DNA Studies: A Novel Structural Transition, Relaxation of Secondary Structure by TOPO I, and Resolution of a PCR Problem

Greg Patrick Brewood

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

Doctor of Philosophy

University of Washington

2006

Program Authorized to Offer Degree: Department of Chemistry

UMI Number: 3224191

INFORMATION TO USERS

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 bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send 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.

UMI®

UMI Microform 3224191

Copyright 2006 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Greg Patrick Brewood

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

J. Michael Schurr

Reading Committee:

J. Michael Schurr

Bruce H. Robinson

Daniel T. Chiu

Date: 5/31/06

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 the 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 Proquest Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, 1-800-521-0600, 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 Jay Brewer

Date 5/31/06

University of Washington

Abstract

DNA Studies: A Novel Structural Transition, Relaxation of Secondary Structure by TOPO I, and Resolution of a PCR Problem

Greg Patrick Brewood

Chairperson of the Supervisory Committee:
Professor J. Michael Schurr
Department of Chemistry

Both theory and experiments were employed previously to investigate the effects of small neutral osmolytes on the twist energy parameter (E_T) and average intrinsic twist (l_0) of p30δ DNA (4932 bp). The results for the supercoiling free energy and intrinsic twist at 37°C, in ethylene glycol (EG) and acetamide (Ac) suggested, and were interpreted in terms of a two-state model. The DNA was assumed to undergo a conformational transition between two states that had different numbers of waters 'associated' with the duplex. Those previous studies, however, only extended to 20 (w/v)% ethylene glycol where the transition is only about half-completed. Hence, quantitative assessments of the twist energy parameter, which governs the supercoiling free energy, and of the intrinsic twist of the more dehydrated state were lacking. In the present experiments, the effects of EG over the range from 0 to 40 (w/v) % are examined and certain properties of the more dehydrated state ascertained for the first time.

The torsion elastic constants of certain DNAs have exhibited a temporal evolution, on a time scale of two weeks or more, subsequent to a change in deformational strain, which implies that an intensive variable, namely the superhelix density, has traversed one or more structural transition boundaries, leaving the DNA in a metastable state. The observed time-scale for these structural transitions is consistent with a model, wherein, any given short sequence of DNA embedded in a longer DNA transiently fluctuates among two (or more) distinct secondary structures that extend over much larger domains of variable position and size, and whose relative stabilities depend on

distant as well as close-lying base pairs. In general, one expects to observe very slow kinetics and metastable states for highly cooperative structural transitions, due to the difficulty of nucleating a small island of one structure in the midst of a large domain of the other.

While attempting to amplify the sequence of DNA from base pairs 233-432 of the plasmid pBR322, several problems and puzzles intrinsic to the Polymerase Chain Reaction (PCR) technique were encountered. Considerable evidence was collected which suggests that the mechanism of amplification for some sequences, including this one, may differ from that of the standard model and indicates that the process may be more complex than previously thought for sequences with non-randomly distributed and high percentages of GC's.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I wish to express my gratitude to Professor J. Michael Schurr for allowing me the opportunity to learn under his tutelage. His attention to detail, tactfulness, and patience are virtues that will serve to guide me for many years to come. Thanks also go to Dr. Bryant Fujimoto and Dr. David Rangel for the stimulating discussions and many collaborations that have steered me around countless pitfalls. To the former group members too numerous to mention whose contributions were invaluable to the work presented herein and to two in particular; Dr. Rogelio Diaz, JD, and Mr. Christopher Sucato, thank you. Also, recognition goes to the Seattle chapter of the ARCS foundation for their generous financial assistance received through a named fellowship that was donated by the Washington Research Foundation (WRF) during my first 3 years of graduate school.

I would like to express appreciation to a few other people who were instrumental in my arrival upon this point. Most notably are two professors from Kansas State University where I completed my undergraduate degree. Thanks go to Dr. Donald W. Setser with whom I had the opportunity to work studying small molecule, gas phase chemical kinetics and to Dr. Pedro L. Muiño who served as my mentor and introduced me to the University of Washington, Department of Chemistry and specifically to Professor Schurr. I would also like to express my sincere appreciation to my brothers, Scott and Todd, whose respective paths have helped shape my journey. Finally, I would especially like to recognize my mother, Patricia, whose kind spirit and quiet resolve have played no small role in this endeavor.

DEDICATION

This work is dedicated to my mother, Patricia, and to my family.

Chapter 1
Resolution of a PCR Problem

1.1 Introduction

The technique of PCR is widely used for DNA sample amplification and preparation, and has been applied for diverse purposes in many laboratories worldwide. It constitutes an *in vitro* method to copy and prepare sufficient quantities of DNA from specific regions of any given DNA whose end sequences are known. When new PCR reactions are contemplated, initial attempts are made to optimize the reaction by stepwise adjustment of the chemical composition until only a single band of DNA is observed upon gel staining using a fluorescent dye that is specific for nucleic acid detection. DNA separation is typically achieved using standard gel electrophoresis techniques that discriminate between DNA molecules based on size (or charge) and shape. The gel acts as a molecular sieve through which the DNAs migrate. While DNAs of different lengths will typically migrate (in an average sense) with distinct mobilities during electrophoresis, Stellewagen has shown that certain sequences will migrate with anomalous rates [1]. She has found and even characterized two 147 bp restriction fragments from an Msp I digest of the plasmid pBR322 that migrate differently due to one of them (the one she refers to as 12A) having a bent conformation [2]. This experimental finding was an important motivator for this work, although it turns out be of little importance directly. The point is that one must use caution when interpreting results from gel-electrophoresis separated DNAs, especially during the optimization process, where reaction conditions are varied widely. The process of finding a set of optimal reaction conditions that favors amplification of a unique sequence, and that yields only a single band of DNA, can be a cumbersome task.

The steps involved in PCR are rather simple according to the standard model [3], but that simplicity may belie more complex actions that are actually taking place. Figure 1.1 is a pictorial representation of the standard view of how a PCR reaction proceeds when well behaved. The ability of the polymerase to exponentially amplify the desired sequence, while only linearly amplifying the parent template sequence (the plasmid in

this case) is what makes the technique so valuable in many applications. The amount of desired exponentially amplified DNA sequence should far exceed any other DNA present. However, because of this exponential aspect of the reactions, adventitious DNA's with sequences complementary to the primers may also be amplified. Depending on the subsequent application for the DNA, or because of difficulties inherent in the optimization process, for many reactions it proves simpler to cut the desired DNA band out of a gel to isolate it rather than to optimize the reaction.

PCR systems of the type employed here are inherently complicated systems. This is due to the fact that they are multi-variate. A mixture of chemicals is prepared by adding at least 9 different chemical components (i.e. DNA template, 2 primer oligonucleotides, 4 different dNTP bases, plus salts and polymerase enzyme) to a PCR tube. The complication only gets compounded when the temperatures and length of time for each PCR step are taken into consideration. While the temperature profile of the thermocycler employed here was not specifically tested, it is known that thermocyclers, in general, can have variant temperature profiles among models [4] and can even vary in response between lanes of a single thermocycler [5]. This fact can play a significant role during the optimization process, although sufficient progress can often be made so that single band DNA can be obtained, depending on the sequence. From a microscopic viewpoint, it is straightforward to see why the optimization of a single PCR reaction can take many weeks to months to achieve, if at all.

A large number of variables are needed to parameterize the PCR process. However, a complete study of this issue would address how each component could contribute to the overall amplification efficiency (which also needs to account for the change in concentration of the template as cycling proceeds) and product quality. Consequently, a large multi-dimensional grid of parameter space must be 'searched'. This is the optimization process. Early optimization strategies relied upon discrete, and incremental, experimental testing for successful navigation to an acceptable product [6]. However, this approach of changing only one variable at a time while holding the others constant (called the univariate approach) is known to be slow and ineffective and also includes a high risk of not being able to identify a set of optimal conditions for the

system. With the widespread availability of computers and software packages, current optimization strategies attempt to narrow the field of accessible parameter space before attempting even a single experiment. While this *may* save some time, another dilemma still remains, since some sequences don't amplify uniquely.

For some DNA template/primer combinations, it has been shown that the addition of certain chemical reagents into the reactions can produce clean, singular bands of DNA upon gel electrophoresis [7-17]. These PCR additives are typically used because they have some advantageous effect on the reaction that arises from fundamental interactions with one or more components of the PCR mixture. For example, solvents such as Dimethyl Sulfoxide (DMSO) dehydrate DNA by replacing water around the helix and forming hydrogen bonds with the DNA bases [7, 18]. In forming hydrogen bonds with DMSO, the attractive forces of the interstrand hydrogen bonds weaken between complementary dinucleotide bases and strand separation becomes more favorable. This effect manifests itself macroscopically as a lowering of the melting temperature of the DNA [10]. Thermodynamic properties, including DNA melting, play a significant role in the successful amplification of PCR reactions.

Chapter one will present a detailed description of the empirical observations, puzzles and subsequent experiments that eventually led to a resolution of the forthcoming PCR problem. This study represents a significant contribution to two larger but related fields of nucleic acid chemistry. The first is thermal denaturation and specifically its role in one PCR reaction. The second is polymerase activity in reactions of the PCR type. Evidence will be presented to show that the polymerase employed in these reactions functions like a helicase enzyme by acting to peel apart unmelted duplex DNA as it elongates from a primer. This activity is manifested most strikingly at potential pause sites that are associated with sequences of high G+C content.

1.2 Materials and Methods

The plasmid pBR322 (4363 bp) was grown and isolated from *Escherichia coli* HB101 host cells in our laboratory using a standard plasmid purification kit (QIAprep® Spin Miniprep Kit) from Qiagen, Inc. and was subsequently used as template DNA in our PCR reactions. The target sequence for amplification was the 200 base pair region from 233-432 and contains 63% G+C. Two primers, a *forward* primer (5' GCGCTATATGCGTTGATGCA 3') and a *backward* primer (5' GCAACCGCACCTGTGG 3'), were chosen to amplify this specific region of the plasmid and ordered, with HPLC purification, from IDT, Inc. Because the elongating action of the polymerase enzyme on each primer occurs in the 5' to 3' direction for the strand of DNA being made, primers can in principle be extended from either end of the target sequence, according to the standard model. It is noteworthy that for the plasmid template used in these PCR reactions, the origin of replication is known as well as the entire DNA sequence. Each primer is unique in the plasmid and occurs only once.

The terminology used extensively throughout this chapter will be defined directly, but where applicable, attempts will be made to comply with commonly used convention. The *forward* designation will be used to refer to the primer that elongates in the forward counting direction (233→432 in these reactions) relative to the numbering scheme that defines the template plasmid; it is called the forward primer. Similarly, the molecule that elongates in the *backward* counting direction (432→233 relative to the template plasmid) will be called the backward primer. The *forward* primer is the primer that anneals *to* the anti-sense strand of DNA and thus *creates* the sense strand in the process. The *backward* primer anneals *to* the sense strand of DNA and creates the anti-sense strand upon elongation. In the figures throughout chapter one, the forward primer is *green* and the backward primer is *blue*.

Although many variations of the PCR reaction were employed throughout the course of this study, most reaction mixtures reported here contained 20 mM Tris-HCl (pH 8.8), 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 60 mM KCl, 0.1 % Triton X-100, 0.2 mM dNTP's and 0.05 units/μl Accuzyme polymerase enzyme (with proofreading capabilities) from Bioline USA, Inc. Approximately 2-10 ng pBR322 template DNA were added and

the initial concentration of each primer was 0.4 μM. Thermal cycling was done with an MJ Research, Inc. PTC-100™ Programmable Thermal Controller using the following program, or variation thereof:

Step	Temp (°C)	Time (minutes:seconds)
1	96	2:00
2	96	0:20
3	60.5	0:30
4	72	0:10
5	Return to step 2 and iterate 35 times	
6	72	7:00
7	4	

All PCR reaction products were run without further cleanup on standard 12% mini-polyacrylamide gels prepared according to the protocol of Maniatis et al [19]. Gel reagents were ordered from Bio-Rad Laboratories, Inc. and diagnostic gels were run at 70V for 4 hours.

The melting temperature (T_m) of each primer was calculated using the OligoAnalyzer 3.0 T_m calculator, accessible on the Integrated DNA Technologies (IDT®) website [20]. The algorithm used in the calculations derives from a nearest-neighbor model used for predicting DNA stability and has 'correction' terms added to bring the T_m predicted by the algorithm into agreement with the T_m's measured empirically in monovalent salt solutions [21]. The equation is found to predict the T_m of Duplex oligomers from 69 mM up to 1.02 M salt more accurately than earlier equations due to the assumption that T_m is not a linear function of $\ln[Na^+]$. Instead, a quadratic term is added to the equation to match the empirical observation. The observation that T_m may not be linearly dependent on $\ln[Na^+]$ was initially suggested by Bond et al. [22] and more recently shown by Owczarzy et al. [23] to apply for each of 92 duplex DNAs investigated, ranging from 10 bp to 30 bp in length. Measurement of the relationship between T_m and $\ln[Na^+]$ revealed that each sample exhibited a reproducible curvature in

the same direction. Thus, linear fits to those plots aren't adequate, within the experimental errors of T_m, and a quadratic term is necessary to adequately describe the trend. The T_m of the forward primer (calculated at 0.4 μM concentration and 60 mM monovalent cations) is predicted to be 58.2 °C and the backward primer is 59.6 °C. An annealing temperature (T_{anneal}) of 60.5 °C was used in these reactions. Annealing temperatures from 50-65 °C were also tried, with no significant improvements to the PCR reaction products observed.

1.3 Results

1.3.1 The Main Problem

After very many attempts to optimize the above PCR system, the best product that could be obtained exhibited two bands upon gel electrophoresis (Figure 1.2). The back band migrates with the same mobility as the 200 base pair band of the standard 100 bp ladder from New England Biolabs, Inc. and is the more prevalent species by about a 4:1 ratio (for the majority of reactions) over the front band which migrates with an ~9% faster gel mobility than the back band. Additionally, for gels of greater resolution, the front band is itself two bands. It is speculated that the two bands comprising the front band are a *hetero-* and *homo-* species of DNA produced by a unique mispriming event in the early steps of the PCR reactions and therefore amplified throughout. By *hetero-*, it is meant two single strands of complementary DNA (for regions of overlap) that differ in length and therefore migrate as a duplex of DNA plus an overhanging tail. The term *homo-* refers to two, fully complementary single strands of DNA identical in length that therefore migrates as a duplex DNA.

Upon first inspection, the unresolved front band was presumed to be an extra band attributable to a non-specific binding event involving one of the primers that was caught and elongated by the enzyme off of the template in one of the early cycle steps of the reaction. To confirm this supposition, and to rule out any undetected contaminant, control reactions were performed during the optimization process. The strategy for eliminating extra bands is to stepwise run separate reactions, changing one component at a time while holding all of the other components constant. This directly tests the effect

that each component has on the reaction specificity, as determined by gel electrophoresis. The process continues until an appropriate set of reaction conditions results in a single band of DNA that migrates on a gel relative to a standard DNA of known length.

For these reactions, multiple tests were attempted in order to locate a combination of variables that resulted in a singular band 200 bp in length. After an extensive effort to get rid of the non-specific bands of DNA through optimization, without significant success, a major effort was undertaken to characterize the two bands produced in our PCR reactions. The DNA of both bands was isolated by gel excision/extraction according to the standard procedures of Maniatis et al. [19]. In characterizing the composition of the two bands, our aim was to diagnose and thus eliminate the cause of the apparent mispriming event in our reactions. Due to other work done in this laboratory (which won't be discussed in this thesis), the specific 200 bp region of pBR322 chosen for amplification was the only product that would be useful. Changing sequences or even the length of this sequence was not an acceptable option.

The results of these characterization experiments have produced some valuable insights into the microscopic workings of the polymerase chain reaction on a macro-molecular level. More importantly, these results have raised numerous questions in regard to the processes taking place, and have shown for at least one sequence in particular that the standard model of PCR almost certainly does *not* apply.

1.3.2 A Puzzle in the Initial Back Band Sequencing Data

In one singular set of sequencing data from the isolated back band DNA, the forward primer was found to have been inserted into the *backward* primer site. During the PCR steps *in the sequencing process*, the primer (which is part of the sense strand) annealed itself *to* the sense strand template by the 3 bases at its 3' end (Figure 1.3) and thus produced a molecule 217 base pairs in length (Figure 1.4b). This singular result, which was *not* observed in subsequent sequencing of the back band, makes no sense at first, but illuminates the kind of unexpected processes that can occur in the PCR. Subsequent sequencing runs have confirmed that the sequence of the back band, the band

that co-migrates with the 200 bp band of the standard ladder, is precisely the 200 bp sequence the amplification reactions were designed for.

The intensity data from sequencing reactions contain additional information about the PCR process. In the sequencing data of Figure 1.4a, which is the original sequence from the combination of backward primer plus isolated template, the peaks appear uniform and clean, and presumably reflect an unhindered extension process, as is expected in PCR reactions by the standard model. This sequence corresponds to the 200 bp antisense strand of pBR322 from 432-233 (the primer sequence is not read as part of the sequence data). In contrast, the sequence data from the forward primer plus isolated template, shown in Figure 1.4b look very different. Not only are the intensity heights non-uniform, but in the sense strand sequence the complement to the forward primer has been inserted (in proper orientation) at the right end of the molecule (the primer sequence is supposed to be at the left end of the molecule only and shouldn't appear in the sequencing data). This is the 217 bp molecule noted above. One can read the majority peaks in the forward primer sequence data as the sense strand of pBR322 from 233-432, with the addition of the complement to the forward primer at the 3' end of the sequence. This latter datum indicates that a microscopic fluctuation involving the forward primer has been captured by the polymerase and shows that the forward primer has been caught doing what it is not expected to do. Subsequent sequencing of the back band did not display this feature, which must have arisen from some kind of singular event.

If one inspects the sequencing data of Figure 1.4b more closely, the sequence from the reaction of the forward primer plus 200 bp template gives two sequences. As stated above, the majority peaks report the sense strand sequence corresponding to 233-432 of pBR322 (the target sequence) plus 17 bp of the complement to the forward primer at its right end. The underlying minority peaks that contribute to coinciding peak signals, however, also correlate to a sequence of DNA and can be read as the antisense strand of DNA, also of length 217 bp (only 1 kind of primer is present in sequencing reactions) that is evidently copied and amplified concurrently. Each strand of this 217 bp duplex contains the complement to the forward primer at its 3' end. The sequencing reactions are also of the PCR variety and from simple intensity comparisons of the data we infer

that the DNA is being copied from left to right (233-432, the sense strand) more efficiently than from right to left (432-233, the antisense strand). The kinetics of the reaction are thus slower for the forward primer amplifying from the sight that isn't its full complement which creates the anti-sense strand of DNA (right to left, 432 – 233) than for the forward primer amplifying from its intended sight, where it binds to its full sequence complement and creates the sense strand of DNA (left to right, 233 – 432). Evidently, the polymerase enzyme has to peel apart the sense/anti-sense template strands of DNA ahead of it as it elongates the sense strand of DNA behind it. It will be shown forthcoming that this is because this particular GC-rich sequence isn't expected to melt completely apart, even at the high temperatures used for melting in these sequencing reactions (96 °C). This is one possible explanation that is consistent with the observed difference in the kinetics between the two directions of elongation (left-to-right vs. right-to-left). Once the forward primer has been inserted into the right end of the molecule the resulting template that will be amplified from has the forward primer present at both ends of the molecule. The sequence data of Figure 1.4a is distinguishable and can be read as a single sequence of DNA. Because the polymerase has to act with some helicase activity to peel apart the DNA as it elongates and the sequence that is complementary to the backward primer remains largely unmelted, the backward primer will not bind to its intended sight until the DNA is peeled apart. Once the DNA is peeled apart, however, the backward primer can bind to its sight and is elongated in an unhindered manner. This gives rise to the clean sequence data observed in Figure 1.4a. While this result may be unexpected, it is not counter-intuitive, and in fact demonstrates that the primer, DNA and polymerase combination is sensitive to its flanking DNA sequence. It is emphasized again that additional subsequent submissions of the back band for sequencing have shown its length to be 200 bp and its sequence to be identical to 233-432.

1.3.3 The Sequence of the Front Band Reveals a Mispriming Error

Sequencing technology was also used to investigate the origin of the front band. After a few ambiguous attempts at sequencing the isolated front band, it was successfully inserted into a sequencing plasmid using standard cloning technology (from Invitrogen

Life Technologies Corporation). The front band is 179 base pairs and is identical in sequence to the plasmid pBR322 from 233-393 (160 bases) with the forward primer inserted in reverse orientation extending the last 19 base pairs (179 bases total, see Figure 1.5, sequence data not shown). At the site where the forward primer inserts itself into the sequence there is a 5 base pair run that is complementary to the forward primer, starting one base removed from the 3' end of the primer. The forward primer has been caught again in the act of mispriming.

DNA polymerases can extend from mismatched terminal base pairs. In the sequence data of the 179 bp front band, the terminal base at the 3' end of the incorrectly inserted forward primer is missing. This indicates that the polymerase was able to bind and extend a mismatched DNA pair in the PCR reactions. Goodman et al. [24] have shown that the DNA polymerase *Thermus Aquaticus* (Taq), with no observable exonuclease activity, can extend from an otherwise complementary primer template pair that simply contains a mismatched bp at the 3' terminal end of the primer, albeit with less efficiency. They measured the extension efficiency for the 16 possible combinations of matched and mismatched base pairs at the template-primer 3' termini. They found that the transition mispairs were extended 10^3-10^4 less efficiently than their correctly paired counterparts and also that extension efficiencies were independent of temperature. The base pair at the 3' terminal end of the forward primer is an Adenine and the mismatched pair in the reactions would correspond to an A(primer)·C(template). The Accuzyme DNA polymerase used here has 3' to 5' proof reading activities which prevents misincorporations during polymerase activity. The very significant problem unique to the PCR technique is that once a primer gets misincorporated into a sequence, it has primer sites at each end that can amplify concurrently. This activity doesn't happen strictly in the late cycles of PCR but can occur at any point during the reactions. Once the misprimed sequence embeds itself in the reaction, it will become amplified.

1.3.4 The Forward Primer Plus Back Band DNA Yields Amplified Front Band

The polymerase chain reaction can occur when only forward primer, and no backward primer is added to *pure backband template* plus reaction mix. PCR reaction mixtures were prepared that included only the isolated 200 bp backband DNA and excluded one or the other of the primers from the reaction mix. In those reactions where only backward primer was added (Figure 1.6, lanes 7-10), the reaction behaved as expected and gave no band of DNA. However, in those reactions involving only the forward primer (with no backward primer present), a band corresponding to the front band was observed (Figure 1.6, lanes 2-6 and 11-14). These data confirm that the forward primer is able to bind and is amplified from a secondary spot that isn't its full, complementary sequence. Somehow, the conditions are being met for the amplification process to occur.

1.3.5 Is the Template Fully Melted?

Why is it favorable for the forward primer to bind at a sequence that isn't its full complement (as appears to be the case)? In PCR it is commonly assumed that the duplex DNA 'completely' melts apart during the initial step of cycling (Figure 1.1, step 1). This assumption may suffice for low-melting DNAs, that is DNA sequences of low GC content, but can the same behavior be expected to occur for DNA of higher GC content, which melts at a much higher temperature? The entire region of pBR322 from 233-432 is 63% GC and includes some 20 bp regions that are as high as 80% (simple inspection). The G+C distribution is non-uniform and it will be shown that the template is likely unable to completely melt apart in certain regions of the sequence, even at 96 °C (T_m in PCR reactions). Benight et al. [25] showed that the ends of a 301 bp fragment of DNA were the first regions to melt apart (or fray) at temperatures above 80 °C but that some interior regions were not melted apart at 90 °C. The scenario where the duplex ends fray open while other regions, even one very near to the end remain closed is consistent with the mispriming event captured in the sequence data.

To investigate how the base pair composition might affect the melting behavior of DNA, a 'melting' curve was calculated for the target sequence of pBR322. The model employed here calculates the 'melting' curve (which is actually a melting probability) using the nearest-neighbor statistical thermodynamic theory of helix-coil transitions in DNA following the procedure of Benight et al. [26]. In the nearest neighbor model, the T_m of each base is assumed to depend not only on the base pair type (i.e. an A·T or G·C) but also contains thermodynamic contributions from the neighbors on each side. The model dictates that each nucleotide base is either hydrogen bonded to its complement on the opposite strand and stacked with its neighbors (state "1"), or is non-hydrogen bonded and unstacked (state "0"). The characterization of the DNA molecule in a two-state manner is similar to the characterization employed in the well-known Ising model of ferromagnetism [27].

The reference state is taken to be the all-helix state and a stability parameter, s_i, which has the form:

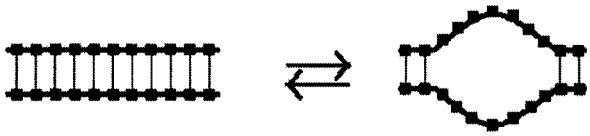
$$s_i = \exp\left(\frac{\Delta S}{RT}(T - T_i)\right) \tag{1.1}$$

is assigned to every base pair in the DNA. A stability parameter of 1.0 is assigned to each intact base pair and $1/s_i$ to each broken base pair. In equation (1.1)

$$T_i = \begin{Bmatrix} T_{AT} \\ T_{GC} \end{Bmatrix} + \frac{\left\{\delta G_{i-1,i} + \delta G_{i,i+1}\right\}}{2\Delta S} \tag{1.2}$$

where T_{AT} and T_{GC} are the average melting temperatures of A·T or G·C base pairs. Frank-Kamenetski [28] found empirically that, in 0.2 M NaCl, T_{AT}=69.6 °C and T_{GC}=110.5 °C. The entropic term, ΔS=-24.8 entropy units (1 e.u. = 1 cal/K·mol) is assumed to be independent of temperature. According to Delcourt and Blake [29], the assumption of constant entropy for the different neighbor pairs derives from the results of calorimetric studies that have been conducted over a wide range of temperatures and base compositions (and also counterion concentrations). Over 30 independently measured

values for the calorimetric enthalpy, ΔH_{cal}, and the corresponding T_m for the samples tested gave a nearly constant value for the transition entropy of 24.85 (±1.74) cal/mol·bp from the relation $\Delta S=\Delta H_{cal}/T_M$. In equation (1.2) the $\delta G_{i,i+1}$ terms are the deviations from the average base pair stacking free energy. There are 10 possible nearest neighbor states due to directional degeneracies and so $\delta G_{i,i+1}$ can have 10 different values. The δG values reported by Doktycz et al [30] and employed in these calculations are given in Table 1.2. The following equilibrium reaction was used to represent the melting of a duplex DNA



(1.3)

The relative probability for a run of **m** base pairs existing in the melted (unbound) state is then calculated to be

$$P^{rel}_{j\rightarrow j+m+1} = (s_j s_{j+1} s_{j+2} \cdots s_{j+m+1})^{-1} \exp\left(\frac{\delta G_{j-1,j} + \delta G_{j+m-1,j+m}}{2RT}\right) \quad (1.4)$$

The relative probability calculated in equation (1.4) basically reports the equilibrium constant for the chemical reaction represented in equation (1.3) and tabulates the values so the ratio of [unmelted]/[melted] is monitored. In this way, P^{rel} reports the likelihood that each run of **m** bases will exist in the entirely bound (or unmelted) states. The sliding window scheme is employed and m=29 bases was used in the following analyses. In the sliding window calculation, a set of boundaries are first placed on the sequence that act to enclose the region of the sequence that will contribute to P_{rel}. The size of the window is actually m+2 to account for the nearest neighbor contributions of the two bases on either end. After P^{rel} is calculated for the window j → j+m+1, the window is iterated by 1 bp down the sequence and the calculation repeated. This is continued for the entire sequence that falls entirely within 232-433 of pBR322. The results are shown in Figure 1.7. Not every window will 'melt' completely apart with the

same probability. The left side of the molecule, which has a higher %A·T composition, is predicted to melt with ~10^9-10^{10} more likelihood than the right side of the molecule. In Figure 1.7, the lower left end (bp < ~30) of the molecule corresponds to the sequence location of the forward primer and the higher right side (bp > ~170) is the location of the backward primer.

The right side of the target molecule, from base pairs 151-200, is 72% G+C. Our hypothesis is that the two strands of duplex DNA are not *completely* melted apart, even at 96°C, but instead the right-end of the sequence remains duplex while the left-end frays open. The plasmid pBR322 is known to contain regions of non-randomly distributed G·C and A·T pairs that affect its thermal melting properties. Falzon et al. [31] showed, using thermal denaturation studies, that the 214 bp region of pBR322 from 4342-192 (pBR322 is circular) melted with a biphasic melting profile due to its non-uniform distribution of G·C's and A·T's in the sequence. The first region to melt was the A·T rich region (bp 4342-80) followed by the higher T_m G·C rich region (bp 80-192). Using the analysis outlined above, a melting curve was generated for the 214 bp region of pBR322 (unpublished calculation) and it confirmed that the left side of this sequence (bp 4342-80) melted with a much higher probability. It is reasonable to think that the region studied here (233-432) could also melt with a biphasic profile, since it also has G·C distributions that are non-uniformly distributed and a characteristic melting profile shown in Figure 1.7. Wartell and Benight also showed through measurement of UV absorbance derivative profiles at 268 nm, that the 192 bp DNA fragment from base pairs 4342 − 173 of pBR322 (prepared by digestion with Hae III restriction enzyme), melted with a biphasic profile [32]. This 192 bp fragment is a section of the 217 bp fragment studied by Falzon et al.

Returning to Figure 1.7, not all sections of the template are predicted to melt to the same extent. There are three distinct regions that have a high probability for being unmelted at 96 °C. This incomplete melting creates regions of the DNA that then act as barriers to the PCR. The DNA from the front band is 179 bp with the forward primer present at both ends of the molecule (with opposite orientations). The mispriming event occurs because of a 5 base sequence overlap from 161-165 (393-397 in the plasmid) on the template. Under ideal circumstances, where the DNA melts completely apart to give

two single strand templates, this binding of the forward primer by 5 bp would be thermodynamically unfavorable compared to either the more stable binding (by 20 bp) of this same primer to its full sequence complement or the binding of the 20 bp backward primer to its full sequence complement. As the melting profile of Figure 1.7 emphasizes, ideal circumstances *do not* apply for this sequence. The bases from 161-165 lie in a region that has a high probability for being melted (or single stranded) but that is also wedged in between two regions that are likely unmelted (double stranded). Thus, the ability of the primer to bind by only a 5 base overlap experiences a dramatic increase in relative favorability. For this sequence, furthermore, the complementary sequence to the backward primer (185-200) is simultaneously blocked (either fully or partially). If, during the initial phase of the elongation step, the backward primer is blocked while the forward primer binds in reverse orientation to the 161-165 site, albeit with lower probability than for its (1-20) binding site, then the backward primer should not be able to compete with the forward primer in PCR thermocycling.

1.3.6 Competition Studies

As the concentration of backward primer is titrated to a level 5 times greater than that of the forward primer in the amplification of purified, 200 bp backband, no change in the relative intensity of the front and back bands is observed (Figure 1.8). That is, as the concentration of backward primer is increased relative to the forward primer in the PCR reactions, there is no visible change in the ratio of the bands (from left to right). This indicates that the backward primer at its 185-200 site does not compete effectively with the forward primer at the (161-165) binding site. This supports the notion that the right side of the molecule is not melting completely apart. Regarding Figure 1.8, it is unknown why the ratio of back band to front band appears to be closer to 1:1, instead of 4:1 for this set of experiments. The crucial observation is that there is no visible change in the initial ratio as one reads left to right, corresponding to increasing amounts of backward primer.

1.3.7 Effect of Ethylene Glycol

A different result is observed for the PCR reactions, when ethylene glycol is added (Figure 1.9). Ethylene glycol has been shown to lower the melting temperature of duplex DNA [20]. By including it in our reactions, the double strand DNA serving as the template melts apart more completely. A large increase in the ratio of back band to front band is seen in these reactions, indicating that the back band has become the strongly favored product. If the backband were diluted to normal intensity, the front band would likely be almost invisible. Presumably increasing the concentration of backward primer would further enhance the back-band/front-band ratio.

All of the additives listed in Table 1.1 have been added to PCR reactions in various other laboratories, because of their ability to lower T_m of the duplex DNA. Most of the successful reagents, however, have been found to work empirically on a case by case basis and so do not necessarily work for all reactions. We found that none of these additives completely eliminated the front band. With EG as the additive, the appropriate combination of variables was not found to *completely* eliminate the front band, but a sufficiently great ratio of backward to forward primer may push the back band/ front band ratio to a satisfactorily large value.

1.3.8 Strand Separation in Incompletely Melted DNA

If the strands are not completely separated by melting, can they be separated by the polymerization process itself? In a separate set of sequencing data for the 200 bp back band another discrepancy between the backward and forward primed sequences emerged, which may provide a clue to the action of the elongating enzyme. Both sequences reported the appropriate 200 bp region from 233-432. However, if one looks at the intensity profiles of the backward and forward sequencing reactions, very different trends are observed. In Figure 1.10, the peak heights of the sequencing reactions are plotted versus base pair position with the terminal 5' base of the primer designated number 1. These observations were inferred by visual inspection and are approximate averages of the data intended only to emphasize a fundamental qualitative difference between the two sequencing reactions. A more formal analysis will be described shortly.

The sequence from the backward primer seems to show a smooth, roughly exponentially decaying intensity, consistent with what one might anticipate for termination reactions of this type. This suggests that extension of the backward primer occurs in an unhindered manner, when it takes place, and that the backward primer probably binds to a completely free single-strand template, when it does bind. The behavior of the forward primer is quite different. Its intensity data appear to exhibit an initial dip, then fluctuate around, and finally level off for the rest of its extension process.

To more formally investigate the sequence data obtained with the forward primer in Figure 1.10, fourier analysis was employed and combined with a χ^2 minimization protocol (see Appendix A). The resulting coefficients, when inserted back into the fourier function, generate a best fit line through the data. The sequence data (peak heights), along with the best fit line through the data are shown in Figure 1.12b. In the figure, the fitting protocol (with 10 components included) exposes three very distinct dips in the sequence data. We speculate that in the case of the forward primer, the enzyme must separate the unmelted duplex strands that act as barriers as it extends in the rightward direction. Hence it moves more slowly along the template. Because it moves more slowly, its exonuclease site can, with greater probability, cut out the dideoxynucleotides that it interprets as an error and incorporate the appropriate deoxynucleotide into the strand before continuing. Such behavior would be consistent with the trends in Figure 1.10. Figure 1.11 presents a schematic illustration of the polymerase extending the forward primer and liberating the initial sense strand, which can then bind the backward primer. The backward primer, in turn, can be extended in an unhindered manner.

There is a correlation between the higher G+C regions of the DNA template that are predicted to remain unmelted, and the dips in the sequence data that coincide with barriers to the elongation process. Both plots are shown in Figure 1.12 to show, explicitly, this correlation.

1.3.9 The Front Band consists of Two Bands

Polyacrylamide gels a typical set of reaction products that were run under conditions of higher resolution (i.e. longer times) revealed that the front band is made up of two bands of DNA as well. It is speculated that these two front bands are a homo-species (referred to herein as the *front*-front band) that likely arises from two fully complementary DNAs 179 bases in length and a hetero- species (referred to as *back*-front band) that consists of two strands of differing lengths (one strand 179 bases and the other 200 bases in length) that migrates as a duplex DNA plus an overhanging tail. Figure 1.13 shows a gel containing the 3 bands (200 bp back band plus the *back*-front band and *front*-front bands) observed upon viewing the reaction products subsequent to staining after electrophoresis. The experiment shown in Figure 1.13 compares three amplification reactions that contained different amounts of polymerase enzyme in the starting mix. For those reactions that had low concentration of enzyme (i.e. the 1x and 2x reactions) both front bands were present. However, comparing lanes 3 and 4, where the concentration of enzyme doubles from 2x to 4x, differential amplification occurs and favors the 200 bp band over either of the front bands. Again, it is this band that contains the products that the reactions were designed to amplify. Thus it can be inferred that the reaction amplifies more efficiently in the presence of excess enzyme. The origin of this behavior is not known, however, it is clear that the anti-sense strand of DNA (which arises from amplification of the backward primer that has bound to its full complement) has been amplified more efficiently in these reactions. This observation is supported by the dramatic increase in the intensity of the back band of DNA that is 200 bp in length. Furthermore, the homo- species (*front*-front band) is absent in the reaction product of lane 4 (compared to those of lanes 2 and 3). The hetero- species is present which indicates that under the conditions, wherein the enzyme is present in excess, the backward primer can bind to its complementary sequence and is extended more readily. The result is an increase in the production of the full anti-sense strand of DNA that binds to the full sense strand and produces a DNA product of length 200 bp. The hetero-species remains because the forward primer is still actively misprimed from its secondary site, albeit with but does so with less efficiency compared to the back band. The single

stranded DNA produced from this unique mispriming site can still bind to the 200 base sense strand that is present in excess subsequent to amplification. The result is a DNA molecule containing a single strand 179 bases long bound to a 200 base single strand (this is the hetero- species) that migrates as a duplex DNA with an overhanging tail. Regardless, the reactions the primers were designed to amplify occur with greater efficiency in the presence of excess enzyme.

