UNIVERSITY OF CALIFORNIA, SAN DIEGO

DNA Tethering Characterization, Enzyme-Mediated DNA Looping Under Tension, and Nucleosome Stability in the Force Measuring Optical Tweezers

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Physics

by

Gregory John Gemmen

Committee in charge:

Professor Douglas E. Smith, Chair
Professor Robert N. Dutnall
Professor James T. Kadonaga
Professor Melvin Y. Okamura
Professor Sunil K. Sinha

2006

The dissertation of Gregory John Gemmen is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2006

DEDICATION

This dissertation and the time, effort, talents, and sacrifices that it represents are dedicated in loving memory to my late grandmother, Jeanne Faasse Karelse.

EPIGRAPH

There are strange instances of it, indeed; an honest man is sometimes, for the sake of being original, ready to do something base. It sometimes happens that one of these luckless men is not only honest but good, is the guardian angel of his family, maintains by his labour outsiders as well as his own kindred, and yet can never be at rest all his life! The thought that he has so well fulfilled his duties is no comfort or consolation to him; on the contrary, it irritates him. "This is what I've wasted all my life on," he says; "this is what has fettered me, hand and foot; this is what has hindered me from doing something great! Had it not been for this, I should certainly have discovered – gunpowder or America, I don't know precisely what, but I would certainly have discovered it!" What is most characteristic of these gentlemen is that they can never find out for certain what it is that they are destined to discover and what they are within an ace of discovering. But their sufferings, their longings for what was to be discovered, would have sufficed for a Columbus or a Galileo.

from *The Idiot*, Fyodor Dostoevsky

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

Symbols that appear consistently throughout this dissertation are listed under the chapter in which they are first introduced and in the order in which they are introduced. Symbols and/or abbreviations that are universally recognized, such as 'λ' for wavelength or 'bp' for base pair, are not listed.

Chapter 1

PCR: Polymerase chain reaction, a technique in which DNA samples are amplified by a repeated sequence of melting, annealing, and polymerization.

DIG-αDIG: The body-antibody reaction between digoxigenin and antidigoxigenin used to tether DNA molecules; the pipetted bead is coated with the latter.

SA: Streptavidin. This coats the optically trapped bead in the experiments and binds with biotin, facilitating DNA tethering.

L1, L2, L3, L4: Plano-convex lenses in the tweezers system.

ND1, ND2: Neutral density filters in the tweezers system.

O1, O2: Objective lenses in the tweezers system.

FC: Flow cell in the tweezers system.

M, DM: Mirror, dichroic mirror in the tweezers system.

BS: Beam splitting cube in the tweezers system.

XYZ: 3-D micrometer stage in the tweezers system.

LP, SP: Long pass, short pass filter in the tweezers system.

PSD: Position sensitive detector in the tweezers system.

I/V: The illumination/visualization aspect of the tweezers system.

CCD: Charge coupled device, a camera in the I/V aspect of the tweezers system.

V1, V2, V3, V4: Valves in the fluidics system.

C1, C2: Crimpers in the fluidics system.

<>: In the discussion of the control program, brackets denote a keystroke. In all other chapters, brackets represent the expected value of that quantity from a distribution.

Chapter 2

Δx: The uncertainty in the tether length due to the rotationally constrained pipette bead.

WLC: Worm Like Chain, a model used to describe DNA dynamics theoretically.

L_0: The measured contour length of a DNA molecule in the WLC model.

P: Persistence length, a parameter in the WLC model characterizing rigidity.

S: Stretch modulus, a parameter in the WLC model characterizing elasticity.

k_{off}: The measured dissociation rate of the DNA tethers under tension.

d: Distance from the free energy minimum to the peak of the barrier for the dissociation.

DFS: Dynamic force spectroscopy.

Chapter 3

ΔL: The length of DNA in a loop, or the observed length increase upon disrupting a loop.

REase: Restriction endonuclease.

Chapter 4

N_{pairs}: The number of possible loops for an enzyme on a DNA template.

N_{sites}: The number of recognition sites for an enzyme on a DNA template.

$\rho_{x,y}$: The linear correlation coefficient between x and y (e.g., between force and ΔL).

d: The protein bridging distance for the enzyme mediated loops.

Chapter 5

k_{eff}: An effective rate constant for the DNA cleavage/looping reaction.

1X: Concentration nomenclature for 1 μL stock enzyme in 99 μL reaction buffer.

t_{inc}: Incubation time for which the DNA molecule was held in the experiments.

x/L_0: Fractional extension at which the DNA molecule was held in the experiments.

N: The number of loops per trial when used as an observable.

P_{meas}, P_{false}, P_{real}: The measured distribution, the distribution of false events, and the corrected distribution (either N or ΔL).

δ(ΔL): The length variable representing the shift from the start of the ΔL tail section.

d: The proposed interaction distance between the protein monomers.

ε: The proposed binding energy between the protein monomers.

Chapter 6

ACF: ATP dependent chromatin assembly factor.

NAP-1: Nucleosome assembly protein.

H2A, H2B, H3, H4: Core histones.

H1, B4: linker histones.

NCP: nucleosome core particle.

D_{max}: The maximal diameter of an NCP in X-ray scattering experiments.

F_{max}: The cutoff force above which a data run had to extend in order to be kept.

Chapter 7

δ: The distance between the entering/exiting DNA in a nucleosome.

HE: Homing endonuclease.

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 5

Chapter 6

Chapter 7

LIST OF TABLES

Chapter 2

Chapter 3

Chapter 4

Chapter 5

ACKNOWLEDGEMENTS

I am indebted to a number of people for helping me both professionally and personally. Regarding the former, I wish to thank my advisor Doug Smith first and foremost. His enthusiasm for the experiments was relentless and contagious. He taught me some very valuable lessons about dedication to the work. He was also patient with me through some very difficult times within and outside of the lab, all the while affording me the freedom to manage the projects in the way that best suited me. Though we did not always agree on things, I consider myself very fortunate to have worked with someone whose approach usually meshed well with my own.

I also wish to thank Rachel Millin. Rachel carried out many of these tweezers experiments, allowing me to attend to project management and data analysis issues fully. She also made it possible to take the data in shifts, allowing the rather statistically intensive results to be collected and analyzed in a reasonable time. I also wish to thank Ron Sim. For nine months he chipped away at the nucleosome assembly which at many times seemed intractable. Regarding that work, we are indebted to the assistance of Karl Haushalter, Dmitry Fyodorov, Vassili Alexiadis, Alex Lusser, Buyung Santoso, Susan Cheng, and Jim Kadonaga. Besides the chromatin assembly, Ron also labeled DNA on many occasions, made pipettes/flow cells, helped collect the chromatin data, and hand-clicked the many thousands of events in that data.

I next wish to thank Karl Haushalter, a former postdoc in Dr. Kadonaga's lab who worked on the nucleosome project. Karl did much of the initial work for the DNA labeling protocols, bead preparations, and chromatin assembly. He taught me the basics of biological bench work, and was always willing and eager to answer my questions no matter how befuddled my thinking. Even after he left UCSD, Karl actively kept up communication with me to facilitate the completion of the projects. I also wish to thank Pu Chun Ke, now of Clemson University. He was a post doc in the group who largely set up the tweezers and taught me hands-on optics. Like Doug, Pu Chun modeled an exceptional work ethic, for which I am likewise grateful to him.

I also wish to thank Al Schweitzer and Peter Rickgauer for their help in characterizing the drift of the tweezers and taking the nucleosome measurements, respectively. It has been a pleasure returning the favor so to speak, sharing with them what I have learned to assist them with their dual beam tweezers setup. I also thank Pierre Recouvreux and Aurélie Dupont (University of Lyon), who initiated the PCR labeling of DNA. Their work facilitated the study of many more of the two site restriction enzymes than otherwise would have been possible. Following from that, I thank Derek Fuller for extending their work to different templates.

I also wish to thank Arun Stephen Rahjkumar for doing a host little things around the lab, Kendra Hailey (UCSD Chemistry Department) for preparing a sample of labeled DNA while on rotation, Brian Rankin (University of Göttingen) for clicking much of the two site restriction enzyme peaks data, and Angela Foudrey (Stanford University) for her assistance in writing the LabVIEW control program some four years ago. Special thanks go to a truly special person, Debra Bomar, who was understanding and fantastically helpful before and during my UCSD tenure. Finally, I wish to thank the National Defense Science and Engineering Graduate

Fellowship program for supporting me my first two years at UCSD. While I have tried to mention all the specific contributions made by others to my work efforts, I apologize to those whose contributions I have overlooked here. No matter how large or small, I appreciated them then and I appreciate them still.

All that said, the science is the easy part. The long hours sequestered in a lab working on something about which very few people care, the insolent poverty, the profound opportunity costs of deferred career and investment, the unrelenting weight of being "unfinished", the relationships stressed and broken beyond repair…these have been the real burdens and sacrifices of graduate school. Some might call them tests of mettle and perseverance; others might call them proof of stubborn idiocy and a depressive miasma. Either way, they form a ledger sheet that I could not have balanced on my own, and I am profoundly grateful for all those who have supported and directed me spiritually and emotionally over the past several years.

First and foremost, I am thankful to God for the talents, direction, opportunity, and, in hindsight, the heavy trials. I am grateful to my family. Despite being far away, their emotional support and belief in me was palpable and unwavering. I believe that they knew that I had to do this no matter what; their understanding in this regard has always been deeply appreciated. I hope that finishing this dissertation vindicates their patience and support, and makes them quietly proud.

I am also grateful to Donna Rottier and her family. Losing her taught me more about myself than I ever wanted to learn, but she stuck with me for as long as she could and I thank her for that. I will love her always and I hope there is some solace in that for her and her family. I wish to thank Amy Files and her family. None of this would have gotten done without her loving me so simply and buoying me up through some really trying times with the work.

I am especially grateful to Joe McGovern, Jared Bradley, Jo Bellanca, Noëlle Yochum, and Emma Neff, whose friendship through some really difficult times will never be forgotten. I am grateful to the Rego family, especially Chris and John, who are the most generous and open-hearted people I have ever known. I wish to thank my friends from Pacific Hope Church and Coastlands Church, especially Pastor Evan Lauer. Another community that I am happy to call my own is TDAC Lokomotiv Soccer Club (in its many manifestations over the years). All the people with whom I have regularly surfed and climbed (Lando Pasimio and Mark "Shreveport" Clement, especially) over the last few years also deserve a solid thank-you. There have been so many people that have had a hand in helping me keep my sanity loosely intact and my spirits adequately high, enriching my life in so many ways outside of the lab and the university; they made the completion of this dissertation possible. Even if not mentioned here by name, I am graciously and sincerely indebted to each of them.

The text of Chapter 2, in part, is a reprint of the material as it appears in Fuller, D.N, Gemmen, G.J., Rickgauer, J.P., Dupont, A., Millin, R., Recouvreux, P., & Smith, D.E. (2006). A General Method for Manipulating DNA Sequences from any Organism with Optical Tweezers. *Nuc. Acids Res.* **34**: e15. The dissertation author was a primary researcher and secondary author for the research which forms the basis of Chapter 2.

The text of Chapter 3, in full, appears in Gemmen, G.J., Millin, R. & Smith, D.E. (2006). Tension-Dependent DNA Cleavage by Restriction Endonucleases: Enzymes Binding at Two Sites are "Switched Off" at Low Force. Submitted to *The Proceedings of the National Academy of Sciences, U.S.A*. The dissertation author was a primary researcher and secondary author for the research which forms the basis of Chapter 3.

The text of Chapter 4, in full, appears in Gemmen, G.J., Millin, R. & Smith, D.E. (2006). DNA Looping by Two-Site Restriction Endonucleases: Heterogeneous Probability Distributions for Loop Size and Unbinding Force. Submitted to *Nucleic Acids Research*. The dissertation author was a primary researcher and secondary author for the research which forms the basis of Chapter 4.

The text of Chapter 5, in part, appears in Gemmen, G.J., Millin, R. & Smith, D.E. (2006). Dynamics of Single DNA Looping and Cleavage by Restriction Endonuclease Sau3AI Measured with Optical Tweezers. Submitted to the *Biophysical Journal*. The dissertation author was a primary researcher and author for the research which forms the basis of Chapter 5.

The text of Chapter 6, in part, is a reprint of the material as it appears in Gemmen, G. J., Sim, R., Haushalter, K.A., Ke, P.C., Kadonaga, J.T., & Smith, D.E. (2005). Forced Unraveling of Nucleosomes Assembled on Heterogeneous DNA Using Core Histones, NAP-1, and ACF. *J. Mol. Biol.* **351**: 89-99. The dissertation author was a primary researcher and author for the research which forms the basis of Chapter 6.

VITA

1994	B.S., Hope College, Holland, MI
1996	M.S., Northwestern University, Evanston, IL
1998 – 2000	Research Assistant, Department of Physics University of California, San Diego
2000 – 2001	Staff Research Associate, Department of Bioengineering University of California, San Diego
2001 – 2006	Research Assistant, Department of Physics University of California, San Diego
2006	Ph.D., University of California, San Diego

PUBLICATIONS

Gemmen, G.J., Millin, R. & Smith, D.E. (2006). Dynamics of Single DNA Looping and Cleavage by Restriction Endonuclease Sau3AI Measured with Optical Tweezers. Submitted to the *Biophysical Journal*.

Gemmen, G.J., Millin, R. & Smith, D.E. (2006). DNA Looping by Two-Site Restriction Endonucleases: Heterogeneous Probability Distributions for Loop Size and Unbinding Force. Submitted to *Nucleic Acids Research*.

Gemmen, G.J., Millin, R. & Smith, D.E. (2006). Tension-Dependent DNA Cleavage by Restriction Endonucleases: Enzymes Binding at Two Sites are "Switched Off" at Low Force. Submitted to *The Proceedings of the National Academy of Sciences, U.S.A*.

Fuller, D.N, Gemmen, G.J., Rickgauer, J.P., Dupont, A., Millin, R., Recouvreux, P., & Smith, D.E. (2006). A General Method for Manipulating DNA Sequences from any Organism with Optical Tweezers. *Nuc*. *Acids Res*. **34**: e15.

Smith, D.E., Gemmen, G.J., Millin, R., Rickgauer, J.P., Schweitzer, A.L., & Fuller, D.N. (2005). Using Optical Tweezers to Study Protein-DNA Interactions. *Proc*. *SPIE Int*. *Soc*. *Opt*. *Eng*. **5930**: 593012.

Fuller, D.N, Gemmen, G.J., Rickgauer, J.P., Dupont, A., Millin, R., Recouvreux, P., Schweitzer, A.L., & Smith, D.E. (2005). A General Method for Manipulating DNA

Sequences from any Organism with Optical Tweezers. *Proc*. *SPIE Int*. *Soc*. *Opt*. *Eng*. **5930**: 593013.

Gemmen, G. J., Sim, R. , Haushalter, K.A., Ke, P.C., Kadonaga, J.T., & Smith, D.E. (2005). Forced Unraveling of Nucleosomes Assembled on Heterogeneous DNA Using Core Histones, NAP-1, and ACF. *J*. *Mol*. *Biol*. **351**: 89-99.

Gemmen, G.J. & Rouze, N. (1996). Analysis of Idealized Zeeman Quantum Beat Experiments in the Advanced Laboratory. *Am*. *J*. *Phys*. **64**(2): 147-150.

TALKS PRESENTED

"Optical Tweezers Measurements of DNA-Protein Interactions". September 2, 2005. Hope College Departments of Biology and Physics Seminar, Invited Speaker.

"DNA Looping by the Two-Site Restriction Enzyme *Sau3AI* Studied Using Optical Tweezers". March 21, 2005. APS March Meeting, Los Angeles, CA.

POSTERS PRESENTED

"A Survey of DNA Looping and Cleavage Properties of Different Restriction Enzymes Using Optical Tweezers." March 22, 2005. APS March Meeting, Los Angeles, CA.

"Optical Tweezers Measurements of DNA-Protein Interactions." March 13, 2005. ACS Annual Meeting, San Diego, CA.

"Mechanical Unraveling of Nucleosomes Assembled Using Core Histones, NAP-1, and ACF." February 16, 2005. Biophysical Society Annual Meeting, Long Beach, CA.

"Mechanical Unraveling of Nucleosomes Assembled Using Core Histones, NAP-1, and ACF." January 2 – 7, 2005. Single Molecule Biophysics Workshop. Aspen, CO.

"Oscillatory and Non-Markov Dynamics in Murine (Sca+/Lin-) Immature Hematopoietic Cell Tracks." January 5 – 10, 2002. Program in Mathematics and Molecular Biology Workshop. Santa Fe, NM.

ABSTRACT OF THE DISSERTATION

DNA Tethering Characterization, Enzyme-Mediated DNA Looping Under Tension, and Nucleosome Stability in the Force Measuring Optical Tweezers

by

Gregory John Gemmen

Doctor of Philosophy in Physics

University of California, San Diego, 2006

Professor Douglas E. Smith, Chair

The force-measuring optical tweezers has afforded scientists unprecedented insights into many of the DNA-protein interactions that regulate genetic processes. This dissertation details the construction of an optical tweezers and the study of several such interactions. A method for generating arbitrary DNA sequences for use in the tweezers is described, including a characterization of tethering as a function of buffer pH, salinity, and DNA loading. The strength of the DNA-tethering molecular bond was measured via the distributions of unbinding times under tension. The interaction of DNA with restriction endonucleases (REases) that require two recognition sites provided a model system to study DNA looping, which is critical to such biological processes as transcription, replication, and recombination. Here it is shown that a few piconewtons of applied force completely inhibited cleavage for fifteen known or suspected two site REases, whereas it had little effect on one site REases. The application of higher tensions either inhibited or catalyzed the activity

of one site REases, depending on the protein induced bend in the DNA. By replacing the Mg^{++} necessary for cleavage with Ca^{++}, the loops were stabilized with the DNA intact. In the tweezers each of these loops was pulled apart, yielding a length and disruption force. Distributions thereof for the same fifteen two site REases revealed strong enzymatic effects on DNA looping. One such enzyme, Sau3AI, was used to study looping as a function of time and applied tension. Comparisons with theories indicate that protein induced bridging and kinking play a profound role in tensioned DNA looping. The extreme sensitivity of looping to tension provides a mechanism by which even extracellular stress may act as a molecular switch. Lastly, arrays of nucleosomes were assembled on arbitrary DNA using an ATP-dependent enzymatic system and then stretched in the optical tweezers. Abrupt events releasing ~ 55 to 95 bp of DNA at forces ranging from ~ 5 to 65 pN were observed, attributable to the unraveling of nucleosomes. The rewrapping of nucleosomes was occasionally observed upon relaxing the DNA. The unraveling and rewrapping of nucleosomes under tension may have an influence on DNA-directed cellular processes.

Chapter 1. Introduction and Experimental Apparatus

1.1. Background and Dissertation Organization

In the last decade, the rapid advancement of single molecule techniques such as fluorescent resonant energy transfer and atomic force microscopy have afforded physicists and biologists unprecedented insights into a host of biophysical systems. The discrete sampling in time of bulk techniques has yielded to detailed, real time measurements of individual molecular processes; the ensemble averages of bulk assays can be replaced by statistical distributions of a population of molecules in the stochastic, Brownian motion regime. The quintessential single molecule technique is the force measuring optical tweezers, which has facilitated numerous breakthroughs in the study of individual molecular motors (1,2,3), rheology at the single molecule level (4,5), and DNA-protein interactions (6,7). The interrogation of a few such DNA-protein interactions with a force measuring optical tweezers is the subject of this dissertation, which will begin with a description of the optical tweezers system.

This dissertation is divided into seven chapters. Each of the scientific chapters (Chapters 2–6) has been submitted or accepted for publication. Thus, each chapter contains an introduction, a methodology section that may be partially redundant with that of other chapters and the first chapter, and a conclusion in which the relevance and possible future directions of the work of that chapter are discussed. Any redundancy has been removed when possible.

The first chapter, of which this Introduction is a part, is a description of the optical tweezers system with a particular emphasis on the elements thereof for which

I was responsible. A brief discussion of optical trapping and force measurement is provided, as well as of the optics of the tweezers used in the experiments described here. The fluidics of the system, including both buffer and enzyme flow, is described, including a summary of flow cell and micropipette construction. A more detailed explanation of the experimental control program and its many features is then presented; the development of the control program was my first significant responsibility in the research group. Lastly, the calibration and noise limitations of the optical tweezers are also characterized. Ultimately the management of the laboratory and the projects therein would become entirely my responsibility, though detailed descriptions of every protocol and procedure are mercifully not presented.

The second chapter describes my part in the development of a DNA labeling scheme by PCR and the characterization of that DNA in the optical tweezers; this work culminated in a paper published in *Nucleic Acids Research*. In this work the tethering efficiency as a function of DNA loading, salt concentration, and buffer pH are described. The observed length distributions for a number of DNA templates from diverse organisms, including human, are also presented. The strength of the DNA tethering molecular bound (DIG-αDIG, specifically) is also characterized in a series of lifetime measurements at different DNA tensions.

The third chapter describes the development of an optical tweezers assay to identify Type II restriction enzymes that need to interact with two copies of their recognition site in order to cleave DNA efficiently; the sites are brought together via DNA looping to form the active enzyme complex. The application of sufficient tension to the DNA inhibits the looping and thus the cleavage of such enzymes. For

comparison, the effects of tension on the cleavage activity on regular one site restriction enzymes are investigated, and an inhibition that correlates with DNA bending angle is discussed. The work comprising this chapter has been submitted to the Proceedings of the National Academy of Sciences.

The fourth chapter extends this survey of two site restriction enzymes to investigate their looping characteristics. By replacing the Mg^{++} that is usually required for cleavage with Ca^{++}, the loop stabilizing enzyme complexes can form but not cut the DNA. These loops can then be pulled apart in the tweezers, revealing a loop length and disruption force. The distributions for these lengths and forces are presented and discussed; this work has been submitted to Nucleic Acids Research.

The fifth chapter of this dissertation uses the two site restriction enzyme Sau3AI to study the thermodynamic looping of DNA as a function of time and tension. DNA looping is utilized by an ever-growing number of biological processes such as transcription and DNA recombination. Extensive comparisons with theories are presented where appropriate, and the relevance of tension as an *in vivo* regulatory mechanism is discussed. This work has been submitted to the Biophysical Journal.

The sixth chapter of this dissertation has been published, in part, in the Journal of Molecular Biology. We assembled nucleosomes on a random template using an ATP-hydrolyzing, motor enzyme system discovered by Jim Kadonaga's group at UCSD. We then pulled apart the nucleosome arrays in the optical tweezers. A nucleosome is the repeating unit of chromatin (i.e., how DNA is packaged in eukaryotes) and consists of a DNA molecule wrapped 1.7 times (~147 base pairs) around a histone protein core. As the basic functions of DNA such as transcription,

replication, and repair are dependent on the accessibility of various portions of the DNA molecule (8), understanding the energy landscape of the assembled nucleosome and how it unravels will provide fundamental insights into such important DNA processes. The effects of buffer salt concentration on unraveling force are presented, as well as the observance of spontaneous rewrapping of nucleosomes.

The final chapter of this dissertation briefly recounts the many projects that were tried preliminarily in my laboratory, but due to circumstances (namely, other projects working successfully) were not pursued further. Two additional optical tweezers have been built in the research group; these projects may find new life on these tweezers if time, opportunity, and manpower permit.

1.2. Development and Characterization of the Optical Tweezers

1.2.1. Optical Trapping and Force Measurement

Integral to the optical tweezers is optical trapping. The apparatus used in this dissertation research relied on two translucent, protein-coated polystyrene beads, one secured onto a micropipette by suction and another optically trapped, which could bind to 'handles' biochemically attached to the ends of a single DNA. The trap was formed by strongly focusing a laser. The dielectric polystyrene bead was polarized by electric field of the focused laser light passing through it according to $\mathbf{p} = \varepsilon\mathbf{E}$. An induced dipole experiences a force according to $\mathbf{F} = \varepsilon\nabla E^2$. Therefore, an induced dipole will be drawn towards the region of the greatest light intensity, which is the focal point of the trapping objective. Alternatively, trapping can be understood in terms of the momentum of the photons, as drawn in Fig. 1.1. The bead refracts the

photon so that its incoming and outgoing directions differ. This momentum change imparts a force to the bead. When the bead is on the optical axis, the symmetry renders no net force. When the bead is displaced from the axis, however, the net force is back towards the axis, resulting in a trap. Along the beam axis, the trapping must overcome the 'downstream' force from reflecting photons. As shown in Fig. 1.1, the force imparted to the bead by the refracting photons along the optical axis is much less than that normal to it, so that the trap was weaker along the optical axis.

When an external force acts on the trapped bead, the bead is displaced from the center of the trap against the trapping force, as shown in Fig. 1.1. As the bead is translucent, this displacement alters the optical path of the laser light passing through it (i.e., the trapped bead acts as a spherical lens). This change in the optical path is manifested in the laser light emerging at a different angle. By positioning a second objective lens at its focal distance downstream of the trapped bead, this light can be collected and recollimated. Because the incident angle into this second objective (i.e., the collector) is not along the optical axis, however, the collimated light emerging from the collector will not be centered around the optical axis (though it will be parallel to it). This shift from the optical axis is measurable, and it is what facilitates the measuring of force on the trapped bead (and hence the tension in the DNA).

1.2.2. Optics of the Tweezers System

The force-measuring optical tweezers used throughout this dissertation is shown schematically in Fig. 1.2. The beam path, shown as a dotted line in the figure, originates at the infrared laser (CrystaLaser, Inc.) with a wavelength $\lambda = 1064$ nm and

a power of 1W. Infrared light was used because it does not harm biological molecules. The light passed through a lens (L1-L4 were plano-convex; Thorlabs, Inc.) which focused it through the aperture of a shutter (Newport, Inc.) placed at the focal point of both L1 and L2; it should be noted that eventually the shutter, which could release trapped beads, was removed from the system. Once past the shutter, the light was recollimated by L2, but with a much larger beam diameter (~ 1 cm). After passing through L2, the beam reflected off a mirror (Thorlabs, Inc.) and passed through a neutral density filter (Edmund Optics). By changing the optical density of ND1, we altered the input power of the optical trap to alter the trap stiffness. Typically we used an input power of ~ 220 mW.

With the input power set, the laser light reflected off a dichroic mirror (Edmund Optics, Inc.) which reflected red light while passing blue. Discussed momentarily, the illumination and visualization of the flow cell required this element as opposed to a regular mirror. From the dichroic mirror, the laser light overfilled the back aperture of the trap forming, 60×, 1.2 NA water immersion objective lens (Olympus, Inc.), hence the beam expansion by L1 and L2. The trap forming objective (O1 in the figure) was built to correct for spherical aberration and had an anti-reflective coating. Back reflections from the objectives – or any optical component – into the laser cavity could cause mode hopping and other instabilities; to prevent this I typically set ND1 slightly askew.

The optical trap was formed within the flow cell (FC in the figure; discussed below). After passing through the trapped bead as discussed above, the laser light was gathered by the collector, O2, an objective nominally identical to O1 and

mounted on a 3-directional micrometer stage to facilitate focusing and alignment. From the collector, the collimated light passed through a polarizing beam splitting cube (Newport, Inc.), which redirected a portion (the portion polarized in the plane of the figure) to the position sensing detector (On-Trak, Inc.); the beam splitting cube could also be tilted about two axes to facilitate alignment. The PSD (and its amplifier) allowed for the measure of any deflection of the laser and, in turn, the force on the trapped bead. The PSD was essentially a 10×10mm silicon photodiode whose analog output (-10 to + 10 V) in both X and Y was proportional to the location of the incident light on the photodiode; the intensity of the incident light was quantified by the 0 to +6 V PSD Sum signal. A primer on the theory of the operation of the PSD is available at www.on-trak.com/theory.html. To facilitate detection and minimize interference from other light sources, another neutral density filter and a long pass filter were mounted over the aperture of the PSD.

Fixed relative to the flow cell was the pipette, which held the second bead needed to tether DNA. The flow cell was mounted on a frame which was mounted on a piezo-driven nanostage (MadCity Labs, Inc.) that moved perpendicular to the optical axis. The nanostage was mounted on a micrometer stage (labeled XYZ in Fig. 1.2). Moving the pipette via the nanostage relative to the optical trap stretched the DNA molecule, exerting a force on the trapped bead which was measured as described above. The layout of the main channel of the flow cell is shown in Fig. 1.3.

The beads that were held on the pipette were coated with antidigoxygenin (αDIG). It binds to digoxygenin, which was one of the handles on the DNA molecules. The optically trapped beads were coated with streptavidin, which binds

very strongly to biotin, a vitamin molecule that was the handle on the end of the DNA opposite the DIG handle. In all of the experiments described in this dissertation the DNA was prebound to the streptavidin beads before use in the tweezers. The biochemical methods by which these handles were affixed to the DNA varied from experiment to experiment and, along with those for bead preparation, are described in greater detail in the methods sections of each project chapter in this dissertation.

To create a DNA tether (as shown in Fig. 1.3), one first moved the αDIG dispenser tube to the optical trap to capture a αDIG bead. Once captured, the bead was moved to the pipette and transferred onto it by suction via a syringe. The SA tube was then moved to optical trap to capture a potentially DNA-loaded bead; once done, the pipette was brought to the optical trap. Rubbing the two beads together lightly for a few seconds allowed for a tether to form if DNA was present on the SA bead; a tether was confirmed if the measured force rose steeply when the two beads were separated sufficiently. The indicated flow direction kept unwanted beads away from the optical trap during experiments. Moving either bead tube to the optical trap was done by the micrometer stage; motions of the pipette about an optically trapped SA bead were performed by the nanostage, the motion of which was controlled by a control program in LabVIEW (discussed below).

To manipulate the beads, the user had to be able to see the beads. This was accomplished by the illumination/visualization (I/V) system in the tweezers. Referring again to Fig. 1.2, the I/V initiated at the blue LED (Digi-Key), which I wired with a protection diode and resistor (in series) and powered with an over-the-counter DC adapter set to ~ 5V. The non-collimated, non-polarized light from the

LED passed through a lens (L3), which focused it past collimation to the beam splitting cube, which passed the 'agreeable' portion of it to the back of the collector objective (O2), slightly overfilling it. The blue light then imaged the pipette and trapped bead, was gathered by O1, and sent towards the dichroic mirror. The DM passed the blue light towards L4, which focused it onto the CCD camera (Watec Co.), which was connected directly to a monitor. A short-pass filter was placed in front of the CCD to eliminate any reflections from the laser. An image of what the user actually saw is shown in Fig. 1.3.

Though the protocols for doing so will not be described in this dissertation, re-alignment of the optical system, at least in part, demanded much of my attention. The pipette, for example, clogged easily and needed replacing, which involved removing the flow cell and removing the collection objective. Additionally, the collection objective, being mounted on a (relatively soft) micrometer stage, would drift slowly and demonstrate hysteresis. Depending on the state of the experiments, it was often more prudent to align the rest of the system to compensate.

1.2.3. Fluidics System

1.2.3.i. Flow Cell

The optics trapped beads, the piezo-driven nanostage and pipette moved the two beads relative to each other, and protein chemistry attached the ends of a DNA molecule to those beads. The part of the apparatus in which this occurred was the flow cell. Its design was developed by Steve Smith in Carlos Bustamante's lab at UC Berkeley, and brought down to UCSD by Doug Smith. Consequently, a detailed

protocol for their construction is not presented in this dissertation. However, the very delicate construction of flow cells (once broken they could not be repaired), the requisite insertion of pipettes, and the realignment of the optics consumed a great deal of my time and efforts in the lab.

A flow cell, shown schematically in Fig. 1.4, consisted of four layers: a 24 × 60 mm #1 microscope slide, two layers of 0.004" thick heat-cured epoxy sheet, and another microscope slide. The top slide is laser drilled with six holes, as shown in the figure. The holes were laser-drilled at UC Berkeley, so conserving slides was important. The epoxy was cut to be the same size as the slides but with three long channels, as shown in the figure. These channels had to be cut with a razor according to a template, also from UC Berkeley. It was important that the channels be cut cleanly so that, once the epoxy was cured, the inside walls of the channel did not have too many jagged protrusions that would interfere with manipulating beads or catch debris. Additionally, the epoxy needed to be cut so that neither the channels nor the holes were closed off when it was cured.

Between the layers of epoxy were the bead dispenser tubes and the pipette insertion tube. The stock for these three tubes was also brought down from UC Berkeley. The pipette insertion tube needed to extend far enough outside of the flow cell to simplify pipette insertion; this nubbin was easy to break, however, so great care was needed in handling flow cells. Also, the pipette insertion tube had to be placed at the midpoint of the cell and far enough into the main channel to keep the inserted pipette stationary once in position. Each bead dispenser tube had to be cut to ~ 1/4" and placed on a diagonal ~ 3/8" downstream of the pipette, but pointing

upstream. This meant that beads coming out of the dispenser tubes would be swept in the flow away from the pipette and thus would not interfere once a tether was established. Once the three tubes were placed on the bottom sheet of the epoxy (done with tweezers), the second sheet of epoxy and microscope slide were placed so that all four layers were aligned. This was the most difficult step in flow cell construction, because almost invariably placing the second epoxy sheet moved the three tubes. Once the four layers were aligned and stacked with the tubes in the proper places, the flow cell was heat cured (125 C, overnight typically).

With a flow cell cured, a pipette had to be inserted, which was also a delicate operation. A pipette, shown schematically in Fig. 1.5, was thin glass tubing that was pulled to a very fine tip (< 1μm diameter). Both the glass tubing stock and pipette puller were brought down from UC Berkeley. Once the pipette was pulled, its blunt end was inserted ~ ½" into a ~18" length of PE-10 tubing. The tubing was sealed around the body of the pipette by heat shrink tubing. A 30 gauge needle on a 1 mL syringe was inserted into the other end of the tubing, allowing for suction or pressure to be applied to the pipette. The pipettes were tested by the 'bubble test,' which involved holding the pipette tip under water and pressuring the 1 mL syringe until bubbles emerged from the tip. Empirically, if bubbles emerged from the tip when the syringe was compressed to ≤ 0.2 mL, it usually meant that the pipette was good.

With a pipette constructed, inserting into the flow cell was done under a microscope, likewise brought down from UC Berkeley. With the flow cell mounted (and leak tested) on its frame and the pipette inserted into a holder designed here, both the pipette insertion nubbin of the flow cell and pipette tip had to be brought

together in X,Y, and Z, the last of which occurred when both were in the same focal plane of the microscope. Under the microscope, this was an iterative process, going to higher and higher magnifications. At the highest magnification, the nubbin could be used test the relative Z displacement of the two by touching it to the side of the pipette. When the nubbin pushed the pipette sideways without pushing it out of focus, the two were aligned in Z for insertion into the tube. If the pipette tip contacted any part of the flow cell, it would break and a new pipette would have to be made. Once the pipette tip was inserted into the main channel of the flow cell, the pipette body (the heat shrink tubing) was fixed to the frame holding the flow cell. In this way, the pipette and flow cell were fixed relative to each other, and the frame on which they were mounted could be moved (by the stages) relative to the optical trap.

1.2.3.ii. Buffer and Enzyme Flow

Critical to the biophysics experiments was keeping the molecules of interest in the appropriate buffers. The buffers varied from experiment to experiment, and even from initially tethering the DNA molecule to incubating it in enzyme. The controllable flow of beads was also extremely important. Solutions to clean the flow cell, water to rinse it, and air to dry it (and keep the micropipette unclogged) all needed to be introduced. Also, the flow system needed to provide a way in which enzyme solutions could be quickly introduced to the tethered DNA. Lastly, the flow rates of both enzyme solutions and buffers needed to be adjustable. Too fast off a flow would often disrupt the DNA tether or sweep the bead out of the optical trap, too slow of a flow made data collection inefficient and increased the risk of other

problems such as debris falling into the trap. In addition to the control program, the development of the flow system was one of my earliest responsibilities in the laboratory. Its upkeep throughout the experiments was a persistent challenge. Its development was progressive and heuristic in that the system was expanded as our needs changed; its final form was not optimal, but it was functional.

The final state of the flow system is shown in Fig. 1.6. The flow cell was coupled to the rest of the fluidics by Tygon tubing (I.D./O.D. = 0.03125”/ 0.09375”) press fit against the glass slide of the flow cell, as shown schematically in Fig. 1.5. The blunt Tygon end was centered over the holes in the microscope slide at the end of each channel. A drilled-out set screw (5/32”, 36 pitch) secured the tubing against the flow cell as it was finger-tightened in its frame. The Tygon tubing was coupled directly to the PE-100 tubing by snugly fitting around it. The top and bottom channels of the flow cell carried the DNA-streptavidin and αDIG beads, respectively. Into one end of each bead channel was connected the 1 mL syringe with the syringe needle (18 gauge with a bluntly cut end) connected to the tubing by a short piece of Tygon. On the other end of each bead channel was a standpipe made of PE-50 tubing. The user controlled the flow of beads into the main channel by adjusting the level of fluid in the standpipe with the syringe pressure. In practice, the bead flow into the main channel was stopped once a bead was captured.

The main channel, in which the pipette was located, carried buffer and enzyme. Beads entered the main channel from the bead dispenser tubes, shown as the short diagonal lines connecting the bead channels to the main channel in Fig. 1.4. The main channel was coupled to the rest of the fluidics system in the same way as

the bead channels. Typically, buffer from the buffer drip syringe (10 mL; backed up by atmospheric pressure) passed through PE-100 tubing to the first crimper (C1), which allowed the user to control the flow rate in the main channel. It then passed through the 8-way valve to the out port (labeled "Out" in Fig. 1.6). The 8-way valve could also turned to select a 10 mL buffer syringe for filling the flow cell, a 10 mL water syringe for rinsing the flow cell, a 10 mL syringe of 1 M KOH or ~5% SDS for cleaning the flow cell, and an air syringe for purging the flow cell. From the 8-way valve, the buffer passed through the valve labeled V3. This valve, which was almost always in the configuration shown, could be turned to isolate the flow cell from the flow cell, which was used to stop the flow when there was only buffer in the system (with V2 set as its drawn in Fig. 1.6). Stopping the buffer flow facilitated the most accurate force measurements because it eliminated non-Brownian viscous forces on the optically-trapped, force-measuring bead. Turning V3 also allowed the use to purge any air bubbles from the 8-way valve (common when replacing the syringes) without disrupting the main channel of the flow cell.

From the output of V3, the buffer passed to an identical valve (V2 in Fig. 1.6). This valve selected either enzyme or buffer flow to the cell. V2 was coupled to the flow cell through PE-10 tubing (I.D. = 0.28 mm); all other tubing in the system was PE-100 (I.D. = 0.86 mm). This smaller tubing was used to reduce the volume of enzyme that would be needed to run experiments. Once through the main channel of the flow cell, the buffer or enzyme solution exited the flow cell toward the valve V4. This valve had three settings: wide open (as shown), constricted (towards the crimper C2), and closed. Typically the wide open setting was used when using buffer,

constricted when using enzyme (the flow of which was controlled by C2), and closed when actually taking data. When using a pressurized syringe (such as the H_20 rinse), it was very important the wide open setting was used or the crimper C2 was all the way open (functionally identical to wide open), lest the flow cell be broken.

Enzyme flow originated from the "Sample Tube" shown in Fig. 1.6. Enzyme aliquots (typically 100 μL) would be prepared in 0.5 mL micro-centrifuge tubes, labeled 'Enz." in the schematic. The 1 mL enzyme uptake syringe would pull the solution into the sample tube up to a prescribed level marked on the tube (with the T-valve V1 as drawn); the timing of enzyme arrival at the tethered DNA molecule relied on getting the initial enzyme level correct. Enzyme did not enter the uptake syringe. V1 also allowed the uptake syringe to be purged without pressurizing the sample tube, which was necessary before uptake. With V4 set to the constricted outflow and C2 fully crimped, the valve V2 would be turned half way (~ 45º) to stop both buffer and enzyme flow from the flow cell and to isolate the tethered DNA molecules from pressure jolts associated with turning the valves. Once the flow cell was isolated by V2, V1 was turned so that air pressure backed up the gravity-fed enzyme flow. Once V1 was turned to air, V2 was slowly turned the rest of the way to feed the sample tube through to the main channel of the flow cell. Once fully turned (and with the tether still intact), the enzyme flow was initiated by slowly releasing C2.

Enzyme flow was measured on the sample tube, which was marked in 5 μL increments, by utilizing the "Flow meter" feature of the control program. Flowing bead solutions periodically allowed us to test when the enzyme reached the pipette and the regularity of the flow rate, ensuring accuracy in timing measurements. Once

the cell was filled with enzyme, the flow was stopped by gently closing V4. With the flow stopped, data could be collected. If, after the experimental trial the DNA tether was intact and could be returned to "normal" (e.g., free of loops), we would flow in another 5 μL of enzyme and repeat the experiment. This allowed us greater data collection efficiency and saved on enzyme solution. After the experiments were done or the tether was lost, the channel was flushed with buffer (after the valves were returned to configuration shown) by syringe and then returned to drip flow.

All couplings that were not Tygon press-seals or tubing snug fits were made by microfluidics parts from Upchurch Scientific (Oak Harbor, WA). All the valves except V1 were from Hamilton, Inc. (Reno, NV). V1 was a plastic valve ordered through Fisher Scientific, Inc. The tubing, as well as all the syringes, were from Becton-Dickinson, Inc. As the pieces were usually single use, keeping the lab stocked with the appropriate parts was a part of my managerial duties.

1.2.4. Optical Tweezers Control Program

1.2.4.i. Overview

In order to measure the deflected laser signal from the PSD and to manipulate the micropipette with the nanopositioning system, a LabVIEW control program was written that integrated these functions and emphasized versatility for the user. As the group started to build the tweezers, writing the control program largely became my responsibility as I had fairly extensive LabVIEW programming experience. Although the control program was consistently upgraded to include new features and eliminate

obsolete ones, the development and the troubleshooting of the core of the program took nearly a year. Accordingly, I would like to describe it in moderate detail.

A desktop computer running Windows 2000 XP Professional (Microsoft) and, as mentioned, LabVIEW 6i (National Instruments) was interfaced to the experimental apparatus thru an IOTech 2001 DAQboard. Of its many features, the DAQboard has sixteen analog input channels, each with sixteen bit resolution over the range -10 to +10 volts, eight 8-bit digital I/O ports, and four 16-bit, -10 to + 10 volts analog output channels. Five analog input channels were utilized to measure the X deflection, the Y deflection, and the Sum signals from the PSD, as well as the X and Y position sensors of the nanopositioning system. One digital output was utilized to trigger the shutter, when that was in the system. Two analog output channels were utilized to produce the voltages that controlled the nanopositioning system in the X and Y directions. All other DAQboard channels were unused. The IOTech board was supplied with a basic set of drivers for LabVIEW and a National Instruments programmers' forum made a host of more advanced, privately written drivers available.

The overall architecture of the program is relatively simple, and is outlined schematically in Fig. 1.7. After the initialization of the DAQboard, the files to be written, and a myriad of variables, a WHILE loop ran continuously until the user ended the program by pushing the <End> key or a fatal error occurred. After initialization, the DAQboard scanned all five input channels at 1000 Hz continuously, writing the data to the onboard buffers from where it was uploaded to and averaged by the program ten scans per iteration. Consequently, the program ran at 100 Hz. After the five input voltages were determined each iteration, they were used to

calculate a force and extension for the DNA molecule. The program utilized this force, extension, and a number of other variables to calculate the DC voltage (-10V to +10V) that was to be sent to the nanopositioning system.

These DAC voltages were also dependent on the user-selected state of pipette control: PSD alignment, force clamp, autostretch, scrolling, timed pulsing, and trackball. Each mode will be considered individually below. As Fig. 1.7 indicates, these different modes were on the leaves of a series of nested TRUE/FALSE cases that were determined by front panel controls linked to keystrokes. Thus the user could easily switch from one mode to another in the control program as needed. For example, the way the figure is drawn indicates that all the modes were off (their Boolean controls FALSE), defaulting pipette control to the trackball mode. If a case were on, the other side (i.e., TRUE) of the corresponding leaf would be activated; subordinate cases were only nested on the FALSE sides of leaves. In this way there was a hierarchy of cases, with PSD Align being the most fundamental, so that some operations were not possible while another mode was operational (e.g.., jumping directly into timed pulsing from force clamp was not possible). In practice, however, this was not at all inconvenient because typically only one mode was used in a particular experiment. Additionally, safeguards against unreasonable jumps up in the hierarchy were affected by disabling the controls for the higher modes while in a lower mode. For example, though possible in the nested architecture, accidentally going to ‘PSD Align’ from ‘Autostretch’ was made impossible by disabling the ‘PSD Align’ control when in this lower mode. On the other hand, a reasonable step up in the hierarchy, such as entering ‘Force Clamp’ from ‘Autostretch’, was kept possible.

Each mode relied on parameters from within the iteration (such as force and extension), as well as parameters from the previous iteration. In LabVIEW, this is accomplished by reading variables from the previous iteration and writing them to the next. These are called 'Shift Registers' and are indicated by the block arrows mounted on the WHILE loop in Fig. 1.7. LabVIEW is an interpreted language, meaning that it does not read/write the values for a particular variable from a particular memory location. Instead, it utilizes floating memory, meaning it rewrites its variables each iteration. In short, it can run very slowly if not programmed carefully. Therefore, to run at 100 Hz, I had to economize the use of passed variables - indeed any variables at all - very diligently. Performance nonwithstanding, engineering the operation of the program with dozens of customized features in an interpreted language also required special consideration, as referencing variables by name is impossible. Without exploring the details (as there were literally thousands of them that went into the development of the program), judicious use of Boolean flags and integer values – instead of floating decimal point of string variable – proved to be the best programming motif. In short, it was computationally less expensive to pass Booleans and 8-bit integers and use them to select longer (in bits) variables only when needed rather than passing the decimal numbers or characters each iteration.

1.2.4.ii. Front Panel and User Interface

The user interacted with the control program via the front panel, which is shown in a typical configuration in Fig. 1.8. The most striking feature of the front panel was the large graph that usually displayed the measured force versus DNA

extension; the user could also display the running time series of force or extension. The control of this graph is discussed below in the section "Graphing Controls". The other important elements of the front panel were the function indicators and the numerical controls. As mentioned above, LabVIEW is an interpreted language. One of the most computationally expensive elements of the control program was the display of graphics such as button controls or numerical variables. Keeping the number of objects on the front panel down to a minimum allowed for more efficient operation. As a control in LabVIEW was disabled when it was not displayed, displaying many objects allowed for the greatest versatility. Consequently, in any given operational mode, all unnecessary controls were disabled (i.e., made invisible). This also helped avoid keystroke mistakes, as altering the parameters of a disabled mode was often impossible from outside of that mode.

The user controlled much of the operation of the program through a large number of keystrokes (see Appendix 1.1) that were linked to LabVIEW controls. These controls were typically Booleans that triggered different functions within the program. When useful, indicators for these Boolean controls were displayed on the front panel; they would light up red when activated. Examples of such indicators can be seen on the lower left of the front panel. The overwhelming majority of active Boolean controls, though enabled, were not displayed in the visible window as not to clutter the front panel.

Other extremely important controls were the numerical controls, a few of which can be seen in Fig. 1.8. These were used by the various functions of the program, and the user could cycle through to the desired numerical variable by hitting

<Tab>. Many of the numerical variables (as well as string variables) were made visible by a keystroke so that they could be changed, but that change would not take effect until the numerical variable was made invisible. Again, this eliminated the computationally expensive variable display. Also, many of the different modules used the same numerical variable, but simply displayed a different name for it in the different modes, reducing the computational expense of displaying variables.

1.2.4.iii. Data File Generation and Handling

Upon starting the control program, a pop-up menu asked the user was for a 'filename' and a 'username' in addition to the aforementioned initial calibration. These were effectively string variables that could be changed while the program was running, which was useful when doing different experiments throughout the course of a day, for example. The 'filename', 'username', and date were all appended to create the working filename. Additionally, the program had safeguards against overwriting an existing working filename before starting the overall 'While Loop.'

At anytime during the operation of the program, the user could engage the 'Keep Data' feature (<F1>). When toggled on, the 'Keep Data' feature wrote the X and Y force and nanostage position sensors, as well as the PSD Sum signal, to a '.dat' file of the working filename. Utilizing boolean flag variables, the first iteration after 'Keep Data' was toggled on wrote into the .dat file a flag of 500 for all five variables. These unreasonable values served as an obvious marker to the beginning of a data run, and subsequent analysis file handling VIs utilized this flag to separate data runs into 'sub-files' for easier analysis. After the first iteration with 'Keep Data' on, the

program recorded this data at one set per iteration, so each variable was recorded at 100 Hz. By keeping the working filename on an active string variable, however, the user was able to change the 'filename' or 'username' variables at any time by pushing <Shift+F1> or <Cntl+F1>, even when writing a data file. When either was selected, a text box opened displaying the current variable. The new working filename was recorded when either box was closed. With the 'filename' or 'username' variable changed, both the data file and the comment file (see below) would have this new file name. If the user changed the 'filename' or 'username' variable to one that already existed, the program automatically incremented the working filename by adding an ordinal number to it. The filename could also be incremented automatically by pressing <Shift+Cntl+F1>. This feature proved to be extremely useful because it allowed the user to produce a multitude of different files (e.g., one per molecule), as opposed to having to parse out individual data runs out of a very large file.

1.2.4.iv. Force and Extension Values

As mentioned, many of the pipette control modes utilized the force and extension measurements each iteration. Also, these two metrics were the data that was recorded. The Daqboard, however, actually registered a series of voltages from the PSD and nanostage. These voltages were converted to the two working metrics by first subtracting off the 'zero' values and then multiplying by a calibration scaling.

The 'zero' values were the voltage values of the PSD and nanostage when the DNA was held very slack or when the beads were touching, respectively. In practice, after bringing a streptavidin bead to the pipette, one first touched the beads together

to establish zero extension. Since all movements of the pipette that were associated with manipulating a DNA molecule were performed by manipulating only the nanostage, this zero was valid insofar as the micrometer stage was not moved nor drifted too far. Because getting new beads required manipulating the micrometer stage, the nanostage had to be rezeroed for each bead. This was done by pressing <F3>, which disabled movement of the nanostage for one second and averaged the nanostage sensor positions for both X and Y over those 100 iterations. This average voltage was then put on a Shift Register and subtracted from the nanostage sensor voltage reading each iteration. This difference was then multiplied by the constant conversion factor of 2.5 μm/volt (i.e., 50 μm range over a 20 V sensor voltage range).

A similar process was necessary for the voltage values from the PSD. If a DNA molecule is held at less than half of its nominal contour length, the force-extension curve is flat and can be regarded (macroscopically) as zero. In truth, simply holding the DNA constrains its Brownian motions, which is manifested as an 'entropic tension', as discussed in Chapters 3, 4, and 5. Nonetheless, these 'entropic tensions' are well below the Brownian limit of detection in real time; an extremely stable tweezers might measure them with sufficient averaging. Once a DNA was tethered, the beads were positioned at an extension less than half the contour length of the DNA, and the user pushed <F4>. This disabled the nanostage (as not to exert any macroscopic tension on the DNA) and averaged the PSD voltage in the pull direction (i.e., Y) for 4 s; the percentage noise level of the bead in the optical trap was much larger than that of the nanostage sensors. After the 400 iterations, this averaged value was subtracted from all subsequent PSD voltage measurements in the pull direction.

This voltage difference was converted to a force using the force calibration scaling, which could also be changed as necessary through the control program utilizing the overstretch of the DNA (discussed below).

1.2.4.v. Operational Modes of the Control Program

1.2.4.v.i. Trackball Control

The default mode of pipette control was the trackball. This mode was on the FALSE side of the innermost TRUE/FALSE leaf. When the control program was started or any other pipette control mode stopped, the control program returned to trackball control. This mode was used to rub the two beads together to form DNA tethers, and was the only mode that allowed for simultaneous pipette control in both the X and Y directions. Two of the sub-VIs (analogous to a subroutine in a line-coded programming) were downloaded from a LabVIEW users forum on National Instrument's website.

On the first iteration of trackball control, the program read the existing DAC voltages in X and Y and set the cursor position (a pair of integers indicating the pixel coordinates of the trackball) commensurate with these voltages. For example, if the DAC voltages were at the center of their range, the cursor would be moved to the center of the screen's range. This was done to maximize the range of the pipette when initializing trackball control. On subsequent iterations, the pixel position in X and Y of the cursor was read and from it was subtracted the previous cursor pixel position. These differences in pixels were converted to a voltage increment to be added to the previous DAC voltages with a scaling that was empirically determined to

result in sufficiently large yet smooth pipette motion. However, this motion was not smooth enough for data collection. Finally, the pipette could be released from pipette control by pressing <F2>, represented by the 'Hold Position' indicator in Fig. 1.8. Though it was never necessary unless a mistake had been made, this allowed the user to use the trackball to highlight a numerical control, for example, without moving the pipette. This 'Hold Position' feature also worked in the 'Autostretch' mode, but not in any of the other pipette control modes.

1.2.4.v.ii. Timed Pulsing

The most experiment specific mode of pipette control was 'Timed Pulsing', which was activated by pressing <F9>. This mode, positioned on the TRUE side of the innermost TRUE/FALSE leaf, was used to measure the cleavage activity of the two site restriction enzymes which required looping (and hence very low tension) to cleave the DNA, as well as the regular one site restriction enzymes at zero tension (Chapters 3,4, and 5). In this mode, the DNA was usually held slack, but probed periodically with a quick tug to a low tension to ascertain if it was still intact. Thus, the operation of this mode required a three-state flag (0,1, or 2), corresponding to resting, pulling, or relaxing.

On the first iteration, the 'Max Force' numeric variable (the same used in 'Autostretch') was set to a test tension of 5 pN, which was above the level of noise and drift over the maximal five minute experimental duration. Likewise, the 'Max Disp' numerical variable (likewise used in the 'Autostretch' mode) was set to 20 μm, which was longer than the longest DNA we used. Most importantly, the 'Preset

Displacement' scrolling mode was enabled (discussed below), which quickly ran the pipette to the value of the 'Min Disp' numerical variable, which would be the location of the pipette during the rest phases of the 'Timed Pulsing' mode.

Once this resting location was reached, the control program began counting the prescribed number of seconds with the pipette stationary (state flag 0), which could be changed by the user at any time by highlighting the numerical 'Pulse Seconds' variable (<Shift+F9>). Once that time had passed, the control program stretched the DNA with a step size of 50 (state flag 1) until either 'Max Force' (i.e., the test tension) or 'Max Disp' was reached. If 'Max Force' was reached, the molecule was deemed to be still intact, so the control program relaxed the DNA (state flag 2) back to its resting location (state flag 0) for another cycle of this pulsing. If the 'Max Disp' was reached, then the molecule was deemed to have been cleaved, at which time the control program would sound a beep (especially useful for user multi-tasking) and exit the 'Timed Pulsing' mode. The control program also displayed a running tally of the number of successfully completed pulses, so that the 'data analysis' of these experiments amounted to nothing more than recording the displayed number (this number was also written to the comment file).

1.2.4.v.iii. Scrolling Modes

The pipette could also be moved to a number of predetermined positions very quickly when one of four different keystroke commands were issued. Collectively, these were the 'Scrolling Modes' of the pipette control, and they were programmed onto the TRUE side of the second innermost TRUE/FALSE leaf. 'Stage Centering'

(<F6>) moved the nanostage to the center of its range by running both DAC voltages to zero (the full input range was -10 to +10 V). This was useful when the largest trackball range was desired for pipette control, such as when forming a tether.

Similar to 'Stage Centering' was 'Stage Cornering' (<Shift+F6>), which ran each DAC voltage to -10 V, the minimum input of the nanostage controller. This reduced the high voltage across the PZT piezoactuators, which was important to prolong their life. Several times in my tenure in the lab the PZT stacks shorted and the nanostage had to be sent to the manufacturer for repairs; fortunately they were very helpful each time. Nonetheless, the down time was detrimental to the progress of the experiments at times. Whenever the control program was running but the nanostage was not being used, it ideally was low-voltage cornered.

