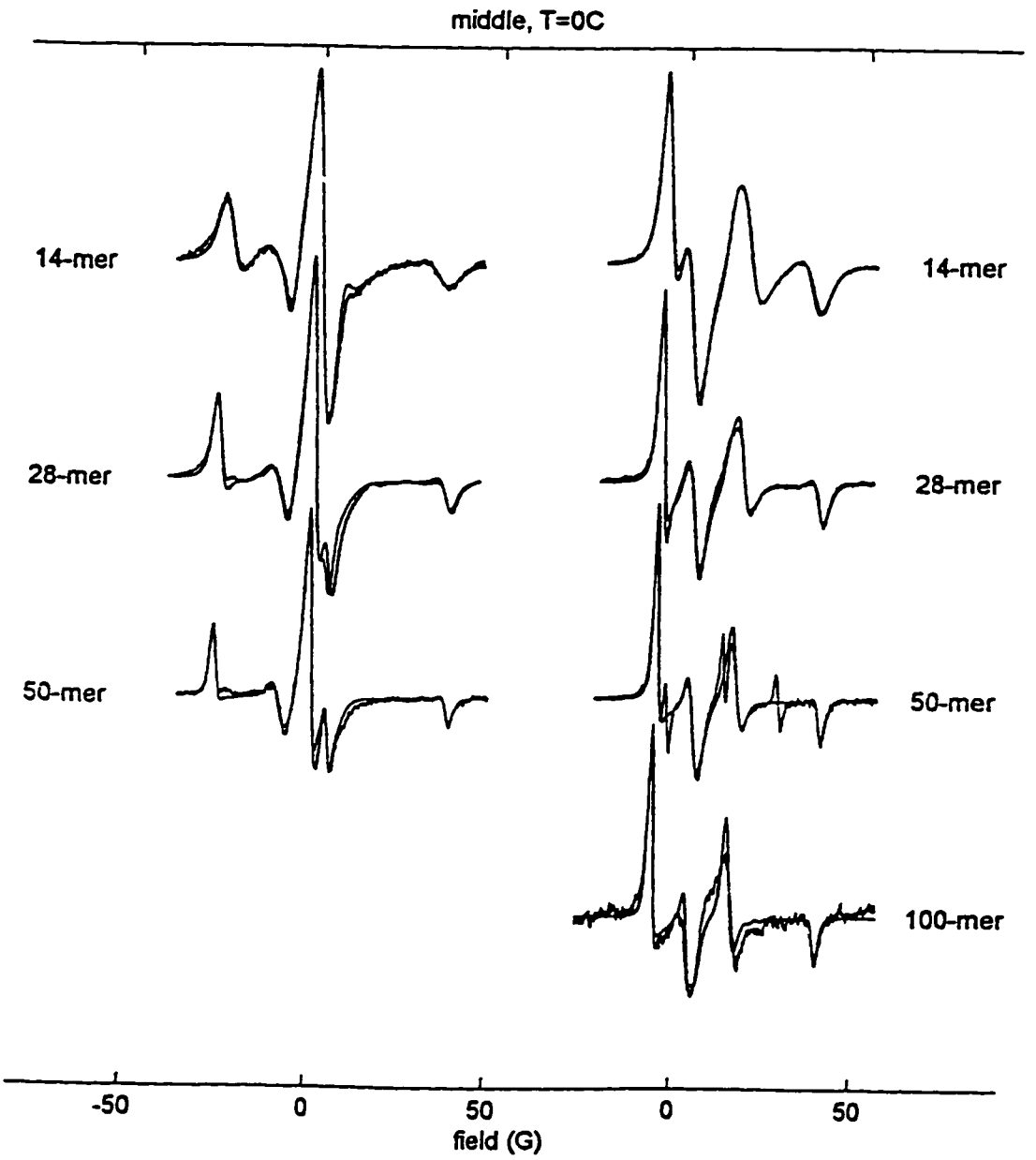


Figure A.6. Middle-labeled DNA, T = 0°C.

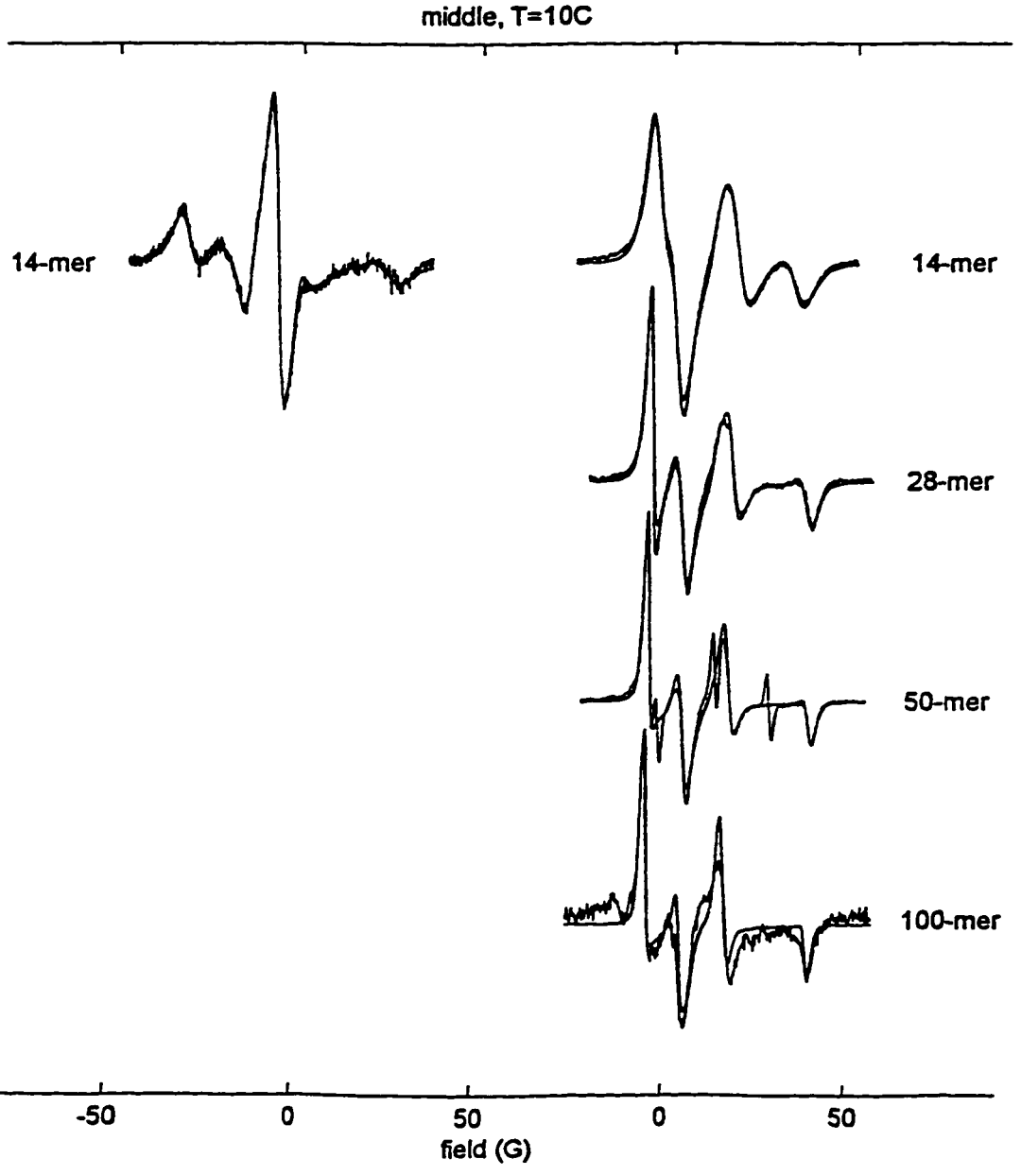


Figure A.7. Middle-labeled DNA, T = 10°C.

Figure A.8. Middle-labeled DNA, T = 20°C.

Figure A.9. Middle-labeled DNA, T = 30°C.

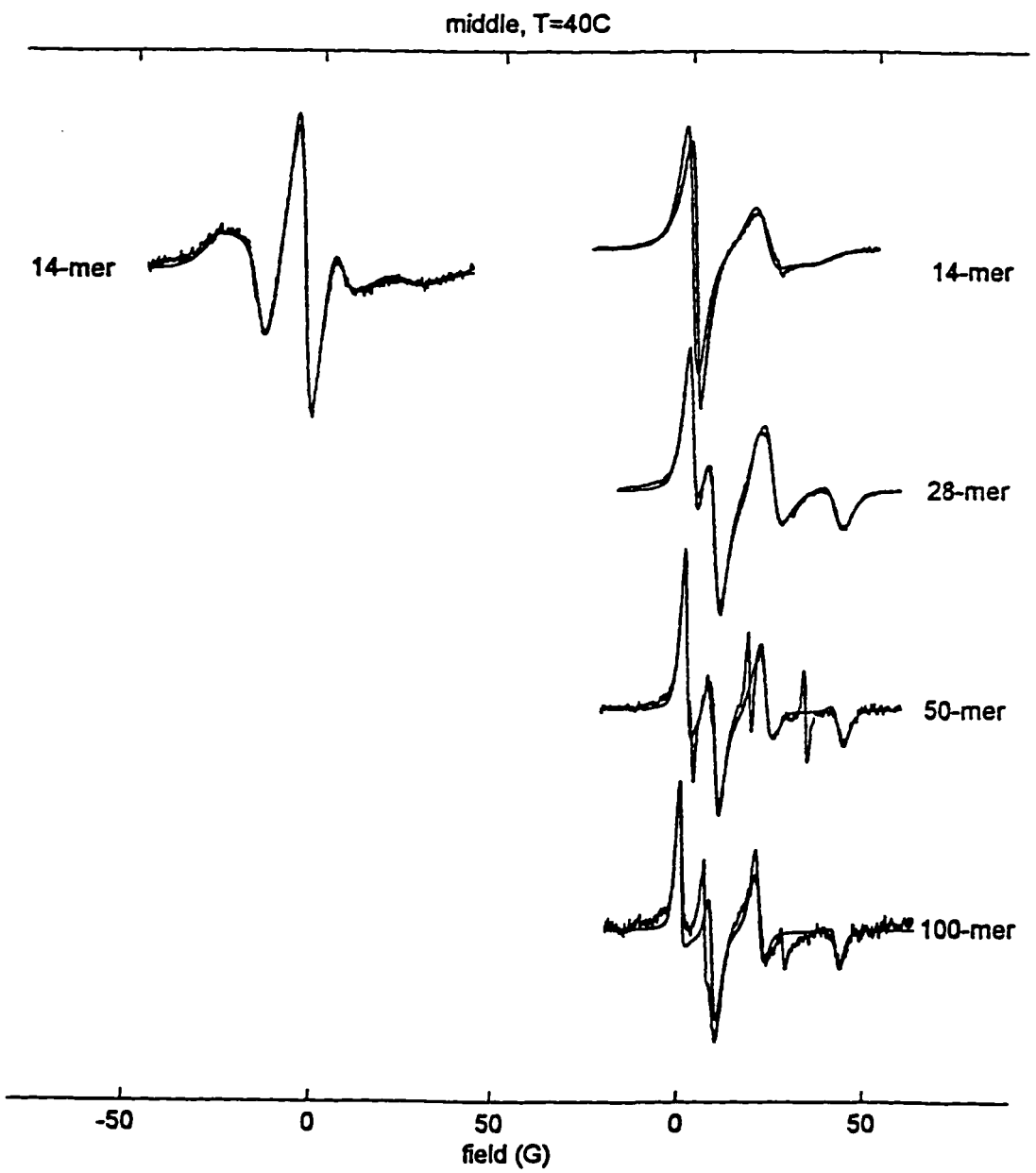


Figure A.10. Middle-labeled DNA, T = 40°C.