1.3.10 Altering the 3'-End of the Sequence Eliminates the Front Band

PCR from a very similar region of the template p30δ can be optimized sufficiently so no front band appears upon gel electrophoresis. The total length of p30δ is 4932 base pairs and the 200 bp region from 233-432 contains 52% G+C, compared to 63% for the plasmid pBR322. Because both the pBR322 and p30δ sequences are identical from bp 233-375 (Figure 1.14), the same forward primer was used in both reactions. The backward primer was adjusted accordingly because of its unique sequence at the right end. No information has been obtained to infer by what model this PCR operates but it is anticipated to more closely resemble that laid out in Figure 1.1 by Mullis et al. (1986) because of its lower GC content. p30δ is a derivative of pBR322 constructed by inserting an 842 bp sequence containing the Bgl II to Sal I fragment from *S. cerevisae* into the region between the BamH I (bp 375) and Sal I (bp 651) sites of pBR322 [33]. This plasmid exhibits less evidence of long-lived metastable secondary structures, and irreproducible behavior than does the parent pBR322 [34, 35]. Figure 1.15 shows the computed melting profile for plasmid p30δ co-plotted with that of pBR322, both from bp 233-432. The plasmid pBR322 has more difficulty melting open to reveal the complementary sequence of the backward primer than the plasmid p30δ. Figure 1.16 shows the results of a PCR optimization, by a similar method, comparing the reactions with p30δ as the template versus those with pBR322. The reactions with the plasmid p30δ as template have been optimized sufficiently to result in a single band of DNA 200 base pairs in length. In the p30δ system, since both ends of the template DNA are predicted to melt completely apart, the primers can bind at their respective sites. A polymerase molecule is then able to elongate from each primer, possibly unwinding the

DNA with some helicase activity as it translates along the template in a manner similar to the pBR322 template. Regardless of how it proceeds, though, it appears that the unique mispriming event can be eliminated, at least in the early cycles of the amplification process.

1.4 Discussion

PCR has become an increasingly important application for laboratories worldwide but is still a technique in development, and many fundamental questions remain. The process is much more complicated than originally anticipated (as many labs are well aware), at least for some sequences of high GC content. McDowell et al. [36] presented data correlating high T_m domains, as predicted from a simple thermal stability profile, with permanent termination sites that prevent the amplification of a DNA mimic, or internal standard in competitive PCR reactions. Primer extension by the enzyme Taq polymerase was inhibited in sequences with localized domains that met the criteria identified by the authors, which was mainly that the T_m be greater than 81 °C. If the T_m of the fragment, specifically the temperature of 50% of the molecules are melted lay entirely below 81 °C, the mimic amplified along with the desired PCR product independent of %G+C and length for those mimics tested. Three mimics were used that were each 304 bp in length but had different predicted thermal stability profiles. Those reactions containing a mimic with 49% G+C that 'spiked' three separate times above 81 °C resulted in *no* band of DNA at 304 bp. Reactions containing a mimic with 53% G+C that 'spiked' two times also resulted in *no* band of DNA. Interestingly, those reactions containing a mimic with 53% G+C that showed no spikes in its thermal stability profile did produce a band at 304 bp in length. There were no termination sites and DNA elongation by the polymerase was unhindered in these reactions.

Pause sites associated with T7 DNA polymerase tend to occur immediately following the incorporation of a deoxyguanosine that is embedded in a G+C rich region and whose general local sequence is of the form *pyrimidine-G-C* [37]. There are 14 such sites of that form in the sense strand of pBR322 from 253-432. The pausing was eliminated by the addition of 2 M betaine into the sequencing reactions. The highest

concentration of betaine added to our PCR reactions, however, was 1.41 M (Table 1.1). Chou found that addition of *E. coli* single-stranded binding protein (EcoSSB) produced differential amplification resulting in a single band the same length as the target. His hypothesis was that the protein bound to and melted the PCR template secondary structure during the PCR reaction, thus allowing elongation to occur unhindered [38].

The inherent complexity in PCR previously discussed is compounded, yet again, by the fact that the temperature cycling in the reactions also serves to provide thermal energy to the 'system'. This allows the 'system' to access higher energy states. However, much still remains unknown about the interaction (binding, extending and fidelity) properties of the polymerase with DNA template. Shibata and Schurr have found that the same secondary bonds responsible for stabilizing the native structure at low temperatures can also promote aggregation in the thermal denaturation region for sufficiently long chains of a biopolymer [39]. While this may not be a problem for melting of the 200 bp template (it is too short), the plasmid pBR322 whose sequence is initially amplified is long enough to experience aggregation problems in the thermal denaturation region. Again, if there is a mispriming event in one of the early cycle steps, it will be amplified throughout. It is also documented that DNA secondary structure in solution can undergo structural changes in response to various perturbations, including rising temperature in the pre-melting region, where native DNA is known to undergo a transition involving B and one or more other right handed states [40]. Each perturbation can play a significant role in the polymerization process and thus affect the final results observed upon gel electrophoresis.

1.5 Conclusions

For PCR reactions involving the sequence from 233 – 432 of pBR322, two bands are consistently observed, one with 200 bp and one with 179 bp in a ratio of approximately 4:1. Numerous variables of the reaction were altered and many different additives were tried to affect this ratio, but both bands persisted under all conditions. The template pBR322 contains 63% G+C with some 20 base pair regions being as high as 80% G+C. The % G+C is non-uniformly distributed and likely has a biphasic melting

profile. It is hypothesized that the 200 bp fragment of DNA isn't melting open on the right side of the molecule, even at 96°C. This is supported in those experiments, wherein the concentration of backward primer is varied over a wide range and gives no indication of competing with the forward primer at the 161 – 165 site in the reactions without ethylene glycol. The forward primer binds in reverse orientation ~160 base pairs into the desired sequence and creates a molecule with a total length of 179 base pairs with the forward primer inserted on both ends. This is consistent with the observed migration of the front band.

In one singular sequencing event, the forward primer was inserted into the right side of the molecule with the proper orientation to give a sequence with a total length of 217 base pairs. The sequencing data indicate that the identical single-strand primer sites at opposite ends of the duplex must be copied at two different rates. For this 217 bp DNA, the polymerase thus extends more efficiently going from left to right than from right to left.

The data for the 200 bp DNA suggest that the backward primer binds to a free single-strand template, which was not melted off, but instead peeled off by extension of the forward primer. In this regard, the polymerase is functioning like a helicase enzyme to separate the unmelted single strands as it extends the forward primer. The sequence data from the backward primer show monotonically and roughly exponentially decaying behavior consistent with the fact that it binds to a free single strand template and can extend at a 'normal' rate. The sequence data from the forward primer, however, do not show the same behavior, but instead exhibit three regions of low intensity in the sequencing data. These regions, associated with barriers created by regions of unmelted DNA, serve as pause sights for the stalled polymerase. In the process of stalling, the exonuclease site has time to excise the dideoxynucleotide, which it interprets as an error, and incorporate the correct base into that position. It is currently thought that the enzyme has to peel off the unmelted regions of DNA which means that it would have helicase capabilities associated with its elongating action. This is a new property of polymerase enzymes and certainly means that the PCR process is more complex than the standard model.

In those reactions that contained 17.5% ethylene glycol the intensity of the back band was much greater than the front band and the ratio of back band to front band intensities increased ~60-fold as the concentration of backward primer was increased in the reaction mix (to arbitrarily high levels). This is due to the ethylene glycol lowering the melting temperature of the duplex DNA which exposes the backward primer binding site and thus allows the backward primer to bind and elongate from its full complement much more readily than in those reactions where the backward primer has to wait for the polymerase to peel the two strands apart before it can bind to its full complement. Hence, more strands are created by the backward primer and the result is that the amplification reaction favors production of the 200 bp DNA (i.e. the DNA that the reactions were designed for).

1.6 Notes to Chapter 1

1. Stellwagen, N.C., *Anomalous electrophoresis of deoxyribonucleic acid restriction fragments on polyacrylamide gels.* Biochemistry, 1983. **22**(26): p. 6186-93.

2. Stellwagen, N.C., *Conformational isomers of curved DNA molecules can be observed by polyacrylamide gel electrophoresis.* Electrophoresis, 2000. **21**(12): p. 2327-34.

3. Mullis, K., et al., *Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction.* Cold Spring Harb Symp Quant Biol, 1986. **51 Pt 1**: p. 263-73.

4. Hoelzel, R., *The trouble with 'PCR' machines.* Trends Genet, 1990. **6**(8): p. 237-8.

5. Linz, U., *Thermocycler temperature variation invalidates PCR results.* Biotechniques, 1990. **9**(3): p. 286, 288, 290-3.

6. Roux, K.H., *Optimization and troubleshooting in PCR.* PCR Methods Appl, 1995. **4**(5): p. S185-94.

7. Varadaraj, K. and D.M. Skinner, *Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases.* Gene, 1994. **140**(1): p. 1-5.

8. Chevet, E., G. Lemaitre, and M.D. Katinka, *Low concentrations of tetramethylammonium chloride increase yield and specificity of PCR.* Nucleic Acids Res, 1995. **23**(16): p. 3343-4.

9. Frackman, S., Kobs, Gary, Simpson, Dan, and Storts, Doug, *Betaine and DMSO: Enhancing Agents for PCR.* Promega notes, 1998. **65**: p. 27.

10. Cheng, S., et al., *Effective amplification of long targets from cloned inserts and human genomic DNA.* Proc Natl Acad Sci U S A, 1994. **91**(12): p. 5695-9.

11. Chakrabarti, R. and C.E. Schutt, *The enhancement of PCR amplification by low molecular-weight sulfones.* Gene, 2001. **274**(1-2): p. 293-8.

12. Kovarova, M. and P. Draber, *New specificity and yield enhancer of polymerase chain reactions.* Nucleic Acids Res, 2000. **28**(13): p. E70.

13. Hung, T., K. Mak, and K. Fong, *A specificity enhancer for polymerase chain reaction.* Nucleic Acids Res, 1990. **18**(16): p. 4953.

14. Winship, P.R., *An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide.* Nucleic Acids Res, 1989. **17**(3): p. 1266.

15. Bachmann, B., W. Luke, and G. Hunsmann, *Improvement of PCR amplified DNA sequencing with the aid of detergents.* Nucleic Acids Res, 1990. **18**(5): p. 1309.

16. Chakrabarti, R. and C.E. Schutt, *The enhancement of PCR amplification by low molecular weight amides.* Nucleic Acids Res, 2001. **29**(11): p. 2377-81.

17. Comey, C.T., J.M. Jung, and B. Budowle, *Use of formamide to improve amplification of HLA DQ alpha sequences.* Biotechniques, 1991. **10**(1): p. 60-1.

18. Lee, C.H., H. Mizusawa, and T. Kakefuda, *Unwinding of double-stranded DNA helix by dehydration.* Proc Natl Acad Sci U S A, 1981. **78**(5): p. 2838-42.

19. Maniatis, T., Fritsch, EF, and Sambrook, J., *Molecular Cloning: A Laboratory Manual.* 1982, New York: Cold Spring Harbor laboratory Press.

20. Spink, C.H. and J.B. Chaires, *Effects of hydration, ion release, and excluded volume on the melting of triplex and duplex DNA.* Biochemistry, 1999. **38**(1): p. 496-508.

21. Owczarzy, R.a.B., Mark. *Calculation of Tm for Oligonucleotide Duplexes.* 2005 [cited; Available from: http://www.idtdna.com/support/technical/TechnicalBulletinPDF/Calculation_of_Tm_for_Oligonucleotide_Duplexes.pdf.

22. Bond, J.P., C.F. Anderson, and M.T. Record, Jr., *Conformational transitions of duplex and triplex nucleic acid helices: thermodynamic analysis of effects of salt concentration on stability using preferential interaction coefficients.* Biophys J, 1994. **67**(2): p. 825-36.

23. Owczarzy, R., et al., *Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures.* Biochemistry, 2004. **43**(12): p. 3537-54.

24. Huang, M.M., N. Arnheim, and M.F. Goodman, *Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR.* Nucleic Acids Res, 1992. **20**(17): p. 4567-73.

25. Benight, A.S., R.M. Wartell, and D.K. Howell, *Theory agrees with experimental thermal denaturation of short DNA restriction fragments.* Nature, 1981. **289**(5794): p. 203-5.

26. Benight, A.S., Schurr, J.M., Flynn, P.F., Reid, B.R., and Wemmer, D.E., *Melting of a Self-complementary DNA Minicircle.* Journal of Molecular Biology, 1988. **200**: p. 377-399.

27. Newell, G.a.M., EW, *On the Theory of the Ising Model of Ferromagnetism.* Reviews of Modern Physics, 1953. **25**(2): p. 353-389.

28. Frank-Kamenetskii, M.D., *Simplification of the Empirical Relationship between Melting Temperature of DNA, Its GC Content and Concentration of Sodium Ions in Solution.* Biopolymers, 1971. **10**: p. 2623-2624.

29. Delcourt, S.G. and R.D. Blake, *Stacking energies in DNA.* J Biol Chem, 1991. **266**(23): p. 15160-9.

30. Doktycz, M.J., et al., *Studies of DNA dumbbells. I. Melting curves of 17 DNA dumbbells with different duplex stem sequences linked by T4 endloops: evaluation of the nearest-neighbor stacking interactions in DNA.* Biopolymers, 1992. **32**(7): p. 849-64.

31. Falzon, L., et al., *PCR generation of large amounts of purified DNA.* J Biochem Biophys Methods, 1994. **29**(3-4): p. 251-7.

32. Wartell, R.M., and Benight, A.S., *Thermal Denaturation of DNA Molecules: A Comparison of Theory with Experiment.* Physics Report (Review Section of Physics Letters), 1985. **126**(2): p. 67-107.

33. Di Mauro, E., et al., *Activation of in vitro transcription and topology of closed DNA domains.* J Biol Chem, 1985. **260**(1): p. 152-9.

34. Clendenning, J.B., et al., *Effect of ethidium binding and superhelix density on the supercoiling free energy and torsion and bending constants of p30 delta DNA.* Biophys Chem, 1994. **52**(3): p. 191-218.

35. Naimushin, A.N., et al., *Effect of ethidium binding and superhelix density on the apparent supercoiling free energy and torsion constant of pBR322 DNA.* Biophys Chem, 1994. **52**(3): p. 219-26.

36. McDowell, D.G., N.A. Burns, and H.C. Parkes, *Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR.* Nucleic Acids Res, 1998. **26**(14): p. 3340-7.

37. Mytelka, D.S. and M.J. Chamberlin, *Analysis and suppression of DNA polymerase pauses associated with a trinucleotide consensus.* Nucleic Acids Res, 1996. **24**(14): p. 2774-81.

38. Chou, Q., *Minimizing deletion mutagenesis artifact during Taq DNA polymerase PCR by E. coli SSB.* Nucleic Acids Res, 1992. **20**(16): p. 4371.

39. Shibata, J.a.S., JM, *A Theory of Aggregation in the Thermal Denaturation Region of Multistrand Biopolymers.* Biopolymers, 1981. **20**: p. 525-549.

40. Schurr, J.M., et al., *The question of long-range allosteric transitions in DNA.* Biopolymers, 1997. **44**(3): p. 283-308.

Table 1.1 Many PCR reactions were run that contained additives in the mixture whose desired effect is to increase the specificity of the reaction. The following table shows the results that were obtained with each additive tested (column 1). The concentration ranges tested (column 2) were found from the literature and the picture shown to the right are of the result observed upon gel electrophoresis (column 3). All gels shown have been stained with SYBR®Gold nucleic acid dye, which is ~10-fold more sensitive than ethidium bromide.

Additive	Concentration	Result
Formamide	0-10%	
Glycerol	0-20%	
DMSO	0-10%	
Tween 20	0-1%	
NP40	0-1%	
Ethanol	0-10%	
BSA	0-1 μg/ml	
PEG	0-10 mM	
TMAC	15-100 mM	
DMF	0-20%	
Acetamide	0.8-1.0 M	
Glycine	0-10 mM	
Betaine	0.88-1.41 M	
Triton X-100	0-1%	
THF	0-10%	
EtBr	1 μg/ml	

Table 1.2 δG values (cal/mol) of the 10 nonunique Nearest-Neighbor Stacking Interactions at 55 mM Ionic Strength

δG_{AT}	-59.8
δG_{TA}	91.6
$\delta G_{AA/TT}$	-128.4
$\delta G_{AC/GT}$	-95.1
$\delta G_{CA/TG}$	-71.7
$\delta G_{TC/GA}$	-122.0
$\delta G_{CT/AG}$	204.4
δG_{CG}	-14.4
δG_{GC}	-298.5
$\delta G_{GG/CC}$	197.2

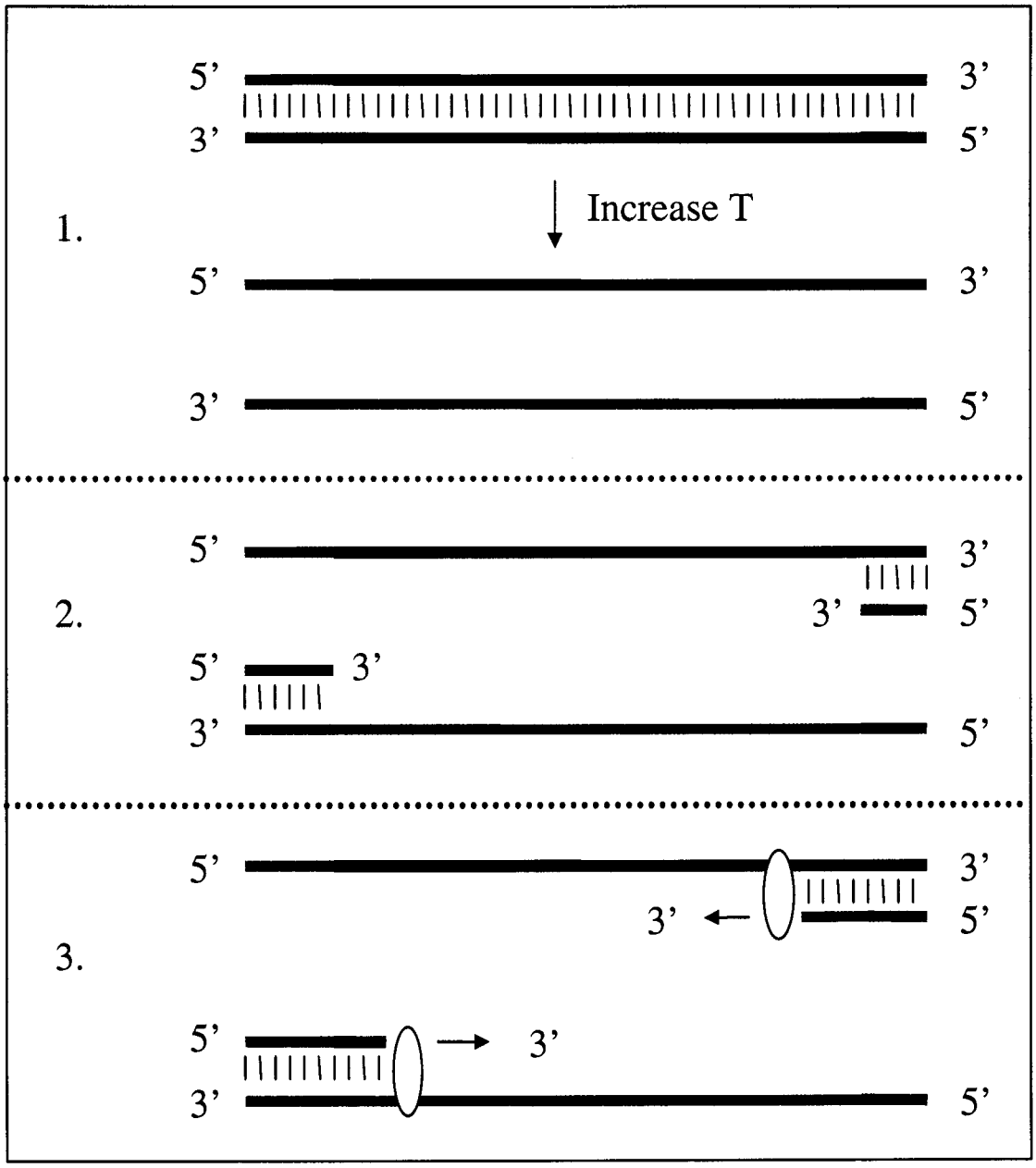


Figure 1.1 The basic steps of PCR according to the standard model. In step 1, the temperature is increased to 96°C to melt apart the template DNA strands. In step 2, the primers *should* anneal to their respective binding sites on the template strands. In step 3 a polymerase sits down and proceeds to extend the primer in the 5' to 3' directions, thus doubling the template DNA with each cycle (for reactions of unit efficiency).

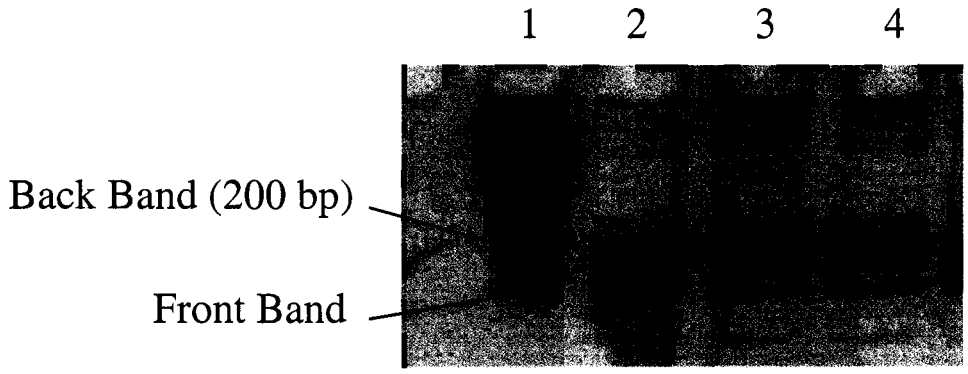


Figure 1.2 A 12% polyacrylamide gel showing the two bands typically seen in these PCR reactions. The back band is more intense by an ~ 4:1 ratio over the front band. This gel was stained with Ethidium Bromide. Lane 1: Standard 100 bp ladder from New England Biolabs. Lanes 2 – 4: PCR reaction products all show the same 2 bands.

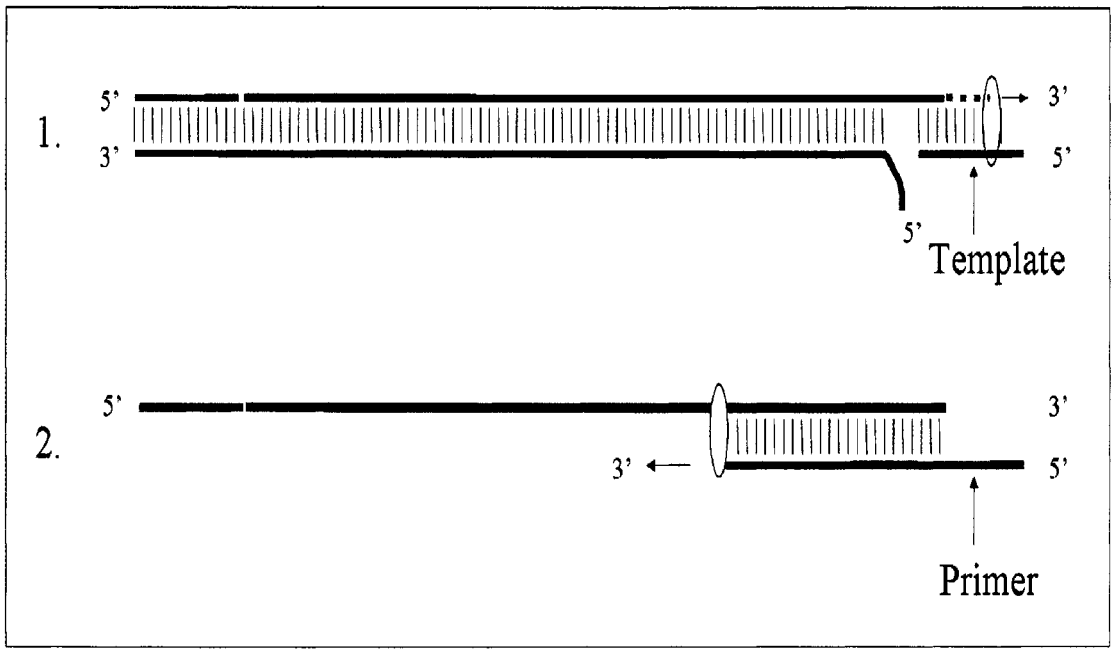
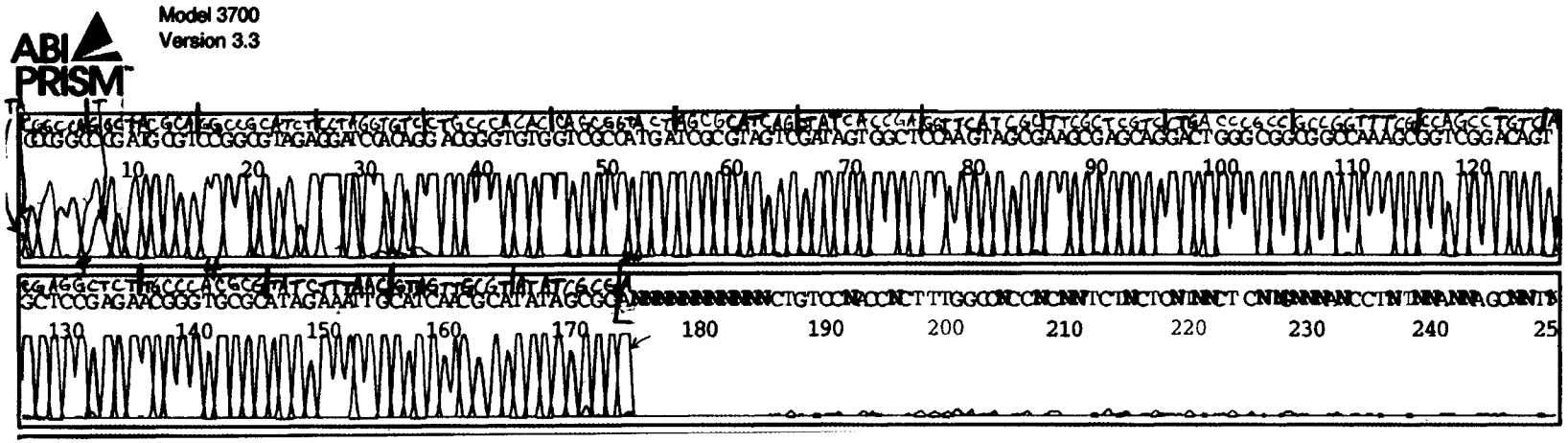


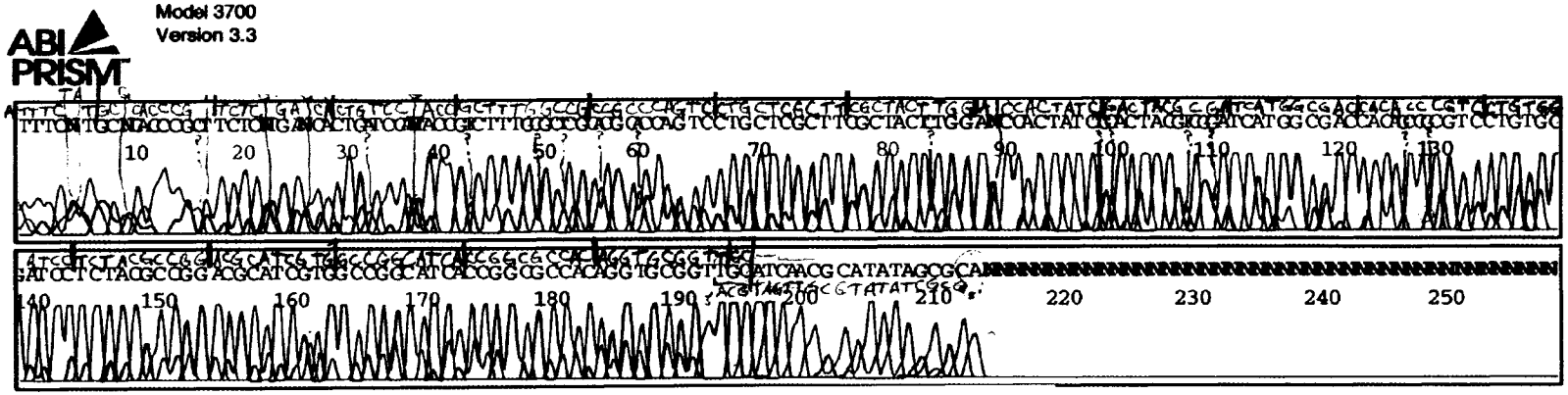
Figure 1.3 Two possible mechanisms of insertion to create a molecule 217 bp in length with the forward primer adhering to only 3 bases of the *sense* strand of DNA. In mechanism 1 (top picture), the forward primer acts as the template to complete the polymerase elongation of the sense strand. In mechanism 2 (bottom picture), the forward primer acts as a primer and gets elongated after binding to only three base pairs in the template site.

Figure 1.4 Sequence data for the 200 bp back band. Sequence A is the back band+backward primer combination which has clean, uniform peaks and therefore no trouble elongating. This is interpreted to mean that the blue primer has a clean, single strand template to elongate from (whenever possible). Sequence B is of the back band+forward primer combination. The data are a superposition such that the majority peaks report the sequence of pBR322 from 233-432 plus the complement to the green primer inserted in reverse orientation. The minority peaks correspond to the complementary sequence elongated in the opposite direction (see Figure 1.3).

A


Model 3700
Version 3.3
ABI PRISM

B


Model 3700
Version 3.3
ABI PRISM

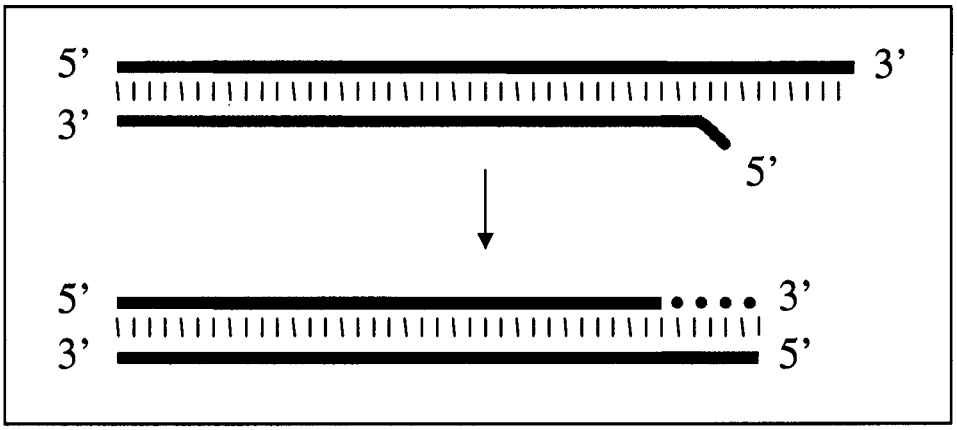


Figure 1.5 Schematic illustrations of the sequence data for the isolated front band inserted into a sequencing plasmid. The top picture shows the approximate location (and orientation) relative to the desired 200 bp template where the forward primer was inserted into the sequence. The sense strand template (shown black) contains a 5 bp run that is complementary to the forward primer, starting 1 bp removed from the 3' end of that primer. Insertion of the forward primer at this non-specific sight results in a DNA 179 bp in length. The bottom picture represents the sequence data corresponding to DNA that is 179 base pairs in length that results from insertion of the forward primer which is present at both ends of the molecule in subsequent steps of the PCR.

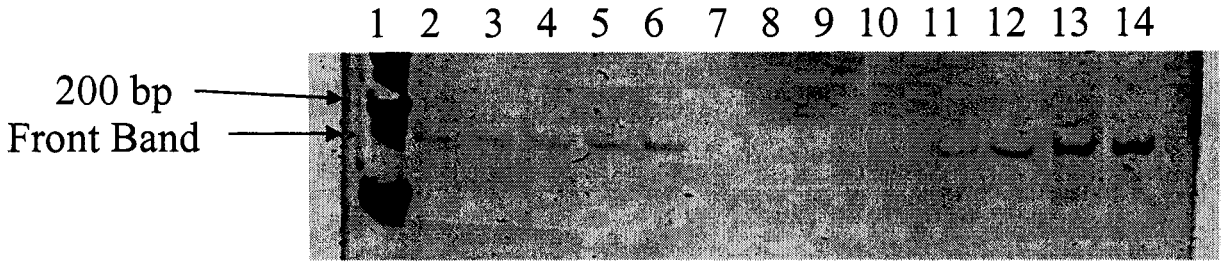


Figure 1.6 Results of PCR amplification reactions with only one primer added to the mix and plasmid pBR322 as the template DNA. Lane 1: 100 base pair standard ladder. Lanes 2-6: Titration of forward primer into mix, concentration of DNA constant, and *no* backward primer added. Lanes 7-10: Titration of backward primer into mix, concentration of DNA constant, and *no* forward primer added. Lanes 11-14: Titration of DNA into mix, forward primer constant, *no* backward primer added.

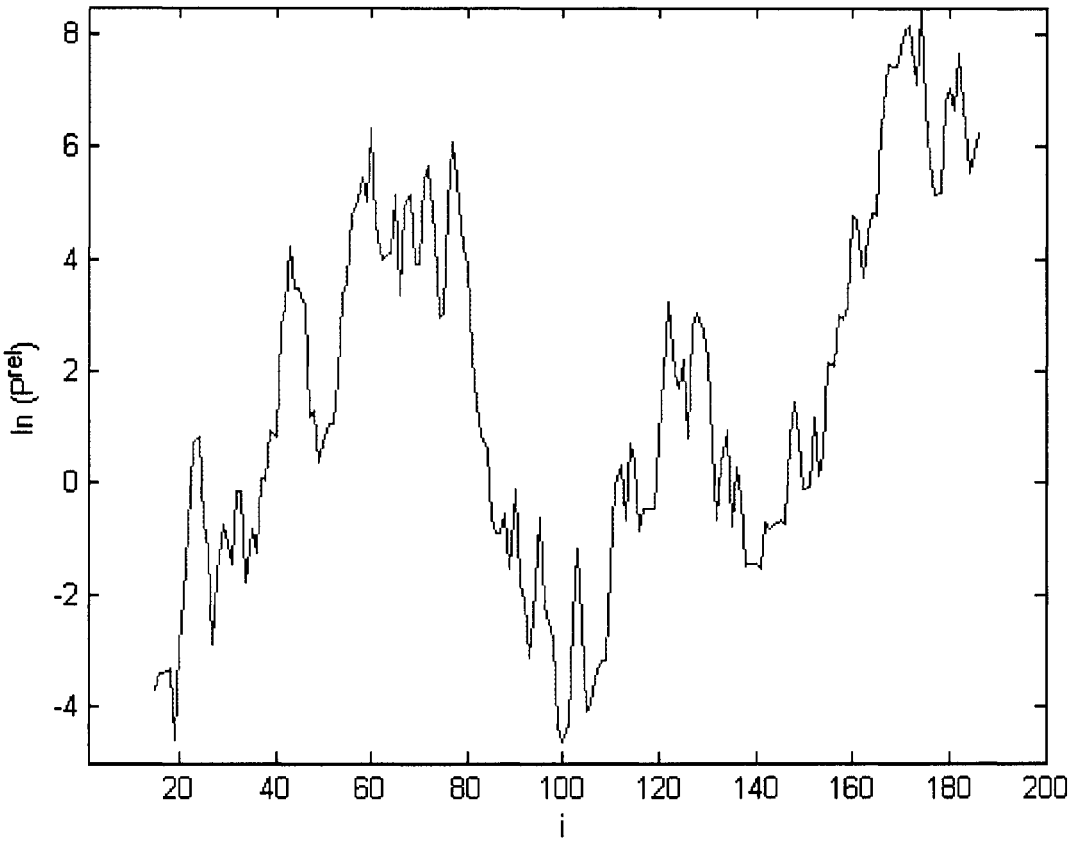


Figure 1.7 pBR322 melting profile; $\ln(P^{rel})$ vs. *i* (basenumber).
Melting probabilities were calculated based on the thermodynamic values listed in Table 2.2, as described in the text. Figure 1.7 was calculated $T_m = 96\ ^{\circ}C$ that corresponds to the temperature of melting in the PCR reactions reported throughout chapter 1. On the abscissa is plotted *i*, which is the base number of the central base for a given window. The window size in this figure are 29 bp length (i.e. for $i = 20$, P^{rel} is calculated, as described in the text, for the bases $n = 6 - 34$).

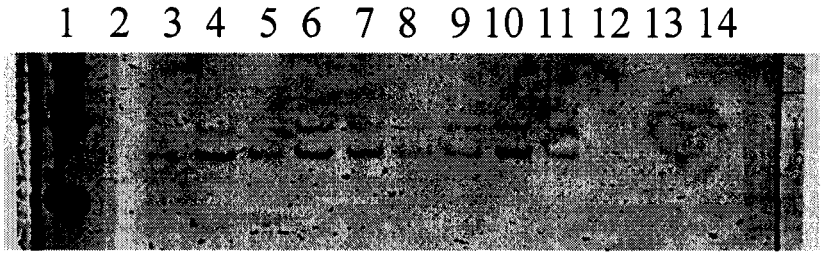


Figure 1.8 Backward primer titration in 0% EG. When the backward primer is titrated into the PCR reactions, no change in the relative intensity (as determined by eye) is observed. The concentration of backward primer increases from left to right (lanes 4 to 11) such that lane 11 is 5x the concentration of backward primer as that added in lane 4. The concentration of forward primer is equal to the concentration of backward primer in lane 4 and is held constant throughout.