The third scrolling mode was 'Current Zero' (<Shift+F3>). Rarely used, this feature scrolled the pipette back to its current zero position (i.e., where the two beads were touched to establish zero DNA extension). Originally, this feature was intended as a way to quantify positional drift over time for long experiments. The experiments discussed in this dissertation, however, all were (fairly) short in time. Additionally, bringing the beads so close together risked forming multiple tethers or establishing the 'velcro' condition, in which the beads stuck together.

The fourth scrolling mode, and the only one used in a data collection protocol, was the 'Preset Displacement' command (<Shift+F2>). In this mode the control program would run the pipette to the preset displacement in the pulling direction specified in the numerical variable 'Min Disp' (i.e., the same number used in the 'Autostretch' mode). The location of the pipette in the off direction remained

unchanged. Once the pipette arrived at the preset displacement, 'Hold Position' was enabled. This feature was useful in the looping experiments especially (Chapters 3 and 4), in which the DNA was held taut during enzyme flow-in but quickly relaxed to the prescribed fractional extension. Additionally, this 'Preset Displacement' scrolling was incorporated as the first step of the 'Timed Pulsing' mode discussed above.

Each of the four scrolling modes utilized a step size of 40 (~3 μm/s). As the longest DNA utilized in the experiments in this dissertation was ~ 8.1 μm, scrolling the pipette with DNA tethered almost always resulted in a broken tether. Thus to guard against a key mis-stroke and a broken tether, the scrolling modes (except for 'Preset Displacement') could not be enabled while 'Keep Data' was on. Likewise, as scrolling typically meant large pipette motions, the graph display was cleared.

1.2.4.v.iv. Autostretch

The feature by which the control program uniformly moved the pipette to generate force-extension curves, called 'Autostretch', was toggled on and off by pressing <F5>. It was programmed on the TRUE side of the third outermost TRUE/FALSE leaf. Upon initiating 'Autostretch', the program would stretch the DNA by increasing the bead separation at the speed determined by the numerical variable 'Step Size'. 'Step Size', which the user could change at any time, was the number of DAQboard steps to increment (or decrement) the DAC voltage sent to the nanostage per control program iteration. A single step corresponded to ~ 0.75 nm; the manufacturer's specified resolution for the nanostage was ~0.33 nm, so this step was feasible. Non-integer values for 'Step Size' were acceptable, as the program

calculated an analog voltage for the controller as opposed to incrementing the DAC outputs. For example, a 'Step Size' of 0.5 meant that the DAC voltage would increment a basic voltage step (20 V/2^{16} steps = 0.31 mV per fundamental step, as per the Daqboard specifications) every two iterations. .

The program would stretch the DNA until either the 'Max Force' or the 'Max Disp' value was encountered; confusion as to which would trigger the turnaround was avoided by setting the non-desired limit to an infeasible value. When this limit was passed, the program would relax the DNA until either the 'Min Force' or 'Min Disp' variable was encountered. Because both minimum variables were used by the different modes of the control program for different functions, setting one to an infeasible value might have had tether breaking consequences in further operations. Consequently, the lower turnaround variable could be toggled by <Cntl+Shift+F5>; its default was 'Min Disp'. The 'Min Disp' variable was especially useful when the lower turnaround was in the largely flat, entropic portion of the force-extension curve, where triggering on a force value would be too widely variable. Safeguards against bumping the beads together for a very low limit were likewise implemented.

The sense of 'Autostretch' was determined by a Boolean flag variable which was TRUE for increasing extension and FALSE for decreasing extension. This flag determined whether the voltage change set by 'Step Size' was to be added or subtracted to the previous DAC voltage and which turnaround limit was to be utilized next. When a turnaround was passed, this flag variable would change (changing the sign of the DAC voltage change per iteration) until the opposite limit was passed. This made the 'Autostretch' feature robust against noise fluctuations near either limit;

before this feature was programmed in this manner, the pipette would often flutter about either limit before moving smoothly towards the other. Additionally, the user could switch the direction of the pipette by pushing <Shift+F5>, which simply changed the state of the directional flag.

1.2.4.v.v. Force Clamping

Another important feature afforded by the control program was the 'Force Clamp', which adjusted the pipette to keep a constant force on the tethered molecule. This feature was on the TRUE side of the second outermost TRUE/FALSE leaf in the program. The underlying premise of the control was simple: if the force was below the desired value, the DNA was stretched by one DAC step from the Daqboard (which corresponded to ~ 0.75 nm), but if the force was above the desired value, the DNA was relaxed by one Daqboard step. Adjusting the clamp each iteration meant that it was responding, in part, to noise. A more sophisticated clamping mechanism (e.g., partial integration-differentiation or moving average), however, proved to be computationally expensive with little noticeable improvement. The resonant frequency of the nanostage was 2.5 kHz, and it could be scanned full amplitude (50 μm) at 400 Hz, so tiny adjustments at 100 Hz were not detrimental, though many of them were superfluous.

The clamp was started by pressing <F10> ('Force Clamp') or <Shift+F10> ('Force Collapse'). The former would tell the program to maintain the force value at which the clamp was initiated. However, when the clamp was started, this value was highlighted on the front panel so that it could be changed easily. Initializing 'Force

Collapse' would tell the program to adjust the tension on the DNA to a preset value; a window for the adjustment of this preset value could be viewed at any time by pressing <Cntl+F10>. In this mode, the initial adjustment to the preset value was achieved by either stretching or relaxing the DNA quickly (~ 7.5 nm). Thereafter, the regular force clamp mechanism maintained the DNA tension. Deactivating the force clamp, no matter how it was initiated, was done by pressing <F10>.

The 'Force Collapse' was especially useful for the one site restriction enzyme cleavage studies (Chapter 3) and the DIG-αDIG bond strength studies (Chapter 2). As time was the measured variable in these studies, a tethered molecule could be stretched to its working tension very quickly, whereas adjustments within the regular 'Force Clamp' mode occurred at one tenth the speed. 'Force Collapse' could even be initiated while the control program was force clamping. Additionally, there were safeguards in the clamping mechanism to prevent the two beads from being pushed together, a feature originally designed for real time DNA compaction studies. Finally, the comment generated upon initiation of the force clamp recorded the initial clamping value, whereas the comment generated upon the cessation of force clamping recorded the final clamping value, if indeed the user changed clamping values.

1.2.4.v.vi. PSD Alignment

To facilitate the alignment of the laser onto the PSD, the control program had a 'PSD Alignment Mode', which was on the TRUE side of the outermost of the nested TRUE/FALSE and was toggled on and off by pushing <F11>. When on, this mode displayed the raw X and Y voltages (neither zero-shifted nor force calibrated)

from the PSD and displayed them on a 2-D graph with a -10 to +10V volt range for each axis; the PSD Sum voltage was displayed on a bar chart. The X and Y display range could be toggled to -1 to +1 V for finer adjustments by pushing <Shift+F11>. With the laser sufficiently centered (by adjusting the second objective and the beam splitting cube), the user could "zero" the PSD by pushing <Cntl+F11>, which averaged the X and Y voltages for 100 iterations (1s). This average would be subtracted from all subsequent PSD voltages, thus serving as a baseline. In practice this was identical to the force zeroing, so it was rarely used. While the 'PSD Alignment Mode' was on, the nanostage was held fixed as not to move the flow cell during alignment. Conversely, the 'PSD Alignment' mode could only be turned on when the pipette was under 'Trackball' control, so as not to accidentally ruin a data run. As with most features, a comment was recorded on the first iteration of this mode, as well as on the first iteration after it was turned off.

1.2.4.vi. Additional Control Program Features

1.2.4.vi.i. Comment File Generation and Handling

A very valuable feature of the control program is the comment file that was generated every time the control program was run. The comment file was a text file of the exact same name as the data file, but with the extension .cmm. If the data file name was changed when the control program was running, the name of the comment file was changed as well. Moreover, the program automatically recorded almost every keystroke command during the program operation, creating a log of the program's operation. This was accomplished by use of Boolean flags. When the user

switched pipette control modes, for example, the appropriate comment was generated on the subsequent iteration (i.e., the first iteration in the new mode) and written to a string variable. Concurrently, a Boolean flag variable was set to TRUE and sent with the comment to the comment file writing module. On all other iterations, an empty string and a FALSE Boolean flag were sent to the comment file writing module. As a consequence of LabVIEW's interpreted nature, every automated comment did this every iteration; consequently, each automatic comment sent out FALSE flags and empty strings on nearly all iterations. The comment writing module, which was likewise visited every iteration, utilized the flags to select which comment was to be written in a "funneling" fashion. First, an OR gate with every flag variable as its inputs, determined if anything would be written at all. If so, the comment writing module determined which string variable (i.e., which comment) would be written by converting every Boolean flag to an integer, 0 or 1 (FALSE or TRUE), and multiplying it by its own unique power of 2. These resulting integers were added and the sum was then used to select which comment variable was piped through to be time stamped and written. Because each flag was multiplied by its own unique power of two, there was no way for this addition to associate the wrong comment (which would be an empty string) with the active flag. The way in which all the commenting generating modules were written insured that the Boolean flag went TRUE and the comment string was written on the same program iteration. Additionally, the comment writing module was programmed to handle the rare instances of two desired active flags, again unique because of the imposed base-2 arithmetic, or two accidental active flags, in which case an error message is written to the comment file. These

automatic comments – usually generated when a feature was turned on and off - proved to be exceptionally useful in locating and interpreting data runs; in a few instances they and the time stamps associated with them were automatically incorporated into the data analysis algorithms.

In addition to the automatic comments, the user could enter a comment regarding the experiment at any time. By pushing <F7>, a text bar was opened and highlighted so that the use could type in whatever he or she wished. Upon closing the text bar (again by pushing <F7>), the comment was written with a time stamp of when the comment bar was first opened, not when it was ultimately sent to the comment file writing module. Throughout the data collection this also proved to be very useful as bad data runs could be easily discarded, good data runs could be easily highlighted, and so forth.

There were also a few additional, heavily used comments that can be sent to the comment file with just one keystroke. They include “Focus Adjust” (<Shift+F7>), “Flow Adjustment” (<Cntrl+F7>), and “Reagents Arrived” (<Cntrl+Shift+F7>). In certain versions of the program designed to perform the timed DNA cleavage by one site restriction enzyme experiments, the “Reagents Arrived” keystroke comment also triggered the “Keep Data On” function (but not vice-versa), and the comment reflected this dual purpose. Subsequent data file handling and analysis VIs that I wrote read these automated comments as a measure of timing for DNA cleavage experiments, for example. The other automated comments were useful in determining the cause for a sudden shift in the data baseline, for example, as was the case with a ‘Focus Adjust’.

A particular feature of the control program that utilized the comments file was the 'Flow Meter', the indicator for which is in the upper right of Fig. 1.8. As discussed below, the introduction of enzyme involved a careful series of steps with the microfluidics system. Too robust of an enzyme flow would result in broken DNA tethers and wasted enzyme. To monitor the flow rate of enzyme then, the user could press <F8> each time the enzyme solution interface passed a 5 μL gradation on the sample tube. This recorded the flow meter number (i.e., the i^{th} gradation from the top of the sample tube) in to the comments file with a time stamp, and incremented the flow meter number. Subsequent VIs that I wrote extracted flow rates form the comment files. If a mark was missed, the flow meter could be incremented without recording the number by pressing <Shift+F8>. Pressing <Cntl+F8> decremented the flow meter similarly, and <Cntl+Shift+F8> reset the flow meter entirely. While all of the data presented in this dissertation was taken with stationary flow, the flow meter allowed us to monitor the performance of the flow system very easily.

1.2.4.vi.ii. Graphing Controls

The user monitored the experiments by the visualization monitor (Fig. 1.3) and the graph on the front panel of the control program (Fig. 1.8). Refreshing and plotting data on each iteration was the most computationally expensive procedure in the LabVIEW control program, so the graphing module actually plotted only every tenth point (i.e., 10 Hz), which was still well more than sufficiently often to monitor the DNA. Additionally, an extremely important feature programmed into the control program was the ability to clear the graph, which was done by pressing <Esc>.

Additionally, many features such as zeroing the force or extension or centering the stage automatically cleared the graph. Clearing the graph emptied the cumbersome arrays to which LabVIEW appended the data of that iteration for plotting. This was a consequence of the interpreted nature of LabVIEW; with no pre-allocated memory locations to write data, the program had to transfer the entirety of the ever-growing arrays each iteration, which could slow down the program and cause a potentially fatal backlog of data scans in the Daqboard's buffers. Though it was never triggered, I programmed in a module that monitored the number of backlogged data scans and would clear the graph if that number reached 100 (i.e., if the control program fell 0.1s behind the Daqboard).

Typically the front panel graph displayed DNA extension on the abscissa and the measured force on the ordinate. Programmed into the control program, however, was the ability to display either the measured tension or extension as a function of time. This feature was designed originally for DNA compaction studies, but proved to be very useful in DNA cleavage experiments as well (Chapter 3). The user could toggle between the force-extension curve and the time series plot by pressing <F12>. Once in the time series mode, the vertical axis could be toggled back and forth between force and extension by pressing <Shift+F12>.

The default scale of the force-extension curve was from -2 to 30 μm in extension and -5 to 120 pN in tension. Such a large sale allowed for the most freedom with long DNA molecules, breaking multiple DNA tethers, or testing the strength of hookups. However, the user could toggle through a series of force and extension ranges for more precise viewing of features. The vertical axis was cycled

by pressing <PgUp>, and the horizontal axis was cycled by pressing <PgDn>. The available force ranges, which were always on the vertical axis, were -5 to 120, -5 to 90, -4 to 60, -3 to 30, and -2 to 10 pN. The extension ranges, which would be on the horizontal axis in force-extension plotting but could be on the vertical axis in time series plotting, were -2 to 30, -2 to 20, -1 to 15, -1 to 10, and -1 to 6 μm. The available time ranges were 300, 120, 60, 40, and 20 s. It should be noted that for the time series modes, when the horizontal range was completed, the graph refreshed but with the next time increment (defined by the range) displayed. For example, if the time range was 20s and the time series plot reached the 20s, the trace would start over again but the range of the graph would be 20 to 40s, and so on.

1.2.4.vi.iii. Controlling the Shutter

My very first task in the lab was to wire the shutter (Newport HP846) so that the user could block the laser via the control program to release a bead from the trap; the program kept track of the state of the shutter. The circuitry that I designed for this is shown in Fig. 1.9. The magnetically activated shutter, which was at the focal point of L1 in Fig. 1.2, was usually open. The shutter opened by a brief current pulse in the direction of Pin 1 to Pin 4, and closed by a pulse in the opposite direction. Thus simply removing the current did not toggle the shutter. In either direction, the current pulse was provided by $+V_0$, an AC to DC adapter that sourced 300 mA at 7.5 volts.

To open the shutter, the user hit the <Space Bar> which instructed the control program to generate an "open" signal and output it as a TRUE signal on the 0^{th} bit of a digital I/O port of the DAQboard; the DAQboard's +5 V output served to power the

digital ICs, as shown. On the rising edge of this TRUE signal, the 74121 One-Shot generated as output a HIGH pulse that was $\tau \approx \ln 2 R_{ext} C_{ext} \approx \ln 2 (5.1 k\Omega)\ (10 \mu f) \approx 35$ ms. This HIGH pulse on pin 9 of the 74121 served as input to the gate of a 2N7000 FET transistor (with a 200kΩ bleeding resistor). With the transistor's source tied to ground, the drain pulled down enough current to energize the coil in the relay (Siemens V23042-A2003-B101), switching it briefly so that the common was tied to pin 9 instead of pin 11. With $+V_0$ connected briefly to the relay's common, the requisite current pulse could surge through the shutter from Pin 1 to Pin 4, opening the shutter. When the relay de-energized after the 35 ms HIGH pulse, the relay's common was tied back again to pin 11 and hence to ground, but the shutter stayed open. The capacitor and resistor placed in parallel with the shutter prevented ringing noise in the relays from switching the shutter.

To close the shutter, the user hit the <Space Bar> again, and the program output a TRUE on the 1st bit of the same digital port. By the exact same mechanism, this switched the other relay, allowing current to surge from Pin 4 to Pin 1 of the shutter. Again, after the 35 ms HIGH pulse, the relay de-energized and its common fell back into contact with ground via pin 11, but the shutter stayed closed. It was the current demands of the relay that necessitated using FET transistors; the duration of the pulse was suggested by the specifications of the shutter, which, like the FET transistors, could overheat with too long of a pulse. Ultimately, the shutter was removed from the optical tweezers system and used on another experimental in the group, but with the same circuitry and programming.

1.2.5. Calibration and Noise Characterization of the Optical Tweezers

1.2.5.i. PSD Calibration

As shown schematically in Fig. 1.1, an external force on the optically trapped bead displaces it from the center of the trap and redirects the outgoing laser. As the bead sits at the focal point of the second objective, this redirection is manifested as a translation of laser beam from the optical axis. As mentioned above, this translation is what is converted into a voltage by the PSD. Associating this PSD voltage with either a known force or bead displacement required a viable method by which to calibrate the optical tweezers. Force calibration was performed in three ways: the application of a known viscous force, measuring the Fourier power spectrum of a bead in the trap (no external force), and setting a measured DNA overstretch to the known force of overstretch. Bead displacement calibration was performed by applying a known displacement to a bead on the pipette and measuring the resulting PSD voltage. Developing the LabVIEW VIs that performed these calibrations was another early responsibility of mine in the laboratory.

The first method of force calibration – applying a known viscous force to a trapped bead – was performed by trapping a bead and then, with the flow cell isolated from the rest of the fluidics, applying a known sinusoidal waveform to flow cell via the nanostage. With the flow cell and the fluid therein moving relative to the trapped bead, the viscous force on the bead is $\mathbf{F} = 6\pi\eta a\mathbf{v}$, where a is the radius of the bead, η is the viscosity of the fluid, and $\mathbf{v}$ is the velocity of the fluid around the bead. The fluid velocity around the bead would be simply the time derivative of the known displacement being applied to flow cell.

Unlike the motion of the pipette in the control program, the motion of the flow cell in this calibration could be preprogrammed into the DAC buffers for extremely fast (i.e., smooth) output, which was set at 10 kHz. The LabVIEW VI allowed the user to center the laser on the PSD by adjusting the second objective. This step proved to be largely optional because of the excellent linearity of the PSD over its range and the relatively small laser displacements relative to the size of the PSD (10×10mm). In other words, there was little danger that the laser would be deflected off the PSD, in which case it would give an erroneous reading. Nonetheless, the force-free PSD reading was obtained to serve as a baseline. With the baseline set, the LabVIEW VI oscillated the stage in the chosen direction (X or Y) at a prescribed frequency and a steadily growing amplitude. The amplitude of the oscillation had to be ramped slowly so that the bead would not be dislodged from the trap. Once the prescribed peak-to-peak amplitude was reached, the program collected the PSD sum voltage, the PSD voltages in X and Y, and the X and Y sensor data from the nanostage at 1 kHz. Because the data was not processed in real time, the calibration program could utilize buffered input, hence the faster ADC rate. After the prescribed number of cycles at the full amplitude, the amplitude was ramped down to zero so that the bead would remain trapped and further calibrations on the same bead (i.e., same radius) could be performed immediately. The PSD sum signal was collected only to be sure that the laser was not at anytime spilling off the photodiode. The sensor voltage in the shaking direction was converted into a fluid velocity (and ultimately a viscous force) by taking the smoothed derivative and multiplying by the 2.5 μm/V scaling factor of the nanostage.

The measured PSD signal versus this applied viscous force for 100 cycles of a 20 Hz sinusoidal input (45 μm peak-to-peak amplitude) is shown in Fig. 1.10. Note the exceptional linearity for this (relatively low) frequency. The slope of the fit line gives the calibration. It should be noted that when the tweezers was properly aligned, the difference in the X and Y calibrations was less than 10%. No experiments in this dissertation relied on pulling in the X direction, though the user could have switched the dominant direction to X by pressing <Home> in the control program.

For low enough frequencies and amplitudes, the calibration was consistent (for a given input laser intensity and alignment). At higher frequencies and large amplitudes, the measured PSD voltage would demonstrate some non-linearities at higher viscous forces. Physically this non-linearity can come from at least two sources: the fact that the bead is not fixed in the trap (the velocity of the fluid around the bead is not simply the prescribed velocity applied by the nanaostage) and the fact that there was some mechanical 'softness' in the coupling of the flow cell to the nanostage. The former could be seen with sufficiently violent shaking; the bead began to move relative to the fixed laser beam whereas for placid shaking it did not. In this sense the viscous calibration via a sinusoidal waveform is a first order approximation. The latter cause of non-linearity was due to the fact that the flow cell was mounted on a frame that was mounted to the nanostage. For a number of practical reasons, this frame was mounted hanging off the side of the nanostage, creating a moment arm with some mass. With no element of the nanostage-frame-flow cell mounting being perfectly rigid, this moment arm would lead to non-

linearities (i.e., the motion of the flow cell would slightly lag the input signal but lead the sensor voltage), especially for large amplitude shaking.

The bead displacement calibration was performed in much the same way as the viscous force calibration, but with a bead 'trapped' on the pipette instead of optically. The laser was first centered on the PSD with no bead in the beam path, and the PSD sum signal was noted. Ideally, with no bead in the trap, the PSD sum signal should already be maximized, indicating proper alignment of the objectives lenses. Then the pipetted bead would be positioned by the micro- and nanostages into the beam path so that the PSD X, Y, and Sum signals all returned to their previous values, indicating that the bead was positioned at the center of the optical trap in all three directions. Once the bead was centered, it was moved sinusoidally through the beam diameter while the displacement and the resulting PSD deflection signal were monitored. For small displacements, the response was linear. For sufficiently large displacements, the beam diameter (diffraction limited to ~ 1 μm) would not wholly pass through the bead and the PSD signal would fall off. However, such large displacements were not relevant to our experiments; in practice the optically trapped bead would have been pulled out of the trap entirely.

The force calibration divided by the displacement calibration yielded the trap stiffness, analogous to the spring constant in Hooke's Law. Though the control program relied only on the PSD force calibration, the trap stiffness is more physical and not equipment specific. The trap stiffness for the tweezers used here was typically in the range 0.10 to 0.18 pN/nm, depending on the alignment, input laser intensity, and so forth.

A direct measure of the trap stiffness could be achieved by recording the Fourier power spectrum of a bead in the optical trap undergoing Brownian motion. The Langevin equation of such a situation is given by $\mathbf{F}(t) = k\mathbf{x} + \gamma\mathbf{v}$, where k is the trap stiffness, γ accounts for viscous forces ($6\pi\eta a$ from the discussion above), and $\mathbf{F}(t)$ is the stochastic forcing, which averages to zero. The power spectrum of this situation is calculated to be $S(f) = k_bT/\gamma\pi^2(f^2+f_c^2)$, where f is frequency and f_c is called the 'corner frequency', given by $f_c = k/2\pi\gamma$ (9). At room temperature, $k_bT \sim 4$ pN nm, which manifested itself in our system as ~ 0.5 pN of noise typically.

An example of a power spectrum measurement from our optical tweezers is shown in Fig. 1.11, with the frequency on a log-10 scale. The fit is the Lorentzian discussed above. The 'corner frequency' is where the linear approximation to the high frequency portion (f > 2000 Hz) crosses the linear approximation to the plateau (10 < f < 200 Hz). For the curve in Fig. 1.11, $f_c \approx 950$ Hz, which translates into k = 0.13 pN/nm, in decent agreement with our other calibration measurements. The deviations from the Lorentzian at low frequencies (f < 10 Hz) indicate non-Brownian noise and indicate systematic drifts in our system, which are discussed below.

The third method by which we calibrated for the force utilized the 'overstretch transition' in DNA. Double-stranded B-form DNA undergoes a fundamental change in structure when subjected to a sufficient force whereupon it increases in length ~ 70%. Once initiated, this transition is highly cooperative, so that there is very little necessary increase in force and the force-extension curve of the DNA demonstrates a plateau. The force level of this plateau has been well studied as a function of ionic conditions (10). By pulling a DNA molecule into 'overstretch' and noting the force

at which it occurred, the user could shift the calibration used by the program so that the observed value was mapped onto the known overstretch value. This was implemented in the control program by pushing <Cntl+F4>. A numeric window would open that prompted the user to enter in the observed overstretch value. By pushing the <Cntl+F4> again, the numeric window disappeared and the new calibration would take effect. In the event of a mistake in recalibration, the keystroke <Cntl+Shift+F4> would revert the control program to the previous calibration. The keystroke <Shift+F4> opened a numeric that allowed the user to change the proper force value of the overstretch plateau. This little used feature was useful in the nucleosome experiments (Chapter 6), wherein the salinity of the buffer was often changed. Additionally, the ability to recalibrate 'on the fly' accounted for short term drifts in laser intensity which affected trap stiffness.

1.2.5.ii. Instrumental Noise and Drift

Alluded to above, the low frequency deviations in the Fourier power spectrum in Fig. 1.11 were due to long term drifts in the system, many of which would not average to zero over the time scale of our experiments. A great deal of time and effort was spent diagnosing the sources of drift, which arose in both force and position measurements. The force measurement drift arose from both instabilities in the laser (pointing and intensity) and thermal drift of the optical components. The positional measurement drift arose primarily from the softness of the microstages on which the nanostage (and ultimately the flow cell/pipette) was mounted.

Regarding the former, little could be done but lightening the flow cell holder by cutting away any extra aluminum. Geometric constraints of the components (the micro and nanostage) necessitated mounting the flow cell off the side of the nanostage. Though this created a moment arm, it kept the liquid (e.g. buffers and objective water) away from the high voltage piezo-stacks in the nanostage. After repositioning the microstage (e.g., after getting a bead from the dispenser tubes and bringing it back to the pipette), a waiting period was observed to allow the microstage to settle. Fortunately, the data presented in this dissertation did not rely on the long term precision of an absolute distance between the pipetted bead and the optical trap. The position of the pipette was often rezeroed, and the length measurements were almost always differential (e.g., the amount of DNA released from a disrupted loop or a nucleosome), meaning that a distance reported was the difference between two points on a force-extension curve, typically taken ~ 1s apart. The positional drift on such time scales was well below the noise limit of the nanostage sensors.

The force measurements, however, were not differential, and thus the long term drift posed a more significant problem as it manifested itself as variability in the stiffness of the optical trap. This variability arose from variations in the laser (intensity and pointing instability) and from thermal drift of the optical components throughout the system. When tested independently apart from the rest of the tweezers system, the laser pointing instability and intensity fluctuations were within manufacturer's specifications, and could not account for the observed drifts. (It should be noted that the laser did require a significant stabilization time (~ 1 hr) after being turned on, and would begin mode hopping if it was left on too long. Simply

observing the warm-up time and turning the laser off at the end of each day significantly reduced the problems that I attributed to mode hopping.) When in the system, however, the observed laser drifts could be significant (~ an amount corresponding to a few piconewtons over a few minutes). Examples of the drift are shown in Fig. 1.12 and Fig. 1.13. For these figures, the second objective was brought into position by manipulating the microstage on which it was mounted, and immediately thereafter monitoring the PSD X and Y force signals with no bead in the trap for twenty minutes. In fact, though the measured PSD Sum was different due to fewer reflections, a similar level of drift was observed even when the flow cell was removed from the beam path. Note the wander of the recorded position of the beam in Fig. 1.12. Note the downward trend in Y in Fig. 1.13 (which can be seen in Fig. 1.12 as the thick collection of lines gradually creeping towards lower Y values), even though these were different drift characterization runs.

The fluctuations in the beginning of the time trace (Fig. 1.13) were attributed to the mechanical settling of the microstage on which the collector was mounted. These are the large 'loops' in Fig. 1.12. The long term drift, which was downward in Y in these examples, was attributed to thermal drift of the optical components in the system. To compensate for these drifts, the force was rezeroed with every bead and recalibrated by overstretching a piece of DNA whenever possible. The short term fluctuations were attributed to vibrations in the optical components and air currents in the beam path (which would change the index of refraction in the beam path). To minimize such drifts, the beam path was shielded two-fold from air currents. This simple solution cut the short term drifts significantly (by ~ 50% typically).

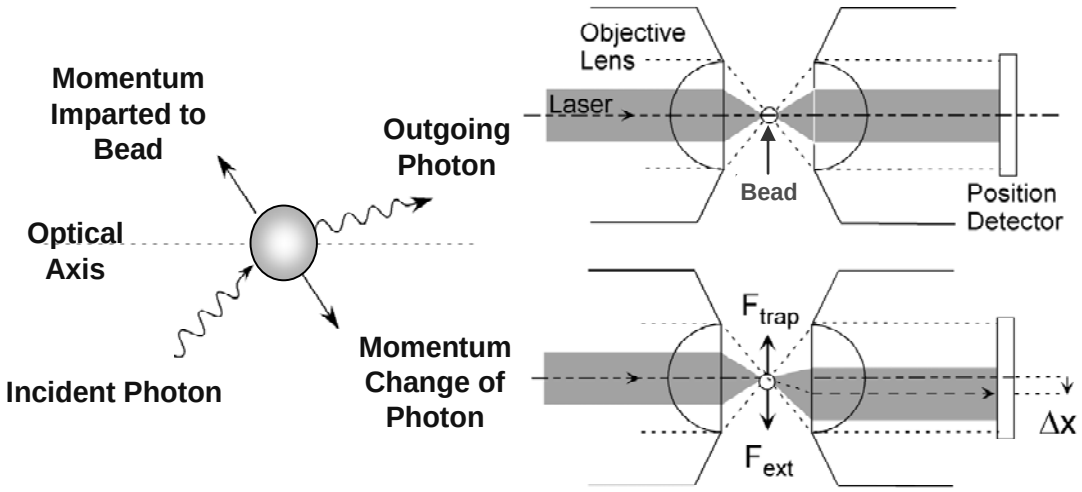

Fig. 1.1. (Left) Schematic of ray tracing interpretation of optical trapping. The incident photon, angled towards the optical axis by the trap-forming objective, is refracted upon passing through the bead. The refraction changes the direction of the photon, which imparts momentum (and hence a force) to the bead. (Right) A bead in the center of the trap effects no net displacement of the outgoing laser beam because the refraction is symmetric relative to the optical axis (top). A bead experiencing an external force is displaced from the center of the trap, changing the optical geometry and displacing the outgoing beam an amount Δx from the optical axis.

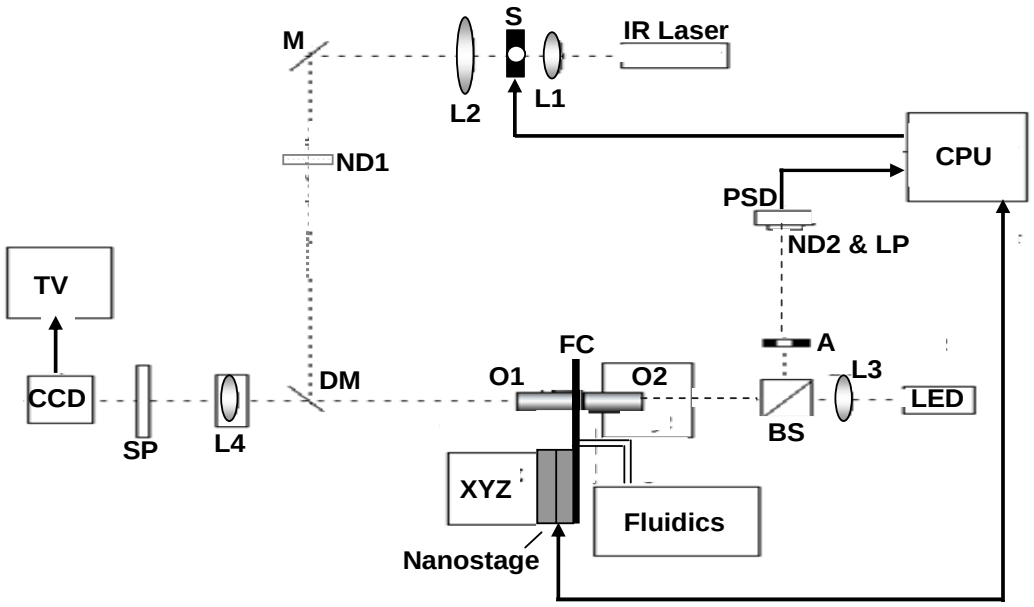

Fig. 1.2. Schematic of the optical tweezers. The beam path is shown as a dotted line. L1, L2, L3, and L4 are plano-convex lenses. S is an. M is a mirror, DM is a dichroic mirror (long reflect/short pass). O1 and O2 are both 60×, 1.2 NA water immersion objectives pointed at each other with the flow cell (FC) sandwiched between. BS is a polarizing beam splitting cube, and A is an adjustable aperture. ND1 and ND2 are both neutral density filters, while LP and SP are long-pass and short-pass filters, respectively. The position sensing detector (PSD), the nanostage, and the electrically triggered shutter (S) are all coupled to the computer (CPU) via a LabVIEW program, with the arrow showing the direction(s) of information flow. The blue LED, the camera (CCD) and the video monitor (TV) are part of the visualization system. The flow cell is mounted on the nanostage, which is mounted on a micrometer stage (XYZ). O2 is mounted on an identical micrometer stage.

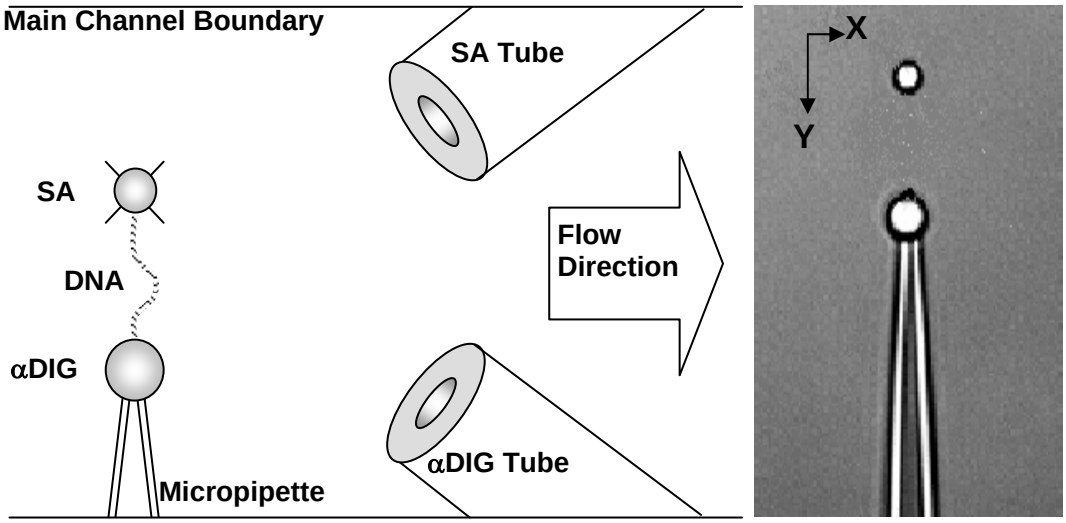

Fig. 1.3. (Left) Schematic of the main channel of the flow cell (not to scale). The optical trap is denoted by the cross hairs about the (trapped) DNA-streptavidin bead (SA). All other elements are fixed relative to the flow cell and move relative to the laser. (Right) Actual image from the monitor in the experiment. In practice, tethered DNA molecules were not seen directly, but inferred from the force measurement. The axes in the upper left denote the X and Y directions referred to in the text; the experiments in this dissertation were all in the Y direction.

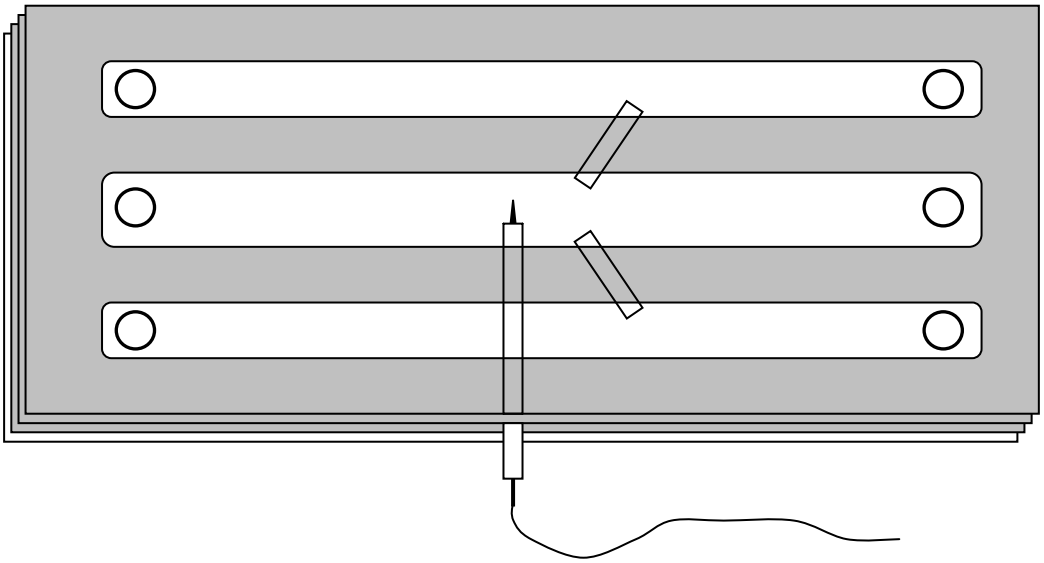

Fig. 1.4. Flow cell schematic (not to scale). The layers, from back to front, are undrilled microscope slide, epoxy sheet, epoxy sheet, and drilled microscope slide, the holes of which are shown as circles. Both slides are 24 × 60 mm, #1 slides, and the heat cured epoxy sheets are each 0.004" thick. Both the bead dispenser tubes (diagonal) and the pipette insertion tubing lie between the epoxy sheets. The pipette is shown in black, and the PE-10 tubing that couples to the syringe (allowing for the application of suction) is shown as the curved gray line. The gray in the flow cell represents the area that the cured epoxy covers. The flow would be from left to right to sweep beads from the dispenser tubes away from the pipette.

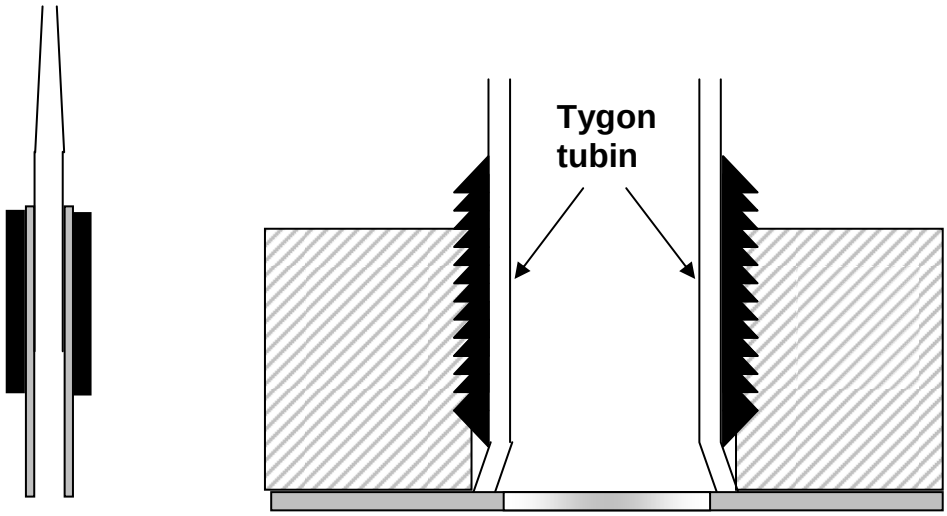

Fig. 1.5. (Left) Schematic of pipette construction in cross section (not to scale). The pulled glass is represented by the thin black lines, the PE-10 tubing by the gray lines, and the heat shrink collar by the thick black lines. (Right) Cross-section schematic of the Tygon press fit that coupled the flow cell to the fluidic system (not to scale). The flow cell is horizontal (gray) and the Tygon tubing is centered over the hole. The set screw is shown in black, and the frame which held the flow cell fixed relative to the nanostage is shown striped. The tubing of the rest of the fluids system fit snugly into the other end of the Tygon section.

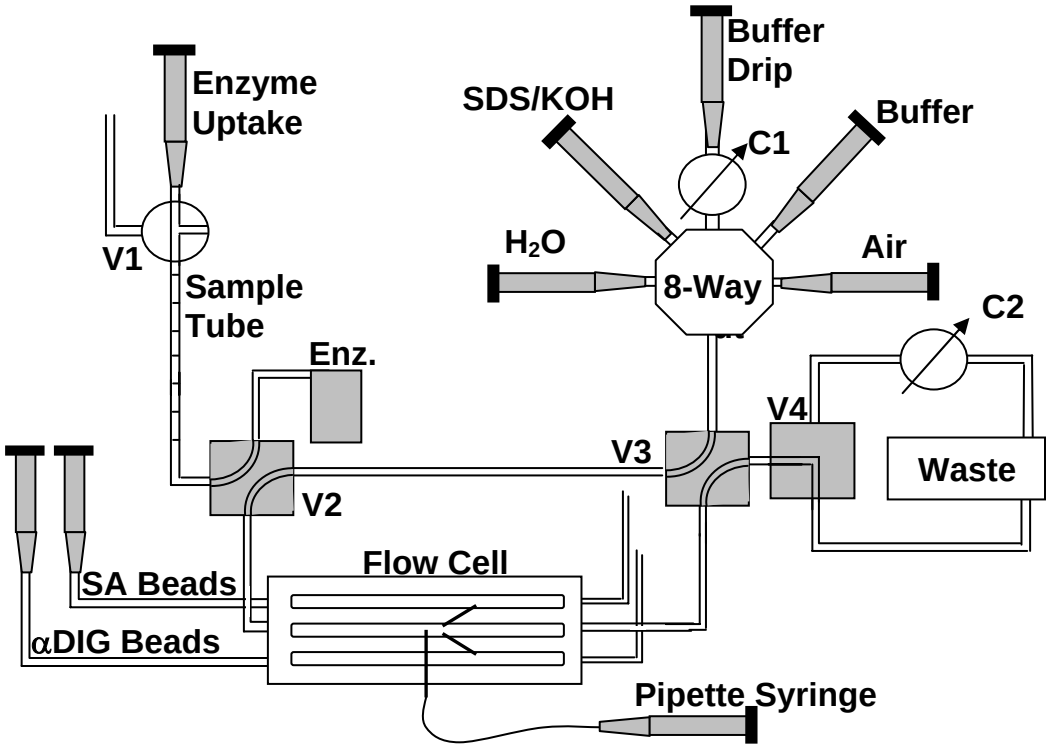

Fig. 1.6. Schematic of the flow system (not to scale). The tapered gray shapes with black caps represent syringes, either 1 mL or 10 mL, and each is labeled according to its purpose/contents. The sample tube is a section of Becton-Dickinson PE-100 tubing with 5 μL gradations marked. The valves (V1, V2, V3, V4; discussed in the text) can each be turned in 90º increments; the valve labeled "8-Way" allows the user to select one of eight inputs to flow out through the "Out" port. In practice, three of the eight were capped and unused. The crimpers (C1 and C2) allowed the user to continuously modulate the flow rate by crimping the tubing. The tubes leading out of the flow cell from the top and bottom channels are stand pipes open to the air. With the bead syringes, the level of fluid in these stand pipes controlled the flow of beads into the main channel of the "Enz."

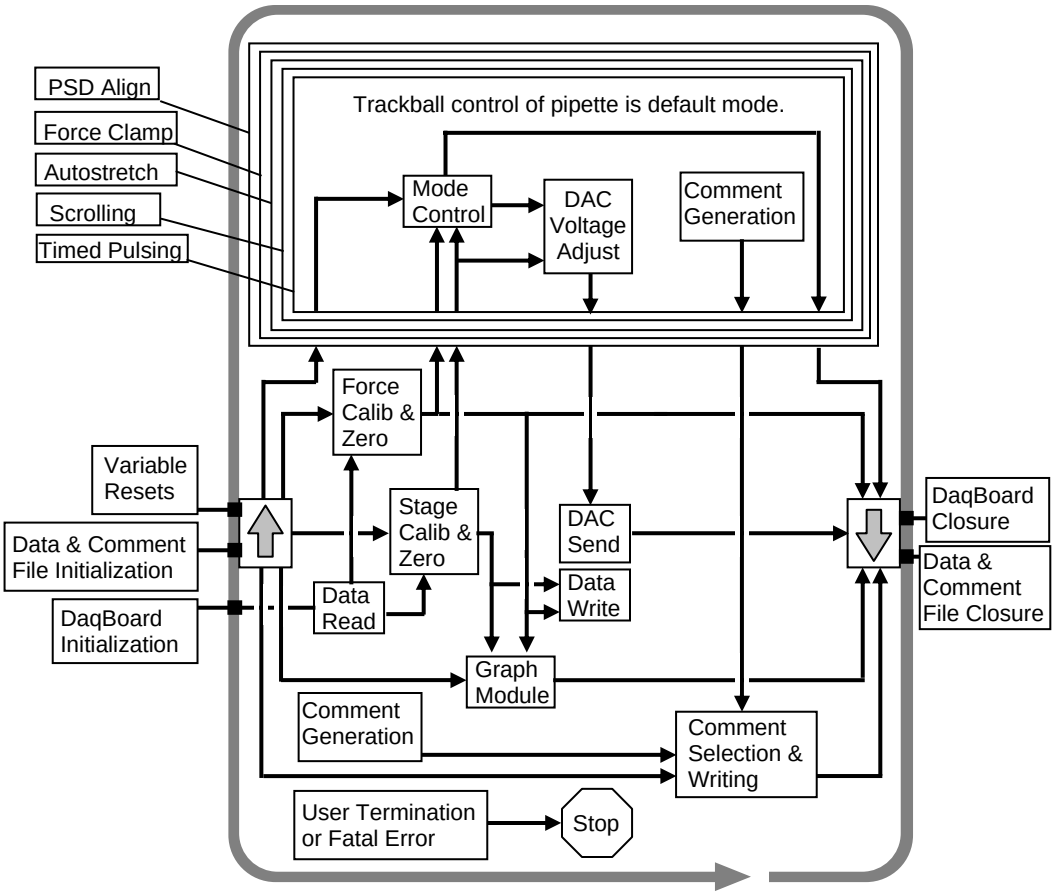

Fig. 1.7. Flow Chart for the Control Program. The 100 Hz WHILE loop is represented by the gray oval with the arrow. The different pipette control modes are throughout the stacked IF/THEN leaves, and are indicated on the left as they were situated in the program. The black arrows represent the flow of information in the program, though this diagram does not represent the many hundreds of information exchanges per iteration. The large block arrows on the sides represent the down-loading (uploading) of information from (to) the previous (next) iteration. This diagram does not reflect all the information that was repeatedly passed in this manner either; every function in the program relied on iterative information. The program could be stopped by the user or a number of fatal errors, most of the Daqboard related.

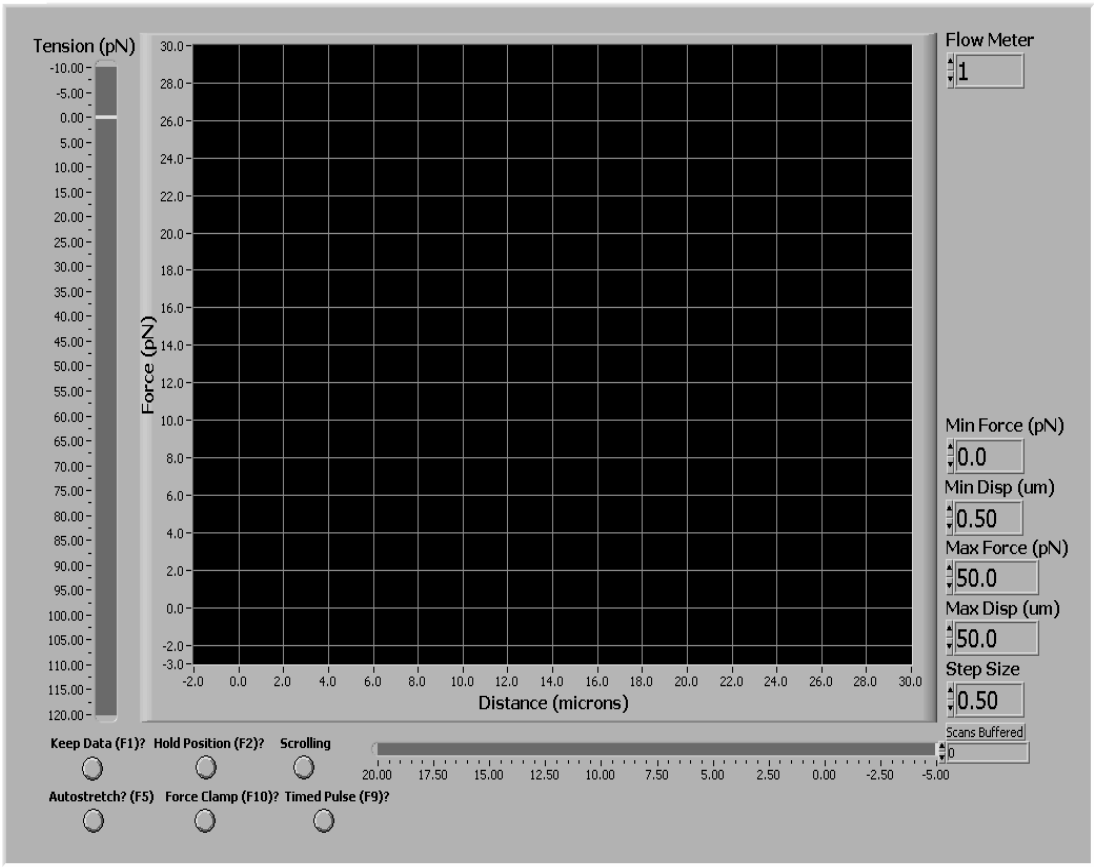

Fig. 1.8. The front of LabVIEW control program in a typical configuration. The large screen (adjustable scales one each axis) showed the force-extension curve or the force/ position trace in real time. The vertical indicator on the left indicated tension in the pulling direction, the horizontal indicator indicated tension in the off direction. The buttons in the lower left indicated in which mode the program was operating. The numerical controls to the right could be changed by the user as needed.

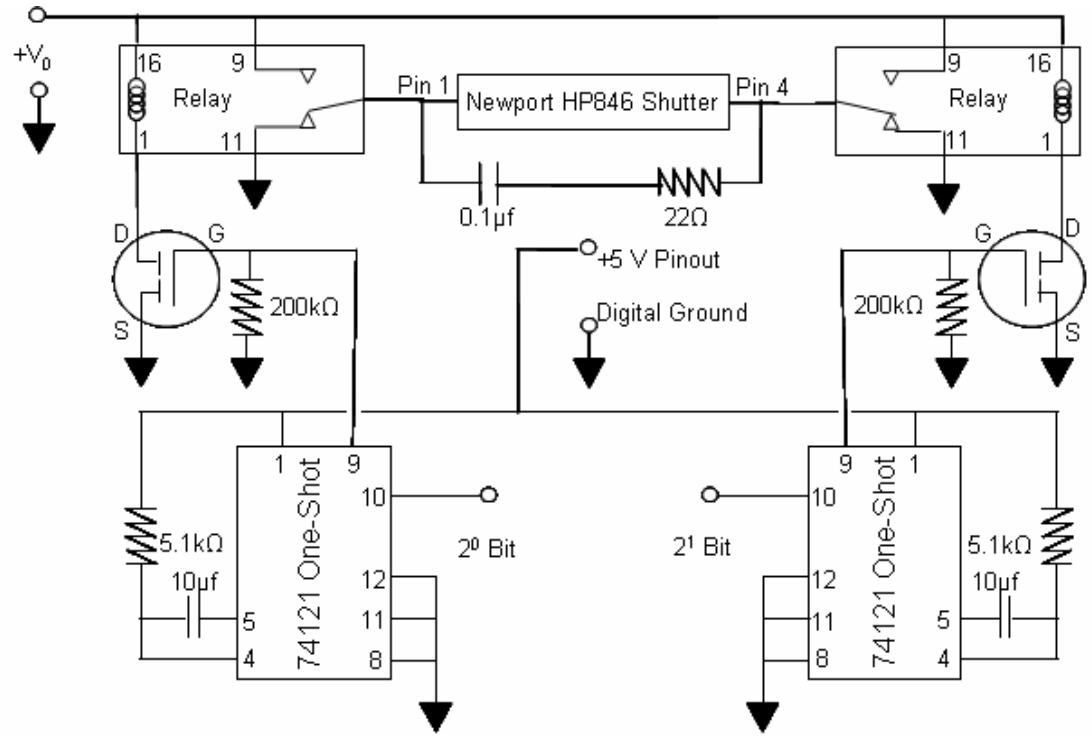

Fig. 1.9. Shutter Circuit Diagram. From this diagram, it is impossible to ascertain if the shutter is open or closed, as both relays are in the de-energized state. As discussed in the text, the shutter opens on a current pulse in one direction (Pin 1 to Pin 4) and closes on a current pulse in the other. The LabVIEW control program determined whether the shutter was to be open or closed, which the user toggled by the space bar.

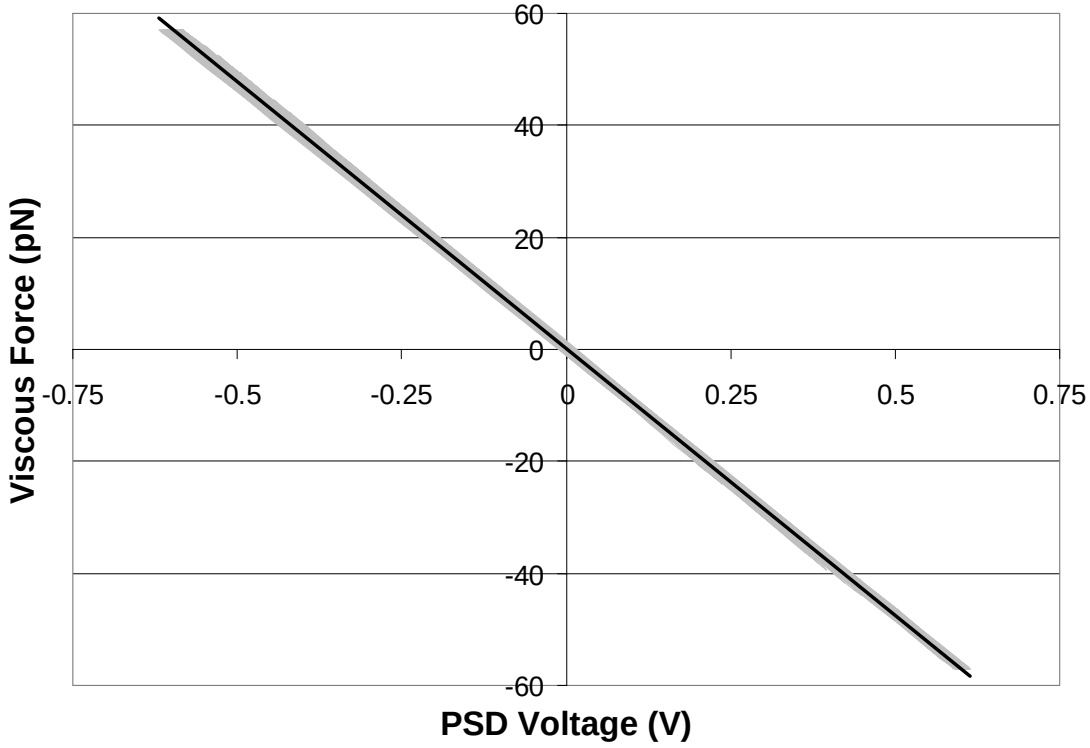

Fig. 1.10. Viscous force calibration curve. The gray lines represent 100 cycles of oscillating the nanostage with a frequency of 20 Hz and a peak-to-peak amplitude of 45 μm. The thin black line represents the linear fit, yielding a calibration for this curve of 95.4 pN/PSD volt.