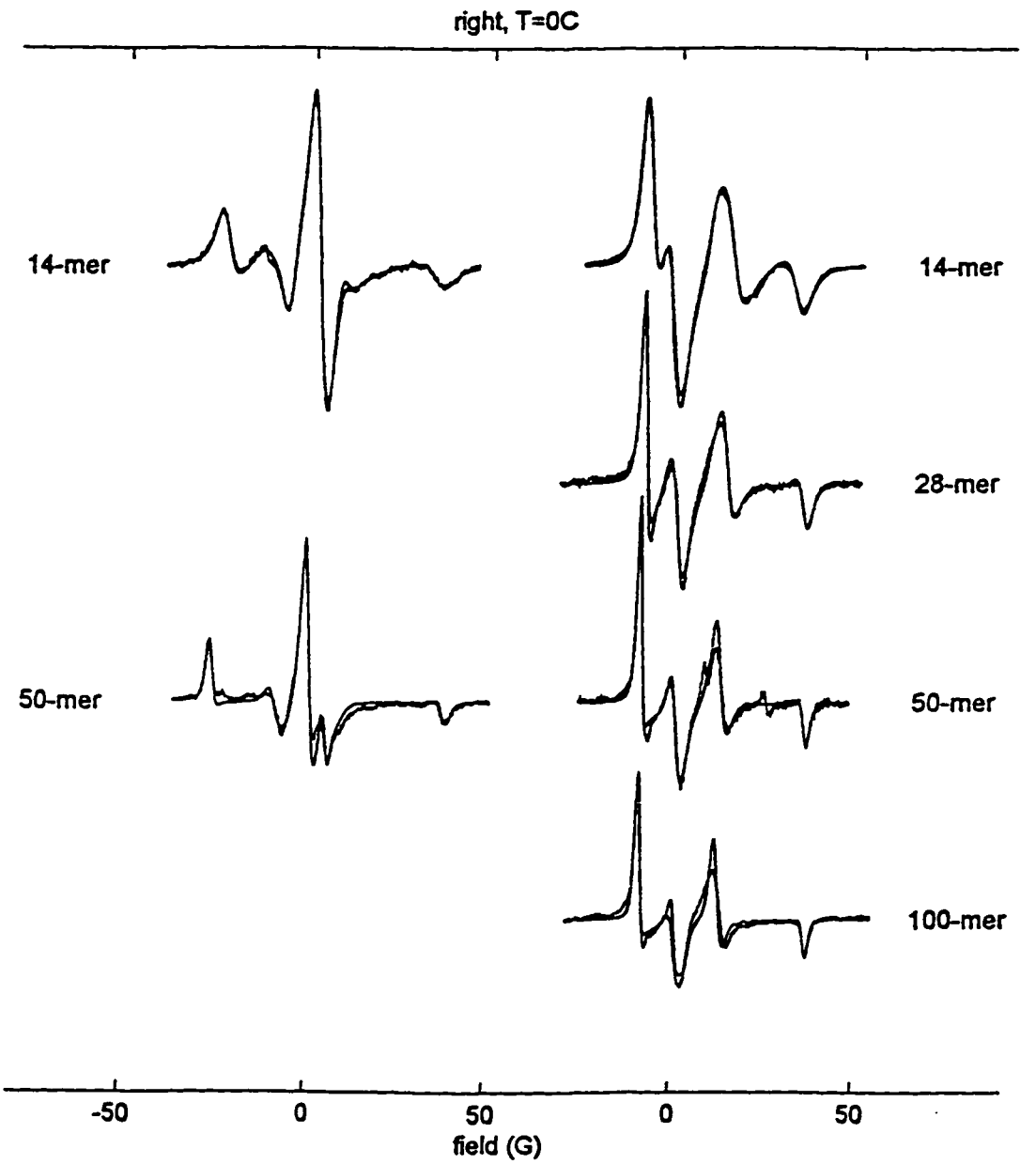


Figure A.11. Right-labeled DNA, T = 0°C.

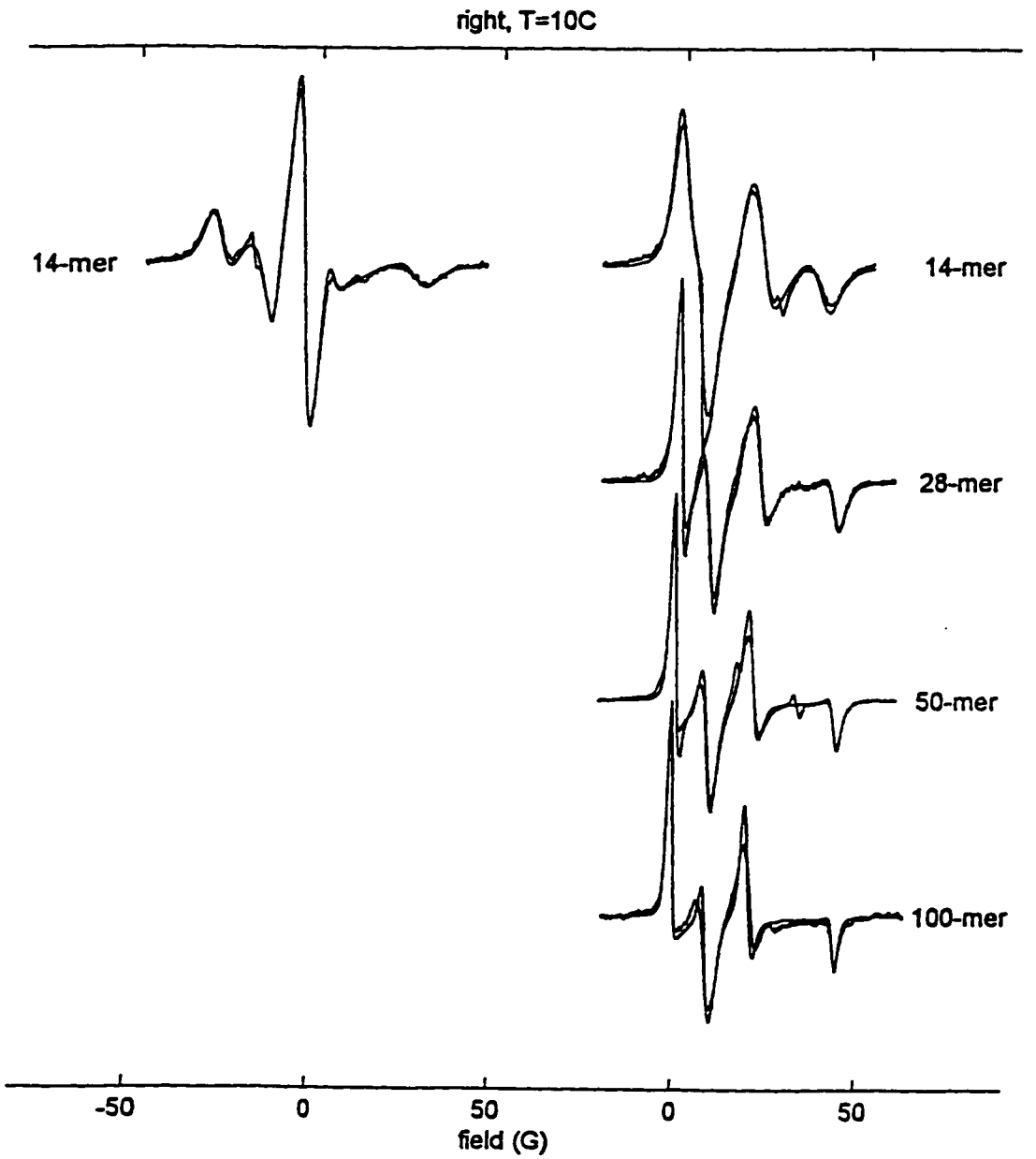


Figure A.12. Right-labeled DNA, T = 10°C.

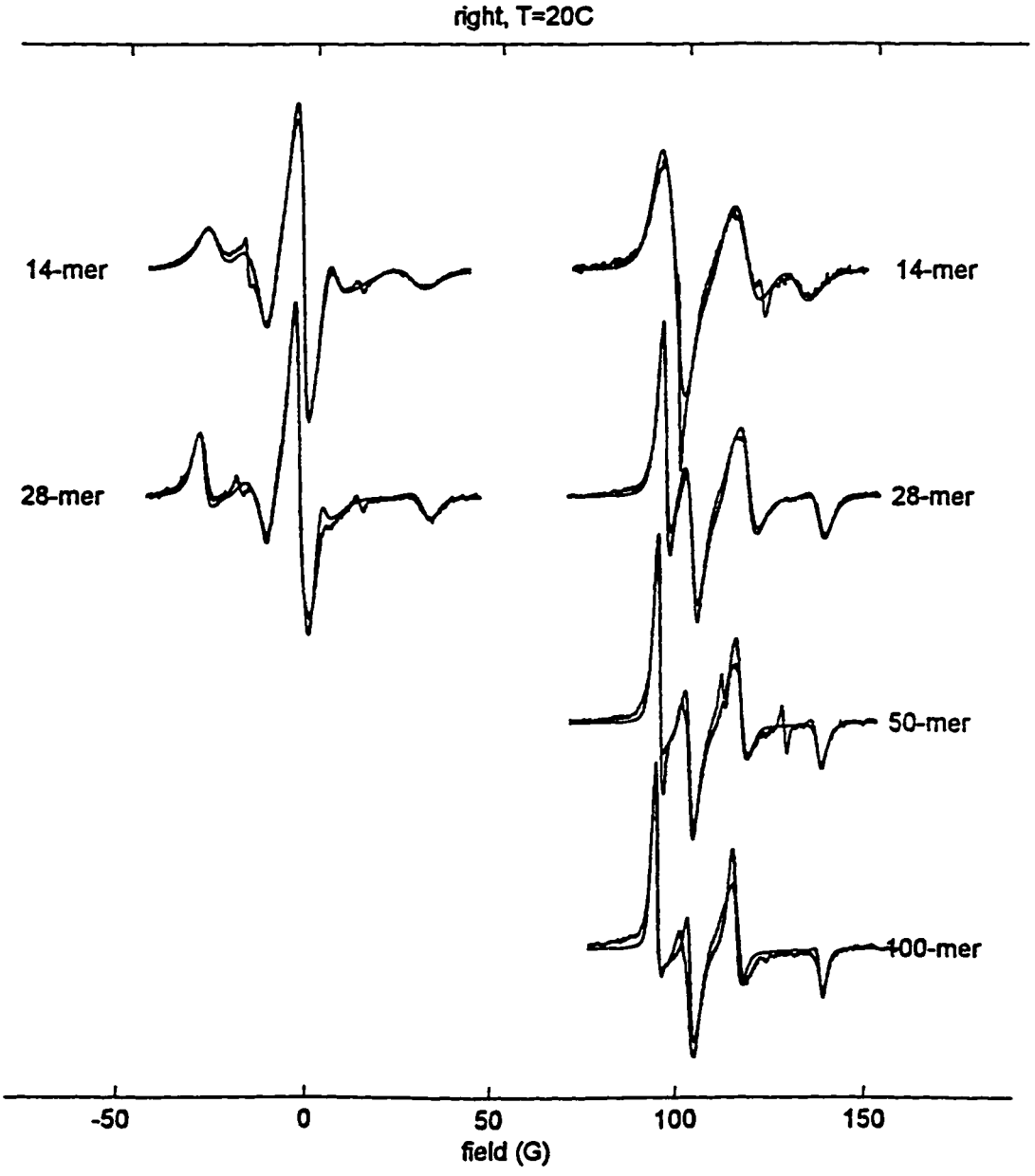


Figure A.13. Right-labeled DNA, T = 20°C.