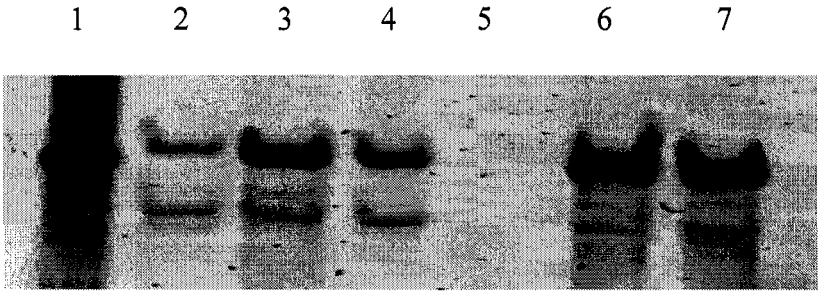


Figure 1.9 Backward primer titration in 17.5% EG. When the backward primer is titrated into the PCR reactions, a clear change in the relative intensity of the two bands is observed. The concentration of backward primer increases from left to right (lanes 2-7) such that lane 7 contains 20x the concentration of backward primer as that of lane 2. Lane 1 shows the 200 base pair band of a standard ladder. The intensity ratios of back band to front band for each lane were found to be: lane 2 (1.0), lane 3 (9.8, lower bound due to saturation of the back band signal), lane 4 (4.5), lane 6 (57.8, lower bound) and lane 7 (42.7, lower bound).

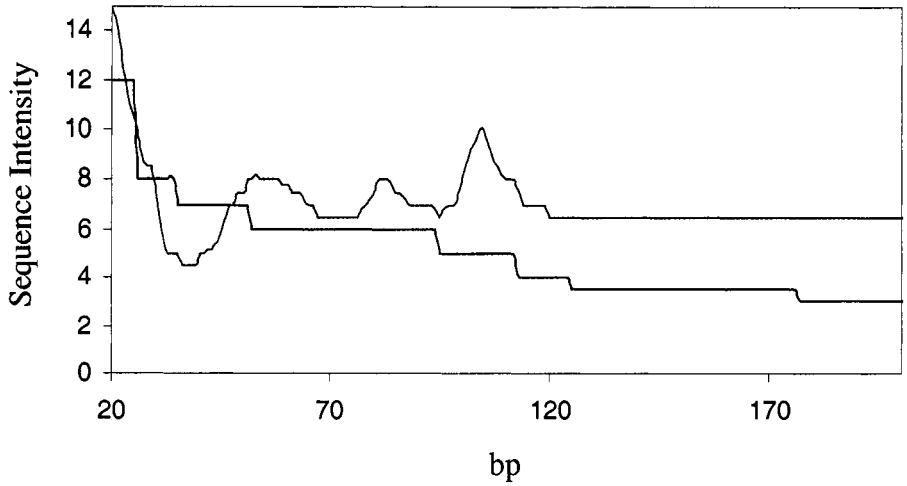


Figure 1.10 Sequence data results for the back band with either the forward primer or backward primer. This plot shows the trend of the data (measured by visual inspection from the intensity peak heights) for the forward and reverse primed sequences obtained from the same back band sample. The blue line is the data for the back band + backward primer combination and shows monotonically and roughly exponentially decaying behavior. The green line illustrates the trend of the data for the back band + forward primer combination. This combination doesn't show the same trend as those with the backward primer and, indeed is far from monotonic. There is an initial dip in the data before it increases, fluctuates and levels off to a nearly constant value.

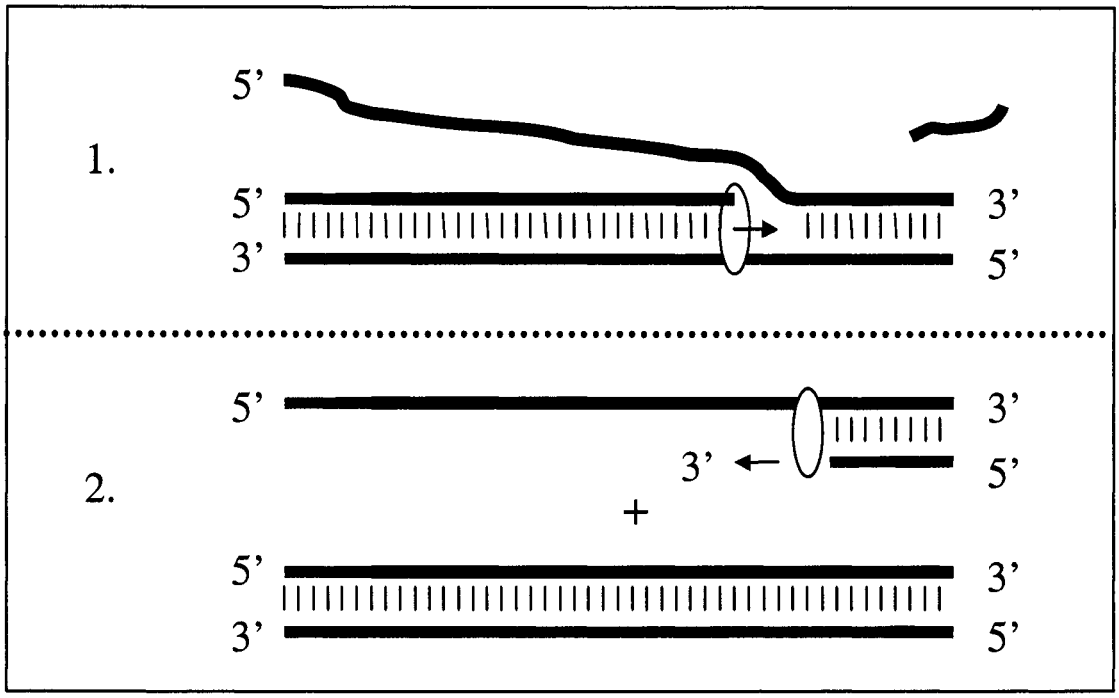
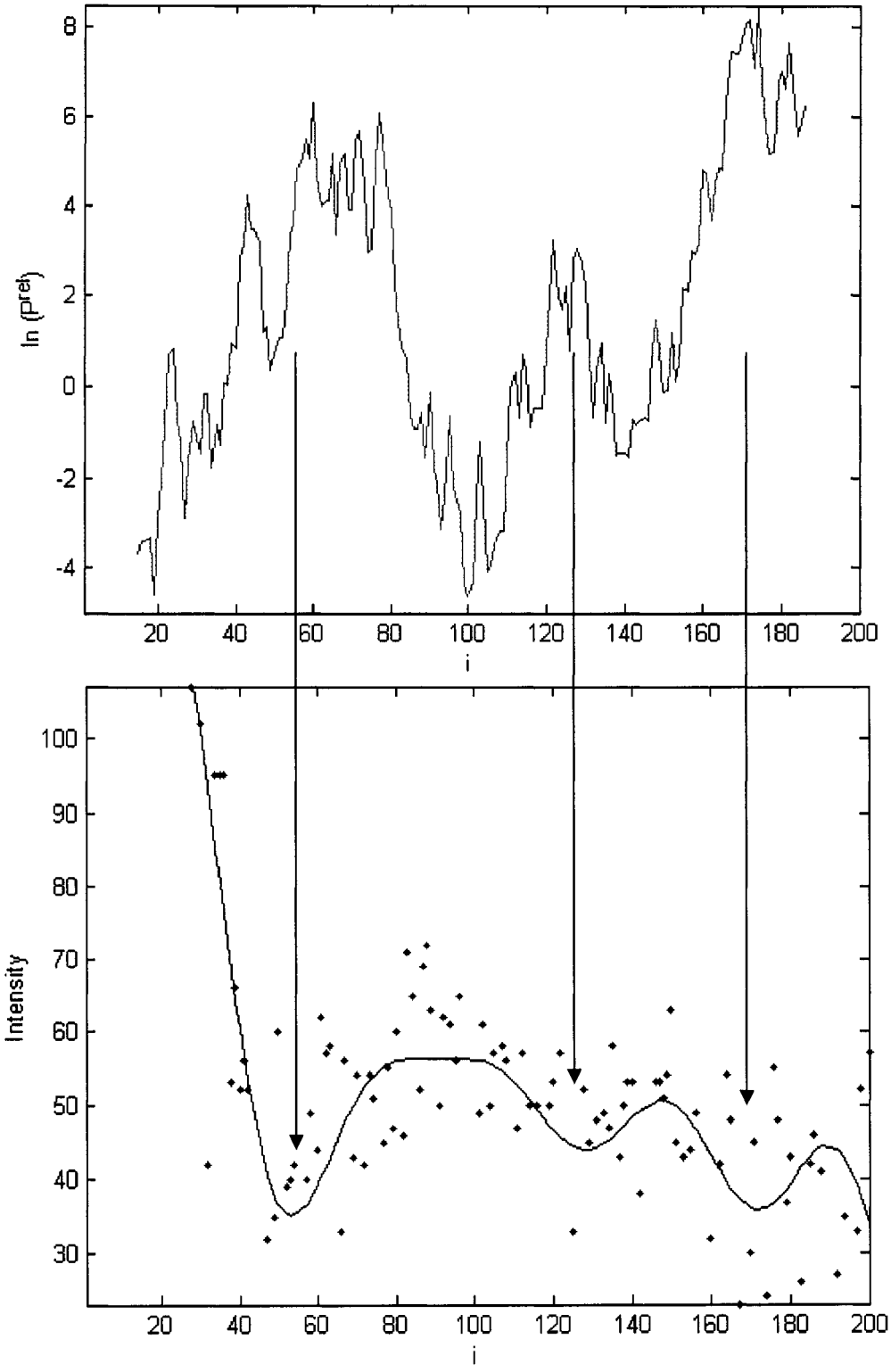


Figure 1.11 Proposed PCR mechanism for amplification of the isolated 200 bp backband. In step 1, the forward primer anneals to its complementary sequence, which has melted apart, and proceeds to amplify the sequence. However, because the entire sequence of DNA isn't able to melt apart, the polymerase has to act with some helicase activity to unwind the DNA as it extends. Once the black strand is liberated, it can bind the backward primer, which in turn is extended in an unhindered manner, as shown in figure 2.

Figure 1.12 Correlation between the melting probability and the sequence data. An apparent correlation exists between those parts of the 200 bp template sequence that melt with a low probability and regions where the polymerase stalls (observed as regions of low intensity in the sequence data). The melting profile is the same as shown in Figure 1.7 and is calculated for a $T_m = 96\ ^{\circ}C$ (the sequencing reactions are also run with a melting temperature of 96 °C). The abscissas for both plots are *i*, which is the base number of the central base for a given window. The window size used in these calculations are 29 bp (i.e. for $i = 20$, P^{rel} is calculated for the window of bases $n = 6 - 34$). The bottom plot are the intensity data from a submission of the back band (see also Figure 1.10) that has regions of low intensity and correspond to places where the enzymes are thought to stall at. The line extending through the data were obtained from a fit calculated by a fourier-type analysis and combined with a χ^2 minimization protocol that can be found in Appendix A.


8
6
4
2
0
-2
-4
ln (P^rel)
20
40
60
80
100
120
140
160
180
200
i
100
90
80
70
60
50
40
30
Intensity
20
40
60
80
100
120
140
160
180
200
i

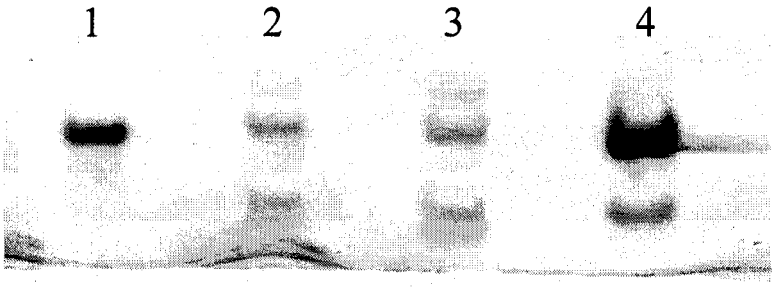


Figure 1.13 PCR products from reactions containing excess polymerase and plasmid pBR322 as template. This experiment showed that preferential amplification favoring the back band over either of the two front bands occurred in the presence of excess enzyme. Lane 1 is the 200 bp band of a standard ladder. The reactions in lane 2 contained 1x the concentration of enzyme as a typical reaction (i.e. this was the normal amount of enzyme added to the reaction) while lane 3 contained 2x the concentration of enzyme. Lane 4 contained 4x the concentration of enzyme. These results show that in the presence of excess enzyme, the backward primer competes more effectively with the forward primer and thus produces a dramatic increase in the intensity of the back band. In lane 4, only a single front band is present. This likely arises because the concentration of both the sense and anti-sense strands (200 bases in length) of DNA is high. The sequences 179 bases long apparently can't find each other (i.e. amplification efficiency is low) and so no homo-species is observed. It is unknown why the backward primer can compete, and amplify, in these reactions

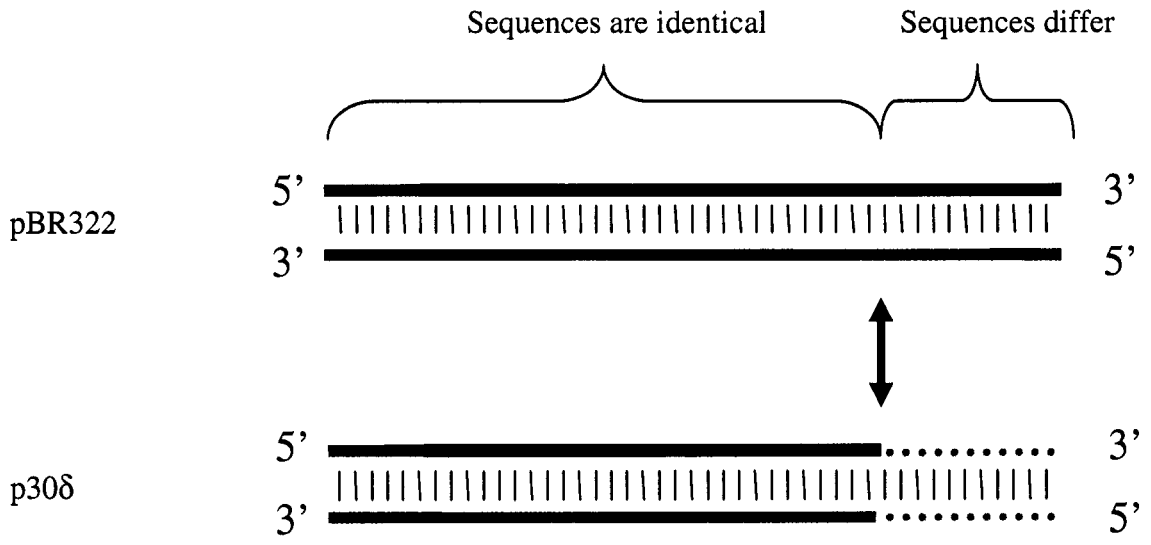


Figure 1.14 pBR322 vs. p30δ sequence homology for the 200 bp region of interest extending from bp 233 to 432 for each sequence; the sequences shown dashed in the p30δ plasmid have a different sequence of DNA than the corresponding sequence of pBR322. The 142 bp on the left side of both molecules are identical. The forward primer used was the same for both reactions (i.e. those with either pBR322 or p30δ) while the backward primer was unique to the anti-sense sequence of each template.

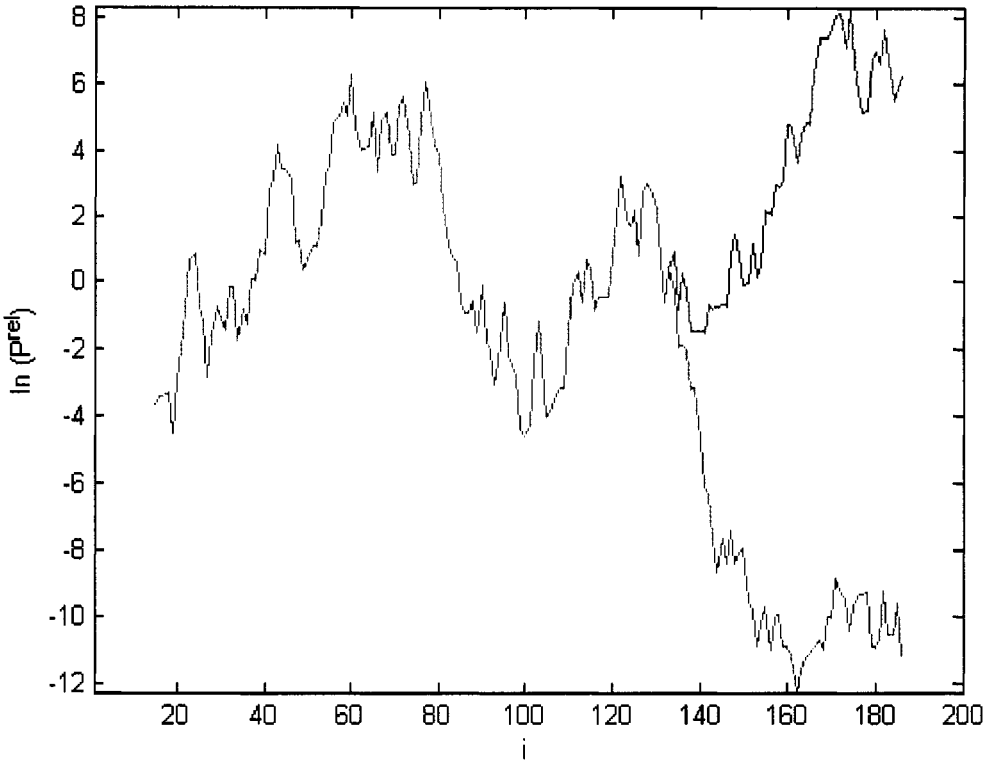


Figure 1.15 $\ln(P^{rel})$ vs. *i* (basenumber)

Figure 1.15 shows a comparison of the melting profiles for pBR322 (blue) and p30δ (red). The sequences of both plasmids are identical from bases 1 – 142 so their melting curves coincide; only that of p30δ has been shown. Melting probabilities were calculated based on the thermodynamic values listed in Table 2.2, as described in the text. Both curves were calculated for a melting temperature of $T_m = 96$ °C that is the melting temperature used in the PCR reactions reported throughout chapter 1. The abscissa is plotted as *i* which is the base number of the central base for a given window, the window size used in these calculations were 29 bp (i.e. for $i = 20$, P^{rel} is calculated for the window of bases n = 6 – 34). From the profile comparison it is evident that p30δ is predicted to exist in the unbound state with a much higher probability than pBR322 in the region of the backward primer (right side of the figure).

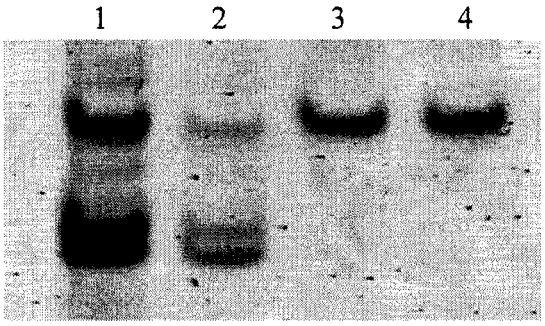


Figure 1.16 PCR reaction products with pBR322 (lanes 1 and 2) and p30δ (lanes 3 and 4) as the template. The reactions for p30δ resulted in 1 band of DNA 200 bp in length. p30δ has a lower %GC and melts with a higher probability than the corresponding sequence of pBR322 in the region of the backward primer binding site (see Figure 1.15).

Chapter 2

A Structural Transition Induced by Ethylene Glycol

2.1 Introduction

DNA secondary structures are typically classified into three main families, specifically the right-handed A and B (including the former C) families, and the left-handed Z-family. Different conformational substates within the right-handed families have also been detected by various methods, although specific structural details are lacking in most cases (Table 1 in [1]). The *average* secondary structure of any DNA may respond to various perturbations, whenever some significant fraction of its sequence undergoes a structural transition between conformational substates. While it is true in principle that nearly any kind of perturbation can alter the average secondary structure, not all perturbations suffice to *significantly* shift the distribution (of sequence among the conformations) that defines the DNA in its reference state. Monitoring such transitions has proven challenging since the differences in properties can be subtle and therefore difficult to quantitate with reliable accuracy. Nonetheless, physical parameters that describe the 'state' of the DNA helix have been reported (Table 2.1) and used to characterize distinctive changes in the duplex conformation that correspond to structural transitions. In chapter 2 it will be shown that the twist energy parameter (E_T) that governs the supercoiling free energy can be measured as a function of the water activity (a_w), specifically $-\ln(a_w)$, to reveal such a transition.

Solution water activity can be varied by addition of neutral solutes (osmolytes) that show little or *no* preferential interaction for the macromolecule relative to the interaction of the macromolecule with water (the solvent). Both are expected to be on the same order of magnitude. However, because standard osmolytes are typically larger than water, they are excluded by the macromolecule from a larger volume than the solvent water molecules, whose spherical centers can approach to a closer distance around the macromolecule. Courtenay et al. determined the preferential interaction coefficients, Γ_{os}, for seven different osmolytes with the protein bovine serum albumen (BSA) [2] and

Schurr et al. used these coefficients to calculate the relative magnitudes of the contributions due to an excluded volume effect on one hand or an osmolyte-water exchange equilibrium on the other with the surrounding solution [3]. They found that in all cases the contribution from the excluded volume exceeded the contribution from the exchange reactions. They also found that under the assumption that the exchange contribution is dominated by sites in the first surface-contiguous layer, the ratio of the average exchange constant to its neutral random value lies in the range 1.0 ± 0.15. The waters near the surface of the molecule are readily exchanged with the surroundings and so are not 'bound' to the macromolecule at all. It is more accurate to think of them as 'associated' with the macromolecule primarily as a consequence of the exclusion of osmolyte centers from a greater volume. The general approach to studying changes in 'associated' waters has come to be known as the osmotic stress method.

The osmotic stress method can be used to study structural transitions in DNA. It will be shown below that the small neutral osmolyte Ethylene Glycol (EG), when added into solution to vary a_w, induces a structural change in the DNA double helix. As briefly stated above, this transition is observed by monitoring the twist energy parameter (E_T) as a function of $-\ln(a_w)$, where a_w is the activity of water, and produces data that exhibit a sigmoidal curve. These data can be fitted by a two-state, two-parameter model, wherein the DNA exhibits two conformations that are involved in a hydration-coupled conformational equilibrium. We interpret the data in terms of independent domains, each undergoing an all-or-none transition, and extract values for the equilibrium constant (K_o) and n, the number of 'associated' water molecules per cooperative domain that are lost during the transition. These data can also be fitted near the midpoint by a continuous range of two-state, three-parameter models, which extend from the non-cooperative to highly cooperative regimes. All such models predict a transition with the same slope and midpoint, provided that the ratio $n'/J = 34$ (where n' = the number of 'associated' waters lost per bp and $1/J \sim$ the domain size in bp). No independent experimental information was available to determine the average size of each cooperative domain. However, Spink and Chaires have shown that 4 waters are lost per bp upon melting of the DNA [4]. The

number of waters lost per base pair is expected to be much less than 4, since this DNA is not melting.

Those waters associated with the DNA have thermodynamic properties (such as compressibility, molar volume and molar enthalpy) that are different from bulk water [5, 6] and have been shown to play a role in the binding reaction of certain ligands with DNA [7-10]. X-ray diffraction experiments have revealed two hydration layers associated with the DNA that contain about 20 total water molecules per base pair [11] and higher resolution diffraction studies reveal up to 4 layers of ordered water molecules in the minor grooves of certain sequences [12, 13]. Spink and Chaires found a linear relationship between T_M^{-1} and $\ln(a_w)$ (Figure 4 of [24]) for the 3 different cosolutes ethylene glycol, acetamide and glycerol with *E. coli* DNA. The data were taken from 0 – 20 w/v % cosolute and all fell on a common line which suggests that the curve is largely determined by the change in waters, Δn_w, and not by preferential interactions of the cosolute for one state over another (i.e. $\Delta\Gamma$ is a constant). This evidence supports a two-state transition model.

Any allosteric transition model whereby the secondary structure of a sizeable domain of DNA is converted from its 'ground' state to some other 'excited' state, regardless of whether it is induced by the binding of a ligand or by dehydration of the DNA due to a change in the activity of water, could serve a particular functional role in remote signaling that occurs in, for instance, regulating gene transcription [1, 14, 15]. Chapter 2 will address the following questions specifically:

Q1: Does dehydration of DNA by EG significantly alter the twist energy parameter that governs the supercoiling free energy?

Q2: Does dehydration of DNA significantly affect its intrinsic twist and circular dichroism (CD) spectrum?

Q3: How many 'associated' waters per cooperative unit are lost during the conformational transition?

Q4: What fractions of the DNA are in state 2 under standard 0.1 M NaCl aqueous buffer conditions?

Q5: Is the loss of 'associated' water at 37°C coupled to a conformational equilibrium involving two duplex states characterized by distinct E_T and δl_0 values (δl_0 is the difference between l_0 values, or intrinsic twist, in the presence and absence of osmolyte)?

Q6: What insights do cooperative and non-cooperative models offer near the midpoint of the transition?

Q7: What elastic properties can DNA in state 2 be expected to exhibit compared to DNA in state 1?

2.2 Theory

The topoisomer distribution method can be used to monitor a structural transition induced by adding the osmolyte ethylene glycol (EG). In this method, supercoiled DNA is relaxed by Calf Thymus Topoisomerase I resulting in a gaussian distribution of topoisomers that typically contain 7 to 9 visible topoisomers. Because this distribution is a gaussian it is characterized by two numbers of significance, its mean ($\mu = l_0$, the intrinsic twist) and variance ($\sigma^2 = N/2E_T$), where N is the number of bp and E_T is the twist energy parameter), a more formal introduction to these parameters follows. The superhelical strain of a circular duplex DNA is characterized by its linking difference, $\Delta l = l - l_0$, where l is the linking number (number of turns of one single-strand around the other), $l_0 = N\phi_0$ is the intrinsic twist, N is the number of bp, and ϕ_0 is the intrinsic succession angle. For any given topoisomer, the number of turns of one strand around the other is a topological constant called the linking number, l and the average intrinsic twist, l_o, is generally a non-integral number. The linking number is partitioned between the twist (t) and writhe (w) according to $l = t + w$ [16, 17] and substitution for l gives $\Delta l = t - l_o + w$.

Both experiments [18] and simulations [19] on long DNAs ($N > 2000$) indicate that the free energy change to vary the linking difference from $\Delta m = m - l_0$ to Δl is

$$\Delta G_{sc} = kT(\frac{E_T}{N})(\Delta\ell^2 - \Delta m^2) \tag{2.1}$$

where k is the Boltzmann constant, E_T is the twist energy parameter, T is the temperature and N is the number of base pairs in the DNA.

The twist energy parameter, as defined in [20], is given as

$$E_T = \left(\frac{(2\pi)^2}{2kT}\right)\frac{\alpha^{eff} B\kappa_\beta{}^{eff}}{\alpha^{eff} + B\kappa_\beta{}^{eff}} \tag{2.2}$$

wherein α^{eff} is the effective torsion elastic constant, and $B\kappa_\beta{}^{eff} = \kappa_w{}^{eff}$ is the effective torque constant *for the writhe*, which should be nearly proportional to the bending constant $\kappa_\beta{}^{eff}$ under any fixed set of solution conditions [20]. Simulations confirm the invariance of B to different κ_β and to the variation of solution conditions from 0 to 20% EG [21]. B is expected to vary with certain solution conditions such as salt concentration, but not with Δl.

The DNA is assumed to consist of cooperative domains, each of which can undergo an all-or-none two-state transition according to the chemical equation:

$$DNA(1)\cdot(H_2O)_N \rightleftharpoons DNA(2)\cdot(H_2O)_{N-n} + nH_2O \tag{2.3}$$

where n is the number of 'associated' water molecules lost per cooperative domain. Previously, data were collected from 0 – 20% EG [22] at which point the transition is about half complete. These data have been extended all of the way to 40% EG and show a trajectory with a sigmoidal shape with two plateaus at different values of E_T. This is clear evidence that a conformational change is occurring. However, it remains to be determined what are the actual states involved in this transition. When the DNA is sufficiently dilute, the equilibrium constant expression can be written as [21, 22]:

$$K_0(a_w)^{-n} = \frac{[DNA(2)\cdot(H_2O)_{N-n}]}{[DNA(1)\cdot(H_2O)_N]} \equiv K \tag{2.4}$$

where K_0 is the equilibrium constant for reaction (2.3), and K is the effective equilibrium constant for the conformational transition.

Neither equation (2.3) nor (2.4) takes *explicit* account of the osmolyte, and must be regarded as non-rigorous. In a more rigorous approach, reaction (2.3) is written without taking explicit account of either water or osmolyte [23],

$$DNA(1) \rightleftharpoons DNA(2) \tag{2.5}$$

but the standard state chemical potentials of DNA(1) and DNA(2) must account for the presence of the osmolyte. Following the formalism already laid out in the theory of preferential interaction coefficients, the macromolecule is referred to as species 2 so that DNA(J) (species 2 in state J, where J = 1 or 2 is the state of the DNA) can be denoted as 2_J (or species 2 in state J). Then in the two-component solvent (with c_w and c_{os} mol/L of water and osmolyte, respectively), we have

$$\begin{aligned}
\mu_{2J}^0(c_w, c_{os}) &= \mu_{2J}^0\left(c_w^0\right) + \int \left(\frac{\partial \mu_{2J}^0}{\partial \mu_w}\right)_{T,P,c_2^\infty} d\mu_w \\
&= \mu_{2J}^0\left(c_w^0\right) + \Gamma_{2J|w}\cdot RT \int_0^{\ln a_w} d\ln a_w \\
&= \mu_{2J}^0\left(c_w^0\right) + RT\ln\left(a_w^{\Gamma_{2J|w}}\right)
\end{aligned} \tag{2.6}$$

where $c_w^0 = 1/\bar{V}_w$ is the concentration of pure H_2O, the concentration of macromolecule is taken in the limit of infinite dilution, $C_{2J} \to 0$ (and is symbolized by C_{2J}^∞), and

$\Gamma_{2J|w} \equiv (\partial\mu_{2J}{}^0/\partial\mu_w)_{T,P,C_{2J}^\infty}$ is a *preferential interaction coefficient.* $\Gamma_{2J|w}$ was previously expressed in terms of DNA-water and DNA-osmolyte pair correlation functions [3], which were in turn evaluated in terms of hard core excluded volume and osmolyte-solvent exchange contributions [3]. $\Gamma_{2J|w}$ may in principle be either positive or negative. In the two-component solvent that surrounds the DNA(J) in its standard state, the Gibbs-Duhem equation gives $d\mu_{os} = -(c_w/c_{os})d\mu_w$, so there is only a single *independent* variable, $d\mu_w$, which accounts for the presence of both water and osmolyte. The standard free energy change for the reaction (2.5) *in the mixed solvent* is then

$$\Delta G^0(c_w, c_{os}) = \mu_{22}^0(c_w^0) - \mu_{21}^0(c_w^0) + RT\ln(a_w{}^{\Delta\Gamma}) \tag{2.7}$$

where $\Delta\Gamma \equiv \Delta\Gamma_{22|w} - \Gamma_{21|w}$ is the difference in preferential interaction coefficients of DNA(2) and DNA(1) with water. The effective equilibrium constant for the reaction (2.5) is

$$K \equiv e^{-\left(\mu_{22}^0(c_w^0) - \mu_{21}^0(c_w^0)\right)/RT} a_w{}^{\Delta\Gamma} = K_o a_w{}^{\Delta\Gamma} \tag{2.8}$$

where $K_o = \exp[(\mu_{22}^0(c_w^0) - \mu_{21}^0(c_w^0))/RT]$ is the equilibrium constant in the absence of ethylene glycol. If we set $n = -\Delta\Gamma$, then the equilibrium constant in the non-rigorous equation (2.4) has the same form as in the rigorous equation (2.8). This indicates that n must rigorously be interpreted as $n = -\Delta\Gamma$, which could still be regarded as the drop in associated water molecules that accompanies the transition. In any event, it is not necessary to take further account of the osmolyte.

The equilibrium fractions of subunits in states (1) and (2) are [21-23]

$$f_1^0 = 1/(1+K);$$
$$f_2^0 = K/(1+K) \tag{2.9}$$

The effective torsion elastic constant for a DNA composed of cooperative domains, each of which experiences an equilibrium like either that in equation (2.3) or equation (2.5), can be rewritten using Eqs. (B8) and (B7) in appendix B of [23] and is found to be given by

$$\alpha^{eff} = \left[\frac{f_1^0}{\alpha_1} + \frac{f_2^0}{\alpha_2} + \frac{f_1^0 f_2^0 \left(\phi_2^0 - \phi_1^0\right)^2}{2kT} \right]^{-1} \tag{2.10}$$

where α_1 and α_2 are the torsion elastic constants, and ${\phi_2}^0$ and ${\phi_1}^0$ are the intrinsic twists, of the type 1 and 2 domains, respectively. The last term in square brackets in equation (2.10) arises from a twisting strain-induced shift in the equilibrium. If the difference, ${\phi_2}^0 - {\phi_1}^0$, is sufficiently small ($\leq 4.7°$), as is the case here, then the equilibrium is effectively frozen during twisting, the last term is negligible compared to the first two terms, and

$$\alpha^{eff} = \left[\frac{f_1^0}{\alpha_1} + \frac{f_2^0}{\alpha_2} \right]^{-1} \tag{2.11}$$

The treatment of planar bending is analogous to that for twisting and can be rewritten using Eqs. (B8) and (B7) in [23] to yield an effective bending elastic constant that is given by

$$\kappa_\beta^{eff} = \left[\frac{f_1^0}{\kappa_{\beta 1}} + \frac{f_2^0}{\kappa_{\beta 2}} + \frac{f_1^0 f_2^0 \left(\beta_2^0 - \beta_1^0\right)^2}{2kT} \right]^{-1} \tag{2.12}$$

where $\kappa_{\beta 1}$ and $\kappa_{\beta 2}$ are the bending elastic constants, and ${\beta_1}^0$ and ${\beta_2}^0$ are the intrinsic bends of the 1 and 2 type domains. Again the last term in square brackets in equation (2.12) arises from a bending strain-induced shift of the equilibrium. If the difference, ${\beta_2}^0-{\beta_1}^0$, is sufficiently small ($\leq 4.7^\circ$), as assumed here, then the equilibrium is effectively frozen during bending, the last term is negligible compared to the first two terms, and

$$\kappa_\beta^{eff} = \left[\frac{f_1^0}{\kappa_{\beta 1}} + \frac{f_2^0}{\kappa_{\beta 2}} \right]^{-1} \tag{2.13}$$

Substitution of equations (2.8), (2.9), (2.11), (2.13) into (2.2) yields [21, 22]:

$$\frac{1}{E_T} - \frac{1}{E_{T1}} = \left(\frac{1}{E_{T2}} - \frac{1}{E_{T1}}\right) \frac{K_0 (a_w)^{-n}}{(1 + K_0 (a_w)^{-n})} \tag{2.14}$$

where E_T is the effective twist energy parameter for the DNA containing mixed domains, E_{T1} applies for DNAs containing only type 1 domains with α_1 and $\kappa_{\beta 1}$, and E_{T2} applies for DNAs containing only type 2 domains with α_2 and $\kappa_{\beta 2}$. E_T is defined as N times the free energy to change Δl from 0 to 1 in units of $k_B T$.

The difference, $\delta l_0 = l_0(c_{os}) - l_0(0)$, between intrinsic twist values in the presence and absence of osmolyte is predicted to be [21, 22]

$$\delta l_0 = -\Delta_a + \frac{(l_{02} - l_{01}) K_0 (a_w)^{-n}}{\left(1 + K_0 (a_w)^{-n}\right)} \tag{2.15}$$

where c_{os} is the osmolyte concentration and Δ_a is a constant that depends upon the fixed buffer components. l_{01} and l_{02} are the intrinsic twists for state 1 and state 2, respectively.

2.3 Materials and Methods

Native supercoiled p30δ DNA (4932 bp) was equilibrated, or relaxed, at 37°C for 6 hours by topoisomerase I in the presence of 0 to 40 w/v% of EG. Topoisomerase I was added via 3 additions at time 0 (reaction start time), 2 and 4 hours to ensure completion of the reaction. After 6 hours, the reaction was terminated by extraction with a phenol-choloroform-isoamyl alcohol mixture that removes the enzyme. The topoisomers were then separated by gel electrophoresis in a medium containing sufficient chloroquine that all visible topoisomers were slightly positively supercoiled. After removing the chloroquine, gels were stained with ethidium bromide (see Figure 2.1A for a representative gel) under conditions, wherein the amount of dye bound to each topoisomer is proportional to its relative concentration. The fluorescence in the gel was imaged and quantitated with a Fluorimager SI gel scanner. Longitudinal sections in the direction of migration down the centers of representative lanes are shown in Figure 2.1B. Typically 12 lanes were run and analyzed for each reaction condition. The integrated intensity of each band was reckoned using the commercial software program ImageQuant 5.1.