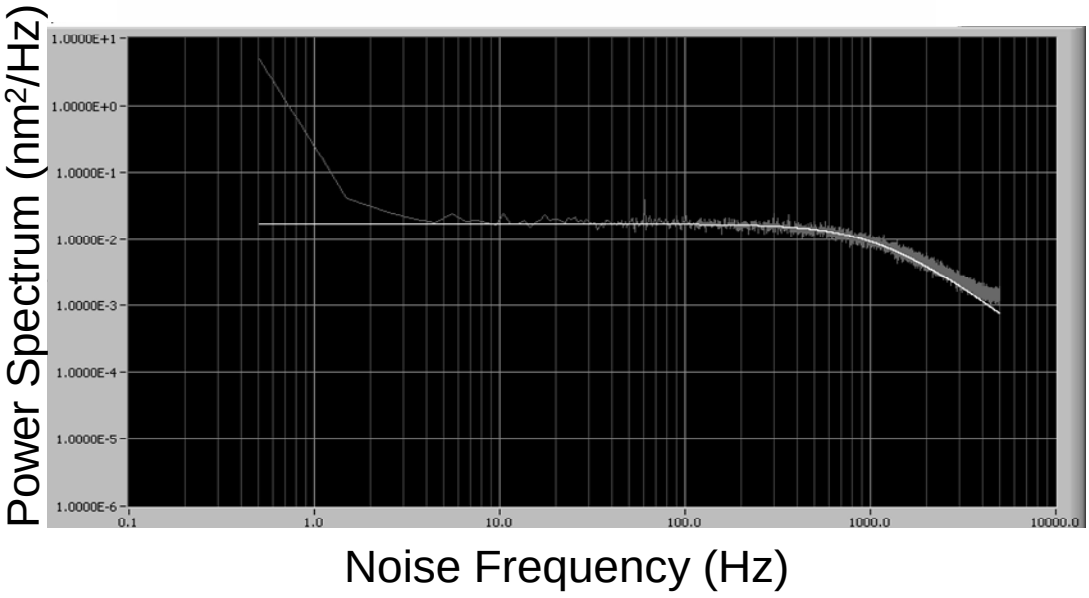

Fig. 1.11. The Fourier power spectrum for a nominally 2 mm diameter bead in the optical trap, with the frequency on a log-10 scale. The gray is the measurement whereas the white is the fit according to the Lorentzian discussed in the text. Where the linear approximations of the high frequency regime and the plateau in frequency cross is called the 'corner frequency', the expression for which contains the trap stiffness. The deviations at low frequency represent mechanical vibrations of components in the system, optical fluctuations, and long term thermal drifts.

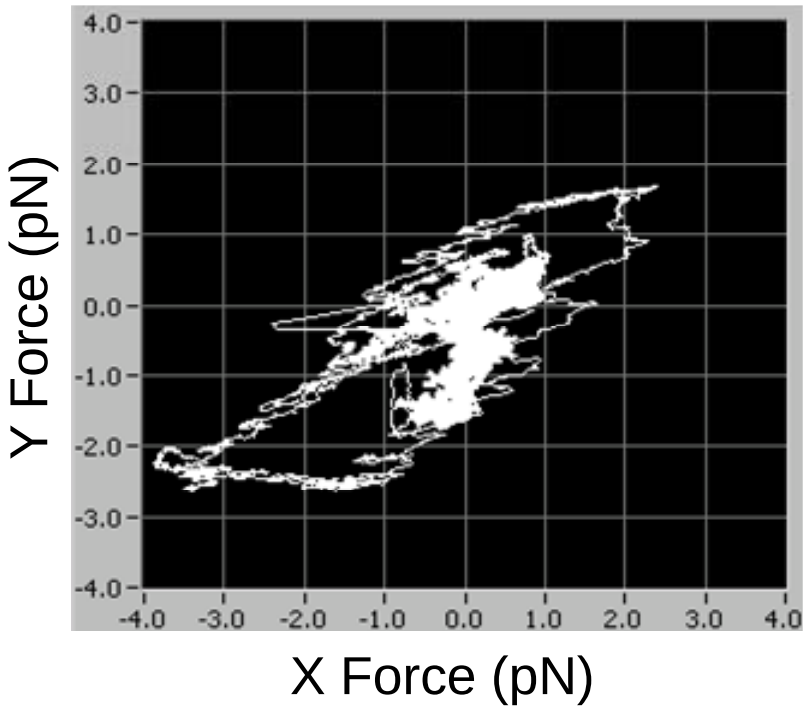

Fig. 1.12. Effective Laser Wander. The white trace is the "force" recorded over twenty minutes at 100 Hz with no bead in the trap and the buffer flow stopped. The PSD calibration during this measurement was 95 pN/PSD volt. The statistical correlation between the X and Y measurements was 0.56.

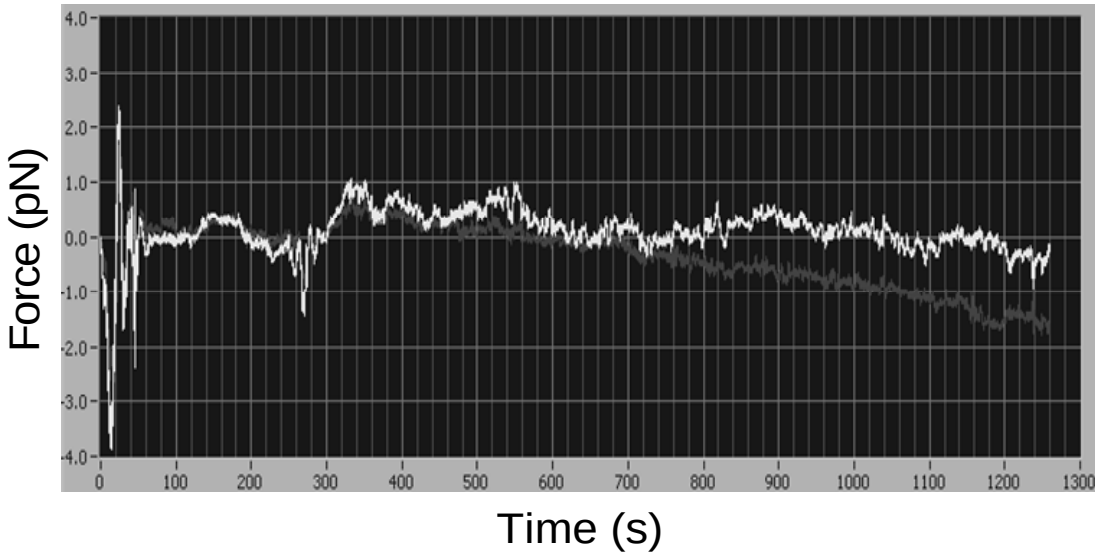

Fig. 1.13. Long time drift measurements of the force in X (white trace) and Y (gray trace) over twenty minutes taken with no bead in the trap and a PSD calibration of 95 pN/ PSD voltage. Note the downward trend of the Y trace especially, which was attributed to mechanical and thermal drift of optical components. The short term wiggles were attributed to vibrations of optical components and laser pointing instability. The large disturbance in the beginning is due to the mechanical manipulation of the stage holding the second objective lens (i.e., the collector).

Appendix 1.1 Summary of Control Program Keystrokes, Version 55

General Controls

Ctrl+r *starts the program, as does clicking the white LabVIEW arrow in the upper left, prompting the user to turn on the stage controller and then for a filename, user name, initial comment, and calibration value.*

Esc *clears the graph.*

Tab *cycles through all the numerical controls on the screen. Hitting "Enter" or a trackball key "dehighlights" a control, returning the program to normal.*

PgUp *cycles through the vertical scale ranges (usually Force, see F12).*

PgDn *cycles through the horizontal scale ranges (usually Extension, see F12).*

Home *toggles the pull direction between X and Y. When it's X, a red indicator with an "X" in it appears on the right side of the screen. X and Y direction zeroes are stored separately, so you will need to rezero when changing pull directions. Defaults to Y.*

End *stops the program, prompting the user to turn off the stage controller.*

F1 pertains to the data file and its name

F1 *toggles the "Keep Data" feature on and off.*

Shift+F1 *allows the user to change the filename. A window appears, highlighted, in which the user types the new name. The change does not take effect until* **Shift+F1** *is pressed again. Don't leave this open because string controls slow the program down.*

Ctrl+F1 *allows the user to change the user's name. A window appears, highlighted, in which the user types the new user. The change does not take effect until* **Ctrl+F1** *is pressed again. As of version 55, this is still in the program but since it has been rarely used, it might be eliminated to free up a key and, more importantly, reduce the number of string variables the program keeps each iteration. Don't leave this open because string controls slow the program down.*

Ctrl+Shift+F1 *automatically increments the filename by adding an ordinal number to it.*

F2 pertains to current or preset positions

F2 *toggles the "Hold Position" feature. When on, the "Hold Position" light will be on, allowing the user to move the trackball without moving the pipette.*

Shift+F2 *scrolls the pipette to the "Minimum Displacement" value. Useful for returning the pipette to a slack position quickly.*

Ctrl+F2 *is unused as of Version 55 (01/10/05).*
Ctrl+Shift+F2 *is unused as of Version 551 (01/10/05).*

F3 pertains to the zero position

F3 *zeroes the position (and clears the screen). As of version 45 or so (with the return of the stage in the August of '03), this records the DAC voltage, and hence it is instantaneous. Use it when the beads are touching.*

Shift+F3 *returns the stage to the position of the Current Zero. This will probably be eliminated as I don't recall ever using it regularly.*

Ctrl+F3 *is unused as of Version 55 (01/10/05).*
Ctrl+Shift+F3 *is unused as of Version 55 (01/10/05).*

F4 pertains to zeroing/calibrating the force

F4 *zeroes the force by averaging the force signal for five seconds (and clears the screen). Hold the beads about one bead diameter apart for this operation, during which all pipette motion is disabled. Due to drift, etc., it's good to rezero often.*

Shift+F4 *is unused as of Version 55 (01/10/05).*

Ctrl+F4 *allows the user to recalibrate on the fly. To use this, get a decent idea where the overstretch plateau is at the current calibration. Hitting* **Ctrl+F4** *opens a window in which the user can enter this value. Upon hitting* **Ctrl+F4** *again, the window closes and the program automatically recalibrates the overstretch plateau to 65 pN.*

Ctrl+Shift+F4 *allows the user to revert to the old calibration in case there was a mistake in recalibrating, like entering the wrong value.*

F5 pertains to autostretch

F5 *toggles "Autostretch" on and off, denoted by the red light beneath the graph.*
Before starting, make sure your values are all acceptable (see **Tab** *instructions).*

Shift+F5 *turns the pipette around when in autostretch mode.*

Ctrl+F5 *is unused as of Version 55 (01/10/05).*

Ctrl+Shift+F5 *toggles the lower limit of the autostretch between "Minimum Displacement" and "Lower Limit". Defaults to "Minimum Displacement".*

F6 pertains to scrolling the stage to set positions

F6 *centers the nanostage in the middle of its range and clears graph upon arrival. Very useful before trying to get hookups (use when moving the micrometer stage).*

Shift+F6 *corners the nanostage at its (-10 V, -10 V) input position. Good for keeping the high voltages on the nanostage's PZTs at a minimum.*

Ctrl+F6 *is unused as of Version 55 (01/10/05).*
Ctrl+Shift+F6 *is unused as of Version 55 (01/10/05).*

F7 pertains to entering comments in the text file

F7 *opens the window to make a comment and highlights it. Comment is recorded once the window is closed by hitting F7 again. Nothing is written is comment window is left blank when closed. Graphing is disabled while comment window is open and it is very important to close it when not in use, as having it open is computationally expensive.*

Shift+F7 *enters the automatic comment "Refocus" into the text file, useful to explain shifts if the data is being kept.*

Ctrl+F7 *enters the automatic comment "Flow Adjust" into the text file, useful to explain shifts if the data is being kept.*

Ctrl+Shift+F7 *enters the comment "Reagents Arrived" into the text file and is to be pressed when the interface in the sample tube crosses the red line. This is extremely important when doing cutting studies as it marks "time zero". This might be moved to Ctrl+Shift+F8 in the near future to keep all sample tube flow command on one key (01/01/05; see below).*

F8 pertains to the flow meter

F8 *marks the flow meter number (upper right of screen) in the text file. User hits it when the interface in the sample tube crosses each line. The number shown on the screen should be the next line the interface will cross.*

Shift+F8 *advances the flow meter number by one without recording the mark in the file. This is very useful.*

Ctrl+F8 *decrements the flow meter number by one without recording the mark in the file. This has not been very useful at all and will probably be removed, making room for the "Reagents Arrived" autocomment.*

Ctrl+Shift+F8 *resets the flow meter to zero. This is done automatically by a few of the other features, such as centering the nanostage and the like.*

F9 pertains to the timed pulse feature

F9 *toggles on and off the timed pulse feature. When activated, the pipette first scrolls to the "Min Disp", and then pulses at a set time (e.g., every N seconds) to the "Max Force" setting. If it encounters the "Max Force", it returns to the "Min Disp". It also pulses and resets its clock if "Reagents Arrived (Ctrl+Shift+F7)" is pushed when this feature is active. This feature is useful for slack cutting measurements...it tells us if the DNA tether is still there.*

Shift+F9 *shows the "Pulse Seconds" control, in which the user can adjust the time between pulses. Need not visible for it to take effect, but it is nice to have it invisible and out of the way when doing something completely different.*

Ctrl+F9 *is unused as of Version 55 (01/10/05).*
Ctrl+Shift+F9 *is unused as of Version 55 (01/10/05).*

F10 pertains to the Force Clamp feature

F10 *toggles the "Force Clamp" on/off. Upon activation, the force clamp light will come on and the force clamp number control will appear highlighted. The program moves the pipette to keep the measured force value (i.e., the tension in the DNA) at the value in the force clamp number control. This value can be changed and the pipette will move accordingly, though slowly.*

Shift+F10 *turns on "Force Collapse", which rapidly moves the pipette to the preset force (see* **Ctrl+Shift+F10***). Upon reaching that force, it clamps it there and acts just like the regular force clamp. This feature is extremely useful in doing anything at tension when flowing in an enzyme, allowing you to put the DNA molecule under tension right before the enzyme arrives. You can even activate this feature from within the regular force clamp, and exit it by just hitting* **F+10***, the same way you exit the regular force clamp.*

Ctrl+F10 *makes the "Force Collapse" value visible so that it can be changed. The function of this control is not affected by this control being either visible or invisible, but it is nice to have it invisible if doing something completely different.*

Ctrl+Shift+F10 *is unused as of Version 55 (01/10/05).*

F11 pertains to the PSD Alignment feature

F11 *toggles the "PSD Align" on and off. Use the X and Y micrometers on the second objective stage to center the white spot, which represents the X and Y PSD voltages. The red spot is simply the X and Y voltages each multiplied by five. The blue bar represents laser intensity. Before using this feature,*

however, make sure the laser spot looks good on the adjustable aperture with the IR viewer. Almost all other program functions are disabled when in "PSD Align" mode.

Shift+F11 *simply toggles the scale of the "PSD Alignment" screen between ±10 volts and ±1 volts. Defaults to the larger scale when "PSD Align" is activated.*

Ctrl+F11 *is unused as of Version 55 (01/10/05).*
Ctrl+Shift+F11 *is unused as of Version 55 (01/10/05).*

F12 pertains to the Time Series plot feature

F12 *toggles the "Time Series" plot on and off, allowing the user to view either Force or Extension as a function of time in real time. "Clear Graph" and the scalings all still work when this is activated.*

Shift+F12 *toggles between Force and Extension as the vertical axis.*

Ctrl+F12 *is unused as of Version 55 (01/10/05).*
Ctrl+Shift+F12 *is unused as of Version 55 (01/10/05).*

These key assignments are subject to change in order to accommodate any forthcoming features that would make sense on a particular key that is already full (e.g., moving "reagent arrival" to F8 in some fashion allowing for a different autocomment).

References for Chapter 1

1. Svoboda, K., Schmidt, C.F., Schnapp, B.J., & Block, S.M. (1993) Direct Observation of Kinesin Stepping by Optical Trapping Interferometry. *Nature* **365**: 721-727.

2. Finer, J.T., Simmons, R.M., & Spudich, J.A. (1994) Single Myosin Molecule Mechanics: Piconewton forces and Nanometre Steps. *Nature* **368** :113-119.

3. Rief, M, Rock, R.S., Mehta, A.D., Mooseker, M.S., Cheney, R.E., & Spudich, J.A. (2000) Myosin-V Stepping Kinetics: A Molecular Model for Processivity. *PNAS* **97**: 9482-9486.

4. Perkins, T., Smith, D.E., & Chu, S. (1994) Direct Observation of Tube-Like Motion of a Single Polymer Chain. *Science* **264**: 819-822.

5. Bustamante, C., Smith, S. B., Liphardt, J., & Smith, D.E. (2000) Single Molecule Studies of DNA Mechanics. *Curr. Opin. Struct. Biol.* **10**:279-285.

6. Bennink, M.L. (2000) *Force Spectroscopy of Single DNA-Protein Complexes.* Doctoral Thesis, University of Twente.

7. Husale, S., Grange, W., & Hegner, M. (2002) DNA Mechanics Affected by Small DNA Interacting Ligands. *Single Molecules* **3**: 91-96.

8. Wolffe, A. (1997) *Chromatin Structure and Function*. Academic Press, San Diego.

9. Gittes, F. & Schmidt, C. (1998) Signals and Noise in Micromechanical Measurements. *Meth. Cell Biol.* **55**: 129-156.

10. Smith, S.B, Cui, Y, & Bustamante, C. (1996) Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science* **271**: 795-799

Chapter 2. A General Method for Manipulating DNA Sequences From Any Organism with Optical Tweezers

2.1. Abstract

Here we describe and characterize a method for manipulating desired DNA sequences from any organism with optical tweezers. Molecules are produced from either genomic or cloned DNA by PCR using labeled primers and are tethered between two optically trapped beads. We demonstrate that human, insect, bacterial, and viral sequences ranging from ~10 to 40 kbp can be manipulated. Force-extension measurements show that these constructs exhibit uniform elastic properties in accord with the expected contour lengths for the targeted sequences. Detailed protocols for preparing and manipulating these molecules are presented, and tethering efficiency is characterized as a function of DNA concentration, ionic strength, and pH. Attachment strength is characterized by measuring the unbinding time distribution as a function of applied force.

2.2. Introduction

Protein-DNA interactions play a critical role in the molecular biology of all organisms. For example, the ~3.3 billion basepair (bp) human genome is estimated to code for several thousand DNA binding proteins, such as transcription factors, nucleases, repair and structural proteins, topoisomerases, and DNA polymerases. Most DNA-protein interactions are sequence specific, meaning that a protein recognizes a particular sequence at which it binds the DNA. A wide variety of methods exist for studying such sequence specific protein-DNA interactions, including DNase footprinting, sucrose gradient sedimentation, gel mobility shifts,

fluorescence spectroscopy, imaging by electron microscopy, and x-ray crystallography. Over the last decade the mechanical manipulation of single DNA molecules has been developed. Manipulation of DNA by optical tweezers was pioneered by Chu and coworkers, and extended by Bustamante and coworkers (1-4). This method has also been applied to study fundamental biochemical processes, including transcription, replication, chromatin unraveling, viral DNA packaging, and helicase translocation (5-11). This method is obviously also the method of this dissertation, used here to study a wide variety of properties of two site restriction enzymes, the dynamics of DNA looping (utilizing a two site restriction enzyme), and the properties of nucleosomes assembled by the ATP-hydrolyzing ACF system on a random DNA template. Each of these will be examined in later chapters, but in this chapter a method for generating any desired DNA sequence (with handles) will be discussed. Also, the characterization of the DNA tethering efficiency, force-extension curves, and attachment strength is presented.

2.3. Methods

2.3.1. Genomic DNA Preparation

E. Coli DNA was obtained by growing a 2 ml culture of DH5α competent cells (Invitrogen Corp.) in LB broth (Miller) (Fisher Scientific) overnight at 37 °C. The cells were pelleted at ~3000*g* in a microcentrifuge tube and the pellet was lightly dabbed with a sterile cotton swab. DNA was then extracted and purified using the QIAamp DNA Blood Mini Kit (Qiagen) following the "buccal swab spin protocol" per the instructions with the following modifications: 10 mg/ml proteinase K (Roche

Biochemicals) was substituted for the supplied protease, a total of 1% SDS was included during the lysis step, and the lysis was incubated overnight. *Drosophila* DNA was obtained by growing embryos as described previously (13). Embryos were frozen in liquid nitrogen and stored at –80 °C. ~100 μl of embryos were thawed and dabbed with a sterile cotton swab. The DNA was then extracted and purified using the QIAamp DNA Blood Mini Kit (Qiagen) following the "buccal swab spin protocol," substituting 10 mg/ml proteinase K for the supplied protease. Human DNA was obtained by firmly scraping the inside of one of our cheeks about ten times with a sterile cotton swab. The DNA was then extracted and purified using the QIAamp DNA Blood Mini Kit (Qiagen) following the "buccal swab spin protocol" without modification. λ DNA purchased from NEB was used directly. Extracted genomic DNA samples were characterized by UV spectroscopy (absorbance at 260 nm) and gel electrophoresis. These measurements indicated that DNA concentration following purification ranged from 5-50 ng/μl, and that fragment lengths were predominantly ~15 to 25 kbp in length. This aspect of the project was carried out by D. Fuller, P. Recouvreux, and A. Dupont.

2.3.2. BAC DNA Preparation

The UCSC genome browser and alignment software was used to identify an appropriate BAC clone bracketing sequence of interest (14). We chose the clone CTD-2240D16 (from Caltech Library D) containing sequences from chromosome 14 that code for human tissue plasminogen activator. A culture of E. coli carrying this clone was obtained from Invitrogen. A pipette tip was dipped in this culture and used

to inoculate 2 ml of LB broth in a 15 ml cell culture tube, which was incubated overnight at 37 °C on an orbital shaker. BAC DNA was then extracted and purified using the FastPlasmid Mini Kit (Eppendorf) per the manufacturers instructions with the following modifications: 600 µl of lysis solution was used, the room temperature incubation was 5 min, 195 µl isopropanol was added to the lysate, and the elution buffer was preheated to 65°C. This aspect of the project was carried out by D. Fuller, P. Recouvreux, and A. Dupont.

2.3.3. Synthesis and Labeling by PCR

The UCSC genome browser was used to identify human and *Drosophila* DNA sequences of interest, and the NCBI databases were used to identify *E. coli* and λ sequences. PCR primers were selected using GeneRunner software (except for those targeting lambda and human sequences, which were recommended by Eppendorf). The selected templates of interest are identified in Table 2.1. All primers were chosen to have a melting temperature of ~62-68 °C to limit necessary modifications of reaction conditions. Forward primers were labeled at the 5' end with biotin-TEG and reverse primers were labeled at the 5' end by digoxygenin (DIG) (via an amino-C6), such that a DNA molecule could be tethered as normal. All labeled primers were purchased from Operon Biotechnologies. All PCR reactions were carried out using the Triplemaster PCR system (Eppendorf), which combines Taq DNA polymerase, a proof reading enzyme, and a processivity enhancing buffer additive.

The basics of the PCR labeling and replication scheme are shown in Fig. 2.1. Presented in any basic biochemistry textbook, PCR relies on the heat denaturation of

DNA (i.e., separation of the strands), followed by the annealing of primers to the desired sequences upon cooling. The primers (typically 20 to 30 bp long) bind to their shorter sequences much more readily than do the two strands to each other because of the large statistical difference of matching dozens of basepairs compared to thousands. With a primer bound to each strand, the Taq polymerase covalently incorporates free nucleotides to complement the nucleotides of the single strand, yielding double stranded DNA. The fact that the polymerase is thermally stable is important because this heating-cooling-reaction (i.e., melting-annealing-polymerization) cycle is repeated many times over to amplify the amount of DNA. After a few cycles the desired product (i.e., the stretch of DNA between the primers) will be amplified overwhelmingly more so than any of the byproducts with long overhangs (i.e., single stranded out to the ends of the original template sequence), though such byproducts may be indistinguishable in the tweezers as they will have the same handle locations as the desired product. Nonetheless, as the primers will mark the ends of the desired amplified sequence, using labeled primers ensures that we have handles at the end of our DNA molecules, as shown in Fig. 2.1.

50 μl PCR reactions were carried out using reagent concentrations recommended by Eppendorf. No supplemental Mg^{++} was added to the reaction buffer. Reactions were run in a 24-well Hybaid PCR Sprint thermocycler using 200 μl thin-walled PCR tubes (Fisher Scientific). Because some reactions required tuning to obtain optimum results, the thermocycling parameters and quantity of template DNA used in each reaction varied; these parameters are listed in Ref. 12.

Table 2.1. Parameters of the different DNA templates.

Organism	Length (bp)	Genes	Source	%GC
λ Phage (virus)	10,051	P9-p20	NEB	42.4
Drosophila (insect)	14,001	Acf1, CG2118	Embryos	45.6
Human	15,138	Tpa	cheek cells	50.8
Human	15,138	Tpa	BAC clone	50.8
E. coli (bacteria)	25,340	TopA-sapA	Liquid Culture	49.6
λ Phage (virus)	40,368	P9-p78	NEB	48.9

2.3.4. Optical Tweezers

Two different optical tweezers instruments were used for measurements. The first is the tweezers used throughout this dissertation. In the second, two beads are held in two separate optical traps and the DNA is stretched by displacing one trap by use of an acousto-optic deflector, similar to the configuration used in studies of single actin-myosin interactions (15). The dual beam tweezers was built here by J.P. Rickgauer and A. Schweitzer, and is described in greater detail in Ref. 12.

2.3.5. Bead Preparation

The following bead preparation procedures were used throughout all of the experiments comprising this dissertation, save for minor adjustments in storage buffer composition such as NaCl or BSA concentration. For the DNA carrying beads, 200 μl of 0.5% (w/v) streptavidin (SA) coated beads (Spherotech) were washed twice to remove any free streptavidin by pelleting at 10,000g in a microcentrifuge tube and resuspending them twice in 200 μl PBS pH 7.4, 1 M NaCl with 0.1 mg/ml BSA. To prepare αDIG coated beads, 200 μl of 0.5% (w/v) Protein G coated polystyrene beads (Spherotech) were washed twice by pelleting at 10,000g in a microcentrifuge tube and resuspending them twice in 200 μl PBS buffer. After the second wash, the beads

were resuspended in 20 μl PBS and 5 μl 200 μg/ml of αDIG (Roche) was added. The beads were incubated on a rotisserie (Barnstead Labquake) at room temperature for at least 30 minutes (usually much longer) and then washed three times in 200 μl PBS and resuspended in 20 μl PBS. 3-6 μl of these beads were loaded into a 1 ml tuberculin syringe (Becton Dickinson, Co.) of the experimental buffer for injection into the microfluidic chamber. Typically, we made new αDIG beads every few days. Streptavidin beads lasted longer; typically a new batch was prepared weekly.

2.3.6. DNA Tethering

Labeled DNA was first attached by one end to the steptavidin beads in a bulk reaction as follows: 3 μl of appropriately diluted DNA (ranging from ~2.5 to 500 ng/μl, such that DNA:bead stochiometry varied from ~1:1 to ~200:1, as discussed in results) was mixed with 27 μl of beads and incubated for 30-60 minutes at room temperature on a slowly rotating rotisserie (Barnstead Labquake). 5-10 μl of these beads were diluted in 0.5 ml of PBS (or the experimental buffer at the time) and loaded into a syringe for injection into the microfluidic chamber. The PCR generated DNA preparations were fairly concentrated. The ligation labeling and fill-in labeling reactions of other experiments in this thesis were less so. Consequently, the DNA: bead loading ratios were different. Indeed, even with different preparations of nominally the same labeled DNA, the best loading ratios were empirically determined and tweaked to account for degradation in both the DNA and the bead preparations.

DNA tethers were formed in the flow chamber as follows: first, a single αDIG bead was trapped in the first optical trap and then transferred onto the tip of the

micropipette by applying suction. Once the αDIG bead was secured on the pipette, the suction was "backed off" so that there was minimal fluid flow into the pipette around the bead (i.e., through imperfections in the crown of the pipette). Too much suction of fluid into the pipette during an experiment interfered with tethering and skewed the force-extension results (i.e., DNA could get sucked into the pipette around the αDIG bead). After the αDIG bead was secured and the suction equalized, a streptavidin bead (carrying DNA) was trapped in the optical trap and brought nearly in contact with the pipetted αDIG coated bead for ~10 seconds in an attempt to form a tether. Usually the same αDIG bead was used in up to five trials before discarding it; typically an αDIG bead was unable to tether DNA after ten trials.

All measurements of tethering efficiency were done using the single beam optical tweezers system. For the trials done with varying salt (NaCl), the binding of the DNA to the streptavidin beads was done in the same manner described above, except these beads were then diluted in 0.5 ml of 20 mM Tris-HCl, pH 7.8, with 0 to 2M NaCl, instead of in PBS buffer. For the trials with varying pH, the following 10 mM buffers were used: acetate (pH 4), citrate (pH 5.6), phosphate (pH 7), Tris-HCl (pH 8.5), carbonate (pH 9.9), phosphate (pH 11.8). An appropriate amount of NaCl was added to each so as to keep the total ionic strength at 150 mM (16).

2.4. Results and Discussion

2.4.1. Manipulation of Desired Sequences

To demonstrate that we could manipulate specific DNA sequences from a variety of organisms, we targeted six arbitrary sequences in the human, *Drosophila*,

E. coli, and bacteriophage λ genomes (Table 2.1). Genomic DNA was purified from each organism and PCR amplification and labeling were carried out as described in methods. Each of these DNA constructs was successfully manipulated by optical tweezers, by following the tethering protocols described above. High precision extension measurements confirmed that the tethered molecules had lengths consistent with the targeted sequences, as discussed below. Based on these results, from a broad variety of DNA sources, we expect that this method can be used to prepare virtually any desired sequence from virtually any organism for use in the optical tweezers.

We also showed that cloned DNA could be used to generate desired sequences. BAC clones (in *E. Coli*) spanning the human genomes, and genomes of many other organisms, have been produced during genome sequencing projects. For example, BAC clones that cover the entire human genome have been produced at Caltech and Children's Hospital of Oakland Research Institute. To demonstrate that these could be used, we targeted the same human sequence that we targeted when using genomic DNA, which contains the gene for tissue plasminogen activator. This construct was successfully manipulated and exhibited similar elastic behavior as the construct generated using genomic DNA, as described in more detail below.

Prior to use in the optical tweezers PCR products were analyzed by agarose gel electrophoresis (shown in Ref. 12). Cycling conditions were tuned to obtain the strongest signal possible in a single band (also shown in Ref. 12), although weaker secondary bands were often observed before optimization of the cycling parameters. In some cases it proved difficult to obtain a completely pure product even after these adjustments (e.g., the 25.3 kbp *E.Coli* sample in Fig. 2.10). Presumably the

amplification of undesired products could be further minimized through further optimization of the primers and reaction conditions. Fortunately, secondary products represented a small fraction of the total and were always of significantly shorter length than the desired products. In practice, tethering of undesired products in the optical tweezers was rare (<10% in the worst case) and could easily be distinguished as having significantly shorter lengths.

2.4.2. Tethering Efficiency

As DNA molecules are tethered by bringing two beads into contact, it is important to optimize the tethering conditions: Too little DNA results in beads failing to tether, while too much DNA results in the tethering of multiple molecules on single beads. As the number density of the beads is very low (~0.5 x 10^9 per ml), most samples of DNA prepared by this PCR labeling method must be heavily diluted to have a high likelihood of obtaining single tethers. Here, DNA is first incubated with streptavidin beads in bulk for ~ 60 minutes to tether the biotin labeled ends. In theory, one could simply set the bead:DNA stoichiometry to ~1:1 and wait for the binding reaction to reach equilibrium, whereupon the distribution of DNAs per bead would be expected to follow a Poisson distribution with ~37% of beads having exactly one DNA tethered. However, our experience indicates that even after ~ 10 hours of incubation fewer tethers are detected than predicted. Some molecules may have been improperly labeled, degraded, or adhered in a manner that prohibits binding to the second bead. Thus, in practice we find it convenient to use a 3-fold excess of DNA and an incubation time of ~30 to 60 minutes. Further binding is

essentially stopped by the ~100-fold dilution prior to injection into the flow chamber. Systematic measurements show that tethering efficiency for the 10.1 kbp λ DNA construct varies from zero to nearly 100% multiple tethers as the DNA:bead ratio is increased from 1:1 to 200:1 (Fig. 2.2). In this particular titration, the 3:1 ratio, yielding ~30% single tethers and <5% multiple hookups, proved to be convenient.

In some experiments one may test whether a single DNA is tethered by elastic measurements in the l tweezers. For these experiments it may be convenient to use a higher DNA:bead ratio. However, when studying protein-DNA complexes, the elastic signature for a single tether may be altered. In this case, it is better to accept a lower tethering efficiency in order to minimize the chance of multiple DNA tethers, which may invalidate certain data sets. We often like to use conditions where about one in five trials yield a hookup. Our results on tethering efficiency (Fig. 2.2) are useful as a guide as to what reaction conditions to use. However, there is often significant variation depending on the particular DNA and bead samples being used. We often find it necessary to titrate the amount of DNA up or down by a factor of 3 to find optimal tethering conditions.

Depending on the biochemical process being studied, one may wish to tether DNA molecules under a variety of solution conditions. Here we have characterized the dependence of tethering efficiency on salt (NaCl) and pH. In these experiments, the first attachment (biotin-streptavidin) was formed during a ~ 60 minute incubation in standard buffer conditions, as described above, while attempts to form the DIG-αDIG linkage were made under conditions of varying salt and pH. As shown in Fig. 2.3, tethering worked quite well from 0 to 2 M NaCl, with the highest efficiency at

150 mM NaCl. This efficiency may be attributed to electrostatic screening, which presumably allows the DNA and beads, which are both negatively charged, to approach more closely. The drop in efficiency at very high salt is consistent with known weakening of antibody-antigen interactions under these conditions (17). We note that problems with tethering often occur when using divalent cations, such as Mg^{++}, above ~5 mM. These conditions tend to cause non-specific adhesion of the beads to each other and/or to the DNA, and in this case it is helpful to include a blocking agent such as bovine serum albumin (BSA). We usually include 0.1 mg/ml BSA during tethering incubations. While BSA mitigates non-specific adhesion, we found that increasing the concentration of BSA further to 1 mg/ml reduced tethering efficiency to inconvenient levels, essentially eliminating it altogether. This is presumedly due to the streptavidin binding sites being overly blocked or, if any DNA is attached, it being tangled in the overabundance of BSA. DNA tethering worked at pH values ranging from 5.6 to 9.9, with optimal results at pH 7 to 8.5 (Fig. 2.4). These experiments assume that the tethering efficiency is not dependent on the buffering agent. While this assumption is potentially faulty, tests between Citrate and Acetate at pH = 5.6 were statistically similar (not shown). Tethering did not work at all at pH = 4 or 11.8, presumably because of the denaturation of the DNA (18) and proteins under these extreme conditions. Non-specific adhesion of the beads to each other and the DNA was often observed at pH 5.6 and 9.9. Indeed, at the lower pH values, the protonation of the DNA led to an electrostatic adhesion between DNA and the beads. This adhesion was used in ‘DNA combing’ experiments in which overhanging single-stranded ends were adhered to microscope slides and then the

DNA was extended in a gentle flow (19). At pH = 4 there was also a precipitation of BSA, which clogs the flow chamber, adhere to the beads, and interfere with optical trapping. A small amount of BSA precipitation was also observed at pH 5.6, although DNA manipulation was still workable at this pH. A summary of the BSA precipitation observations is presented in Table 2.2.

Table 2.2. Summary of BSA Precipitation Observations

0.1 mg/ml	Tris·HCl (7.8)	Tris·Acetate (7.8)	Citrate (5.6)	Acetate (5.6)	Acetate (4.0)
No BSA	Clean	Clean	Clean	Clean	Clean
BSA	Clean	Clean	A little 'dirty'	A little 'dirty'	Very dirty

Comparing the two sets of similar pH trials indicates that the precipitation was not due to anything inherent in the acetate or acetic acid. Furthermore, the Acetate (5.6) and Acetate (4.0) results indicate that the BSA precipitation is exacerbated by lowering the pH, which could have consequences in the *in situ* experiments and even bead preparations if one is using the attachment scheme of Allemand, *et al*. (19).

2.4.3. Force-Extension Measurements

To test the repeatability of optical tweezers measurements on the prepared DNA constructs, force-extension measurements were performed on ensembles of molecules for all five of the templates produced by the PCR methodology using both the single beam-pipette tweezers and the dual beam tweezers. The higher precision results of the dual beam tweezers are discussed in Ref. 12. The measurements taken by the single beam-pipette tweezers, however, help characterize the precision of the instrument used throughout this thesis. An example of each of the five constructs is

shown in Fig. 2.5. These are pulls to 50 pN; another set of measurements in which each molecule was pulled at ~ 75 nm/s until the DIG bond broke are discussed below. The nominal contour lengths of each type of molecule are 10,051 bp (3.42 μm) for the 10 kbp λ fragment, 14,001 bp (4.76 μm) for the *drosophila*, 15,138 bp (5.15 μm) for the human-pBAC, 25,340 bp (8.62 μm) for *E.coli*, and 40,368 bp (13.73 μm) for the 40 kbp λ fragment. Experimentally, the contour length was quickly determined by fitting the nearly linear, high force portion (~20 to 50 pN) of the force-extension curve. Where the fit crossed the the F = 0 axis was called L_0. Comparisons of this method with a full fit of the Worm Like Chain model (discussed below) reveal less than 10% difference even in the worst case. The variation in L_0 from molecule to molecule is due to the fact that DNA molecules may attach at any point on a bead and a bead held by the micropipette is not free to rotate. This effect leads to an uncertainty in the absolute molecular extension on the order of the radius of the bead (~1000 nm or ~3000 bp), as illustrated schematically in Fig. 2.6. Also, the pipette may not be at the same depth as the trapped bead (i.e., different focal planes), or the pipette may not be directly 'under' the trapped bead (see Fig. 2.6). In the dual beam tweezers both beads are free to rotate and both traps share the same focal plane (i.e., both beams are formed by the same objective), so these concerns are avoided. An additional small amount of variation in L_0 may come from the inherent variation in the diameter of the beads. They have a reported standard deviation in diameter of ~2% (~50 nm), as determined by the manufacturer via transmission EM data.

The distributions of measured contour lengths and DIG-αDIG bond rupture forces are shown for each template in Figs. 2.7 through 2.11. We compared

measurements on the 15.3 kbp human DNA construct prepared in two different ways: from the genomic DNA from cheek cells versus from the BAC clone. The sets of force extension curves for these two samples fall closely together (after correcting for L_0 differences), and the distributions of measured L_0 values were very similar. It also worth noting that the measured rupture forces for every template were similar, consistent with the notion that the tether breakage is indeed the DIG-αDIG bond.

For every template, the measured L_0 typically had peak widths of several hundred nanometers, as discussed above. Comparable length measurements on the dual beam tweezers yield a L_0 distribution with a width of only ~ 65 bp (~ 22 nm), which can be attributed to the variance in bead diameter. More importantly, these measurements indicate that lengths observed are indeed the lengths expected. An exception is the *E. Coli* template, the results for which are shown in Fig. 2.10. In the gel electrophoresis to analyze the final product of the PCR reaction (shown in Ref. 12), there was a faint, shorter band present. We believe that the small peak in the L_0 distribution centered about a lower L_0 value represents this secondary product. Such a product could originate from one of the primers binding to a site that *almost* complements the primer, say 18 or 19 of 20 base pairs for example, and represents a shorter interprimer distance, as discussed above. Evidence for this is also present for the human-pBAC trials in Fig. 2.9.

Next we checked that the molecules were behaving in a manner consistent with the targeted construct lengths. The elasticity of DNA molecules has been shown to agree with the behavior predicted by the Worm-Like Chain (WLC) model (4,20-23). In this model, the fractional extension of a molecule (defined as the end-to-end

distance divided by the contour length) is a universal function of the applied force. The extension of molecules of different lengths at a given force is therefore expected to be proportional to the contour length (24). As discussed above, the difference in tethering length means that effectively the contour length of the DNA is also a fit parameter for force-extension curves on the one-beam tweezers. On the two beam trap, the contour length can be regarded as a constant, as long as the template length is known. Nonetheless, with the contour length as a fit parameter, the WLC model can be well fit to force-extension data from the one beam tweezers, yielding very reasonable parameter values (fits not shown). With the two beam trap, fits of the WLC to the force-extension data confirmed that our method of preparing and tethering DNA sequences yields optical tweezers measurements that are highly repeatable and consistent with the targeted sequences.

It is known that for <100 bp different DNA sequences have different conformations and bending/torsional rigidities (25,26). For example, significant bending occurs when AT tracts are repeated in phase with the helical pitch of the DNA (27). Differences in curvature of the DNA due to local sequence variations are of relevance to both DNA looping and nucleosome stability (28). However, the measurements noted here probe the global elasticity of long DNA molecules with tens of thousands of basepairs. Two long, random DNA sequences with equal GC-AT content would not have different global elastic properties because local variations would average out (29). The constructs measured here are not purely random sequences, but they have fairly balanced GC versus AT content (Table 2.1). Even to within the measurement precision of the dual beam tweezers, we find that, as has been shown for λ DNA,

these different sequences exhibit the same elasticity (24). With further improvements in instrument resolution of the dual beam tweezers, it may be of interest to compare elasticity measurements for sequences with skewed GC-AT content, or for sequences having long stretches of repeated DNA bending motifs. Due to the mechanical and thermal coupling of the pipette to the 'outside world', the rotational constraint on the pipetted bead, and the fact that the single beam tweezers does not perform a differential force measurement (i.e., it is especially sensitive to laser drift), such high precision measurements with a single beam tweezers are likely impossible.

2.4.4. Attachment Strength

When a tension > 30 pN is applied to a single DNA we find that it usually unbinds from the beads in less than a minute. Molecules can sometimes be stretched to the overstretching transition point at ~65 pN, but at such forces the link usually breaks within a few seconds. In cases where one wishes to study protein-DNA interactions under high force, this unbinding may interfere with measurements. However, many experiments do not require application of such high forces for long periods. As it has been shown that unbinding does not happen as fast when DNA is tethered at each end by biotin-streptavidin (4), we attribute the weak link to the DIG-αDIG bond or the protein G-αDIG interaction on the surface of the bead. Previous unpublished results, however, indicated that cross-linking the αDIG and protein G did not yield much stronger tethers than without crosslinking. While connecting each end of the DNA via biotin-streptavidin provides a strong linkage, it has the disadvantage that both ends of the DNA are likely to bind to the same bead when tethering the

DNA. This problem may be avoided by tethering the DNA in a flow, but this requires flowing in a solution of free DNA molecules into the chamber, which is inconvenient and very inefficient with regards to time and materials. In any case, the DIG-αDIG link usually provides a sufficiently strong linkage for many protein-DNA interaction experiments to be carried out.

We characterized the strength of the DIG-αDIG link by sharply ramping the force to a certain value and measuring the time it took for the tether to break. This measurement was repeated on an ensemble of 10.1 kbp λ molecules at each force to determine the distribution of unbinding times. An example is shown in Fig. 2.12. Note the sharp force ramp initially and the sudden drop when the DIG-αDIG bond broke. The time is reckoned from the initial instant at the prescribed force to the drop, as denoted by the arrows in the figure. As the unbinding events are thermally activated events analogous to a Kramer's tunneling of sorts, the time intervals for unbinding are expected to follow an exponential distribution $P(t) \sim exp(k_{off}\, t)$. The distributions of unbinding times are shown in Fig. 2.13. Indeed, each distribution was well fit by this distribution, yielding a value of k_{off} at each force. The fitted values of k_{off} were $0.16 \pm 0.05\ s^{-1}$, $0.26 \pm 0.05\ s^{-1}$, $0.34 \pm 0.03\ s^{-1}$, $0.51 \pm 0.03\ s^{-1}$ at $F = 30$, 40, 50, and 60 pN, respectively. These k_{off} are plotted as a function of force in Fig. 2.14. Also shown in Fig. 2.14 are the fractions of tethers that did not reach the prescribed force as a function of that prescribed force. Naturally, the percentage rises sharply with increasing force; the line on the plot is intended only as a visual aid and is not intended to imply a particular functional form.

These measured dissociation rates k_{off} as a function of forces can also be used to estimate the dissociation rate at zero force, k_0. In the simplest model, the rupture of a molecular bond may be described as occurring along a single reaction coordinate and across a single activation energy barrier. If an external force, F, is applied the activation energy is decreased by an amount Fd where d is the distance from the free energy minimum to the peak of the barrier. An applied force is therefore predicted to lead to an exponential amplification of the dissociation kinetics: $k_{off}(F) = k_0 \exp(Fd/k_BT)$ (30). As shown in Fig. 2.14, our data fits this expression reasonably well for forces ranging from 30 to 60 pN, yielding apparent values of $d = 0.15$ nm $k_0 = 6 \times 10^{-2}$ s^{-1}. This value of k_0 is considerably higher than the value of ~2 x 10^{-6} s^{-1} reported for biotin-streptavidin (31), a result compatible with our observations. Its incompatibility with low force observations (~5-10 pN), however, where tethers hold for >5 min, suggests the model's limitations: although a simple one barrier model may describe the data at high forces, the binding energy landscape may be more complex in the low force regime (32).

As was shown in Figs. 2.7 through 2.11, the DIG-αDIG bond ruptures at forces ~ 30 to 65 pN (the overstretch transition) when the DNA is pulled at ~ 75 nm/s; occasionally the tether would last into and past overstretch. The speed at which the molecule is pulled, however, plays an integral part in determining the breaking force of a tether. A naïve way to understand this is to consider that faster pulling gets a molecule to higher forces very quickly, so that the molecule spends less time at each force. The shorter times means there is less likelihood that the "fatal" thermal fluctuation would occur to disrupt the bond; in the opposite extreme, a statically held

(i.e., the limit of slow pulling) tether will eventually break even at very low force due to such a large fluctuation. The measurements above were essentially equilibrium measurements, once the prescribed force (30, 40, 50 or 60 pN) was reached.

Alternatively, one could find the force free dissociation rate and barrier distance *d* by using dynamic force spectroscopy (DFS). Used to explore a growing number of biomolecular problems including ligand-receptor bonds, protein unfolding, and nucleosome unraveling, DFS relates the most probable unbinding force to the natural log of the force loading rate (32,33). Implementing a constant force rate is difficult in that it requires the control program to use the derivative of a measured variable (force) as a control parameter. Very fast feedback and PID control (partial integration and differentiation) would make this possible, but such a scheme is not viable in the comparatively slow LabVIEW. However, the nearly linear force-extension profile in the typical force regime of the DIG-αDIG bond disruption may mean that a constant separation velocity, which is easy to implement, may suffice. One could measure both the disruption force and the force rate at disruption and relate the latter to the former in the framework of DFS.

Preliminary experiments with a factor of two between separation speeds yielded mixed results, either because the rate difference was not sufficient to effect a different force distribution or the drift in the pipette tweezers washed out the force difference. Consequently, an intriguing future direction of this project one the dramatically more precise and stable dual beam tweezers would be to measure the rupture forces at several different speeds, apply DFS, and compare the results to the results above (or to comparable equilibrium measurements on the dual beam trap).

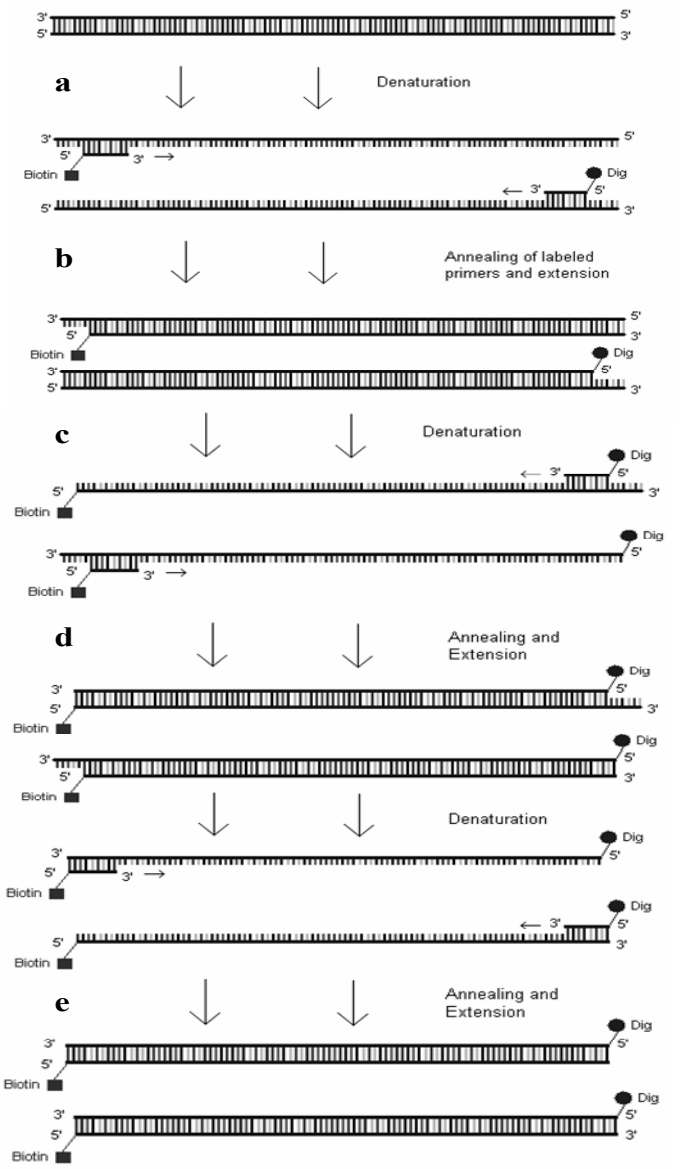

Fig. 2.1. Schematic of the PCR labeling scheme for DNA used in the optical tweezers. The first step (a) is the heat denaturation of the DNA (i.e., strand separation). Primers 20 to 30 bp in length and labeled with either a biotin or DIG were selected for high specificity. Upon the primers annealing to the separated strands (b), Taq polymerase catalyzes the fill-in of the single strands, leaving overhangs. A second denaturation (c), primer annealing, and fill-in (d) leaves a labeled end but with an overhang on the other end. A third cycle (e) leaves the desired product with a label on each end. From this point on, the desired product grows as ~2n, whereas products with overhangs (not necessarily unusable) grow as ~n. Typical PCR reactions involve 20 – 30 cycles.

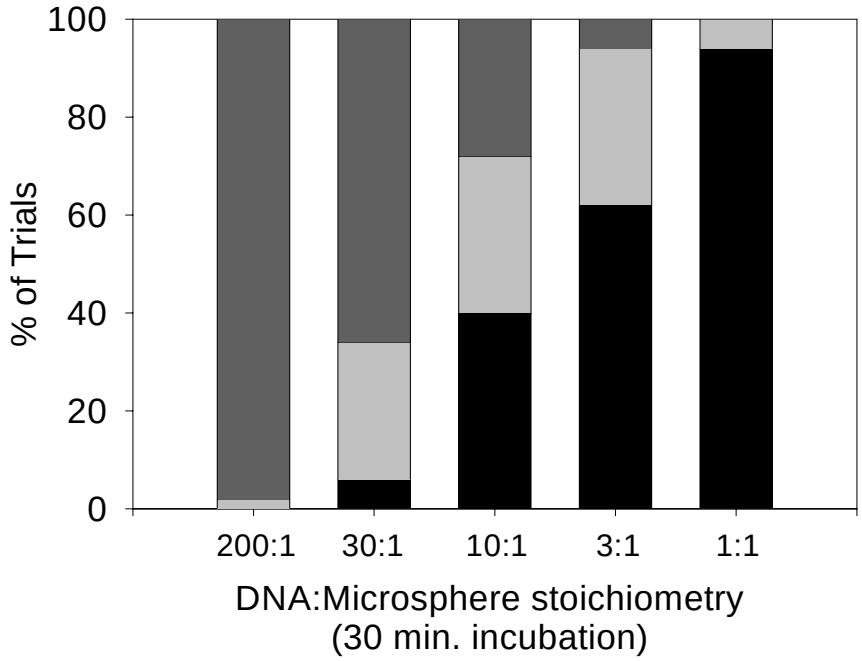

Fig. 2.2. Efficiency of DNA tethering following 30 min. incubation with streptavidin microspheres versus stoichiometry. Black, light gray, and dark gray bars indicate percentages that yielded zero, one, or multiple DNA tethers, respectively. These experiments were carried out using the 10 kbp λ fragment. Fifty different streptavidin beads were tested at each stoichiometry, with the αDIG bead being replaced every three to five streptavadin beads.

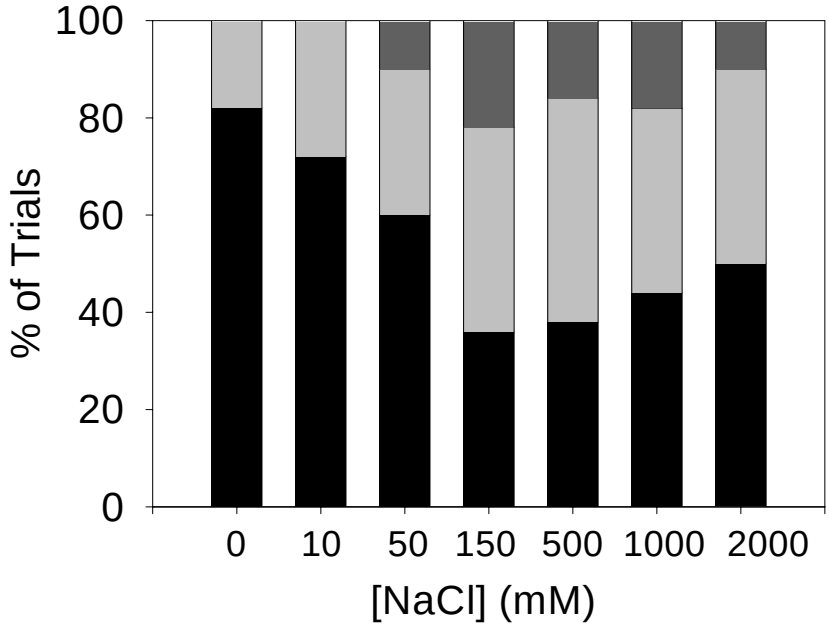

Fig. 2.3. Efficiency of DNA tethering in the microfluidic chamber versus NaCl concentration. The buffering agent was TRIS·HCl (pH = 7.8). The black, light gray, and dark gray bars indicate percentages that yielded zero, one, or multiple DNA tethers, respectively. Fifty different streptavadin beads were tested at each salt concentration, with the αDIG bead being replaced every three to five trials.