Figure A.14. Right-labeled DNA, T = 30°C.

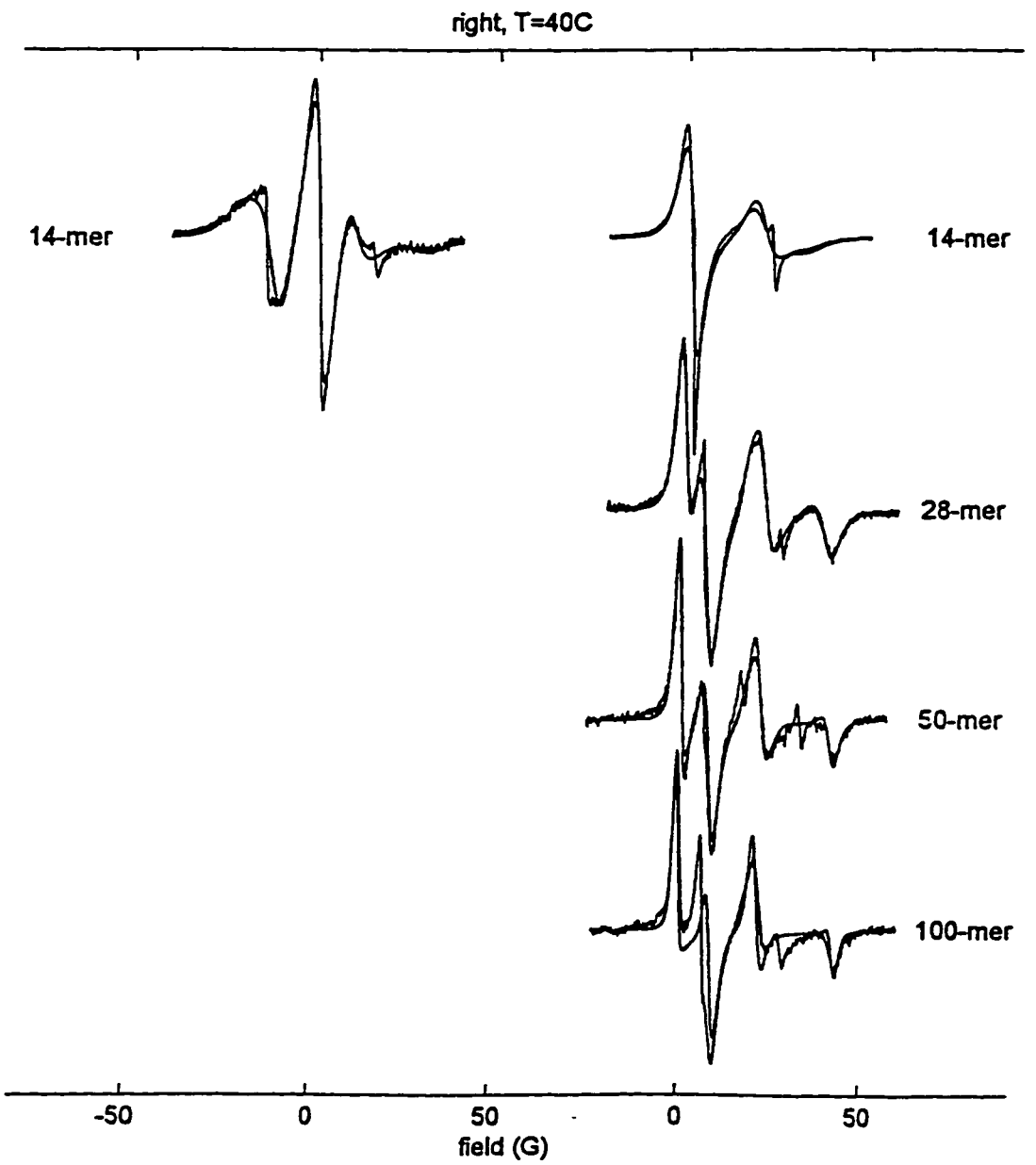


Figure A.15. Right-labeled DNA, T = 40°C.

APPENDIX B: MATLAB CODE FOR LINEWIDTH ANALYSIS.

The following MATLAB code was used in the analysis of the linewidth data described in Chapter 2. All the code presented here except buildp.m was written by Bruce H. Robinson.