The linking number of the most populous topoisomer (i.e. the most intense band) is the nearest integer to the intrinsic twist, $l_{mp} = NINT(l_0)$, under the conditions of the topoisomerase reaction. Each topoisomer band is characterized and indexed by its excess linking number, $l_{ex} \equiv l - NINT(l_0)$ which simply centers the peaks around the most populous topoisomer ($l_{ex} = 0$). For the majority of reactions 7-9 topoisomers are present and l_{ex} is indexed as . . . -3, -2, -1, 0, +1, +2, +3 . . . . Because the intensity ratio of each band is proportional to the amount of DNA present in that band, the relative intensity of each band was equated to the equilibrium population ratio, c_{lex}/c_0, for the relevant topoisomers [24]. These ratios were fitted by the relation

$$c_{lex} / c_0 = \exp[-(\frac{E_T}{N})((l_{ex} + \Delta l_0)^2 - \Delta l_0{}^2)] \qquad (2.16)$$

where $\Delta l_0 = l_{mp} - l_0$ is the linking difference of the most populous topoisomer. In equation (2.16), the adjustable parameters are E_T and Δl_0. Figure 2.2 shows a typical best-fit 'curve' for a single lane co-plotted with the corresponding experimental values obtained using equation (2.16) for 0% (black line) and 40% (gray line) EG. A distinct shift of l_0 has occurred between 0% and 40% EG which indicates that a change in secondary structure of the DNA has taken place.

The change in intrinsic twist upon varying the osmolyte concentration from 0 to c_{os} is [21, 22]

$$\begin{aligned} \delta l_0 &\equiv l_0(c_{os},T) - l_0(0,T) \\ &= NINT(l_0(c_{os},T)) - NINT(l_0(0,T)) - \Delta\Delta l_0 \end{aligned} \tag{2.17}$$

where $\Delta\Delta l_0 = \Delta l_0(c_{os}) - \Delta l_0(0)$ is the difference between the linking differences of the most populous topoisomers in the presence and absence of osmolyte, and $NINT(l_0(c_{os})) - NINT(l_0(0))$ is readily obtained from gel data fits like that in Figure 2.2. Any time reaction products at two different concentrations of osmolyte are made, the two sets of topoisomer reactions should be run on the same gel prior to comparison.

Measurements of a_w were made using vapor pressure osmometry in the laboratory of Charles H. Spink (unpublished data) who measured the water activites up to 20% EG. a_w's from 20-40% EG were obtained from freezing point depression measurements in the engineering literature [25] (for sample calculation see Appendix C).

If the conclusions are correct, and if E_{T1}, E_{T2}, l_{01}, and l_{02} correctly describe the two different conformational states of the DNA, then it is not difficult to appreciate the challenges that need to be overcome with the current method of data collection. The magnitude of the transition being measured (through E_T) is small and the change associated with this transition is nearly as small as the error bars for any individual data set, meaning these measurements are essentially being made in the noise. Collection of data is not a timely event either and at the current rate the reproducibility error of the measurements is *not* sufficiently averaged out. The error bars associated with any single

set of data are given by the error of the mean and is related to the standard deviation, for that set, through the factor $1/\sqrt{N}$. ($\sigma_m = \sigma/\sqrt{N}$; σ_m is the standard error of the mean, σ is the standard deviation of the data set and N the number of data points collected). This is essentially a problem of signal averaging and poses a concern since 4 times as much data needs to be collected to gain just a factor of 2 in resolution.

2.4 Results

2.4.1 Dehydration of DNA by EG significantly alters the twist energy parameter that governs the supercoiling free energy

DNA undergoes structural transitions as the activity of water is varied and exists as B-form in those solutions where the water activity is high [26]. a_w of the bulk solution is altered by addition of neutral osmolytes that are not expected to preferentially interact with the DNA relative to water. At the very least these osmolytes are excluded from a larger volume by the DNA and in effect create a region whereby only water is accessible. These are the 'associated' waters that are not bound at all. In addition, the DNA may exhibit some longer-range weak attraction or repulsion for the osmolytes relative to water. In any case, EG acts to reduce a_w of the bulk solution and, in doing so, induces a structural change in the DNA.

In Figure 2.3, E_T values are plotted versus $-\ln(a_w)$ for two different samples of DNA. The measurements taken on the most recent preparation of DNA are represented with black circles, while the data depicted with red triangles are from a separate, identically prepared sample produced following the same protocol for growth and isolation of the plasmid DNA [22]. The data presented herein (the black circles) are a continuation of those previous experiments and extend the data past 20% EG to 40% EG to test the two-state model and assess properties of the final state. The previous data were interpreted in terms of a two state model, even though the transition is only half completed at 20% EG. The data of the most recent sample are initially somewhat flat from 0 to 15% EG ($-\ln(a_w) = 0.011$ to 0.062), then decrease with a gentle slope from 17.5% to 25% EG ($-\ln(a_w) = 0.075$ to 0.12) and finally begin to flatten out to a second, lower plateau. This lower plateau remains relatively constant up to 40% EG

($-\ln(a_w) = 0.245$). The topoisomerase enzyme begins to lose its activity at 43% EG and is completely inactive by 47% (Figure 2.4). The experimental E_T data complete the sigmoidal trajectory that is characteristic of the two-state model. This is the first time that the twist energy parameter has been used to map out such a structural transition in DNA.

The data from the previous sample (red triangles) show the same trend as this sample, even if it has a somewhat different shape. This difference in shape is real, and presumably arises because the DNA samples have different histories, which includes exposure to topoisomerase enzymes of differing activities. The data sets were collected several months apart and used multiple sources of enzyme stock. The producer of topoisomerase I, Invitrogen Life Technologies, can offer no guarantee that the individually purchased enzyme samples will have the same activity, or that a single sample will maintain constant activity over a time frame so long. The enzyme likely loses some activity each time that it goes through a freeze-thaw cycle (-20°C to 4°C and back to -20°C) which happens every time a set of reactions are prepared.

The following strategy was adopted to obtain quantitative information about the four adjustable parameters (E_{T1}, E_{T2}, n and K_o) in Equation 2.14. Initially, the most recent EG data (black circles) in Figure 2.3 was fitted by a four-dimensional grid search that resulted in best fit values for all four of the parameters. Then, to make more precise interpretations for the parameters n and K_o, a second fitting routine was employed that used the values for each of the two plateaus, $E_{T1} = 979$ and $E_{T2} = 890$, that resulted from the four-dimensional fit. This 2nd fit of equation 2.14 to the data, which was a two parameter fit, used the Levenburg-Marquardt protocol from the curve fitting toolbox of Matlab®7.0, along with the analysis found in Appendix D to find the variances and covariances and resulted in best fit values of $n = 34 \pm 2.65$ and $K_o = 0.056 \pm 0.013$. The variances for n and K_o were $\langle \delta n^2 \rangle = 7.04$ and $\langle \delta Ko^2 \rangle = 1.73 \times 10^{-4}$, respectively, while the covariance $\langle \delta n \delta K_o \rangle = -0.0276$. The fitted line is shown co-plotted with the data in Figure 2.3 and gives a reasonable fit to the data points. From this fit, it can be concluded that dehydration by EG alters the twist energy parameter, E_T, governing the supercoiling free energy. Because the lower plateau on the right side of the fit appears arbitrarily high

compared to the three data points in that region, another fit (identical to the 2nd fit above) was done with a lower $E_{T2} = 885$ that checked the robustness of the fit parameters. The same protocol was used and produced nearly identical results, within the curve fitting bounds. The precise value of the lower plateau does not significantly affect the conclusions derived from the fit.

2.4.2 *Dehydration of DNA significantly affects its intrinsic twist and circular dichroism (CD) spectrum*

In Figure 2.5, δl_0 (from equation (2.15)) is plotted versus $-\ln(a_w)$. The sigmoidal shape of the data is more or less apparent. Moreover, the data can be described by the two-state model predicted by equation (2.14). The values of $n = 34$ and $K_0 = 0.056$ from the fit shown in Figure 2.3 were subsequently inserted into equation (2.15), along with $(l_{02}\text{-}l_{01}) = -0.8358$ (i.e. $\delta l_0(40\%) - \delta l_0(0\%) = -0.8358$). The resulting sigmoidal curve had a trajectory that followed the general trend of the data and is consistent with the two-state assumption. The error bars associated with any one data point are very small and it is of concern that no single curve could be placed close to all the data. For unknown reasons there is a reproducibility error between different sets of reactions (i.e. different reaction preparations) that exceeds the magnitude of the statistical error (or the error from different tubes from the same reaction preparation). The red data points are the previous data from Rangel et al. [22]. An astonishing result from the current measurements, though, is that less than one turn (~ 0.8 turn) of the helical axis can be distinguished out of 474 turns present in p30δ. This amounts to a measurement with better than 0.2% (~ 0.17%) statistical precision!

A structural transition is captured by CD measurements made on the DNA in 0% (solid black line) and 40% (solid gray line) EG (Figure 2.6). These data support the notion that a change in secondary structure has occurred over a sizeable fraction of the DNA. The intent of Figure 2.6 is only to show that a structural change has occurred. The black spectrum has a characteristic B-DNA shape and the gray spectrum is not so different, so it appears that the DNA is engaging in a structural change, perhaps within

the 'B-family'. This is consistent with the other data presented in Figures 2.3 and 2.5 where a subtle transition is seen to occur.

The smoothed data in Figure 2.6 were generated using a standard fourier analysis that can be found in any textbook on the subject. Computer codes written, tested, and implemented to carry out this analysis were created in-house. A thorough explanation of the fourier analysis implemented for analysis of these CD data, including the explicit code employed can be found in Appendix B. The curves shown in Figure 2.6 include 7 fourier terms total (sine + cosine). The fourier series was truncated by a χ^2 protocol that monitored χ^2 as a function of *kincluded* (ktotal = $2 \cdot$*kincluded* + 1) and allowed an empirical determination to be made based on the slope of the χ^2 vs. *kincluded* plot. Typically, an independent measurement reveals where this truncation should occur but in this instance no such data were available.

A word should also be said about the structural transition from the CD data. A definite change in the shape of the CD spectrum has occurred from 0 to 40% EG, however it is not clear how much of that change is due to a change in secondary structure of the DNA and how much is due to a change in the local environment of the DNA bases such as medium dielectric constant. No independent assessment was made to address this issue. CD measurments were made for 0, 10, 20, 30 and 40% EG (see Appendix B Figures B.1 and B.2) to probe for a structural change. While a complete characterization of the transition process is not expected from only 5 measured data sets, no clues regarding the validity of the two-state assumption are apparent. There are no obvious isosbestic points and the middle 3 sets (10, 20 and 30% EG) do not cluster in comparison to the two boundaries (0 and 40% EG) as might be expected from a sigmoidal transition. Regardless, it can be concluded that dehydration of the DNA significantly affects its CD spectrum as well as its intrinsic twist and twist energy parameter.

2.5 Discussion

In this section indirect results that derive from the data along with related experiments will be described.

2.5.1 *How many 'associated' waters per cooperative unit are lost during the conformational transition?*

Approximately 34 associated H_2O per cooperative domain are lost during this conformational transition. Unfortunately, the length of the cooperative domain is unknown for the current DNA. Spink and Chaires have found that about 4 bound waters/bp are lost upon melting [4] so it is anticipated that the domain size is greater than 34/4 = 8.5 since this sample has not melted and is likely greater than 34 bp which corresponds to less than 1 water molecule lost/bp during the transition. The transition is a very subtle transition and both states are suspected to be from the 'B-Family'.

2.5.2 *What fractions of the DNA are in state 2 under standard 0.1 M NaCl aqueous buffer conditions?*

Equation (2.9) gives a function describing the behavior of the equilibrium fraction of subunits as a function of a_w. By inserting the values $n = 34$ and $K_0 = 0.056$ from the most recent preparation of DNA (black circles) one can address the question, "what is the fraction of subunits in state 2 when $a_w = 0.996$ which corresponds to standard 0.1 M NaCl aqueous buffer conditions?" Upon doing so, one finds that $f_2^0 = 0.06031$. In other words, when $a_w = 0.996$, 6.03% of the base-pairs are in state 2. That is 1 out of 20 base pairs on average, which is a *not insignificant* amount.

2.5.3 *Is the loss of 'associated' water at 37°C coupled to a conformational equilibrium involving two duplex states characterized by distinct E_T and δl_0 values?*

While the data do support the general trends predicted by a two-state model, other models have not been eliminated from consideration by the current data. The two-state model is simply the most basic model that one can reasonably use to describe the observed behavior. One other alternative would be a continuous model where the DNA changes from one structure to another in a (more or less) continuous manner. The data for this model would look like a straight line, with negative slope, starting at 0% EG and monotonically decreasing to 40% EG. If there were no structural change occurring, a

straight line positioned at the average values ($E_T \sim 935$ and $\delta l_0 \sim -0.4$) and with slope = 0 would result. The current data clearly decrease from 0% up to 40% EG and also have a point of inflection separating two different and opposite concavities in the data, which is consistent with a two-state model.

2.5.4 Near the Midpoint for this Transition both Cooperative and Non-cooperative Models Produce Similar Slopes and Midpoints

As shown in Appendix E, a simplistic model similar to the Ising model of ferromagnetism [27] was derived to analyze this structural transition. This model is very similar to the all-or-none transition model outlined previously with the exception that it has incorporated a cooperativity parameter, J, whose inverse is proportional to the size of either domain at the midpoint. When J = 1 the transition is non-cooperative and the results simplify to the all-or-none transition model outlined in the theory section above. In Figure 2.7 transitions for both non-cooperative (J = 1) and cooperative (J = 1/34) models were generated using the E_{T1} and E_{T2} plateau values from the previous fits. By design, fits from both models produce a curve with the same slope and midpoints near the midpoint of the transition. This will happen as long as the ratio of p (the number of water lost/bp; this number is different than the aforementioned n which was the number of water lost/domain) to J (1/J ~ domain size) was held constant at p/J = 34. The quantity p/J is the decrease in associated water per cooperative domain. An expression relating the fraction of subunits in state 2 (θ_2) to a_w can also be found from the analysis in Appendix E and the results are shown in Figure 2.8. Again, the cooperative and non-cooperative models predict similar results near the midpoint of the transition. In the figure, as a_w approaches 0.996 (i.e. standard buffer conditions) θ_2 does not go to zero but remains at some value that is ~ 5 – 10%. This is a significant result because it means that for any standard buffer conditions where a_w is reasonably close to 0.996, approximately 5 – 10% of the DNA can be expected to reside in state 2.

2.5.5 *What Elastic Properties apply for the DNA in state 2 relative to the DNA in state 1?*

Rangel et al. [22] used fluorescence polarization anisotropy (FPA) to perform measurements of the torsion elastic constant from 0 to 19.4 w/v% EG and found that α increased by a factor of 1.37 over this range. It is a bit fortuitous that 19.4% EG also corresponds very closely to the midpoint of the observed transition, which lies at ~ 20% EG. This means that the E_T data for the most recent preparation of DNA (shown in Figure 2.3) can be combined with the FPA data of Rangel et al. to deduce the elastic properties for any DNA existing in state 2.

By equating $\alpha_{eff} = 1.37 \cdot \alpha_1$ and setting $f_1 = f_2 = 0.5$, as should be the case at the midpoint of the transition, equation (2.11) can then be used to solve for the ratio of α_2 to α_1 (see Appendix F). It is found that $\alpha_2 = 2.175 \cdot \alpha_1$. In other words, near 40% EG the DNA can be expected to exhibit a ~ 2-fold increase in torsion elastic constant. To continue, equation (2.2) was used, along with α_1, E_{T1}, α_2, and E_{T2} to solve for the ratio of the bending rigidity of the DNA in state 2 to that of DNA in state 1 and produced $\kappa_{\beta 2}/\kappa_{\beta 1} = 0.833$. The bending rigidity of the DNA is weakened for state 2 compared to state 1. If the persistence length of state 1 is $P_1 = \kappa_{b1}/k_BT = 500$Å ($k_B$ is Boltzmann's constant and T is the temperature in Kelvin), then $P_2 = 416.5$Å.

The evidence for a structural transition, which is exposed through E_T, combined with the finding that DNA(2) has a torsion elastic constant that is 2.175 times that of DNA(1) is an important finding that will have implications in various related areas. Table 2.2 shows the torsional rigidities ($C = h\alpha$; where h = 3.4 Å is the rise per base pair and α is the torsion elastic constant) obtained by four different experiments; Fluorescence Polarization Anisotropy (FPA), Cyclization Kinetics (CK), the Topoisomer Ratio (TR) method, and Single Molecule Pulling Experiments (SMP). For those experiments, wherein the DNA has a bending strain imposed on it, but is put under no tension by a pulling apparatus, the torsional rigidity increases by ~ 1.5 to 2.0-fold. Likewise, those DNAs that have no bending strain imposed on them but are put under tensile forces applied through a pulling apparatus, the torsional rigidity increases by ~ 2-fold. The

torsional rigidity of p30δ in 40% ethylene glycol is evidently similar to the values obtained for sufficiently bent or pulled DNAs. This may imply that a structural transition is occurring in any experiments, wherein the DNA is subjected to sufficient external perturbative forces.

2.5.6 Can topoisomerase I operate on previously relaxed DNA to enhance the rate of equilibration of metastable secondary structure?

In these experiments, the largest source of error was the lack of reproducibility among different data sets while the statistical errors within a data set (for a successful reaction) were very small. This indicates a source of variability in the reactions that should be taken into account. The hypothesis proposed to account for this observation is that topoisomerase I has two functions. Its main function is to nick and reseal the 'equilibrated' DNA as advertised. The second function could be to bind to relaxed DNA, still nicking and resealing as before, but in doing so, now it would be acting to relax any metastable secondary structure on a much shorter time scale than the previously observed timescale of weeks to months [1, 28]. This action could significantly affect the observed E_T values and will be addressed more thoroughly in chapter 3.

2.6 Conclusion

At 37°C, a reduction in water activity from 0.989 to 0.783 (or equivalently an increase in $-\ln(a_w)$ from 0.011 to 0.245) by adding 40 w/v% EG *decreases* E_T by 0.91 fold. That same reduction in water activity diminishes l_0 by ~0.8 turns out of 474 turns and changes the corresponding CD spectrum. Approximately 34 H_2O are lost per cooperative domain during this conformational transition but the length of the cooperative domain is unknown. This loss of 'associated' water appears to be coupled to a conformational equilibrium involving two duplex states that exhibit somewhat different elastic properties constants for both torsion and bending. The torsion elastic constant for state 2 (α_2) is 2.175 times that of state 1 (α_1). The bending constant of state 2 ($\kappa_{\beta 2}$) is 0.833 times that of state 1 ($\kappa_{\beta 1}$). If the persistence length (P_1) of the DNA in state 1 is

assumed to be 500 Å, as is found for DNA in similar buffers, then the persistence length of the DNA in state 2 (P_2) is 416.5 Å. This transition can be fitted by a range of cooperative or non-cooperative transition models as long as the ratio of the number of water molecules/bp to the inverse domain size (1/J) is held constant and equal to 34, the number of bound water lost per domain for the transition, which was found by a fit of the data using equation (2.14). In fact, the fraction of subunits in state 2 never approaches zero even in standard buffer conditions which means that some portion of the DNA will still reside in state 2 in any reasonably buffered solution, which usually applies since buffers are prepared in order to mimic the environment found inside the cell. From $K_0 = 0.056$, it can be inferred that 6.03% of the base-pairs (~ 1 bp in 20) of p30δ exists in the alternative state under $a_w = 0.996$ standard 0.1 M NaCl aqueous buffer conditions.

2.7 Notes to Chapter 2

1. Schurr, J.M., et al., *The question of long-range allosteric transitions in DNA.* Biopolymers, 1997. **44**(3): p. 283-308.

2. Courtenay, E.S., et al., *Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants in vivo and for the interpretation of "osmotic stress" experiments in vitro.* Biochemistry, 2000. **39**(15): p. 4455-71.

3. Schurr, J.M., D.P. Rangel, and S.R. Aragon, *A contribution to the theory of preferential interaction coefficients.* Biophys J, 2005. **89**(4): p. 2258-76.

4. Spink, C.H. and J.B. Chaires, *Effects of hydration, ion release, and excluded volume on the melting of triplex and duplex DNA.* Biochemistry, 1999. **38**(1): p. 496-508.

5. Chalikian, T.V., et al., *Influence of base composition, base sequence, and duplex structure on DNA hydration: apparent molar volumes and apparent molar adiabatic compressibilities of synthetic and natural DNA duplexes at 25 degrees C.* Biochemistry, 1994. **33**(9): p. 2394-401.

6. Rentzeperis, D., D.P. Kharakoz, and L.A. Marky, *Coupling of sequential transitions in a DNA double hairpin: energetics, ion binding, and hydration.* Biochemistry, 1991. **30**(25): p. 6276-83.

7. Robinson, C.R. and S.G. Sligar, *Changes in solvation during DNA binding and cleavage are critical to altered specificity of the EcoRI endonuclease.* Proc Natl Acad Sci U S A, 1998. **95**(5): p. 2186-91.

8. Sidorova, N.Y. and D.C. Rau, *Differences in water release for the binding of EcoRI to specific and nonspecific DNA sequences.* Proc Natl Acad Sci U S A, 1996. **93**(22): p. 12272-7.

9. Vossen, K.M., et al., *Role of macromolecular hydration in the binding of the Escherichia coli cyclic AMP receptor to DNA.* Biochemistry, 1997. **36**(39): p. 11640-7.

10. Qu, X. and J.B. Chaires, *Hydration changes for DNA intercalation reactions.* J Am Chem Soc, 2001. **123**(1): p. 1-7.

11. Schneider, B. and H.M. Berman, *Hydration of the DNA bases is local.* Biophys J, 1995. **69**(6): p. 2661-9.

12. Shui, X., et al., *The B-DNA dodecamer at high resolution reveals a spine of water on sodium.* Biochemistry, 1998. **37**(23): p. 8341-55.

13. Shui, X., et al., *Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations.* Biochemistry, 1998. **37**(48): p. 16877-87.

14. Parekh, B.S. and G.W. Hatfield, *Transcriptional activation by protein-induced DNA bending: evidence for a DNA structural transmission model.* Proc Natl Acad Sci U S A, 1996. **93**(3): p. 1173-7.

15. Shibata, J.H., et al., *Structures and dynamics of a supercoiled DNA.* Biochemistry, 1984. **23**(6): p. 1188-94.

16. White, J., *Self-Linking and the Gauss Integral in Higher Dimensions.* American Journal of Mathematics, 1969. **91**: p. 693-728.

17. Fuller, F., *The Writhing Number of a Space Curve.* Proceedings of the National Academy of Sciences of the United States of America, 1971. **68**(4): p. 815-819.

18. Depew, R.a.W., JC, *Conformational Fluctuations of DNA Helix.* Proceedings of the National Academy of Sciences of the United States of America, 1975. **72**(11): p. 4275-4279.

19. Gebe, J.A., et al., *Monte Carlo simulations of supercoiling free energies for unknotted and trefoil knotted DNAs.* Biophys J, 1995. **68**(2): p. 619-33.

20. Wu, P.G., et al., *Interaction of chloroquine with linear and supercoiled DNAs. Effect on the torsional dynamics, rigidity, and twist energy parameter.* Biochemistry, 1988. **27**(21): p. 8128-44.

21. Rangel, D.P., *Effects of neutral osmolytes on DNA, Ph.D. Thesis*. 2005, University of Washington.

22. Rangel, D.P., et al., *Effects of small neutral osmolytes on the supercoiling free energy and intrinsic twist of p30delta DNA.* Biopolymers, 2004. **75**(4): p. 291-313.

23. Delrow, J.J., P.J. Heath, and J.M. Schurr, *On the origin of the temperature dependence of the supercoiling free energy.* Biophys J, 1997. **73**(5): p. 2688-701.

24. Clendenning, J.B., et al., *Effect of ethidium binding and superhelix density on the supercoiling free energy and torsion and bending constants of p30 delta DNA.* Biophys Chem, 1994. **52**(3): p. 191-218.

25. Flick, E., *Industrial Solvents Handbook.* 5th ed. 1998: William Andrew Publishing/Noyes. Chapter 7.

26. Westhof, E., *Water: an integral part of nucleic acid structure.* Annu Rev Biophys Biophys Chem, 1988. **17**: p. 125-44.

27. Newell, G.a.M., EW, *On the Theory of the Ising Model of Ferromagnetism.* Reviews of Modern Physics, 1953. **25**(2): p. 353-389.

28. Song, L., et al., *Evidence for allosteric transitions in secondary structure induced by superhelical stress.* J Mol Biol, 1990. **214**(1): p. 307-26.

29. Schurr, J.M., Fujimoto, B.S., Wu, P. & Song, L., *Fluorescence Studies of Nucleic Acids. Dynamics, Rigidities, and Structures, in Topics in Fluorescence Spectroscopy vol. 3, Biochemical Applications*, ed. J.R. Lakowicz. Vol. 3. 1992, New York: Plenum press. 137-229.

30. Delrow, J.J., et al., *Effect of temperature on DNA secondary structure in the absence and presence of 0.5 M tetramethylammonium chloride.* Biopolymers, 1998. **45**(7): p. 503-15.

31. Fujimoto, B.S. and J.M. Schurr, *Dependence of the torsional rigidity of DNA on base composition.* Nature, 1990. **344**(6262): p. 175-7.

32. Heath, P.J., et al., *Effect of bending strain on the torsion elastic constant of DNA.* J Mol Biol, 1996. **260**(5): p. 718-30.

33. Naimushin, A.N., B.S. Fujimoto, and J.M. Schurr, *Dynamic bending rigidity of a 200-bp DNA in 4 mM ionic strength: a transient polarization grating study.* Biophys J, 2000. **78**(3): p. 1498-518.

34. Gebe, J.A., et al., *Effects of Na+ and Mg2+ on the structures of supercoiled DNAs: comparison of simulations with experiments.* J Mol Biol, 1996. **262**(2): p. 105-28.

35. Naimushin, A.N., et al., *Effect of ethidium binding and superhelix density on the apparent supercoiling free energy and torsion constant of pBR322 DNA.* Biophys Chem, 1994. **52**(3): p. 219-26.

36. Taylor, W.H. and P.J. Hagerman, *Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure. II. NaCl-dependence of DNA flexibility and helical repeat.* J Mol Biol, 1990. **212**(2): p. 363-76.

37. Shimada, J.Y., H., *Ring closure probabilities for twisted wormlike chains. Application to DNA.* Macromolecules, 1984(17): p. 689-698.

38. Shore, D. and R.L. Baldwin, *Energetics of DNA twisting. I. Relation between twist and cyclization probability.* J Mol Biol, 1983. **170**(4): p. 957-81.

39. Clendenning, J.B. and J.M. Schurr, *Circularization of small DNAs in the presence of ethidium: a theoretical analysis.* Biopolymers, 1994. **34**(7): p. 849-68.

40. Shore, D. and R.L. Baldwin, *Energetics of DNA twisting. II. Topoisomer analysis.* J Mol Biol, 1983. **170**(4): p. 983-1007.

41. Frank-Kamenetskii, M.D., et al., *Torsional and bending rigidity of the double helix from data on small DNA rings.* J Biomol Struct Dyn, 1985. **2**(5): p. 1005-12.

42. Horowitz, D.S. and J.C. Wang, *Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling.* J Mol Biol, 1984. **173**(1): p. 75-91.

43. Shimada, J. and H. Yamakawa, *Statistical mechanics of DNA topoisomers. The helical worm-like chain.* J Mol Biol, 1985. **184**(2): p. 319-29.

44. Kahn, J.D. and D.M. Crothers, *Measurement of the DNA bend angle induced by the catabolite activator protein using Monte Carlo simulation of cyclization kinetics.* J Mol Biol, 1998. **276**(1): p. 287-309.

45. Kahn, J.D., E. Yun, and D.M. Crothers, *Detection of localized DNA flexibility.* Nature, 1994. **368**(6467): p. 163-6.

46. Roychoudhury, M., et al., *Global structure and mechanical properties of a 10-bp nucleosome positioning motif.* Proc Natl Acad Sci U S A, 2000. **97**(25): p. 13608-13.

47. Vologodskaia, M. and A. Vologodskii, *Contribution of the intrinsic curvature to measured DNA persistence length.* J Mol Biol, 2002. **317**(2): p. 205-13.

48. Zhang, Y. and D.M. Crothers, *High-throughput approach for detection of DNA bending and flexibility based on cyclization.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3161-6.

49. Zhang, Y. and D.M. Crothers, *Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization.* Biophys J, 2003. **84**(1): p. 136-53.

50. Moroz, J.D. and P. Nelson, *Torsional directed walks, entropic elasticity, and DNA twist stiffness.* Proc Natl Acad Sci U S A, 1997. **94**(26): p. 14418-22.

51. Strick, T., et al., *Twisting and stretching single DNA molecules.* Prog Biophys Mol Biol, 2000. **74**(1-2): p. 115-40.

52. Strick, T.R., et al., *The elasticity of a single supercoiled DNA molecule.* Science, 1996. **271**(5257): p. 1835-7.

53. Strick, T.R., et al., *Behavior of supercoiled DNA.* Biophys J, 1998. **74**(4): p. 2016-28.

54. Bouchiat, C.M., M., *Elasticity theory of a supercoiled DNA molecule.* Phys. Rev. Lett., 1998. **80**: p. 1556-1559.

55. Bryant, Z., et al., *Structural transitions and elasticity from torque measurements on DNA.* Nature, 2003. **424**(6946): p. 338-41.

Table 2.1 Physical properties used to monitor changes of DNA secondary structure *in solution*

- Helix Repeat
- CD Spectrum
- Gel Electrophoretic Mobility Compared to Normal Standards
- Torsional Rigidity as Monitored by Time-Resolved Fluorescence Polarization Anisitropy of Intercalated Ethidium
- Intrinsic Binding Constant and Cooperativity Parameter for Ethidium and Daunomycin Binding
- Bending Rigidity as Monitored by Dynamic Light Scattering
- Transient Electric Birefringence
- Transient Photodichroism of Bound Methylene Blue
- Heat Capacity and Enthalpy
- Optical Melting Profile
- Susceptibility to Enzymatic Attack
- Resonance Raman Spectrum

Table 2.2 Measured torsional rigidities ($C = h\alpha$) for DNAs under bending or tensile strain obtained by 4 different methods.

DNA	Bend (deg/bp)	Tension (pN)	Method C(fJ fm)	Reference(s)
linear viral DNAs	0	0	FPA 150-170	[29]
linearized plasmids	0	0	FPA 190-220	[23, 29-33]
linear 181 bp DNAs	0	0	FPA 220	[32]
circular plasmids	≤ 0.4	0	FPA 200-230	[19, 23, 24, 29-32, 34, 35]
circular (340-350 bp)	1.0	0	CK 200	[36]
circular (237-254 bp)	1.45	0	CK 240-300	[37, 38]
circular (247 bp)	1.45	0	TR 410-420	[32, 39, 40]
circular (205-217 bp)	1.7	0	TR 320-330	[32, 41-43]
circular (150-200 bp)	1.8-2.4	0	CK 220-340	[44-49]
circular (181 bp) (fresh)	2.0	0	FPA 310-330	[32]
circular (181 bp) (8 mos.)	2.0	0	FPA 400	[32]
linear	none	0.1-2.0	SMP 300	[50-53]
linear	none	0.1-5.0	SMP 350	[51, 54]
linear	none	0.3-8.0	SMP 450	[50-53]
linear	none	15-45	SMP 410-440	[55]

FPA: Fluorescence Polarization Anisotropy
CK: Cyclization Kinetics
TR: Topoisomer Ratio
SMP: Single Molecule Pulling Experiments

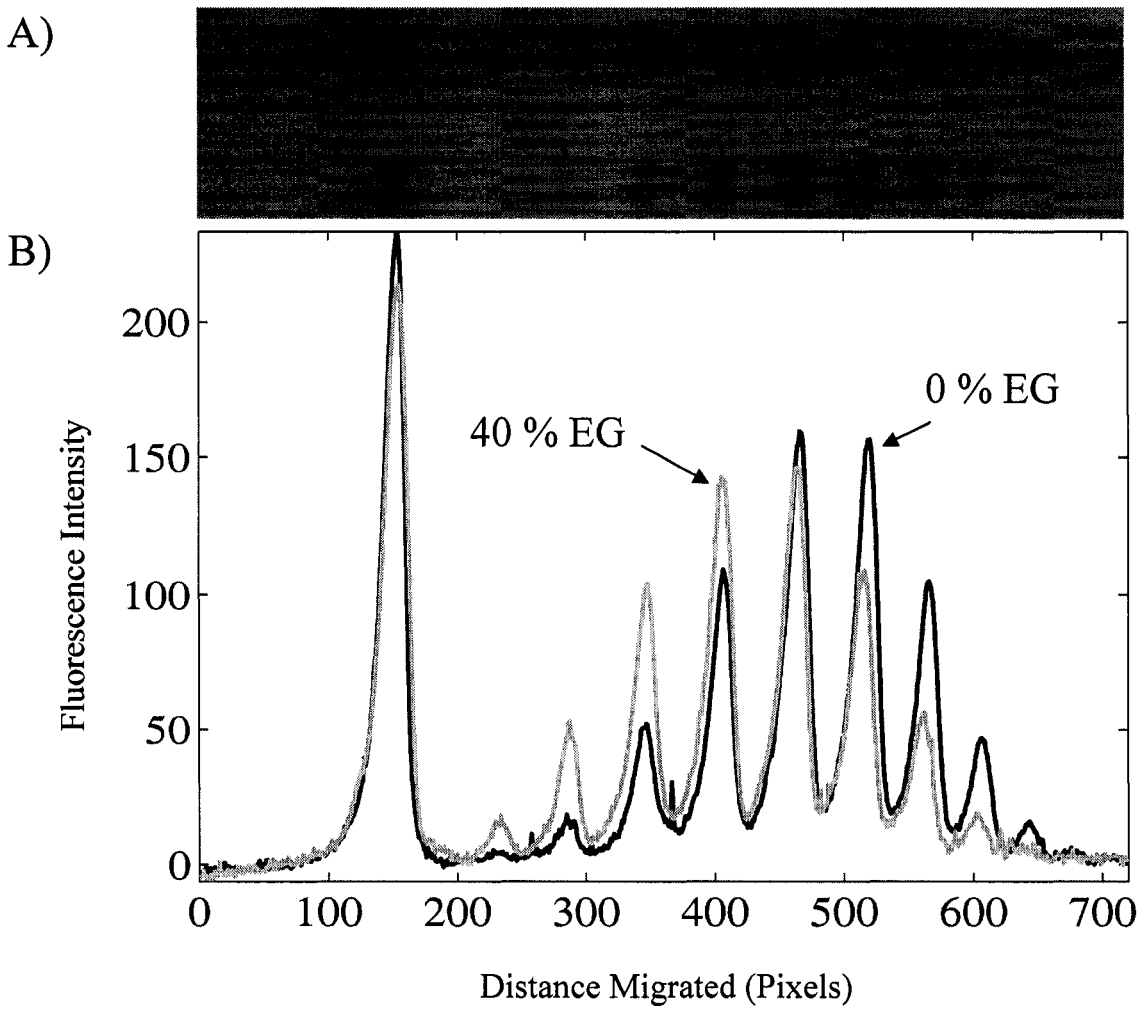


Figure 2.1 A) 0.8% Agarose gel showing DNA relaxed using topoisomerase I. The DNA on bottom is from a reaction with 0% added osmolyte, and the DNA on top is from a reaction with 40% EG. B) Lane traces showing the band intensity for each of the individual topoisomer reactions. The 0% reaction is shown solid black while the 40% reaction is a gray line. The band at 150 pixels (both reactions) is nicked DNA. The distribution describing the DNA from the 40% reaction has been shifted back nearly **one** turn.