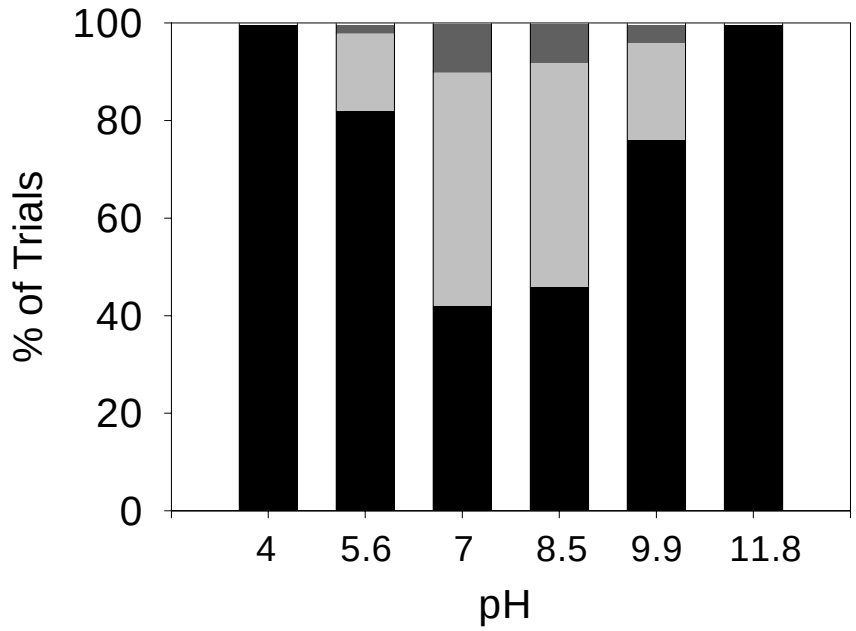

Fig. 2.4. Efficiency of DNA tethering in the microfluidic chamber versus pH. Black, light gray, and dark gray bars indicate percentages that yielded zero, one, or multiple DNA tethers, respectively. As discussed in the text, different buffering agents had to be used in each case, but the overall ionic concentration (150 mM) was kept constant. Fifty different streptavadin beads were tested at each pH, with the aDIG bead being replaced every three to five trials.

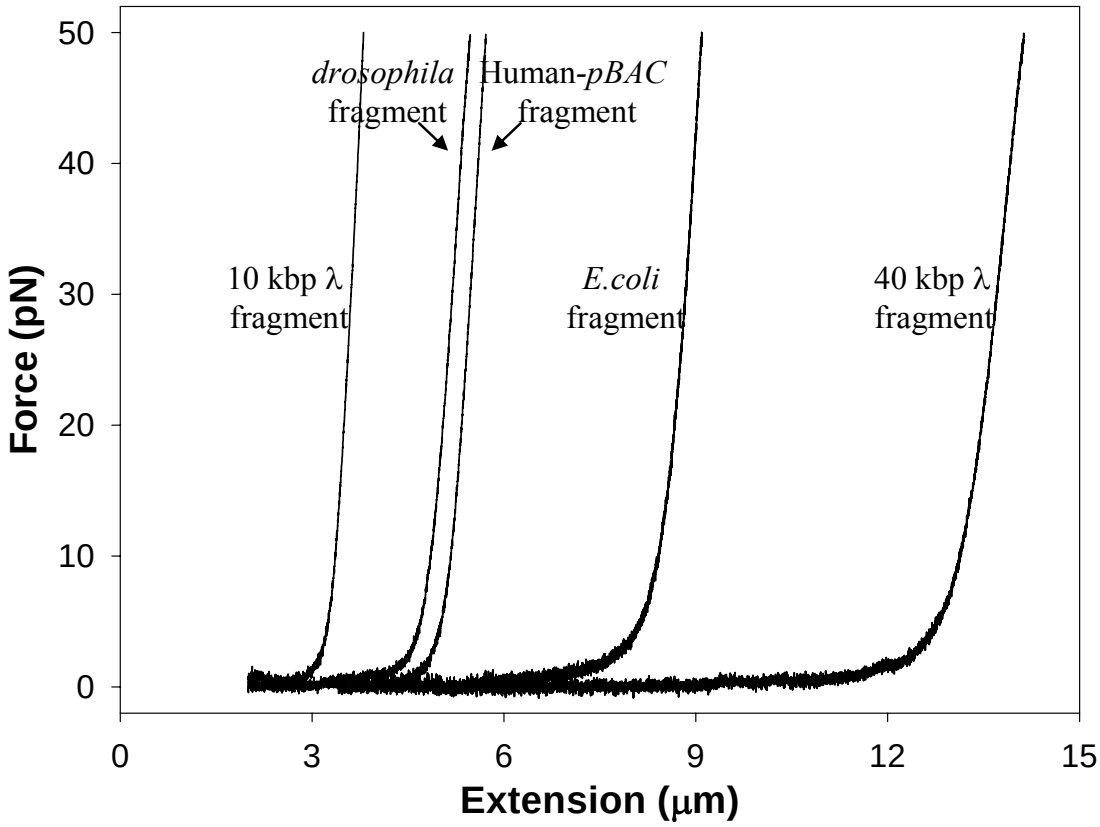

Fig. 2.5. Example pulls for the five templates generated and labeled via the PCR methodology. From left to right, the templates are the 10 kbp λ fragment (10,051 bp; 3.42 mm), *drosophila* (14,001 bp; 4.76 mm), human-pBAC (15,138 bp; 5.15 mm), *E.coli* (25,340 bp; 8.62 mm), and the 40 kbp λ fragment (40,368 bp; 13.73 mm), as labeled on the figure.

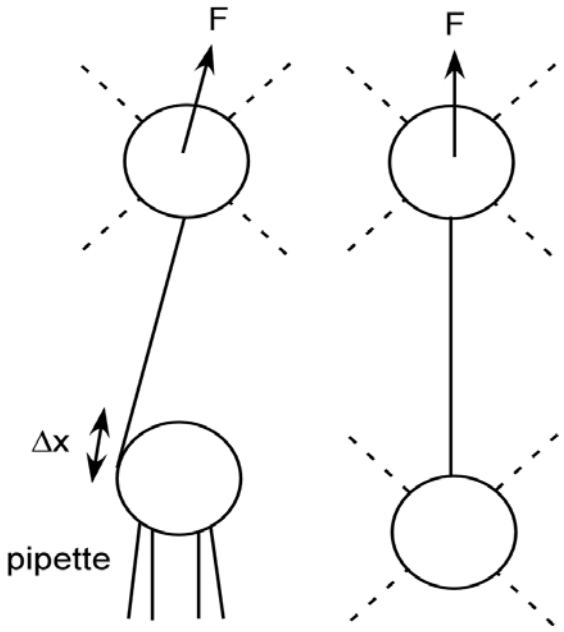

Fig. 2.6. Schematic drawing of a tether stretched in the single beam optical tweezers versus the dual optical tweezers, indicating the uncertainty in extension measurement introduced by using the pipette (where the bead is not free to rotate). The dashed line across the pipette bead indicates the range of equivalent positions having uncertainty Δx where a DNA may be tethered.

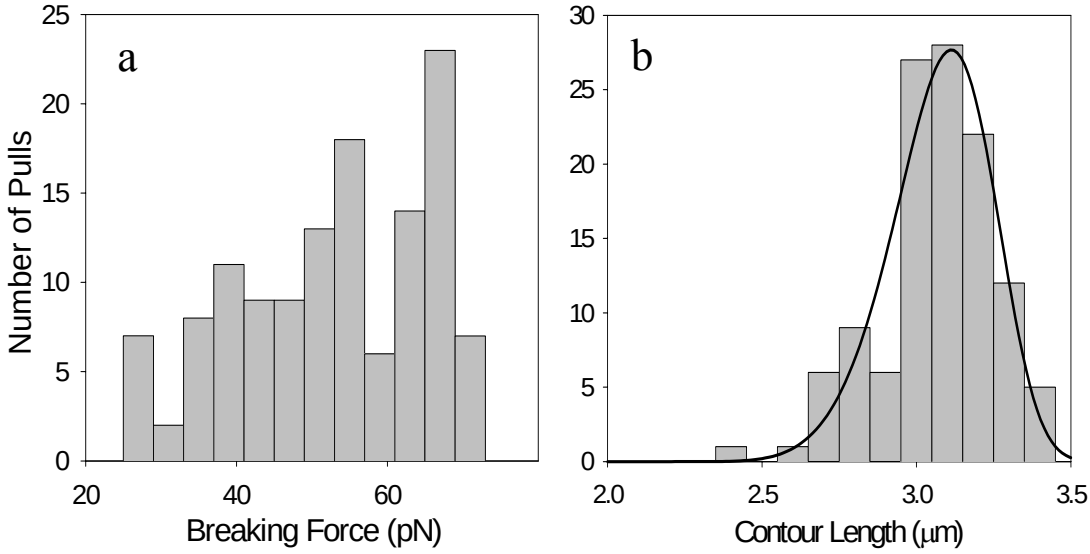

Fig. 2.7. The distribution of breaking forces (a) and measured contour lengths (b) for the 10 kbp λ fragment. As explained in the text, each molecule was pulled at a constant speed (~ 75 nm/s) until the tethers broke. The working contour length of each is determined by extrapolating the nearly linear, high force portion of the curve ($f > 15$ pN; see Fig. 2.5) across the $f = 0$ axis. Accordingly, only those tethers that survived past $f = 35$ pN provided an adequate range to fit linearly and were included in the distribution of contour lengths. The average breaking force is 53.8 pN with a $\sigma_f = 18.2$ pN calculated from $N_f = 134$ events. Seven events beyond 80 pN were observed but not shown. The contour length distribution is calculated from N = 124 events with a Weibull fit peak of 3.11 μm and a full width half-max of 0.39 μm. The nominal contour length is 3.42 μm.

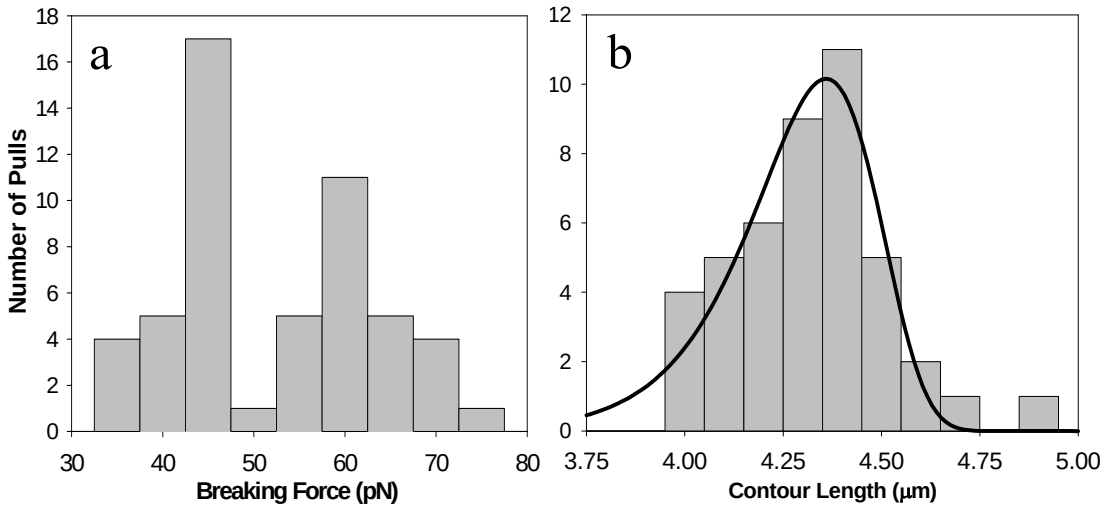

Fig. 2.8. The distribution of breaking forces (a) and measured contour lengths (b) for the *drosophila* fragment. Again, only those tethers that survived past f = 35 pN were included in the distribution of contour lengths. The average breaking force is 49.6 pN with a σ_f = 10.6 pN calculated from N_f = 53 events. Two tethers did not break before the streptavidin bead was pulled completely out of the optical trap (not included in force statistics). The contour length distribution is calculated from N = 55 events with a Weibull fit peak of 4.36 μm and a full width half-max of 0.39 μm. The nominal contour length is 4.76 μm.

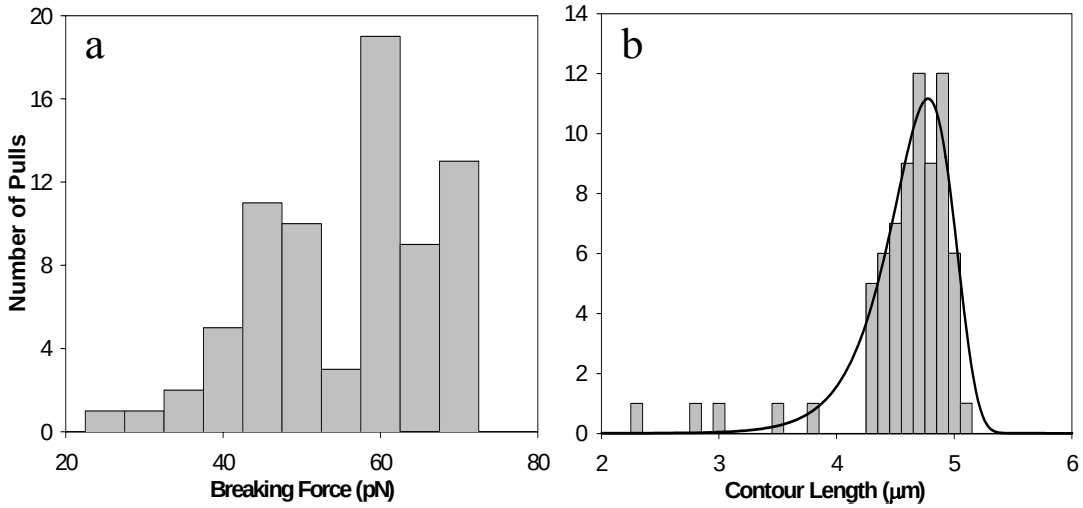

Fig. 2.9. The distribution of breaking forces (a) and measured contour lengths (b) for the human-pBAC fragment. Again, only those tethers that survived past f = 35 pN were included in the distribution of contour lengths. The average breaking force is 53.6 pN with a σ_f = 12.5 pN calculated from N_f = 75 events. One tether did not break before the streptavidin bead was pulled completely out of the optical trap (not included in force statistics), while another was above 80 pN. The contour length distribution is calculated from N = 72 events with a Weibull fit peak of 4.77 μm and a full width half-max of 0.66 μm. The nominal contour length is 5.15 μm.

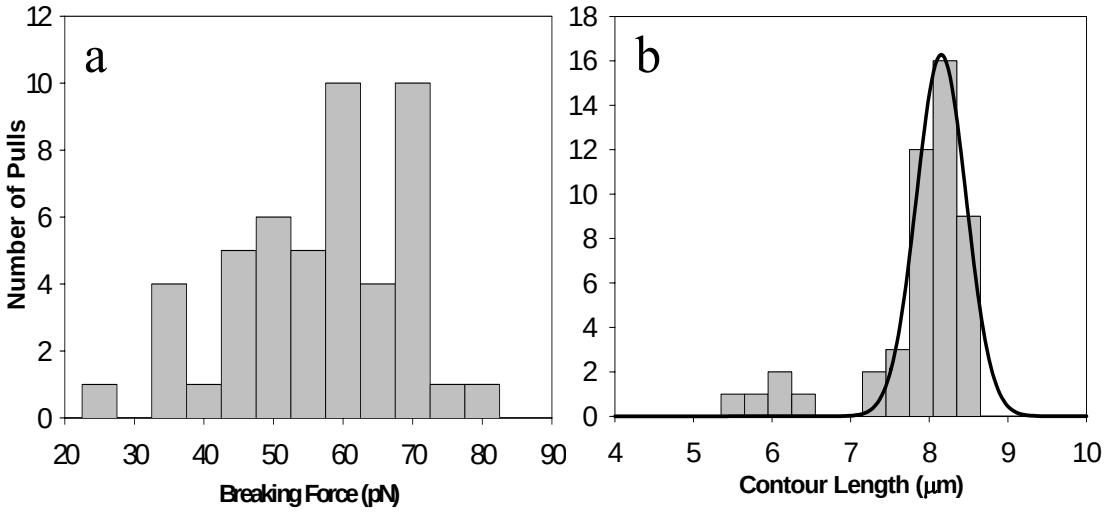

Fig. 2.10. The distribution of breaking forces (a) and measured contour lengths (b) for the *E. coli* fragment. Again, only those tethers that survived past f = 35 pN were included in the distribution of contour lengths. The average breaking force is 53.9 pN with a σ_f = 12.3 pN calculated from N_f = 48 events. Four tethers did not break before the streptavidin bead was pulled completely out of the optical trap (not included in force statistics). The contour length distribution is calculated from N = 47 events with a Gaussian fit peak of 8.15 μm and a width of 0.32 μm. The nominal contour length is 8.62 μm. The five events around ~ 6 μm are believed to be due to a secondary, shorter band detected in the gel in this sample, as mentioned in the text.

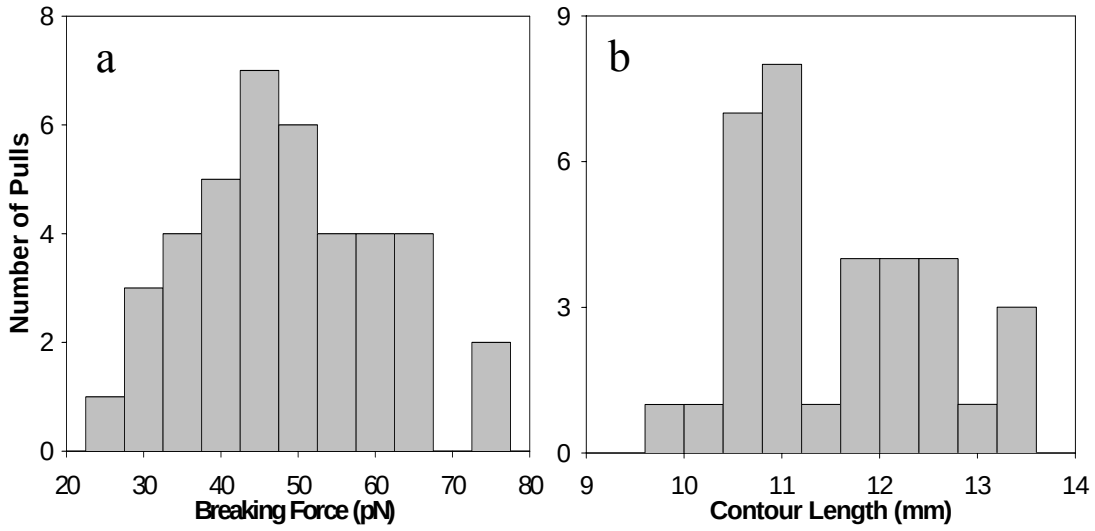

Fig. 2.11. The distribution of breaking forces (a) and measured contour lengths (b) for the 40 kbp λ fragment. Again, only those tethers that survived past f = 35 pN were included in the distribution of contour lengths. The average breaking force is 47.3 pN with a σ_f = 15.5 pN calculated from N_f = 41 events. Two tethers did not break before the streptavidin bead was pulled completely out of the optical trap (not included in force statistics), while another was above 80 pN. The contour length distribution is calculated from N = 35 events with an average of 11.35 μm and a standard deviation of 0.94 μm. One event was below 9 μm. The nominal contour length is 13.73 μm.

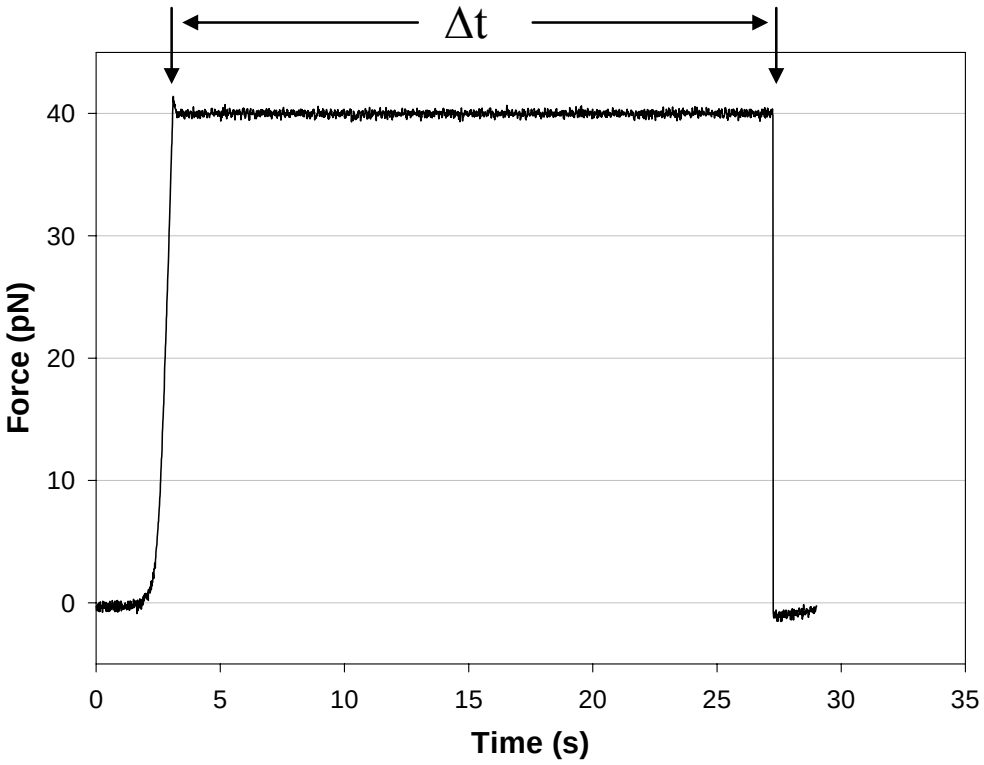

Fig. 2.12. Example data set for the unbinding time experiments. Once a tether is made (before t = 0), the force is quickly ramped up to the preset level, in this case 40 pN. Upon reaching the preset level (left arrow), the force is maintained at the limit by a 100 Hz force clamp. When the tether breaks (right arrow), the measured force drops to zero and Δt is measured. Not all trials made it to the preset force, as discussed in the text.

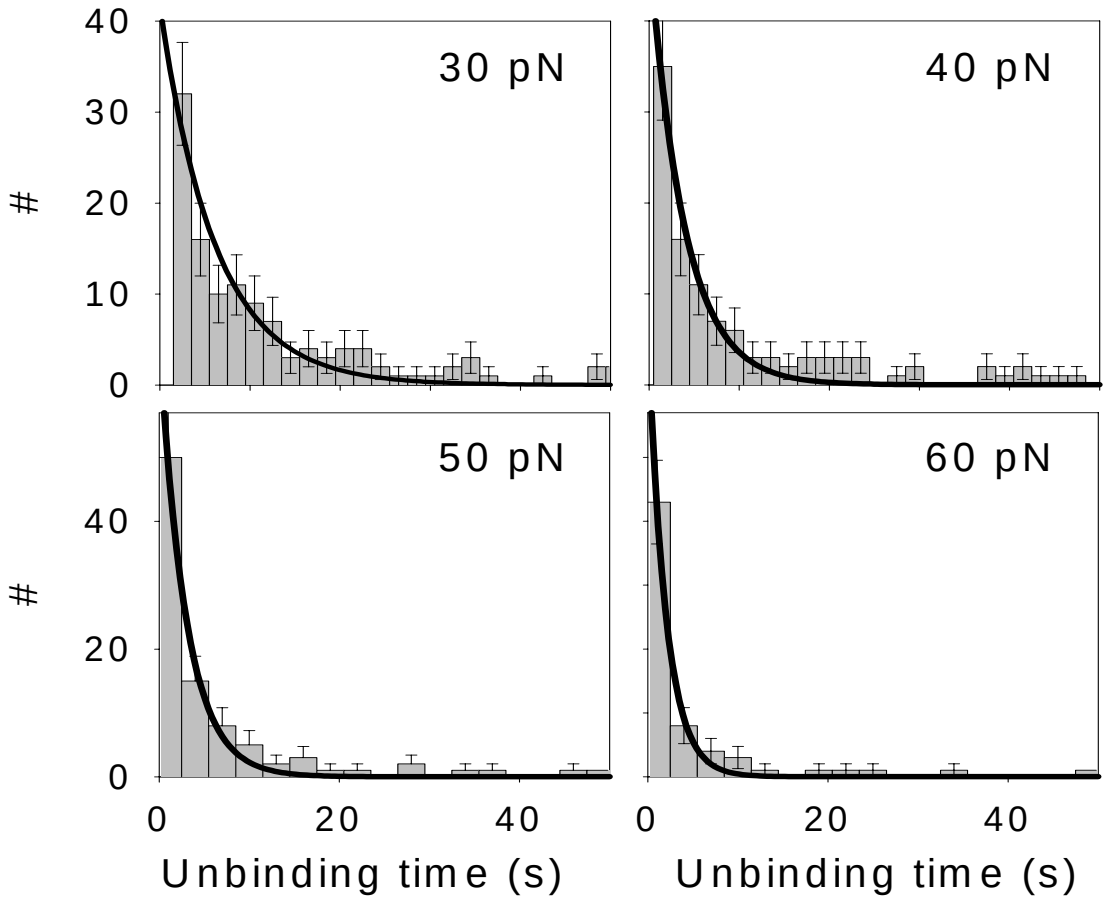

Fig. 2.13. Histograms of unbinding times of a tethered DNA held at 30, 40, 50, and 60 pN. Each data set was fit by a decaying exponential of the form $P(t) \sim \exp(-k_{off} t)$, where P(t) is the probability density and k_{off} is a dissociation rate. The number of trails were N_{trials} = 309, 263, 319, and 170 for F = 30, 40, 50, and 60 pN, respectively.

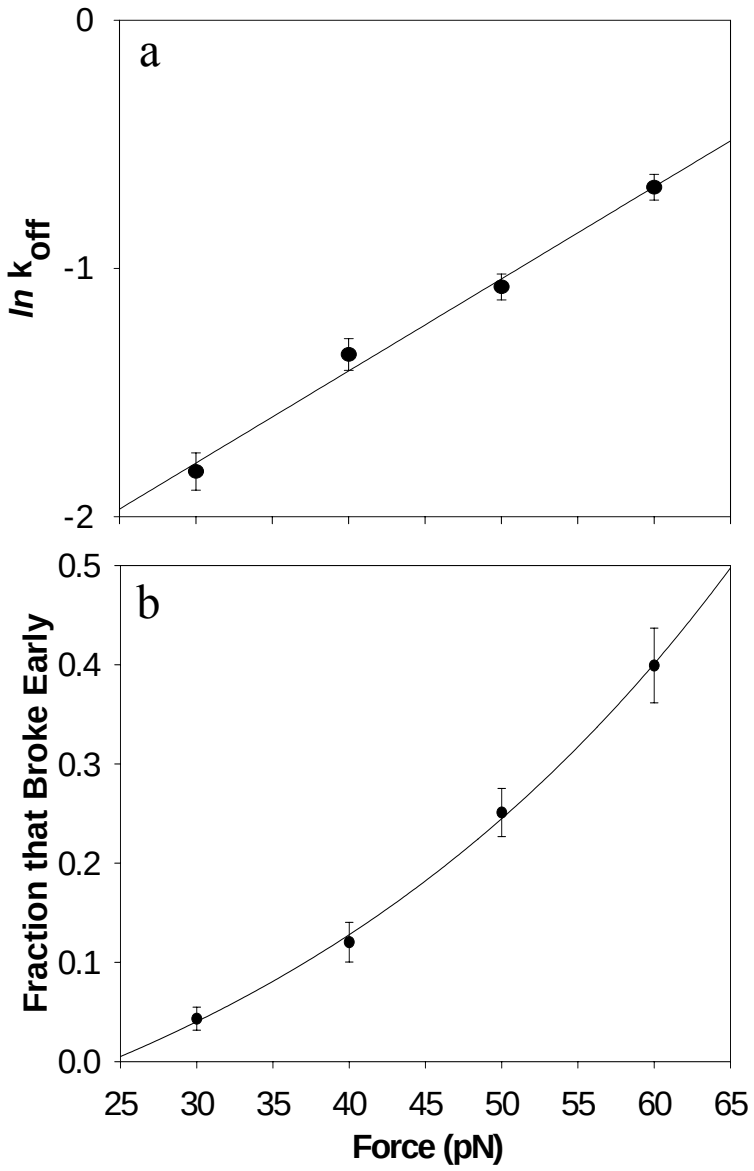

Fig. 2.14. (a) Semilog plot of the dissociation rate *k*off versus force. The line is a linear fit as suggested by a simple unbinding model discussed in the text. (b). The fraction of trials that did not reach the force level as a function of force. The uncertainties were determined by the square root of the number of "failures" divided by the total number of trials ($N_{failures}$ = 14, 36, 107, 113 and N_{total} = 323, 299, 426, 283 for F = 30, 40, 50, 60 pN, respectively). The line is a visual aid and does not imply a specific functional form.

References for Chapter 2

1. Chu, S. (1991) Laser trapping of neutral particles. *Science*. **253**: 861.

2. Perkins, T.T., Quake, S.R., Smith, D.E. & Chu, S. (1994) Relaxation of a single DNA molecule observed by optical microscopy. *Science*. **264**: 822-826.

3. Perkins, T.T., Smith, D.E. & Chu, S. (1994) Direct observation of tube-like motion of a single polymer chain. *Science*. **264**: 819-822.

4. Smith, S.B., Cui, Y. & Bustamante, C. (1996) Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*. **271**: 795-799.

5. Yin, H., Wang, M.D., Svoboda, K., Landick, R., Block, S.M. & Gelles, J. (1995) Transcription against an applied force. *Science.* **270**: 1653-1657.

6. Wang, M.D., Schnitzer, M.J., Yin, H., Landick, R., Gelles, J. & Block, S.M. (1998) Force and velocity measured for single molecules of RNA polymerase. *Science* **282**: 902-907.

7. Wuite, G.J., Smith, S.B., Young, M., Keller, D. & Bustamante, C. (2000) Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature*. **404**: 103-106.

8. Cui, Y. & Bustamante, C. (2000) Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. *Proc Natl Acad Sci USA*. **97**: 127-132.

9. Brower-Toland, B., Smith, C., Yeh, R., Lis, J., Peterson, C.L. & Wang, M.D. (2002) Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proc Natl Acad Sci USA*. **99**: 1960-1965.

10. Smith, D.E., Tans, S.J., Smith, S.B., Grimes, S., Anderson, D.L. & Bustamante, C. (2001) The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature*. **413**: 748-752.

11. Perkins, T.T., Li, H.W., Dalal, R.V., Gelles, J. & Block, S.M. (2004) Forward and reverse motion of single RecBCD molecules on DNA. *Biophys. J.* **86**: 1640-1648.

12. Fuller, D.N, Gemmen, G.J., Rickgauer, J.P., Dupont, A., Millin, R., Recouvreux, P., & Smith, D.E. (2006) A General Method for Manipulating DNA Sequences from any Organism with Optical Tweezers. *Nuc. Acids Res*. **34**: e15.

13. Fyodorov, D.V. & Levenstein, M.L. (2002), *Current Protocols in Molecular Biology*. John Wiley & Sons, New York, pp. 21.27.21-21.27.

14. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. & Haussler, D. (2002) The human genome browser at UCSC. *Genome Res*. **12**: 996-1006.

15. Simmons, R.M., Finer, J.T., Chu, S. & Spudich, J.A. (1996) Quantitative measurements of force and displacement using an optical trap. *Biophys. J*. **70**: 1813-1822.

16. Beynon, R.J. & Easterby, J.S. (1996) *Buffer Solutions: The basics*. BIOS Scientific Publishers, Oxford.

17. Janeway, C.A., Travers, P., Walport, M. & Shlomchik, M.J. (2001) *Immunobiology*. 5th ed. Garland Publishing, New York.

18. Bunville, L.G. & Geiduschek, E.P. (1965) Kinetics and Equilibria in the Acid Denaturation of Deoxyribonucleic Acids from Various Sources. *Biopolymers.* **3**: 213-240.

19. Allemand, J.F., Bensimon, D., Jullien, L., Bensimon, A., & Croquette, V. (1997) pH-Dependent Specific Binding and Combing of DNA. *Biophys. J.* **73**: 2064-2070.

20. Marko, J.F. & Siggia, E.D. (1995) Stretching DNA. *Macromolecules*. **28**: 8759-8770.

21. Bustamante, C., Marko, J.F., Siggia, E.D. & Smith, S. (1994) Entropic elasticity of lambda-phage DNA. *Science.* **265**: 1599-1600.

22. Wang, M.D., Yin, H., Landick, R., Gelles, J. & Block, S.M. (1997) Stretching DNA with optical tweezers. *Biophys. J*. **72**: 1335-1346.

23. Bouchiat, C., Wang, M.D., Allemand, J., Strick, T., Block, S.M. & Croquette, V. (1999) Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.* **76**: 409-413.

24. Sakata-Sogawa, K., Kurachi, M., Sogawa, K., Fujii-Kuriyama, Y. & Tashiro, H. (1998) Direct measurement of DNA molecular length in solution using optical tweezers: detection of looping due to binding protein interactions. *Eur. Biophys. J.* **27**: 55-61.

25. Hogan, M., LeGrange, J. & Austin, B. (1983) Dependence of DNA helix flexibility on base composition. *Nature*. **304**: 752-754.

26. Hogan, M.E. & Austin, R.H. (1987) Importance of DNA stiffness in protein-DNA binding specificity. *Nature*. **329**: 263-266.

27. Koo, H.S., Wu, H.M. & Crothers, D.M. (1986) DNA bending at adenine-thymine tracts. *Nature*. **320**: 501-506.

28. Widom, J. (2001) Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys*. **34:** 269-324.

29. Odijk, T. (1995) Stiff chains and filaments under tension. *Macromolecules*. **28**: 7016-7018.

30. Bell, G.I. (1978) Models for the specific adhesion of cells to cells. *Science*. **200**: 618-627.

31. Piran, U. & Riordan, W.J. (1990) Dissociation rate constant of the biotin-streptavidin complex. *J. Immunol. Methods.* **133**: 141-143.

32. Evans, E. (2001) Probing the relation between force--lifetime--and chemistry in single molecular bonds. *Annu. Rev. Biophys. Biomol. Struct.* **30**: 105-128.

Chapter 2 Acknowledgement

The text of Chapter 2, in part, is a reprint of the material as it appears in Fuller, D.N, Gemmen, G.J., Rickgauer, J.P., Dupont, A., Millin, R., Recouvreux, P., & Smith, D.E. (2006). A General Method for Manipulating DNA Sequences from any Organism with Optical Tweezers. *Nuc*. *Acids Res*. **34**: e15. The dissertation author was a primary researcher and secondary author for the research which forms the basis of this chapter.

Chapter 3. Tension-Dependent DNA Cleavage by Restriction Endonucleases: Enzymes Binding at Two Sites are "Switched Off" at Low Force

3.1. Abstract

DNA looping occurs in many important enzyme-DNA interactions, including gene transcription. Recent theoretical studies predict that a few piconewtons applied to DNA would almost completely inhibit looping. Here we study restriction endonucleases that require interaction at two separated sites for efficient cleavage. Using the tweezers we measured the dependence of cleavage activity on tension with fifteen known or suspected two site enzymes (BfiI, BpmI, BsgI, BspMI, Cfr9I, Cfr10I, Eco57I, EcoRII, FokI, HpaII, MboII, NarI, SacII, Sau3AI, and SgrAI) and six one site enzymes (BamHI, EcoRI, EcoRV, HaeIII, HindIII, and DNaseI). The one site enzymes were largely unaffected by 5 pN, whereas all of the two site enzymes were completely inhibited. At higher forces, cleavage was partially inhibited for EcoRV and HaeIII, enhanced for HindIII, and largely unaffected for BamHI, EcoRI, and DNaseI. These findings correlate with structural data showing that EcoRV bends the DNA while BamHI, EcoRI, and DNaseI do not. Thus, enzyme activity involving either DNA looping or bending can be tension modulated. Such a mechanism could facilitate mechanosensory transduction *in vivo*, and such DNA looping enzymes may thus constitute an example of a tension sensing "molecular switch".

3.2. Introduction

Restriction endonucleases (REases) are prokaryotic enzymes that act to “restrict” invasion of foreign DNA by cleaving it at specific recognition sequences (1). Used in

procedures such as DNA cloning, fingerprinting, mapping, and sequencing, these enzymes serve as indispensable tools in molecular biology research (2). From the perspective of molecular biophysics, these enzymes are also excellent model systems for studying basic principles of specific protein-DNA interactions (3).

The most commonly studied REases are Type II, which cleave within or near single recognition sites, usually require Mg^{2+}, and do not hydrolize ATP. More than 3500 different Type II REases having over 200 different binding specificities have been identified (4). Of interest in our present study are the many unorthodox Type II REases that do not cleave DNA efficiently if the template contains only one recognition site (5,6). Efficient cleavage is only observed with templates containing two or more sites, indicating that that the active complex binds at two sites and the intervening DNA is looped (6). This phenomenon of DNA looping is of importance in molecular biology and plays a role in many key processes including DNA transcription, replication, recombination, and repair (7-14). Looping by *lac* and *gal* repressor proteins in *E.coli*, for example, is well known and has recently been studied *in vitro* via tethered particle assays (15-17).

Several recent theoretical studies have considered the effect of DNA tension on looping (18-20). As discussed by Marko and others, the probability of forming a loop of size ΔL via thermal fluctuations against an applied tension F is expected to be suppressed by a factor proportional to $exp(-F\Delta L/kT)$. Detailed calculations based on the worm-like chain (WLC) model predict that a few piconewtons of tension would, in most cases, decrease the probability of looping by at least several orders of magnitude over the tension-free probability. An independent theoretical study by

Blumberg, *et al.,* reached the same conclusion (20). Free energy differences between looped and unlooped DNA were calculated in a simple two-state model. For loop sizes >100 bp, a few piconewtons of tension is predicted to increase the time required for loop closure by at least two orders of magnitude and the degree of inhibition is predicted to increase rapidly with loop size.

Here we study the effect of tension on DNA cleavage by one site and two site REases by manipulation of single DNA molecules with force-measuring optical tweezers. Partial inhibition of one site cleavage by EcoRV at very high DNA tension was recently reported in an independent study (21). Here we confirm and extend these findings on one site enzymes and report a much stronger and universal effect of tension on activity with two site REases. In agreement with the recent theoretical predictions, we find that these enzymes constitute an example of a tension sensing "molecular switch."

3.3. Methods

3.3.1. DNA Constructs

pBAC-A was prepared by ligating a 4282 bp digoxygenin (DIG)-labeled PCR fragment to a 10845 bp biotin-labeled restriction fragment of pBACe3.6. The PCR fragment was generated by amplification of a sequence from pFastBac HT-b (Invitrogen) using the primers 5'-GTGGTATGGCTGATTATGATC and 5'-GCAGCCTGAATGGCGAATGG-3' and was labeled by incorporation of 20 μM of dUTP-11-DIG (Roche) and 200 μM each of dATP, dCTP, dGTP, dTTP in the PCR. The 10845 bp fragment was produced by digesting pBACe3.6 (Children's Hospital of

Oakland Research Institute) with BsrGI (New England Biolabs, "NEB") and end labeling by using the Klenow fragment of *E. coli* DNA polymerase I, exo- (NEB) to incorporate dATP-14-biotin (Invitrogen). Both fragments were purified using the Qiagen PCR purification kit and digested with XhoI (NEB). To isolate the desired product the samples were run on a 1% agarose gel in 1X TAE buffer and purified using the Qiagen Gel Extraction kit. The two fragments were then ligated by use of T4 DNA ligase (NEB).

½-λ-L was prepared by using the Klenow fragment of DNA polymerase I, exo- to fill in the ends of methyladenine-free DNA (NEB) with biotin-dATP and dCTP (Invitrogen). The DNA was then XbaI-digested and purified with a DNA clean up kit. A second fill-in was then done with digoxygenin-labeled dUTP (Roche), resulting in two fragments of ~8.1 μm. The sample was then digested with XhoI to to prevent tethering of the unwanted right end. The 15 kbp human DNA sequence, 14 kbp *Drosophila* sequence, and 10 kbp bacteriophage λ sequence were generated by PCR amplification with genomic DNA using biotin and digoxygen labeled primers, as described previously (22).

3.3.2. DNA Tethering

Streptavidin coated beads (200 μl of 0.5% w/v, 2.2 μm diameter, Spherotech) were washed by centrifuging at 10,000 ×g and resuspended twice in 200 μl of phosphate buffered saline (PBS) pH 7.4 (Fisher Scientific) and 0.1 mg/ml Bovine serum albumin (BSA) (New England Biolabs). 5 μl of diluted DNA (~10-100 ng/μl) was mixed with 5 μl of beads and incubated for ~45 min. at room temperature on a

slowly rotating mixer. 5-10 μl of these beads were diluted in 0.5 ml of PBS and loaded into a 1 ml tuberculin syringe for injection into the sample chamber. Protein G beads (200 μl of 0.5% w/v, 2.8 μm diameter, Spherotech) were likewise washed and resuspended in 20 μl PBS. Then, 5 ul 200 μg/ml of antidigoxygenin (Roche) was added. The beads were incubated on the mixer for ~ 1 hr, washed twice, and resuspended in 20 μL PBS. 5 μl of the beads were loaded into a syringe for injection into the flow cell. The αDIG bead was held by a micropipette while the DNA bead was trapped with the optical tweezers. The two beads were brought into proximity such that the DIG-labeled end of one DNA molecule bound to the αDIG coated bead, forming a DNA tether between the beads (23).

3.3.3. Enzymes

We examined as many two site REases as was feasible given our DNA templates. Evidence in the literature for two site behavior of REases comes from variety of studies. In a few cases looped complexes have been directly imaged by electron microscopy (NaeI, Cfr10I, EcoRII, and Sau3AI) (24-27), while DNA recombination assays and gel mobility shift assays have been used to infer DNA looping with SfiI and Cfr10I (28-30). Fluorescence resonance energy transfer (FRET) has been used to detect conformational changes consistent with DNA loop formation by NgoMIV (31), and looping by BspMI was detected by magnetic tweezers (32). Further evidence comes from the comparison of DNA cleavage rates on templates containing two sites versus only one site, which has been reported for SgrAI, BsgI, BpmI, MboII, and NarI (33-35). Finally, stimulation of activity upon addition of

short oligo-nucleotide duplexes containing the recognition sequence has been reported for Eco57I, FokI, HpaII, Ksp632I, Cfr9I, and SacII. Such stimulation provides evidence that an enzyme complex can form *in trans* (36-39).

Based on the above considerations, we purchased BamHI, BpmI, BsgI, BspMI, EcoRI, EcoRV, FokI, HaeIII, HindIII, HpaII, MboII, NarI, SacII, Sau3AI, SfiI, and SgrAI from New England Biolabs. BfiI, Cfr9I, Cfr10I, and Eco57I were purchased from Fermentas, EcoRII and Ksp632I were purchased from Roche, and bovine pancreatic DNase I was purchased from Calbiochem. Six of these (BamHI, EcoRI, EcoRV, HaeIII, HindII, and DNaseI) are common one site endonucleases, which were chosen for comparison with the two site enzymes. With each enzyme we used the reaction buffer recommended by the manufacturer. The DNaseI buffer contained 50 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$ and 50 μg/ml BSA. All measurements in the tweezers were made at room temperature (~20 °C). Relevant data on the enzymes are given in Table 3.1.

3.4. Results and Discussion

3.4.1. Tension-Dependent Inhibition of Two Site Enzymes

Our experiment for measuring DNA cleavage activity was carried out as shown schematically in Fig. 3.1. We held the DNA molecule taut at an end-to-end extension of ~95% (~5 pN) while the enzyme solution was flowed into the sample chamber. The flow was continued until the entire chamber was uniformly filled with enzyme solution, and then stopped. Measurements were made with the DNA either held stretched at 5 pN or relaxed to an end-to-end extension of 35%, corresponding to

a ~~t~~ension of ~0.06 pN. When the DNA was held taut cleavage events were detected as a sudden drop in the measured tension, rendering the event observable in real time. The signature would closely resemble Fig. 2.12. With the DNA relaxed, we tested for an intact DNA tether every 30 seconds by quickly separating the beads. If the molecule had been cleaved, the measured force remained zero as the beads were separated. If the molecule remained intact, however, a tension of up to a few piconewtons was sensed and the molecule was then quickly relaxed for a subsequent incubation period. This procedure was continued until either the molecule was cleaved or until 5 minutes had elapsed. This process was repeated many times for both one site and two site endonucleases.

The cleavage statistics are shown in Fig. 3.2. One site enzymes were largely unaffected by an applied tension of 5 pN, whereas all of the known or suspected two site enzymes were strongly inhibited. The two site enzymes differ from the one site enzymes in that they must loop the DNA, so this inhibition is consistent with the theoretical predictions for tension dependent looping (18-20).

Three of the fifteen enzymes studied, EcoRII, NarI, and SgrAI, did not show complete cleavage even with the DNA relaxed; note that there was still a clear drop in activity between the relaxed and stretched trials. For SgrAI and NarI, the relative lack of activity may be attributed to the DNA template having a relative paucity of recognition sites (only four and five sites, respectively, with ~~a~~ minimum separations of 1321 bp and 428 bp). On the other hand, incomplete cleavage was also observed with EcoRII, for which the DNA template contained a similar number of sites and a

similar distribution of separations as those of the enzymes for which complete cleavage was observed (40).

Control experiments with the two site enzymes were done using templates with zero or one site. Cfr9I, NarI, SacII, SfiI, and SgrAI were tested on a template with zero sites and showed no activity. BfiI was tested on a template with only one site and also showed no activity. SacII cleaved a template with one copy of its recognition sequence very infrequently. Cleavage by two site enzymes on templates containing only one site has been previously reported in bulk studies with Acc36I, BpmI, BsgI, BspMI, and FokI. In accord with our finding with SacII, it was found that these enzymes cleave more than an order of magnitude more slowly than they do on a template containing two sites (33).

We note that a small fraction of tethers appeared to be cleaved with two site enzymes even when the DNA was held under tension. This finding, however, can be completely attributed to spontaneous unbinding of the DNA from the α-DIG bead rather than actual DNA cleavage by the enzyme. As shown in Fig. 3.2, roughly the same background level of breakage was observed in control experiments in which the reaction buffer without enzyme was flowed into the sample chamber. Thus our results are consistent with complete inhibition of cleavage by two site REases at 5 pN of tension.

Several thousand Type II REases have been discovered and more discoveries are occurring rapidly. While only a small number of these are currently known or suspected to operate by a two site mechanism, it is likely that many more do as well. We note that our assay provides a way to rapidly screen enzymes for two site

behavior without needing engineered DNA templates. Indeed, for the enzymes FokI, Cfr9I, Eco57I, and HpaII there have been no previous definitive assertions that each utilized DNA looping, though they were suspected (J. Ryu, Loma Linda Univ., *private communication*).

3.4.2. Effect of Tension on One Site Enzymes

Our main purpose in studying one site REases was to contrast their behavior with that of two site enzymes. However, as some REases are known to induce sharp bending of DNA (41), we hypothesized that cleavage activity might also be inhibited in such cases. Tension would be expected to interfere with the binding of enzymes that induce DNA bending because it would increase the free energy change required for binding.

Although all of the one site enzymes were largely unaffected by 5 pN of tension, some were sometimes affected by higher tensions (Fig. 3.3). Clear inhibition of EcoRV was observed at 20 and 40 pN, whereas very little inhibition was seen with BamHI, EcoRI, or DNaseI. These findings agree with the crystal structure data which shows that EcoRV bends its cognate DNA through an angle of ~50°, whereas neither EcoRI nor BamHI bends DNA significantly (41-43). Our findings on BamHI and EcoRV are in agreement with findings on these two enzymes recently reported in an independent study (21). Here we provide additional results on four other one site enzymes. We observed a small inhibitory effect of tension with HaeIII, but only at 40 pN. No structure for a HaeIII-DNA complex has been reported, but on the basis of our data we predict that it bends DNA, though to a lesser extent than EcoRV.

For HindIII we clearly observed that the cleavage was enhanced by tension rather than inhibited. A tension of 40 pN dramatically increased the rate of cleavage, and control experiments confirmed that this effect was not merely due to increased unlinking of the DNA tether from the beads. We interpret this result as indicating that there is no protein induced DNA bending with HindIII, and that the applied stress is directly transmitted to the phosphodiester backbone of the DNA. Such stress would accelerate cleavage as it favors the formation of the final products of the reaction. Furthermore, upon close inspection of the data with BamHI, EcoRI, and DNaseI, one sees a small degree of tension induced catalysis, although it is not as dramatic as that observed with HindIII. No crystal structures for HindIII-DNA complexes have been reported, but we predict that it does not involve significant DNA bending. Our findings with DNaseI are in accord with crystal structure data which show that this enzyme binds in the minor groove of DNA and causes only minor distortions (44).

As DNase I cleaved rapidly even at high tension, and as it is a non-specific endonuclease that binds in the minor grove of DNA, we wondered what effect extreme distortion of the double helix would have on its activity. When held at a tension greater than ~65 pN, dsDNA undergoes an "overstretch" in which the molecule lengthens by 70% and the helix unwinds. Interestingly, when the DNA was held in this overstretched configuration at 75 pN, we observed partial inhibition of DNaseI. Thus, DNaseI appears to be very permissive in its interaction with DNA. Controls without DNase I showed that the observed cleavage events were real – that they were not simply the tethers giving way.

3.5. Conclusions

Cleavage of single DNA molecules by type II restriction enzymes was studied via optical tweezers. A number of different one and two site endonucleases were examined. The specific activity of two site enzymes was universally "switched off" by 5 pN of DNA tension, whereas it had virtually no effect on one site enzymes. This finding agrees with several theories that predict a strong tension dependence of DNA looping. Inhibition was also observed with certain one site enzymes, but only at much higher tensions. This inhibition was correlated with protein induced DNA bending. Our results indicate that DNA looping provides a mechanism for a tension-sensing switch. It is conceivable that this mechanism could act *in vivo* to facilitate intracellular or extracellular mechanosensory transduction.

Table 3.1. Properties of the REases studied in the cleavage experiments as reported in REBASE. Also listed are the DNA templates used, number of recognition sites, and concentration of enzyme. Type IIE REases bind at two sites, but only one is cut, whereas type IIF cleave coordinately at both binding sites. Type IIG have restriction and modification activities in the same subunit and Type IIS enzymes recognize asymmetric sequences and cleave at least one strand outside of the recognition sequence. An * denotes an unknown entry.

Enzyme	Type	MW (kD)	Form in Soln.	Active Complex	Recognition Sequence	DNA (# sites)	C (U/ml)
BfiI	IIS	40	Dimer	Dimer	ACTGGG (5/4)	Human 15kb (5)	1
BpmI	IIE, G, S	117	*	*	CTGGAG$(N)_{16}$↓	½-λ-L (17)	20
BsgI	IIE, G, S	121	*	*	GTGCAG$(N)_{16}$↓	½-λ-L (31)	30
BspMI	IIE	222	Tetra-mer	Tetra-mer	ACCTGC$(N)_4$↓	½-λ-L (24)	20
Cfr9I	IIE	37	*	*	C↓CCGGG	Human 15kb (7)	200
Cfr10I	IIF, P	320	Tetra-mer	Tetra-mer	R↓CCGGY	½-λ-L (56)	100
Eco57I	IIE, G	117	Mono-mer	*	CTGAAG$(N)_{16}$↓	½-λ-L (25)	50
EcoRII	IIE, P	92	Dimer	Dimer	↓CCWGG	½-λ-L (36)	50
FokI	IIS	66	Mono-mer	Dimer	GGATG$(N)_9$↓	pBAC-A (26)	40
HpaII	IIE	41	*	*	C↓CGG	pBAC-A (49)	100
Ksp632I	IIE	*	*	*	CTCTTC (1/4)	½-λ-R (20)	100
MboII	IIS	49	Mono-mer	Dimer	GAAGA (8/7)	pBAC-A (27)	50
NarI	IIE	*	*	*	GG↓CGCC	*Drosophila* 14kb (5)	80
SacII	IIE	*	*	*	CCGC↓GG	*Drosophila* 14kb (5)	100
Sau3AI	IIE	56	Mono-mer	Dimer	↓GATC	pBAC-A (55)	40
SgrAI	IIP	38	Dimer	Tetra-mer	C(A/G)↓CCGG(C/T)G	λ 10kb (4)	200

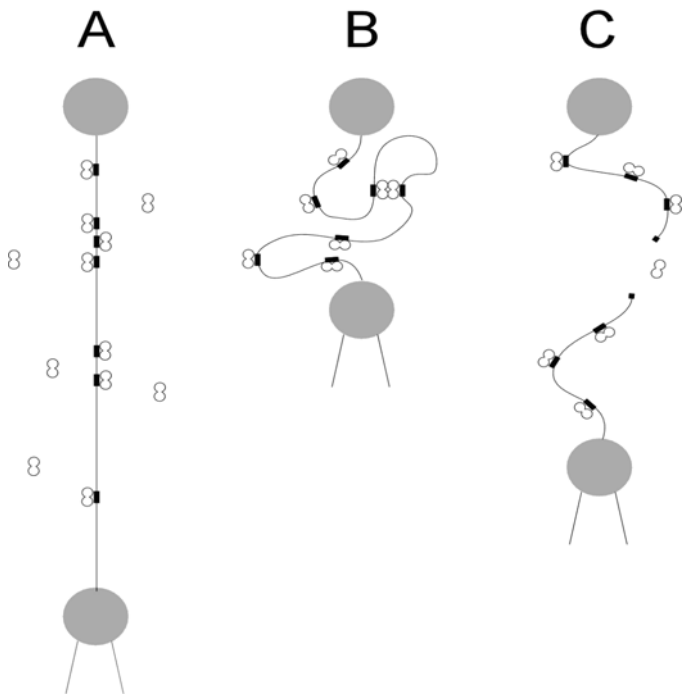

Fig. 3.1. Schematic of the single DNA cleavage measurements for two site REases. DNA is tethered between two beads, one manipulated by optical tweezers and the other manipulated by a micropipette positioned by a piezoelectric nanopositioning stage. (A) The DNA is held taut at an end-to-end extension of 95% (~5 pN) while a solution containing the enzyme is introduced. (B) The molecule is then relaxed to 35% extension (~ 0.06 pN) so that the active enzyme complex may form via DNA looping. (C) Upon separating the beads we detect whether or not cleavage has occurred.