code name	code function
rot_c13	builds and calculates correlation coefficient of simulation to data
buildp	builds P and rot_phase from input parameters
line_c13	builds complicated Lorentzian spectrum
lor_fm	generates individual Lorentzian line including MF and MA in calculation
lor_full	generates individual Lorentzian line including MF, MA, h_1, and R_{1e} in calculation
c13	incorporates ^{13}C peaks into Lorentzian spectrum
g_conv	builds Gaussian and convolve with Lorentzian spectrum
rotate	rotates spectrum
zeeman	rotates spectrum to account for Zeeman modulation
see_c13	builds and scales simulation so that simulation and data can be plotted together
coef	calculates coefficients of covariance matrix between data and simulation

```
function [irstat] = rot_c13(po,nxover,lorprog,nequiv,px,signal)
% function [irstat] = rot_c13(po,nxover,lorprog,nequiv,px,signal);
% calculates spectrum from parameters in po and px and
% compares to data, calculating irstat = 1-R^2
% analyzes signal w/ n xovers -- generalized!
% MUST global Gauss and Z_Data before the call
%  Gauss = X axis variable;  Z_Data is complex and contains quadrature too
% po, px = COLUMN vectors of parameters, defined in buildp, to build P
%  po = [ x1 {a1} x2 {a2} ... floating_variables]'
%  px = [ n1 {a1} n2 {a2} ... dummy fixed_variables]'
%   P = [x1 ... xn lor1 ... lom sigma MF MA R1 h1 rot_phase c13a c13ra]';
% nxover = # crossovers (1 for single line, 2 for N15, 3 for N14)
% lorprog = flag to determine which line-generating program to use
%     1 = lor_simp; 2 = lor_freq; 3 = lor_mod; 4 = lor_fm; 5 = lor_full
% nequiv = # equivalent C13s to contribute to the C13 peaks
%   to not include the C13 peaks in the analysis, set nequiv = 0,
%      and assign dummy #s to c13a and c13ra
% optional:  signal is 'imag' ie absorption A  (default)
%                or 'real' ie dispersion D
%                or 'magn' ie absolute magnitude sqrt(A^2 + D^2)
% NOTE: if signal is to be sent, px MUST be sent also (px = [0] is fine)
% see buildp.m for how to build po and px ('help buildp')

% includes rotation of abs/disp phase angle
% computes the R statistic, and returns (1 - R^2) for fit of model to data
% fits the data from the parameters in po and px
% calls buildp,line_c13,rotate,zeeman
% called by cwfitc13

% for MATLAB 4.0 versions
     global Gauss Z_Data

% default will compare absorption spectra and in phase
     sigtype = 'imag';
     phase = 1;
% if signal sent, then change sigtype from 'imag' to 'real' or 'abs'
     if (nargin == 6)              % signal exists
       if (signal == 'real' ), sigtype= 'real';,end
       if (signal == 'magn' ), sigtype = 'abs';,end
     end
% send to buildp to make the necessary P
     [P,rot_phase]=buildp(po,px,nxover);
% commence fitting
     lor = line_c13(Gauss,P,lorprog,nequiv);
     lor = rotate(lor,rot_phase);
     [lor,phi1] = zeeman(lor,0,'imag');
     [lor,phi2] = zeeman(lor,phi1,'imag');
     Y = eval( [ sigtype , '(Z_Data(:,phase))' ] ) ;
     Y_hat = eval( [ sigtype , '(lor(:,phase))' ] ) ;
     r_sq = ( corrcoef(Y,Y_hat) ) ;
     r_sq = r_sq(1,2) ^ 2 ;
     irstat = 1 - r_sq ;
     disp([irstat po' ] )
```

```
function [P,rot_phase] = buildp(po,px,nxover)
% function [P,rot_phase] = buildp(po,px,nxover)
% buildp builds the P vector for the lor_* programs can use it to build the lorentzian lineshape.
% this is really inelegant code which could be replaced by a double-column matrix
%    where the first column is fix/float (1/0) and the second contains the values
% first build Po and Px from po, px
%  po = [ x1 {a1} x2 {a2} ... floating_variables]'
%  px = [ n1 {a1} n2 {a2} ... dummy fixed_variables]'
%    xj = left-most crossover for each N-isotope
%    aj = distance between related lines (a1 = N14, a2 = N15, for example)
%    nj = number of lines IN ADDITION to the one specified by position xj.
%      if n15, then add 1 more line, so n=1;  if n14, n=2 etc.
%     the sign of nj controls whether aj comes from po or px
%      if aj comes from po (floats), then nj > 0
%      if aj comes from px (fixed),  then nj < 0
%  if aj is floated or nj=0 then NO number is needed below nj.
%  nxover = total number of crossovers = sum(abs(nj)))
%  Po = [x1 ... xn floating_variables]'
%  Px = [dummy fixed_variables]'
% then build P and rot_phase from Po,Px; Px may be []
%   modified 6/13/94 AR to include option to fix gaussian sigma
%   modified 3/23/95 AR to include x, C13 and MF,MA,R1,h1 options
%   Po,Px logic and lorprog flags tested 3/24/95 AR
% Po = COLUMN vector!!
%   = [x1 ... xn l1 ... ln sigma MF MA R1 h1 rot_phase c13a c13ra]';
%  x = xover = crossover (G); l = lor = lorentzian width (G)
%  sig = Gaussian width, sigma (G); MF = modulation frequency (G)
%  MA = modulation amplitude (G); R1 = 1/(T1e) (G)
%  h1 = observer power (G); rot_phase = rotation angle (degrees)
%  c13a = distance btwn low and hi-field c13 peaks (G)
%  c13ra = C13 ratio, approx equal to natural C13 abundance
%      height of C13 peak will be ratio * nequiv
% Po can be optimized by fmins; Px is kept fixed; lPo + lPx = 2*n+9
% to fix arguments, use Px: (Px = COLUMN!)
%          1st element of Px = flag
% fix nothing -> Px = [0] (MUST DO if want signal ~= 'imag')
% fix      MF MA R1 h1          -> Px = [ 1      MF MA R1 h1]';
% fix      MF MA R1 h1 c13a c13ra -> Px = [ 2      MF MA R1 h1 c13a c13ra]';
% fix  sig MF MA R1 h1          -> Px = [ 3  sig MF MA R1 h1]';
% fix  sig MF MA R1 h1 c13a c13ra -> Px = [ 4  sig MF MA R1 h1 c13a c13ra]';
% fix  sig   MA R1 h1 c13a c13ra -> Px = [ 5  sig   MA R1 h1 c13a c13ra]';
% fix  sig MF   R1 h1 c13a c13ra -> Px = [ 6  sig MF   R1 h1 c13a c13ra]';
% fix  sig MF MA   h1 c13a c13ra -> Px = [ 7  sig MF MA   h1 c13a c13ra]';
% fix  sig MF MA R1   c13a c13ra -> Px = [ 8  sig MF MA R1   c13a c13ra]';
% fix  sig MF      h1 c13a c13ra -> Px = [ 9  sig MF      h1 c13a c13ra]';
% fix  sig MF MA      c13a c13ra -> Px = [10  sig MF MA      c13a c13ra]';
% fix  sig         h1 c13a c13ra -> Px = [11  sig         h1 c13a c13ra]';
% fix x    MF MA R1 h1 c13a c13ra
%          -> Px = [12 MF MA R1 h1 c13a c13ra]';
% fix x sig MF MA R1 h1 c13a c13ra
%          -> Px = [13 sig MF MA R1 h1 c13a c13ra]';
% called by rot_c13 and see_c13 to build P for use in line_c13
% build Po and Px from po and px
% pointers to where we are in the po and px vectors.
npo=1;, npx=1;
```

```
nxs=0;  % the new/updated number of crossovers found as we go.
Po=[];
while (nxs < nxover)
    nj = abs(px(npx));
    if(px(npx) == 0),  Pj= po(npo) ;,end
    if(px(npx) > 0), v= (0:1:nj)';, npo=npo+1;, Pj = po(npo-1)+ v*po(npo);,end
    if(px(npx) < 0), v= (0:1:nj)';, npx=npx+1;, Pj = po(npo)+ v*px(npx);,end
    nxs=nxs+nj+1;, npx=npx+1;,npo=npo+1;, Po = [ Po; Pj];
end
lpo = length(po);, lpx = length(px);
Po = [Po;po(npo:lpo) ];, Px = px(npx:lpx);
% build P and rot_phase from Po and Px
      n2 = nxover * 2;
      lPo = length(Po);
      if( isempty(Px)),              % Px does not exist
        if( lPo == n2+8),              % must be true
         rot_phase = Po(n2+6);
         P = [ Po(1:n2+5); Po(n2+7:n2+8) ];
         else, error('lPo ~= 2*nxover+8!')   %jump out if lPo ~= n2+8
        end
      else                          % Px does exist
        lPx = length(Px);
        if((lPo+lPx) == n2+9 ),        %lPo+lPx = n2+9 to work!
         if (Px(1) == 0), rot_phase = Po(n2+6);
          P = [ Po(1:n2+5); Po(n2+7:n2+8) ];, end
         if (Px(1) == 1), rot_phase = Po(n2+2);,
          P = [ Po(1:n2+1); Px(2:5); Po(n2+3:n2+4) ];, end
         if (Px(1) == 2), rot_phase = Po(n2+2);,
          P = [ Po(1:n2+1); Px(2:7)];, end
         if (Px(1) == 3), rot_phase = Po(n2+1);,
          P = [ Po(1:n2); Px(2:6); Po(n2+2:n2+3)];, end
         if (Px(1) == 4), rot_phase = Po(n2+1);,
          P = [ Po(1:n2); Px(2:8)];, end
         if (Px(1) == 5), rot_phase = Po(n2+2);,
          P = [ Po(1:n2); Px(2); Po(n2+1); Px(3:7)];, end
         if (Px(1) == 6), rot_phase = Po(n2+2);,
          P = [ Po(1:n2); Px(2:3); Po(n2+1); Px(4:7)];, end
         if (Px(1) == 7), rot_phase = Po(n2+2);,
          P = [ Po(1:n2); Px(2:4); Po(n2+1); Px(5:7)];, end
         if (Px(1) == 8), rot_phase = Po(n2+2);,
          P = [ Po(1:n2); Px(2:5); Po(n2+1); Px(6:7)];, end
         if (Px(1) == 9), rot_phase = Po(n2+3);,
          P = [ Po(1:n2); Px(2:3); Po(n2+1:n2+2); Px(4:6)];, end
         if (Px(1) ==10), rot_phase = Po(n2+3);,
          P = [ Po(1:n2); Px(2:4); Po(n2+1:n2+2); Px(5:6)];, end
         if (Px(1) ==11), rot_phase = Po(n2+4);,
          P = [ Po(1:n2); Px(2); Po(n2+1:n2+3); Px(3:5)];, end
         if (Px(1) ==12), rot_phase = Po(n2+2);,
          P = [ Po(1:n2+1); Px(2:lPx)];, end
         if (Px(1) ==13), rot_phase = Po(n2+1);,
          P = [ Po(1:n2); Px(2:lPx)];, end
% P = [x1 ... xn l1 ... ln sigma MF MA R1 h1 c13a c13ra]';
        else, error('lPo+lPx ~= 2*n+9!') % jump out if lPo+lPx ~= 2*n+9
        end
      end
```

```
function [ line ] = line_c13(x,P,lorprog,nequiv)
% function [ line ] = line_c13(x,P,lorprog,nequiv);
% computes the complex lineshape for lorentzians & convolves w/ gaussian (width sig)
% x = x-axis, ie, Gauss
% P = COLUMN vector!! built in buildp.m
%   = [x1 ... xn l1 ... ln sigma MF MA R1 h1 c13a c13ra]';
%  x = xover = crossover (G); lor = lorentzian width (G)
%  sig = Gaussian width, sigma (G); MF = modulation frequency (G)
%  MA = modulation amplitude (G); R1 = 1/(T1e) (G)
%  h1 = observer power (G)
%  c13a = distance from low-field C13 peak to hi-field peak (G)
%  c13ra = ratio of C13:N peak height, approx equal to natural C13 abundance
%      height of C13 peak will be ratio * nequiv
% note that xover,lor,MF,MA,R1,and h1 are all in units of x (eg, Gauss)
%   MF = 10 kHz == 3.57mG == 0.00357G
% lorprog = flag to determine which line-generating program to use
%     1 = lor_simp; 2 = lor_freq; 3 = lor_mod; 4 = lor_fm; 5 = lor_full
% nequiv = # equivalent C13s to contribute to the C13 peaks
%   to NOT include the C13 peaks in the analysis, set nequiv = 0,
%   and assign dummy #s to c13a and c13ra
% calls lor_simp,lor_freq,lor_mod,lor_fm,lor_full(which calls lor_ra),c13,g_conv
% called by rot_c13, see_c13
% Nt is the type of NO radical:  Nt=2 is for 15N ; Nt=3 for 14N
     Nt = round ( (max(size(P)) - 8 ) / 2);, Nt2 = Nt*2;
     n = length(x) ;
     dx = ( x(n) - x(1) ) / (n-1);
% linegen = name of program to generate lorentzian lineshape
% pspec = spectrometer parameters required for the generating program
%         lor_simp, lor_freq --> x,l,MF
%         lor_mod,  lor_fm   --> x,l,MF,MA
%         lor_full           --> x,l,MF,MA,R1,h1
     lorstrg = ['lor_simp';'lor_freq';'lor_mod ';'lor_fm  ';'lor_full'];
     pspecmat = ['[P(Nt2+2)     ]'
                 '[P(Nt2+2)     ]'
                 '[P(Nt2+2:Nt2+3)]'
                 '[P(Nt2+2:Nt2+3)]'
                 '[P(Nt2+2:Nt2+5)]'];
     linegen = lorstrg(lorprog,:);
     pspec = eval(pspecmat(lorprog,:));
     Pline = [P(1); P(1+Nt); pspec];
     eval(['y = ',linegen,'(x,Pline);'])  %y = linegen(x,[xover;lor;pspec]);
     if (nequiv ~= 0)
       y = c13(n,dx,P(Nt2+6:Nt2+7),nequiv,y);
     end
     if(Nt > 1)
      for k = 2:Nt,
        Pline = [P(k); P(k+Nt); pspec];
        eval(['yk = ',linegen,'(x,Pline);'])
        if (nequiv ~= 0)
          yk = c13(n,dx,P(Nt2+6:Nt2+7),nequiv,yk);
        end