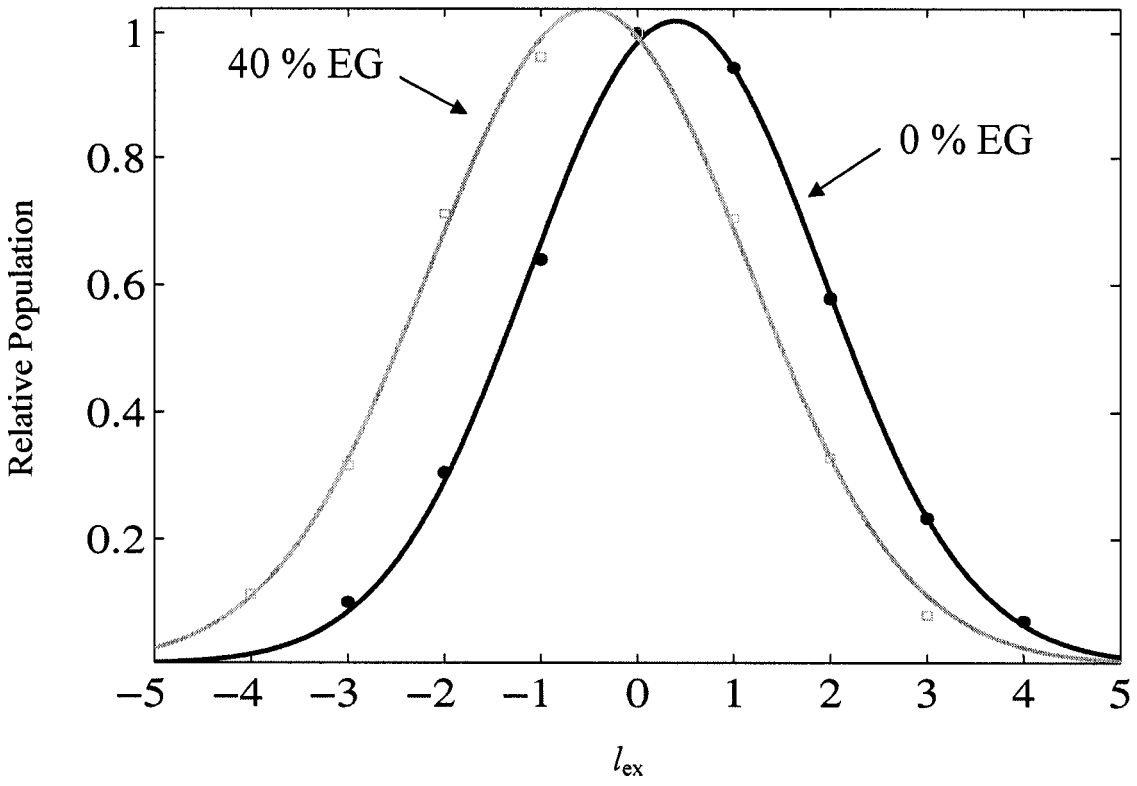


Figure 2.2 In any individual reaction, the areas of each of the topoisomer peak are normalized to the most intense band and plotted versus the topoisomer number. The data are fitted with the gaussian curve in Equation 2.16. The data from a representative 0% reaction is shown with black circles and its corresponding fit with a solid black line while the data from a 40% EG reaction is represented by the open squares and a solid gray line.

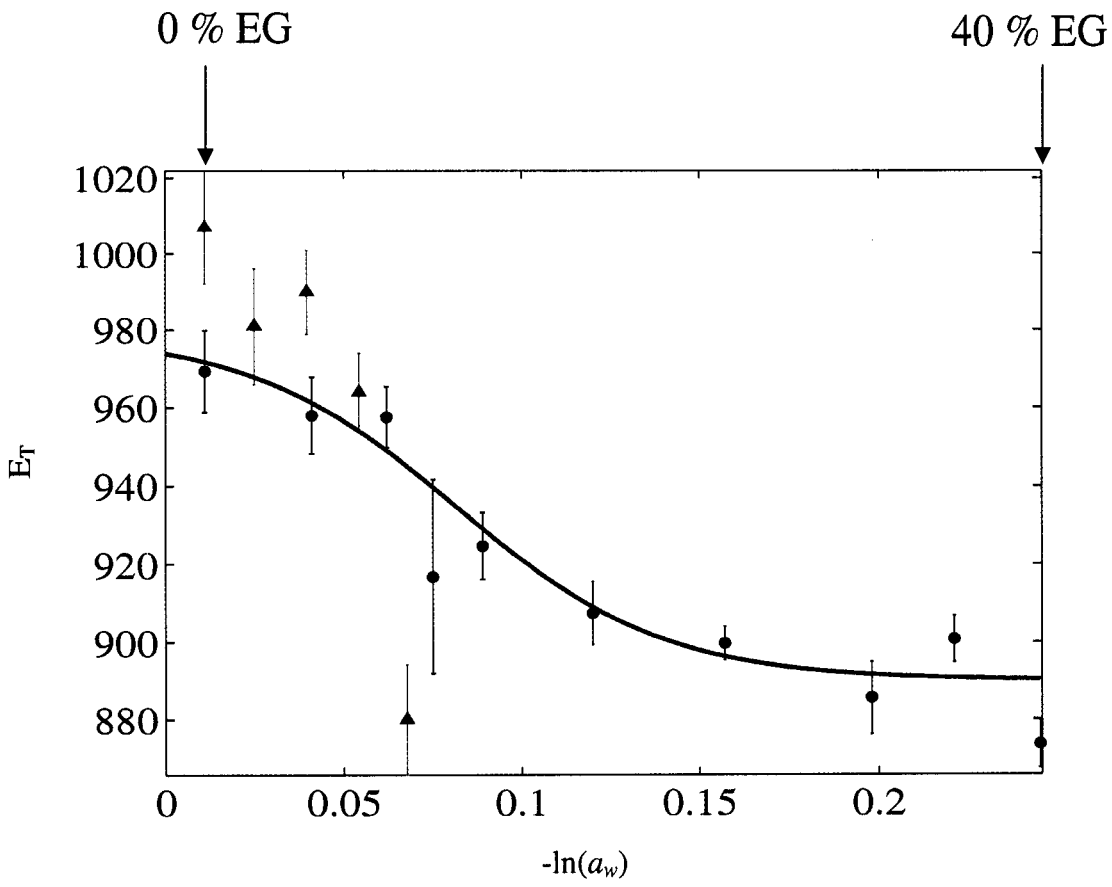


Figure 2.3 Plot of E_T versus $-\ln(a_w)$. The black circles (with errorbars) are the data from the most recent topoisomerase reactions and are believed to show the characteristic sigmoidal shape that result from a two state model for the DNA. The data shown in red are the previous data from Rangel et al. (2004) that have been extended out to 40% EG in these experiments. The solid black curve is a fit of Equation 2.14 to the black data.

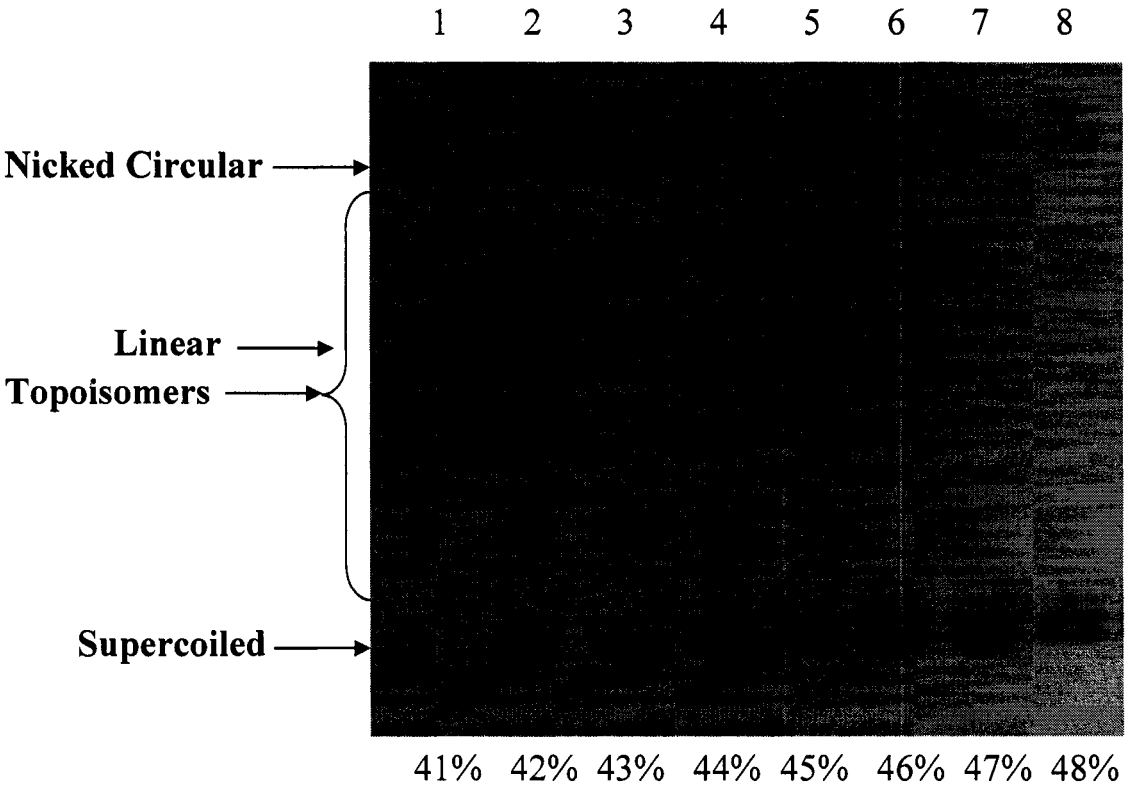


Figure 2.4 At concentrations above 40% EG, Topo I begins to lose its activity and becomes inactive at 47% EG. Lanes 1-8 (numbers above each lane) show the results of reactions run with 41–48% (numbers below each lane) ethylene glycol, respectively. The different types of DNA present on the gel are labeled in the left margin.

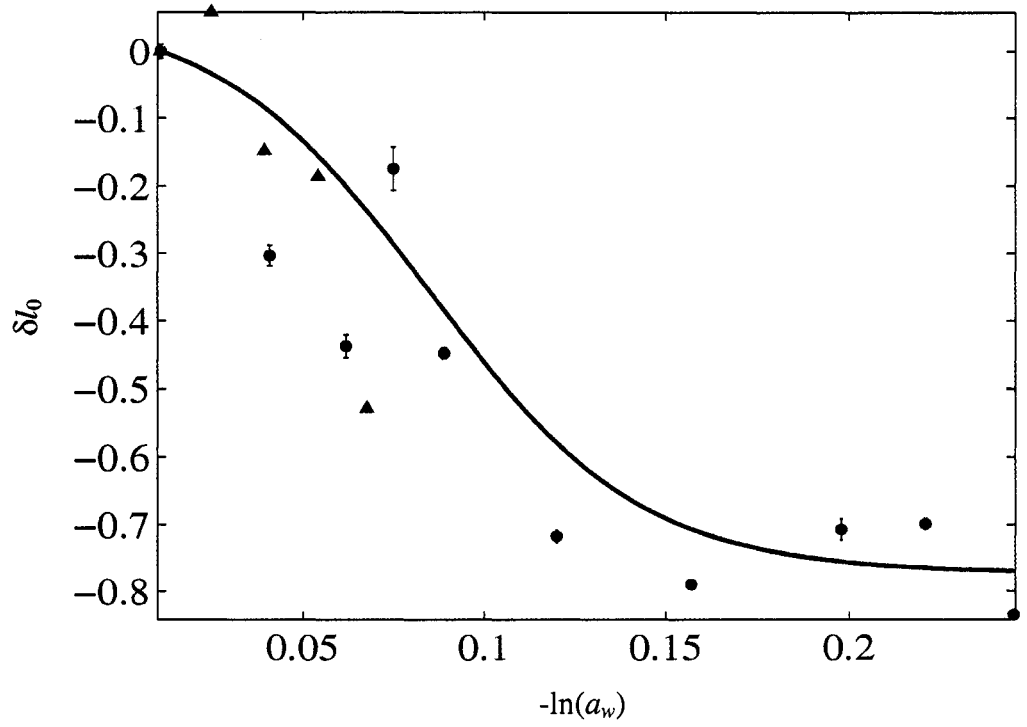


Figure 2.5 Plot of δI_0 versus $-\ln(a_w)$. The parameters n and K_0 (obtained from the fit shown in Figure 2.3) were inserted into Equation 2.15 and resulted in a curve (solid black line) that ran through the middle of the data. The data from the most recent preparation of p30δ DNA is shown with black circles while the data from a previous preparation are shown in red.

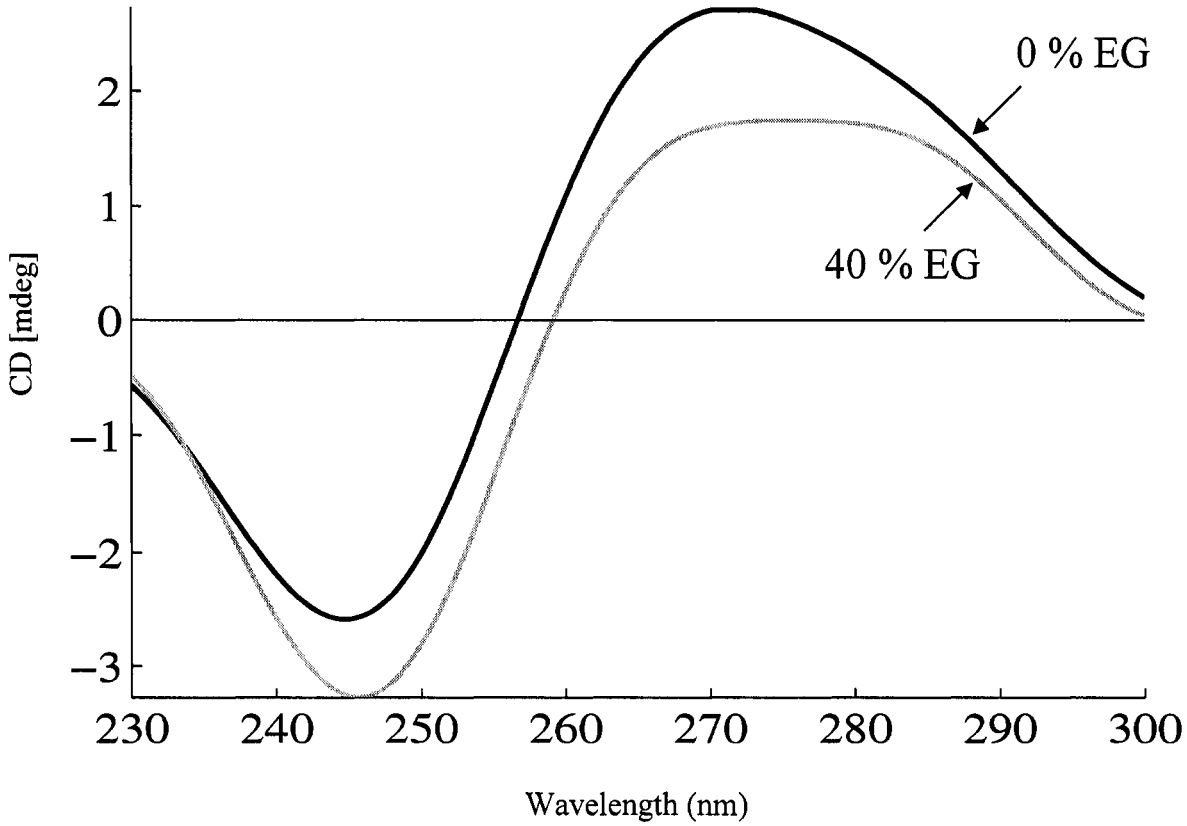


Figure 2.6 CD measurements taken on linear p30δ in 0% (solid black curve) and 40% EG (solid gray curve). The CD spectrum has clearly shifted indicating a change in the secondary structure of the DNA.

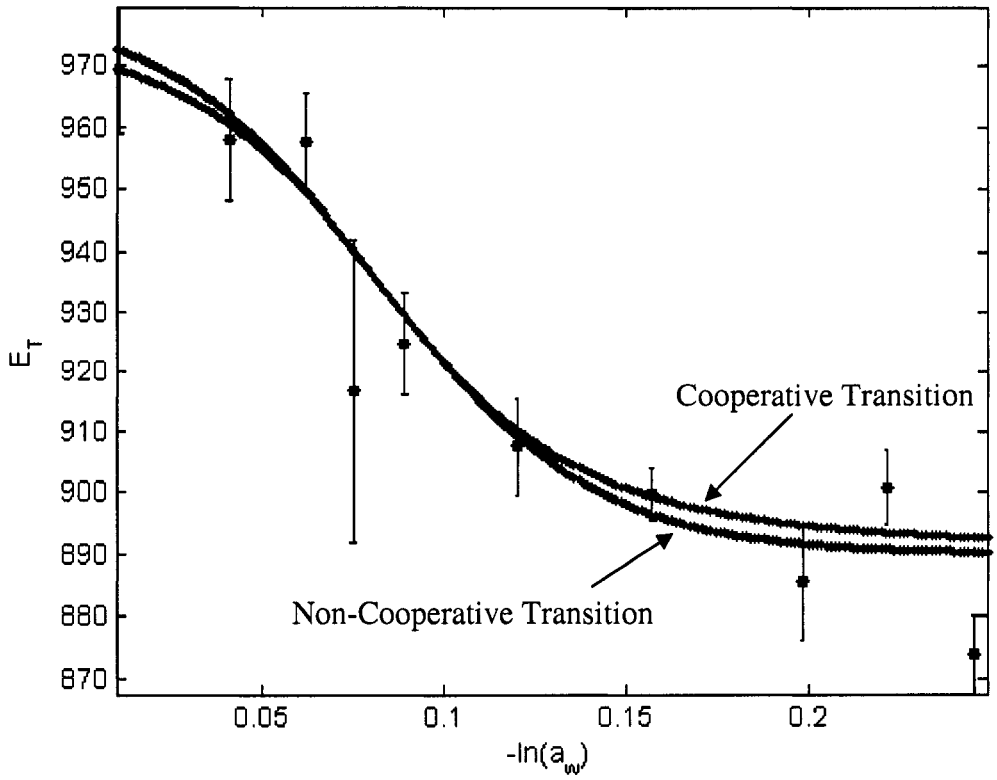


Figure 2.7 E_T vs. $-\ln(a_w)$; Cooperative vs. Non-Cooperative Transitions. Models were derived wherein a two-state transition can occur in the DNA by either a cooperative or non-cooperative mechanism (see Appendix D). A fit of this theory to the data requires that the ratio p/J = 34. The blue line is a fit from a cooperative model with p = 1, J = 1/34. The black line is a fit from the non-cooperative model with p = 34, J = 1. Near the midpoint of the transition the curves will coincide independently of the model as long as the ratio p/J held constant at 34.

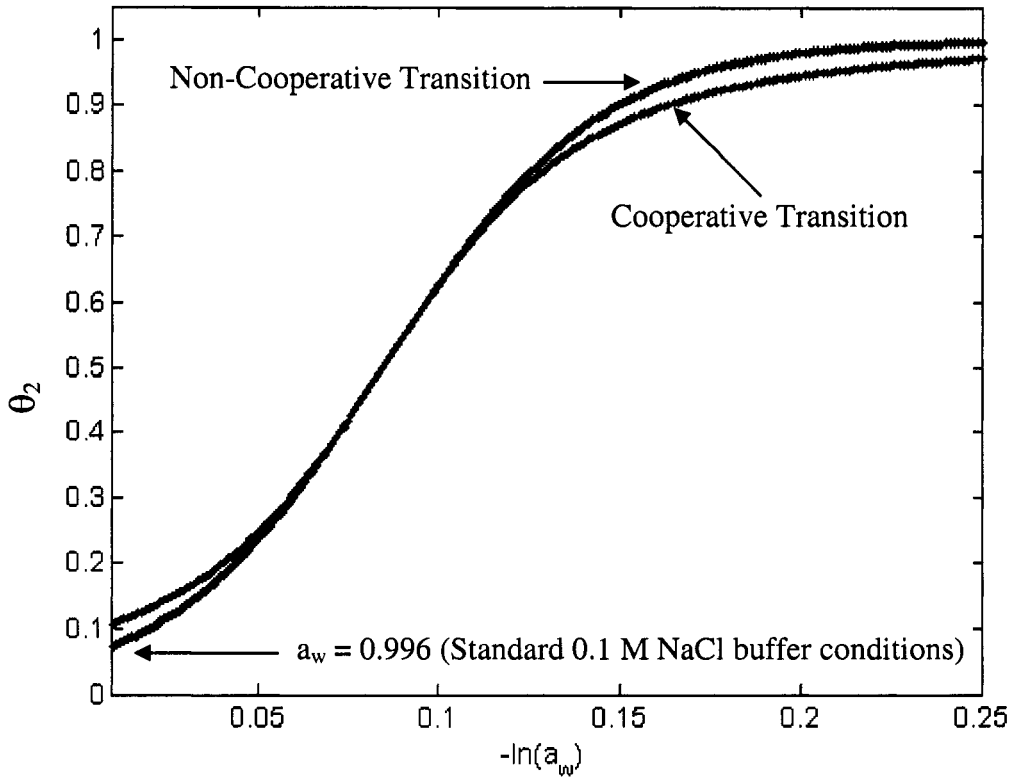


Figure 2.8 θ_2 vs. $-\ln(a_w)$; cooperative vs. non-cooperative Transitions. From the two-state model, the fraction of subunits in state 2 can be found for the transition as a function of $-\ln(a_w)$. The blue line is for the cooperative transition with p = 1, J = 1/34. The black line is for the non-cooperative transition and has p = 34, J = 1. Both curves coincide near the midpoint of the transition. In standard buffer conditions, ~ 5 – 10% of the DNA are predicted to reside in state 2.

Chapter 3

Relaxation of Meta-Stable Secondary Structure by Calf Thymus Topoisomerase I

3.1 Introduction

For three different methods of relaxing superhelical stress, namely linearization by a restriction enzyme [1], binding of *E. coli* single strand binding protein (SSB) [2], or the binding and unwinding by intercalating dyes [3], it was found that low values of the torsion elastic constant (α) resulted upon release of the super-helical strain and prevailed for times as long as two or more weeks. For DNAs linearized by restriction enzymes (Figure 3.1), α eventually evolved upward to values comparable to, but slightly lower than, those measured for the native plasmids. The molar ellipticity at 273 nm of p30δ DNA in 0.01 M ionic strength underwent a similar decrease upon linearization followed by upward evolution on a similar time-scale (unpublished data reported by Song et al. [1]). However, in the case of bound intercalating dyes, the upward evolution at ~ 4 − 8 weeks was *not* observed, and the DNA instead existed in one or more metastable states that seemed more frozen in. These observations imply the existence of metastable secondary structure(s) in freshly relaxed supercoiled DNAs.

Metastable secondary structure refers to the conformation(s) of DNA sequences that occupy an 'excited' state and belong to a distinctly different free energy basin than that which has the lowest free energy (see Figure 3.2). The minima of the two free energy basins occur at different values of one of the (non-integrated) coordinates in a potential of mean force. Such a DNA molecule can undergo a structural transition according to a two-state model. Excited subunits of the DNA will typically *not* occupy the higher energy state indefinitely but will equilibrate back to the 'ground' state. This process is expected to occur extremely slowly for *highly cooperative* structural transitions. The observed time scale of two weeks or more for equilibration subsequent to relieving the superhelical stress suggests that the transition is highly cooperative. This will be expanded upon below. Ample evidence suggests that DNA meets the qualifications necessary to exhibit metastable states [1-13]. Shibata et al. [9] laid the

foundation for the possibility of metastability in DNA by noticing that different preparations (four from different replicates of the same standard preparation procedure, one from an alternate preparation, and two from irreversible conversions of such forms to daughter products) of the plamid M13mp7 resulted in seven distinct long-lived structural conformers. These seven conformers exhibited either of two different torsion elastic constants, depending upon the buffer type (Tris vs. Citrate) and had one of three different gel mobilities. In that work, they proposed a specific model for remote control of gene expression wherein binding of a site-specific regulatory protein, which typically binds several tens or more base pairs away from the specific polymerase binding site, can activate transcription through a secondary structure transmission effect (see Figure 3.3). But for this to occur, the DNA must meet three requirements: 1) the DNA should exhibit two or more distinct structural states that provide a molecular basis for transmitting information through the intervening n base pairs between the regulatory protein and the polymerase, 2) the free energies per subunit of these two secondary structures must be nearly equal and, 3) the junction free energies must be positive and large to ensure propagation of a given secondary structure over a distance of at least n, $n >> 1$, base pairs beyond the binding site. If these three criteria are met, they suffice to induce metastability and hysteresis in any system.

The primary function of a topoisomerase is to relax superhelical stress. While there are many classifications of topoisomerase enzymes, they typically fall into one of two categories that are either type I or type II [14, 15]. Topoisomerases that are type I transiently break a single strand of the DNA duplex backbone and allow one strand of the DNA helix to pass around the other resulting in a net unwinding of the DNA helix. Prior to departure of the enzyme from the plasmid, the nicked DNA is resealed by the enzyme. Type II enzymes break the pair of strands simultaneously to allow unwinding of the DNA helix to occur before the DNA strands are re-ligated. Calf Thymus Topoisomerase enzymes, like those employed in these experiments, are known to be type I. There are also DNA gyrases that are type II topoisomerases. They too act to change the superhelix density but their primary action is to introduce negative supercoils into the plasmid in a

reaction that requires dATP and Mg^{2+} ions. For DNA that is negatively supercoiled the function of a topoisomerase is counteractive to that of a gyrase [16].

For the transition observed in Figure 2.3, the largest source of error in the mean of the means of several data sets for the same conditions was lack of reproducibility between different data sets. The σ_{E_T} shown (as error bars in Figure 2.3) were calculated according to $\sigma^2_{\langle \bar{E}_T \rangle} = \sum_{j=1}^{m} \left(\bar{E}_{T_j} - \langle \bar{E}_T \rangle \right)$, where $\bar{E}_{T_j}$ is the mean of the jth data set (~ 12 gel lanes) and $\langle \bar{E}_T \rangle = \left(\frac{1}{m} \right) \sum_j \bar{E}_{T_j}$ is the mean of the means. The variance of each single data set is calculated according to $\sigma_{\bar{E}_T} = \sum_{i=1}^{12} \frac{\left(E_{T_i} - \bar{E}_T \right)^2}{12}$. We find that $\sigma_{\langle \bar{E}_T \rangle}$ substantially exceeds the $\sigma_{\bar{E}_{T_j}}$ for all j = 1...m. The statistical errors within any single data set for a successful reaction were small compared to the variations between data sets for the same reaction. This indicates a source of variability in the reactions that requires some explanation. The hypothesis proposed to account for this observation is that topoisomerase I has two functions. Its main function is to nick and reseal the 'equilibrated' DNA as is widely known. The second function could be to bind to relaxed DNA, still nicking and resealing as before, but in doing so, it could act to relax any metastable secondary structure on a much shorter time scale than the previously observed timescale of weeks to months [1, 8]. This action could significantly affect the observed E_T values and indeed has manifested itself through a series of time-resolved measurements of E_T. The following questions will be addressed in this chapter:

Q1: Are metastable states found subsequent to relaxation with topoisomerase I?

Q2: Are there any perturbations that accelerate equilibration of the metastable secondary structure?

3.2 Materials and Methods

The plasmid p30δ (4932 bp) was isolated from *E. coli* HB101 host cells and separated from nicked circles and chromosomal DNA by equilibrium banding in a density gradient according to the protocol of Langowski et al. [2]. The plasmid was dialyzed into a storage buffer containing 50 mM KCl, 50 mM Tris, 0.1 mM EDTA, pH = 7.5 and kept at 4°C. The A_{260}/A_{280} ratio was ~ 1.9 and analysis of the band intensities following gel electrophoresis indicated a supercoiled:nicked ratio greater than 4:1. p30δ is a derivative of pBR322, constructed by inserting an 842 bp fragment containing the Bgl II to Sal I fragment of *S. cerevisae* into the region between the BamH I and Sal I sites of pBR322 [17]. This plasmid exhibits less evidence of long-lived metastable secondary structures and irreproducible behavior than does the parent pBR322 [3, 18].

3.2.1 Time-Resolved Measurements of E_T

Time resolved experiments using native supercoiled p30δ relaxed by CT Topo I were performed using the topoisomer distribution method by stopping reactions at 10 minute intervals with a phenol-chloroform-isoamyl alcohol mixture and then separating the topoisomers using gel electrophoresis. Gels were run in a medium containing sufficient chloroquine that all visible topoisomers were slightly positively supercoiled. After removing the chloroquine upon completion of electrophoresis, gels were stained with ethidium bromide (see Figure 2.1A) under conditions, wherein the amount of dye (ethidium bromide) bound to each topoisomer is proportional to its relative concentration. The fluorescence in the gel was imaged and quantitated with a Fluorimager SI gel scanner and the integrated intensity of each band was reckoned using commercial software (ImageQuant 5.1) by taking longitudinal sections in the direction of migration down the centers of representative lanes (see Figure 2.1B). Fits were made to the data using equation (2.14) and gave values for E_T and δl_o. A typical best-fit 'curve' can be found in Figure 2.2 for a single lane co-plotted with its corresponding experimental values.

The activity of Topo I was tested *first* in a set of experiments, where fresh Topo I was added to the DNA sample with 0% added EG, at t = 0, 120, and 240 minutes, and data were collected at various times from t = 10 to 360 minutes. In a *second* test, the DNA sample received a single addition of Topo I at t = 0, and data were collected from 10 to 180 minutes. In a *third* test, the DNA sample received the first addition of Topo I at t = 0. At t = 60 minutes the sample was divided into two aliquots, one of which received a second addition of fresh Topo I, and the other of which received the same volume of buffer without the Topo I. Data were then collected from 10 to 120 minutes. In *all* of these experiments in 0% EG, anomalously low E_T-values were observed at the earliest times after the first addition of Topo I, and a second addition was *required* in order for E_T to increase up toward its equilibrium plateau value. In a *fourth* test, the DNA samples contained 20% EG. Two samples received Topo I at t = 0, but only one received a second addition at t = 120 minutes, and data were collected for both samples from 20 to 240 minutes. In a *fifth* test, the DNA sample contained 40% EG, and Topo I was added at t = 0, 120, and 240 minutes and data were collected for 10 to 360 minutes.

3.3 Results

3.3.1 Are metastable states found subsequent to relaxation with topoisomerase I in 0% EG?

The action of Topo I was investigated over time intervals from 10 minutes to three hours (Figure 3.4). At 0% EG, E_T starts out with low values (~ 400 – 500) that are distinct and approximately two-fold smaller than the higher 'equilibrium' values (~ 1000 – 1100) that are observed as the reactions proceed, and more additions of the enzyme are made. For all reaction times, the distribution of relaxed topoisomers exhibited a symmetrical 'equilibrium' shape and the position of the most populous topoisomer in the gel did not vary with reaction time or amount of enzyme added, even for the earliest times analyzed.

Essentially no topoisomers appeared between the initial supercoiled band and the symmetrical distribution of relaxed topoisomers (Figure 3.5). These findings suggest that the enzyme acts processively to relax supercoiled DNAs from their initial native

supercoiled state to their equilibrium distribution of relaxed topoisomers via single 'encounters' with each DNA. Thus, in 0% EG, the Topo I acts processively.

Despite the absence of significant superhelical strain in the relaxed topoisomers, the initial E_T-value was initially anomalously low, but rose after the second addition of active enzyme at ~ 120 minutes to reach its equilibrium plateau value at ~ 150 minutes, and was not further affected by the third addition of Topo I at 240 minutes (Figure 3.4). When only one addition of Topo I is made at t = 0, not all of the initial native supercoiled DNA is converted to relaxed topoisomers, which indicates that the enzyme loses its activity within ~ 120 minutes after addition (data not shown). In this case of a single addition of Topo I at t = 0, E_T remained low for the entire 180 minutes examined (Figure 3.6). This indicates that additional active enzyme is required in order to cause the increase in E_T toward its equilibrium plateau value. The different temporal trajectories of E_T for two different aliquots of the same DNA sample, one with and one without a second addition of Topo I at t = 60 minutes are presented in Figure 3.7. For the sample with the additional Topo I, E_T rose toward the equilibrium value, whereas for the sample without additional Topo I, E_T remained low, as expected from the results in Figure 3.6.

From the evidence cited above we conclude that (i) metastable secondary structure with a ~ 0.5-fold too low value of E_T exists subsequent to relaxation of the native supercoiled DNA by Topo I, and fresh Topo I enzyme can bind to a previously relaxed DNA and relax some of the metastable secondary structure.

3.3.2 Are there any perturbations that accelerate equilibration of the metastable secondary structure?

In Figure 3.8, the time-resolved behavior of E_T for the same DNA in 0% (black), 20% (blue) and 40% (gray) EG is shown. As stated previously, in 0% EG the enzyme acts processively to relax the superhelical strain of the circular DNA in a single encounter. The circles and solid black curve are taken from Figure 3.4. The dotted line extending from t = 120 to t = 180 minutes represents the data that were collected for a set of reactions where only 1 addition of enzyme was made (Figure 3.6).

In 20% EG, intermediate topoisomers were observed between the native supercoiled band and the distribution of relaxed topoisomers (data not shown), so the

processivity is clearly not as great as in 0% EG. Because the amplitudes of the most relaxed topoisomers significantly exceed those of the intermediate topoisomers, we infer that the Topo I still acts somewhat processively in 20% EG. The earliest measured E_T's fall near those values found for the equilibrium DNA in 40% EG. However, based on the trajectory of the data, it is proposed that the very earliest E_T's (not measured) would lie between those of the 0 and 40% EG values (~ 410 and 710, respectively) and evolve in time such that the 'equilibrium' values ($t \rightarrow 360$ minutes) would lie between 0 and 40% EG. Topoisomerization reactions in 20% EG are indifferent to the amount of enzyme added. Figure 3.9 shows the results of two experiments run at 20% EG. One reaction had only 1 addition of enzyme at $t = 0$ and the other had 2 additions of enzyme at $t = 0$ and $t = 120$ minutes. Data were collected for 240 minutes, or 4 hours. The differences between the two curves are within the error of the measurement itself. They are shown to nearly overlay the projected curve from the data shown in Figure 3.8 for 20% EG. We infer that equilibration of secondary structure is aided by the addition of 20% EG.

In 40% EG, a larger amplitude of intermediate topoisomers was observed between the supercoiled band and the distribution of relaxed topoisomers (data not shown), so the processivity, if any, is now much lower. The E_T values of the DNA in 40% EG start out around ~ 750 – 800 at the earliest measured times and show no sign of changing over the entire 10 – 360 minute time span. The presence of 40% EG has evidently pushed the DNA into an alternate equilibrium state with little or no help from the Topo I enzyme.

The initially low value of E_T increases somewhat as progressively more ethylene glycol is added into the reactions. By 40% EG, the DNA has been equilibrated with little or no help from the topoisomerase enzyme. Finally, the differences between the long-time equilibrium values of E_T in these 0, 20 and 40% EG solutions are similar to those in the sigmoidal transition of Figure 2.3.

3.4 Discussion

The time-resolved E_T measurements presented in this chapter provide an important link to those values previously reported in the literature, wherein a two-fold discrepancy in E_T was found to exist between those values found using the ligation

method ($E_T \sim 950 - 1150$, see Table 3.1) and those found using the dye-binding method ($E_T < 550$, see Table 3.2). There are differences in the techniques that might account for this apparent discrepancy. In the topoisomer distribution method, states accessible by thermal fluctuation from the relaxed condition are generated which means that this method *in principle* samples states that are near the relaxed normal B conformation. The dye-binding method samples a much wider range of higher superhelix densities, where allosteric transitions in secondary structure have likely occurred. Song et al. [1] found that those quantities sensitive to the secondary structure, namely the torsion constant (α), plateau diffusion coefficient (D_{plat}), and molar ellipticity ([θ]) of the pUC8 dimer exhibited a dip to rather low values at $\sigma = -0.025$ and had a subsequent rise to more normal values at greater negative superhelix densities. The torsion constant and molar ellipticities for this sample also evolved over time (~ 50 days) toward more normal, but still somewhat low, values. The reactions reported here for Topo I relax p30δ through this region of the superhelical density and so, in fact, provide a clue as to the events taking place. In 0% EG, low values of E_T ($\sim 400 - 500$) are obtained for $t < 120$ minutes while high values ($\sim 1000 - 1100$) result upon equilibration for longer times. This evolution of E_T, coupled to the fact that the enzyme acts processively to relax the DNA in a single encounter (the gel position of the most intense band does not change throughout the reaction) indicates that the enzyme is acting to equilibrate metastable secondary structure. This may be a biologically significant second function of the Calf-Thymus Topo I enzyme.

3.5 Conclusion

Calf Thymus Topo I relaxes superhelical strain *processively* in aqueous buffer. It is found here that Calf Thymus (CT) Topo I binds to pre-relaxed DNA and also is able to relax metastable secondary structure that persists in the relaxed DNA in 0 and 20% ethylene glycol, which appears to be a new, second function for Calf Thymus Topo I. CT Topo I is able to function up to 43 w/v % ethylene glycol and is inactive by 47%. Processivity is found to decline with increasing w/v % ethylene glycol.

In 0 w/v % ethylene glycol, metastable secondary structure with an anomalously low value of $E_T \sim 400 - 500$ prevails immediately after relaxation of the superhelical strain. This is consistent with previous observations of metastable secondary structures that exhibit anomalously low torsional rigidities immediately after relaxation by other means. Calf Thymus Topo I acts to equilibrate such metastable secondary structure prevailing immediately after release of the superhelical strain. The value of E_T observed immediately after topoisomerization increases with w/v % ethylene glycol, and sufficient ethylene glycol is able to equilibrate the metastable secondary structure with little or no help from Topo I.

3.6 Notes to Chapter 3

1. Song, L., et al., *Evidence for allosteric transitions in secondary structure induced by superhelical stress.* J Mol Biol, 1990. **214**(1): p. 307-26.

2. Langowski, J., et al., *Change of conformation and internal dynamics of supercoiled DNA upon binding of Escherichia coli single-strand binding protein.* Biochemistry, 1985. **24**(15): p. 4022-8.

3. Clendenning, J.B., et al., *Effect of ethidium binding and superhelix density on the supercoiling free energy and torsion and bending constants of p30 delta DNA.* Biophys Chem, 1994. **52**(3): p. 191-218.

4. Di Mauro, E., *Topological Evidence for Allosteric Transitions in DNA Secondary Structure (Personal Communication).*

5. Heath, P.J., et al., *Effect of bending strain on the torsion elastic constant of DNA.* J Mol Biol, 1996. **260**(5): p. 718-30.