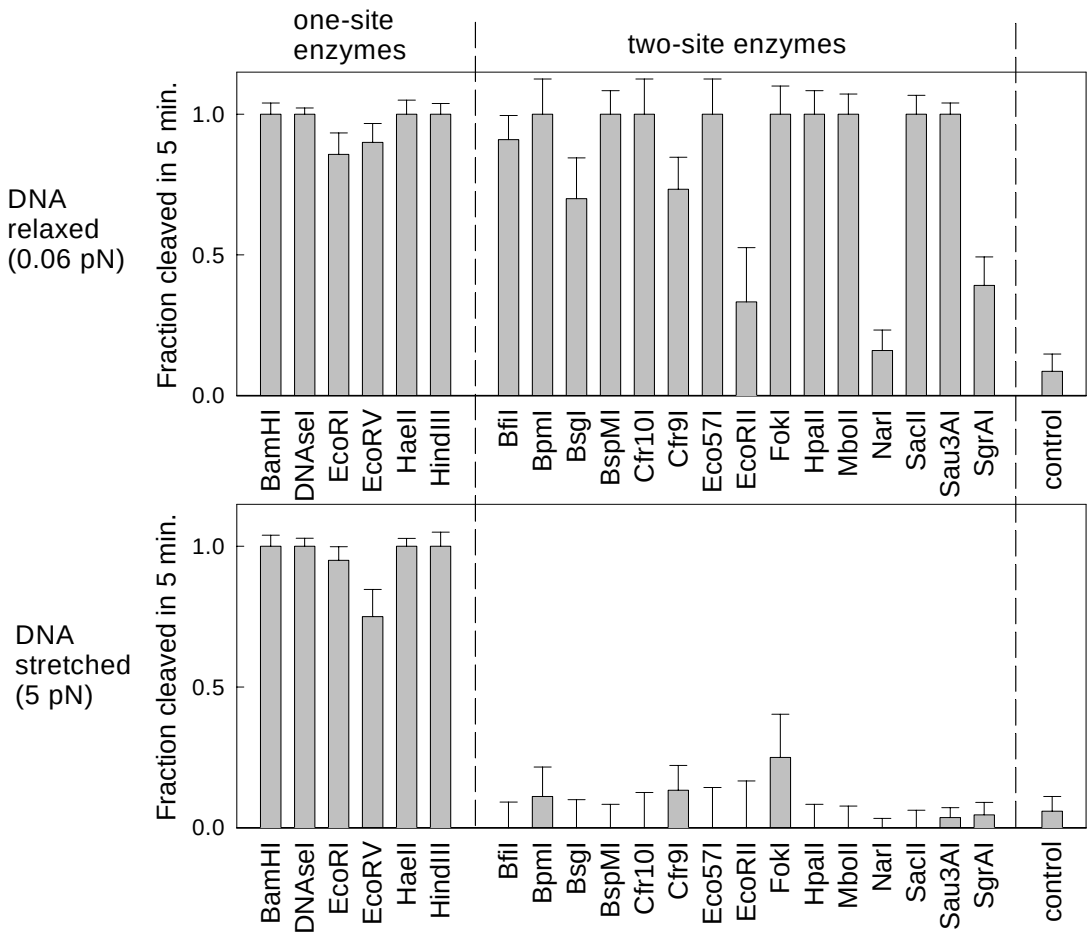

Fig. 3.2. Dependence of cleavage activity on DNA tension. Fraction of molecules that were cleaved after a 5 minute incubation with the one site enzymes on the left side of each graph and the known or suspected two site enzymes on the right side of each graph. The upper graph shows the results with relaxed DNA, while the lower graph shows the results with the DNA held stretched at an end-to-end extension of 95% (~ 5 pN). The bar on the far right labeled "control" are measurements with no enzyme, showing that a small fraction of molecules spontaneously detach from the beads even at zero tension.

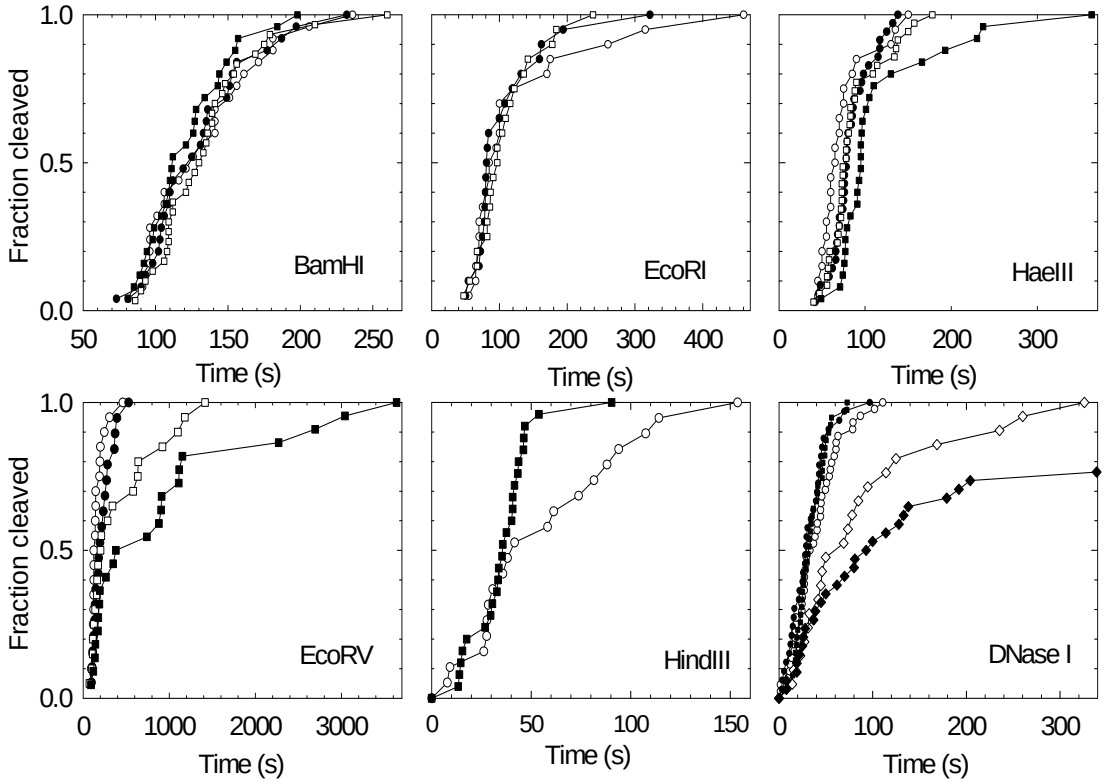

Fig. 3.3. Tension dependence of DNA cleavage with one site endonucleases. Measurements with relaxed DNA (fractional extension 35%, 0.06 pN, open circles), and tensions of 5 pN (filled circles), 20 pN (open squares), 40 pN (filled squares). Note that HaeIII and EcoRV show tension induced inhibition while HindIII and, to a lesser extent, EcoRI show tension induced enhancement. DNAse I (lower right plot) showed a slight catalysis at 5 and 40 pN, and inhibition at a very high tension of 75 pN (open diamonds). To quantify spontaneous unbinding from the microspheres at high force, negative controls without enzyme were also carried out at 75 pN (filled diamonds).

References for Chapter 3

1. Halford, S.E. (2001) Hopping, jumping and looping by restriction enzymes. *Biochem. Soc. Trans.* **29**: 363-374.

2. Roberts, R.J. (2005) How restriction enzymes became the workhorses of molecular biology. *PNAS* **102**: 5905-5908.

3. Pingoud, A. & Jeltsch, A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res*. **29**: 3705-3727.

4. Roberts, R.J., Vincze, T., Posfai, J. & Macelis, D. (2005) REBASE--restriction enzymes and DNA methyltransferases. *Nucleic Acids Res*. **33**: D230-232.

5. Kruger, D., Barcak, G., Reuter, M. & Smith, H. (1988) EcoRII can be activated to cleave refractory DNA recognition sites. *Nucleic Acids Res*. **16**: 3997-4008.

6. Halford, S.E., Welsh, A.J. & Szczelkun, M.D. (2004) Enzyme-mediated DNA looping. *Ann. Rev. Biophys. Biomol. Struct.* **33**: 1-24.

7. Schleif, R. (1992) DNA Looping. *Ann. Rev. of Biochemistry* **61**: 199-223.

8. Dunn, T.M., Hahn, S., Ogden, S. & Schleif, R.F. (1984) An Operator at -280 Base-Pairs That Is Required for Repression of Arabad Operon Promoter - Addition of DNA Helical Turns between the Operator and Promoter Cyclically Hinders Repression. *PNAS.* **81**: 5017-5020.

9. Mukherjee, S., Erickson, H. & Bastia, D. (1988) Enhancer Origin Interaction in Plasmid-R6K Involves a DNA Loop Mediated by Initiator Protein. *Cell* **52**: 375-383.

10. Su, W., Middleton, T., Sugden, B. & Echols, H. (1991) DNA Looping between the Origin of Replication of Epstein-Barr-Virus and Its Enhancer Site - Stabilization of an Origin Complex with Epstein-Barr Nuclear Antigen-1. *PNAS* **88**:10870-10874.

11. Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B. & Mullerhill, B. (1994) Quality and Position of the 3 Lac Operators of Escherichia-Coli Define Efficiency of Repression. *Embo J.* **13**: 3348-3355.

12. Weiner, B.M. & Kleckner, N. (1994) Chromosome Pairing Via Multiple Interstitial Interactions before and during Meiosis in Yeast. *Cell* **77**: 977-991.

13. Blackwood, E.M. & Kadonaga, J.T. (1998) Going the distance: A current view of enhancer action. *Science* **281**: 60-63.

14. Chen, Y. & Rice, P.A. (2003) New insight into site-specific recombination from Flp recombinase-DNA structures. *Ann. Rev. Biophys Biomol. Struct.* **32**: 135-159.

15. Finzi, L. & Gelles, J. (1995) Measurement of Lactose Repressor-Mediated Loop Formation and Breakdown In Single DNA-Molecules. *Science* **267**: 378-380.

16. Wong, O.K., Guthold, M., Erie, D. & Gelles, J. (2000) Multiple conformations of lactose repressor-DNA looped complexes revealed by single molecule techniques. *FASEB J.* A1445-A1445.

17. Lia, G., Bensimon, D., Croquette, V., Allemand, J.F., Dunlap, D., Lewis, D.E.A., Adhya, S.C. & Finzi, L. (2003) Supercoiling and denaturation in Gal repressor/heat unstable nucleoid protein (HU)-mediated DNA looping. *PNAS.* **100**: 11373-11377.

18. Sankararaman, S. & Marko, J.F. (2005) Formation of loops in DNA under tension. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys*. **71**: 021911.

19. Yan, J., Kawamura, R. & Marko, J.F. (2005) Statistics of loop formation along double helix DNAs. *Phys. Rev. E* **71**: 061905.

20. Blumberg, S., Tkachenko, A.V. & Meiners, J.C. (2005) Disruption of protein-mediated DNA looping by tension in the substrate DNA. *Biophys J*. **88**: 1692-1701.

21. van den Broek, B., Noom, M.C. & Wuite, G.J. (2005) DNA-tension dependence of restriction enzyme activity reveals mechanochemical properties of the reaction pathway. *Nucleic Acids Res*. **33**: 2676-2684.

22. Fuller, D.N, Gemmen, G.J., Rickgauer, J.P., Dupont, A., Millin, R., Recouvreux, P., & Smith, D.E. (2006) A General Method for Manipulating DNA Sequences from any Organism with Optical Tweezers. *Nuc. Acids Res*. **34**: e15.

23. Gemmen, G.J., Sim, R., Haushalter, K.A., Ke, P.C., Kadonaga, J.T. & Smith, D.E. (2005) Forced unraveling of nucleosomes assembled on heterogeneous DNA using core histones, NAP-1, and ACF. *J. Mol. Biol*. **351**: 89-99.

24. Topal, M.D., Thresher, R., Conrad, M. & Griffith, J. (1991) NaeI endonuclease binding to pBR322 DNA induces looping. *Biochemistry* **30**: 2006-2010.

25. Siksnys, V., Skirgaila, R., Sasnauskas, G., Urbanke, C., Cherny, D., Grazulis, S. & Huber, R. (1999) The Cfr10I restriction enzyme is functional as a tetramer. *J Mol. Biol*. **291**: 1105-1118.

26. Mucke, M., Lurz, R., Mackeldanz, P., Behlke, J., Kruger, D.H. & Reuter, M. (2000) Imaging DNA loops induced by restriction endonuclease EcoRII. A single amino acid substitution uncouples target recognition from cooperative DNA interaction and cleavage. *J. Biol. Chem*. **275**: 30631-30637.

27. Friedhoff, P., Lurz, R., Luder, G. & Pingoud, A. (2001) Sau3AI, a monomeric type II restriction endonuclease that dimerizes on the DNA and thereby induces DNA loops. *J. Biol. Chem*. **276**: 23581-23588.

28. Szczelkun, M.D. & Halford, S.E. (1996) Recombination by resolvase to analyse DNA communications by the SfiI restriction endonuclease. *Embo J*. **15**: 1460-1469.

29. Milsom, S.E., Halford, S.E., Embleton, M.L. & Szczelkun, M.D. (2001) Analysis of DNA looping interactions by type II restriction enzymes that require two copies of their recognition sites. *J Mo.l Biol.* **311**: 515-527.

30. Watson, M.A., Gowers, D.M. & Halford, S.E. (2000) Alternative geometries of DNA looping: an analysis using the SfiI endonuclease. *J Mol. Biol.* **298**: 461-475.

31. Katiliene, Z., Katilius, E. & Woodbury, N.W. (2003) Single molecule detection of DNA looping by NgoMIV restriction endonuclease. *Biophys. J.* **84**: 4053-4061.

32. Yan, J., Skoko, D. & Marko, J.F. (2004) Near-field-magnetic-tweezer manipulation of single DNA molecules. *Phys Rev E Stat Nonlin Soft Matter Phys*, **70**: 011905.

33. Bath, A.J., Milsom, S.E., Gormley, N.A. & Halford, S.E. (2002) Many type IIs restriction endonucleases interact with two recognition sites before cleaving DNA. *J. Biol. Chem*. **277**: 4024-4033.

34. Soundararajan, M., Chang, Z., Morgan, R.D., Heslop, P. & Connolly, B.A. (2002) DNA binding and recognition by the IIs restriction endonuclease MboII. *J. Biol. Chem*. **277**: 887-895.

35. Gowers, D.M., Bellamy, S.R. & Halford, S.E. (2004) One recognition sequence, seven restriction enzymes, five reaction mechanisms. *Nucleic Acids Res*. **32**: 3469-3479.

36. Reuter, M., Kupper, D., Pein, C.D., Petrusyte, M., Siksnys, V., Frey, B. & Kruger, D.H. (1993) Use of specific oligonucleotide duplexes to stimulate cleavage of refractory DNA sites by restriction endonucleases. *Anal. Biochem*. **209**: 232-237.

37. Kupper, D., Moncke-Buchner, E., Reuter, M. & Kruger, D.H. (1999) Oligonucleotide stimulators allow complete cleavage of agarose-embedded DNA by particular type II restriction endonucleases. *Anal. Biochem*. **272**: 275-277.

38. Vanamee, E.S., Santagata, S. & Aggarwal, A.K. (2001) FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.* **309**: 69-78.

39. Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S., Dryden, D.T., Dybvig, K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res*. **31**: 1805-1812.

40. A sixteenth putative two site enzyme, Ksp632I, was also studied but we did not observe any cleavage events in the optical tweezers measurments, although cleavage with the same sample of enzyme was observed following a one-hour bulk reaction as assessed by agarose gel electrophoresis.

41. Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. & Wilson, K.S. (1993) The Crystal-Structure of Ecorv Endonuclease and of Its Complexes with Cognate and Non-Cognate DNA Fragments. *Embo. Journal* **12**: 1781-1795.

42. Kim, Y.C., Grable, J.C., Love, R., Greene, P.J. & Rosenberg, J.M. (1990) Refinement of Eco Ri Endonuclease Crystal-Structure - a Revised Protein Chain Tracing. *Science* **249**: 1307-1309.

43. Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. & Aggarwal, A.K. (1995) Structure of BamHI Endonuclease Bound to DNA - Partial Folding and Unfolding on DNA-Binding. *Science* **269**: 656-663.

44. Suck, D. (1994) DNA recognition by DNase I. *J. Mol. Recognit.* **7**: 65-70.

Chapter 3 Acknowledgement

The text of Chapter 3, in full, appears Gemmen, G.J., Millin, R. & Smith, D.E. (2006). Tension-Dependent DNA Cleavage by Restriction Endonucleases: Enzymes Binding at Two Sites are "Switched Off" at Low Force. Submitted to *The Proceedings of the National Academy of Sciences, U.S.A*. The dissertation author was a primary researcher and secondary author for the research which forms the basis of this chapter.

Chapter 4. DNA Looping by Two Site Restriction Endonucleases: Heterogeneous Probability Distributions for Loop Size and Unbinding Force

4.1. Abstract

Proteins interacting at multiple distant sites via DNA looping play an important role in many fundamental biochemical processes. Restriction endonucleases that require interaction at two recognition sites for efficient activity are a model system for studying such interactions. Here we used single DNA manipulation to study sixteen known or suspected two site endonucleases. For BpmI, BsgI, BspMI, Cfr10I, Eco57I, EcoRII, FokI, HpaII, NarI, Sau3AI, and SgrAI, we found that changing Ca^{2+} for Mg^{2+} blocked cleavage and allowed us to observe stable DNA loops. Forced disruption of these loops allowed us to measure the rate of looping and probability distributions for loop size and unbinding force. For four enzymes we observed bimodal force distributions, indicating conformational heterogeneity and/or complex binding energy landscapes. Measured unlooping events ranged in size from 7-7500 bp and the most probable size ranged from ~75 to ~500 bp, dependent on enzyme. In most cases the size distributions were in closer agreement with theoretical models that assume sharp DNA kinking than with classical models of DNA elasticity. Our findings indicate that DNA looping is highly dependent on the specific protein and does not depend solely on the mechanical properties of DNA.

4.2. Introduction

A wide range of biological processes, including DNA transcription, replication, and repair, involve formation of protein complexes that bind at multiple

sites on the same DNA molecule. DNA looping allows for interaction between such proteins bound at distant sites, and has been shown to have important implications for the regulation of DNA directed biochemical processes (1-8). For example, DNA looping provides a mechanism for enhancing or repressing gene expression (9).

Many theories have made predictions for the dependence of the rate of loop formation on loop size (10-15). Polymer physics suggests there should be an optimum size loop. Very small loops are unfavorable due to DNA bending rigidity, whereas very large loops are unfavorable due to entropy. Considering only DNA mechanics, an optimum loop size of ~500 bp is predicted by models which treat DNA as a homogeneous worm like chain with a persistence length of 150 bp. Such models are in decent agreement with the elastic behavior of DNA as deduced in cyclization measurements (16). While cyclization in principle involves only DNA mechanics, theories for protein induced DNA looping must account for additional effects such as protein interaction distance, protein induced DNA bending and kinking, loop-loop interactions, elasticity of the protein complex, sequence dependent DNA bending and flexibility, and DNA helicity (Fig. 4.1). Notably, models with DNA kinking predict considerably smaller optimum loop sizes, down to ~100 bp. (17,18,19).

Two site restriction endonucleases (REases) acting on long linear DNA molecules provide a model system for studying protein mediated DNA looping (20). For example, with NaeI, Cfr10I, EcoRII, and Sau3AI, stable loops are formed when Ca^{2+} is substituted for the normal cofactor Mg^{2+} and have been directly imaged by electron microscopy (21-24). Evidence of looping with SfiI and Cfr10I comes from DNA recombination and gel mobility shift measurements (25-27), with NgoMIV

from fluorescence resonance energy transfer measurements (28), with BspMI from magnetic tweezers measurements (29), and with NarI and NaeI from tethered particle assays (30). For other enzymes, measurements show that efficient cleavage only occurs on templates containing two or more sites (31-34). Finally, cleavage with other enzymes has been found to be stimulated by short oligonucleotide duplexes that have the recognition sequence, suggesting that the enzyme complex is able to bind at two sites in *trans* (i.e. on two different molecules) (35-38).

In Chapter 3, cleavage of single DNA molecules by twenty-one different one site and two site REases was studied. A tension of 5 pN inhibited the enzymes that required DNA looping. Here we report measurements of stable DNA looping with ten of these two site REases when Ca^{2+} was substituted for Mg^{2+}. These measurements allow us to characterize the rate of looping, distribution of loop sizes, and binding strengths of loops. Our data indicates that DNA looping depends on the structural specifics of the protein-DNA complex and does not depend solely on the mechanical properties of DNA. We compare our measurements of the distribution of loop sizes with many different theoretical predictions in the literature. We find that in most cases our data agree more closely with predictions of models that postulate sharp DNA kinking rather than with predictions of classical worm like chain models.

4.3. Methods

4.3.1. Endonucleases

BamHI, BpmI, BsgI, BspMI, EcoRI, EcoRV, FokI, HaeIII, HpaII, MboII, NarI, SacII, Sau3AI, SfiI, and SgrAI were obtained from New England Biolabs

(NEB). BfiI, Cfr9I, Cfr10I, and Eco57I were obtained from Fermentas. EcoRII and Ksp632I were obtained from Roche. The two sites enzymes and their properties are listed in Table 4.1.

4.3.2. DNA Constructs

pBAC-A was prepared by ligating a 4282 bp digoxygenin (DIG)-labeled PCR fragment to a 10845 bp biotin end-labeled restriction fragment of pBACe3.6. The PCR fragment was generated by amplification of a sequence from pFastBac HT-b (Invitrogen) using the primers 5'-GTGGTATGGCTGATTATGATC and 5'-GCAGCCTGAATGGCGAATGG-3' and was labeled by incorporation of 20 μM of dUTP-11-DIG (Roche) and 200 μM each of dATP, dCTP, dGTP, dTTP in the PCR. The multiple DIG labeling was used to provide a stronger attachment in some experiments but was not necessary. The 10845 bp fragment was made by digesting pBACe3.6 (Children's Hospital of Oakland Research Institute) with BsrGI (NEB) and end labeling by using the Klenow fragment of *E. coli* DNA polymerase I, exo$^-$ (NEB) to incorporate dATP-14-biotin (Invitrogen). Both fragments were purified using the Qiagen PCR purification kit and digested with XhoI (NEB). To isolate the desired product the samples were run on a 1% agarose gel in 1X TAE buffer and purified using the Qiagen Gel Extraction kit. The two fragments were then ligated by use of T4 DNA ligase (NEB). *pBAC-B* was prepared by labeling the aforementioned biotin labeled, 10,845 bp XhoI fragment of pBACe3.6 by using the Klenow fragment of DNA polymerase I, exo- to incorporate dUTP-11-DIG.

½-λ-L was prepared by using the Klenow fragment of DNA polymerase I, exo$^-$ to fill in the ends of methyladenine-free λ DNA (NEB) with biotin-dATP and dCTP (Invitrogen). The DNA was then digested by XbaI and purified using the Promega Wizard DNA clean up kit. A second fill-in was done with digoxygenin-labeled dUTP (Roche). The fragments were then XhoI digested to select the left end (24,508 bp). *½-λ-R* was prepared in the same manner, except ApaI was used instead of XhoI to select the right end (23,994 bp). *PhiX174 DNA* was purchased from NEB and was labeled by digesting with XhoI and end labeling with dATP-14-biotin (Invitrogen) and purified using the Promega Wizard DNA clean up kit. The DNA was then digested with StuI and end labeled with dUTP-11-DIG.

4.3.3. DNA Tethering

Streptavidin coated microspheres (200 µl of 0.5% w/v, 2.2 µm diameter, Spherotech) were washed by twice centrifuging at 10,000×*g* and resuspended in 200 µl of phosphate buffered saline (PBS) pH 7.4 (Fisher Scientific) and 0.1 mg/ml Bovine serum albumin (BSA) (NEB). 5 µl of diluted DNA (~10-100 ng/µl) was mixed with 5 µl of beads and incubated for ~45 min. at room temperature on a slowly rotating mixer. 5-10 µl of these beads were diluted in 0.5 ml of PBS and loaded into a 1 ml tuberculin syringe for injection into the sample chamber.

Protein G coated beads (200 µl of 0.5% w/v, 2.8 µm diameter, Spherotech) were washed in the same manner and resuspended in 20 µl PBS. Then, 5 ul 200 µg/ml of antidigoxygenin (Roche) was added. The beads were incubated on the

mixer for ~45 minutes and then washed three more times and resuspended in 20 μL PBS. 5 μl of the beads were loaded into a syringe for injection into the flow cell.

4.4. Results

4.4.1. Formation and Disruption of Stable DNA Loops

The measurement technique is illustrated in Fig. 4.2. A single DNA molecule was held stretched (fractional extension 95%, corresponding to a tension of 5 pN) while the enzyme solution was flowed into the sample chamber. The enzymes (Table 4.1.) were diluted in their standard reaction buffers except that Ca^{2+} was substituted for Mg^{2+}. The flow was continued for 30 seconds to insure that the entire chamber was uniformly filled with the enzyme and the DNA was then relaxed (to a fractional extension of 35%, corresponding to a low tension of ~0.06 pN). The molecules were incubated for two minutes to allow enzyme binding to occur. The DNA was then stretched at a rate of 150 nm/s while the force was recorded at 100 Hz.

If the DNA remained tethered after reaching a tension of 60 pN it was relaxed back to 5 pN. Incubation and stretching were then repeated. If the tether broke, which typically occurred after 1 to 10 stretch cycles due to detachment of the DNA from the beads, the enzyme was drained from the flow cell, a new DNA molecule was tethered, and a new batch of enzyme solution was introduced. Measurements were repeated ~70 to 300 times with each enzyme depending on the activity level.

Typical force-extension data is shown in Fig. 4.3. Prior to introducing the enzyme, the measured elasticity was as expected for a single DNA molecule (39). After interacting with the enzyme, the tether was often shortened by a variable length,

consistent with loop formation. Upon stretching we recorded sudden drops in the measured force, each followed by an increase in tension. These "sawteeth" indicate events in which sequestered lengths of DNA are suddenly released, consistent with opening of the DNA loops. Analysis then yielded the number of loops formed, and the disruption force and DNA length change associated with each. By performing many measurements we obtain distributions for these quantities for each enzyme.

The observed length changes were clearly within the range of the possible loop sizes that could form given the separations of recognition sites on the DNA templates. In measurements with templates containing relatively few recognition sites, we typically saw only zero or one event per trial. An example is shown in Fig. 4.3a, in which BsgI was tested on a template containing only four recognition sites. The plot shows the only event recorded in over a hundred trials. This event indicated a length change of 946 bp, which corresponds closely to one of the possible site separations on the DNA template (the 945 bp distance between sites at 8342 and 9287 bp). To obtain better statistics with BsgI, we switched to a template with a larger number of sites. Multiple sawteeth were generally observed when DNA template contained many sites. An example is shown in Fig. 4.3b, in which Sau3AI was tested on a template containing 55 recognition sites.

Four types of control experiments were done (Fig. 4.3c). First, DNA was stretched in the reaction buffers with no enzymes added to confirm that these buffers do not cause events. Second, several one site REases (BstNI, HaeIII, and MspI) were tested and no events were observed. Two of these, BstNI and MspI, were chosen for controls because they are isoschizomers of the two site enzymes EcoRII and HpaII.

Third, several two site enzymes (NarI, SacII, SgrAI, and Sau3AI) were tested with templates containing no recognition sites and, with the exception of SgrAI, no events were observed. SgrAI was unusual in that it appeared to cause frequent non-specific looping, which may be connected with a previous report that SgrAI can bind certain non-cognate sites (40). Fourth, several two site enzymes (SfiI and NaeI, SacII) were tested on a DNA template containing only one site and no looping was detected.

Stable loops in solutions with Ca^{2+} were not detected with five of the sixteen two site enzymes studied. For BfiI and SacII, loops were not possible to assess because DNA cleavage occurred despite the absence of Mg^{2+}. Such behavior has been reported previously for BfiI (41). However, in previous work we found that tension inhibited cleavage with both BfiI and SacII, a finding that is strongly indicative of DNA looping. No stable loops were detected in 50 trials with Ksp632I, which had 20 sites on the DNA template, despite the use of extended incubations of 5 minutes. We also detected no stable looping with either Cfr9I or MboII, although we had previously observed inhibition of cleavage by tension, indicating a looping mechanism. Although evidence for Ca^{2+} dependent binding of MboII to DNA in *trans* has been reported (33) we did not detect looping with MboII in Ca^{2+}.

Disruption of loops with the enzymes that supported stable looping in Ca^{2+} occurred at forces ranging from ~ 5 to ~60 pN (Fig. 4.4). The most probable force varied from ~10 to 30 pN (Fig. 4.5), similar to other measured protein-DNA interactions, such as the forces required to disrupt nucleosomes (42). Statistics for nine of the enzymes are given in Table 4.2. NarI and SgrAI were not included in the table because of insufficient statistics and due to their tendency to form non-specific loops

(30,40). The unlooping force distributions for different enzymes had different shapes. For example, those for EcoRII and BpmI were broad and roughly Gaussian whereas those for HpaII, Cfr10I, and EcoRII were broad but skewed to low force.

Interestingly the force distributions for FokI, Sau3AI, BspMI, and BsgI were bimodal. Such distributions may have several explanations. As DNA looping generally involves simultaneous protein-DNA and protein-protein interactions we first thought that the two force regimes simply corresponded to these two separate interactions. However, this could not be correct in most cases because if each loop were held together by two links of unequal strengths, the unbinding of the weaker link would be more frequent. Only in the case of Eco57I was there evidence of a smaller peak at high forces. The opposite was seen with BspMI, BsgI, FokI, and Sau3AI, so an alternative explanation is needed. Recent theoretical studies suggest that bimodal force distributions can arise from conformational heterogeneity and/or multibarrier binding energy landscapes (43).

4.4.2. Dependence of Loop Formation on Divalent Cations

Many protein-DNA interactions are strongly dependent on divalent cations, and Mg^{2+} as a cofactor for cleavage activity is a hallmark of the majority of Type II REases. To investigate the divalent cation requirements for DNA looping we carried out experiments in which 1 mM EDTA was substituted for Ca^{2+}. In the case of BpmI, BsgI, BspMI, Cfr10I, Eco57I, and HpaII, we did not detect any loop formation, consistent with the aforementioned expectations,. Suprisingly, however, we did detect stable loop formation with EcoRII, FokI, and Sau3AI (Fig. 4.6).

Distributions of loop disruption forces with Ca^{2+} or EDTA are shown in Fig. 4.6. In all cases fewer loops were formed in EDTA than in Ca^{2+} and with Sau3AI and EcoRII the average unbinding force was lower with EDTA, in accord with the expectation that divalent ions increase the binding specificity. In the case of FokI, however, the unbinding force was unexpectedly higher with EDTA. With Sau3AI the force distribution was bimodal with Ca^{2+} but became monomodal in EDTA, which suggests qualitatively different binding modes dependent on ionic conditions.

Measurements with HpaII were carried out by varying [Ca^{2+}] from 1 μM to 300 mM (Fig. 4.7). No events were observed for this enzyme with EDTA and events were rarely observed with [Ca^{2+}] concentrations below 0.1 mM. The number of loops increased ~10-fold as [Ca^{2+}] was increased from 0.1 to 5 mM, but then dropped ~10-fold as the calcium was increased up to 300 mM. The optimum point for looping was at ~5-10 mM, the same range reported to be optimal for cleavage. Surprisingly, while the number of loops decreased with [Ca^{2+}] above 5 mM, suggesting nonspecific salt inhibition, the unbinding force increased steadily with [Ca^{2+}] up to 100 mM. This effect may indicate that individual complexes can have different numbers of bound ions resulting in different sub-populations having different binding strengths (44).

4.4.3. Distributions of Loop Sizes

Although many theoretical models have predicted the dependence of the probability of loop formation on loop size, little experimental data is available for comparison with these theories. Also, little is known about how loop sizes depend on the protein involved. An advantage of our method is that loop sizes are measured

directly from the unlooping events in the force-extension data (Fig. 4.3), and loop size distributions (Fig. 4.8) may be obtained from an ensemble of such measurements.

The separations between sites on the DNA dictate the loop sizes that can form in our experiment. Plots of these distributions are shown for several of the enzymes in Fig. 4.9. The use of long DNA templates with many binding sites yields distributions that are generally broad and extend out to ~10-20 kbp. Comparisons of measured to possible loop sizes are shown in Fig. 4.8. We use a bin size equal to the persistence length of DNA (150 bp), a characteristic measure of DNA flexibility (39). Several features are evident. The distributions of possible sizes are continuous and nearly flat over the range from 0 to 3000 bp, while the measured distributions are strongly skewed towards shorter loops. The finding of few long loops is consistent with the theoretical notion that long loops are unlikely due to entropic effects.

While all distributions indicate strong suppression of large loops, the distributions of short loops vary dramatically for different enzymes. Among the four examples shown in Fig. 4.8 the shortest loops were formed with EcoRII and Cfr10I, although the distribution is clearly broader with Cfr10I. The distribution is also broad with BsgI, but is shifted to larger loop size. To determine the inherent probability distributions for loop size and to correct for the influence of DNA template, we normalized each distribution by dividing the number of measured events in each bin by the number of possible loop sizes. We also reduced the bin widths, which enabled us to identify optimum loop sizes in most cases (Fig. 4.11). The finding of few very short loops is consistent with the theoretical notion that very short loops are unfavorable due to the bending rigidity of DNA. As shown in Fig. 4.11, the optimum

loop size ranged from <75 bp up to ~500 bp, dependent on enzyme. Interestingly, with BpmI and BspMI the size distributions had small secondary peaks. Such behavior may indicate different possible modes for these enzymes. Statistics for all ten enzymes are given in Table 4.2.

Individual unlooping events ranged from as small as 7 bp to as large as ~7500 bp. We were limited in resolving small loops not by instrument resolution (~ 5 bp) but by the distribution of sites in the DNA template, as very few sites were separated by less than 10 bp. Minimum sizes measured with different enzymes ranged from 7 to 60 bp (Table 4.2). Our findings clearly show that loops substantially smaller than the persistence length are readily formed. In Fig. 4.11 the rates of formation of short loops with different enzymes are compared. In the case of the very small loops (shorter than ~30 bp) the use of the term "loop" is perhaps not appropriate. In these cases the two proteins are likely bound adjacent to each other with the DNA simply traversing across them, rather than looping freely through the solution. The elasticity of the protein may be quite important in facilitating the formation of such complexes.

Comparisons of the experimental loops size distributions with predictions of a number of different theoretical models are shown in Figs. 4.10 and 4.11. On clearly sees that, in all cases except BsgI and possibly BspMI, the measured loop sizes agree better with models that assume sharp DNA kinking (Fig. 1) than with models that assume classical DNA elasticity (homogeneous bending). Also it is evident that no single model was able to describe all of our data sets. Systematic comparisons of all of the data sets with all of the published models were performed but are not shown.

4.4.4. Frequency of Loops

The number of loops formed in a single DNA molecule after an incubation time can be directly counted in the force-extension data (Fig. 3). In our experiments an incubation of two minutes was used because the majority of looping events occurred in this time interval. Longer times did not increase the number of loops proportionally, possibly due to the gradual "self-tensioning" of DNA. The total number of loops that could form in a given molecule is equal to $N_{sites}/2$, truncated to the nearest integer. In an ensemble of measurements, however, the number of different possible loops equals the number of combinations of pairs of sites $N_{pairs} = N_{sites}(N_{sites}-1)/2$. To assess the relative rates of looping with different enzymes, we calculated the mean number of loops per number of possible pairs of sites ($<N>/N_{pairs}$) (Table 4.2). Significant differences with different enzymes are evident.

4.5. Discussion

4.5.1. Evidence for DNA looping

We detected DNA looping in buffers containing Ca^{2+} with eleven of the sixteen two site enzymes studied. With BspMI, Cfr10I, EcoRII, NarI, and Sau3AI, this corroborates previous evidence for DNA looping (22-24,26,29,30). Our observations with BpmI, BsgI, and SgrAI confirm that these enzymes operate by DNA looping, as postulated based on the finding that two sites are required for efficient cleavage (32,45). We report the first evidence for DNA looping with Eco57I, FokI, and HpaII, which were previously reported to have activity stimulated by interaction with DNA in *trans* (35-37).

4.5.2. Loop Disruption Forces

In our experiments loop opening is accelerated by sudden application of force and the disruption forces provide a measure of binding strength. One advantage of this approach is that that we can rapidly probe high affinity complexes that would take an extremely long time to dissociate under equilibrium conditions (46). Average unlooping forces measured in our experiment ranged from ~15 to 30 pN. These values may be compared with forces measured in optical tweezers studies of other protein-DNA interactions. Similar forces of were measured for unraveling of DNA bound in nucleosomes and shown to correspond to a very slow zero force dissociation rate of $\sim 10^{-7}\ s^{-1}$ assuming a single energy barrier model. However, DNA loops formed by Cfr10I were reported to have lifetime of only ~90 seconds (26), which, taken together with our data, would imply a very different binding energy landscape.

Cfr10I and FokI are the only two site enzymes for which binding affinities have been reported, and they were found to be similar (47,48). In contrast we found a bimodal force distribution with FokI, suggesting multiple binding modes, and only the higher of the two peak forces was nearly the same as that measured with Cfr10I. Crystal structure data shows that FokI has a small dimerization interface and is therefore predicted to have relatively weak association compared with other REases (49). Our measurements support this prediction as FokI had the lowest average disruption force (~15 pN).

We analyzed the unbinding forces for correlations with enzyme properties. One might expect larger proteins and/or those recognizing longer sequences to bind more strongly. Little correlation was observed with molecular weight ($\rho_{F,MW} = -0.04$),

but a positive correlation ($\rho_{F,Seq} = 0.5$) was observed between average force and recognition sequence length. This correlation was not universal. For example, although BspMI recognizes a six base pair sequence and Sau3AI recognizes a four base pair sequence, they exhibited similar unlooping forces.

We also searched for possible correlations between unbinding force and loop size. Due to the rigidity of DNA, loops are under mechanical stress; such stress would be expected to accelerate the opening of the loop. However, the applied force could act in a direction to relieve this stress (for an antiparallel loop geometry, Fig. 1), or to increase it by pulling the loop smaller (for a parallel loop geometry). Such effects are expected to be more significant for small loops. Interestingly, for loops less than one persistence length there was a small negative correlation between unbinding force and loop size for most enzymes (Table 4.2).

4.5.3. Bimodal Force Distributions

The unlooping force distributions for FokI, Sau3AI, BspMI, and BsgI were bimodal. The mean sizes of loops within the two different peaks in the force distribution were not significantly different. Therefore, a simple explanation for the bimodality is to postulate molecular heterogeneity in the enzyme complexes. Individual complexes could exhibit different binding modes with different binding strengths, resulting from different geometries, binding surfaces, or coordinated ions. Alternative loop geometries have been postulated previously for SfiI based on varying mobilities observed in gel shift assays (27). As many of these enzymes form dimers or pseudodimers the weaker and stronger events could correspond to binding

of complexes at one versus two surfaces. Molecular heterogeneity could also stem from variable protein-DNA interactions. Some REases, such as NaeI, have been reported to contain two different DNA binding clefts (50) while others, such as SgrAI, have been reported to be able to bind one cognate site and one secondary site differing slightly in sequence (40). However, no such effects have been reported in the cases of FokI, Sau3AI, BspMI, and BsgI.

Therefore, we find an alternative explanation from a recent theory to be attractive. This theory states that a bimodal force distribution can arise during forced disruption of homogeneous complexes if the interaction has a multiple energy barrier landscape (43). Bimodal distributions were recently observed in the unbinding of lipid molecules from a bilayer (46) and in the unbinding of RAN protein from nuclear importin receptor B1 (51). To our knowledge the present findings constitute the first observations of bimodal unbinding forces for a protein-DNA complex.

4.5.4. Dependence on Ionic Conditions

Most Type II REases require divalent cations as a cofactor. Ca^{2+} ions facilitate specific binding, but Mg^{2+} ions facilitate specific binding and cleavage of DNA. This dependence on ions for specific binding is known to be dramatic in certain cases. For example, the dissociation constant of the one site REase PvuII is reported to be ~60,000 times lower in 10 mM Ca^{2+} than in 1 mM EDTA, according to a filter binding measurement (52). Studies of the two site enzyme SfiI by recombination assays showed a similar effect. The bound complex lasted >7 hours in Ca^{2+} versus ~4 minutes in EDTA (26).

Notably, we observed stable looping for EcoRII, FokI, and Sau3AI with no divalent ions. Previous studies for EcoRII and Sau3AI by electron microscopy utilized solutions containing Ca^{2+} but did not report whether it was required for looping. A recent study of FokI complexes by analytical ultracentrifugation concluded that FokI is a monomer in solution and dimerizes only in the presence of divalent cations (37). Although we observed clear looping with FokI in ETDA, these complexes exhibited the lowest observed average disruption force. Apparently the looped complex we detected with our method is not stable enough to be detectable by analytical ultracentrifugation. We note that our method allows loop formation to be probed on time scales down to ~1 second.

4.5.5. Differences in Frequency of Loops

Significant differences in the average frequency of looping (measured by $<N>/N_{pairs}$) with different enzymes are clearly evident (Table 4.2). For example, one may compare the results on EcoRII and Eco57I. Both were used at the same concentration and incubated with the DNA for the same amount of time, yet the looping rate with Eco57I was nearly an order of magnitude higher than with EcoRII. On this basis, one may be tempted to conclude that Eco57I has a higher affinity to bind DNA in loops. Surprisingly, however, a lower unbinding force was measured with Eco57I than with EcoRII (Table 4.2). The major difference between these two cases was that Eco57I exhibited a dramatically larger optimum loop size than EcoRII (~250 vs. < 75 bp) (Fig. 4.10). Therefore, in the comparison between these two enzymes, bending energy appeared to be more important in dictating the rate of

looping. However such a trend is not strictly universal because across different enzymes we measured no statistically significant correlation between average frequency of looping and optimum loop size ($\rho_{N,\Delta L} = 0.21$), frequency and average unbinding force ($\rho_{N,F} = 0.37$), or size and force ($\rho_{F,\Delta L} = 0.07$). Thus, looping properties are strongly dictated by the structural specifics of each protein-DNA complex.

4.5.6. Loop Sizes

Many studies of DNA have found quite good agreement with the predictions of the worm like chain (WLC) model. Measurements of the elasticity of single DNA molecules, in particular, are in excellent agreement with this model and indicate a persistence length $P \cong 150$ bp (53-56). In early work, Stockmayer and Yamakawa calculated that a "teardrop" shape, corresponding to a loop opening angle of ~81°, would minimize bending energy in the WLC model. Shimada and Yamakawa subsequently considered thermal fluctuations and derived an expression for the cyclization probability with molecular length. Although there are clearly differences between cyclization and looping, it has been proposed by a number of authors that the Shimada-Yamakawa expression can be used as an approximation in the case of DNA looping (10,13,14). Further modifications to the simple WLC model have also been proposed to account for finite interaction distance (i.e. the span of the protein fixing the loop) (11,57,58), the effect of DNA tension (14,29), different elastic potentials (16), and entropic interactions between loops (12). More detailed molecular models have also been proposed that can account for sequence dependent DNA curvature and flexibility and protein geometry and flexibility (18,19).

However, while classical models of DNA elasticity have been successfully applied to describe many experimental findings, a recent study surprisingly reported much faster cyclization of 94 bp DNA molecules than is anticipated by this model (59,60). Subsequently, a number of new theoretical models were proposed that allow for the possibility of spontaneous sharp kinks in DNA (16,61,62). However, reinterpretation of the cyclization data along with further measurements and calculations have led Du, *et al.,* to conclude that spontaneous kinks are unlikely and that the data can be reconciled with the WLC model (16). On the other hand, DNA kinking may be quite relevant in the case of protein-mediated DNA looping as many proteins, including some REases such as EcoRV, induce sharp bends in DNA (63). Several theoretical models have analyzed the effect of such permanent kinks (14,15).

In most cases the observed loop size distributions were in better agreement with models that assume sharp DNA kinking than with models that assume classical DNA elasticity (e.g., Figs. 4.9 and 4.10). Only with BsgI and BspMI were non-kinked models in better agreement, although in both cases there were a significant number of loops that were shorter than predicted. The classical Shimada-Yamakawa model predicts an optimum size of ~500 bp and very few loops <200 bp, whereas most of the experimental distributions indicated optimum values of <200 bp. Within classical models a number of possible effects may reduce loop sizes. First, it is possible that the DNA persistence length could be shorter than the often assumed value of 150 bp. Values as low as ~120 bp have been reported in solutions containing divalent cations (55). Second, the finite span of the protein complex lowers the predicted sizes. Third, entropic compression can occur when multiple loops form. In

the case of BsgI and BspMI these considerations led to better agreement with theoretical predictions. Predicted distributions for entropically compressed loops and for loops with a 10 nm protein span agreed more so with the data than did the Shimada-Yamakawa model or models assuming sharp kinks. However, in the other cases these corrections were insufficient to reconcile the data with classical models.

The finite span of the protein complex may also lead to an underestimation in the measured loop sizes. When a loop is disrupted we measure the change in length ΔL of the DNA but this does not exactly correspond to the distance between the two sites. The initial length of a complex is given by $L_{initial} = (L_{DNA}-L_{loop}) + d$, where L_{DNA} is the total length of the DNA molecule, d is the span of the protein, and L_{loop} is the length of DNA sequestered. As the final length $L_{final} = L_{DNA}$, we have $\Delta L = L_{loop} - d$. Thus, when comparing measured and predicted "loop sizes," one must consider the magnitude of d. Based on the molecular weights of the enzyme complexes we estimate that d may be ~5 to 20 bp. Thus, the measured loop sizes are expected to slightly underestimate the length of DNA sequestered inside the loop. However, this effect cannot reconcile the small loops we observe with the classical models.

Another experimental consideration concerns DNA tension. Two groups have recently modeled this effect theoretically and a shift to smaller loop size is predicted with increasing DNA tension. Such predictions, however, cannot reconcile our findings of loop sizes smaller than predicted by classical theory. The DNA was held an extension of 35% during loop formation, which corresponds to a tension of ~0.06 pN. However this tension is relatively small and the calculations of Sankararaman and Marko predict that it would have negligible effect on the loop sizes.

In the cases of BpmI, Cfr10I, Eco57I, EcoRII, FokI, HpaII and Sau3AI the observed loop size distributions were in better agreement with models that assume sharp DNA kinking. The model that often fit best over sizes ranging from 0 to 1000 bp was the 90° kink model of Ref. 14. However, the expressions proposed by Rippe for 30-120° kinks and a 10 nm protein span fit better to most data sets over the range from 150 to 600 bp. These models postulate a permanent kink at the apex of the loop and were in better agreement with the data than those that assume spontaneous DNA kinking. Yan, *et al.*, have considered the possibility of formation of "thermally activated hinges" in DNA due to localized strand separation (61). Such events were predicted to be rare, however, with an average separation of >1 kbp. Furthermore, on the basis of cyclization experiments and Monte-Carlo simulations, Du *et. al.* conclude that the probability for spontaneous kinking of a 100 bp segment of DNA is less than 0.02%. Thus, we conclude that it is unlikely that spontaneous kinks would occur often enough between closely spaced sites (a small fraction of the overall DNA length) to explain the shortness of our optimum loop sizes. Even the models that postulate permanent kinks do not fit perfectly and cannot account for the preponderance of short loops observed in some cases (e.g. with EcoRII and FokI).

In our experiment DNA kinking between two recognition sites could potentially occur if additional protein binding sites occurred between the two sites in question. Certain REases are known to induce sharp bending of DNA. For example, EcoRV bends the DNA by ~50°. Of the enzymes we observed to form short loops, FokI is the only one for which a structure of the enzyme-DNA complex has been determined and this structure shows that FokI does not induce significant bending.

Furthermore, an analysis the distribution of recognition sites for these enzymes on our DNA template shows that there is a very low probability of having one or more sites between two nearby sites. Thus, it appears that in the case of FokI we can rule out protein induced DNA kinking inside loops as a mechanism.

In cases where binding of the enzyme at the recognition site did induce kinks, it is possible that these kinks at the base of the loop could have a similar effect on facilitating short loops as a kink at the apex. Additionally, to explain the smallest events on the order of 20 to 50 bp one could imagine a complex in which DNA is essentially wrapped across the surfaces of the proteins, akin to how DNA is wrapped on the histone octamer in chromatin. In this scenario the DNA would be in close contact with the protein complex along a substantial fraction of its length, rather than looping freely through the solution. Flexibility of the protein complex is also anticipated to be important in facilitating the formation of such short loops.

4.5.7. Comparison of Loop Sizes with Previous Studies

Only a few previous experimental studies have examined the dependence of loop formation probability on loop size. Of the nine type II REases formation of loops of varying size on linear DNA has only been studied for EcoRII and BspMI. The most extensive comparisons can be made with previous results on EcoRII. Reuter, *et al.*, studied the dependence of cleavage rate on distance for site separations of 5, 10, 21, 31, 63, 73, 191, and 952 bp (64,65). Remarkably, they found that highest activity occurred on the 10 bp template and no activity was observed with separations greater than ~1000 bp. In our experiment the DNA template allowed for

loops ranging in size from ~30 to 16,000 bp and the observed loop sizes (Fig. 4.10) indicated an optimum size smaller than 70 bp. This finding of is thus consistent with that reported by Reuter, *et al.*, although we could not pinpoint the exact value of optimum due to limitations on loop sizes dictated by our DNA template. On the other hand, we detected loops as large as 4.1 kbp, which is significantly longer than what was previously reported to be possible.

We may also compare our results with recent findings on BspMI. Cleavage activity has been observed with site separations ranging from 38 to 2644 bp, however these measurements were done with supercoiled DNA. As looping is theoretically predicted to be quite different between the linear and supercoiled forms (58), is it not appropriate to compare with these data. Looping in linear DNA has been studied using magnetic tweezers and loop sizes ranging from ~90 to 1500 bp were reported (29). However, no statistics on loop sizes were reported in this previous study. In our measurements, we detected loops ranging from ~25 to 4500 bp, and found that the optimum size was ~340 ± 40 bp (Table 4.2.). Interestingly, we also observed a second peak in the size distribution, indicating that formation of loops <100 bp is also favorable, a finding which suggests an alternative binding mode for BspMI.

Our measurements with BpmI, BsgI, Cfr10I, Eco57I, FokI, HpaII, Sau3AI are novel results, as relative rates of formation of different loop sizes have not been reported for these enzymes. Our measurements find large variability in the distribution of loop sizes depending on the protein. This finding underscores that one may not describe protein-mediated DNA looping by considering only DNA mechanics. As emphasized by Zhang, *et al.*, detailed models will have to consider the specific

structural details such as loop geometry, sequence dependent DNA bending, protein induced DNA bending, protein span, and protein elasticity. At present, development of more detailed theories will have to await experimental determination of structural information for these enzymes. Conversely, however, it should be possible to use our data to place constraints on the possible structures of looped complexes.

4.6. Conclusion

In summary, protein induced DNA looping was studied via optical tweezers manipulation of single DNA molecules. A wide number of different two site REases were examined. Stable DNA looping was observed in many cases in solutions containing Ca^{2+}. Forced unlooping measurements permitted quantification of rates of looping and unbinding force and loop size distributions. Distinct differences in behavior were observed with different enzymes. Heterogeneous behavior, such as bimodal force and length distributions, were also observed within the behavior of certain enzymes. In many cases loops substantially shorter than the classical DNA persistence length were prevalent. Such loops appear to necessitate models that include effects such as protein induced DNA kinking, bridging, and wrapping.

Table 4.1. Properties of the REases as reported in REBASE. Type IIE REases bind at two sites, but only one is cleaved, whereas type IIF cleave at both binding sites. Type IIG have restriction and modification activities in the same subunit and Type IIS enzymes recognize asymmetric sequences and cleave at least one strand outside of the recognition sequence. Entries with an asterisk are those for which no information was available. For the "Previous Evidence" column, 'cis' refers to a bulk assay in which a template with two sites was cut more efficiently than a template with only one site, 'trans' refers to a bulk assay in which cleavage was stimulated by the addition of an oligo with the recognition site, 'MT' refers to magnetic tweezers, and 'EM' refers to electron microscopy, and 'OT' refers to optical tweezers. In the "Loops?" column, 'N/C' means no observed cutting, and "NS" means non-specific looping. In the "Recognition Sequence" column, 'N' means any nucleotide, 'R' means A or G, 'Y' means C or T, and 'W' means A or T.

Enzyme	Prev. Evid.	Loops?	Type	MW (kD)	Form in Solution	Active Complex	Recognition Sequence
BfiI	cis	N/C	IIS	40	dimer	dimer	ACTGGG (5/4)
BpmI	cis	Yes	IIE, G, S	117	*	*	CTGGAG$(N)_{16}$↓
BsgI	cis	Yes	IIE, G, S	121	*	*	GTGCAG$(N)_{16}$↓
BspMI	MT	Yes	IIE	222	tetramer	tetramer	ACCTGC$(N)_4$↓
Cfr9I	trans	No	IIE	37	*	*	C↓CCGGG
Cfr10I	EM	Yes	IIF, P	320	tetramer	tetramer	R↓CCGGY
Eco57I	trans	Yes	IIE, G	117	monomer	*	CTGAAG$(N)_{16}$↓
EcoRII	EM	Yes	IIE, P	92	dimer	dimer	↓CCWGG
FokI	trans	Yes	IIS	66	monomer	dimer	GGATG$(N)_9$↓
HpaII	trans	Yes	IIE	41	*	*	C↓CGG
Ksp632I	*	No, N/C	IIE	*	*	*	CTCTTC (1/4)
MboII	cis	No	IIS	49	monomer	dimer	GAAGA (8/7)
NarI	OT	NS	IIE	*	*	*	GG↓CGCC
SacII	trans	N/C	IIE	*	*	*	CCGC↓GG
Sau3AI	EM	Yes	IIE	56	monomer	dimer	↓GATC
SgrAI	cis	NS	IIP	38	dimer	tetramer	CR↓CCGGYG

Table 4.2. Experimental conditions and results. N_{site} is the number of recognition sites on the DNA template; C is the enzyme concentration in units/ml; N_P is the number of pairs of sites that can form loops on the DNA template; N/N_P is the number of observed loops per molecule normalized by N_P and standard error; <F> is the mean loop disruption force and standard deviation in pN; <L> is the mean observed loop size in basepairs and standard deviation for the normalized distributions (see text); L_S is the shortest observed event size in the ensemble of data; L_L is the longest observed event size; ρ_S is the correlation coefficient between disruption force and event size for events <150 bp and the number in parentheses is the standard deviation expected for uncorrelated data. In the "DNA" column, L refers to the ½-λ-L template discussed in the text, while A and B refer to the respective varieties of the pBAC template.

Enz.	**C U/ml**	**DNA**	**N_{site}**	**N_P**	**N/N_P ×100**	**<F>**	**<L>**	**L_S**	**L_L**	**ρ_S %**
BpmI	20	L	17	136	2.3 (0.42)	29.6 (11.2)	777 (208)	60	3634	18 (12)
BsgI	30	L	31	465	1.7 (0.22)	26.9 (12.3)	645 (83)	45	3344	-3.2 (12)
BspMI	20	L	24	276	0.91 (0.14)	23.2 (10.3)	613 (124)	25	4494	-16 (9.0)
Cfr10I	100	L	56	1540	1.2 (0.052)	22.5 (12.4)	470 (89)	22	7444	-23 (5.9)
Eco57I	50	L	25	300	2.7 (0.19)	21.8 (11.2)	573 (103)	19	7512	-6.7 (11)
EcoRII	50	L	36	630	0.29 (0.026)	25.6 (12.2)	665 (176)	21	6940	-24 (10)
FokI	40	B	26	325	0.79 (0.038)	14.8 (11.0)	301 (61)	18	4120	-22 (5.6)
HpaII	100	B	49	1176	0.51 (0.041)	16.6 (9.92)	325 (65)	20	2500	-23 (6.9)
Sau3AI	40	A	55	1485	0.66 (0.041)	25.0 (10.8)	423 (85)	7	2682	-11 (5.1)

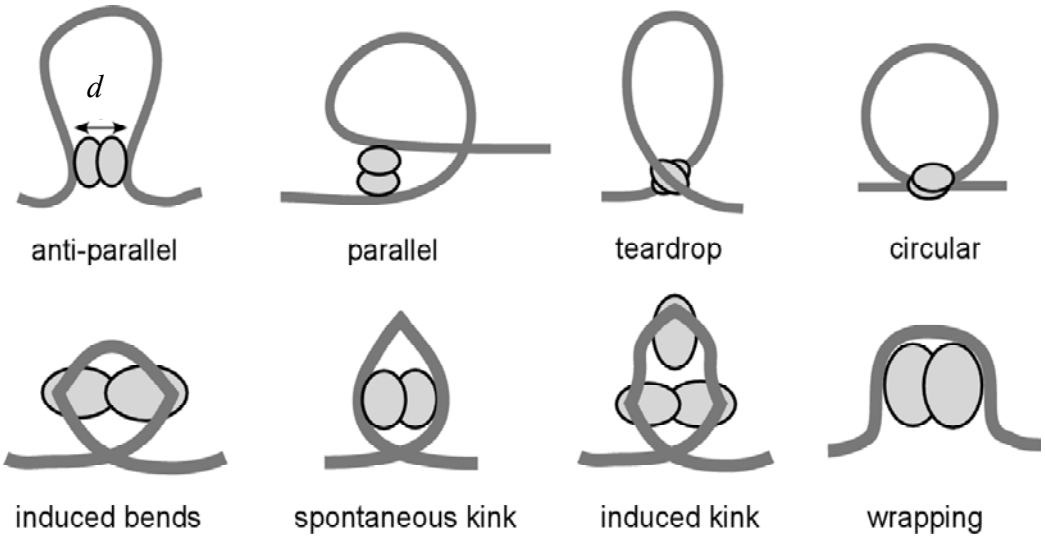

Fig. 4.1. Schematic illustrations of various possible looping geometries. The parameter *d* indicates the span of the protein.