        y = y + yk;                %y = y + linegen(x,[xover;lor;pspec]);
      end
     end
     line = g_conv(y,dx,P(Nt2+1));   % convolve and give the answer
```

```
function [y] = lor_fm (x,P,rmax) ;
% function [y] = lor_fm (x,P,{rmax}) ;
% computes the complex lorentzian at the first harmonic for any
%   modulation amplitude and frequency
%   note: when the modulation amplitude is large the line is not lorentzian
% y = [ y1 y2], both are complex real == dispersion  imag == absorption
%     y1 == in-phase and y2== phase-quadrature wrt the zeeman modulation
% x == x axis variable in Gauss (or Hertz or ...)
% P == parameter list; all variables in P must be in the same units as x
% P == [xnot R2 modfreq modamp]'
% P(1) == xnot, the center of the line
% P(2) == R2 = 1/T2, the spin-spin relaxation rate.
% P(3) == omega_m, the modulation frequency
% P(4) == hm, the peak to peak modulation amplitude.
% yields identical results as lor_full when observer power = 0
% called by line_c13

      d2 = (P(4)/4)^2;
      if(nargin == 2), rmax =1;, end
      r = rmax;
% initial defintions of a+/-
      ap = (x - P(1) ) - i * P(2) + r*P(3);
      am = (x - P(1) ) - i * P(2) - r*P(3);
% boundary conditons for g+/-
      gp =  ap .* ( 1 + sqrt( 1  - ( P(4) ./ ( 2* ap) ) .^2 ) ) / 2 ;
      gm =  am .* ( 1 + sqrt( 1  - ( P(4) ./ ( 2* am) ) .^2 ) ) / 2 ;
% recurrence as needed
      for r = (rmax-1):-1:1
        ap = (x - P(1) ) - i * P(2) + r*P(3);
        am = (x - P(1) ) - i * P(2) - r*P(3);
        gp = ap - d2 ./ gp;
        gm = am - d2 ./ gm;
      end
% compute first harmonic signal
      a = (x - P(1) ) - i * P(2) ;
      g = a - d2 * ( ( 1 ./ gp ) + ( 1 ./ gm ) );
      ap = (x - P(1) ) - i * P(2) ;
      y = [ ( g.* gm )   ( g .* gp )  ];
      y = ( -1/ (2*pi) ) ./ y;
      y = y * [1 -i; 1 i] ;
```

```
function [y] = lor_full(x,P,rmax) ;
% function [y] = lor_full(x,P,{rmax}) ;
% computes the lineshape for a single spin at the first harmonic
% (with minor modifications any harmonic may be accessed)
% Results are two complex signals: Disp+i*Abs for both in-phase and quadurature
% x == x axis variable in Gauss (or Hertz or ...)
% P == parameter list; all variables in P must be in the same units as x
%      P == [ xnot R2 mod_freq h_mod R1 h_obs]';
% P(1) == xnot == center of resonance
% P(5) and P(2)== R1 and R2 are the rates of spin-lattice and spin-spin relax
% P(3) == mod_freq is the Zeeman modulation frequency  e.g. 100kHz = 0.0357 gauss
% P(4) == h_mod is the peak-to-peak amplitude of the Zeeman modulation
% P(6) == h_obs is the amplitude of the observer rf in the rotating frame
% calls lor_ra
% called by line_c13

    M = [ 1 0 0 0; 0 1 0 0 ; 0 0 0 0; 0 0 0 0];, E4 = eye(4);
    if(nargin == 2), rmax = 1;,  end
    rmax = max(rmax,1);
    h2 = 2*P(4)/4;
    P2 = (P(4)/4)^2;
% set up all the RA matricies (from P) that will be needed for the calculation
    for r=0:(rmax-1)
      name = [ 'RA', int2str(r) ];
      eval([name, ' = lor_ra(r,P);'])
    end
    [T , lamda ] = eig(lor_ra(rmax,P));
    lamda = diag(lamda);
    TI = inv(T);
% set up all the RG matricies that are needed for the problem.
    RG4 = [];
    for value = 1:4
        lam = lamda(value);
        a = x - P(1) + lam ;
        gam =  a .* ( 1 + sqrt( 1 - (h2 ./ a) .^ 2 ) ) / 2  ;
        RG4 = [ RG4 gam ];
    end
% now march through the values of x
    y = [];
    for point = 1:length(x)
      delta = ( x(point) - P(1) ) * E4;
      RG = T*diag(RG4(point,:))*TI;
% the starting point for downward recurrence
% now get the  G1 from RA(r)
      for r = (rmax-1):-1:1
        RG = (delta + eval(['RA',int2str(r)]) ) - P2*inv(RG);
      end
% from RG(r=1)  find X(r=1)
      RG = inv(RG);
      Y = inv( (delta + RA0 ) - (2*P2)*M*RG*M );
      Y = RG*Y(:,2);
      y = [ y  Y ];
    end
    y = [ -i 1 0 0 ; 0 0 -1 -i]*y;
    y = (-1. / pi )  * y';
```

```
function RA = lor_ra(r,P) ;
% function RA = lor_ra(r,P) ;
% P == [ xnot R2 mod_freq h_mod R1 h_obs ]';
% lor_ra only uses 4 elements of P: P(2:3) and P(5:6) but pass them all anyway
% called by lor_full to generate the matrix RA
% Set up an RA matrix  from P, r = rth harmonic of interest
      v = [P(5) (-r*P(3)) ]';
      v = (P(6)^2/(v'*v))*v;
      v = [ P(2)  r*P(3)]' + v;
      v = [ P(2) ; v ; r*P(3) ];
% the elements of v are in order [  R2 R2' r*omega' r*omega ]
      RA = [ 0 v(2) v(3) 0; -v(1) 0 0 v(4); v(4) 0 0 v(1); 0 v(3) -v(2) 0]';
```

```
function [cc] = c13(n,dx,pc13,nequiv,y)
% function [cc] = c13(n,dx,pc13,nequiv,y)
% builds C13 spectra and adds into the spectrum in y
% n = length(x);, dx = (x(n)-x(1))/(n-1);
% pc13 = [c13a c13ra]';
%   c13a = distance from low-field C13 peak to hi-field peak
%   c13ra = ratio of C13:N peak height, approx equal to natural C13 abundance
%       height of C13 peak will be ratio * nequiv
% nequiv = # equiv C13s that contribute to the C13 peaks
% y = y-vector to which C13 spectra are to be added
% called by line_c13.m
   c13a = pc13(1)/2;
   c13offset = round(c13a/dx);       %x offset for each C13 peak
   zz = zeros(c13offset-1,1);
   c13height = nequiv*pc13(2);
   c = [c13height;zz;1;zz;c13height];
% now convolve vector c with each column of y and put together in vector cc
   cc1 = conv(c,y(:,1));, cc2 = conv(c,y(:,2));
   cc = [cc1 cc2];
% truncate cc to the same length as y
   cc = cc(c13offset+1:n+c13offset,:);
```

```
function [c] = g_conv(A,dx,sigma)
% function [c] = g_conv(A,dx,sigma);
% builds gaussian of width sigma, then convolves it with complex signal A
%   (A is usually the sum of the complex lorentzian lines)
% dx = distance between points on the x axis
% returned signal c is in register with A and of identical length
% called by line_c13.m
% decide on range say 5 * sigma
      [NA,nb]  = size(A) ;
      range = 5;,range = range * sigma ;  % going +/- 5 standard deviations
      NC = round(range/dx) ;
      if(NC < 1)              % nothing to convolve
        c = A;
       else
% build gaussian c
%    make an integer range for the computed x axis
%    integers guarantee that there are 2*NC+1 elements in c
         range = NC * dx ;
         c = [ - range : dx : range ];
         c = c' ;             % make c a column vector
         w = 1. / ( sqrt ( 2.) * sigma ) ;
         g = ( dx * w / sqrt(pi) ) .* exp ( (-1.) .* ( ( w .* c ) .^ 2 ) );
% convolve A and gaussian g
         c = [];
         for n = 1:nb
           c = [ c conv(A(:,n),g) ];
         end
% truncate signal to length = NA
         c = c( NC+1 : NA+NC , :);
      end
```

```
function [ Z_out ] = rotate( Z,phi )
% function [ Z_out ] = rotate( Z,phi );
% rotate the complex vector by phi in degrees
% Z may be N by 1 or N by 2 if the quadrature signal is hung on it
% called by rot_c13, see_c13
      a = exp ( i * pi * phi /180.);
      Z_out = a * Z;
```

```
function [Z,phi_opt] = zeeman(Z_to_rot,phi,type)
% function [Z,phi_opt] = zeeman (Z_to_rot,phi,'type')
% rotate the complex inphase and out of phase vector
%   by a phase angle for the zeeman modulation
% rotate both real and imag (or disp and abs) by the same amount
% type is 'real','imag', or 'abs' and used to determine
%   phi_opt, which is the angle needed to complete the rotation
%   to an optimum angle to minimize the quadrature component
% called by rot_c13, see_c13

      c = cos(pi*phi/180);
      s = sin(pi*phi/180);
      rot = [ c s ; -s c ];
      Z = Z_to_rot * rot;
% now find null position for spectrum of type (= real, imag, or abs)
      Z_abs = eval( [ type , '(Z) '] );
      Z_bar =  [ mean(Z_abs(:,1))  mean(Z_abs(:,2)) ];
      for i = 1:2
        Z_abs(:,i) = Z_abs(:,i) - Z_bar(i);
      end
      V = Z_abs' * Z_abs ;
      phi_opt = atan2( (-2.*V(1,2)) ,  (V(1,1) - V(2,2)) ) / 2. ;
      phi_opt = 180 * phi_opt / pi;
```

```
function [std_err,r_sq,z_fit] = see_c13(po,nxover,lorprog,nequiv,signal,px)
% function [std_err,r_sq,z_fit] = see_c13(po,nxover,lorprog,nequiv,signal,px)
% computes fit for the n xover case, generated from the parameters in po,px
% MUST global Gauss and Z_Data before the call
% po, px = COLUMN vectors of parameters, defined in p1, p2, p3.m, to build P
%   po = [ x1 {a1} x2 {a2} ... floating_variables]'
%   px = [ n1 {a1} n2 {a2} ... dummy fixed_variables]'
%     P = [x1 ... xn lor1 ... lom sigma MF MA R1 h1 rot_phase c13a c13ra]';
% nxover = # crossovers (1 for single line, 2 for N15, 3 for N14)
% lorprog = flag to determine which line-generating program to use
%       1 = lor_simp; 2 = lor_freq; 3 = lor_mod; 4 = lor_fm; 5 = lor_full
% nequiv = # equivalent C13s to contribute to the C13 peaks
%    to not include the C13 peaks in the analysis, set nequiv = 0,
%    and assign dummy #s to c13a and c13ra
% signal is 'imag' ie absorption A
%        or 'real' ie dispersion D
%        or 'magn' ie absolute magnitude sqrt( A^2 + D^2)
% calls buildp,line_c13,rotate,zeeman,coef
% called by integc13

% include for Matlab 4.0 versions
      global Gauss Z_Data

% send to buildp to make the necessary P
      [P,rot_phase] = buildp(po,px,nxover);
% generate lorentzian
      lor = line_c13(Gauss,P,lorprog,nequiv);
      lor = rotate(lor,rot_phase);
      [lor,phi1opt ] = zeeman(lor,0.0,'imag');
      [lor,phi2opt ] = zeeman(lor,phi1opt,'imag');

      c_d = cov([ real(Z_Data) real(lor) ]);
      c_a = cov([ imag(Z_Data) imag(lor) ]);
      c_m = cov([ abs(Z_Data)  abs(lor)  ]);

% would be more elegant to have the next 4 lines here rather than in the
%   following if-then loop, but that results in error messages of
%   'dividing by zero!'  It still works but is annoying
%       rr = ( coef(c_d,1,3) ) .^ 2;
%       ri = ( coef(c_a,1,3) ) .^ 2;
%       sr = sqrt( (ones(rr) - rr ) .* [ c_d(1,1) c_d(2,2) ] );
%       si = sqrt( (ones(ri) - ri ) .* [ c_a(1,1) c_a(2,2) ] );

% assign sigtype, r_sq, and std_err according to signal
      if (signal == 'real')
         sigtype = 'real';, cmat = c_d;
         rr = ( coef(c_d,1,3) ) .^ 2;
         sr = sqrt( (ones(rr) - rr ) .* [ c_d(1,1) c_d(2,2) ] );
         r_sq = rr;, std_err = sr;
      elseif (signal == 'imag')
         sigtype = 'imag';, cmat = c_a;
         ri = ( coef(c_a,1,3) ) .^ 2;
         si = sqrt( (ones(ri) - ri ) .* [ c_a(1,1) c_a(2,2) ] );
         r_sq = ri;, std_err = si;
      else                        %signal == 'magn'
```

```
        sigtype = 'abs';, cmat = c_m;
        rr = ( coef(c_d,1,3) ) .^ 2;
        ri = ( coef(c_a,1,3) ) .^ 2;
        sr = sqrt( (ones(rr) - rr ) .* [ c_d(1,1) c_d(2,2) ] );
        si = sqrt( (ones(ri) - ri ) .* [ c_a(1,1) c_a(2,2) ] );
        r_sq = rr + i*ri ;, std_err = sr + i*si ;
    end
% generalized calc
    Y_bar = eval(['mean(',sigtype,'(Z_Data))']);
    Y_hat_bar = eval(['mean(',sigtype,'(lor))']);
    scale = [cmat(1,3)/cmat(3,3) cmat(2,4)/cmat(4,4) ];
    base = Y_bar - scale .* Y_hat_bar ;
    z_fit =  [ ( scale(1) * lor(:,1) + base(1) )  ...
             ( scale(2) * lor(:,2) + base(2) ) ];
% for ex., Y_bar = mean(real(Z_Data));
%       Y_hat_bar = mean(real(lor));
%       scale = [ c_d(1,3)/c_d(3,3)  c_d(2,4)/c_d(4,4) ];
```

```
function  reg = coef(c,k,l)
% function  reg = coef(c,k,l);
% computes an ordered pair of regression coefficients from covariance matrix c
    c1 = c(k,l) / sqrt( c(k,k) * c(l,l) );
    k1= k+1;
    l1 = l+1;
    c2 = c(k1,l1) / sqrt( c(k1,k1) * c(l1,l1) );
    reg = [ c1 c2 ];
```