6. Kim, U.S., et al., *Dynamics and structures of DNA: long-range effects of a 16 base-pair (CG)8 sequence on secondary structure.* Biopolymers, 1993. **33**(11): p. 1725-45.

7. Naimushin, A.N., et al., *Effect of polyethylene glycol on the supercoiling free energy of DNA.* Biopolymers, 2001. **58**(2): p. 204-17.

8. Schurr, J.M., et al., *The question of long-range allosteric transitions in DNA.* Biopolymers, 1997. **44**(3): p. 283-308.

9. Shibata, J.H., et al., *Structures and dynamics of a supercoiled DNA.* Biochemistry, 1984. **23**(6): p. 1188-94.

10. Wu, P.G., et al., *Effect of ethidium on the torsion constants of linear and supercoiled DNAs.* Biophys Chem, 1991. **41**(3): p. 217-36.

11. Rangel, D.P., *Effects of neutral osmolytes on DNA, Ph.D. Thesis.* 2005, University of Washington.

12. Díaz, R.P., *Fast and slow internal dynamics of 1^3C labeled DNA oligomers in solution.* 2002, University of Washington.

13. Wu, P.G., et al., *Interaction of chloroquine with linear and supercoiled DNAs. Effect on the torsional dynamics, rigidity, and twist energy parameter.* Biochemistry, 1988. **27**(21): p. 8128-44.

14. Wang, J.C., *Cellular roles of DNA topoisomerases: a molecular perspective.* Nat Rev Mol Cell Biol, 2002. **3**(6): p. 430-40.

15. Wang, J.C., *DNA topoisomerases.* Annu Rev Biochem, 1985. **54**: p. 665-97.

16. Bauer, W.R., *Structure and reactions of closed duplex DNA.* Annu Rev Biophys Bioeng, 1978. **7**: p. 287-313.

17. Di Mauro, E., et al., *Activation of in vitro transcription and topology of closed DNA domains.* J Biol Chem, 1985. **260**(1): p. 152-9.

18. Naimushin, A.N., et al., *Effect of ethidium binding and superhelix density on the apparent supercoiling free energy and torsion constant of pBR322 DNA.* Biophys Chem, 1994. **52**(3): p. 219-26.

Table 3.1 E_T values for numerous different plasmids determined using the Ligation Method; as reported in [13]. The reported data fall in the range of 1000 ± 115.

DNA	[salt]	E_T
PM2	10	1000
fd	10	920
SV40	10	1110
PM2	200	985
ColE1	200	1115
Minicol	200	1065

Table 3.2 E_T values for numerous different plasmids determined using the Dye-Binding Method; as reported in [13]. With the exception of the value of 1246, all E_T values are less than 550.

DNA	$-\sigma$	[salt]	E_T
PM2[a]	0.110	3M CsCl	537
λ2b5c[a]	0.051	3M CsCl	527
SV40[a]	0.065	5.8 M CsCl	495
λ[a]	0.072	0.1 M NaCl	1246
λ[a]	0.078	0.4 M NaCl	533
λ[a]	0.090	1.0 M NaCl	507
pBR322[a]	0.048	0.1 M NaCl	280±70
pBR322[a]	0.083	0.1 M NaCl	347±50
pBR322[b]	0.048	0.1 M NaCl	360
pBR322[b]	0.083	0.1 M NaCl	460

[a] Experiments with Ethidium Bromide
[b] Experiments with Chloroquine

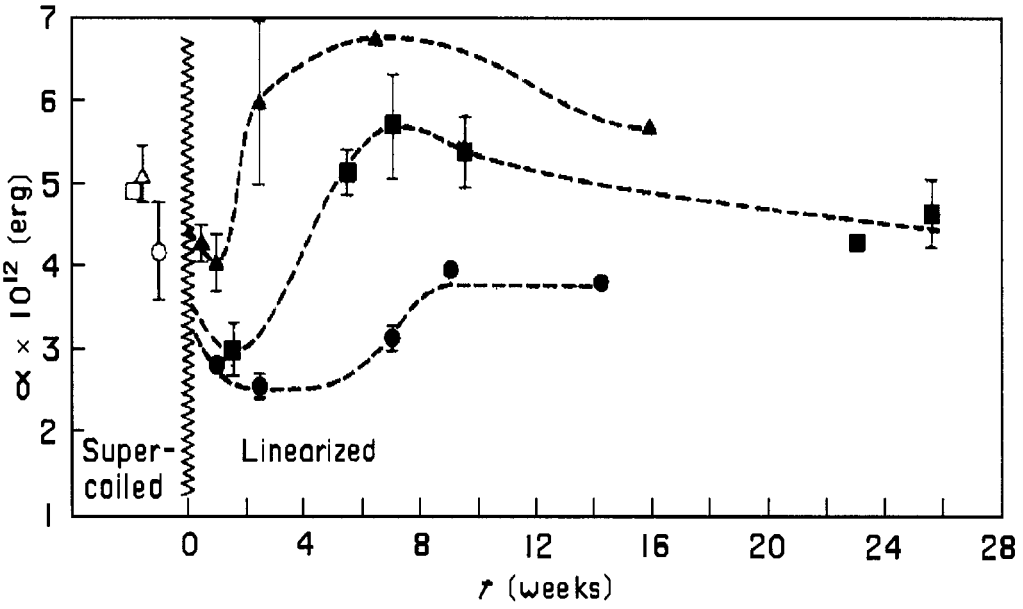


Figure 3.1 Torsion Constant (α) as a function of time after linearization for 3 supercoiled DNAs. The open and filled symbols are, respectively, the values of α before and after linearization.
(○) M13mp7 in 2 mM-Tris, 10 mM-NaCl, 0.2 mM EDTA (pH=7.0)
(Δ) pBR322 in 2 mM-Tris·HCl, 10 mM-NaCl, 2 mM EDTA (pH=7.0)
(□) pUC8 in 10 mM-Tris·HCl, 10 mM-NaCl, 1 mM EDTA (pH=7.0).
[1, 2].

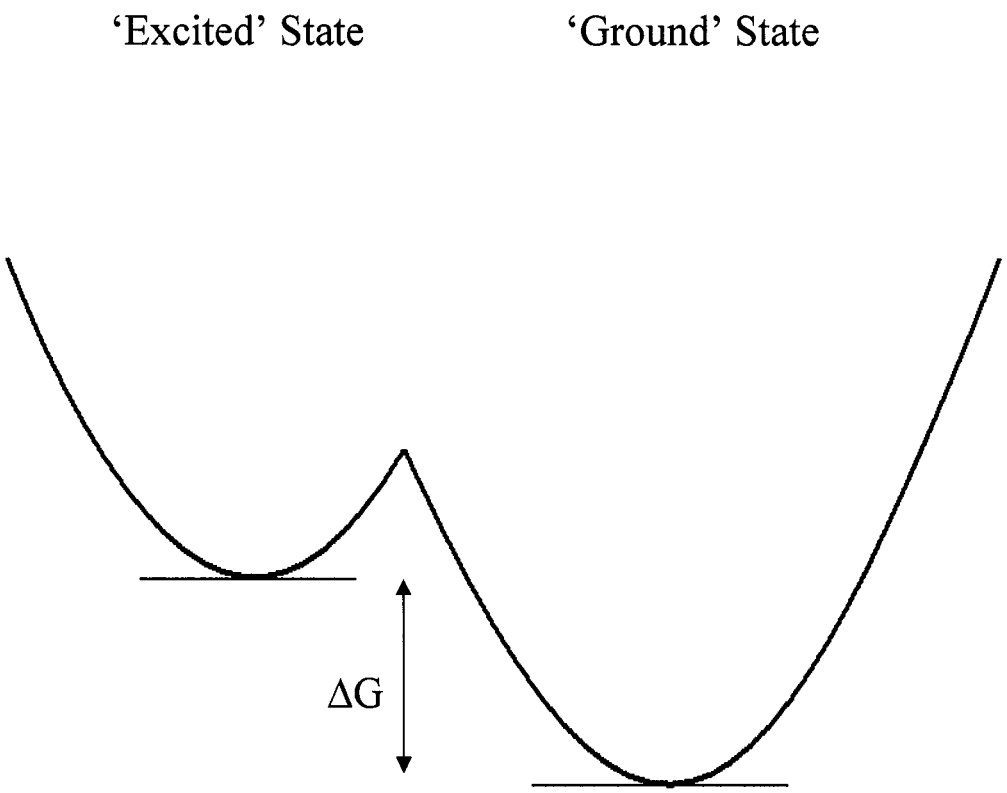


Figure 3.2 Free Energy Diagram necessary for a system to exhibit Metastability. If some fraction of the DNA occupies an 'excited' state, it can relax to the bottom of that basin but will eventually equilibrate to the lower energy 'ground' state. At equilibrium some fraction of the DNA will exist in both states.

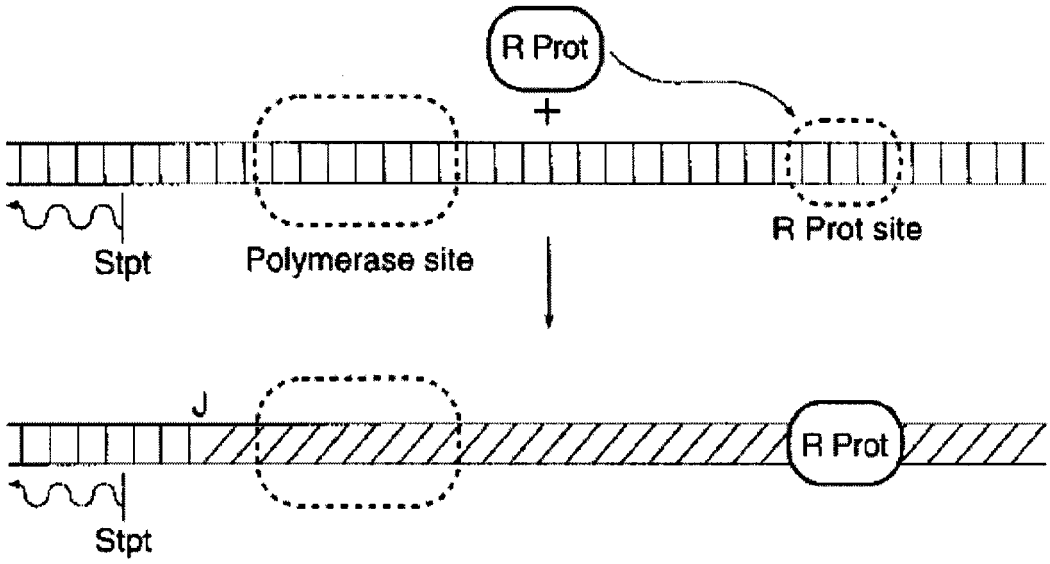


Figure 3.3 How are Allosteric Transitions Important in Biology?
This schematic illustration, from Schurr et al. [8], shows how an allosteric transition of the DNA secondary structure might play a role in remote cell signaling. When the regulatory protein (R) binds to its specific site, the DNA secondary structure over a large domain is converted from its normal (for those conditions) 'ground' state to the slightly less stable 'excited' conformation preferred by the regulatory protein. Provided that either the binding or the kinetics of RNA polymerase is sensitive to the difference in secondary structure between the 'ground' and 'excited' states, a basis for transcriptional regulation via a structural transmission effect is established.

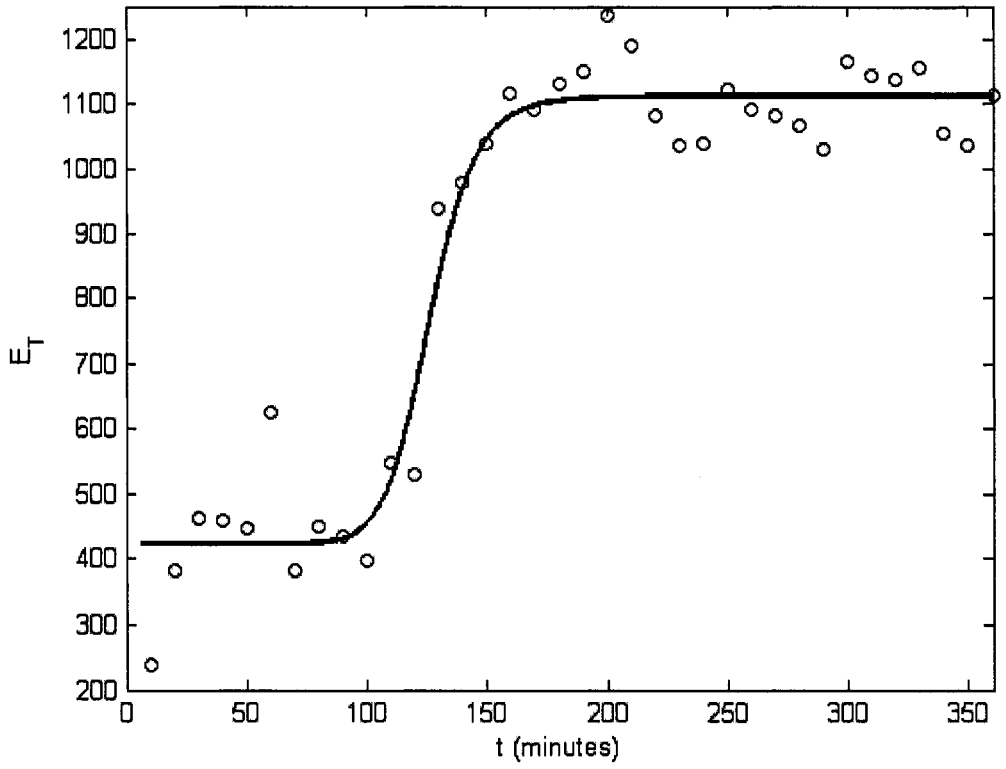


Figure 3.4 The twist energy parameter (E_T) versus time (○) for reactions with 0% (w/v) EG. The magnitude of the E_T values start low at early times just after initiation of the reaction and evolve such that there is a two-fold increase in E_T. Fresh enzyme was added at t = 0, 120, 240 minutes.

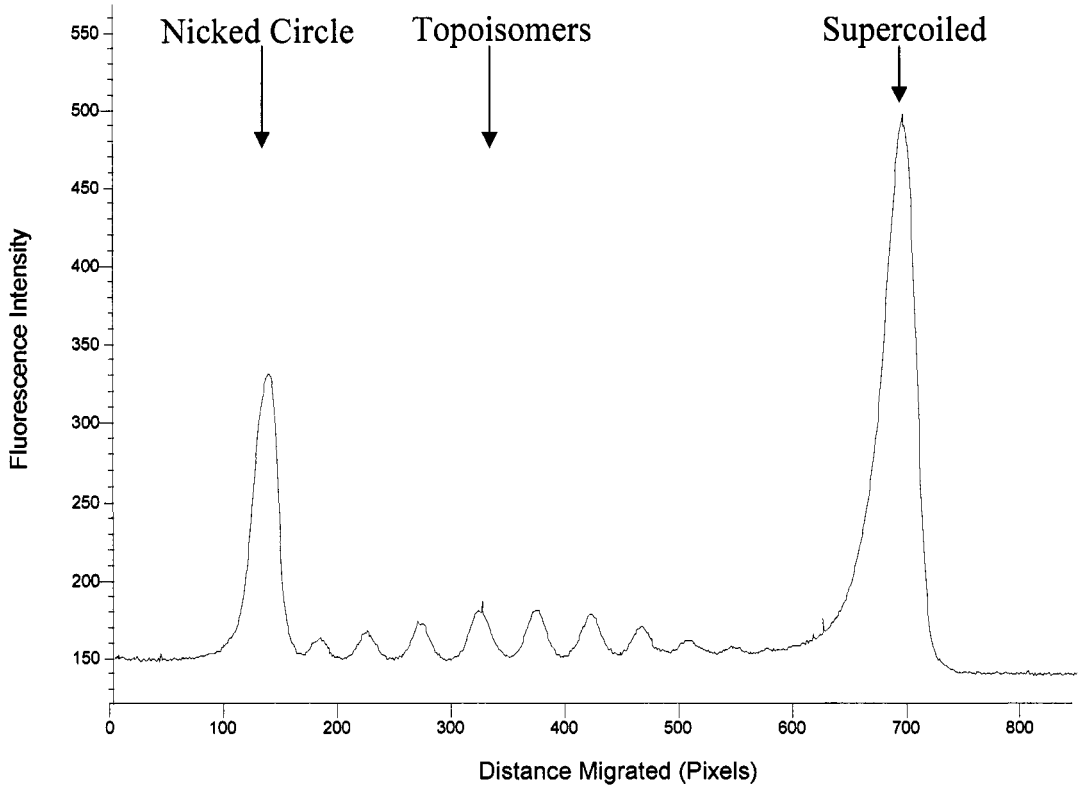


Figure 3.5 Evidence that Calf Thymus Topoisomerase I is a processive enzyme. This lane trace shows gel intensity data for a reaction that was stopped at t = 40 minutes. The topoisomer products have just begun to accumulate and the supercoiled DNA is still present in large excess quantities. The center of the distribution is nearly the same as those reactions taken at long times (i.e. t = 360 minutes) indicating that the Topo I enzyme is a processive enzyme and completely relaxes the superhelical strain in one encounter with the DNA. If this enzyme was a non-processive enzyme, a series of peaks between the supercoiled DNA and the equilibrated distribution would be observed at these short times.

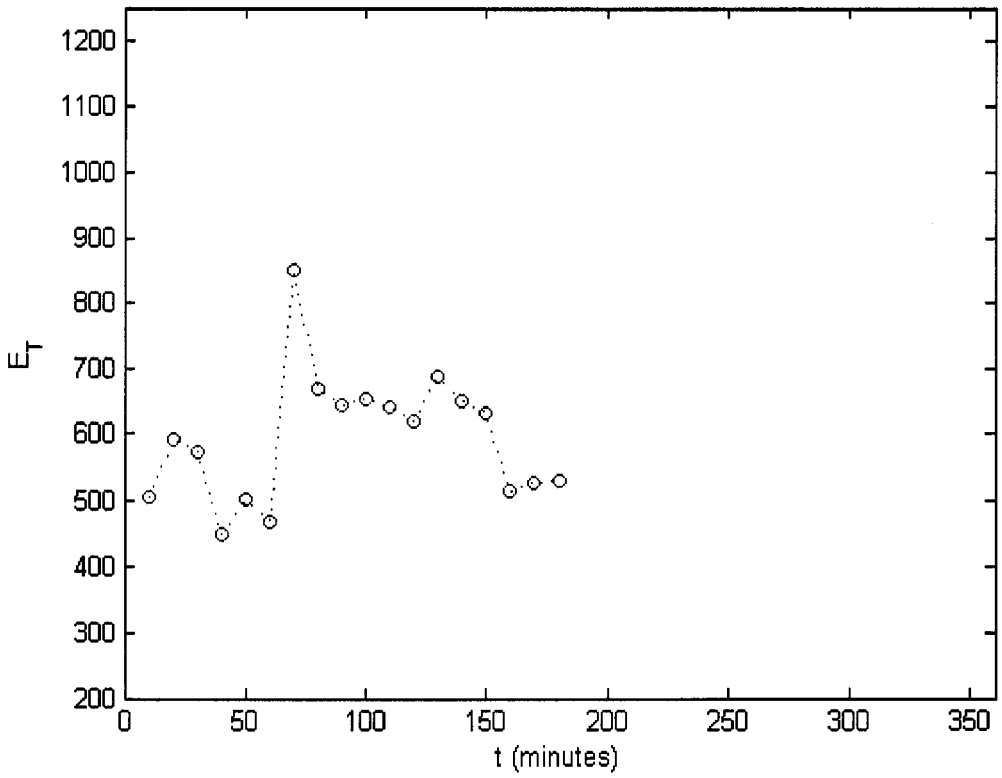


Figure 3.6 E_T vs time (minutes) for reactions with 0% added EG (blue ○) and 1 addition of Topo I enzyme. The trend of the data show that the twist energy parameter starts low and stays low throughout the reaction. This indicates that the enzyme loses activity within the first 2 hours at 37 °C. No 2nd addition of enzyme was made at t = 120 minutes in this experiment.

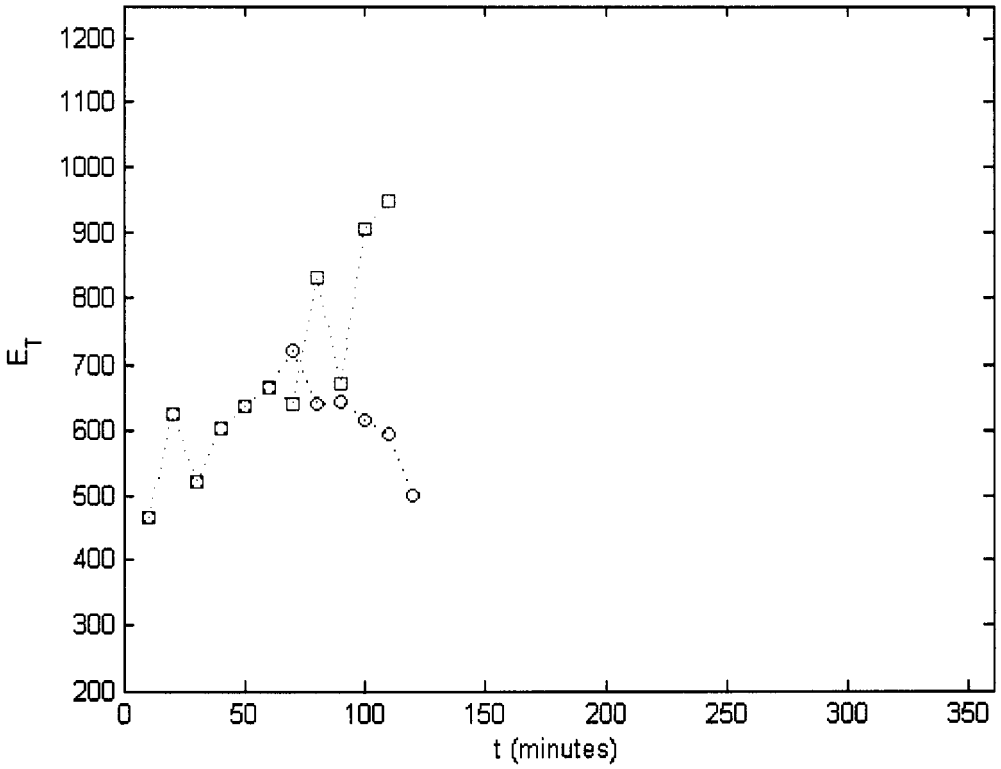


Figure 3.7 E_T vs. time (minutes) comparing those reactions exposed to only 1 addition of Topo I enzyme (blue ○) with those that received a 2nd addition at t = 60 minutes (red □) for reactions that contained 0% EG. The reactions that received a 2nd aliquot of enzyme began to ascend toward their equilibrium values. For $t \leq 60$ minutes the data are identical because the same preparative techniques were to be used for each set of reactions. The data therefore apply to both reactions and are used to describe each of them.

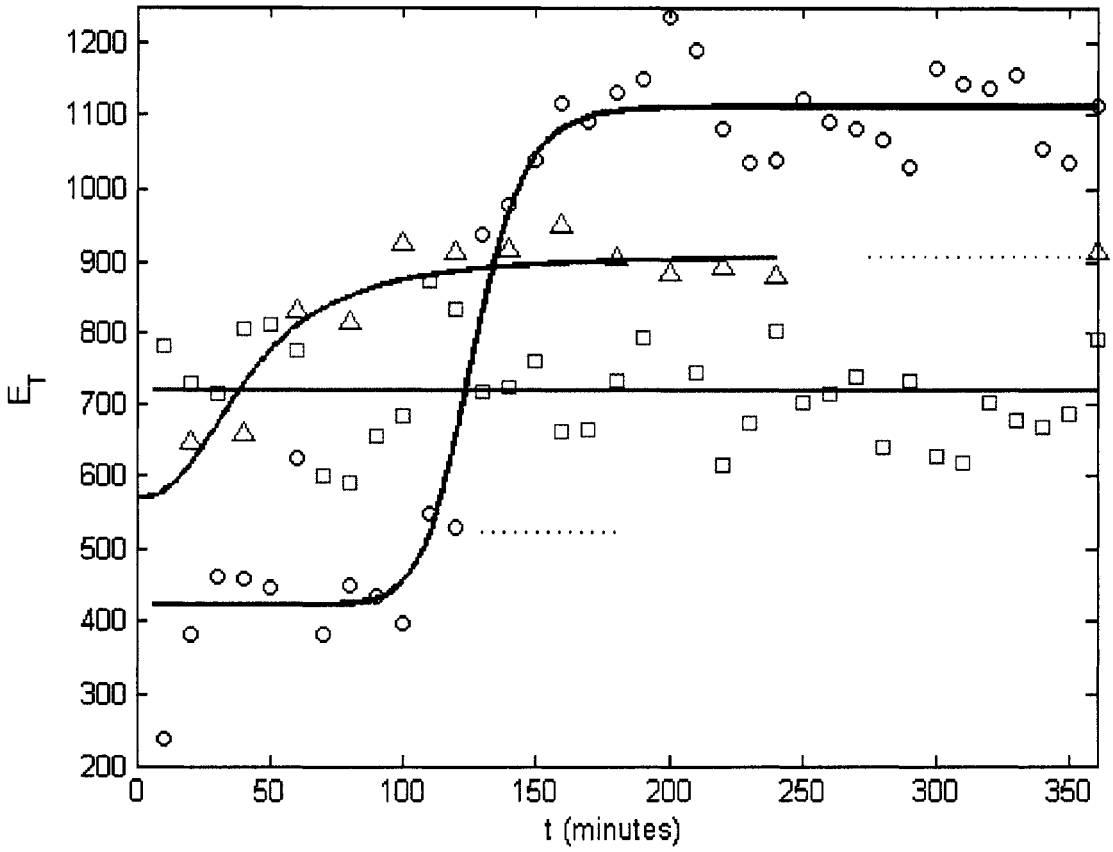


Figure 3.8 E_T vs time for 0% (black ○), 20% (blue Δ), and 40% (gray □) EG. At 0% and 20% EG, E_T starts low at short times and evolves to the higher values associated with equilibrium as the DNA is exposed to more Topo I. At 40% EG, E_T shows no signs of evolving and is constant throughout. The DNAs from reactions with 0% and 40% EG are in different structural states at t = 360 minutes where equilibration of the DNA has occurred.

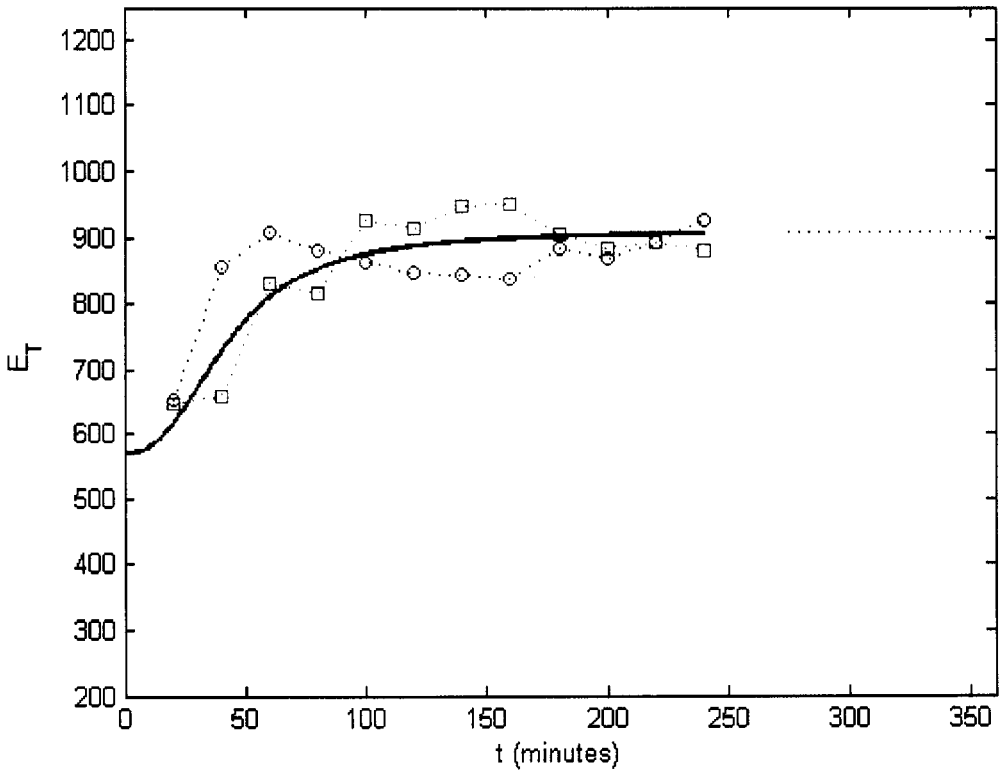


Figure 3.9 E_T vs. time (minutes) comparing reactions with 20% EG exposed to only 1 addition of Topo I enzyme (blue ○) with those that received a 2^{nd} addition at t = 120 minutes (red □). These data extend to 240 minutes (4 hours). Within the error of the experiments there is no difference between the data sets. The solid blue line is the line from Figure 3.8 and is overlaid on this data only to show that the curves at 20% EG all show the same observed behavior. E_T values at short times are low and evolve in time upon exposure to the enzyme Topo I to those values that are associated with 'equilibrated' DNAs.

List of References

Bachmann, B., W. Luke, et al. (1990). "Improvement of PCR amplified DNA sequencing with the aid of detergents." Nucleic Acids Res **18**(5): 1309.

Bauer, W. R. (1978). "Structure and reactions of closed duplex DNA." Annu Rev Biophys Bioeng **7**: 287-313.

Benight, A. S., Schurr, J.M., Flynn, P.F., Reid, B.R., and Wemmer, D.E. (1988). "Melting of a Self-complementary DNA Minicircle." Journal of Molecular Biology **200**: 377-399.

Benight, A. S., R. M. Wartell, et al. (1981). "Theory agrees with experimental thermal denaturation of short DNA restriction fragments." Nature **289**(5794): 203-5.

Bevington, P. R. a. R., D.K. (1992). Data Reduction and Error Analysis for the Physical Sciences. Boston, Massachusetts, The McGraw-Hill Companies, Inc.

Bloomfield, V. A., Crothers, D.M., Tinoco, Jr., I (2000). Nucleic Acids: Structures, Properties, and Functions. Sausalito, California, University Science Books.

Bond, J. P., C. F. Anderson, et al. (1994). "Conformational transitions of duplex and triplex nucleic acid helices: thermodynamic analysis of effects of salt concentration on stability using preferential interaction coefficients." Biophys J **67**(2): 825-36.

Bouchiat, C. M., M. (1998). "Elasticity theory of a supercoiled DNA molecule." Phys. Rev. Lett. **80**: 1556-1559.

Brewood, G., Rangel, DR, Aliwarga, T, and Schurr, JM (2005). "Alternative Conformations of DNA Induced by Osmotic Stress." Biophysical Journal **88**(1): 567A.

Bryant, Z., M. D. Stone, et al. (2003). "Structural transitions and elasticity from torque measurements on DNA." Nature **424**(6946): 338-41.

Camilloni, G., E. Di Martino, et al. (1988). "Eukaryotic DNA topoisomerase I reaction is topology dependent." Nucleic Acids Res **16**(14B): 7071-85.

Camilloni, G., E. Di Martino, et al. (1989). "Regulation of the function of eukaryotic DNA topoisomerase I: topological conditions for inactivity." Proc Natl Acad Sci U S A **86**(9): 3080-4.

Chakrabarti, R. and C. E. Schutt (2001). "The enhancement of PCR amplification by low molecular-weight sulfones." Gene **274**(1-2): 293-8.

Chakrabarti, R. and C. E. Schutt (2001). "The enhancement of PCR amplification by low molecular weight amides." Nucleic Acids Res **29**(11): 2377-81.

Chalikian, T. V., A. P. Sarvazyan, et al. (1994). "Influence of base composition, base sequence, and duplex structure on DNA hydration: apparent molar volumes and apparent molar adiabatic compressibilities of synthetic and natural DNA duplexes at 25 degrees C." Biochemistry **33**(9): 2394-401.

Cheng, S., C. Fockler, et al. (1994). "Effective amplification of long targets from cloned inserts and human genomic DNA." Proc Natl Acad Sci U S A **91**(12): 5695-9.

Chevet, E., G. Lemaitre, et al. (1995). "Low concentrations of tetramethylammonium chloride increase yield and specificity of PCR." Nucleic Acids Res **23**(16): 3343-4.

Chou, Q. (1992). "Minimizing deletion mutagenesis artifact during Taq DNA polymerase PCR by E. coli SSB." Nucleic Acids Res **20**(16): 4371.

Clendenning, J. B. (1993). The Interaction of Unwinding Ligands with Circular DNAs. Chemistry. Seattle, Washington, University of Washington. **PhD**.

Clendenning, J. B., A. N. Naimushin, et al. (1994). "Effect of ethidium binding and superhelix density on the supercoiling free energy and torsion and bending constants of p30 delta DNA." Biophys Chem **52**(3): 191-218.

Clendenning, J. B. and J. M. Schurr (1994). "Circularization of small DNAs in the presence of ethidium: a theoretical analysis." Biopolymers **34**(7): 849-68.

Comey, C. T., J. M. Jung, et al. (1991). "Use of formamide to improve amplification of HLA DQ alpha sequences." Biotechniques **10**(1): 60-1.

Courtenay, E. S., M. W. Capp, et al. (2000). "Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants in vivo and for the interpretation of "osmotic stress" experiments in vitro." Biochemistry **39**(15): 4455-71.

Delcourt, S. G. and R. D. Blake (1991). "Stacking energies in DNA." J Biol Chem **266**(23): 15160-9.

Delrow, J. J. (1996). Evidence of Alternative Secondary Structure States in DNA. Chemistry. Seattle, Washington, University of Washington. **PhD**.

Delrow, J. J., P. J. Heath, et al. (1998). "Effect of temperature on DNA secondary structure in the absence and presence of 0.5 M tetramethylammonium chloride." Biopolymers **45**(7): 503-15.

Delrow, J. J., P. J. Heath, et al. (1997). "On the origin of the temperature dependence of the supercoiling free energy." Biophys J **73**(5): 2688-701.

Depew, R. a. W., JC (1975). "Conformational Fluctuations of DNA Helix." Proceedings of the National Academy of Sciences of the United States of America **72**(11): 4275-4279.

Di Mauro, E. Topological Evidence for Allosteric Transitions in DNA Secondary Structure (Personal Communication).

Di Mauro, E., M. Caserta, et al. (1985). "Activation of in vitro transcription and topology of closed DNA domains." J Biol Chem **260**(1): 152-9.

Díaz, R. P. (2002). Fast and slow internal dynamics of 1^3C labeled DNA oligomers in solution, University of Washington. **Ph. D. Thesis**.

Doktycz, M. J., R. F. Goldstein, et al. (1992). "Studies of DNA dumbbells. I. Melting curves of 17 DNA dumbbells with different duplex stem sequences linked by T4 endloops: evaluation of the nearest-neighbor stacking interactions in DNA." Biopolymers **32**(7): 849-64.

Engel, T. a. R., P. (2006). Thermodynamics, Statistical Thermodynamics and Kinetics. San Francisco, California, Pearson Education, Inc.

Falzon, L., C. Kirk, et al. (1994). "PCR generation of large amounts of purified DNA." J Biochem Biophys Methods **29**(3-4): 251-7.

Flick, E. (1998). Industrial Solvents Handbook, William Andrew Publishing/Noyes.

Frackman, S., Kobs, Gary, Simpson, Dan, and Storts, Doug (1998). "Betaine and DMSO: Enhancing Agents for PCR." Promega notes **65**: 27.

Frank-Kamenetskii, M. D. (1971). "Simplification of the Empirical Relationship between Melting Temperature of DNA, Its GC Content and Concentration of Sodium Ions in Solution." Biopolymers **10**: 2623-2624.

Frank-Kamenetskii, M. D., A. V. Lukashin, et al. (1985). "Torsional and bending rigidity of the double helix from data on small DNA rings." J Biomol Struct Dyn **2**(5): 1005-12.

Franklin, S. E. and R. G. Gosling (1953). "Molecular configuration in sodium thymonucleate." Nature **171**(4356): 740-1.

Fujimoto, B. S. and J. M. Schurr (1990). "Dependence of the torsional rigidity of DNA on base composition." Nature **344**(6262): 175-7.

Fuller, F. (1971). "The Writhing Number of a Space Curve." Proceedings of the National Academy of Sciences of the United States of America **68**(4): 815-819.

Gebe, J. A., S. A. Allison, et al. (1995). "Monte Carlo simulations of supercoiling free energies for unknotted and trefoil knotted DNAs." Biophys J **68**(2): 619-33.

Gebe, J. A., J. J. Delrow, et al. (1996). "Effects of Na+ and Mg2+ on the structures of supercoiled DNAs: comparison of simulations with experiments." J Mol Biol **262**(2): 105-28.

Heath, P. J., J. B. Clendenning, et al. (1996). "Effect of bending strain on the torsion elastic constant of DNA." J Mol Biol **260**(5): 718-30.

Hoelzel, R. (1990). "The trouble with 'PCR' machines." Trends Genet **6**(8): 237-8.