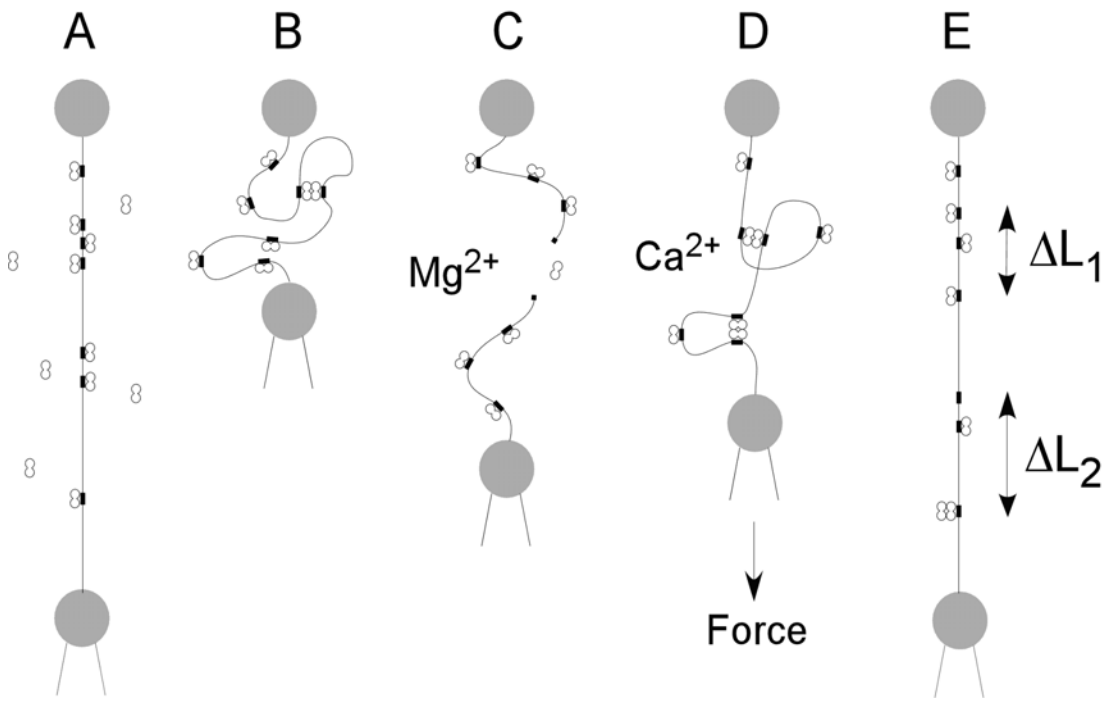

Fig. 4.2. Schematic of the DNA looping measurements. The DNA is tethered between two beads, one manipulated by optical tweezers and the other manipulated by a piezoelectrically positioned micropipette. (A) The DNA is held taut while a solution containing enzyme is introduced. (B) The molecule is then relaxed for a prescribed incubation time, during which the active complex may form via DNA looping. (C) In the presence of Mg^{2+} ions, the DNA molecule is generally cleaved. (D) When Ca^{2+} is substituted for Mg^{2+} cleavage is blocked and stable DNA looping is detected. (E) Upon separating the beads, the looped complexes are disrupted, causing sudden increases in the tether length ΔL and drops in the DNA tension.

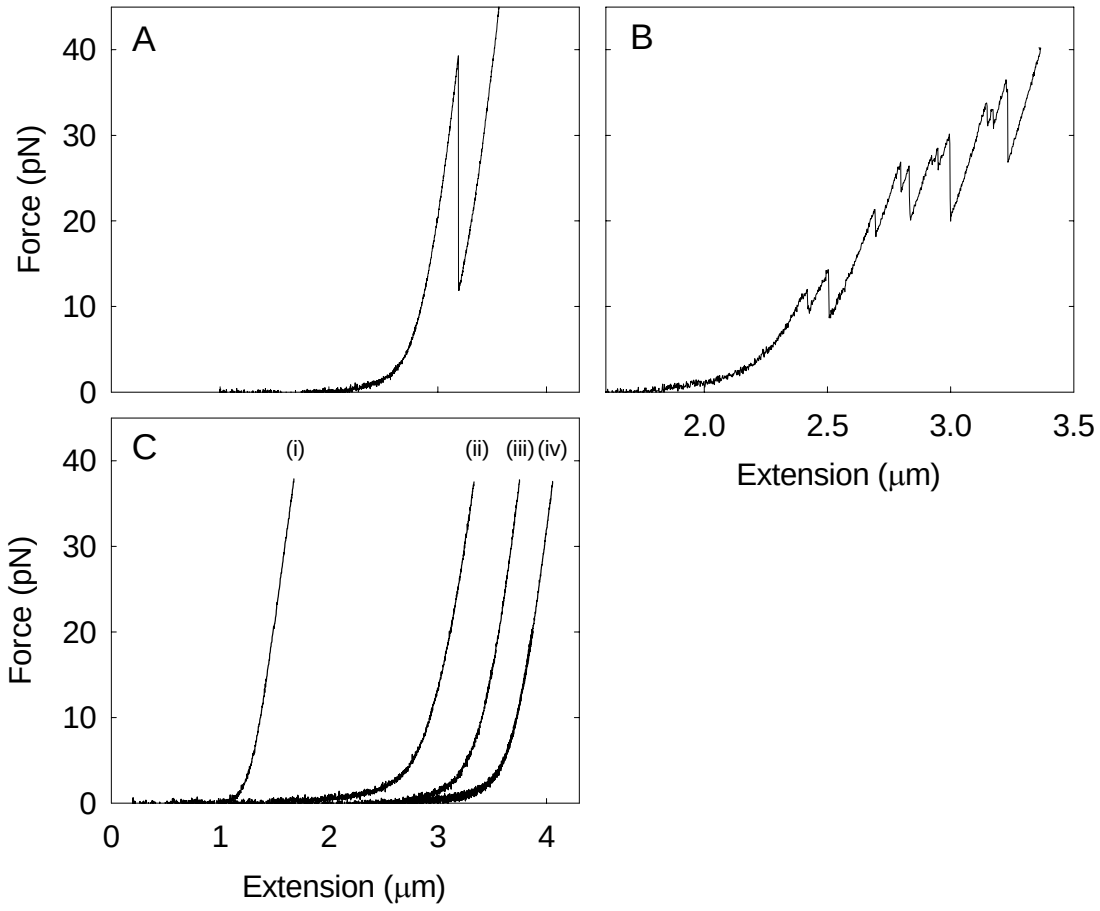

Fig. 4.3. Typical DNA force-extension plots. (A) The two site enzyme BsgI tested on the pBAC template, which has four recognition sites. The one detected loop had a measured length of 946 bp, in excellent agreement with the separation between the binding sites at positions 8342 and 9287 bp (a distance of 945 bp). (B) Sau3AI on the pBAC-A template, which has 55 recognition sites. (C) Control experiments: (i) the two site enzyme Sau3AI was tested on bacteriophage phiX174 DNA, which contains no copies of its recognition sequence; (ii) the one site enzyme HaeIII was tested on pBAC-B DNA containing 36 copies of its recognition sequence; (iii) the two site enzyme SfiI was tested on pBAC-B DNA containing one copy of its recognition sequence; and (iv) pBAC-B DNA, containing 26 FokI binding sites, incubated in the FokI reaction buffer, but no FokI added.

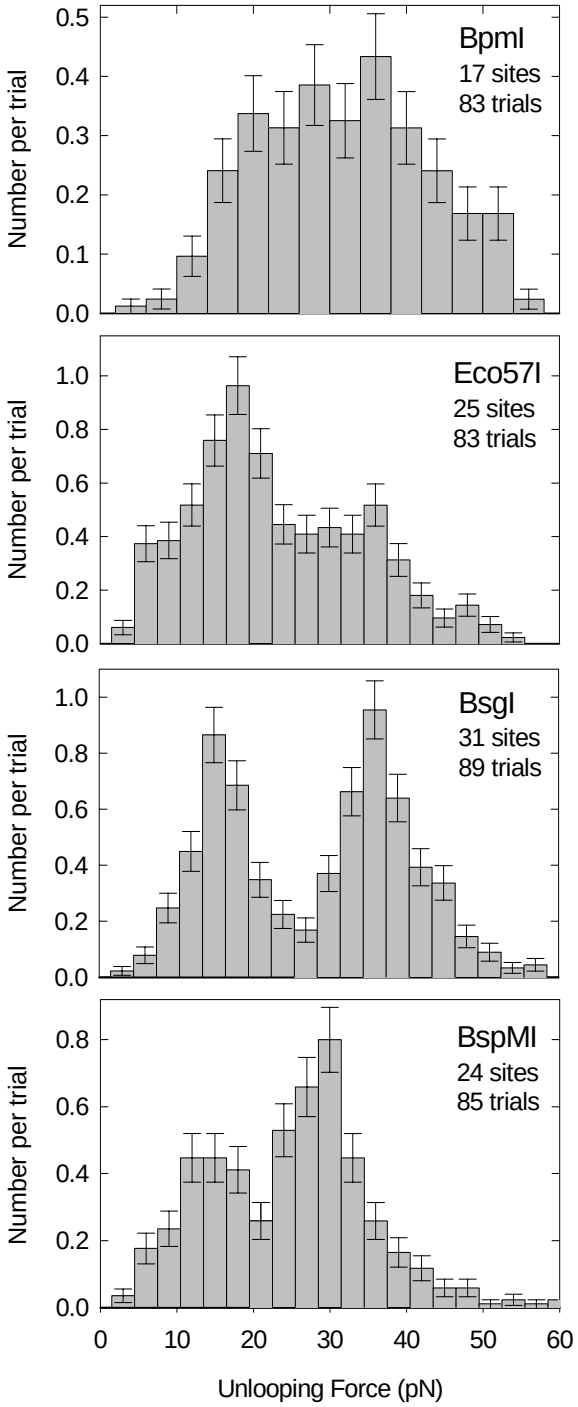

Fig. 4.4. Histograms of the measured forces required to disrupt loops for different two site enzymes. Number of events is reported as number of events per trial. So as not to low bias the force distributions, trials in which the tether broke before reaching 40 pN were not included. The number of recorded events ranged from 256 to 1330 depending on the enzyme.

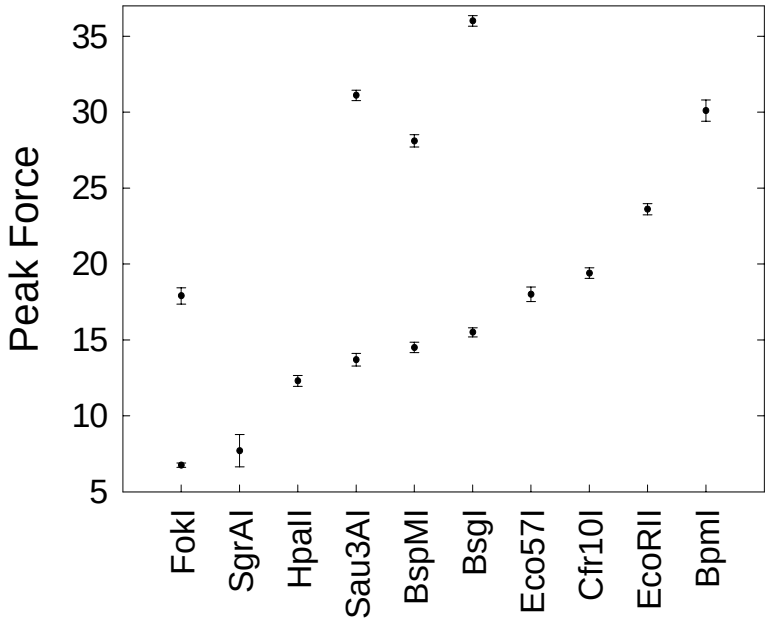

Fig. 4.5. Most probable loop disruption force for each enzyme. For the bimodal distributions, values for both the high and low peaks are plotted. The error bars report the standard error.

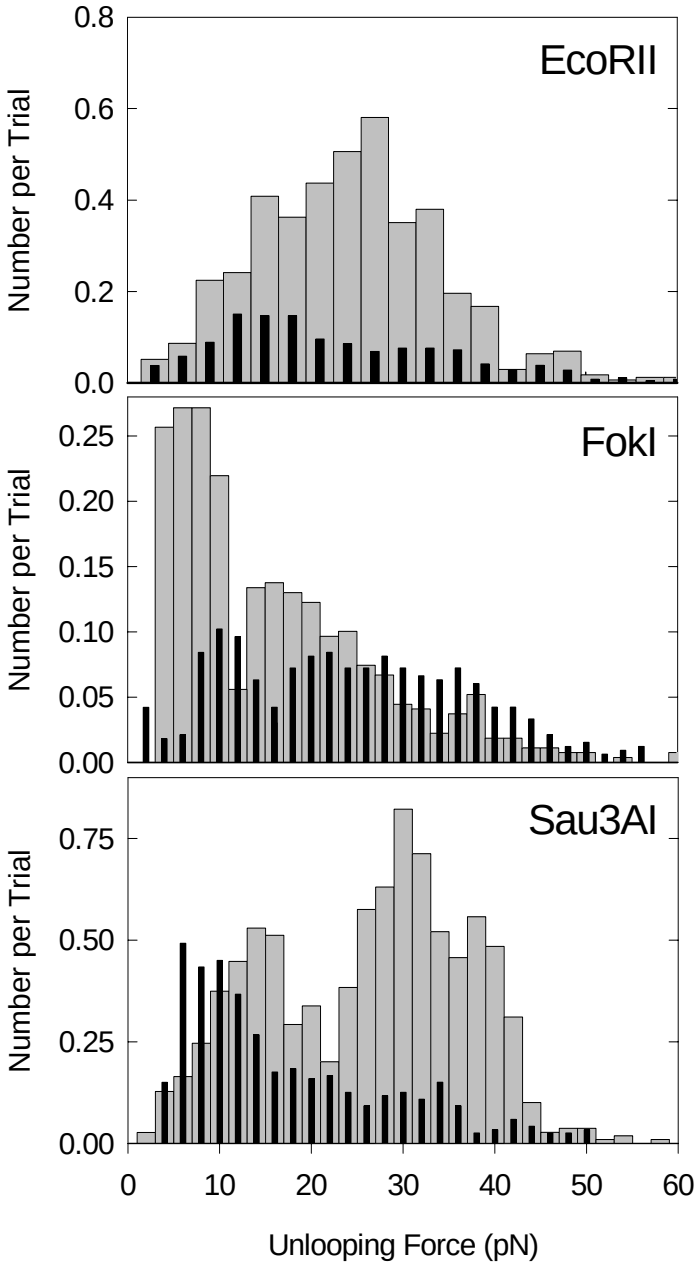

Fig. 4.6. Distribution of forces at which loop disruption occurred in buffers containing Ca^{2+} (gray bars) versus no divalent cations and 1 mM EDTA added (black bars).

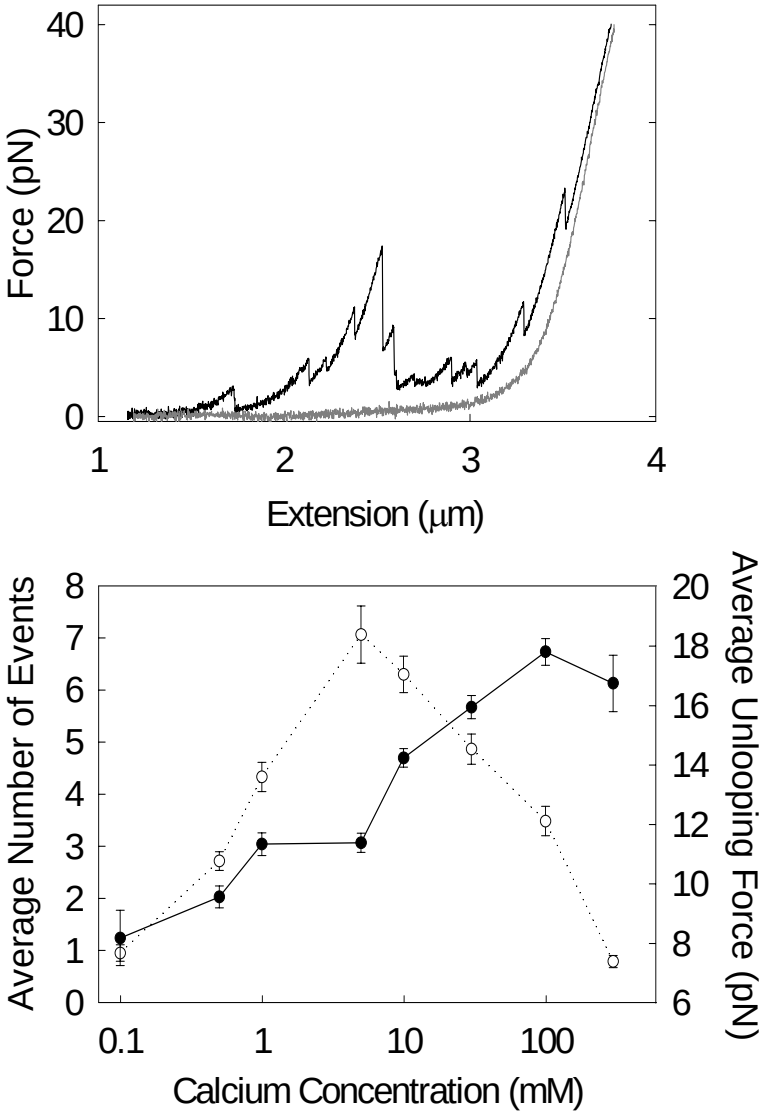

Fig. 4.7. Dependence of DNA looping by HpaII on Ca^{2+}. (Top) Typical force-extension curves with 5 mM Ca^{2+} (black line, having many peaks) and with no divalent cations and 1 mM EDTA added (gray line, no peaks). (Bottom) Mean number of loops (open circles, dashed line, left vertical axis) and mean loop disruption force (filled circles, solid line, right vertical axis) versus Ca^{2+} concentration.

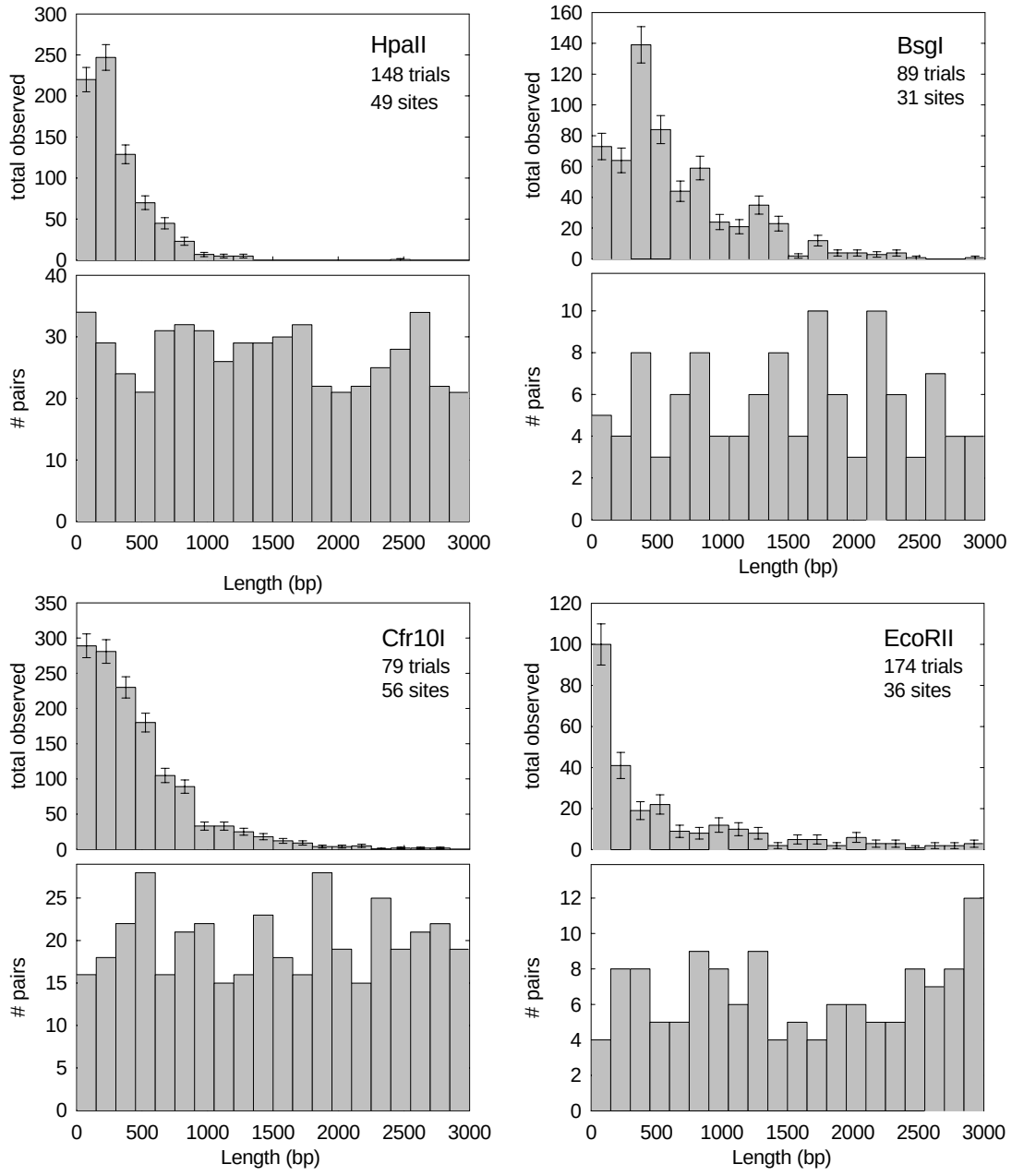

Fig. 4.8. Histograms of observed loop sizes formed with various two site enzymes. Below each graph is a histogram of the possible loop sizes plotted over the same range. The bin size is equal to 150 bp, which is the approximate persistence length of DNA.

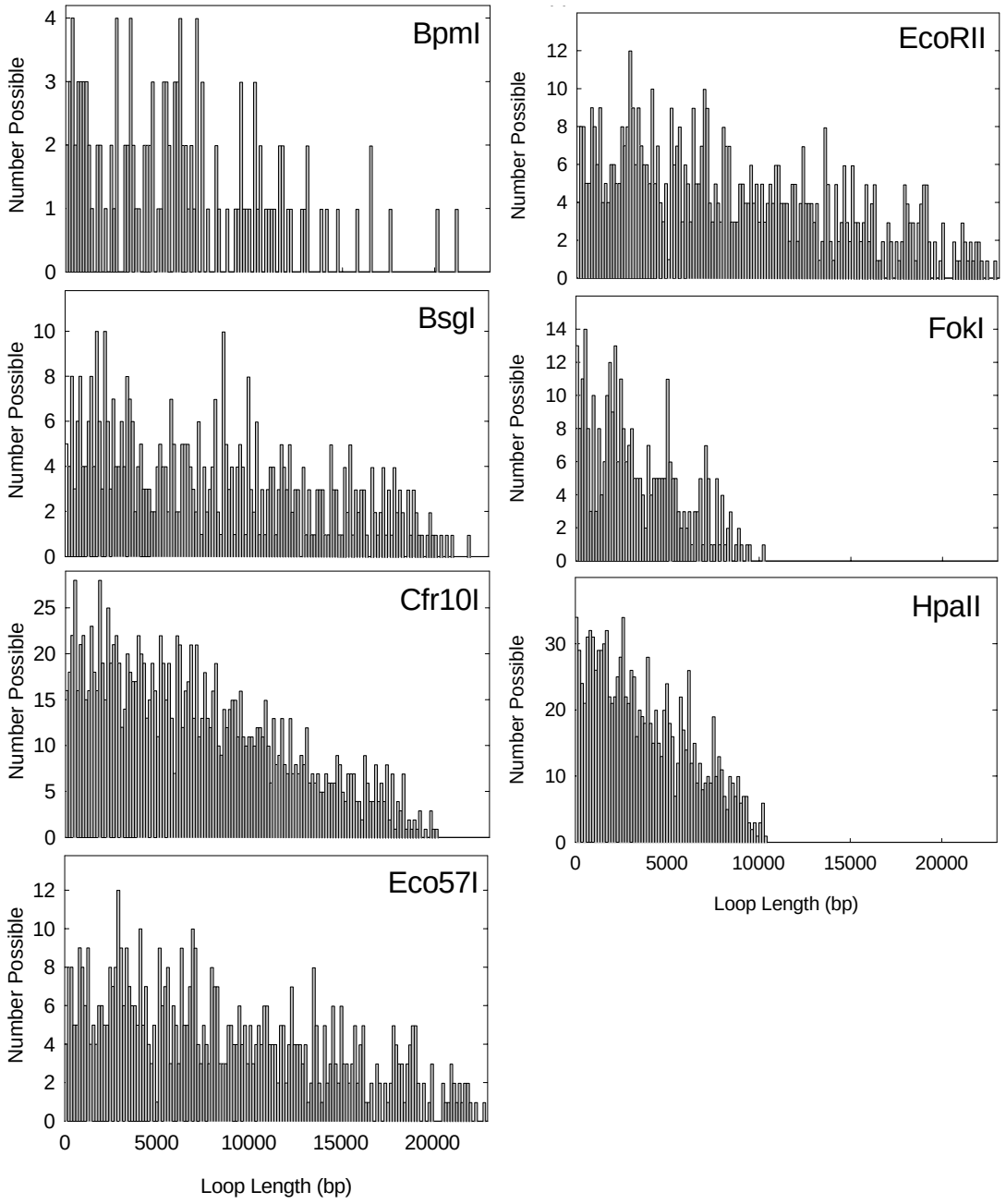

Fig. 4.9. Histograms of possible loop sizes that could be formed by different two site enzymes computed by considering all possible distances between pairs of binding sites in the DNA template.

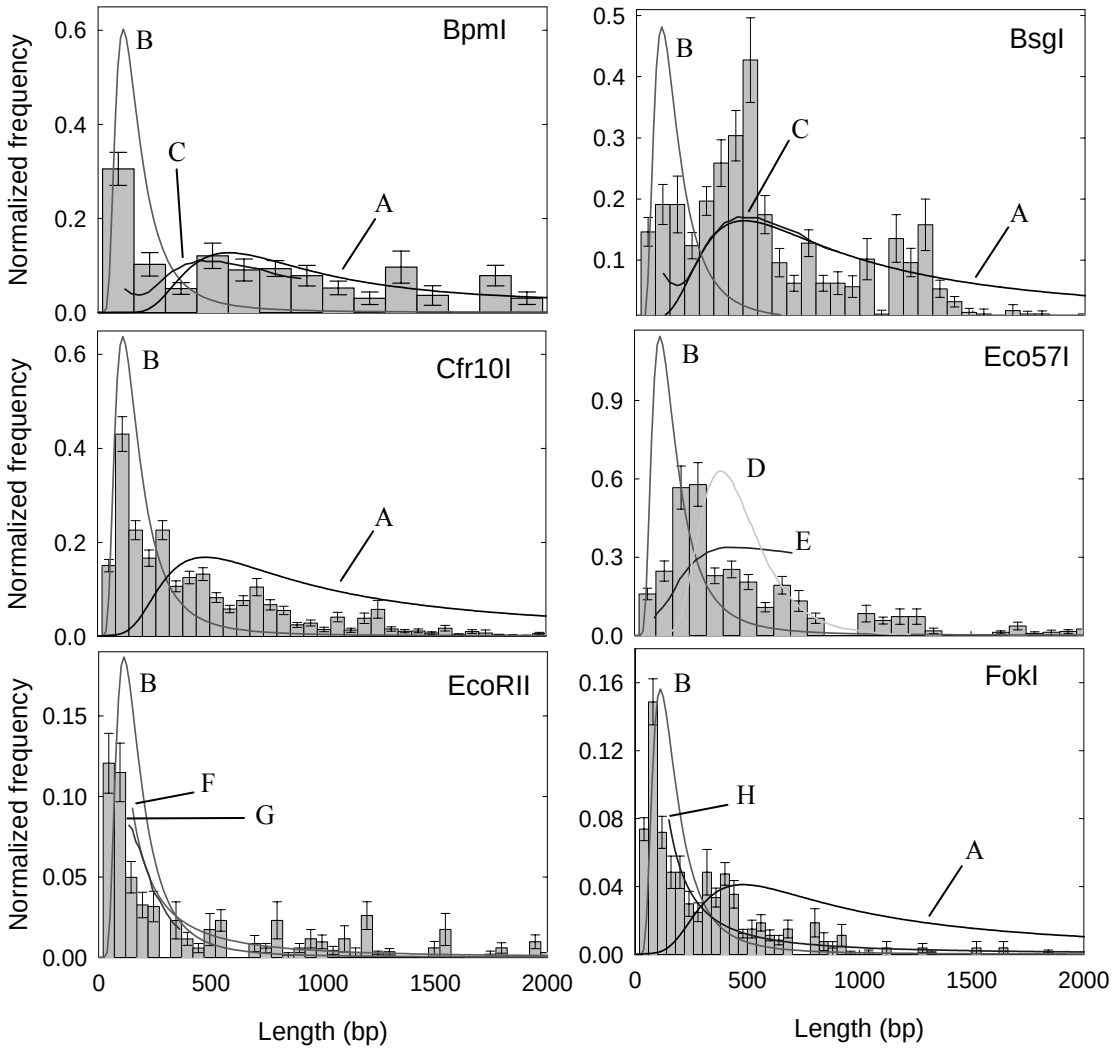

Fig. 4.10. Normalized distributions of observed lengths of unlooping events following incubation of DNA with enzyme solution for 2 minutes. To normalize for differences in DNA templates, histograms of the number of events (per molecule) in each length bin were divided by the number of pairs of sites on the DNA template in the bin. The solid lines are comparisons with theoretical distributions taken from the following references: Ref. 57 with r=10 nm (A); Ref. 14, 90° kink (B); Ref. 15 with hinge m=11 and free ends (C); Ref. 12 with $\beta\nu$=0 and $\beta\varepsilon$=15 (D); Ref. 16 90° kink with p=0.002 (E); Ref. 10 120° curvature with r=10 nm (F); Ref. 15 90° kink and hinge with m=11 and free ends (G); Ref. 10 90° curvature with r=10 nm (H). The theoretical predictions were normalized so that their integrated area over the range given in the aforementioned references was equal to the area under the corresponding data (see text).

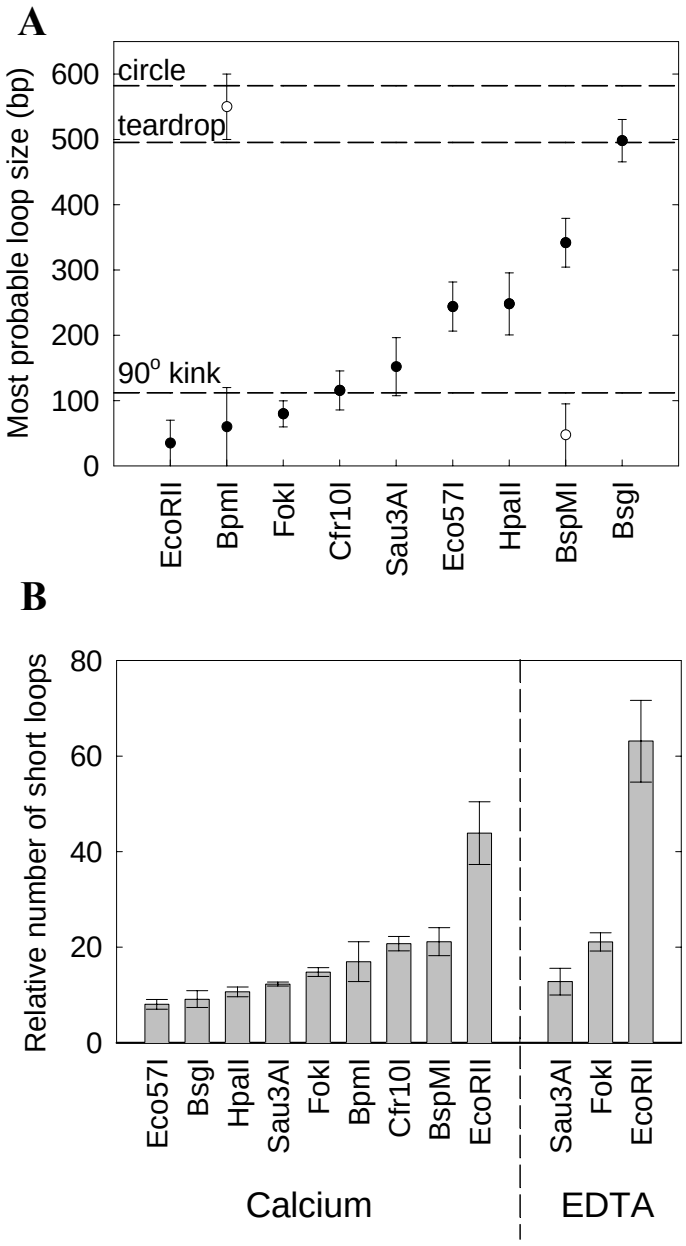

Fig. 4.11. Loop sizes data. (A) Most probable loop sizes for each enzyme. For those distributions with two discernible peaks, the smaller peak is plotted as an empty circle. The dashed lines in this plot indicate the most probable length predicted by the circle, teardrop, and 90° kinked models in Ref 14. The error bars on EcoRII and BpmI extend to zero size because peaks occurred in the first bin of the histograms. In these cases we were not limited by the resolution of our instrument, but by the available sites on the DNA. (B) Relative tendencies for short loop formation by different enzymes. The relative number of short loops was calculated as the number of loops shorter than one persistence length per number possible divided by total number observed per total possible.

References for Chapter 4

1. Schleif, R. (1992) DNA looping. *Annu. Rev. Biochem.* **61**: 199-223.

2. Dunn, T.M., Hahn, S., Ogden, S. & Schleif, R.F. (1984) An operator at -280 base pairs that is required for repression of araBAD operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl. Acad. Sci. U. S. A.* **81**: 5017-5020.

3. Mukherjee, S., Erickson, H. & Bastia, D. (1988) Enhancer Origin Interaction in Plasmid-R6K Involves a DNA Loop Mediated by Initiator Protein. *Cell* **52**: 375-383.

4. Su, W., Middleton, T., Sugden, B. & Echols, H. (1991) DNA looping between the origin of replication of Epstein-Barr virus and its enhancer site: stabilization of an origin complex with Epstein-Barr nuclear antigen 1. *Proc Natl Acad Sci U S A*, **88**, 10870-10874.

5. Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B. & Muller-Hill, B. (1994) Quality and position of the three lac operators of E. coli define efficiency of repression. *EMBO. J.* **13**: 3348-3355.

6. Weiner, B.M. & Kleckner, N. (1994) Chromosome Pairing Via Multiple Interstitial Interactions before and during Meiosis in Yeast. *Cell* **77**: 977-991.

7. Blackwood, E.M. & Kadonaga, J.T. (1998) Going the distance: A current view of enhancer action. *Science* **281**: 60-63.

8. Chen, Y. & Rice, P.A. (2003) New insight into site-specific recombination from Flp recombinase-DNA structures. *Annu Rev Biophys. Biomol. Struct*. **32**: 135-159.

9. Ogata, K., Sato, K. & Tahirov, T.H. (2003) Eukaryotic transcriptional regulatory complexes: cooperativity from near and afar. *Curr. Opin. Struct. Biol.* **13**: 40-48.

10. Rippe, K., von Hippel, P.H. & Langowski, J. (1995) Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem. Sci*. **20**: 500-506.

11. Merlitz, H., Rippe, K., Klenin, K.V. & Langowski, J. (1998) Looping dynamics of linear DNA molecules and the effect of DNA curvature: a study by Brownian dynamics simulation. *Biophys. J.* **74**: 773-779.

12. Sankararaman, S. & Marko, J.F. (2005) Entropic compression of interacting DNA loops. *Phys. Rev. Lett*. **95**: 078104.

13. Blumberg, S., Tkachenko, A.V. & Meiners, J.C. (2005) Disruption of protein-mediated DNA looping by tension in the substrate DNA. *Biophys. J*. **88**: 1692-1701.

14. Sankararaman, S. & Marko, J.F. (2005) Formation of loops in DNA under tension. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys*. **71**: 021911.

15. Yan, J., Kawamura, R. & Marko, J.F. (2005) Statistics of loop formation along double helix DNAs. *Phys. Rev. E* **71**: 061905.

16. Du, Q., Smith, C., Shiffeldrim, N., Vologodskaia, M. & Vologodskii, A. (2005) Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proc. Natl. Acad. Sci. U. S. A.* **102**: 5397-5402.

17. Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. & Zhurkin, V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U. S. A.* **95**: 11163-11168.

18. Zhang, Y.L. & Crothers, D.M. (2003) Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization. *Biophys. J.* **84**: 136-153.

19. Zhang, Y.L., McEwen, A., Crothers, D.M. & Levene, S. (in press, 2006) Statistical-mechanical theory of DNA looping. *Biophys. J.*

20. Halford, S.E., Welsh, A.J. & Szczelkun, M.D. (2004) Enzyme-mediated DNA looping. *Annu. Rev. Biophys. Biomol. Struct.* **33**: 1-24.

21. Topal, M.D., Thresher, R.J., Conrad, M. & Griffith, J. (1991) NaeI endonuclease binding to pBR322 DNA induces looping. *Biochemistry* **30**: 2006-2010.

22. Siksnys, V., Skirgaila, R., Sasnauskas, G., Urbanke, C., Cherny, D., Grazulis, S. & Huber, R. (1999) The Cfr10I restriction enzyme is functional as a tetramer. *J. Mol. Biol*. **291**: 1105-1118.

23. Mucke, M., Lurz, R., Mackeldanz, P., Behlke, J., Kruger, D.H. & Reuter, M. (2000) Imaging DNA loops induced by restriction endonuclease EcoRII. A single amino acid substitution uncouples target recognition from cooperative DNA interaction and cleavage. *J. Biol. Chem*. **275**: 30631-30637.

24. Friedhoff, P., Lurz, R., Luder, G. & Pingoud, A. (2001) Sau3AI, a monomeric type II restriction endonuclease that dimerizes on the DNA and thereby induces DNA loops. *J. Biol. Chem*. **276**: 23581-23588.

25. Szczelkun, M.D. & Halford, S.E. (1996) Recombination by resolvase to analyse DNA communications by the SfiI restriction endonuclease. *EMBO. J.* **15**: 1460-1469.

26. Milsom, S., Halford, S.E., Embleton, M. & Szczelkun, M.D. (2001) Analysis of DNA looping interactions by type II restriction enzymes that require two copies of their recognition sites. *J. Mol. Biol.* **311**: 515-527.

27. Watson, M., Gowers, D.M. & Halford, S.E. (2000) Alternative geometries of DNA looping: an analysis using the SfiI endonuclease. *J. Mol*. *Biol.* **298**: 461-475.

28. Katiliene, Z., Katilius, E. & Woodbury, N.W. (2003) Single molecule detection of DNA looping by NgoMIV restriction endonuclease. *Biophys. J.* **84**: 4053-4061.

29. Yan, J., Skoko, D. & Marko, J.F. (2004) Near-field-magnetic-tweezer manipulation of single DNA molecules. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **70**: 011905.

30. van den Broek, B., Vanzi, F., Normanno, D., Pavone, F.S. & Wuite, G.J. (2006) Real-time observation of DNA looping dynamics of Type IIE restriction enzymes NaeI and NarI. *Nuc. Acids Res*. **34**, 167-174.

31. Kruger, D.H., Barcak, G.J., Reuter, M. & Smith, H.O. (1988) EcoRII can be activated to cleave refractory DNA recognition sites. *Nuc. Acids Res*. **16**, 3997-4008.

32. Bath, A.J., Milsom, S.E., Gormley, N.A. & Halford, S.E. (2002) Many type IIs restriction endonucleases interact with two recognition sites before cleaving DNA. *J. Biol. Chem.* **277**: 4024-4033.

33. Soundararajan, M., Chang, Z., Morgan, R.D., Heslop, P. & Connolly, B.A. (2002) DNA binding and recognition by the IIs restriction endonuclease MboII. *J. Biol. Chem.* **277**: 887-895.

34. Gowers, D.M., Bellamy, S.R. & Halford, S.E. (2004) One recognition sequence, seven restriction enzymes, five reaction mechanisms. *Nuc. Acids Res*. **32**: 3469-3479.

35. Reuter, M., Kupper, D., Pein, C.D., Petrusyte, M., Siksnys, V., Frey, B. & Kruger, D.H. (1993) Use of specific oligonucleotide duplexes to stimulate cleavage of refractory DNA sites by restriction endonucleases. *Anal. Biochem.* **209**: 232-237.

36. Kupper, D., Moncke-Buchner, E., Reuter, M. & Kruger, D.H. (1999) Oligonucleotide stimulators allow complete cleavage of agarose-embedded DNA by particular type II restriction endonucleases. *Anal. Biochem*. **272**: 275-277.

37. Vanamee, E.S., Santagata, S. & Aggarwal, A.K. (2001) FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.* **309**: 69-78.

38. Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S., Dryden, D.T., Dybvig, K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nuc. Acids Res*. **31**: 1805-1812.

39. Bustamante, C., Smith, S.B., Liphardt, J. & Smith, D. (2000) Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* **10**: 279-285.

40. Wood, K.M., Daniels, L.E. & Halford, S.E. (2005) Long-range communications between DNA sites by the dimeric restriction endonuclease SgrAI. *J. Mol. Biol.* **350**: 240-253.

41. Lagunavicius, A., Sasnauskas, G., Halford, S.E. & Siksnys, V. (2003) The metal-independent type IIs restriction enzyme BfiI is a dimer that binds two DNA sites but has only one catalytic centre. *J. Mol. Biol.* **326**: 1051-1064.

42. Gemmen, G.J., Sim, R., Haushalter, K.A., Ke, P.C., Kadonaga, J.T. & Smith, D.E. (2005) Forced unraveling of nucleosomes assembled on heterogeneous DNA using core histones, NAP-1, and ACF. *J. Mol. Biol.* **351**: 89-99.

43. Derenyi, I., Bartolo, D. & Ajdari, A. (2004) Effects of intermediate bound states in dynamic force spectroscopy. *Biophys. J.* **86**: 1263-1269.

44. Pingoud, A., Fuxreiter, M., Pingoud, V. & Wende, W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life. Sci.* **62**: 685-707.

45. Bilcock, D.T., Daniels, L.E., Bath, A.J. & Halford, S.E. (1999) Reactions of type II restriction endonucleases with 8-base pair recognition sites. *J. Biol. Chem.* **274**: 36379-36386.

46. Evans, E. & Williams, P. (2004) *Dynamic Force Spectroscopy*. Springer Verlag, Berlin.

47. Skirgaila, R., Grazulis, S., Bozic, D., Huber, R. & Siksnys, V. (1998) Structure-based redesign of the catalytic/metal binding site of Cfr10I restriction endonuclease reveals importance of spatial rather than sequence conservation of active centre residues. *J. Mol. Biol*. **279**: 473-481.

48. Waugh, D.S. & Sauer, R.T. (1993) Single amino acid substitutions uncouple the DNA binding and strand scission activities of Fok I endonuclease. *Proc. Natl. Acad. Sci. U. S. A*. **90**: 9596-9600.

49. Wah, D.A., Bitinaite, J., Schildkraut, I. & Aggarwal, A.K. (1998) Structure of FokI has implications for DNA cleavage. *Proc. Natl. Acad. Sci. U. S. A*. **95**: 10564-10569.

50. Huai, Q., Colandene, J.D., Topal, M.D. & Ke, H.M. (2001) Structure of NaeI-DNA complex reveals dual-mode DNA recognition and complete dimer rearrangement. *Nat. Struct. Biol.* **8**: 665-669.

51. Nevo, R., Stroh, C., Kienberger, F., Kaftan, D., Brumfeld, V., Elbaum, M., Reich, Z. & Hinterdorfer, P. (2003) A molecular switch between alternative conformational states in the complex of Ran and importin beta 1. *Nat. Struct. Biol.* **10**: 553-557.

52. Bowen, L.M. & Dupureur, C.M. (2003) Investigation of restriction enzyme cofactor requirements: A relationship between metal ion properties and sequence specificity. *Biochemistry* **42**: 12643-12653.

53. Bustamante, C., Marko, J.F., Siggia, E.D. & Smith, S. (1994) Entropic Elasticity of Lambda-Phage DNA. *Science* **265**: 1599-1600.

54. Vologodskii, A. (1994) DNA Extension under the Action of an External Force. *Macromolecules* **27**: 5623-5625.

55. Wang, M.D., Yin, H., Landick, R., Gelles, J. & Block, S.M. (1997) Stretching DNA with optical tweezers. *Biophys. J.* **72**: 1335-1346.

56. Bouchiat, C., Wang, M.D., Allemand, J.F., Strick, T., Block, S.M. & Croquette, V. (1999) Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.* **76**: 409-413.

57. Ringrose, L., Chabanis, S., Angrand, P.O., Woodroofe, C. & Stewart, A.F. (1999) Quantitative comparison of DNA looping in vitro and in vivo:

chromatin increases effective DNA flexibility at short distances. *EMBO. J.* **18**: 6630-6641.

58. Rippe, K. (2001) Making contacts on a nucleic acid polymer. *Trends in Biochemical Sciences* **26**: 733-740.

59. Cloutier, T.E. & Widom, J. (2004) Spontaneous sharp bending of double-stranded DNA. *Molecular Cell* **14**: 355-362.

60. Cloutier, T.E. & Widom, J. (2005) DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. *Proc. Natl. Acad. Sci. U. S. A*. **102**: 3645-3650.

61. Yan, J. & Marko, J.F. (2004) Localized single-stranded bubble mechanism for cyclization of short double helix DNA. *Phys. Rev. Lett.* **93**: 108108.

62. Wiggins, P.A., Phillips, R. & Nelson, P.C. (2005) Exact theory of kinkable elastic polymers. *Phys. Rev. E* **71**: 021909.

63. Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K. & Wilson, K.S. (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO. J*. **12**: 1781-1795.

64. Reuter, M., Kupper, D., Meisel, A., Schroeder, C. & Kruger, D.H. (1998) Cooperative binding properties of restriction endonuclease EcoRII with DNA recognition sites. *J. Biol. Chem*. **273**: 8294-8300.

65. Kruger, D.H., Kupper, D., Meisel, A., Tierlich, M., Reuter, M. & Schroeder, C. (1995) Restriction endonucleases functionally interacting with two DNA sites. *Gene* **157**: 165.

Chapter 4 Acknowledgement

The text of Chapter 4, in full, appears in Gemmen, G.J., Millin, R. & Smith, D.E. (2006). DNA Looping by Two-Site Restriction Endonucleases: Heterogeneous Probability Distributions for Loop Size and Unbinding Force. Submitted to *Nucleic Acids Research*. The dissertation author was a primary researcher and secondary author for the research which forms the basis of this chapter

Chapter 5. Dynamics of Single DNA Looping and Cleavage by Restriction Endonuclease Sau3AI Measured with Optical Tweezers

5.1. Abstract

We use optical tweezers to measure single molecule DNA cleavage and looping by the restriction enzyme Sau3AI, which requires interaction at two sites via looping. An applied tension of ~1 pN inhibits cleavage completely, whereas 0.03 pN has little effect. Cleavage events are detected in as little as 5 s, and nearly all molecules are cleaved within 5 min. Substituting Ca^{2+} for Mg^{2+} blocks the cleavage, allowing for the detection and forced disruption of the enzyme-stabilized loops. Distributions of loop number and length are measured versus time and tension. The loop number distributions grow increasingly non-Poisson as looping increases, suggesting cooperativity in loop formation. Both the loop number and average length saturate with time, but decrease dramatically with increasing tension. Short loops are significantly less tension-inhibited than long loops. A large number of loops shorter than expected from DNA flexibility theory indicates that protein bridging must be considered and that protein induced and/or spontaneous DNA kinking may occur. Accordingly, our results agree more so with recent theories of tensioned DNA looping that invoke such mechanisms. Our results also support the notion that looping may be an effective mechanism for tension-dependent genetic regulation.

5.2. Introduction

5.2.1. How Nature Utilizes DNA Looping

Regulating a cellular process requires that the appropriate signals be given to the appropriate proteins at the appropriate times. With a regulatory protein bound to

its recognition sequence on the DNA (e.g., an RNA polymerase bound to a promoter site), only a few other proteins can bind immediately adjacent to it in order to transmit conformational signals to it. However, proteins that regulate the process can bind thousands of base pairs away and be brought into contact via DNA looping. This allows for many more proteins to be involved throughout the different stages of a regulatory process than would be possible if only 'nearest neighbor' interactions were utilized (1). A second reason why nature may utilize DNA looping is the cooperativity inherent in a process that requires two binding sites on the DNA. If a protein is bound to one site, it is much more likely to bind to the second because, as long as it is bound to the first, it will be no farther from the second than the distance between the two sites. This effectively increases the concentration of the protein about the second site, buffering the system from drastic changes in the concentration of any of its components. This phenomenon is called the 'chelate effect' (2). Additionally, this binding cooperativity effectively increases the net affinity that the protein has for the DNA; in order to escape the DNA completely, it must be dissociated from both sites simultaneously (1,3). This allows the binding sites on the DNA to be saturated despite the overall concentration of the protein being rather low, an important point when there may be thousands of regulatory proteins in the cell nucleus. A third reason why nature may use DNA looping is the redundancy in sequence recognition that it effects. If the proteins bound at two DNA sites need to be brought together by looping to be functional, the discrimination against different sequences at one site will be multiplied by the discrimination at the other, providing a much more stringent net discrimination against different sequences than at either site individually (4,5).

5.2.2. Examples of DNA Looping in Nature

Reviewed in Ref. 1, thermodynamic DNA looping plays an important role in many biological processes such as genetic recombination (6,7), DNA replication (8,9), and transcription regulation (10-13). Examples of looping that either promote or repress transcription can be found in both prokaryotes and eukaryotes, involving stretches of DNA from thousands of base pairs long to less than one hundred base pairs long. Examples of 'short loop' systems in *E.coli* include the 113 bp between O_E and O_I in the *gal* operon (14), the 93 bp stretch between the BoxE and BoxB sites in the nagE-B operon (15), and the 92 bp stretch between O1 and O3 of the *lac* operon (11). *In vitro* experiments on the *lac* repressor show that the repression was maximized for a 71 bp loop length and still ample even for a 58 bp loop (16).

A growing number of Type II restriction enzymes that require two recognition sites also rely on DNA looping. Reviewed in Refs. 17-18, such enzymes operate generally by one of two motifs. The first is one in which each enzyme unit can bind the DNA twice, cleaving the DNA only when both binding clefts are occupied. *EcoRII* is a dimer that operates by this motif (19), and *BspMI* is a tetramer that does likewise (20). The second motif by which these 'two site' restriction enzymes function dictates that individual enzyme subunits each bind to a recognition site. These DNA-bound subunits (monomers or dimers) are then brought together by DNA looping, forming the active enzyme multimer (dimer or tetramer) that cleaves the DNA. It is believed that *Sau3AI* is a free monomer in solution with a recognition site of GATC that cleaves the DNA at one site upon dimerization with another DNA-bound monomer (21).

5.2.3. Bulk Methods to Study Looping

Many bulk methods have been utilized to study DNA looping, including helical twists assays, the electrophoresis of looped circles, a DNase digestion assay to assess DNA groove distortion, cooperativity assays, and experiments using interlocked circles of DNA, called *catenanes* (1). Helical twist assays rely on the torsional stress and cyclization of DNA, specifically that the looping depends on relative phase of each strand (22,23). The insertion of 5 bp (half a helical turn) between the two sites on a sufficiently short DNA template leads to a change in activity of the system that relies on the looping, as in the *arabinose* operon in *E.coli* (10). The electrophoresis of looped circles relies on the fact that a supercoiled circle of DNA with protein bridging two sites migrates faster through a gel than a similar circle with no protein bound or with protein bound to only one site (24). The Dnase digestion assay relies on the fact that for DNA trapped in a small loop, the resulting torsional stress distorts the minor groove enough to modulate the activity of minor groove binding Dnase (25,26). Cooperativity assays rely on the chelate effect mentioned above and have been used to study the *lac* operon (27). The catenane experiments have been used to discern between looping through three-space and one-dimensional 'searching' along the DNA for two site restriction enzymes (17,18).

5.2.4. Theoretical Descriptions of DNA Looping

The length of DNA in a loop, the closure geometry of the loop, and any tension on the DNA are three important variables in the theoretical description of looping. Based largely upon the semiflexible wormlike chain model (WLC) (28,29),

classical looping theory states that entropy inhibits the formation of long loops in that the two ends have difficulty finding each other in 3-space (30), while the intrinsic rigidity of DNA makes forming a short loop energetically expensive (31,32). The measure of this rigidity is called the *persistence length* of DNA, defined as the length scale over which tangent vectors to the contour of the DNA remain significantly correlated (i.e., e^{-1}). In the presence of divalent cations (our experiments are done in 10 mM Mg^{2+} or Ca^{2+}), the persistence length of DNA is ~ 40 nm, which is roughly 120 bp (33). The increasing number of systems in nature that rely on forming short loops, such as those in *E.coli* mentioned above, are of particular interest because the classical understanding of DNA mechanics dictates that such loops should not readily form (31,32,34). This apparent contradiction motivated recent cyclization studies of ~100 bp fragments (35). The results indicated that such short loops form orders of magnitude more readily than classically predicted, though others have interpreted these results to be consistent with the WLC after all (36). Nonetheless, the existence of such short loops have motivated the incorporation of DNA defects, such as kinks or hinges, into looping theories. Such defects bring about regions of greatly enhanced flexibility, making the DNA more readily able to overcome the energetic barriers to forming loops less than the persistence length (37,38).

Such defects also play a very important role in the theories of looping under tension (39-41). These theories are most relevant to our optical tweezers studies, because holding the ends of the DNA molecules constrains the fluctuation of the molecule, which exerts an 'entropic tension' on the molecule. These theories all share in common the prediction that although tension inhibits the formation of longer

loops more so than shorter loops due to the extra work needed to pull in the DNA against the applied tension. As will be discussed below, the incorporation of flexible defects in the DNA can greatly enhance looping efficiency under tension, likewise allowing for a population of surprisingly short looping events.

5.2.5. Single Molecule Studies of DNA Looping

A variety of single molecule techniques have also been used to study DNA looping. Electron microscopy experiments have imaged DNA loops for a number of systems, notably for the two site restriction enzymes *EcoRII* (19) and *Sau3AI* (21). These experiments are limited, however, in that non-specific aggregates can form loops, and it is difficult to ascertain loop lengths precisely if the DNA templates have many sites. Fluorescent Resonant Energy Transfer (FRET) has been used to demonstrate loops for the two site restriction enzyme *NgoMIV* (42). Finzi and Gelles measured the Brownian motion of beads tethered to a microscope slide by looped and unlooped DNA to ascertain the lifetime of loops trapped by the lactose repressor protein of *E.coli* (43). Similarly, Lia, *et al.*, used the magnetic tweezers to observe looping on negatively supercoiled DNA utilizing the gal repressor and the HU protein (44). In the optical tweezers, looping was inferred from the observed shortening of the DNA tether due to the binding of transcription factor Sp1 (45). Recently, the opening of loops by the two site enzyme *BspMI* on λ-DNA was observed using a near-field magnetic tweezers, showing loops from 30nm to 500 nm (46). An extensive single molecule study of the statistics of DNA looping as a function of DNA tension and time, however, has not been reported yet.