APPENDIX C: MATLAB CODE FOR ANALYSIS OF DNA DATA

The following MATLAB code was used in the analysis of the DNA data discussed in Chapters 4 and 5. All the code presented here was written by Bruce H. Robinson.

Chapter 4: Length and Position Dependence of DNA Dynamics

code name	code function
saswu	uses bendm to calculate flexural τ for motional modes
bendm	calculates amplitude and rate for flexural motional modes
bend_wc	calculates $\langle \eta_i^2 \rangle$ for length analysis with variable weighting
bend_amp	calculates $\langle \eta_i^2 \rangle$ for length analysis with no weighting
twist	calculates $\langle \varphi_i^2 \rangle$ with no weighting
twistamp	calculates torsional τ for motional modes

Chapter 5: Sequence Dependence of DNA Dynamics

code name	code function
mod_beta	calculates $\langle \eta_i^2 \rangle + \langle \beta_0^2 \rangle$ and compares to $\langle \beta_i^2 \rangle$
amp_fc	calculates $\langle \eta_i^2 \rangle$
forcek	calculates force constants string for given sequence
oseen	generates friction factor matrix (Oseen matrix) for A-tract
rot	rotates Euler angles to generate rotation matrix
rot_inv	generates Euler angles from rotation matrix
spiral	back-calculates coordinates for all spheres via Euler rotations

```
function [tau] = saswu(Pl,visc,T,N)
%function [tau] = saswu(Pl,visc,T,N)
% finds rates from bendm program (using Song's formalism)
%   but assumes that they are the rates (from Wu) for the slowest modes of Q2
%   and that that faster modes are too fast to reduce the effectiveness of those amplitudes.
% Pl = persistance length / step height (eg, 1250 A / 3.4 A)
% visc = viscosity in poise (usu. 0.01 p)
% T = temperature in C
% N = length of DNA in base pairs

T = T + 273;              % temperature in K
fc = Pl * 1.38e-16*(T);   % force constant = (Pdf / h) * kboltz * T
l_bp = round(N/2);        % position (irrelevant, but need a #)
                          % rates apply to all the positions, Mode_amp only to position l_bp
%[phi,Mode_amp,rate] = bendm( time,  0,fc,visc,Temp,position,num_beads)
  [phi,Mode_amp,rate] = bendm([0:10]', 0,fc,visc,T   ,   l_bp      ,N);
        tau = (1./rate)';
```

```
function [phi,Mode_amp,rate] = bendm(time,ff,fc,eta,Temp,l_bp,N_bp)
% [phi, Mode_amp, rate] = bendm(time,friction_factor,force_constant,viscosity,Temp,l,N_bp)
% bend mute (no echo)
% equations from Song Allison and Schurr (SAS) Biopolymers 29, 1990
% compute the bending decorrelation at position l, for DNA
%   with a single base friction factor (ff)  = gamma_perp
%   and a hookean spring force constant fc ( Kappa sub beta in SAS)
% the DNA has N_bp base pairs.  (nb. N_bp = N+1 ala JMS)
% all quantities in cgs units please.
      k_boltz = 1.38e-16;          % boltzmann's constant in cgs
      h_bp = 3.4e-8;               % rise per base pair in cm
      h_sub = 31.8e-8;             % diameter of the effective (mesoscopic) sphere
% number of subunits equivalent to the length of dna chosen
      N_sub = round( ( h_bp/h_sub) * N_bp );
      h_sub = h_bp * N_bp / N_sub;
% redo the sphere diameter so that the lengths are indeed the same
       l = round( (h_bp/h_sub) * l_bp );
      if(l < 1), l=1;, end
      if(N_sub <= l ), l = N_sub-1;,end
% the bending persistence length is related to the bending force constant
      P_bend = h_bp * fc  / ( k_boltz * Temp );  % in cm
         ff_eff = (3*pi*eta)*(h_sub^3);     %the rotational friction factor of the effective sphere
% calculate phi_infinity
      phi_infinity = ( ( N_bp * h_bp )/ (12 * P_bend) ) * ( 1 + 3* ( ((2*l_bp - 1 - N_bp)/N_bp) ^ 2 ) );

% now do the dynamic part using the scaled sizes
% construct eigenvalues and eigenvectors lamda and Q.
      N1 = N_sub-1;
      D1 = diag(ones(N1-1,1),1) - diag(ones(N1,1),0) ;
      D1 = [ D1 zeros(N1,1) ];
      D1(N1,N_sub) =1;
      D2 = D1* D1';
      D2(1,1) = 1;, D2(N_sub-1,N_sub-1) = 1;
      D4 = D1'*D2*D1; % D4 is a 4th derivative in centeral difference form
% use the Toeplitz form to generate the modified Oseen Tensor H
      Toe  = (1:(N_sub-1));
      Toe = 1 ./ Toe ;
      H =  [ 1   ( ((3/8) * Toe) .* ( 1 + (Toe .* Toe)/6 ) ) ];
      H = toeplitz(H);
      [ Q, Lamda ] = eig(H*D4);
      Lamda =diag(Lamda);
      [ Lamda, Index ] = sort(Lamda);
      Q = Q(:,Index);

% rationalize and normalize the two uniform modes
      Q1 = ones(N_sub,1);
      Q(:,1) = Q1 ;
      Q2 = (1:N_sub)' - (N_sub+1)/2;
      Q(:,2) = Q2;
% this implies that b = 1 of eqn 56 SAS
% normalize the uniform modes by assuming Q'*D4*Q = 1.
      m1 = Q1'*Q1;
      m2 = Q2'*Q2;
      mu =  diag(Q'*D4*Q)';
      mu = sqrt([ m1 m2 mu(3:N_sub) ]);
```

```
      mu = ones(N_sub,1) * mu;
      Q = Q ./ mu;

% develop the alternative way to Q, called Qa:
% Qa is the same as Q up to a sign of each column vector.
% testing this shows that it adds nothing to the above
% it can differ only by a sign of each column.
%       [ C, C_lam ] = eig(D4);
%       C_lam = diag(C_lam)
%       [ C_lam , index] = sort(C_lam)
%       C = C(:,index);
%       C_lam(1) = 0;,C_lam(2) = 0;
%       Cl12 =  sqrt(C_lam)
%       D12 = C * diag(Cl12) * C';
%       [ B , DHD_lam ] = eig(D12'*H*D12)
%       DHD_lam = diag(DHD_lam)
%       [ DHD_lam,index] = sort(DHD_lam)
%       B = B(:,index)
%       DHD_lam(1) = 0;,DHD_lam(2) = 0;
% test that B' * B = 1
% test that C' * C = 1.
%       ns = length(C)
% shrink C for later use
%       C = C(:,3:ns);
%       C_lam_inv =  1 ./ Cl12(3:ns)
%       CLC = C * diag(C_lam_inv) * C'
% shorten B and DHD_lam  compare DHD_lam to Lambda
%       B = B(:,3:ns);
%       Qa = H*D4*CLC*B*diag( (1 ./ DHD_lam(3:ns)) )
%       Qa = [ Q(:,1:2) Qa ]  % add in the uniform modes

% obtain the translation and rotational diffusion coefficients for the two uniform modes
      Hinv = inv(H);
      qhq = 1 / ( Q2' * Hinv * Q2 );
% estimate the size of the sphere 31 A above  Should be self consistent
      H_sub = ( ff*N_bp*qhq/(3*pi*eta) ) ^ (1/3);
      Mode_amp = ( Q(l+1,:) - Q(l,:) ) .^ 2;
      Mode_amp = (h_sub/P_bend) * Mode_amp(3:N_sub);
%       tau = ( (N_bp*qhq)*(h_sub/h_bp) * (ff/fc) ) ./ Lamda(3:N_sub);
      rate = Lamda(3:N_sub) * ( (fc/ff_eff)*(h_bp/h_sub));

      phi =  exp( time * ( - rate' ) ) * Mode_amp' ;
      phi = sum(Mode_amp) - phi;
```

```
function [eta_sq,eta_weighted] = bend_wc(n,Pl,ratio,Temp,wt)
% [eta_sq,eta_weighted] = bend_wc(n,Pl,ratio,Temp,wt)
% computes the bending amplitude of all the normal bending modes
% modified from bend_amp with a correction factor to weight the modes
%    to reduce weight of long modes for long length DNAs
% eta_sq is the mean square amplitude of oscillation at infinite time for each base_pair
% eta_weighted is eta_sq, weighted
% n = number of base pairs of DNA = number of "beads"
% Pl = persistence length = ratio of persistence length: bead diameter;
%   (bead diam = h = 3.4A) for example 1250/3.4
% ratio = [ kappa'/kappa  begin_bead end_bead; change force constant over a subregion of DNA
        r = ratio(1);, n1 = ratio(2);, n2 = ratio(3);
        o = ones(n-1,1);
        o(n1:n2-1) = r * o(n1:n2-1) ;
        A =  diag(o,1) + diag(o,-1);
        t = sum(A);
        A = ( diag(t) - A );
        [Q,lamda] = eig(A);              %  A = Q * lamda * Q'
        lamda = diag(lamda);             % sort eigenvalues; remove zero
        [ lamda, Indx] = sort(lamda);
        Q = Q(:,Indx);
        drop = 2;
        Q = Q(:,drop:n);
        lamInv = 1 ./ lamda(drop:n);
        Q2 = Q .* Q;
        eta_sq = 2 * ( Q2 * lamInv ) /Pl;
% amp_sq = amplitude (squared) at each base from each mode, size = row x col
% row index = position on the DNA; column index = particular mode at that point
        [ nq , mq ] = size(Q2);
        amp_sq =  2 * ( Q2 .* (ones(nq,1)*lamInv' ) ) /Pl;
% get the weighting factors for the individual amplitudes
% [phi,Mode_amp,rate] = bendm( time,  0, fc , visc , Temp, position, n)
% rates apply to all the positions, Mode_amp only to position l_bp
% here, only want rates (taus proportional to 1/rates)
        Temp = Temp + 273;               % temperature in K
        fc = Pl * 1.38e-16*(Temp);       % force constant = (Pdf / h) * kboltz * T
        visc = 0.01;                     % viscosity in poise
        l_bp = round(n/2);               % position (irrelevant, but need a #)
        [phi,Mode_amp,rate] = bendm([0:10]', 0, fc , visc , Temp, l_bp, n);
        Azzhz = 48 * 2.8e6;                              % Azz in Hz
        Aperphz = ( (9.223 + 6.982)/2 ) * 2.8e6;         % Aperp = (Axx+Ayy)/2
        taus = zeros(1,mq);
        tau = (1./rate)';
        taus(1:length(tau)) = tau;
        taus = ones(nq,1)*taus;
% use the rates from bend program (using Song's formalism)
%  but assume that they are the rates (from Wu) for the slowest modes of Q2
%  and that that faster modes are too fast to reduce the effectiveness of those amplitudes.
% fract will be the fractional weighting of the amplitudes as given by Q2 for each element of Q2.
% Q2 is more or less square, coupling every position along the chain to each and every
%  normal mode of the chain.
% use 2 * abs(...) so that fractn -> 1 for isotropic motion
        fractn =  2*taus .* abs( -Azzhz + Aperphz) .* ( sin(sqrt(amp_sq)).^2 );
        weighting = exp(-wt * fractn);
        eta_weighted  =  ( weighting .* amp_sq )* ones(mq,1);
```

```
function [eta_sq] = bend_amp(n,Pl,ratio)
% [eta_sq] = bend_amp(n,persistence_length,ratio)
% computes the bending amplitude of all the normal bending modes
% eta_sq is the mean square amplitude of oscillation at infinite time for each base_pair
% n = number of base pairs of DNA = number of "beads"
% Pl = persistance length = ratio of persisitence length: bead diameter;
%    (bead diam = h = 3.4A) for example 1250/3.4
% ratio = [ kappa'/kappa  begin_bead end_bead ]
% ratio will change the force constant over a subregion of the DNA
        r = ratio(1);, n1 = ratio(2);, n2 = ratio(3);
        o = ones(n-1,1);
        o(n1:n2-1) = r * o(n1:n2-1) ;
        A =  diag(o,1) + diag(o,-1);
        t = sum(A);
        A = ( diag(t) - A );
        [Q,lamda] = eig(A);              %  A = Q * lamda * Q'
        lamda = diag(lamda);             % sort eigenvalues; remove zero
        [ lamda, Indx] = sort(lamda);
        Q = Q(:,Indx);
        drop = 2;
        Q = Q(:,drop:n);
        lamInv = 1 ./ lamda(drop:n);
        Q2 = Q .* Q;
        eta_sq = 2 * ( Q2 * lamInv ) /Pl;
```

```
function  phi = twist(time,ff,fc,Temp,l,N_bp)
% phi = twist(time,friction_factor,force_constant,Temp,l,N_bp)
% computes the twisting decorrelation at position l, for DNA with a single base friction factor (ff)
%             and a hookean spring force constat (fc also called alpha)
% DNA has N_bp base pairs.  (note: not N+1 ala JMS)
% all quantities in cgs units please.
        k_boltz = 1.38e-16;
        L = 1:(N_bp-1);
        L = L';
        lamda = 4 *( sin( L * (pi/(2*N_bp)) ) .^ 2);
        rate = - lamda' * (fc/ff);
        d_squared = ( k_boltz * Temp / fc ) ./ lamda;
        Q_squared = (2/N_bp) * ( cos( ( pi*(2*l-1)/(2*N_bp) ) * L ) .^ 2);
        phi_infinity = d_squared' * Q_squared;
        Qd = ( d_squared .* Q_squared );
        phi = phi_infinity - ( exp( time * rate ) * Qd ) ;



function [phi_inf_squared,tau,phi_truncated] = twistamp(fc,ff,T,N)
%[phi_inf_squared,tau,phi_truncated] = twistamp(force_constant,friction_factor,Temp,N_bp)
% computes the mean square twisting angle and the relaxation times
%       for the internal twisting modes.
% uses the force_constant (fc) and the friction factor for a single base (ff)
% D = Kb*T/ff, and tau = 1/(6D)  = (viscosity/Kb*T) * Sphere_Volume
% T = Temperature in Kelvin
% fc and ff are in cgs units, typically 4e-12,6e-23 for 20C respectively
% there are N bases

        K_b = 1.38e-16;         %ergs/K
        t = - ones([ N-1,1]);