Horowitz, D. S. and J. C. Wang (1984). "Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling." J Mol Biol **173**(1): 75-91.

Huang, M. M., N. Arnheim, et al. (1992). "Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR." Nucleic Acids Res **20**(17): 4567-73.

Hung, T., K. Mak, et al. (1990). "A specificity enhancer for polymerase chain reaction." Nucleic Acids Res **18**(16): 4953.

Kahn, J. D. and D. M. Crothers (1998). "Measurement of the DNA bend angle induced by the catabolite activator protein using Monte Carlo simulation of cyclization kinetics." J Mol Biol **276**(1): 287-309.

Kahn, J. D., E. Yun, et al. (1994). "Detection of localized DNA flexibility." Nature **368**(6467): 163-6.

Keller, W. (1975). "Determination of the number of superhelical turns in simian virus 40 DNA by gel electrophoresis." Proc Natl Acad Sci U S A **72**(12): 4876-80.

Kim, U. S., B. S. Fujimoto, et al. (1993). "Dynamics and structures of DNA: long-range effects of a 16 base-pair (CG)8 sequence on secondary structure." Biopolymers **33**(11): 1725-45.

Kovarova, M. and P. Draber (2000). "New specificity and yield enhancer of polymerase chain reactions." Nucleic Acids Res **28**(13): E70.

Langowski, J., A. S. Benight, et al. (1985). "Change of conformation and internal dynamics of supercoiled DNA upon binding of Escherichia coli single-strand binding protein." Biochemistry **24**(15): 4022-8.

Lee, C. H., H. Mizusawa, et al. (1981). "Unwinding of double-stranded DNA helix by dehydration." Proc Natl Acad Sci U S A **78**(5): 2838-42.

Linz, U. (1990). "Thermocycler temperature variation invalidates PCR results." Biotechniques **9**(3): 286, 288, 290-3.

Maniatis, T., Fritsch, EF, and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. New York, Cold Spring Harbor laboratory Press.

McDowell, D. G., N. A. Burns, et al. (1998). "Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR." Nucleic Acids Res **26**(14): 3340-7.

McQuarrie, D. A. (2000). Statistical Mechanics. Sausalito, California, University Science Books.

Moroz, J. D. and P. Nelson (1997). "Torsional directed walks, entropic elasticity, and DNA twist stiffness." Proc Natl Acad Sci U S A **94**(26): 14418-22.

Mullis, K., F. Faloona, et al. (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." Cold Spring Harb Symp Quant Biol **51 Pt 1**: 263-73.

Mytelka, D. S. and M. J. Chamberlin (1996). "Analysis and suppression of DNA polymerase pauses associated with a trinucleotide consensus." Nucleic Acids Res **24**(14): 2774-81.

Naimushin, A. N., J. B. Clendenning, et al. (1994). "Effect of ethidium binding and superhelix density on the apparent supercoiling free energy and torsion constant of pBR322 DNA." Biophys Chem **52**(3): 219-26.

Naimushin, A. N., B. S. Fujimoto, et al. (2000). "Dynamic bending rigidity of a 200-bp DNA in 4 mM ionic strength: a transient polarization grating study." Biophys J **78**(3): 1498-518.

Naimushin, A. N., N. Quach, et al. (2001). "Effect of polyethylene glycol on the supercoiling free energy of DNA." Biopolymers **58**(2): 204-17.

Newell, G. a. M., EW (1953). "On the Theory of the Ising Model of Ferromagnetism." Reviews of Modern Physics **25**(2): 353-389.

Owczarzy, R., Y. You, et al. (2004). "Effects of sodium ions on DNA duplex oligomers: improved predictions of melting temperatures." Biochemistry **43**(12): 3537-54.

Owczarzy, R. a. B., Mark. (2005). "Calculation of Tm for Oligonucleotide Duplexes." from http://www.idtdna.com/support/technical/TechnicalBulletinPDF/Calculation_of_Tm_for_Oligonucleotide_Duplexes.pdf.

Parekh, B. S. and G. W. Hatfield (1996). "Transcriptional activation by protein-induced DNA bending: evidence for a DNA structural transmission model." Proc Natl Acad Sci U S A **93**(3): 1173-7.

Press, W. H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. (1986). Numerical Recipes: The Art of Scientific Computing. Cambridge, Cambridge University Press.

Qu, X. and J. B. Chaires (2001). "Hydration changes for DNA intercalation reactions." J Am Chem Soc **123**(1): 1-7.

Rangel, D. P. (2005). Effects of neutral osmolytes on DNA, Ph.D. Thesis, University of Washington. **Ph.D. Thesis**.

Rangel, D. P., C. A. Sucato, et al. (2004). "Effects of small neutral osmolytes on the supercoiling free energy and intrinsic twist of p30delta DNA." Biopolymers **75**(4): 291-313.

Rentzeperis, D., D. P. Kharakoz, et al. (1991). "Coupling of sequential transitions in a DNA double hairpin: energetics, ion binding, and hydration." Biochemistry **30**(25): 6276-83.

Robinson, C. R. and S. G. Sligar (1998). "Changes in solvation during DNA binding and cleavage are critical to altered specificity of the EcoRI endonuclease." Proc Natl Acad Sci U S A **95**(5): 2186-91.

Roux, K. H. (1995). "Optimization and troubleshooting in PCR." PCR Methods Appl **4**(5): S185-94.

Roychoudhury, M., A. Sitlani, et al. (2000). "Global structure and mechanical properties of a 10-bp nucleosome positioning motif." Proc Natl Acad Sci U S A **97**(25): 13608-13.

Saenger, W. (1984). Principles of Nucleic Acid Structure. New York, Springer-Verlag.

Schneider, B. and H. M. Berman (1995). "Hydration of the DNA bases is local." Biophys J **69**(6): 2661-9.

Schurr, J. M., J. J. Delrow, et al. (1997). "The question of long-range allosteric transitions in DNA." Biopolymers **44**(3): 283-308.

Schurr, J. M., Fujimoto, B.S., Wu, P. & Song, L. (1992). Fluorescence Studies of Nucleic Acids. Dynamics, Rigidities, and Structures, in Topics in Fluorescence Spectroscopy vol. 3, Biochemical Applications. New York, Plenum press.

Schurr, J. M., D. P. Rangel, et al. (2005). "A contribution to the theory of preferential interaction coefficients." Biophys J **89**(4): 2258-76.

Shibata, J. a. S., JM (1981). "A Theory of Aggregation in the Thermal Denaturation Region of Multistrand Biopolymers." Biopolymers **20**: 525-549.

Shibata, J. H., J. Wilcoxon, et al. (1984). "Structures and dynamics of a supercoiled DNA." Biochemistry **23**(6): 1188-94.

Shimada, J. and H. Yamakawa (1985). "Statistical mechanics of DNA topoisomers. The helical worm-like chain." J Mol Biol **184**(2): 319-29.

Shimada, J. Y., H. (1984). "Ring closure probabilities for twisted wormlike chains. Application to DNA." Macromolecules(17): 689-698.

Shore, D. and R. L. Baldwin (1983). "Energetics of DNA twisting. I. Relation between twist and cyclization probability." J Mol Biol **170**(4): 957-81.

Shore, D. and R. L. Baldwin (1983). "Energetics of DNA twisting. II. Topoisomer analysis." J Mol Biol **170**(4): 983-1007.

Shui, X., L. McFail-Isom, et al. (1998). "The B-DNA dodecamer at high resolution reveals a spine of water on sodium." Biochemistry **37**(23): 8341-55.

Shui, X., C. C. Sines, et al. (1998). "Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations." Biochemistry **37**(48): 16877-87.

Sidorova, N. Y. and D. C. Rau (1996). "Differences in water release for the binding of EcoRI to specific and nonspecific DNA sequences." Proc Natl Acad Sci U S A **93**(22): 12272-7.

Sinden, R. R. (1994). DNA Structure and Function. San Diego, California, Academic Press, Inc.

Song, L., B. S. Fujimoto, et al. (1990). "Evidence for allosteric transitions in secondary structure induced by superhelical stress." J Mol Biol **214**(1): 307-26.

Spiegel, M. R. (1974). Schaum's Outlines of Theory and Problems of Fourier Analysis with Applications to Boundary Value Problems. New York, McGraw-Hill.

Spink, C. H. and J. B. Chaires (1999). "Effects of hydration, ion release, and excluded volume on the melting of triplex and duplex DNA." Biochemistry **38**(1): 496-508.

Stellwagen, N. C. (1983). "Anomalous electrophoresis of deoxyribonucleic acid restriction fragments on polyacrylamide gels." Biochemistry **22**(26): 6186-93.

Stellwagen, N. C. (2000). "Conformational isomers of curved DNA molecules can be observed by polyacrylamide gel electrophoresis." Electrophoresis **21**(12): 2327-34.

Strick, T., J. Allemand, et al. (2000). "Twisting and stretching single DNA molecules." Prog Biophys Mol Biol **74**(1-2): 115-40.

Strick, T. R., J. F. Allemand, et al. (1996). "The elasticity of a single supercoiled DNA molecule." Science **271**(5257): 1835-7.

Strick, T. R., J. F. Allemand, et al. (1998). "Behavior of supercoiled DNA." Biophys J **74**(4): 2016-28.

Taylor, W. H. and P. J. Hagerman (1990). "Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure. II. NaCl-dependence of DNA flexibility and helical repeat." J Mol Biol **212**(2): 363-76.

Varadaraj, K. and D. M. Skinner (1994). "Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases." Gene **140**(1): 1-5.

Vologodskaia, M. and A. Vologodskii (2002). "Contribution of the intrinsic curvature to measured DNA persistence length." J Mol Biol **317**(2): 205-13.

Vologodskii, A. V. and J. F. Marko (1997). "Extension of torsionally stressed DNA by external force." Biophys J **73**(1): 123-32.

Vossen, K. M., R. Wolz, et al. (1997). "Role of macromolecular hydration in the binding of the Escherichia coli cyclic AMP receptor to DNA." Biochemistry **36**(39): 11640-7.

Wang, J. C. (1985). "DNA topoisomerases." Annu Rev Biochem **54**: 665-97.

Wang, J. C. (2002). "Cellular roles of DNA topoisomerases: a molecular perspective." Nat Rev Mol Cell Biol **3**(6): 430-40.

Wartell, R. M., and Benight, A.S. (1985). "Thermal Denaturation of DNA Molecules: A Comparison of Theory with Experiment." Physics Report (Review Section of Physics Letters) **126**(2): 67-107.

Watson, J. D. and F. H. Crick (1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid." Nature **171**(4356): 737-8.

Weast, R. C., Ed. (1980). CRC Handbook of Chemistry and Physics. Boca Raton, Florida, CRC Press, Inc.

Westhof, E. (1988). "Water: an integral part of nucleic acid structure." Annu Rev Biophys Biophys Chem **17**: 125-44.

White, J. (1969). "Self-Linking and the Gauss Integral in Higher Dimensions." American Journal of Mathematics **91**: 693-728.

Winship, P. R. (1989). "An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide." Nucleic Acids Res **17**(3): 1266.

Woodford, K., M. N. Weitzmann, et al. (1995). "The use of K(+)-free buffers eliminates a common cause of premature chain termination in PCR and PCR sequencing." Nucleic Acids Res **23**(3): 539.

Wu, P. G., B. S. Fujimoto, et al. (1991). "Effect of ethidium on the torsion constants of linear and supercoiled DNAs." Biophys Chem **41**(3): 217-36.

Wu, P. G., L. Song, et al. (1988). "Interaction of chloroquine with linear and supercoiled DNAs. Effect on the torsional dynamics, rigidity, and twist energy parameter." Biochemistry **27**(21): 8128-44.

Zhang, Y. and D. M. Crothers (2003). "High-throughput approach for detection of DNA bending and flexibility based on cyclization." Proc Natl Acad Sci U S A **100**(6): 3161-6.

Zhang, Y. and D. M. Crothers (2003). "Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization." Biophys J **84**(1): 136-53.

Appendix A: Fourier Series Plus χ^2 Minimization Protocol for Fitting Finite Data Sets

A Fourier-based method of analysis was combined with a χ^2 minimization routine to produce a best fit smooth line describing the data in Figure 1.12 via a matlab computer code (not shown) that subsequently implemented this protocol. The code will be omitted from this appendix but the framework should be sufficiently derived so that one may follow and implement it for simplistic fits whereby the qualitative shape of the data is important. This code can be applied to any data set where a best fit smooth line is all that is desired.

In equation (A.1) the average of the data is subtracted off from every data point to simplify the analysis. In doing so, the constant fourier component (a_o) is removed and the matrices are made symmetric due to equal numbers of sine and cosine fourier components. To regenerate a curve through the data, one only needs to add the average value back into $\overline{f}_{th}$ once the fourier coefficients (a_n and b_n) are calculated.

This calculation assumes that there are N data points $\left\{i.e.\,\overline{f}\left(x_j\right); j=1,\ldots N\right\}$, and $M \le N$ fourier components where M and N are even numbers since M/2 and N/2 must be whole numbers in the subsequent summations.

$$\overline{f}^{th}(x) = \sum_{n=1}^{M/2}\left[a_n \cos(2\pi nx/L) + b_n \sin(2\pi nx/L)\right] \qquad \text{(A.1)}$$

Equation (A.2) defines χ^2, wherein each data point is compared to that point predicted by equation (A.1) above for the same x_j. If the σ_j are not known for each data point, they may be set to 1 in order to proceed with the analysis.

$$\chi^2 = \frac{\sum_{j=1}^{N}\left(\bar{f}(x_j)-\bar{f}^{th}(x_j)\right)^2}{2\sigma_j^2} = \frac{\sum_{j=1}^{N}\left(\bar{f}(x_j)-\left(\sum_{n=1}^{M/2}\left(a_n\cos(2\pi nx_j/L)+b_n\sin(2\pi nx_j/L)\right)\right)\right)^2}{2\sigma_j^2} \tag{A.2}$$

Now χ^2 is minimized with respect to the cosine fourier coefficients, a_n, and set to zero to find their optimum. The same will also be done with respect to the sine coefficients, b_n but will not be explicitly shown.

$$\frac{\delta\chi^2}{\delta a_m} = 2\sum_{j=1}^{N}\left(\frac{\left(f(x_j)-\sum_{n=1}^{M/2}(a_n\cos(2\pi nx_j/L)+b_n\sin(2\pi nx_j/L))\right)\cdot(-\cos(2\pi mx_j/L))}{2\sigma_j^2}\right)=0 \tag{A.3}$$

$$\frac{\delta\chi^2}{\delta a_m} = \sum_{j=1}^{N} f(x_j)\cos(2\pi mx_j/L) - \sum_{n=1}^{M/2} a_n \sum_{j=1}^{N}\cos(2\pi nx_j/L)\cos(2\pi mx_j/L) - \sum_{n=1}^{M/2} b_n \sum_{j=1}^{N}\sin(2\pi nx_j/L)\cos(2\pi mx_j/L) = 0 \tag{A.4}$$

In order to find the coefficients that will be substituted back into equation (A.1) and which will be necessary to reconstruct a curve through the data, a matrix relation will be set up that can be solved by matrix inversion as follows. Let, the 1 x M/2 matrix F(m) be given by

$$F(m) = \sum_{n=1}^{M/2} a_n(X_{nm}+Y_{nm}); m = 1,\ldots M/2 \tag{A.5}$$

where

$$F(m) = \sum_{j=1}^{N} f(x_j)\cos(2\pi m x_j / L); m = 1,\ldots M/2 \quad \text{(A.6)}$$

and X_{nm} and Y_{nm} are defined to be.

$$X_{nm} = \sum_{j=1}^{N} \cos(2\pi n x_j / L)\cos(2\pi m x_j / L) \quad \text{(A.7)}$$

$$Y_{nm} = \sum_{j=1}^{N} \sin(2\pi n x_j / L)\cos(2\pi m x_j / L) \quad \text{(A.8)}$$

respectively. In a smiliar manner, a matrix G(m) is written as follows for the sine coefficients

$$G(m) = \sum_{n=1}^{M/2} b_n(U_{nm} + V_{nm}) = \sum_{j=1}^{N} f(x)\sin(2\pi m x_j / L); m = 1,\ldots M/2 \quad \text{(A.9)}$$

and has U_{nm} and V_{nm} defined to be

$$U_{nm} = \sum_{j=1}^{N} \cos(2\pi n x_j / L)\sin(2\pi m x_j / L) \quad \text{(A.10)}$$

$$V_{nm} = \sum_{j=1}^{N} \sin(2\pi n x_j / L)\sin(2\pi m x_j / L) \quad \text{(A.11)}$$

The sums of X_{nm}, Y_{nm}, U_{nm}, and V_{nm} extend over the ranges m = 1,...M/2 and n = 1,...M/2.

A new matrix, H(m), will be constructed from F(m) and G(m) as a result of a redefinition of the sine coefficients in terms of only one variable $\left\{b_1,\ldots b_{\frac{M}{2}+1}\right\}=\left\{a_{\frac{M}{2}+1},\ldots a_M\right\}$. This produces an M x 1 column vector

$$H(m)=\begin{vmatrix} F(1) \\ \vdots \\ F(M/2) \\ G(M/2+1) \\ \vdots \\ G(M) \end{vmatrix};m=1,\ldots M \tag{A.12}$$

which will subsequently be defined as

$$H(m)=|W|\cdot\begin{vmatrix} a_1 \\ \vdots \\ a_M \end{vmatrix} \tag{A.13}$$

where the M x M matrix |W| is constructed by building the four other matrices (X_{nm}, Y_{nm}, U_{nm}, and V_{nm}) into it as follows

$$W\left(1:\frac{M}{2},1:\frac{M}{2}\right)=X_{nm}$$
$$W\left(\frac{M}{2}+1:M,1:\frac{M}{2}\right)=Y_{nm}$$
$$W\left(1:\frac{M}{2},\frac{M}{2}+1:M\right)=U_{nm}$$
$$W\left(\frac{M}{2}+1:M,\frac{M}{2}+1:M\right)=V_{nm}$$

so that

$$|W| = \begin{vmatrix} X_{nm} & U_{nm} \\ Y_{nm} & V_{nm} \end{vmatrix} \tag{A.14}$$

Thus the problem of finding the fourier coefficients that also minimize χ^2 has been simplified to a matrix problem. The fourier coefficients are found by matrix inversion as follows

$$\begin{vmatrix} a_1 \\ \vdots \\ a_M \end{vmatrix} = |W|^{-1} \begin{vmatrix} F(1) \\ \vdots \\ F(M/2) \\ G(M/2+1) \\ \vdots \\ G(M) \end{vmatrix} \tag{A.15}$$

Substitution of these coefficients back into equation (A.1) over the range of x_j and adding back on average that was subtracted off initially produces a best fit line to the data. This method was used to find the best fit curve through the data in the bottom panel of Figure 1.12.

Appendix B: Standard Fourier Analysis Fitting Routine

A standard fourier method of fourier analysis was employed to analyze the CD data of p30δ in solutions containing 0 – 40% EG, whose fits are shown in Figure 2.6. The protocol employed can be found from any standard text. The computer code, along with some examples of the results it produces is included in Appendix B. The analysis uses relationships from Appendix A, and it is necessary to specify the number of fourier coefficients that are included (*kincluded*) explicitly in the code. While it has the capacity to perform a χ^2 analysis (the variable called *chiswich* turns this feature on and off) its main utility is simply to take the data and plot a fitted line, given by the fourier coefficients found for a given number of components included. If the χ^2 portion of the program is on, a plot of χ^2 vs. *kincluded* will be plotted so that a determination can be made as to the appropriate number of components to include in the analysis (see Figure B.3). The selection of *kincluded* simply truncates the series of all higher frequency contributions and produces a curve that fits the data to the subjective specification. The series truncation acts as a filter in the process.

An independent source of information typically determines the number of components to be included in a fit to the data, however, no such information was available for the CD experiments described in chapter 2. Increasing the number of sine and cosine terms in any fit results in lower χ^2 values since the data will eventually be reproduced when the number of fourier components is equal to the number of data points. The plot of χ^2 vs. *kincluded* will decay toward zero when this happens. When *kincluded* equals the number of data points, the generated fourier fit line will be driven through each original data point. The utility of the χ^2 vs. *kincluded* plot is to reveal, by inspection, an acceptable place in the graph where the χ^2 is at a plateau (for those plots that show one). Above a certain threshold of *kincluded*, a drastic improvement to the data is not apparent and the fourier fit begins to accomodate any noise that may be present in the data (see Figure B.3). A section of code was added to produce a plot of the slope of χ^2 vs. *kincluded* that can be used to reveal an appropriate place to cutoff the fourier series when *no* plateau is apparent. Figure B.4 shows such a plot of the slope of χ^2 vs. *kincluded* for

the data in Figure B.1A. The slope of χ^2 vs. *kincluded* is calculated for each adjacent set of points. This curve also converges to 0 as *kincluded* approaches the total number of data points, but offers more resolution than the plot of χ^2 vs. *kincluded.* The goal of this program is to find an appropriate number of terms to include in the fourier series so that an accurate but smoothed representation of the data can be made.

B.2 Matlab Code: FOURIER COEFFICIENT CALCULATOR

```
clear all
close all
clc
%This section of the code imports the data to be analyzed by the method of fourier series.
        It calculates the fourier components and plots the data along with the curve
        containing the desired number of fourier components
%cddata1 is the storage file that contains the data (any data) to be analyzed with this
%program. Column 1 – wavelength (i.e. x-axis) data and Columns 2 – numberdatasets
has %the cd (or y-axis) data
cddata1=load('C:\ . . . cddata1.m');
numberdatasets=6; %The data loaded into this program is stored in the file given above
        by the pathname, the x-data is stored in column 1 and the y-data (potentially from
        multiple runs) is stored in columns 2 and higher. This program loads the data and
        does a fourier analysis (or fourier filtering) on as much data as is declared by
        'columndata'.
%columndata == the number of columns to be analyzed in the program.
l=length(cddata1);
numberfouriercomponents=80; %floor(l/2); %number of fourier components to include
numberfouriercurves=numberdatasets*numberfouriercomponents;
swich=1; %This switch turns on the plotting section of the program and selects out all
        fourier wavelengths less than the k-th component of the fourier series to plot
        superimposed on the data
      %swich == 0 Plotting program off; swich == 1 Plotting program on
plotallswich=1; %If plotallswich=0, plot of data specified with dataselect and kincluded
        is plotted otherwise if plotallswich=1, all of the plots specified by kincluded is
        plotted
chiswich=1; %If chiswich=0, chisquare calculator/slope plotter is off. If chiswich=1, it is
        on.
slopeswich=1; %If slopeswich=0, slopeswich plotting program is off. If slopeswich=1, it
        is on.
dataselect=2;%5;
kincluded=3;
currentk=0;
m=(l/2)-0.5;
n=m+1;
xmean=(sum(cddata1(:,1))/l);
fouriercurve2(1:l,1:numberfouriercurves)=0;
chisquareslope(:,1)=0;

        for f=1:1:numberfouriercomponents;
          kcutoff=f;
              %This section calculates the coefficients of the cosine(a's) and sine(b's)
              fourier components
```

```
        for i=1:1:numberdatasets;
                if i==1;
                        currentk=1+currentk; %counter
                end
                for k=1:1:l;     %k is the fourier wavelength cutoff, no terms of
                        higher order k are calculated
                        asum=0;
                        bsum=0;
                        for j=-m:1:m;
                                %lambda=xmean+(0.5*j);
                                w=(2*pi)/(l); %w=fundamental frequency
                                x(j+n,1)=cddata1(j+n,1); %wavelength
                                y(j+n,i)=cddata1(j+n,i+1); %data
                                asum=(y(j+n,i)*cos((2*pi*k*j)/(l)))+asum;
                                bsum=(y(j+n,i)*sin((2*pi*k*j)/(l)))+bsum;
                        end
                        a(k,i)=(2*asum)/(l);
                        b(k,i)=(2*bsum)/(l);
                end
        end
        %This section generates the data to be plotted for i=1:1:numberdatasets;
        currentcurve=i+(currentk-1)*numberdatasets;
        %currentcurve is the column vector that will store the fourier series data as
                %it is calculated for a given dataset and corresponding frequency
        anot(1,i)=(sum(cddata1(:,i+1)/l)); %calculation of fourier component
                %'constant' term
        fouriercurve(1:l,1)=0;
        for k=1:1:kcutoff; %filter
                for j=-m:1:m;
                        fouriercurve(j+n,1)=((a(k,i)*cos((2*pi*k*j)/(l)))+(b(k,i)*si
                                n((2*pi*k*j)/(l))))+fouriercurve(j+n,1);
                        fouriercurve2(j+n,currentcurve)=fouriercurve(j+n,1)+(anot(
                                1,i));
                end
        end
end
%This section plots all of the generated data from the previous section
color=strvcat('k','g','b','c','m','r');
if f==kincluded;
        if swich==0;
                'Plotting Program has been turned off'
        elseif swich==1;
                'Plotting Program is on'
                for i=1:1:numberdatasets;
                        if plotallswich==1;
```

```
                        if i==1;
                                continue
                        end
                        currentcurve=i+(kincluded-1)*numberdatasets;
                        datadotcolor=[color((i),1) '.'];
                        fitlinecolor=[color((i),1) '-'];
                        %plot(cddata1(:,1),cddata1(:,i+1),datadotcolor);
                        hold on;
                        plot(x(:,1),fouriercurve2(:,currentcurve),fitlinecolor
                                );
                        hold on;
                        if i==numberdatasets;
                                xlabel('wavelength');
                                ylabel('cd');
                                title('Fit (by fourier analysis) to CD data');
                                text(270,-2.5,'kincluded =')
                                text(287,-2.5,{kincluded})
                        end
                elseif plotallswich==0 && i==dataselect;
                        graphtoprint=dataselect+(kincluded-
                                1)*numberdatasets;
                        datadotcolor=[color((i),1) '.'];
                        fitlinecolor=[color((i),1) '-'];
                        plot(cddata1(:,1),cddata1(:,dataselect+1),datadotcol
                                or);
                        hold on;
                        plot(x(:,1),fouriercurve2(:,graphtoprint),fitlinecolor)
                                ;
                        hold on;
                        axis tight;
                        xlabel('wavelength');
                        ylabel('cd');
                        title('Fit (by fourier analysis) to CD data');
                        text(270,-2.5,'kincluded =')
                        text(287,-2.5,{kincluded})
                end
            end
        end
    end
%-------------------------------------------------------------------------
%This section of the code is a chi-squared analysis to determine the number of fourier
    %components to keep when drawing the best fit curve to the data with the fourier
        %series method
    if chiswich==0;
    elseif chiswich==1;
```

```
            for i=1:1:numberfouriercomponents;
                chisquare(i,1)=i;
                chisquare(i,2)=0;
                chisquareslope(i,1)=i;
                dataseries=dataselect+(i-1)*numberdatasets;
                for j=1:1:l
                    chisquare(i,2)=((fouriercurve2(j,dataseries)-
                        cddata1(j,dataselect+1))^2)+chisquare(i,2);
                    if i>1 && j==l;
                        chisquareslope(i,2)=((chisquare(i,2)-(chisquare(i-
                            1,2)))/(chisquare(i,1)-(chisquare(i-1,1))));
                    end
                end
                if i==numberfouriercomponents;
                    hold off;
                    if slopeswich==0;
                        plot(chisquare(:,1),chisquare(:,2),'k-')
                        axis tight;
                        xlabel('k included');
                        ylabel('chisquare');
                        title('chisquare vs. k included');
                    elseif slopeswich==1;
                        plot(chisquareslope(:,1),chisquareslope(:,2),'k.-')
                        axis ([0 70 -5 0.2]);%tight;
                        xlabel('k included');
                        ylabel('slope');
                        title('slope vs. k included');
                    end
                end
            end
            if f==numberfouriercomponents;
                chisquare=chisquare
                slope=double(chisquareslope)
            end
        end
end
```

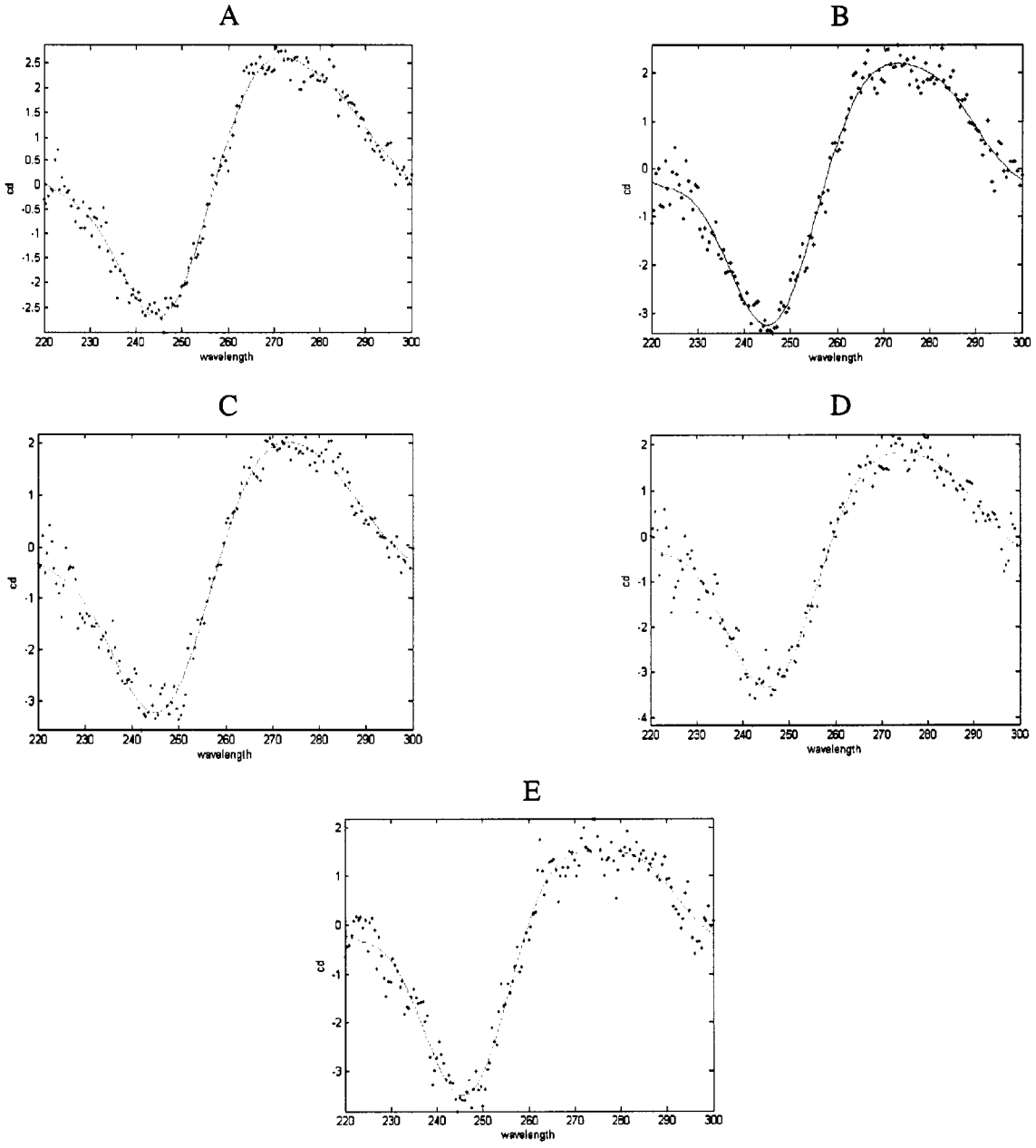


Figure B.1 cd vs. wavelength (λ)
A-E are samples of fits generated by the computer code included in Appendix B (above) to the CD data. In each plot, the dots are experimental data and the solid lines are the fits to the data using a standard fourier series with *kincluded* = 3 (i.e. 7 fourier components total).
(A) 0% EG (B) 10% EG (C) 20% EG (D) 30% EG (E) 40% EG

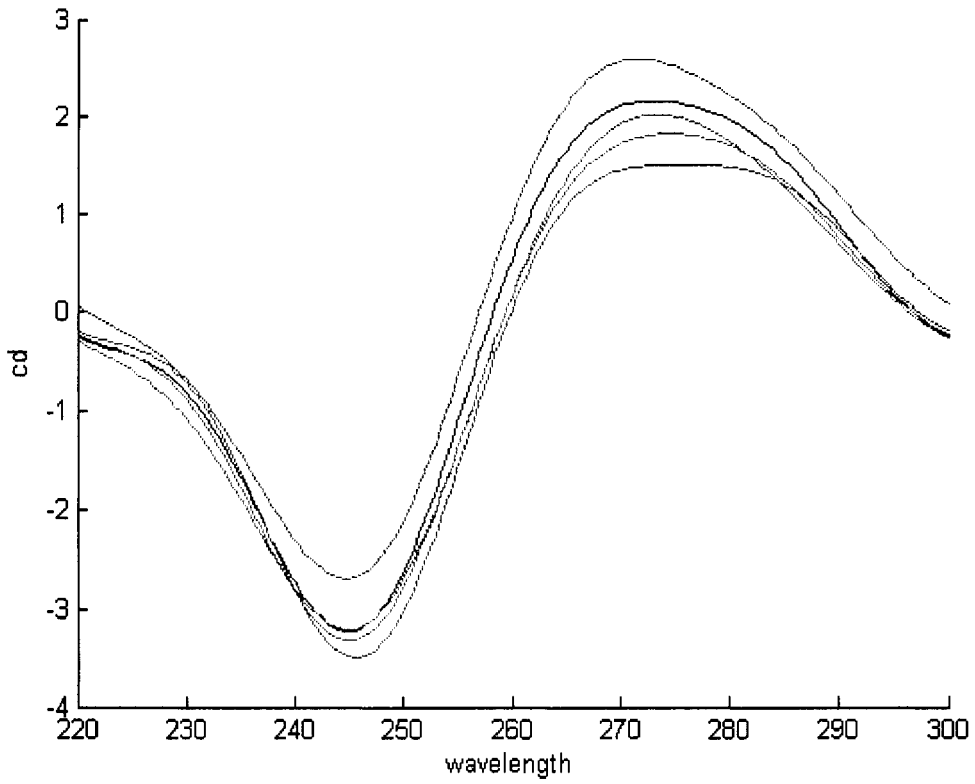
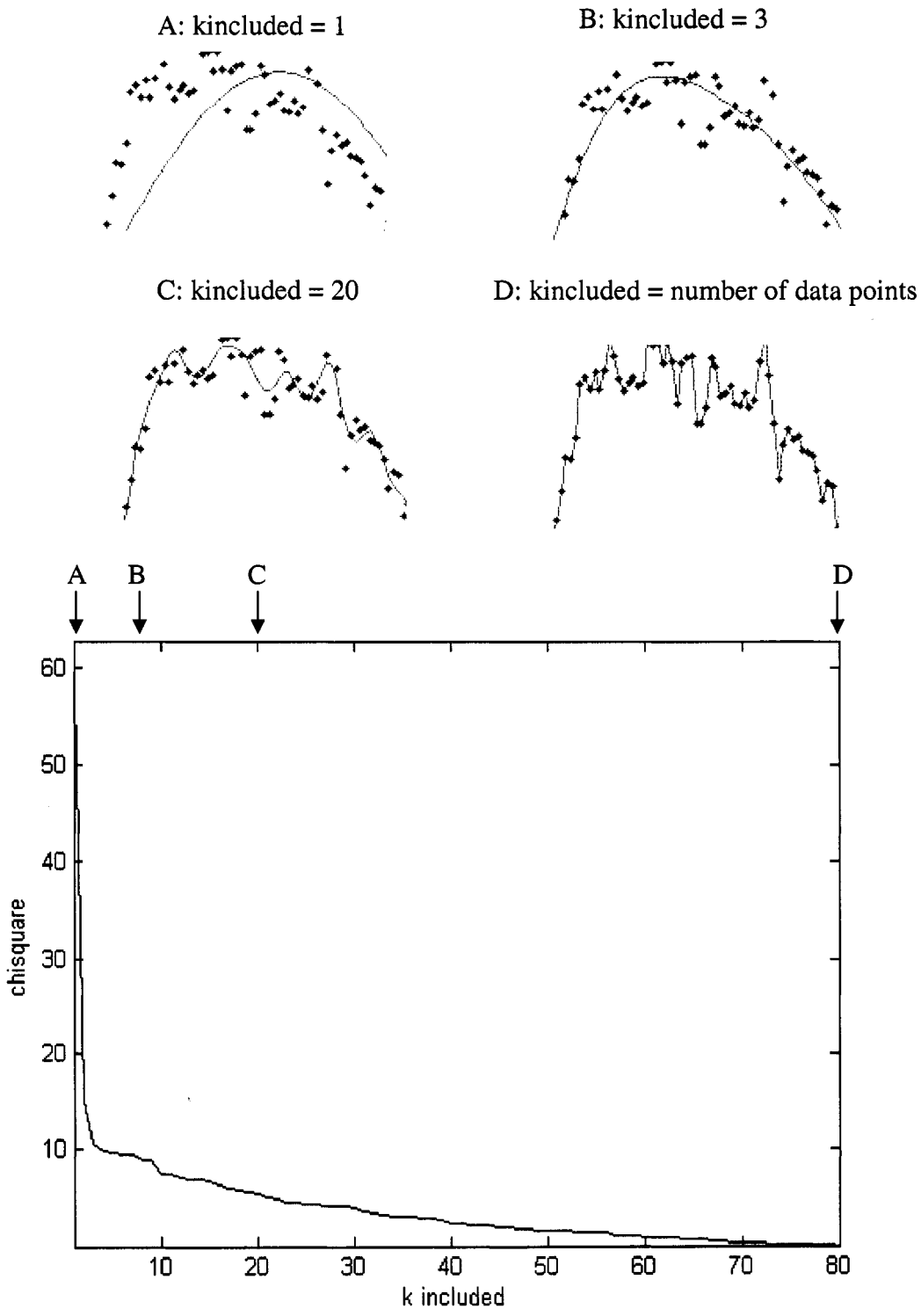


Figure B.2 cd vs. wavelength (λ)
This figure shows the same plots from Figure (B.1) superimposed. A structural change is clearly occurring as observed from the small shift in the fitted CD spectra from 0 – 40% EG. Because the shift is small, this transition likely leaves the DNA in the 'B'-family.