5.2.6. Advantages of Using Optical Tweezers to Study DNA Looping

Towards the end of an extensive single molecule study of the statistics of DNA looping, the optical tweezers afforded us many advantages. First, DNA looping by thermodynamic fluctuations is fundamentally a statistical process. Using the optical tweezers allowed us to build up statistical distributions of the looping metrics as opposed to the ensemble averages of bulk techniques; there is more information available in the statistical distributions for comparison to theory. Second, the optical tweezers made it possible to inhibit looping completely by exerting moderate tensions (a few piconewtons) on the DNA molecule. This afforded exquisite control over the timing of the experiments that can not be obtained in bulk studies and many other single molecule techniques. Third, in the optical tweezers it was possible to exert an entropic tension (<1 pN) on a DNA molecule because of the constraint imposed by holding the molecule by its ends. This entropic tension we expected to effect looping efficiency significantly. A fourth advantage to studying looping in the optical tweezers was that it allowed us to observe many loops of different lengths simultaneously and precisely measure the length of each loop, ΔL. A final advantage afforded by using the optical tweezers was that the formed loops can be pulled apart upon the application of sufficient force, disrupting the DNA-protein and/or protein-protein bonds that initially trapped the loops. Though not utilized in this study directly, the forces necessary to bring about these disruptions are likewise distributed statistically and contain information about DNA-protein and protein-protein interactions. These disruption force distributions have been measured for a wide variety of two site restriction enzymes, including *Sau3AI*, in a separate study (47).

5.3. Methods

5.3.1. DNA Preparation

The preparation of DNA required the ligation of a 4282 bp PCR-generated DIG body-labeled fragment to a 10845 bp *XhoI*-digested, single biotin labeled, 10845 bp pBAC fragment. To generate the former, 30 μL of 10X PCR buffer (no $MgCl_2$), 18 μL of 25 mM $MgCl_2$, 7.5 μL of 20 pmol/μL primer (5'-GTGGTATGGCTGATTA TGATC-3'), 7.5 μL of 20 pmol/μL primer (5'-GCAGCCTGAATGGCGAATGG-3'), 21.5 μL of dNTP mix (2.5 mM each dATP, dCTP, dGTP, dTTP), 2.4 μL of dNTP Labeling Mix T (2.5 mM each dATP, dCTP, dGTP), 6 μL of 1 mM dUTP-11-DIG, 6 μL of 50 ng/μL pFastBac HT-b (Invitrogen), and 195 μL of H_20 were mixed on ice. 6 μL of 5 U/μL Taq DNA polymerase (Promega) were then added. 100 μL of this PCR mixture were distributed to three thin-walled PCR tubes and run as follows: 94 C for 5 hrs, 94 C for 30 min, 56 C for 30 min, 72 C for 5 hrs (25 cycles), 72 C for 10 hrs, and held at 4 C. The 10,845 bp fragment was made by digesting 3 μL of pBACe3.6 (BAC-PAC-Resources) with 3 μL of 10 U/μL *BsrGI* (New England Biolabs) in 15 μL of 10X NEB Buffer 2, 1.5 μL of 100X BSA, and 127.5 μL of H_20 for 1 hr at 37 C, then 5 min at 70 C, and then cooling to 4 C for 5 min. For the biotin fill-in, 3 μL of dNTP Labeling Mix A (2.5 mM each dCTP, dGTP, dTTP), 4.5 μL of 0.4 mM dATP-14-biotin (Invitrogen), and 1.5 μL of 5 U/μL Klenow, Exo-(NEB) were added and the solution was incubated for 15 min at 37 C. 4 μL of 0.5M EDTA was added to stop the reaction. Both the PCR fragment and the pBACe3.6 were purified by Qiaquick PCR purification kit according to manufacturer's protocol and

eluted with EB to give 150 μL of each. Each was then digested with 3 μL of 20 U/μL *XhoI* (NEB) for 1 hr at 37 C in 17 μL of 10X NEB Buffer 2 and 1.7 μL of 100X BSA. To isolate products of the correct length, the samples were run on a 1% GTG agarose/1X TAE gel, and a Qiaquick Gel Extraction kit was used according to the manufacturer's protocol. Each sample was eluted to 150 μL. The two fragments were then ligated by combining 123 μL of the gel-purified, *XhoI* digested PCR product (0.59 of final volume), 60 μL of the gel-purified, XhoI/BsrGI digested pBACe3.6 (0.29 of final volume), 21 μL of the 10X T4 DNA ligase buffer (NEB), and 6 μL of T4 DNA ligase (400 U/μL) and incubating for 16-18 hrs. The reaction was then stopped by the addition of 9 μL of 0.5 M EDTA. The ligation was verified by analyzing 10 μL aliquots on a 0.8% LE agarose/1X TAE gel, visualized by EtBr. The final concentration of product was 5-10 ng/μL. This was loaded onto prewashed streptavidin beads (Spherotech, 2.0–2.9 μm, 0.5% w/v) at a volume ratio of 1:9 and incubated for ~ 1 hr at 25 C. The DNA-bead ratio was empirically set to yield about one hookup every four streptavidin beads.

5.3.2. Experimental Protocols

The DNA cleavage experiments were performed by hooking up each molecule in 1X PBS and then pulling it to 40 pN to break any non-specific stickiness between the DNA and the polystyrene beads and to establish a contour length, L_0, for that molecule. L_0 was determined by fitting the nearly linear, high force portion of the force-extension curve (~15 pN to 40 pN), and ascertaining where that line crossed the F=0 axis; this point is very nearly $0.95L_0$ according to the exact solution to the

extensible WLC (49). Determining L_0 for every molecule accounts for variation in the tethering of the DNA molecule due to the rotationally constrained bead on the micropipette. With L_0 established, a 100 λ reaction aliquot was flowed into the flow cell while the DNA was held tense (5 pN) to prevent any premature looping. When the flow cell was filled, the enzyme flow was stopped. Care was taken to minimize any residual flows due to the sensitivity of DNA to elongational flows. The reaction aliquot typically consisted of 1λ of stock *Sau3AI* enzyme (4 units/λ, where a unit is the amount of enzyme required to digest 1 μg of λ DNA in 1 hr @ 37C in a 50λ reaction) in 99λ of NEB's *Sau3AI* Reaction buffer (10mM Bis-Tris-Propane-HCl (pH = 7.0 @ 25C), 100mM NaCl, 10 mM $MgCl_2$, 1mM dithiothreitol, 0.1 mg/ml BSA). With the flow cell filled with enzyme and the flow stopped, the molecule was relaxed quickly to an end-to-end distance, x, as to maintain a fractional extension of x/L_0. After an incubation time t_{inc}, the beads were quickly separated (~ 4 μm/s) to ascertain if the molecular tether was still intact, as shown in Fig. 5.1. The activity is then quantified as the percentage of molecules cleaved within t_{inc}.

Looping experiments were carried out in much the same manner as the cutting experiments, but with 10mM $CaCl_2$ instead of $MgCl_2$. As shown schematically in Fig. 5.1, the Ca^{++} in the buffer allows the dimers to form but not cleave the DNA (18). By taking advantage of this property, we can use *Sau3AI* to measure both the number of loops (N) and the lengths of those loops as a function of time and DNA tension, allowing a more detailed comparison with looping theories than DNA cleavage alone would facilitate. After holding each molecule at a prescribed x/L_0 for t_{inc} in the presence of the enzyme, the beads were separated at 150 nm/s. Upon

pulling the beads apart, the force-extension curve showed sudden drops corresponding to the bound *Sau3AI* dimers being pulled apart (or the dimer being pulled from the DNA molecule). Each drop was followed by another rise in the force extension curve as the beads were separated further and the predrop tension was restored to the DNA. The distance between the drop and the point at which an equal tension was subsequently reached corresponds to a sudden increase in tether length, ΔL. This method of analysis was tested to be accurate for simulated data, and can discern length increases as small as ~ 5 bp (48). By repeating numerous times at different t_{inc} and x/L_0, we were able to build statistical distributions of N and ΔL at each experimental condition. These distributions were analyzed phenomenologically; the results of this analysis allow for extrapolation to $x/L_0 = 0$ (i.e., force free), as well as to $t_{inc} = 0$ and $t_{inc} = \infty$, for comparison with theory when appropriate. Looping was not studied as a function of enzyme concentration as it was for cutting. Additionally, control experiments using a labeled φX174 template, which has no *Sau3AI* recognition sites, showed no cutting or peaking activity.

5.4. Reaction Kinetics Scheme

The reaction kinetics scheme for either the DNA cleavage or looping reactions is represented in Fig. 5.1. The first step in the reaction is the binding of the *Sau3AI* monomers to the single DNA strand non-specifically, characterized by the rate constants k_{ns+} and k_{ns-}. This step is followed by the monomers finding their specific sites, characterized by the rate constants k_{sp+} and k_{sp-}. It is widely believed that enzymes find their specific recognition sites through a combination of 3-D diffusion

from one part of the DNA molecule to another and 1-D sliding along the DNA. While k_{ns+} and k_{ns-} are likely independent of tension (i.e., $k_{ns\pm} = k_{ns\pm}(c)$, where c is the enzyme concentration), holding the DNA molecule by its ends interferes with the random diffusion of enzymes from one part of the DNA molecule to another (50). As pointed out above, we hold the DNA molecule taut until the enzyme solution fills the cell. With the flow stopped, the molecule is relaxed partially, assumedly restoring *some* 3-D diffusion from site to a site.

Once at a specific site, if the *Sau3AI* monomer needs to bend the DNA significantly in order to bind to the DNA, the so-called *induced fit mechanism* might also become rate limiting. The extent to which *Sau3AI* bends the DNA is not known. For *EcoRV*, which bends the DNA ~55º, tension did not seem to impact the induced-fit mechanism significantly until ~ 10 pN (50), which is much higher than the tensions in our experiments. Additionally, k_{sp+} will depend strongly on how many of the 55 GATC recognitions sites are occupied, which is convoluted by the peculiar stoichiometry. One might expect that with only one DNA molecule, all the recognition sites would be easily filled no matter what the enzyme concentration. As shown below, however, the observed cleavage rates are very sensitive to enzyme concentration. Therefore, it is reasonable to believe that $k_{sp+} = k_{sp+}(f,n)$, where f is the tension in the molecule, and n is the number of the sites filled, which is in turn enzyme concentration dependent.

The third step in the model kinetics scheme is the looping step. As presented below, tension in the DNA molecule strongly influences k_{loop+}. Likewise, because two monomers need to dimerize for looping to be observable, k_{loop+} also strongly depends

on the number of sites occupied and the length of intervening DNA. Therefore, $k_{loop+} = k_{loop+}(f,n,\Delta L)$. For the looping experiments where *Sau3AI* dimers capture the loops but do not cleave the DNA, the results of Friedhoff, *et al*., show that for experimental time scales, $k_{loop+} >> k_{loop-}$ (21). For the cleavage experiments, k_{cut-} is taken to be zero, and the k_{cut+} is taken to be fast compared to all other steps, an assertion supported by our data.

As a whole, the kinetics scheme of the looping or cutting of a single tensioned DNA molecule with multiple recognition sites is very complex, involving individual steps themselves enigmatic. Nonetheless, by empirically determining dynamic ranges of our control variables (i.e., time, tension, and enzyme concentration) and fitting reasonable phenomenological models to the data, we have been able to quantify the *in situ* looping in terms of effective rate constants and characteristic scalings for time, length, and tension.

5.5. Results and Discussion of the Sau3AI-DNA Cleavage Experiments

5.5.1. The Effect of Fractional Extension on DNA Cleavage

For the DNA cleavage experiments, the fractional extensions studied were $x/L_0 = 0.2$, 0.35, 0.5, 0.6, 0.7, 0.8, 0.9, and 0.93. According to the solution to the wormlike chain model (WLC) put forth by Bouchiat, *et al*., these fractional extensions correspond approximately to tensions of 0.03, 0.06, 0.11, 0.17, 0.29, 0.65, 2.5, and 5 pN, respectively (49). In practice, we held the molecule at 5 pN for the $x/L_0 = 0.93$ trials because in this region the force-extension curve steepens and reliable force clamping is possible. Each DNA was held at a particular x/L_0 (or 5 pN)

for t_{inc} = 30s, 60s, or 300s, and the fraction of molecules cut within that time were recorded. The results are shown in Fig. 5.2.

The lines in Fig. 5.2 represent sigmoidal fits (done in SigmaPlot) of the form

$$y = \frac{1}{1 + \exp((f - f_0)/b)}. \qquad \text{Eq. 5.1}$$

In the absence of an analytical model, this phenomenological model was chosen because it captured the observed behavior reasonably well and it can incorporate tension rather easily. To form a DNA loop of length ΔL against the applied tension f, an amount of work, $f\Delta L$, must be done to pull in the DNA. In the aforementioned theories of tensioned DNA looping, this work is incorporated into the partition function calculation via $e^{-\beta f \Delta L}$. The *b* parameter can thus be interpreted as k_BT divided by a length scale. In reality, there is a distribution of possible loop lengths for DNA cleavage on our template, each with its own looping probability. It should be noted that the f = 2.5 pN (x/L_0=0.9) point for the 300s data was regarded as an outlier.

The resulting fit parameters from Eq. 5.1 can help us qualitatively assess the cutting activity as $t_{inc} \rightarrow 0$ and ∞. As $t_{inc} \rightarrow 0$ we expect no cleavage activity, whereas we expect complete cleavage activity as $t_{inc} \rightarrow \infty$. The parameter f_0 is the tension at which one-half of the DNA molecules are cleaved, and it can be estimated from Fig. 5.2. For example, $f_0 \sim 0.2$ pN for the 30s data and ~0.5 pN for the 300s data. If f_0 continues to creep up slowly as $t_{inc} \rightarrow \infty$, complete cleavage can be achieved insofar as *b* stays finite, which seems to be the case. Likewise, letting f_0 go towards zero as $t_{inc} \rightarrow 0$ keeps the numerator of the exponent in Eq. 5.1 positive, which coupled with $b \rightarrow 0$, returns zero activity as $t_{inc} \rightarrow 0$. Although small, the *b*

parameter does not trend towards zero as $t_{inc} \rightarrow 0$ for the points we have. Instead, it has a finite size (~0.10 for the 30s data), reflecting the relative "softness" of the distribution at 30s compared to the others (i.e., the activity is less than one at lower tension for shorter t_{inc}). For this simplistic model to be sufficient then, the *b* parameter must turn over and go to zero at some lower time. In the cutting as a function of time data discussed below, there is a hint of this "turnover" in that the activity is peaked at times less than 30s. As t_{inc} increases the *b* parameter seems to plateau to a value of ~ 0.07 pN for the 300s data; indeed, longer incubation periods (~15 min) did not yield significantly higher observed cleavage percentages (data not shown). Using $k_bT \sim 4.1$ pN nm, this implies a length scale of ~ 59 nm for the DNA looping. As will be shown below, this value is very much in agreement with the observed expected ΔL values at comparable tensions. Additionally, the sharp drop in cleavage activity for all three t_{inc} demonstrates how strongly the activity depends on the entropic tension, a point that will be reiterated throughout our data.

5.5.2. The Effect of Enzyme Concentration on DNA Cleavage

To investigate the dependence of cutting on enzyme concentration, similar experiments to those performed above were repeated not with varying fractional extension, but with varying enzyme concentration. The flow cell was filled with a reaction mixture of 99λ of *Sau3AI* reaction buffer and 1λ of enzyme dilution; 1λ of the enzyme at stock concentration in 99λ of reaction buffer was called "1X", the concentration used in the fractional extension and timed pulsing cutting experiments. A ten-fold dilution of the stock enzyme, using NEB's Diluent Buffer A, added 1:99 to

the *Sau3AI* reaction buffer was called "0.1X," and so on. In this nomenclature, "2X" referred to 2λ of enzyme in 98λ of reaction buffer, and "5X" referred to 5λ of stock enzyme in 95λ of reaction buffer. Statistics were compiled at the following concentrations: 0.001X, 0.01X, 0.1X, 0.25X, 0.5X, 0.75X, 1X, 2X, and 5X. Each DNA molecule was held at $x/L_0 = 0.5$ and it was probed after 30s, 60s, and 300s.

The results are shown in Fig. 5.3. A prominent feature is the apparent sigmoidal dependence of the cutting activity on enzyme concentration, saturating as the enzyme concentration gets higher (≥ 2X). Both the low enzyme non-linearity and the high enzyme saturating behaviors are consistent with previous observations in bulk for *Sau3AI* (21). Additionally, the saturation towards complete cleavage at high enzyme concentration is consistent with dimerization model for *Sau3AI*. If each *Sau3AI* monomer bound two copies of its recognition sequence, like *BspMI*, the cleavage activity would fall off after peaking at an optimal enzyme concentration (5).

As with the fractional extension data, we expect no cleavage as $t_{inc} \rightarrow 0$ and full cleavage as $t_{inc} \rightarrow \infty$, no matter what the (non-zero) enzyme concentration. For lower t_{inc} fewer loops form, and thus the likelihood of two monomers finding each other and cutting the DNA increases with more bound monomers. The number of possible loops grows quadratically with the number of bound monomers (with the likelihood of each loop forming modulated by its length), so at low site occupancies the system is more sensitive to increases in enzyme concentration. As the sites become filled, however, the fractional change in the number of possible loops decreases as $\sim n^{-1}$. Thus we expect the behavior to be sigmoidal in enzyme concentration. For a sufficiently long t_{inc}, just two bound monomers will eventually

find each other, meaning that the activity will be less sensitive to an increase in enzyme and more in accord with the non-sigmoidal Michaelis-Menten kinetics. As longer incubation times did not yield significantly higher cleavage percentages, however, collecting statistics for $t_{inc} > 300$s was not practical.

The model used to assess this supposed sigmoidal behavior is a Hill function

$$y = \frac{c^b}{c^b + c_0^{\ b}}. \qquad \text{Eq. 5.2}$$

The Hill function is a generalization of Michaelis-Menten kinetics that accounts for cooperativity in the binding of multiple reaction sites, the quintessential example of which is oxygen binding to hemoglobin. It should be noted that in order to get a sensible fit to the 300s data, the 0.01X point had to be ignored as an outlier. Cutting was so scarce down at these low concentrations that it very well might be just a few tethers giving way or something comparable that pushed this point away from the general trend of the data. It should also be noted that the arbitrary notation for enzyme concentration presents no difficulties for the model of Eq. 5.2.

Plotting these parameters against t_{inc} shows us the behavior in the short and long time limits, as shown in the inset of Fig. 5.3. The *b* parameter, called the Hill coefficient, is the power that governs the activity, and it is on the order of unity. The error bars are large, so it is difficult to say whether the values are constant or trending downwards with increasing t_{inc}. The dotted line is the average value of 1.202 ± 0.347. Though there is no known cooperativity in the binding of *Sau3AI* monomers, that $b > 1$ reflects the "effective cooperativity" necessary for DNA cleavage brought about by looping. Increasing the enzyme concentration leads to more bound monomers,

increasing the likelihood of two bound monomers finding each other and cleaving the DNA. This is especially important for lower t_{inc} during which fewer loops form, as we will see. As $t_{inc} \rightarrow 0$, the Hill coefficient seems to be trending towards ~3/2. With increasing t_{inc}, however, we see a downward trend towards unity for b, in accord with simple Michaelis-Menten kinetics. As mentioned above, even just two bound monomers will eventually find each other after enough time, meaning that the measured activity will be less sensitive to an increase in enzyme concentration. The c_0 parameter represents the concentration at which half the molecules are cleaved after t_{inc}. It grows large as $t_{inc} \rightarrow 0$, in accord with the expected zero cleavage activity for $t_{inc} = 0$. As t_{inc} grows, we see c_0 trending downwards with a $\sim t_{inc}^{-1}$ dependence (dashed line). The downward trend and its simple form may be tractable theoretically. Finally, this Hill model underestimates the activity for lower enzyme concentrations (≤ 0.1X), reflecting perhaps the aforementioned 'chelate effect.'

5.5.3. The Effect of Incubation Time on DNA Cleavage

The third variable by which we studied the *Sau3AI* cutting was the time. For these experiments we used the 1X concentration and an $x/L_0 = 0.5$. Instead of t_{inc} = 30s, 60s, or 300s, the molecule was pulsed quickly every 5 seconds to 4 pN. If the molecule was still intact, it was returned to its relaxed position for another 5 seconds up to 300s. If the molecule was cut, the test tension would never be reached as the beads were separated. The number of completed intact pulses for each molecule was recorded and binned to form a cutting time distribution (not shown). We did these experiments for 1X (N=160) and ¼X (N=250), and the distributions are shown in Fig.

5.4. Each distribution shows a steep increase to a peak at low time (~15s and ~30s, respectively). This peaking at $t < 30$s, which we ascribe to enzyme loading, may be the "turnover" mentioned in the analysis of the tension cleavage experiments. Additionally, the statistical expectations values for these pulsed timing distributions are $\langle t_{1X} \rangle = 59$s and $\langle t_{¼X} \rangle = 109$s, different by nearly a factor of two. Each peak is followed by a decaying tail due to the recursive nature of the distribution – the number cut in any particular 5s bin is a function of how many molecules survived all the previous 5s bins. In order that probabilities sum to unity, the number cut in any particular bin is normalized by the total number of DNA molecules studied. Under this normalization, the fact that there are progressively fewer and fewer DNA molecules left uncut as time progresses will naturally yield a tail to the time distribution. To reflect this recursion, we propose:

$$P(i) = f_{experimental}(i) \times \left(1 - \sum_{j=0}^{j=i-1} P(j) \right), \qquad \text{Eq. 5.3}$$

where the second term represents the probability that the DNA molecule made it to the i^{th} bin in the first place. If there were no enzyme decay or depletion due to sticking to the flow cell walls, and the DNA-enzyme system had already achieved steady state equilibrium, one might expect that $f_{experimental}(i)$ is a constant, depending only on time independent factors such as temperature. Realistically, however, one might expect $f_{experimental}(i)$ to increase over low times as the enzyme comes into equilibrium with the DNA strand, and then slowly fall off as more and more enzyme

becomes non-viable. Due to the recursive dependence of the P(i) in Eq. 5.3, $f_{experimental}(i)$ and the looping kinetics become difficult to discern from the time data.

A way in which we can eliminate the recursive nature of the measured probability is not by relating the probabilities to the total number of trials, but to the total number remaining uncut at any particular time. In other words, we can consider each run in parallel, counting how many of the molecules have been cut after each interval. The $f_{experimental}(i)$ of Eq. 5.3 can be approximated by the number that cut within any particular interval divided by the number that were available to be cut. One drawback, however, is that these "available probabilities" do not sum to unity over time. One needs only to consider the last molecule of the sample to understand this distinction; in whichever time interval the last molecule is cut, the "available probability" of that interval will be one, albeit with an uncertainty of the same size.

The results of this "available probability" calculation are shown to t = 100s in Fig. 5.5. We see a slight downward trend throughout the bulk of both data sets, consistent with the beginnings of enzyme decay/depletion. We can also see that the probability for a cut in the case of 1/4X is genrally less than that of 1X. Calculating the ratio of one to the other (1X to ¼X), eliminating 0's and ∞'s (since the numerator-denominator selection is arbitrary and zero values are due only to a lack of statistics), and weighting the individual ratios according to the product of the $N_{availables}$ for both samples, the average is 1.97 ± 0.37. The same procedure but inverted (taking the ratio of ¼X to 1X) yields 0.56 ± 0.19. The uncertainties on these average ratios were calculated by weighting the terms in the variance the same way, and then taking the

square root thereof to obtain a weighted standard deviation. This is another instance of the recurrent factor of two between the cutting activity at 1X and ¼X.

A second way in which we can circumvent the recursive nature of the measured pulsed timing distributions is to look at the cumulative distributions, shown in Fig. 5.6. The cumulative fraction of the molecules cut by a particular time is analogous to DNA cleavage assays in bulk. As with the other cleavage experiments, the observable activity seems to have leveled sufficiently by t_{inc} = 300s to make longer times impractical. In theory, however, one expects that cleavage would be complete as $t_{inc} \rightarrow \infty$. Therefore, these fractions were fit with an ever-increasing double saturating exponential of the form

$$y = a(1\text{-}exp(\text{-}k_{eff,1}\, t)) + (1\text{-}a)(1\text{-}exp(\text{-}k_{eff,2}\, t)). \qquad \text{Eq. 5.4}$$

Note that the $t = 0, \infty$ conditions are met for this model. The results of this fit are also shown in Fig. 5.6; the fit parameters are in Table 5.1. The numbers in parentheses are the time scales implied by the effective rate constants. Note that no weighting was used for these fits, and that a single saturating exponential fit the data poorly.

Though the uncertainties are large on $k_{eff,2}$, there is an order of magnitude split between $k_{eff,1}$ and $k_{eff,2}$, indicating two disparate time scales. The coefficient defining the magnitude of the terms relative to each other (a) favors the fast term by a

Table 5.1. Cumulative Distribution Fit Parameters for Pulsed Timing Cleavage

	¼X	1X	Ratio (1X/¼X)
Dbl Expo a	0.71 ± 0.21	0.80 ± 0.07	1.13
$k_{eff,1}$ (s^{-1})	1.8e-2 ± 3.4e-3 (56s)	2.9e-2 ± 2.0e-3 (35s)	1.61
$k_{eff,2}$ (s^{-1})	1.5e-3 ± 1.6e-2 (690s)	4.3e-3 ± 4.7e-3 (234s)	2.95

factor of ~ 4. However, this term is the same (within error) for each concentration, showing again that the effects of concentration are in the effective rate constants. The ratio of the fast rate constant for 1X to that of ¼X is 1.6, whereas the same ratio for the slow rate constants is 2.95, implying that the slow process is more influenced by the four-fold concentration difference than the fast process. The ratio of the fast term to the slow term for 1X is 6.7, whereas it is 12.3 for ¼X, indicating that the splitting of time scales is nearly twice as great for the four-fold decrease in concentration.

The effective the half-life can also be inferred from the cumulative distributions. They are ~ 30s and ~ 64s for the 1X and ¼X data sets, respectively. These times are relatively short compared to the experimental duration (for which cutting was not complete), indicating that these cumulative fractions are changing most rapidly at low times. As shown in the inset of Fig. 5.6, the lowest data points can be approximated as linear in time, and a rate constant can be inferred from the slope of those points. Those rates are 2.18e-2 ± 1.06e-3 s^{-1} for the 1X data and 9.97e-3 ± 1.73e-4 s^{-1} for the 1/4X data. The ratio of these two rates is 2.18 ± 0.11.

Throughout the analysis of the timed cleavage experiments, the ratio of activity between the 1X and ¼X data sets was shown to be very nearly two. That the activity is just halved despite the fourfold decrease in concentration is an indication of the chelate effect discussed in the *Introduction*. Conversely, a fourfold increase in enzyme concentration yields only a two-fold increase in cutting activity. In this way, our data demonstrates how DNA looping may buffer a system from drastic changes in the concentration of one of its enzymatic components (3).

5.6. Results and Discussion of the DNA Looping Experiments

5.6.1. Detection of DNA Looping

In previous sections, we discussed the cutting dynamics of *Sau3AI* on the 55 site template as a function fractional extension, enzyme concentration, and time. While the integrity of the DNA is a clear signal - either the molecule is still intact or it is not - no information about the number or the length of the loops formed can be gleaned from cutting experiments. As mentioned in the *Introduction*, replacing the Mg^{++} in the reaction buffer with Ca^{++} traps the loops in the DNA without cleaving it (Fig. 5.1). EM studies have shown that these loops are rather stable for *Sau3AI* (21), indicating that an equilibrium between loop formation and dissociation had not yet been realized in our experiments, even for t_{inc} = 300s. As we will see, however, the average number of observed loops as a function of time has largely leveled off by our longest incubation time. Gathering statistics for longer t_{inc} was therefore impractical. It should also be noted that in addition to the aforementioned φX174 control experiments, trials utilizing the one site Type II restrictions enzymes *HaeIII* (36 sites), *MspI* (49 sites), and *BstNI* (26 sites), each in its NEB buffer made with Ca^{++} instead of Mg^{++}, were done to ascertain if the observed peaks were endemic to using *Sau3AI* (and indeed two site restriction enzyme in general). These enzymes were chosen because of their large number of recognition sites on the template. Despite more than 50 attempts in each, no peaking activity beyond that of the enzyme-free control set (discussed below) was observed (not shown).

To build distributions of the two looping activity metrics (i.e., the number (N) and length (ΔL) of the loops), individual DNA molecules were tethered between the

beads and exposed to 1X enzyme solution. To investigate the effects of fractional extension on looping, the DNA molecule was held for one minute at $x/L_0 = 0.2$, 0.5, 0.7, or 0.8, which correspond to forces of 0.03, 0.11, 0.29, and 0.65 pN, respectively. The experimental results are usually shown here as a function of fractional extension because it was the experimentally controlled variable. Maintaining a constant force via real time feedback is very hard in the entropic region, where the force-extension curve is macroscopically flat and below the Brownian noise limit. Where comparison with the theories of tensioned DNA looping is more direct, however, the results are presented in terms of these entropic tensions. It should be noted that the data for this study was collected before the publication of the tensioned looping theories, thus our experimental tensions do not exactly match up with those of the theories.

To accumulate statistics, 219, 315, 219, and 110 pulls were taken for $x/L_0 = 0.2$, 0.5, 0.7, and 0.8, respectively. We collected 98, 186, 315, 233, and 130 pulls for $t_{inc} = 10$, 30, 60, 120, and 300s, respectively ($x/L_0 = 0.5$). Example pulls are shown in Fig. 5.7, along with a distribution of disruption forces; these forces for many of two site REases were the subject of Chapter 4 and a separate study (47), and are comparable to the forces observed for unraveling nucleosomes (48) and many other DNA-protein interactions.

From these data sets we formed distributions of the number of observed loops, using the total number of pulls at each experimental condition as a normalization factor. However, the distribution needs to be corrected for false events due to the DNA handle, the streptavidin, or the αDIG proteins peeling slightly from the beads. To that end, 540 pulls of the template without enzyme were analyzed as a control set

and the number of events recorded. The distribution of the number of false events was distributed in a roughly Poisson manner, with $\lambda = 0.33$. This distribution of false events was used to correct the measured data by

$$P_{real}(N) = P_{meas}(N)P_{false}(0) + \sum_{j=1}^{\infty} P_{meas}(N+j)P_{false}(j). \qquad \text{Eq. 5.5}$$

The first term represents the probability that if N loops were measured in a particular pull, all N loops were real. Because the complete absence ($i = 0$) of events must be real in the measured data, this first correction is applied only for $i > 0$; for $i = 0$ the first term is simply the $P_{meas}(0)$. The summation term represents the probability that pulls with more measured events were actually comprised of enough false events to give them N real events. The results of this calculation for the 120s data set are shown as the gray bars in Fig. 5.8. Because there can be only false events and not the false absence of events, the correction shifts the distribution to lower N values as a whole. After this correction the distribution are renormalized so that the probabilities sum to unity. All N distributions presented henceforth have been corrected in this manner, though it should be noted that this is a first order approach to the correction. As we will see, the false events also have a length distribution that in general is much shorter than the "real" events. In principle it would be possible to correlate this control set ΔL distribution to the measured ΔL and N distributions. In other words, a long event is less likely to be a false event than is a short event, and a pull with one long event is more likely to be a pull with one genuine event than is a pull with one very short. Eq. 5.5 makes no such distinction. However, the ΔL distribution of the control set will be considered in a future section.

5.6.2. Distributions of N Loops per Trial

The N distributions, corrected according to Eq. 5.5, are shown in Fig. 5.9 for t_{inc} = 10s, 30s, 60s, 120s, and 300s. As t_{inc} is increased, the number of loops seems to increase significantly and the distribution widens. Being largely a result of thermodynamic fluctuations, DNA looping might be supposed to be a Poisson process. To investigate this, the expectation value <N> of each distribution was calculated and used to create a Poisson curve, overlaid as the black line on each distribution. For the first three times (10s, 30s, and 60s), the N distributions are close to the Poisson prediction, but favoring slightly lower N values. For the lowest time (10s), the shift to lower N may be related to enzyme loading, as was mentioned in the analysis of the timed cutting. The same data implies that the enzyme loading has (nearly) saturated by 30s. For the longer times (120s and 300s), the N distributions are wider than the corresponding Poisson curve. As t_{inc} grows, more loops form and, consequently, independence of individual loop formation is lost. On one hand, with each loop formed there are two fewer sites available for subsequent loop formation. On the other hand, when a loop is formed, two sites that otherwise would have been far apart are constrained to be closer together and thus will form a loop more readily, as sketched in Fig. 5.9. This cooperativity in loop formation will shift N to higher values than the Poisson prediction, which can be seen in the 300s distribution. This cooperativity, however, is not just an *in situ* curiosity. It has been shown that the DNA looping ability of the tetrameric protein RXR facilitates transcription *in vitro* by bringing the promoter and enhancer sites closer together when the two RXR sites lie between the promoter and enhancer sites. In the mouse genome there are 172

promoter regions that contain within them two RXR sites spaced 30-500 bp apart (51). In this way, the looping efficiency of the transcription systems will be dependent on the state of the looping state of the architectural RXR system.

The N distributions for the extension data sets are shown in Fig. 5.10. With decreasing x/L_0, the number of loops increases and the distribution thereof widens. Again, the Poisson curves calculated with the <N> values are overlaid on each distribution. For the $x/L_0 = 0.8$ data, the process is nearly Poisson. With the fewer loops formed at this x/L_0, it is reasonable that any particular loop occurs independently; indeed, ~85% of the pulls at this fractional extension had N = 0 or 1 loops. As x/L_0 is reduced, however, the effects of site saturation and looping cooperativity are manifested as deviations from Poisson dynamics at high N values.

A less sophisticated explanation for the deviations from Poisson dynamics is the inherent uncertainty in selecting runs for inclusion in the distribution. For the most active conditions especially, runs that had very low N values that were included in the distribution might have been "bad" runs, not looping well due to a number of unaccountable variables in the experiment (e.g., transient flows due to bead channel pressures or micropipette suction, unusually high enzyme depletion to the flow cell walls and tubing, etc.). Such "bad" runs could not be *a priori* discarded because there was no accepted distribution to which to compare the data. Falsely low N values would lower <N>, which would falsely shift the Poisson curve to lower N. Because the distributions are normalized to the total number of pulls, the inclusion of falsely lower N pulls would actually decrease the relative contribution of the higher N pulls, further exacerbating the difference from the Poisson curve.

5.6.3. Statistics of the Number of Loops Distributions

Shown in Fig. 5.11a and b are the expectation values, <N>, and the second moments, σ_N, of the N distributions shown in Figs. 5.9 and 5.10, respectively. For the time data sets, <N> was fit with a saturating exponential function $<N> = a(1-\exp(-k_{<N>}t))$. The resulting parameters are a = 5.80 +/- 0.57 loops and $k_{<N>}$ = 1.13e-2 +/- 2.63e-3 s^{-1}. The time scale implied by $k_{<N>}$ is ~ 88s, more than twice as large as the shorter time scale from the timed pulse cutting (see Table 5.1). This is not surprising in that the time scale from the cutting experiments refers fundamentally to forming one loop, whereas the ~ 88s represents a time after which one might expect ~3 loops. This simple description also implies that one might expect ~ 6 loops as $t_{inc} \rightarrow \infty$, a prediction that depends on to what extent the two competing effects of site saturation and looping cooperativity have balanced each other after the five minute incubation. The width of the distribution is quantified by the second moment, σ_N. For the time data, a double saturating exponential fit these values nicely. The two rate constants from this fit imply time scales of ~ 6s and ~ 83s, and a long time asymptote of ~ 3.33 loops. We might expect that σ_N actually decreases as nearly all the sites are incorporated into loops, reducing the variance in the number of loops. As with the dimer formation-dissociation equilibrium, realizing such site saturation experimentally might be practically impossible due to the difficulties inherent in keeping a DNA molecule tethered for such a long time and enzyme depletion.

The <N> and σ_N values of the x/L_0 data sets were both fit linearly. The <N> fit yields an intercept, y_0 = 5.22 ± 0.15 loops, and a slope, m = -5.68 ± 0.26 loops.

For σ_N, the fit parameters are $y_0 = 4.12 \pm 0.11$ loops and $m = -3.81 \pm 0.19$ loops. The <N> line falls to zero at $x/L_0 \sim 0.9$ (~ 2.5 pN). While thermodynamically a large enough fluctuation to invoke looping is always possible no matter what the tension, this result is reasonable and illustrates how sensitive looping is to applied tension.

5.6.4. Tails of the N Distributions

The N distributions shown in Figs. 5.9 and 5.10 each show a decaying tail section. Unlike the results of the timed cutting experiments, there is no recursive nature in the N distributions. The bins for the timed cutting experiment were arbitrarily set to be 5s and those for the loop length ΔL, were likewise arbitrarily set to 10nm; the bins for N are very natural The only subjective decision then is what range of N to include in the section to be fit. Referring to the distributions shown in Fig. 5.9, the selected tails ranges were $N \geq 0$, $N \geq 1$, $N \geq 1$, and $N \geq 3$ for t_{inc} = 10s, 30s, 60s, and 120s, respectively. No well fit tail section for the 300s distribution could be discerned; the distribution shifts to larger N and falls off steeply, as is evident in the leveling of σ_N in Fig. 5.11. As an example of a well fit tail, the fit to the 30s N distribution is shown in Fig. 5.12a, and the parameters to all the fits are plotted against time in Fig. 5.12b. Referring to the four distributions in Fig. 5.10, the following ranges were selected: $N \geq 3$, $N \geq 1$, $N \geq 0$, and $N \geq 0$ for x/L_0 = 0.2, 0.5, 0.7, and 0.8, respectively. The parameters from these fits are plotted versus x/L_0 in Fig. 5.12d. Shown in Fig. 5.12c is the fit to the 0.2 data set as an example.

For the tail fits of the incubation time N distributions, both the *a* and *b* parameters trend downwards with time, which means that the tail in general is getting

shallower and, consequently, the magnitude of the exponential describing it (i.e., the *a* parameter) diminishes. As t_{inc} grows and more and more loops form, however, we expect that the tail section might actually steepen as the N distributions narrows about the ever increasing <N>. We may be seeing this already for the 300s distribution. The σ_N of the distribution as a whole has leveled off, and thus the monotonic flattening of the tail has subsided as well. As $t_{inc} \rightarrow 0$, we expect no looping at all (i.e., all the runs have N = 0). This would correspond to an extremely fast decay of the tail (i.e., $b \rightarrow \infty$) and thus *a* tending towards unity. We do not clearly see that feature, though the lines on Fig. 5.12b are intended only as visual aids to capture the general trends and are not necessarily intended to imply real behaviors as $t_{inc} \rightarrow 0$. The anticipated narrowing effect as looping increases is a consequence of site saturation; mathematically capturing that behavior is well beyond the scope of this phenomenological treatment.

For the tails of the fractional extension data sets, an offset power law of the form $y = y_0 + c(x/L_0)^d$ fits both the *a* and *b* fit parameters quite well; the values for these fit parameters are listed in the caption. Both *a* and *b* increase sharply with increasing fractional extension, as evidenced by the 4.09^{th} and 3.36^{th} powers governing them, respectively. A very large *b* value as $x/L_0 \rightarrow 1$ ensures that the N distribution falls off very sharply at such fractional extensions; the increase in *a* is a consequence of this increasingly sharp fall off and the fact that the N distributions are all normalized to unity. The leveling off of both *a* and *b* as $x/L_0 \rightarrow 0$ to nonzero values ensures that there would indeed be an exponential tail to the N distribution for zero fractional extension. A zero value of *b* would dictate that all N values would be

equally likely. Clearly this can not be true with a finite number of sites and, moreover, a finite incubation time. However, the increase in looping activity also shifts the distribution to higher N values and renders it with a broad peak, shifting the bulk of the data out of the tail section. These facts make it more difficult to select the tail the section and assess its importance.

5.6.5. Cumulative Loops per Trial Distributions

As with the pulsed cutting data, the cumulative N distributions circumvent many of the difficulties in examining the N distributions directly. The cumulative distributions for the incubation time and fractional extension data sets are shown in Fig. 5.13. Unlike the cumulative distributions for the timed pulsing, which were zero at $t = 0$, the cumulative N distributions do not go to zero for N = 0. This introduces an offset into any phenomenological fit function that is not translationally invariant. Consequently, a sigmoidal function of the form

$$y = \frac{1}{1+\exp\left(-(N-N_0)/b_{sig}\right)} \qquad \text{Eq. 5.6}$$

was used; note that it is invariant to a translation in the N variable. The fits to the cumulative N distributions using are also shown in Fig. 5.13. Qualitatively, an offset saturating exponential of the form $y = a(1-\exp(-b(N-N_{offset})))$ and an offset logistic function each gave better fits. However, the introduction of the offset parameter made the parameters vary wildly with t_{inc} and x/L_0. The N_0 and b parameters resulting from the sigmoidal model varied smoothly, however, and trended similarly to $\langle N \rangle$ and σ_N (Fig. 5.11), respectively, for both the time and x/L_0 data sets.

For the time data sets, N_0, which represents the point at which the cumulative N distributions reaches one-half, trends monotonically from less than one-half of a loop at 10s to ~ 5 loops at 300s. In reality, the number of loops, and hence N_0, may saturate for a template with a finite number of sites. The simplistic sigmoidal model of Eq. 5.6, however, can not account for such discrete effects. The *b* parameter, which reflects the "softness" of the cumulative distribution and scales as the width of the N distribution, starts at ~ 1 loop for 10s and increases slowly to ~ 2 loops at 300s. In reality, as $t_{inc} \rightarrow \infty$, we might again expect that the N distribution narrows as all the sites become incorporated into loops. As $t_{inc} \rightarrow 0$, that both the N_0 and *b* parameters can be smoothly interpolated to zero allows for the expected zero looping.

For the fits to the x/L_0 data sets, the N_0 parameter decreases sharply from ~3.5 loops at $x/L_0 = 0.2$ and becomes negative at about $x/L_0 \sim 0.75$, meaning that more than half of the trials will have no loops at all if the molecule is held at a higher x/L_0. This is consistent with cutting results after one minute (see Fig. 5.2), for which the "half activity" fractional extension is ~0.7. As $x/L_0 \rightarrow 0$, N_0 approaches 5, meaning that half of the pulls would have more than five loops at zero fractional extension; recall that $\langle N \rangle \sim 5$ in this case as well. The *b* parameter decreases from its force-free interpolated value of ~ 2.5 loops to zero loops ultimately, reflecting that the N distribution narrows as the looping is suppressed with increasing x/L_0.

5.7. Comparison of Cleavage and Looping Data

A natural comparison between the cutting and looping data is the fraction of unlooped runs in the former compared to the fraction of uncut molecules in the latter.

The sole difference is that the Mg^{++} in the cutting is replaced by Ca^{++} in the looping. Although very few restriction enzymes efficiently cleave DNA without Mg^{++}, we believe that the affinity of *Sau3AI* for the DNA and its ability to form specific DNA complexes is not critically reduced upon this replacement. A similar equivalency (to within a factor of three) has been shown for both *EcoRV* and *EcoRI* (52,53), as well as the two site restriction enzyme *SfiI* (18). Therefore, the fractions of uncut and unlooped molecules should be roughly equal for comparable experiments.

The comparison between the different experiments (Fig. 5.14a) indicates that the pulsed timing cutting experiments are systematically less active, with the fraction uncut being significantly higher after all three times. This systematic discrepancy is discussed below. For the time being, however, the averages have been calculated without the pulsed timed cutting data to serve as a reference to which to compare the fractions. Calculating these averages (also shown in Fig. 5.14a) implicitly assumes that the cleaving of the DNA happens much faster than the looping or formation of the dimer. Because the uncut fraction is not at all higher than the N = 0 fraction (except for the timed cutting), it is safe to conclude that the cleaving of the sugar-phosphate backbones of the dsDNA happens very quickly relative to the looping/ dimer formation. Furthermore, there is decent agreement among the different experiments after 30s and 300s, but poor agreement after 60s. For the cutting experiments, 300s represents a saturated regime in which the cutting is nearly complete and the looping is nearly assured. The 30s and 60s data sets, however, represent more active regimes, and thus are more sensitive to the difference in activity. Referring to Figs. 5.2 and 5.3, we see that the 60s points lie off the fit lines

for both the fractional extension and enzyme concentration experiments. This may in part explain the apparent disagreement between the looping and cutting for the 60s time scale. Considering the data after 30s, where the activity is least saturated in all three experiments (the pulsed cutting is not included in this comparison), the uncut fraction from the fractional extension experiment is 0.24 ± 0.09, the uncut fraction from the enzyme concentration experiment is 0.27 ± 0.09, and the unlooped fraction is 0.28 ± 0.04. Because the uncut fractions are not higher than the N = 0 fraction, it is safe to conclude that the cleaving of the sugar-phosphate backbone of the DNA is not the rate determining step of the entire cleavage reaction.

The comparison of the fractional extension cutting and looping experiments (Fig. 5.14b) reveals fair agreement between the uncut and N = 0 fractions. The lines on Fig. 5.14b are intended only to elucidate the general trend of the data. Consequently, as $x/L_0 \rightarrow 0$ one can argue that the N = 0 and uncut fractions both do indeed go to zero (e.g., a linear fit to the four "N=0" points actually has a negative *y*-intercept). Whatever small differences there are may in part reflect the fact that there are many more trials for the looping than the cutting experiments (on the order of hundreds to roughly thirty). The discrepancy at the low tension end is estimated to be about 0.05, reasonably close to N_{trials}^{-1} for the cutting experiments. As $x/L_0 \rightarrow 1$, we expect that the uncut fraction and the unlooped fraction should both approach unity. The uncut trendline actually approaches one at a slightly lower fractional extension (~0.85). The N = 0 trendline, however, approaches ~ 0.75 as $x/L_0 \rightarrow 1$, indicating that even at such extensions one trial in four will show unlooping events. We feel that this is due to enzyme monomers that occasionally facilitate DNA adhesion to the

beads, which would be mistaken for a loop upon probing the DNA. The discrepancy between these cutting and looping results may also reflect the difficulty in deciphering "good" trials from "bad" trials, discussed originally in the context of the differences between the N distributions and the Poisson curves. While it is difficult to decipher the legitimate N = 0 runs from systematically erred looping trials, it is largely impossible to do so in the cutting experiments, especially in less active cutting conditions. This difficulty would lead to an unduly large percentage of uncut molecules in the final data set.

The comparison of the timed cutting experiments in Fig. 5.14b reveals that every uncut fraction is greater than the corresponding N = 0 fraction, indicating that the pulsed timing systematically inhibits the cutting activity. In order to examine the cutting with fine temporal resolution, the molecule was probed every 5s from its incubation position ($x/L_0 = 0.5$). The probing was a quick (~ 4 μm/s) tug out to 4 pN, a tension well above the noise/drift level over the five minute experiment time. If this test tension was reached, the molecule was returned to $x/L_0 = 0.5$. Faster tugging was empirically deemed too coarse, resulting in a jerky pipette motion. Despite the fast tugging, the molecule spent ~15% of each cycle at fractional extensions greater than 0.7, where the cutting activity is significantly reduced (see Fig. 5.2). Because both the cutting and looping activities saturate with time, this systematic inhibition is most pronounced for low times. Consequently, every uncut fraction is greater than the corresponding N = 0 fraction, with the difference between the two diminishing with time. As $t_{inc} \rightarrow 0$, we expect both fractions to trend towards unity because no incubation time should correspond to no cutting/looping. The uncut fraction does so

nicely, but the N = 0 fraction trends towards 0.5. Such a large discrepancy can not be explained by statistics. We feel that unresolved false events and the difficulties in eliminating erred cutting runs lead to this difference.

5.8. Loop Length ΔL

5.8.1. Distributions of Number of Loops per ΔL

In addition to the number of loops, the distribution of the lengths of those loops can shed light on the looping dynamics. Even when the DNA molecule is under no tension, entropy inhibits the formation of large loops while the intrinsic rigidity of DNA resists the formation of short loops, leading to a peak in the probability distribution. When the DNA is under tension, additional work must be done to loop out the length ΔL against the constraining force. Referring to example pulls shown in Fig. 5.7, we measure ΔL as the distance between a peak on a force-extension curve (where the loop is disrupted by the applied force) and the point at which the force-extension curve returns to the same force value. This sudden increase in tether length corresponds to the length of DNA in the loop. Additionally, the force extension curve would often drop again before the original drop force was reached, corresponding to a second disrupted loop. In these instances the increasing portion of the force-extension curve after the initial drop was extrapolated to the original drop force to obtain ΔL for the original loop disruption.

The ΔL values at each experimental condition were binned (10 nm), and the counts were divided by the total number of events to form a normalized distribution. As with the N distributions, these distributions must be corrected for false events.

The same 540 control pulls had a total of 181 false events. The ΔL values of these control events were likewise 10 nm binned and normalized. This control ΔL distribution, modulated by the probability of any false events existing per trial, was subtracted from the measured distribution using the first-order correction equation

$$P_{real}(\Delta L) \cong P_{meas}(\Delta L) - \sum_{j=1}^{\infty} P_{false}^{(N)}(j)\, j\, P_{false}(\Delta L)\,. \qquad \text{Eq. 5.7}$$

The summation term represents the ΔL distribution of the false events multiplied by the probability of an event being false. However, since the there exists the possibility of more than one event per pull being false, the sum is calculated over the entire distribution of N false events per pull (i.e., the $P_{false}(j)$ from the N Loops per Pull discussion), with each term contributing j false events. The resulting ΔL distributions are then renormalized so that the sum of each equals unity. The results of this calculation on the 300s data set are shown in Fig. 5.15, with the distribution of the false ΔL values is shown as triangles and uses the right vertical axis. The final distribution shows a peak at moderate ΔL and a decaying tail at higher ΔL. As will be discussed below, there is a surprisingly large number of short events (i.e, ΔL < 40 nm, the persistence length of DNA in 10 mM Ca^{++}) for every distribution.

The corrected and renormalized ΔL distributions were then rescaled according to the total number of observed events and trials to reflect the number of loops formed per ΔL. These distributions are shown in Figs. 5.16 and 5.17 for the time and fractional extension data sets, respectively. For subsequent analysis, these distributions are also renormalized by the number of possible loops per ΔL bin afforded by

our template. Though the spectrum of possible ΔL is fairly constant, this step minimizes the discretization of the ΔL distribution brought about by our template. As noted in the *Introduction*, the theoretical models for DNA looping focus on the steady-state formation probability of single loops of a particular size ΔL. We feel that this renormalized distribution of the number of loops formed per ΔL bin most closely reflects this theoretical probability.

Most of the distributions in Figs. 5.16 and 5.17 show a peak at finite ΔL and a decaying tail at higher ΔL. As discussed below, there is a surprisingly large number of short events (i.e, ΔL < 40 nm) for every distribution. Also noteworthy is that with increasing t_{inc} (at $x/L_0 = 0.5$; Fig. 5.16), a significantly increasing number of loops form (i.e., the distributions grow vertically). The ΔL distribution of those loops, however, changes only slightly. With decreasing x/L_0 (t_{inc} = 1 min; Fig. 5.17), an increasing number of loops form, the distributions broaden significantly in ΔL, and shift towards longer ΔL values. Conversely, this shift towards shorter ΔL values with increasing tension is a hallmark of the tensioned looping theories.

5.8.2. Statistics of the Observed Loop Lengths

The statistics for the observed ΔL values as a function of t_{inc} are shown in Fig. 519a. As $t_{inc} \rightarrow 0$ we expect no looping, which renders $\langle \Delta L \rangle$ meaningless. We therefore make no assumptions about the behavior of $\langle \Delta L \rangle$ for t_{inc} < 10s, and fit an offset saturating exponential to the data. The implied time scale is ~ 44 s and the asymptote is ~ 104 nm (~ 306 bp). The standard deviation, $\sigma_{\Delta L}$, as a function of time

did not fit well. The statistics for the observed ΔL values as a function of x/L_0 are also shown in Fig. 5.19b. Note how much more the ΔL statistics are affected by fractional extension than incubation time. Both $<\Delta L>$ and $\sigma_{\Delta L}$ are fit linearly. For the $\sigma_{\Delta L}$ values, $y_0 = 128 \pm 1$ nm and $m = -107 \pm 2$ nm. For the $<\Delta L>$ fit, $y_0 = 177 \pm 6$ nm and $m = -166 \pm 10$ nm. This reduction of the looping length-scale with increasing applied tension is a characteristic of all three tensioned-DNA looping theories, as alluded to in the *Introduction*.

The force-free $<\Delta L>$ (i.e., the y_0 parameter) is substantially lower than the expectation value for the classical cyclization theory by Shimada and Yamakawa of 262 nm. This value was calculated using a model for DNA molecules between 100 and 1500 bp (a range encompassing most of the range of our observed events) and using a protein interaction distance, d, of $0 < d < 20$ nm (31,32). In accord with these calculations, the interaction distance for the case of two *Sau3AI* monomers (discussed below) may be as large as 10 nm. However, the cyclization condition of this classical theory (i.e., that the DNA bends ~360°) is more energetically costly than a less severe bend such as a teardrop (~270°) or a hairpin (~180º). The loop trapped by the gal repressor tetramer in *e.coli*, for example, is a hairpin, aided by the histone-like protein HU (14). The *Sau3AI* dimer-DNA loop structure has not yet been published, and it very well might adopt a protein-bridged hairpin configuration, which would lead to a substantially lower $<\Delta L>$. Additionally, hinge and kink defects in the DNA are believed to play a profound roll in DNA looping by dramatically lowering the energy stored in the looped DNA (37,38). Discussed below, recently published models of

tensioned-DNA looping utilizing such defects are in better agreement with our data than those that do not (39,40). For example, the long-time <ΔL> value of ~ 104 nm reported above is longer than the <ΔL> of 67.4 nm calculated from the digitization of the probability distribution for a π/2-kinked loop at f = 0.15 pN ($x/L_0 \sim 0.5$) shown in Fig. 8 of Ref. 39, but shorter than the estimated <ΔL> = 215 nm for their tear-drop loop at a comparable tension (Fig. 6; 0.08 pN). This is the first indication in our data that 'non-classical' effects such as kinking afford better agreement with theory.

5.8.3. Peaks of Loops per ΔL Distributions

While the published tensioned-looping models are largely numerical and thus do not provide an equation by which to calculate <ΔL>, they do show how the peak ΔL values (i.e., the most probable ΔL) are affected by tension. Accordingly, the peaks of each ΔL distribution for our data were determined. Referring to Fig. 5.15 for reference, a small portion of the distribution bracketing the peak was selected (e.g., 20 to 110 nm). With spuriously 'underfilled' or 'overfilled' bins thrown out, the remaining points were fit with a parabola, a Gaussian, or a Weibull function and the ΔL of the peak determined. Since range selection can significantly affect peak finding, however, a number of different yet reasonable ranges were fit and the results averaged. The final results were checked qualitatively against the shape of the distribution. The error bars are the quadrature averages of the error bars reported for each fit range.

The peaks of the ΔL distributions are shown in Fig. 5.19. For the peak ΔL values versus t_{inc}, an offset saturating exponential was fit as to make no assumptions

about the peak ΔL value as t_{inc} goes to zero. The effective rate constant implied by this fit is $k_{eff,peak} = 3.83 \times 10^{-3}$ s^{-1} (~260 s), and the long time asymptote is ~ 94 nm (~ 276 bp). This effective rate constant is in decent agreement with the slow time scale from the timed cutting (Table 5.1), and the long-time asymptote falls between the predicted steady-state values estimated from Fig. 8 of Ref. 39 of ~ 147 bp and ~ 455 bp at f = 0.10 pN ($x/L_0 = 0.5$) for the π/2-kinked and teardrop loop, respectively.