        A = diag(t,1) + diag(t,-1);
        center = sum(A);
        A = A - diag(center);
        % A is the force matrix:  U = (1/2) fc * phi'*A*phi
        [ Q, Lamda ] = eig(A);   % Q'*A*Q = Lamda;  Q' = inv(Q)
% need to sort Lamda into ascending order and remove lowest eigenvalue
        [ rate, index  ] = sort(diag(Lamda));
        Q = Q(:,index);
        Q = Q(:,2:N);
        rate = rate(2:N);
        Lamda = diag(rate);
        phi_inf_squared = (K_b*T/fc)* diag(Q*inv(Lamda)*Q');
        tau = (ff/fc) ./ rate;
% now generate phi_truncated = phi calculated w/ only the first 4 modes
        Q2 = Q .* Q;
        d2 = (K_b*T/fc) ./ rate;
        max_modes = min(4,N-1);
        Q2 = Q2(:,1:max_modes);
        d2 = d2(1:max_modes);
        phi_truncated = Q2*d2;
```

```
function [beta_hat_sq] = mod_beta(X,Fc_vec,Sequence)
% function [beta_hat_sq] = mod_beta(X,Fc_vec,Sequence)
% X =  counter 1 to N, N = # sequences to be analyzed.  Rseq = sequence(s), N (# seq) x m (bp)
% Fc_vec is the force constant vector, which we need to turn into a matrix.

Fc = reshape(Fc_vec,4,4);
%now we make it symmetric and use only the lower diagonal
fct = Fc';
for k=1:3
  Fc(k,k+1:4) = fct(k,k+1:4);
end

spin_probe_position = 6;
beta_hat_sq = [];
N = length(X);

for k = 1:N
  Fc_seq = forcek(Sequence(k,:),Fc);

% use the sequences now to generate the beta_hat_sq
  [eta,etaw] = amp_fc(Fc_seq');
  beta_hat_sq = [ beta_hat_sq ; eta(spin_probe_position)];
%  beta_hat_sq = [ beta_hat_sq ; [eta(spin_probe_position) etaw(spin_probe_position)] ];
end

% correct beta_hat_sq for persistence length and intercept
P_length = 1250;
beta_sq_not = 0.0214;
beta_hat_sq = (3.4/P_length)*beta_hat_sq + beta_sq_not;
```

```
function [eta_sq,eta_weighted] = amp_fc(fc_vec)
% [eta_sq,eta_weighted] = amp_fc(fc_vec)
% calculates etasq = mean square amplitude of oscillation at infinite time for each basepair
% eta_weighted is weighted for a 50mer, wt = 4, SL in position 6 ONLY

        n = 1 + length(fc_vec);          % n == number of beads (i.e. base pairs) of DNA
% form the A force matrix
        A =diag( [ 0; fc_vec] + [fc_vec;0 ] )  - diag(fc_vec,1) - diag(fc_vec,-1);
        [Q,lamda] = eig(A);              %  A = Q * lamda * Q'
        lamda = diag(lamda);             % sort eigenvalues; remove zero
        [ lamda, Indx] = sort(lamda);
        Q = Q(:,Indx);
        drop = 2;
        Q = Q(:,drop:n);
        lamInv = 1 ./ lamda(drop:n);
        w = ones(size(lamInv) );
        w(1:3) = [ .8516 ;.9939; .9993 ]; % weighting for 50mers at position 6, wt = 4
        Q2 = Q .* Q;
        eta_weighted = 2 * ( Q2 * (w.* lamInv));
        eta_sq = 2 * ( Q2 * lamInv);
```

```
function [Fc] = forcek(Sequence,Fc_table)
% function [Fc] = forcek(Sequence,Fc_table)
% generates the string of pairwise force constants from the sequence
% eg  Sequence = 'AATTATATGCGCG';  etc
% then you get back Fc = [ k1 k2 k3 k4  .....  k(N-1) ]
% from N basepairs get N-1 force constants.

% Table: (of force constants)  interpretation of table
%
%                T    A    C    G
%       A       (1,1) (1,2) (1,3) (1,4)
%       T
%       G
%       C
% e.g.  AT is fc_table(1,1) etc.
% Of the 16 force constants only 10 are independent because
%       AA = TT  GG = CC  AC = GT  AG = CT  TC = GA  TG = CA
%       12 = 21  34 = 43  13 = 31  14 = 41   23 = 32  24 = 42
% The 10 elements are positioned as:
%       3  1  5  4
%       1 10  7  6
%       5  7  8  2
%       4  6  2  9
% This is a symmetric matrix
%  The 5' lead in is in order ATGC and the 3' exit side is in order TACG

N = length(Sequence);

Sequence = Sequence';
Targ1 = [ 'ATGC' ]';
Targ2 = [ 'TACG' ]';
Val1 = zeros(size(Sequence));
Val2 = Val1;
for k=1:4
        index = find( Sequence == Targ1(k));
        Val1(index) = k*ones(size(index));
        index = find( Sequence == Targ2(k));
        Val2(index) = k*ones(size(index));
end

Fc =[];
for k = 2:N
        Fc = [ Fc Fc_table(Val1(k-1),Val2(k)) ];
end
```

```
function [ff] = oseen(r,a);
% function [ff] = oseen(r,[a]);
% constructs the Modified Rotne-Prager Oseen tensor for a chain of
%  contiguous spheres where the spheres are centered at coordinates r.
% r is 3 across [ x y z] by N down
% r is the set of the 3 cartesian coordinates of the center of each sphere
% ff == the friction factor tensor in REDUCED (i.e. dimensionless) units
% ff_true = zeta*ff;  see Garcia de-la Torre Bipolymers 1990.
%   zeta = 6*pi*viscosity*a;
% a == sphere radius, cgs units, only supply a if only one bead (N=1)

% eta = viscosity in cgs = poise
% Temp = temp in K
% D = the diffusion tensor
% The Modified Oseen tensor, will be 3N by 3N when finished.
%       Kb = 1.38e-16;  % cgs Boltzman constant
%       KbT = Kb * Temp;
%       zeta = 6*pi*eta*a;

        [ N , m ] = size(r);        % m must be 3
        Dyad =[];
        I3 = eye(3)      ;          % a 3 by 3 idendity matrix
        O = ones(N,1);  % column vector of 1s
        Z = zeros(N,1);
        I_N = kron(O,I3);

% determine that all beads are the same distance apart and find the total length
if( N > 1)
        dr = diff(r);
        dnorm = sqrt(diag(dr*dr'));        % the nearest neighbor norms
%the values of dnorm should all be the same and be 2*a, where a is the
% radius of the sphere.
        a = mean(dnorm)/2;
        range =  max(dnorm)/2 - min(dnorm)/2;
        disp([ 'sphere size = ' num2str(a) '   range = ' num2str(range) ] )
end
        r = ( r - O * mean(r) )/a;
% Coordinates are moved to center of mass and scaled to sphere size.
% now form the modified oseen tensor matrix:
        Dyad = [];
        for k=1:N
         rk = r(k,:);
         rmk = (r - O*rk); %distances are already scaled by a
% rmn are the relative distances between all beads and the kth bead, scaled to a
         rmk(k,:) = [ 1 1 1]*sqrt(2/3);
         r_norm = sqrt( diag(rmk*rmk') );
         r_unit = rmk ./  ( r_norm * [ 1 1 1 ] );
         Dyadik = [];
         for I=1:N
          rI = r_unit(I,:);
          A = rI'*rI;
          A = (3/(4*r_norm(I)))* ( (I3 + A) + (2/(r_norm(I)^2))*(I3/3 - A ) );
          Dyadik = [ Dyadik; A ];
         end
         Dyad = [ Dyad Dyadik ];
```

```
        end

% now the Dyad Oseen tensor is developed
% have r the set of positions r is N by 3
% Dyad is 3N by 3N.

% for the center position r = [ 1 1 1 ] * sqrt(2/3); r'r' = 2
% then zeta*Tensor = I3* 1/sqrt(2)

        Dyad = Dyad + ( 1 - (1/sqrt(2)) )*eye(3*N);
% ensures that the diagonal terms are indeed all 1
        S = inv(Dyad);

% S  is the S matrix of de la Torre

% set up R_cross
% works so that B x r  =  B * R_cross
        R_cross = [ Z r(:,3) -r(:,2) -r(:,3) Z r(:,1) r(:,2) -r(:,1) Z ];
        R_cross = reshape(R_cross',3,3*N)';

        ff_t = I_N' * S * I_N ;
        ff_c = I_N' * S * R_cross;

% use the above friction factors to find the hydrodynamic center
% according to d-I-T:

        Zc = ff_c - ff_c';
        Zc = [ Zc(3,2) ; Zc(1,3) ; Zc(2,1) ];
        r_o = ( trace(ff_t)*I3 - ff_t ) \ Zc;

% now shift the r's by r_o and reconstruct R_cross
        r = r - ( O * r_o' );
        R_cross = [ Z r(:,3) -r(:,2) -r(:,3) Z r(:,1) r(:,2) -r(:,1) Z ];
        R_cross = reshape(R_cross',3,3*N)';

% the full 6 by 6 ff  (ff is the friction factor tensor sans zeta)
% by moving to the center of hydrodynamic stress the
% coupling should be almost nil, or ff should be block diagonal

        ff =  [ I_N R_cross ]' * S * [ I_N R_cross ] ;

% correct the ff by including on the diagonal part of  ff  N times the
% rotational ff for a single bead (which is 4/3).
        iso = N*4/3;
        ff = ff + diag( [ 0 0 0 iso iso iso]);
% at this point ff is the reduced friction factor

% rescale ff back so that it only need be multipled by zeta
%       ff = ff .* kron  ( [ 1 a ; a (a^2) ] ,ones(3,3) );
%       D = (KbT/zeta) * inv(ff);

% for ROTATION
% by defining tau = 1/6D for the rotational part of D (no coupling to translation)
% tau = ff /6kT = reduced_ff * a^2 * zeta / 6KbT
% for a single bead tau_not  = Vol * eta /KbT and the reduced_ff = 4/3
```

```
% therefore  Tau = Tau_not * reduced_ff * (3/4)
% where Tau_not  is the rotational corelation time for a single bead

%for TRANSLATION
% the translational diffusion coeffient for a single bead is D_not = KbT/zeta
% therefore  D = D_not * inv(reduced_ff)
% D is the 6 by 6 diffusion tensor.
```

```
function [R] = rot(euler)
% function [R] = rot(euler)
% develops the 3 by 3 rotation matrix for two or three euler angles
% euler = [ phi theta psi] as defined in Goldstein (or [theta psi])
%       euler in radians please
% phi rotates around z
% theta rotates down from z around y
% psi rotates around the new z axis

m = length(euler);

if(m == 2)
%  equivalent to the m=3 case assuming phi=0
        c = cos(euler(1));, s = sin(euler(1));
        R = [ 1 0  0; 0 c s; 0  -s c ] ;
        c = cos(euler(2));, s = sin(euler(2));
        R = [ c s 0; -s c 0; 0 0 1] * R;
end

if(m == 3)
% euler = [ phi theta psi] as defined in Goldstein
        c = cos(euler(1));, s = sin(euler(1));
        R = [ c s 0; -s c 0; 0 0 1] ;
        c = cos(euler(2));, s = sin(euler(2));
        R = [ 1 0  0; 0 c s; 0  -s c ] * R;
        c = cos(euler(3));, s = sin(euler(3));
        R = [ c s 0; -s c 0; 0 0 1] * R;
end
```

```
function [euler] = rot_inv(R)
% function [euler] = rot_inv(R)
% takes the R rotation matrix as defined in  Goldstein pg 109
% and back-figures the euler angles = [phi theta psi], in degrees

        theta = acos(R(3,3));
        st = sin(theta);
if (abs(st) < 1.e-4)
%  ie, if theta < 0.01 degrees, call it either pi or 0, then cos(theta) = -1 or 1
        psi=0;% arbitrary choice
        phi = atan2(R(1,2),R(1,1))
        if(phi < 0) , phi = 2*pi + phi;, end
else
        R = R/st;
        phi = atan2(R(3,1),-R(3,2));
        if(phi < 0) , phi = 2*pi + phi;, end

        psi = atan2(R(1,3),R(2,3));
        if(psi < 0) , psi = 2*pi + psi;, end
end
        euler = (180/pi) * [ phi theta psi ];
```

```
function [s,sori] = spiral(euler)
% [coordinates,effective_sphere_coordinates] = spiral(euler)
% euler (angles) == a 3 by N set of rotation angles in degrees
% the top line is phi(k)
%    the middle line is the theta angle (rot about y axis off z axis), and
%    the bottom line is the psi angle (rot about new z axis) for the kth base
%                    relative to the k-1 base.
% coord = the 3 cartesian coordinates by N coordinates of the centers of the N bases
% coord assumed to be coord_z = 1:N
% effective_sphere_coordinates = coordinates of the mesoscopic spheres
%    (radius 15.9 A)
% rotates Nth base first relative to the N-1 origin, then iteratively works backwards
%    along the chain to determine where the first origin is.
%    (first rotation is not applied)

          [m N] = size(euler);
          euler = euler * (pi/180);
          euler = -flipud(euler);     %the inverse of the euler rotations
          s0 = [ 0 0 1]';
          s =s0;
for k = N:-1:2
          R = rot(euler(:,k));
          s = [ s0 R*s];
end
          s = cumsum(s');

% find the places on the line that are a distance l apart
% l is dimensionless, (mesoscopic sphere diameter = 2*15.9A) / (rise per bp = 3.4A)
          l = 2*15.9/3.4;
          L2 = l*l;
          ori = [ 0 0 -(l+1)/2];        % to find the first origin of mesoscopic sphere
          index = 1;
          sori =[];
while( index < N-l/2)
       dist=sqrt(sum( ((s-ones(N,1)*ori).^2)' ) );
          index = find( (dist <= l) );
          index = max(index);
if( index > N-1), index = N-1;, end

          orig = ori;
          ori = s(index,:)-orig;

          L02 = ori*ori';
          del = s(index+1,:) - s(index,:);
          eps = ( sqrt( ((del*ori')^2) + (del*del')*(L2-L02) ) -del*ori' ) / (del*del');
          ori = s(index,:)+ eps*del;

          sori = [ sori ; ori];
end
```