Figure B.3 χ^2 vs. *kincluded*

This figure shows the results from the χ^2 section of the computer code. As more fourier components are included, the fits begin to resemble the actual data. The total number of fourier components included always equals 2·kincluded + 1. A-D show examples from a portion of the curve in Figure B.1A (i.e. 0% added EG) and the arrows below show where these would fall at on the χ^2 vs. *kincluded* curve.

(A) *kincluded* = 1
(B) *kincluded* = 3
(C) *kincluded* = 20
(D) *kincluded* = number of data points


A: kincluded = 1
B: kincluded = 3
C: kincluded = 20
D: kincluded = number of data points
A
B
C
D
chisquare
k included
10
20
30
40
50
60
70
80

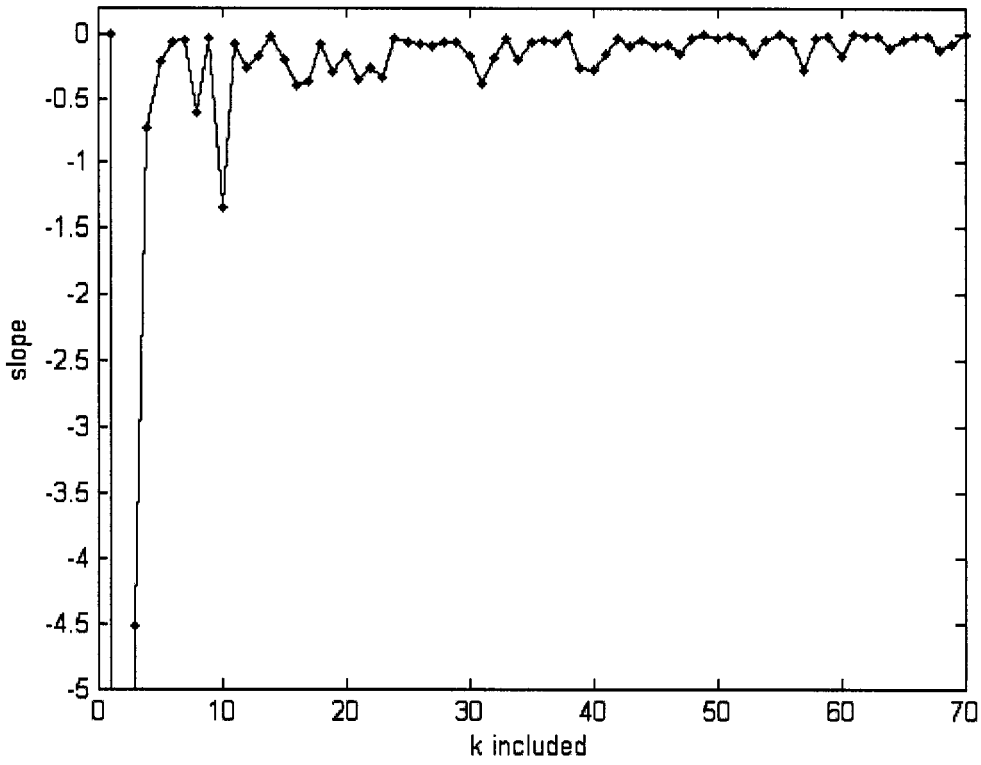


Figure B.4 slope (of χ^2) vs. *kincluded*
This plot shows the results for the slopes calculated from the χ^2 vs. *kincluded* curve shown at the bottom of Figure B.3. The advantage of a plot like this is that it offers more resolution.

Appendix C: Thermodynamic Derivation of the Water Activity (a_w) used in the Analysis of Freezing Point Depression Data

For the topoisomerase I reactions discussed in chapter 2 that extended from 20 – 40% ethylene glycol, experimental measurements of the water activity for the reaction conditions were unavailable. Thus, in order analyze the E_T data as a function of $-\ln(a_w)$, a thermodynamic derivation was employed to obtain a_w from the freezing point depression data that were readily available from the engineering literature (Flick 1998). The result is an analytical expression that is readily solved with the appropriate data (Table C.1). Experimental measurements were made for solutions with 0, 4.8, 9.7, 14.5 and 19.4% EG (Table C.2). For experimental data taken on the most recent preparation of p30δ from 0 – 15% EG, a fitted line to the experimental data of Table C.2 was used to find $-\ln(a_w)$. For those from 15 – 40% EG the derivation below produced the values that were subsequently fitted and relied upon to find $-\ln(a_w)$.

Now, the derivation: At equilibrium

$$\mu_{ice}(T) = \mu_{soln}(T) \tag{C.1}$$

where $\mu_{ice}(T)$ and $\mu_{soln}(T)$ are the chemical potentials of ice and water, respectively, at temperature, T (in Kelvin). Substitution of the Henry's law chemical potential for each phase gives

$$\mu_{ice}^{o}(273) + \int_{273}^{T_f} \frac{\partial \mu_{ice}}{\partial T} dT = \mu_{w}^{o}(273) + \int_{273}^{T_f} \frac{\partial \mu_{w}}{\partial T} dT + RT_f \ln a_w \tag{C.2}$$

which rearranges to

$$-RT_f \ln a_w = \mu_w^o(273) - \mu_{ice}^o(273) + \int_{273}^{T_f} \left(\frac{\partial \mu_w}{\partial T} - \frac{\partial \mu_{ice}}{\partial T} \right) dT \qquad \text{(C.3)}$$

and can be rewritten in the following form as

$$-RT_f \ln a_w = \mu_w^o(273) - \mu_{ice}^o(273) + \int_{273}^{T_f} \left[\frac{\left(\bar{H}_w - \bar{H}_{ice} \right)}{T} \right] dT \qquad \text{(C.4)}$$

The chemical potentials of each phase are *the same at equilibrium* and cancel out to leave

$$-RT_f \ln a_w = -\int_{273}^{T_f} \frac{\left(\bar{H}_w(T) - \bar{H}_{ice}(T) \right)}{T} dT \qquad \text{(C.5)}$$

substituting the following enthalpic relations

$$\begin{aligned} \bar{H}_w(T) &= \bar{H}_w(273) + C_p^w(T - 273) \\ \bar{H}_{ice}(T) &= \bar{H}_{ice}(273) + C_p^{ice}(T - 273) \end{aligned} \qquad \text{(C.6)}$$

produces

$$-RT_f \ln a_w = -\int_{273}^{T_f} \frac{\left[\bar{H}_w(273) - \bar{H}_{ice}(273) + C_p^w(T - 273) - C_p^{ice}(T - 273) \right]}{T} dT \qquad \text{(C.7)}$$

The enthalpy of fusion is defined as

$$\Delta \bar{H}_{fus}(T) = \bar{H}_w(T) - \bar{H}_{ice}(T) \qquad \text{(C.8)}$$

and substituting into the numerator on the right hand side of equation (C.7)

$$\Delta\bar{H}_{fus}(T) = \Delta\bar{H}_{fus}(273) + \left(C_p^w - C_p^{ice}\right)\cdot(T - 273) \qquad \text{(C.9)}$$

and leads to equation (C.10) relating $-\ln(a_w)$ to the readily available $\Delta\bar{H}_{fus}(273), C_p^w, C_p^{ice}$ and R

$$-RT_f \ln a_w = -\int_{273}^{T_f} \frac{\Delta\bar{H}_{fus}(273)}{T} dT - \int_{273}^{T_f} \left(C_p^w - C_p^{ice}\right) dT + \int_{273}^{T_f} \left(C_p^w - C_p^{ice}\right)\cdot\frac{273}{T} dT \qquad \text{(C.10)}$$

The constants R, C_p^w, C_p^{ice}, and $\Delta\bar{H}_{fus}(273)$ are given in Table C.1.

Table C.1 Freezing Point Depression Data from the Industrial Solvents Handbook (Flick 1998). From this data, a_w and $-\ln(a_w)$ were calculated using the derivation in Appendix C. The constants at the bottom of the table were obtained from the CRC Handbook of Chemistry and Physics.

EG%	Freezing Point (°C)	Freezing Point (K)	$-\ln a_w$	a_w
0	0	273.15	0	1.0
2	-0.6	272.55	0.0058230	0.99419
4	-1.3	271.85	0.012636	0.98744
6	-2.0	271.15	0.019469	0.98072
8	-2.7	270.45	0.026323	0.97402
10	-3.5	269.65	0.03418	0.96640
12	-4.4	268.75	0.043057	0.95786
14	-5.3	267.85	0.051966	0.94936
16	-6.3	266.85	0.061907	0.93997
18	-7.3	265.85	0.071891	0.93063
20	-8.0	265.15	0.078907	0.92413
22	-9.0	264.15	0.088966	0.91488
24	-11.0	262.15	0.10922	0.89653
26	-12.0	261.15	0.11941	0.88744
28	-13.0	260.15	0.12965	0.87840
30	-15.0	258.15	0.15026	0.86049
32	-17.0	256.15	0.17106	0.84277
34	-18.0	255.15	0.18153	0.83399
36	-20.0	253.15	0.20261	0.81660
38	-22.0	251.15	0.22388	0.79941
40	-24.0	249.15	0.24534	0.78244

$R = 1.9872$ cal/(mol·K)

$C_p^w = 18.18$ cal/(mol·K)

$C_p^{ice} = 8.64$ cal/(mol·K)

$\Delta\bar{H}_{fus}(273) = 1435.6$ cal/mol

Table C.2 Experimental a_w's measured by vapor pressure osmometry (Rangel et al. 2004)

% EG	$-\ln(a_w)$	a_w
0	0.0110	0.98906
4.8	0.0249	0.97541
9.7	0.0396	0.96117
14.5	0.0543	0.94715
19.4	0.0678	0.93445

Appendix D: Error Analysis to find $<\delta K_o^2>$, $<\delta n^2>$ and $<\delta n \delta K_o>$

A fit of equation (2.14) to the experimental data given in Table D.1, (which corresponds to the black circles in Figure 2.3) resulted in the best fit line shown in Figure 2.3. The plateau values $E_{T1} = 979$ and $E_{T2} = 890$ resulted from a four-dimensional grid search and were held fixed for subsequent analysis. Because these parameters were set to distinct values, the error in the fitted data was attributed to the n and K_o parameters. This analysis was employed to find the variances and covariances in these two parameters. Subsequent to fixing E_{T1} and E_{T2}, the Levenburg-Marquardt fitting routine (with only 2 adjustable parameters) from Matlab®7.0 produced best fit values (that minimized χ^2) of $n = 34$ and $K_o = 0.056$. Because these values of n and K_o minimized χ^2, and because this measure-of-fit parameter will serve as the reference for the error analysis below, it was defined as χ_o^2 (this term is the first term on the right hand side of equation (D.1)). Equation (D.1) is the series expansion of χ^2 through second order in δK_o and δn.

$$\chi^2(K_o,n) = \chi_o^2 + \frac{1}{2}\left(\frac{\partial^2\chi^2}{\partial(\delta K_o)^2}\right)_0 (\delta K_o)^2 + \frac{1}{2}\left(\frac{\partial^2\chi^2}{\partial(\delta n)^2}\right)_0 (\delta n)^2 + 2\cdot\left(\frac{1}{2}\cdot\frac{\partial^2\chi^2}{\partial(\delta K_o)\partial(\delta n)}\right)_0 \delta n \delta K_o \qquad \text{(D.1)}$$

The factors in each term on the right hand side of equation (D.1) that contain second derivatives of χ^2 are the unknowns. These curvatures describe a surface of χ^2 as a function of δK_o and δn and are denoted A, B, and C

$$A = \left(\frac{\partial^2\chi^2}{\partial(\delta K_o)^2}\right)_o$$

$$B = \left(\frac{\partial^2\chi^2}{\partial(\delta n)^2}\right)_o \qquad \text{(D.2)}$$

$$C = \left(\frac{\partial^2\chi^2}{\partial(\delta n)\partial(\delta K_o)}\right)_o$$

To solve for A, δn was set to 0 and a plot of χ^2 vs. $\delta(K_o)^2$ was constructed whose slope is proportional to A. The factor of ½ is added to equation (D.1) to simplify the analysis below that is defined in terms of gaussian integrals (which contain a factor of ½). An analogous plot was constructed for χ^2 vs. $\delta(n)^2$ by setting δK_o to 0 and finding the slope that is proportional to B. C was found by making a plot of χ^2 (after first subtracting the 3 terms on the right hand side of equation (D.1) for different values of δn and δK_o) vs. $\delta n \delta K_o$; the slope of this plot is proportional to C.

In this analysis, the probability distribution function is assumed to be a gaussian in χ^2. Therefore, the probability of observing a particular combination of K_o' and n' for a particular data set, knowing that K_o and n are the values that minimize χ^2, can be expression as the ratio of probability functions by

$$\frac{P\left(K_o', n'\right)}{P\left(K_o, n\right)} = \frac{e^{\left(-\frac{\chi^2}{2}\right)}}{e^{\left(-\frac{\chi_o^2}{2}\right)}} = e^{-\frac{1}{2}\left(\frac{\partial^2 K_o^2}{\partial K_o^2}\right)_o \delta K_o^2} e^{-\frac{1}{2}\left(\frac{\partial^2 K_o^2}{\partial n^2}\right)_o \delta n^2} e^{-\frac{1}{2}\left(\frac{\partial^2 K_o^2}{\partial n \partial K_o}\right)_o \partial n \delta K_o} \tag{D.3}$$

The probability of observing a particular χ^2 for the system is

$$P\left(\delta K_o, \delta n\right) = \frac{e^{-A\delta K_o^2} e^{-B\delta n^2} e^{-C\delta n \delta K_o}}{N} \tag{D.4}$$

where $\delta K_o = K_o' - K_o$, $\delta n = n' - n$ and N is a normalization constant defined as

$$N = \int_{-\infty}^{\infty} d\left(\delta K_o\right) d\left(\delta n\right) \cdot e^{-\frac{1}{2}\left(A\delta K_o^2 + B\delta n^2 + 2C\delta n \delta K_o\right)} \tag{D.5}$$

Equation (D.5) can be converted to a matrix representation with the following form

$$N = \int_{-\infty}^{\infty} dx_1 dx_2 e^{\left(-\frac{\tilde{x}^T M \tilde{x}}{2}\right)} \tag{D.6}$$

where a change in variables substitutes x_1 and x_2 for δK_o and δn, respectively. Thus the x-vector is defined to be

$$\tilde{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; x_1 = \delta K_o, x_2 = \delta n \tag{D.7}$$

and the matrix, M can be defined in terms of the curvatures A, B and C as

$$\tilde{M} = \begin{bmatrix} A & C \\ C & B \end{bmatrix} = \begin{bmatrix} 15428 & 60.5 \\ 60.5 & 0.3794 \end{bmatrix} \tag{D.8}$$

The matrix on the right hand side expresses the 3 variables in terms of the data (i.e. actual numbers). Where applicable in this analysis, the actual numbers will be explicitly written. The matrix Q that diagonalizes M is represented by

$$\tilde{Q} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \begin{bmatrix} -0.0039214 & 1 \\ 1 & 0.003914 \end{bmatrix} \tag{D.9}$$

where the two columns are the eigenvectors that result from diagonalization of M. The x-vectors are redefined in terms of Q and ρ as follows

$$\begin{aligned} \tilde{x} &= \tilde{Q}\tilde{\rho} \\ \tilde{x}^T \tilde{M}\tilde{x} &= \left(\tilde{Q}\tilde{\rho}\right)^T \tilde{M}\tilde{Q}\tilde{\rho} \\ &= \tilde{\rho}^T \tilde{Q}^T \tilde{M}\tilde{Q}\tilde{\rho} \\ &= \tilde{\rho}^T \left[\tilde{Q}^{-1}\tilde{M}\tilde{Q}\right]\tilde{\rho} \end{aligned} \tag{D.10}$$

where ρ is a linear combination of δK_o and δn. Equation (D.6) is expressed in terms of the eigenvalues of M and the elements of the ρ vector by

$$N = \int_{-\infty}^{\infty} dx_1 dx_2 e^{\left(-\sum_{l=1}^{2} \frac{\lambda_l \cdot \rho_l^2}{2}\right)} \tag{D.11}$$

which is rewritten in terms of Q and ρ as

$$N = \int_{-\infty}^{\infty} \frac{d\rho_1 d\rho_2}{|Q|} e^{\left(-\sum_{l=1}^{2} \frac{\lambda_l \cdot \rho_l^2}{2}\right)} \tag{D.12}$$

and simplifies to

$$N = \int_{-\infty}^{\infty} \frac{d\rho_1 d\rho_2}{1.0} \cdot e^{-\frac{1}{2}\left(\lambda_1 \rho_1^2 + \lambda_2 \rho_2^2\right)} \tag{D.13}$$

where |Q| = 1.0 is substituted into equation (D.12) and the two exponential factors are written out explicitly. Upon solving the integral in equation (D.13) the following expressions result for the normalization constant.

$$
\begin{aligned}
N &= \frac{1}{1.0}\sqrt{\frac{2\pi}{\lambda_1}}\sqrt{\frac{2\pi}{\lambda_2}} \\
&= \frac{2\pi}{\sqrt{\lambda_1\lambda_2}} \\
&= \frac{2\pi}{\sqrt{|\lambda_1\lambda_2|}} \\
&= \frac{2\pi}{\sqrt{|\tilde{M}|}}
\end{aligned}
\tag{D.14}
$$

Thus, the normalization constant, N, is inversely related to the determinant of the matrix, M.

The variance in K_o is defined as

$$
\left\langle \delta K_o^2 \right\rangle = \frac{\int_{-\infty}^{\infty} d\delta K_o d\delta n \left(e^{-\frac{\left(A\delta K_o^2 + B\delta n^2 + C\delta n \delta K_o\right)}{2}} \right) \delta K_o^2}{N} \tag{D.15}
$$

which gives equation (D.16) upon transformation

$$
\left\langle \delta K_o^2 \right\rangle = \frac{1}{N}\int_{-\infty}^{\infty} d\rho_1 d\rho_2 e^{-\left(\frac{\lambda_1\rho_1^2}{2}\right)} e^{-\left(\frac{\lambda_2\rho_2^2}{2}\right)} \delta K_o^2 \tag{D.16}
$$

The deviations in K_o and n are represented as

$$\begin{vmatrix} \delta K_o \\ \delta n \end{vmatrix} = \tilde{Q} \begin{vmatrix} \rho_1 \\ \rho_2 \end{vmatrix} = \begin{vmatrix} Q_{11}\rho_1 + Q_{12}\rho_2 \\ Q_{21}\rho_1 + Q_{22}\rho_2 \end{vmatrix} \tag{D.17}$$

which implies that

$$\begin{aligned} \delta K_o &= Q_{11}\rho_1 + Q_{12}\rho_2 \\ \delta n &= Q_{21}\rho_1 + Q_{22}\rho_2 \end{aligned} \tag{D.18}$$

The outcome of substituting equation (D.18) into equation (D.16) and performing the integral is equation (D.19)

$$\left\langle \delta K_o^2 \right\rangle = \frac{1}{N}\left[Q_{11}^2 \cdot \sqrt{\frac{2\pi}{\lambda_1}} \cdot \frac{1}{\lambda_1} \cdot \sqrt{\frac{2\pi}{\lambda_2}} + Q_{12}^2 \cdot \sqrt{\frac{2\pi}{\lambda_2}} \cdot \frac{1}{\lambda_2} \cdot \sqrt{\frac{2\pi}{\lambda_1}} \right] \tag{D.19}$$

Thus, the variance in K_o (as well as the variance in n and the covariance) is found by finding the eigenvalues and eigenvectors of the matrix, M (that was defined in terms of the curvatures for χ^2). Expression (D.19) simplifies to

$$\left\langle \delta K_o^2 \right\rangle = Q_{11}^2\left(\frac{1}{\lambda_1}\right) + Q_{12}^2\left(\frac{1}{\lambda_2}\right) = 1.73 \cdot 10^{-4} \tag{D.20}$$

An analogous path is followed to find the variance in n and yields the following expression

$$\left\langle \delta n^2 \right\rangle = Q_{21}^2\left(\frac{1}{\lambda_1}\right) + Q_{22}^2\left(\frac{1}{\lambda_2}\right) = 7.04 \tag{D.21}$$

The covariance is given by

$$\left\langle \delta n \delta K_o \right\rangle = Q_{11}Q_{21}\left(\frac{1}{\lambda_1}\right) + Q_{12}Q_{22}\left(\frac{1}{\lambda_2}\right) = -0.0276 \tag{D.22}$$

Table D.1 Experimental E_T's measured by the topoisomer distribution method. This data corresponds to the most recent topoisomerase reactions that are shown as black circles in Figure 2.3. a_w's were found using the method of Appendix C.

a_w	$-\ln(a_w)$	E_T	σ_{ET}
0.9891	0.011	969.29	10.49
0.9598	0.041	958.01	9.860
0.9399	0.062	957.59	7.820
0.9277	0.075	916.86	24.86
0.9148	0.089	924.72	8.470
0.8690	0.120	907.56	8.100
0.8547	0.157	899.81	4.300
0.8204	0.198	885.63	9.390
0.8017	0.221	900.84	6.030
0.7827	0.245	873.68	6.250

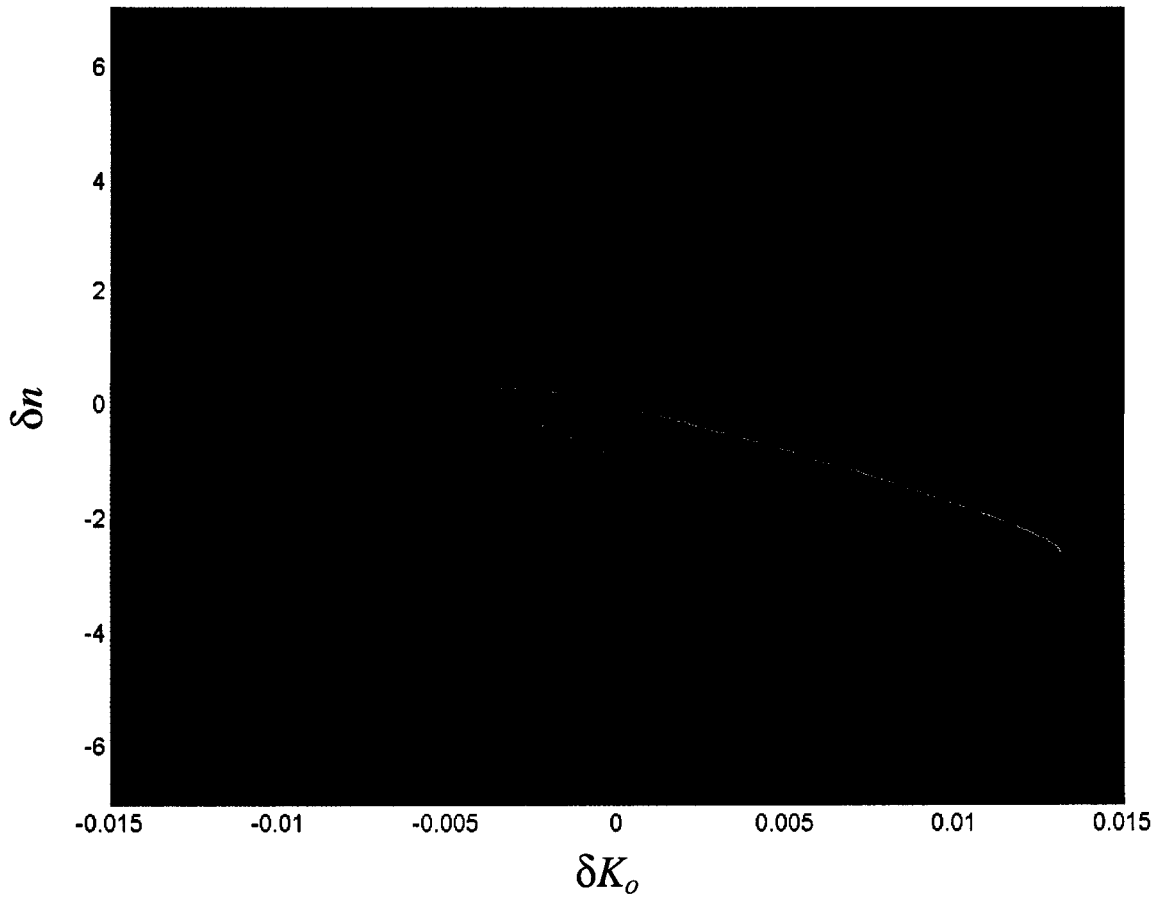


Figure D.1 δn vs. $\delta K_o - \chi^2$ landscape

This figure shows the δn vs. δK_o surface of χ^2. δn and δK_o are plotted in the plane of the page and the point $\delta n = 0$, $\delta K_o = 0$ represents the parameters $n = 34$ and $K_o = 0.056$ that minimized χ^2 in the fitting protocol. The axis coming out of the plane is a measure percent difference in χ^2 between the measured values n' and K_o' with those that minimized χ^2 to give χ_o^2 (i.e. $n = 34$ and $K_o = 0.056$). The three colored regions are distinctly different and are classified by the following scheme; the blue area corresponds to χ^2 values that were greater than 10% higher than χ_o^2, the red area is where χ^2 was between 0 and 10% larger (inclusive) and the green area is where χ^2 was less than χ_o^2. Apparently, the Levenburg-Marquardt fitting protocol located a minimum that wasn't the global minimum (at the level of resolution it used to search the grid). The global minimum from this search, that lies in the region shaded green, was less than 1% different than χ_o^2 and is still within the error bars of the fitted n and K_o parameters found by the fitting protocol. Inserting the global minimum would provide very little improvement to those conclusions reported herein. The two perindicular black lines correspond to the standard deviations (found by taking the square root of the variances reported in Appendix D) in n and K_o which are $n \pm \delta n = 34 \pm 2.7$ and $Ko \pm \delta Ko = 0.056 \pm 0.013$.

Appendix E: Cooperative vs. Non-Cooperative Models for Evaluating Structural Transitions

Hydration-coupled structural transitions in DNA can be analyzed by a two-state model wherein the DNA is represented as a linear chain of subunits that can adopt either of two conformations. An example of such a model is

...1111222222211111111111222222221122211222...

where a '1' represents the DNA in any state 1 and a '2' represents the DNA in other state 2. The basic requirement is that states 1 and 2 reside in different potential wells. The transfer matrix (M) associated with a transition of this type would be

$$M = \begin{vmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{vmatrix} = \begin{vmatrix} 1 & JB \\ J \cdot 1 & B \end{vmatrix} \tag{E.1}$$

where B is the equilibrium constant for the unimolecular reaction occurring in a structural transition between states 1 and 2 at a single base-pair,

$$\begin{aligned} B &= \frac{[2]}{[1]} \\ &= K_o a_w^{-\Delta\Gamma} \end{aligned} \tag{E.2}$$

and $J \equiv \exp\left[\frac{-F_J}{RT}\right]$, where F_J is the junction free energy. J is a parameter that is inversely related to the average length of each domain (d) in the molecule, $\frac{1}{J} \sim (d-1)$, at the midpoint of the transition. From this foundation, an equation relating the fraction of molecules in state 2 (θ_2) to the parameters of the model is given by

$$\langle \theta_2 \rangle = \frac{B}{\lambda_+}\frac{\partial \lambda_+}{\partial B}$$
$$= \frac{1}{2}\left(1 + \frac{B-1}{\left[(B-1)^2 + 4 \cdot B \cdot J^2 \right]^{\left(\frac{1}{2}\right)}} \right) \tag{E.3}$$

where λ_+ is the eigenvalue from the determinant$|M - \lambda I| = 0$. Substitution of equation (E.2) into the above equation gives

$$\langle \theta_2 \rangle = \frac{1}{2}\left(1 + \frac{K_o a^{-\Delta\Gamma} - 1}{\left[\left(K_o a^{-\Delta\Gamma} - 1 \right)^2 + 4 \cdot K_o a^{-\Delta\Gamma} \cdot J^2 \right]^{\left(\frac{1}{2}\right)}} \right) \tag{E.4}$$

The twist energy parameter is related to the fraction of the molecules in each state by

$$E_T = \left[\frac{1-\theta_2}{E_{T1}} + \frac{\theta_2}{E_{T2}} \right]^{-1} \tag{E.5}$$

The slope near the midpoint of the θ_2 vs. ln(B) plot can be found by taking the derivative

$$\frac{d\theta_2}{d\ln B} = B \cdot \frac{d\theta_2}{dB}$$
$$= \frac{B \cdot (B+1) \cdot J^2}{\left(4 \cdot J^2\right)^{\frac{3}{2}}} \tag{E.6}$$

Near the midpoint, $\theta_2 = 1/2$ and B = 1 becomes

$$\left.\frac{d\theta_2}{d\ln B}\right|_{\theta_2 = \frac{1}{2}} = \frac{1}{4J} \tag{E.7}$$

After substituting equation (E.2) for B and introducing $\Delta\Gamma \equiv -p$ equation (E.7) becomes

$$\left.\frac{d\theta_2}{d\ln a_w}\right|_{\theta_2 = \frac{1}{2}} = \frac{-p}{4J} \tag{E.8}$$

When $\theta_2 = 1/2$ and B = 1, $K_o\left(a_w\right)^{-p} = 1$ and $\ln\left(K_o\right) - p\ln\left(a_w\right) = 0$, so the relation

$$\left.\ln a_w\right|_{\theta_2 = \frac{1}{2}} = \left(\frac{1}{p}\right)\ln K_o \tag{E.9}$$

is obtained. The midpoint values of $\left.\ln a_w\right|_{\theta = \frac{1}{2}}$ and $\left.\frac{d\theta_2}{d\ln a_w}\right|_{\theta_2 = \frac{1}{2}}$ together with equations (E.8) and (E.9) provide *two* relations among the *three* variables, p, K_o, and J. Hence, the central part of the transition provides no unique determination of all three parameters. However, for any given trial value of J, a unique determination of p and K_o can be made. Near the midpoint, one finds

$$\frac{dE_T}{d\ln a_w} = \frac{dE_T}{d\theta_2}\frac{d\theta_2}{d\ln B}$$
$$= \left(\left.\frac{dE_T}{d\theta_2}\right|_{\theta_2=\frac{1}{2}}\right)\cdot\left(\left.\frac{d\theta_2}{d\ln a_w}\right|_{\theta_2=\frac{1}{2}}\right)\cdot\left(\left.\frac{d\ln a_w}{d\ln B}\right|_{\theta_2=\frac{1}{2}}\right) \tag{E.10}$$

where the first term in the product can be found using equation (E.6), the second term arises from equation (E.9) and the last term is unity. Near the midpoint this simplifies to

$$\frac{dE_T}{d\ln a_w} = \left(\frac{1}{E_{T1}} - \frac{1}{E_{T2}}\right)\cdot\left(\frac{0.5}{E_{T1}} + \frac{0.5}{E_{T2}}\right)^{-2}\cdot\left(\frac{-p}{4J}\right) \tag{E.11}$$

The average number of junctions at the midpoint is given by

$$\langle n_J\rangle = \left(\frac{\partial \ln \lambda_+^N}{\partial \ln J}\right)_{\theta_2=1/2}$$
$$= N\frac{J}{\lambda_+}\left(\frac{\partial \lambda_+}{\partial J}\right)_{\theta_2=1/2} \tag{E.12}$$
$$= N\frac{J}{J+1}$$

Finally the average domain size between junctions at the midpoint is given by

$$d = \frac{N}{\langle n_J\rangle}$$
$$= 1+\frac{1}{J} \tag{E.13}$$

hence 1/J = d−1 and

$$\frac{p}{4J} = \frac{1}{4}\left(\frac{p}{d-1}\right) \tag{E.14}$$

An estimate of the quantity $\frac{p}{4J}$ from the midpoint slope in equation (E.11) yields ¼ of the $\Delta\Gamma$ per cooperative domain. Unfortunately, the cooperative domain is not known. The midpoint slope, $\left.\frac{dE_T}{d\ln a_w}\right|_{\theta_2=\frac{1}{2}} \cong 763$, in this case yields via equation (E.11) $\frac{p}{4J} = 8.5$, whence $\frac{p}{J} \simeq 34$. If one assumes that $p \leq 1.0$, then $J \leq \frac{1}{34}$, and $d - 1 \geq 34$.

Appendix F: Sample Calculation of α_2 and $\kappa_{\beta 2}$

Appendix F shows a sample calculation of the elastic properties of states 1 and 2 from FPA measurements near the midpoint of the transition observed in Figure 2.3. Rangel et al. found that the torsion elastic constant increased by 1.37-fold between 0 and 19.4 % EG, which corresponds closely to the midpoint of the transition. This means that $\alpha_{eff}\Big|_{\theta=\frac{1}{2}} = (1.37)\cdot\alpha_1$. The ratio of α_2 to α_1 was found as follows:

$$\alpha_{eff}\Big|_{\frac{1}{2}} = 2\left(\frac{1}{\alpha_1}+\frac{1}{\alpha_2}\right)^{-1} = 2\left(\frac{\alpha_1+\alpha_2}{\alpha_1\alpha_2}\right)^{-1} \tag{F.1}$$

from equation (2.11) with $f_1 = f_2 = 0.5$ at the midpoint of the transition. Let R_1=1.37 so that

$$R_1\alpha_1 = \frac{\alpha_1\alpha_2}{\left(\frac{\alpha_1+\alpha_2}{2}\right)} \tag{F.2}$$

and rearrange to solve for α_2

$$\alpha_2 = \frac{R_1\left(\frac{\alpha_1}{2}\right)}{\left(1-\left(\frac{R_1}{2}\right)\right)} = \frac{1.37\left(\frac{\alpha_1}{2}\right)}{\left(1-\left(\frac{1.37}{2}\right)\right)} \tag{F.3}$$

which gives

$$\alpha_2 = 2.175\alpha_1 \tag{F.4}$$

DNA(2) has a torsion elastic constant that is 2.175-fold larger than that of DNA(1). The significance of this finding is discussed in chapter 2.

Knowing the ratio of α_2 to α_1 allows one to solve for the ratio of the bending elastic constants of states 1 and 2 via equation (2.2) where $E_{T1} = 979$ and $E_{T2} = 890$ are known from the fit in Figure 2.3.

$$E_{T1} = \left(\frac{2\pi^2}{2k_BT}\right)\left(\frac{\alpha_1 B\kappa_{\beta 1}}{\alpha_1 + B\kappa_{\beta 1}}\right)$$
$$E_{T2} = \left(\frac{2\pi^2}{2k_BT}\right)\left(\frac{\alpha_2 B\kappa_{\beta 2}}{\alpha_2 + B\kappa_{\beta 2}}\right) \tag{F.5}$$

Taking the ratio of these two equations and rearranging to find $\frac{\kappa_{\beta 1}}{\kappa_{\beta 2}}$ gives

$$\frac{\kappa_{\beta 2}}{\kappa_{\beta 1}} = \left(\frac{\alpha_2}{\alpha_1}\right)\left(\frac{E_{T2}}{E_{T1}}\right)\frac{\left(\left(\frac{(2\pi)^2}{2k_BT}\right)\alpha_1 - E_{T1}\right)}{\left(\left(\frac{(2\pi)^2}{2k_BT}\right)\alpha_2 - E_{T2}\right)} \tag{F.6}$$

which simplifies to

$$\kappa_{\beta 2} = 0.833\kappa_{\beta 1} \tag{F.7}$$

Thus, DNA(2) has an 0.833-fold lower bending elastic constant than does DNA(1). The persistence length decreases due to the relation

$$P = \frac{\kappa_\beta}{k_bT} \tag{F.8}$$

$$P_2 = \frac{\kappa_{\beta 2}}{\kappa_{\beta 1}} P_1 \tag{F.9}$$

so for a $P_1 = 500$Å, $P_2 = 416.5$Å.

VITA

Greg P. Brewood was born September 17, 1976 in Kansas City, Kansas to Dale and Patricia Brewood. He graduated from Shawnee Mission South High School in 1995. In 1997, Greg graduated from Johnson County Community College with an Associate of Arts (AA) degree. He then attended Kansas State University in Manhattan, Kansas where he double majored in Chemistry/Pre-medicine and graduated Cum Laude with a Bachelor of Science (BS) in the spring of 2000. He is three credits shy of earning a minor in Spanish as well. Greg subsequently entered graduate school at the University of Washington in the fall of 2000. In 2002, he earned a Master of Science (MS) degree in Physical Chemistry from the University of Washington and received his Doctor of Philosophy (PhD) in Physical/Biophysical Chemistry in the spring of 2006.