APPENDIX D: FAST MOTION THEORY

Longstanding theory[1-4] predicts that the electron spin-lattice relaxation rate R_{1e} and spin-spin relaxation rate R_{2e} should both depend on the nuclear spin quantum number, m. However, experiment shows that R_{2e} depends strongly on m,[2] while R_{1e} does not.[3,5] To resolve this discrepancy, a new theory has been developed, which reconciles theory and experiment.

The original theory was based upon Redfield theory,[4] which considers the density matrix elements. The new unified theory is based on Abragam's approach to relaxation, which considers the relaxation of coupled observables.[6] Full calculations using either approach yield the same results. However, approximations made to simplify calculations using the Redfield theory resulted in the incorrect prediction for R_{1e}. The new theory, which starts from Abragam's approach, permits and predicts that R_{1e} is independent of m and R_{2e} is dependent on m.

To prepare the way for the complete computer simulations used to test and verify the new theory, a number of preliminary simulations were run. This very early code generated and checked the tensors from the Wigner-Eckart theorem and Racah's rules.[7,8] The code also checked the tensor properties of commutation, orthogonality, and orthonormality. Once the tensors were correctly generated and their properties checked, the full simulations were coded by Dr. Robinson and run.

After much development and revision by Drs. Robinson and Mailer, the theory was presented at the 18th International EPR Symposium in Denver (1995) and subsequently submitted to, and rejected by, the Journal of Chemical Physics. The paper is under revision again at this time.

NOTES TO APPENDIX D

1. Kivelson, D. *Journal of Chemical Physics* **33**, 1094-1106 (1960).

2. Hwang, J.S., Mason, R.P., Hwang, J.H., & Freed, J.H. *Journal of Physical Chemistry* **79**, 489-511 (1975).

3. Percival, P.W., & Hyde, J. S. *Journal of Magnetic Resonance* **23**, 249-257 (1976).

4. Redfield, A.G. *Physical Review* **98**, 1787-1809 (1955).

5. Robinson, B.H., Haas, D.A., & Mailer, C. *Science* **263**, 490-3 (1994).

6. Abragam, A. *The Principles of Nuclear Magnetism*. Oxford University Press, Oxford, 1961.

7. Edmonds, A.R. *Angular Momentum in Quantum Mechanics*. Princeton University Press, Princeton NJ, 1974.

8. Rose, M.E. *Elementary Theory of Angular Momentum*. John Wiley, NY, 1957.

VITA

Annabelle Wey Reese

University of Washington

1996

BS in Chemistry, Yale University, Cum Laude, June 1991.