The peak ΔL values for the fractional extension experiments are also shown in Fig. 5.19, but with the horizontal axis cast in terms of tension for comparison with theory. The solid line through the data is a decaying exponential with a force scale of ~ 300 fN. The lines through the upper right corner represent the theoretical results for a teardrop shaped loop from Ref. 41 (solid line) and from Ref. 39 (dashed line). Both theory predictions are much higher than our observed peak values, though it should be noted that the theoretical calculations utilize a persistence length of 50 nm. For the force-free probability distribution for a teardrop loop (Eq. 8 of Ref. 39), the peak shifts to lower ΔL by ~ 20 % if P is reduced to 40 nm from 50nm. The probability distributions as a function of force and persistence length, however, were not presented analytically; as many DNA binding proteins require divalent cations to bind to DNA efficiently, the tensioned looping probability distributions with P = 40 nm would be of great interest.

Additionally, neither accounts for significant protein bridging. As we will see, the lengths scale of the *Sau3AI* dimer is ~ 10 nm, which leads to a significant increase in the likelihood of looping as calculated by Rippe, *et al.* (32). Assuming that this correction is largely force independent, we can apply it to the distributions in

Figure 6 of Ref. 39. For example, a 10 nm protein bridge shifts the calculated peak of the 0.01 pN probability distribution from ~190 nm to ~184 nm. It shifts the calculated peak of the 0.2 pN probability distribution from ~122 nm to ~109 nm. Also, a protein bridge, when manifested experimentally, would shift the theoretical curves towards lower ΔL by an amount roughly equal to the bridge; the measured increase in tether length is the length of the DNA sequestered in the loop less the size of the protein bridge, as discussed below. After applying this second 'correction', the theoretical peak values for teardrop loops are still very much longer than what we observe. Thus, while it is important, protein bridging alone can not rectify the theoretical teardrop loop results with our data.

An additional shortening of the measured loop distribution can be attributed to the entropy of forming multiple loops, as are seen in many of our trials at lower tensions especially. Cooperativity in loop formation manifested as nested loops (discussed below in greater detail) was already discussed as a possible reason as to why the observed N distributions were non-Poisson. Recently, however, it has been argued that even for non-nested loops the presence of an energetic cooperativity between loops or the binding energy of the proteins stabilizing the loops can shift the probability distribution to lower ΔL values (54). While we do not assert that non-nested loops in our experiments utilize a formation cooperativity (i.e., we assume the "hard core" repulsion of Ref. 54, the binding energy, ε, of the *Sau3AI* dimers stabilizing the loops can dramatically increase the so-called "entropy factor", shifting the ΔL distribution to lower values. As pointed out in the discussion of Fig. 5.7, our observed disruption forces for *Sau3AI*-mediated loops are comparable to those

observed for many different protein-DNA interactions, therefore we assume a typical $\varepsilon \sim 10\ k_bT$. Figure 2a of Ref. 54 indicates that the shift of the peak of the looping probability distribution at f = 0 to lower ΔL values may be as large as ~ 15% for $\varepsilon = 10\ k_bT$ (for circular loops). Assuming a comparable shift for the teardrop shaped loop, this shift to lower ΔL, when combined with that induced by protein bridging, is significant, but not enough to rectify the theory with our data. The ε-driven shift to lower ΔL is not shown as a function of force in Ref. 54.

Figure 8 from Ref. 39 shows the peak values from the probability distribution when there is a $\pi/2$-kink in the tensioned DNA; this curve is represented in our Fig. 5.19b by the dashed line running through the middle of the plot. Such sharp kinks are hypothesized to be induced by the binding of proteins to the DNA (e.g., the TATA-box binding protein (55). For very low tensions, our measured peak values are comparable to these calculated values. Projecting our peak ΔL values back to zero tension yields ~ 93 nm (274 bp). Though there is some discrepancy between Fig. 4 and Fig. 8b of the theoretical paper, our projected value is much longer than the ~110 bp max for the $\pi/2$-kinked DNA value in Fig. 4 of Ref. 39. It is reasonable that a less severe kink in the numerical calculations would yield a force-free peak ΔL that is near our observed value.

Unfortunately, the numerical results for the case of hinged DNA are not presented for comparison. However, they state that the presence of hinges in the DNA allows for hairpin loop configurations to form more easily, shifting the peak ΔL below 100 bp (40). These hinges are postulated to occur about once in every one thousand base pairs, or about four times in our template (37,39). These postulated

hinges, as well as the effects of protein-bridging, may explain why our results fall to lower peak ΔL values much faster with increasing force than do the theoretical, π/2-kink calculations. Nonetheless, our results emphasize the importance of including DNA defects such as kinks and hinges in theories of tensioned DNA looping. It should also be noted that the theoretical results are usually in terms of forming one loop; the effects of forming several loops simultaneously, as is often the case in our results, will be addressed below in detail. Presently, it suffices to say that multiple loops can lead to a few falsely shortened ΔL values.

5.8.4. Tails of the Loops per ΔL Distributions

The $<\Delta L>$, $\sigma_{\Delta L}$, and ΔL_{peak} are all statistics that describe the loops per ΔL bin distributions, which, as mentioned, each have a surprisingly large number of short events, a peak of varying centroid and width, and a decaying tail. The peaks and their response to time and tension were discussed above; the short events and explanations thereof will be discussed below. The tails, however, characterize the formation of longer loops (typically with $\Delta L > P$) with regards to t_{inc} and tension. These tail sections are more relevant to those biological systems that that rely on the looping of hundreds or even thousands of base pairs. For example, *in vitro* experiments on the *lac* repressor show a repression decrease as the operator spacing is increased that is qualitatively similar to the tails of the loops per ΔL distributions observed here (16).

To characterize this 'long' loop regime, the fraction of observed events with $\Delta L > 40$ (P), 80 (2P), 120 (3P), 200 (5P), and 300 (7.5P) nm are shown in Fig. 5.20. For all t_{inc}, the fraction of events with $\Delta L > 200$ nm is small (<10%); recall that these

measurements are taken at $x/L_0 = 0.5$ where the shortening effect of tension is already pronounced. For every t_{inc} except 30s (for which there is an anomalously high number of long events), the fractions are spaced roughly equally up to 200 nm (5P). Additionally, each fraction largely levels off between 120 and 300s, even decreasing slightly in this time interval for the $\Delta L > 300$ nm curve. This mirrors the evolution of both $<\Delta L>$ and $\sigma_{\Delta L}$ in time (Fig. 5.18a). The same fractions as a function of x/L_0 are also shown in Fig. 5.20b. Most notably, the increase of tension shifts the distribution to shorter events, as predicted by the theories of tensioned-DNA looping. Roughly half the events have $\Delta L < P$ (40 nm) for $x/L_0 = 0.8$ (0.65 pN), whereas the long events (5P and 7.5P) are inhibited completely by such tensions.

The effect of time and tension on long events can be further compared to theory by phenomenologically fitting the region with an appropriate function. As evident in Figs. 5.16 and 5.17, the tail of each distribution decays monotonically in ΔL and thus can be fit with an exponential. The foremost difficulty in analyzing the tails of the loops per ΔL distributions in this manner, however, is selecting the region that constitutes the tail. For example, fits could be significantly altered by including an extra bin on the low ΔL end or excluding an anomalously large high ΔL bin. To circumvent this, a few different yet reasonable ranges were fit and assessed qualitatively. The time data sets all yielded tail sections of $\Delta L > 90$ nm. The fractional extension data had tail sections that shifted to lower ΔL with increasing x/L_0, as did the peaks. Decaying exponentials allow the tail sections of each distribution to be shifted so that $\delta(\Delta L) = \Delta L - \Delta L_{tail} = 0$ (where ΔL_{tail} is the first bin in

the tail section), and normalized so that the height of the first bin was unity. Effectively this reduced a two parameter fit (i.e. $y = ae^{-b\Delta L}$) to a one parameter fit (i.e., $y' = e^{-b\delta(\Delta L)}$), without changing the resulting b values (i.e., a length decay rate).

The fits to the shifted and renormalized tails of both data sets are in Fig. 5.21, as well as the resulting b parameters as a function of time and fractional extension. Note that the tails of the time data sets lie nearly on top of one another, whereas the fractional extension tails show a significant change. This is seen most clearly in the b parameter; note that the fractional extension uses the upper horizontal axis. The b parameter varies little with time compared to fractional extension, reflecting that the tension affects the formation of long ΔL loops most significantly due to the extra work needed to pull in the DNA against the force (39-41). For visual aid, a curve was fit through the five time points, leveling off at about $b = 0.02$ nm^{-1}. As a function of x/L_0, a simple exponential was fit for interpolation purposes and is also shown in Fig. 5.21c. As $x/L_0 \rightarrow 1$, b grows quickly but stays finite, ensuring at least the possibility of looping at the comparatively lower ΔL values that would comprise the tail section at higher tensions. This is consistent with the notion that there is always the possibility of a large thermal fluctuation to cause looping, even as the DNA is pulled nearly taut. As $x/L_0 \rightarrow 0$ in Fig. 5.21c, the length decay rate is $\sim 1 \times 10^{-2}$ nm^{-1}.

To compare these values with theory, we again turn to the tension-free, $\pi/2$-kinked looping model presented in Ref. 39, as that provided the best agreement in the peak ΔL analysis. It should be noted that protein bridging is of ever-decreasing importance with increasing ΔL and thus not relevant to the ΔL tail analysis. The

exponential coefficient governing the decay of the ΔL tail was determined by generating a P(ΔL) from their Eq. 8, selecting the tail as ΔL > 90 nm, shifting and normalizing as described above, and then fitting an decaying exponential. The result is a theoretical length decay rate of ~0.02 nm^{-1}, nearly twice as great as our interpolated force-free value. That the length decay rate of the theoretical ΔL tail for kinked-DNA is larger than our interpolated force-free value again indicates that a π/2-kink is too severe to describe our data. A less severe kink would agree more so with our results, if indeed the kinked model is applicable to *Sau3AI*. Its structure might mandate hairpin loops, teardrop loops, or allow for any looping geometry, the so-called free boundary condition. The theoretical result for length decay rate of the teardrop shaped loop is actually ~ 0.006 nm^{-1}, in decent agreement with our interpolated value. The theoretical teardrop tail ΔL, however, occurs for ΔL > 300 nm and with a peak ~ 150 nm, very much larger than our typical ΔL values (39). No matter what the looping configuration utilized by *Sau3AI*, the fact that our data yields peaks and tail parameters only a factor of four different than those predicted by theory invoking severe kinks (and in the same ΔL range) demonstrates the necessity of utilizing such non-classical constructs to describe our *in situ* DNA looping.

5.8.5. Analysis of Short Looping Events (ΔL < 40 nm)

Classical theories of DNA looping predict that loops with ΔL < P, the persistence length of DNA, should rarely form due to the intrinsic rigidity of DNA (31). Our data indicates, however, that they form in this system rather easily, dominating the loops per ΔL distributions at the two highest tensions studied. To

illustrate further this short loop regime ($\Delta L < 40$nm) in our data, the fraction of observed loops with $\Delta L \leq 10$, 20, 30, and 40 nm are plotted as a function of t_{inc} and x/L_0 in Fig. 5.22. For the former, the 30s data set was treated as an outlier in establishing the visual aids on the graph. As t_{inc} grows, the fraction of events with ΔL less than the particular length decreases steadily, indicating that with increasing t_{inc} more long loops form and diminish the relative contribution of the short loops. That the shorter loops form relatively quickly reflects their lower entropic cost, especially if any affinity exists between bound *Sau3AI* monomers. The fraction of short events increases significantly with increasing x/L_0 for all the ΔL limits considered, reflecting that DNA tension favors the formation of short loops relative to long loops. At the higher tensions, the loops per ΔL distributions are dominated by short events, at $x/L_0 = 0.8$, nearly two-thirds of the events have $\Delta L \leq 40$ nm. Additionally, the steepness of the curve as $x/L_0 \rightarrow 1$ decreases as we consider longer ΔL limits; the curves are becoming increasingly linear as the ΔL limit approaches 40 nm.

5.8.6. Discussion of Short Looping Events

In the classical theories of DNA looping, the geometry of the loops plays a critical role in their formation. A complete circular bend is the most energetically expensive, and thus the most unlikely. Tear-drop shaped loops relax this energetic cost, and free-boundary conditions (i.e., no preferred looping geometry) are comparatively still easier to form (31,40). If the loops are captured by proteins, the free energy of the monomer-monomer interaction may help overcome this cost. Additionally, the size and compliance of the 'protein bridges' plays an important role

in facilitating loop formation. The interaction distance, *d*, of this 'bridge' is often set to zero in theoretical calculations for simplicity. Experimentally, requiring that $d = 0$ corresponds to the full cyclization of closed loops unaided by proteins. Although the structure of *Sau3AI* has not been solved and its looping geometry is unknown, electron microscopy images of *Sau3AI*-trapped loops are consistent with the hairpin geometry (21). Additionally, the MW of the *Sau3AI* monomer is 56, 468, which is larger than most restriction enzymes (21). Using a typical specific volume of 0.73 cm^3/g, and the estimator for the volume of the protein, $V \approx (1.21 \times MW)$ $Å^3$ per molecule, an estimate for the diameter of the *Sau3AI* monomer (taken to be spherical for simplicity) is ~ 5 nm (56). This implies $d \sim 10$ nm, a bridging distance that Rippe, et al., have shown results in more than a ten-fold increase of looping probability for $\Delta L < 150$ bp (32). If the proteins can conformationally compensate to dimerize despite less than ideal interfacing geometry, the energetic cost of forming small loops could be reduced dramatically. *FokI* and *MboII* are two TypeIIs restriction enzymes that require two sites to cleave the DNA and undergo dramatic conformational changes upon binding to the DNA to facilitate dimerization (57,58). It is not unreasonable then to envision that *Sau3AI* monomers could conformationally change to facilitate dimerization, though not necessarily to the extent of *FokI* or *MboII*.

In the cell nucleus, the size of the regulatory proteins, typically larger than *Sau3AI* monomers, facilitates the formation of short loops. For example, the *lac* repressor tetramer has a length of ~13 nm (59), yielding to a minimal semicircular arc length of ~ 60 bp. Indeed, for loops formed on linear DNA *in vitro*, DNase digestion indicates a loop as short as 63 bp for the *lac* system (26). *BspMI*, the afore-mentioned

two site restriction enzyme that is a tetramer with a MW of ~215 kD (d ~ 8 nm), was able to cleave DNA with sites just 38 bp apart, albeit on a supercoiled template (20).

Protein bridging also leads to a systematic shortening of the measured ΔL. We measure the tether length increase for an unlooping event, which is shorter than the actual DNA spacing. For large loops, the fractional difference is negligible, but for the semicircle arc length of ~16 nm for a ~ 10 nm protein bridge, the measured ΔL is ~ 6 nm (i.e., $d(\pi/2\text{-}1) \sim 0.6d$). Although there are several possible loops less than 10 nm on our template, we believe that the lowest two bins in our loops per ΔL distribution especially reflect this systematic shortening.

Beyond favorable geometry and large protein bridges, the formation of very short loops depends on defects in the DNA in the form of very flexible hinges (37) and/or protein induced kinks in the DNA (39). Discussed above in terms of semi-quantitative predictions to which we compared our data, numerical calculations show that the kinks can dramatically shorten the ΔL distribution of loops and facilitate their formation against tension (39,40). For very short loops, however, a *Sau3AI*-induced kink at the apex of the loop might not be practical. First, there is no structural study asserting that the monomer bends the DNA. Second, even if the induced kink were a given, the size of the *Sau3AI* monomer makes it unlikely that one monomer can sit at the apex of the loop and induce the kink while two other monomers dimerize in close proximity (< 10 nm). However, any induced kinks at the dimer may help reduce the energetic penalty of bending the DNA 'entering' and 'exiting' the loop (41). In such a manner, the abundance of structural proteins might facilitate the formation of short loops, as was shown for the cyclization of ~100 bp fragments with the HU protein

(60). Likewise, the apparent persistence length of chromatin has been measured using FLP recombination to be just 27 nm (61). Even for underassembled nucleosome arrays in the absence of linker histones and other higher order structural proteins, the persistence length was recently measured to be ~ 20 nm (48). Such a shortened persistence length greatly enhances the formation of short loops *in vivo*.

Hinges in the DNA might have also facilitated the formation of the many short ΔL events in our data, and they present no spatial occlusion difficulties as *Sau3AI*-induced kinks might. Recently proposed to explain the unexpectedly large cyclization rates of 94 and 116 bp fragments (35), these hinges are postulated to be transient 'bubbles' of ssDNA formed by thermal fluctuations (37). The effective persistence length of such a bubble is less than 1 nm, greatly reducing the energetic cost of a very short loop. The energetic cost of forming a bubble is estimated to be ~10 kT, resulting in about 1 defect per micron (40). Our template is ~ 3.7 μm, and more than 15% of it lies between sites separated by less than 40 nm. Therefore, we can expect about three 'short loop' hinges every five runs; hinges did not significantly alter the calculations for looping of longer (ΔL > 100 nm) DNA segments (40). It is worth noting that the observed phenomenon of base flip-out, where individual bases transiently rotate outward on millisecond time scales, is tantamount to forming these single-stranded bubbles and carries an energetic cost of ~ 12 kT (62).

Such defects also facilitate the (un)twisting of the DNA necessary to align the proteins that close the loops by reducing the torsional modulus of the DNA. The helical repeat length of double stranded DNA is ~ 10.5 bp, meaning that enzyme sites separated by integer multiples thereof will be optimally aligned. While flexibility of

the *Sau3AI* may help in this regard, it has recently been shown in cyclization experiments, which do not rely on protein bridging, that strongly bent DNA is ~ 400 times more twistable than classically predicted (63). Similar findings regarding torsion reduction have been reported for shorter loops utilized by transcription factors *ara*C (64) and lac repressor (12,16). For increasing loop length, the dependence on helicity gradually disappears (16). To create smoother distributions for all ΔL, our data was 10nm binned (i.e., about three optimal lengths per bin). Combined with the multitude of possible ΔL values on our template, this makes discerning helical variations in looping difficult. However, as our data seems to mandate the inclusion of highly flexible regions of DNA, so too do the torsional bulk studies.

5.8.7. Theoretical Comparison of the Loops per ΔL Distributions

As mentioned above, the theories of DNA looping, tensioned or not, typically cast their results in terms of the formation probability for a single loop of a given geometry as a function of loop length. In our experiments, we often observed multiple loops of many different lengths. Consequently, our distributions of the number of loops per ΔL bin are the most analogous to the theoretical probability distributions. We have already examined the ΔL statistics, peaks, and tails of these distributions; the surprisingly large fraction of short events (ΔL < P) have also been discussed. To compare directly the entire experimental distributions with the theoretical probability distributions, however, requires special consideration. First, in order to minimize bib-to-bin noise, we regroup the distributions into 20 nm bins. Second, the theoretical probability distributions for a π/2-kinked loop (which

provided the best agreement for the peaks and tails) are calculated for nine discrete tensions (39); our data was collected at four distinct fractional extensions. Consequently, we chose the four closest tensions (f = 0.04, 0.15, 0.30, and 0.60 pN) to our four 'tensions' and digitized the theoretical curves. Third, the digitized theoretical curves need to be rescaled for comparison with our data. This was done so that the expectation value of the theoretical distribution equaled the expectation value of the experimental distribution at each tension, facilitating a comparison of the shapes of the distributions. However, this method destroys the relative scaling of the theoretical curves. Consequently, they are also rescaled so that the expectation value of the 0.04 pN curve equals the experimental expectation value at $x/L_0 = 0.2$, but then the other three theoretical curves maintain their size relative to it.

The results of this are shown in Fig. 5.23. Note that the observed distribution shifts relative to the theoretical curve from higher ΔL at $x/L_0 = 0.2$ to lower ΔL at $x/L_0 = 0.8$. At each tension, the lowest bin is higher than expected, as discussed in the 'Discussion of Short Looping Events.' Interestingly, for the two lowest tensions, there are a greater number of longer events ($\Delta L > 300$ bp) than theoretically predicted. At the two highest tensions, there is a significantly greater number of events at all ΔL relative to lower tensions than theoretically predicted, as indicated by the dotted lines in Fig. 5.23. As was seen with the peaks, and as will be seen with many other comparisons, this indicates that the *Sau3AI* looping system, although inhibited by DNA tension, is more robust to it than theory predicts. It should be noted that in the footnote of their paper, Sankararaman and Marko indicate that their tensioned looping theory gives less trustworthy results for $f < 0.1$ pN (39), yielding a

force-free peak ΔL that is actually slightly longer than that of non-tensioned theories. Given that, the force-free distributions from Eq. 8 of Ref. 39 for the π/2-kinked and teardrop loops with P = 40 nm are shown in the top panel of Fig. 5.23 for comparison.

5.8.8. Loop Length Absorption Rate

Because the loops per ΔL bin distributions (Figs. 5.16 and 5.17) are analogous to the probability of forming a loop at a particular ΔL, the expectation values of the distributions are analogous to the integral $\int \Delta L\, p(\Delta L)\, d(\Delta L)$. In the theory of tensioned DNA looping, this integral is described as the length absorption rate for non-specific loops (37). Our template has 55 sites, which provides a very large number of looping possibilities. Considering also that we normalized each of our distributions by the number of possible loops per ΔL bin, we can regard *Sau3AI* as a non-specific looping enzyme on our template for comparison purposes. Fig. 5.24a shows that the integral saturates for increasing time, as did the expectation value <ΔL>. The offset saturating exponential yields an effective rate constant of $k_{\text{eff},<N/\Delta L>}$ = 7.9e-3 ± 3.4e-3 s^{-1} (~127s) with a long time value of 654 ± 98 nm. This $k_{\text{eff},<N/\Delta L>}$ value represents a time scale that is longer than that of the time-evolution of <N>, σ_N, or <ΔL>, but shorter than that of ΔL_{peak}; it falls between the two time scales from the timed cleavage (Table 5.1). The asymptote value can be considered the expected amount of DNA incorporated into loops for extremely long times. This amount represents only ~18 % of the full DNA template length. Given the fact that many of the 300s pulls had ten and more events, this percentage again emphasizes how many surprisingly short events were measured in our experiments. Additionally, this

percentage is in good agreement with the value of 14% "loop coverage" inferred from Fig. 3 of Ref. 54 for 0.11 pN, zero loop interaction, and a protein binding energy of $\varepsilon = 10\ k_bT$, as utilized previously.

To facilitate comparison with Figure 10 of Ref. 39, the expectation values of the loops per ΔL are cast as a function of tension and the vertical axis is logarithmic in Fig 5.24b. The theoretically predicted form of $\int \Delta L\ p(\Delta L)\ d(\Delta L)$ is a stretched exponential with a asymptotic form of $exp((-(f - f^*)/f_0)^{1/2})$, where f_0 and f^* are force scales. The values are calculated to be $f_0 = 0.0051$ pN and $f^* = 0.0012$ pN for a teardrop loop, and $f_0 = 0.0134$ pN and $f^* = 0.0372$ pN for a $\pi/2$-kinked loop (39). With just four data points total - only two of which are in the asymptotic regime – fitting the stretched exponential is not meaningful. However, a double decaying exponential yielded the two force scales 0.054 pN and 0.499 pN with magnitudes 898 and 162 nm, respectively, for a force-free value of 1060 nm. Normalizing the afore-mentioned theoretical curves to this force-free value indicates that the applied tension dampens this loop length absorption for the *Sau3AI* system much less than for the kinked loop, as shown in Fig. 5.24b. Nonetheless, the form of our data suggests the stretched exponential form (39). Additionally, the $f = 0$ value of 1060 nm represents ~ 29% of the template length, in decent agreement with the force-free, fractional length absorbed for $\varepsilon = 10\ k_bT$ and no loop interaction from Fig. 3 of Ref. 54.

5.8.9. Cumulative Loops per DL and Normalized Looping Time

These loops per ΔL distributions can be further quantified in the same ways in which the N distributions were analyzed. The cumulative loops per ΔL bin

distributions, shown in Fig. 5.25, show the familiar saturation with increasing ΔL. Each is fit with a saturating exponential function, $y = a(1-e^{-\Delta L/\lambda})$, where λ represents a characteristic length scale and *a* is a magnitude with units of "loops." More sophisticated phenomenological models provided better fits qualitatively to every data set, but the resulting parameters varied wildly and provided little insight. With regards to t_{inc}, the parameter *a* starts at a value of ~ 2 loops (10s), and saturates with increasing t_{inc} to a value of ~ 6 loops, very similarly to how <N> behaved. The length scale λ started at a value of ~ 75 nm (at 10s) and crept monotonically upward to a long time asymptote of ~ 110 nm, similarly to how <ΔL> behaved with t_{inc}. With regards to x/L_0, both *a* and λ decreased nearly linearly with x/L_0, again similarly to how <N> and <ΔL> behaved with x/L_0. The parameter *a* has a force-free value of ~ 6 loops and goes to zero as x/L_0 approaches one. The length scale λ has a force-free value of ~ 185 nm, falls off to ~ 40 nm at $x/L_0 = 0.8$, and is nearly zero as $x/L_0 \rightarrow 1$. This length scale shortening is predicted by the recent theories of looping under tension, which predict that tension will hinder the formation of longer loops more so than it does the formation of short loops (39-41).

These loops per ΔL cumulative distributions facilitate comparison with the results presented in Fig. 7 of Ref. 39, which are the loop formation probabilities as a function of tension and ΔL for a π/2-kinked loop. As the theoretical results are extended only to 500 bp (170 nm), however, we consider only those measured loops with ΔL < 170 nm. This is easily done by considering the cumulative number of loops per ΔL distributions at ΔL = 170 nm. These values are shown as a function of

tension in Fig. 5.26a. When integrated over ΔL, the theoretical probability distributions should be analogous to the number of loops formed with $\Delta L \leq 170$ nm. Performing the digitization and integration of the curves for the nine tensions shown, and normalizing the trend to have the same force-free interpolation as our data, we see that the curve falls off more quickly than does our data, as was also the case with the peak ΔL values. That the *Sau3AI* system forms many more loops than predicted at high tension reflects, we feel, the important role of protein bridging in looping. Also, any affinity that exists between two bound *Sau3AI* monomers would be especially manifested at higher tensions.

The saturating exponential fits to these cumulative loops per ΔL distributions and the resulting parameters also allow for an idealized calculation of the measured distribution value at any ΔL. Because ΔL was one of our *dependent* variables (not a constant as in the theoretical calculations), and because there were only four fractional extensions studied in our experiments, this idealization facilitates less noisy comparison with the theoretical results. The first comparison that can be made is with Fig. 12 of Ref. 39, where the probability of forming a loop of a particular length (100, 125, or 150 bp) is shown as a function of tension for a $\pi/2$-kinked loop. The idealized loops per ΔL distribution values for those loop lengths are shown in Fig. 5.26b as a function of applied tension. The lines on Fig. 5.26b (bottom) are the corresponding lines from Fig. 12 of Ref. 39, but normalized to the same force-free value as that interpolated from our data for each loop length. It should be noted these separate normalizations lead to the 'uncrossing' of the theoretical curves discussed in

Ref. 39. As with many of the previous analyses of the data, we observe many more loops than theoretically predicted at higher entropic tensions, even with theoretically kinked DNA. This is especially true throughout the analysis for shorter loop lengths, emphasizing again the importance of protein bridging and the incorporation of 'defects' in describing *Sau3AI*-mediated DNA looping.

These idealized loops per ΔL distributions also allow us to calculate how tension affects the normalized looping time, defined as the ratio of the looping time under tension to the force-free looping time. This quantity is calculated as a function of DNA tension for a 100 bp, 200 bp, 500 bp and 1 kbp loop, utilizing either a hairpin or circular looping geometry (41). The idealized experimental distribution values divided into 60s (i.e., t_{inc}) to yield the expected looping times. Dividing these times by the extrapolated zero-tension value yields the normalized looping times. The results of this are shown in Fig. 5.27a with the theoretical lines from Fig. 3 of Ref. 41 for a hairpin loop, as it is the geometry closest to our situation. Note that our absolute looping times varied from ~ 2300 s (100 bp at 0.03 pN) to ~ 1.5×10^7 s (1 kbp at 0.65 pN); no absolute looping times were quoted for the theory because "loop formation is dependent on the biochemical and structural details of the protein bridge" (41).

For our data, the normalized looping time increases far less with tension than it does theoretically for all ΔL, with the effect being less pronounced for the shorter loops. That the looping in our experimental system is less sensitive to force than the theoretical calculations reflects a potentially different looping geometry, as well as the effects of protein bridging, flexible hinges, and protein induced kinks in the DNA. As mentioned in the "Discussion of Short Looping Events", kinks at the *Sau3AI*

dimer would reduce the energetic penalty of the bending of the DNA outside of the loop, a key component in the calculations of Blumberg, *et al*. Again, such kinking would reduce the sensitivity of the system to applied tension.

5.9. Additional Discussion and Analysis Summary

5.9.1. DNA Tension as a Regulatory Mechanism

In the cell nucleus tensions are exerted on sections of DNA from the action of molecular motors, protein binding, nucleosome positioning , and the like. RNA and DNA polymerases can exert forces of up to ~ 40 pN (65-67). Many proteins bend DNA significantly; the *EcoRV*, for example, bends its cognate DNA ~ 55° (68). Tension on the scale of tens of piconewtons reduces its ability to cleave the DNA (47,50); it is not unreasonable then that its binding can exert tension on partially constrained DNA, such as that between nucleosomes. This sort of tensioning has been described theoretically in the context of tension-mediated coupling of binding proteins (69). Additionally, extracellular mechanical stress can be transferred into the cell nucleus and converted into genetic regulatory signals (70). No matter what the source, DNA tensions would serve to modulate DNA looping, which in turn modulates transcription and the like.

Our data indicates that tensions < 1 pN drastically affect the ability of DNA to loop, so tensions on the order of >1 pN would effectively eliminate looping and the processes that rely on it. This possibility has been called a ‘genetic switch’, and has been characterized thoroughly by Blumberg, *et al*., in calculations of the free-energy of looping DNA under tension (41). The disruptive tension at which the looping is

effectively "turned off" is defined in their calculations as that which brings about a 100-fold increase in the mean looping time from that of tension-free DNA for a particular ΔL and looping geometry. This level was chosen because the *lac* repressor is only ~1% functional when looping is prevented by removing both auxiliary operators from the system. For example, this critical tension is calculated to be 430 fN (corresponding to a fractional extension of ~ 0.77) for a hairpin loop of 100 bp, and decreases with increasing ΔL. For our data, we can extrapolate the normalized looping time to 100 for the four loop lengths in Fig. 5.27a and solve for the disruptive tension. The results of this analysis are also shown in Fig. 5.27b For all loop lengths, the disruptive tension for our data is significantly higher than theoretical predictions using DNA with no defects or protein bridging. Unfortunately, the disruptive tensions were not calculated incorporating such considerations. However, they explicitly state that, according to their calculations, a tension of ~ 200 fN should have 'significant effects' biologically (41).

In our data analysis a significant switch in the data or in the parameters of a model describing it are evident near this 'critical' tension. The first example is the cleavage results of Fig. 5.2. When fit with a sigmoidal model, these results yield f_0 values (the tension at which activity is halved) of ~ 300 fN for the 60s data. The second example of a critical fractional extension is the N_0 parameter in the sigmoidal curves used to describe the cumulative N distributions (Fig. 5.13). Again, N_0 represents the number of loops per trial at which the cumulative distribution is one-half. As a function of x/L_0, N_0 goes to zero at $x/L_0 \sim 0.75$ (370 fN). At higher tensions, more than half of the pulls would have no loops. A third example is the

force scale of ~ 300 fN from the exponential decay of the peak ΔL values with tension (Fig. 5.19b). In each of these examples, the so-called critical entropic tensions are higher than the 200 fN discussed above, especially when considering that in our data the 'critical' level was often 50% activity instead of 1% activity. Considering the other comparisons made to tensioned DNA looping theory throughout this manuscript, this is additional evidence that protein-bridging and DNA defects are critically important in describing *Sau3AI*-mediated looping. Such considerations facilitate looping that is more robust against tension, and thus would allow for higher critical forces. Nonetheless, considering both theory and our data, the effectiveness of tension as a potential 'genetic switch' is evident.

5.9.2. Possibility of Nested Loops

Looping geometry and defects affect the formation of individual loops. As shown in the N Loops analysis, however, we often observe several loops in a trial, which presents the possibility of *nested loops*. Discussed as a possible explanation for the increasingly non-Poisson nature of the N distributions, cooperativity in loop formation manifests itself as nested loops when we probe the DNA. As shown in Fig. 5.28, nested loops would contribute a shortened ΔL value equal to the sum of the distances spanned from each 'outer' site to the inner intact dimer, not the full distance between the sites along the DNA. Because the interior dimer feels no applied tension until the outer dimer is disrupted, it is unlikely that it would dissociate first. Unfortunately, with our ΔL resolution and site-rich template, discerning between nested loops and full size loops by utilizing this summation is not readily possible.

We can more quantitatively assess the extent of nested loop formation by grouping the pulls from each experimental condition (i.e., varying t_{inc} or x/L_0) according to the number of observed events and then investigate the observed length distribution of each subgroup. Because nested loops give necessarily shorter ΔL values and form more readily as N increases, a negative correlation between the ΔL sub-distributions and the N Loops per trial may be expected if loop nesting is occurring. An example of this grouping for the entire t_{inc} = 120s data set is shown in Fig. 5.29. Using a 2σ criteria to eliminate outliers, the remaining values were averaged and are shown in Fig. 5.29 as well. Compared to the overall $\langle \Delta L \rangle = 87 \pm 4$ nm value for the entire 120s data set, the decrease in the average ΔL for increasing N is clear. This trend was likewise evident for nearly every experimental condition, especially the more actively looping ones wherein nested loops would be more prevalent. Considering each data point in Fig. 5.29 as an ordered pair (N, ΔL), we can calculate the linear correlation coefficient, ρ, between the ΔL and the N data sets. For the 120s data set, $\rho \sim -0.10$, and shows similar values for every especially active looping condition ($-0.07 < \rho < -0.12$). Because of the discrete nature of the N data set and the fact that it does not tail off, standard tests to assess the significance of ρ do not apply. However, the large number of points entering the calculation for each data set almost assuredly renders these ρ values meaningfully different from zero. Also, the average ΔL values for each subgroup steadily decrease with increasing N. A caveat to these assertions, however, is that as more loops form, there is less DNA

from which to form unnested loops. A negative correlation between N and ΔL does not prove the existence of nested loops, it is simply consistent with it.

5.10. Analysis Summary and Conclusion

We have studied the thermodynamic looping of DNA under tension facilitated by the two site restriction enzyme *Sau3AI*. Cutting experiments (i.e., with Mg^{++}) revealed that the activity is very sensitive to the tension on the DNA molecule. Cutting experiments in which the enzyme concentration was varied indicated a highly non-linear dependence on enzyme concentration and saturation in activity at high enzyme concentrations. This saturation is consistent with the dimerization model for DNA cleavage by *Sau3AI*. Cutting experiments as a function of pulsed-timing, when analyzed with a double saturating exponential model, gave an indication of two disparate time scales present in the looping dynamics. These timed-pulsing cutting experiments also demonstrated how looping buffers the activity of a system against large changes in enzyme concentration, the so-called 'chelate effect.'

Looping experiments (i.e., with Ca^{++}) allowed for the formation of distributions in the number of loops per trial, N, and the length of individual loops, ΔL, for different fractional extensions and incubation times. Decreasing x/L_0 increased the $\langle N \rangle$ and σ_N statistics significantly, with a force-free projection of ~ 6 loops (after 1 min). Lowering x/L_0 also led to an increased deviation from Poisson dynamics; the N distribution at $x/L_0 = 0.8$ was nearly Poisson in form, but was much wider than the simplistic prediction at $x/L_0 = 0.2$. Part of this deviation was attributed to difficulties such as enzyme depletion, but part of it was ascribed to site saturation and

cooperativity in loop formation, giving rise to *nested loops*. Increasing the t_{inc} led to the formation of more loops, with a saturation evident in both $\langle N \rangle$ and σ_N. The cumulative distributions were analyzed against both x/L_0 and t_{inc}. Comparison of the looping and cutting results yielded decent agreement and clearly demonstrated the systematic inhibition to cutting in the pulsed-timing experiments.

The loops per ΔL distributions facilitated the most direct comparison with theory. Increasing t_{inc} did not alter these distributions nearly as strongly as did increasing x/L_0. The effect of increasing t_{inc} was most clearly seen in the number of loops formed at every ΔL value. Increasing x/L_0 led to a nearly linear decrease in $\langle \Delta L \rangle$ and a comparable narrowing of the loops per ΔL distribution. The peaks of our loops per ΔL distributions shift to lower ΔL values as tension increases with a force scaling commensurate with other 'genetic switch' levels in our data. While the effect of protein bridging is believed to be significant, it alone could not rectify the theoretical predictions of 'classical' looping with our data. Conversely, the theoretical predictions involving DNA kinks agreed more closely with what we observed. Likewise, the parameters governing the decay of our measured ΔL-distribution tails, when extrapolated to a force-free value, indicate again that DNA kinks yield the best theoretical agreement. The loop length absorption rate, defined as the integral of the looping probability over all ΔL, showed two force scales in our data and arguably a stretched exponential form in the asymptotic limit, both consistent with theory. The phenomenological analysis of the loops per ΔL distributions indicated that tension inhibits the formation of long loops more so than short

loops due to the extra work needed to pull in the DNA against the applied tension, a prediction common to all the tensioned-looping theories. Additionally, the number of loops observed for all ΔL exceeded the theoretical prediction for increasing force (as did the number of loops formed at 100, 125, and 150 bp specifically) even when accounting for protein induced DNA kinks. The *normalized looping times* for our data were less than theoretically predicted for hairpin loops, and consequently the 'disruptive tensions' for our data were significantly higher than predicted.

Surprisingly, our loops per ΔL distributions had many events shorter than the persistence length of DNA, which classical theory says should not readily form. At the higher tensions studied, loops with $\Delta L < 40$ nm dominate the distribution. Protein bridging assuredly plays a significant role in the formation of such loops, as the length scale of the *Sau3AI* dimer may be as large as ~ 10 nm. The presence of protein induced kinks and flexible, single-stranded 'bubble' hinges in the DNA also facilitates the formation of short loops. Although the structure of *Sau3AI* has not been solved, proteins that bend their cognate DNA significantly (> 45º) are not uncommon. Furthermore, calculations indicate that on our template one might expect about three single-stranded 'bubble' hinges in five trials. These facts combined with many of the other findings again emphasize the importance of considering DNA defects when describing the statistical distributions of DNA looping.

The use of two site restriction enzymes may provide a number of unique opportunities to study different aspects of DNA looping. Although the structures of few of these two site restriction enzymes are known, one might expect a variety in DNA bending angles, as is seen across one site restriction enzymes. This would

allow for a more direct quantification of the effects of DNA kinking. Though electron microscopy images support the notion that *Sau3AI* traps hairpin loops, one might expect that different two site enzymes facilitate different looping geometries. The knowledge of the sizes of these enzymes might facilitate a more quantitative study of the effects of protein bridging. We chose *Sau3AI* to approximate a continuum of possible loop lengths on our most robust labeled template, but engineering DNA templates with a strict spacing of recognition sites, or engineering a number of templates each with different site spacing, would facilitate a more strictly quantitative comparison with tensioned DNA looping theories.

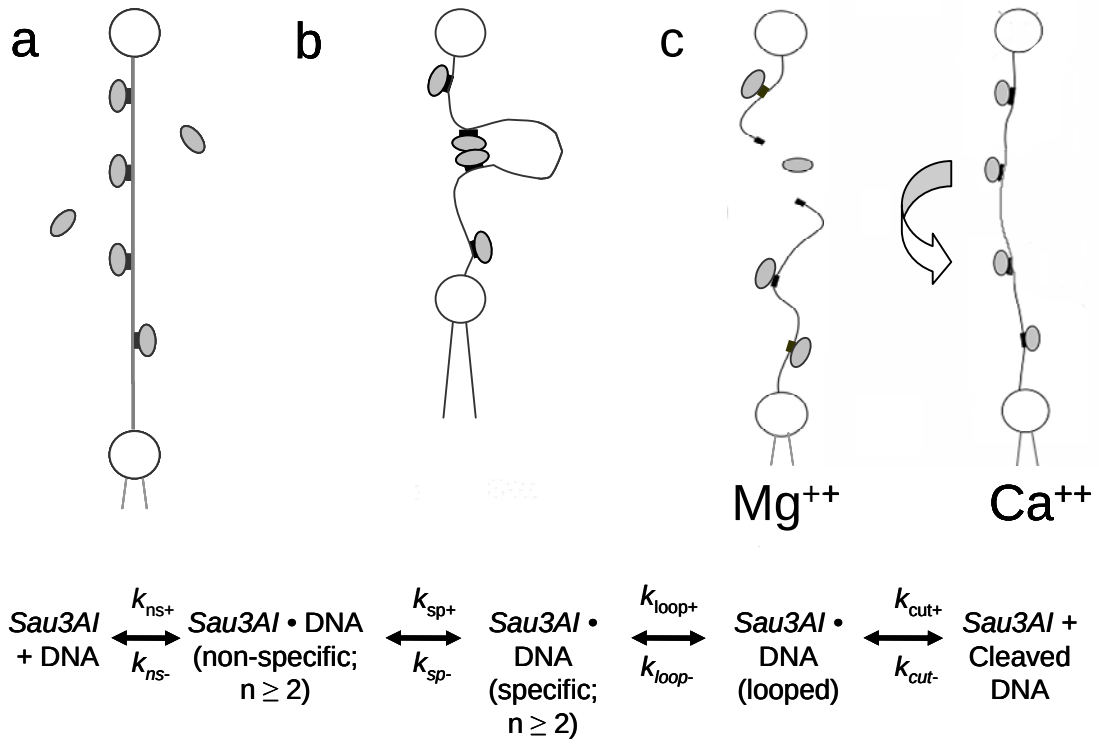

Fig. 5.1. (Top) Schematic of the Cutting and Looping Experiments. The gray ovals represent the *Sau3AI* monomers. The DNA is attached to the optically trapped bead (top) by a biotin-streptavidin linkage. The DNA is attached to the pipetted bead (bottom), which can move relative to the optically traped bead, by a digoxigenin-antidigoxigenin linkage. (a) With the DNA molecule taut (~5 pN) to prevent looping, the flow cell is filled with the enzyme solution. When the cell, is full, the flow is stopped. (b) The molecule is relaxed to prescribed fractional extension for a prescribed incubation time, during which it can loop and form the active *Sau3AI* dimer. (c) In the presence of Mg^{++} ions, the DNA molecule will be cleaved, which is detected as the lack of tether upon separating the beads. In the presence of Ca^{++} ions, the dimers form but do not cleave the DNA. Upon separating the beads, the dimer is disrupted (or removed from the DNA molecule), which results in a sudden drop in the force extension curve. This disruption is represented by the arrow. This figure is not meant to imply that every *Sau3AI* recognition site on the template was occupied. (Bottom) The reaction kinetics scheme for the cleavage or looping reactions.

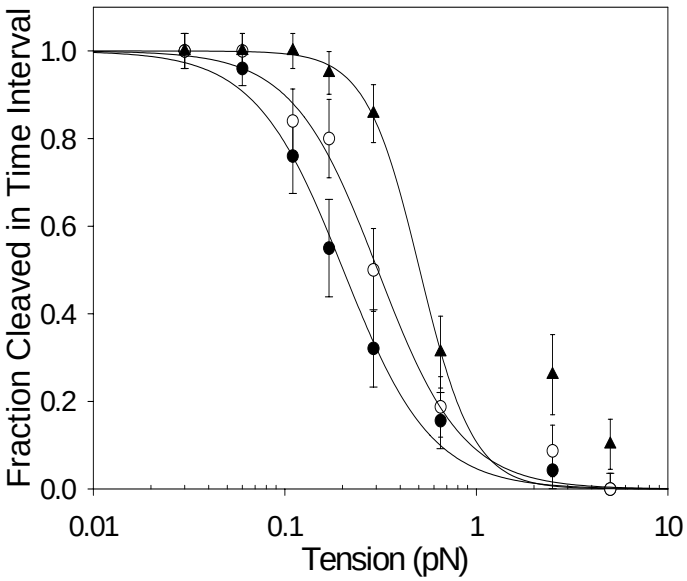

Fig. 5.2. Cutting Activity vs. Tension. As discussed in the text, the cutting activity is defined as the fraction of DNA molecules cut within *t*inc = 30s (filled circles), 60s (empty circles), and 300s (triangles). The error bars are calculated as the standard deviation of the binomial distribution $(p(1\text{-}p)/N)^{1/2}$, where p is the probability of a molecule being cut and N is the number of trials ($N \sim 30$ for each point). The lines are the fits according to the sigmoidal model of Eq. 5.1. The f = 2.5 pN ($x/L_0 = 0.9$) for the 300s data set was removed before fitting as an outlier.

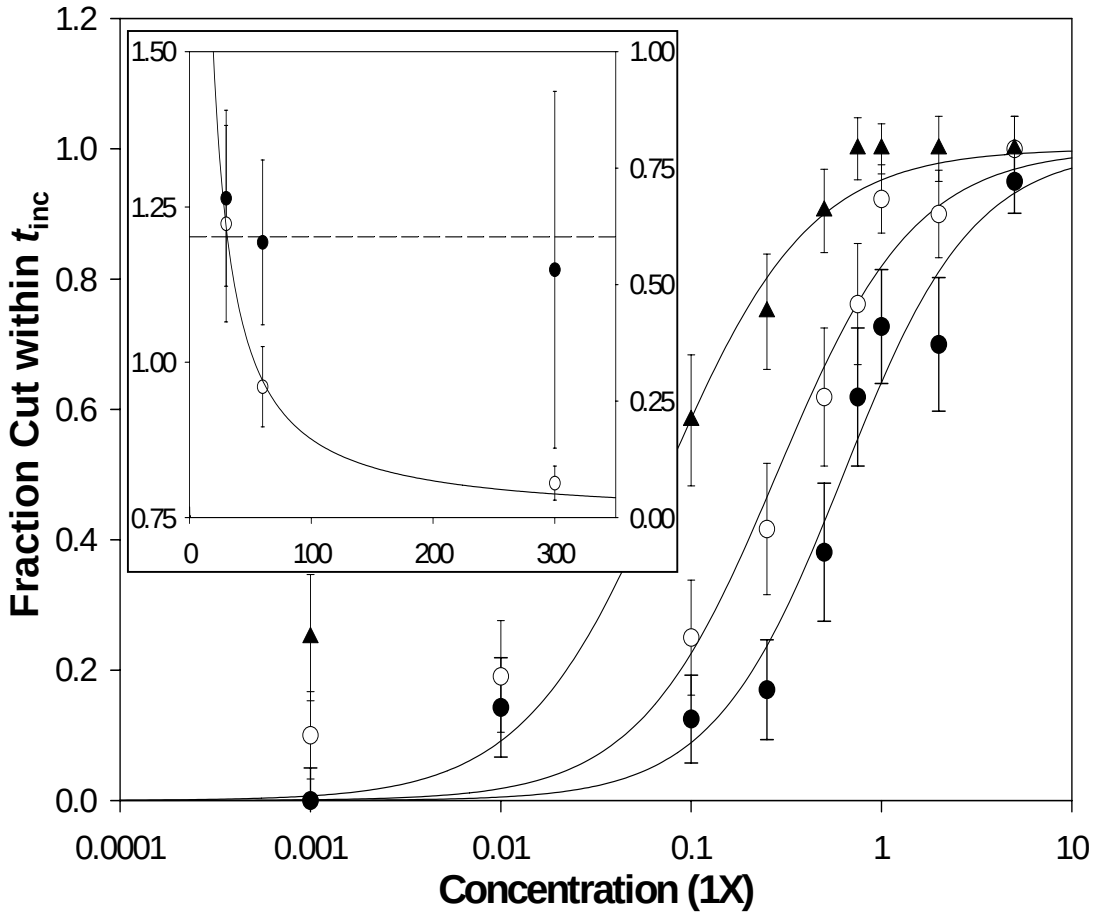

Fig. 5.3. Cutting Activity vs. Enzyme Concentration. Enzyme concentration axis shown on a log scale, and is relative to 1X as discussed in the text. Each DNA molecule held at x/L0 = 0.5 for 30s (filled circles), 60s (empty circles), or 300s (triangles). For the 300s data, the 0.01X point (~ 0.5) is not shown to accommodate the inset. Error bars are calculated using the width of the binomial distribution $(p(1-p)/N)^{1/2}$. The lines represent the best of Eq. 5.2. For the 300s data, the 0.01X point had to be removed to get a sensible fit. (Inset) The parameters of the Hill equation fits. Filled circles represent the Hill coefficients and use the left vertical axis. The average thereof, ~ 1.2, is the dashed line. The empty circles represent the c_0 parameter, which uses the right vertical axis. The horizontal axis of the inset is time with units of seconds.

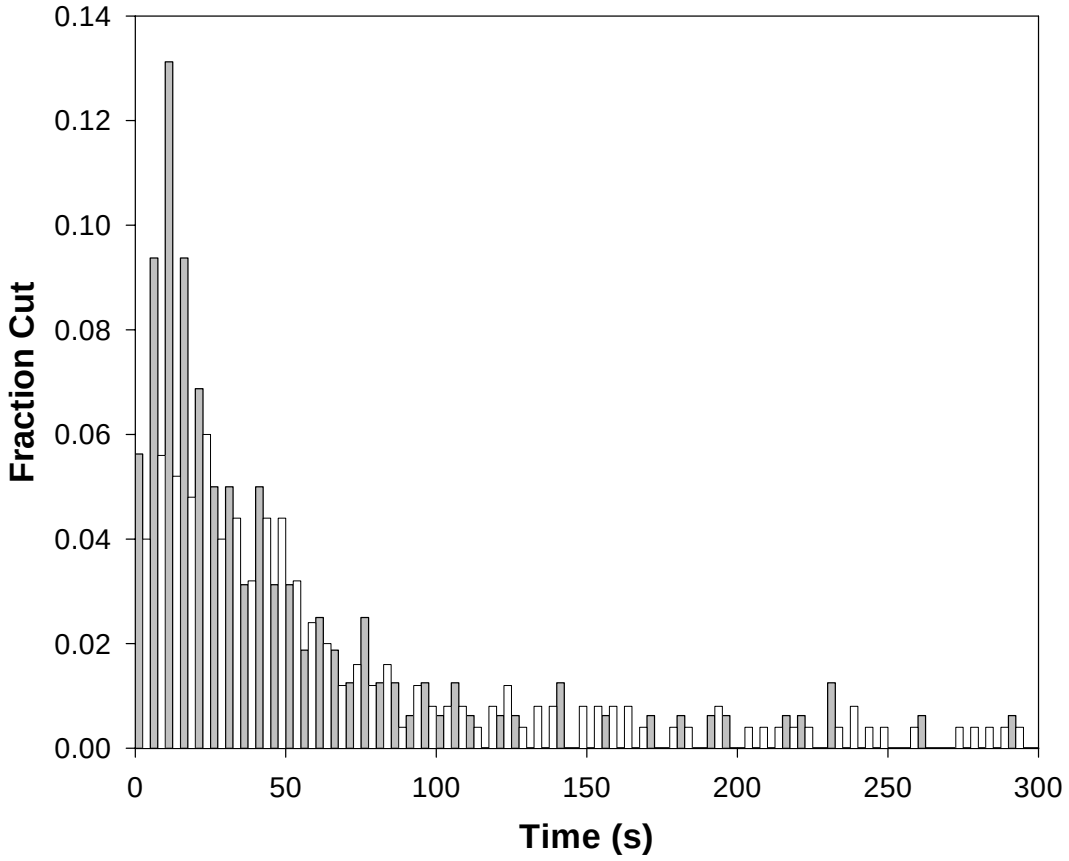

Fig. 5.4. *Sau3AI* Cutting as a Function of Time. The gray bins represent the relative frequency of cutting in that 5s bin for 1X, the white bins for ¼X. Not shown are the bins representing those molecules that did not cut within 300s (5% of the data for 1X, 17.6% for ¼X).

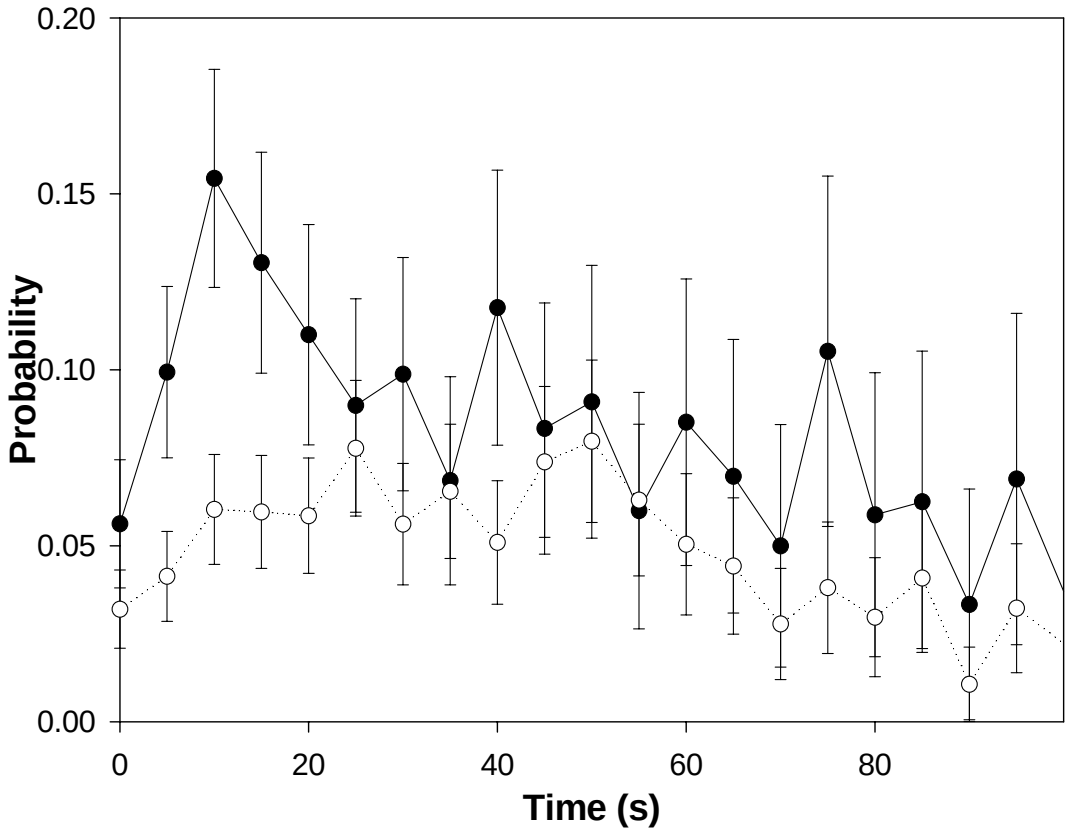

Fig. 5.5. *Sau3AI* Available Cutting Probability vs Time. The filled circles and solid lines represent the 1X data ($N_{initial}$ = 160), and the empty circles and dotted lines represent the ¼X data ($N_{initial}$ = 250). The errors were calculated using the formula $(p(1-p)/N_{available})^{½}$, where p is the number of cuts in a particular time interval divided by $N_{available}$, the number available to be cut at that time interval. As more than 75% of the trials for each concentration are cut by the 100s bin, the scale is only extended to that time for clarity.

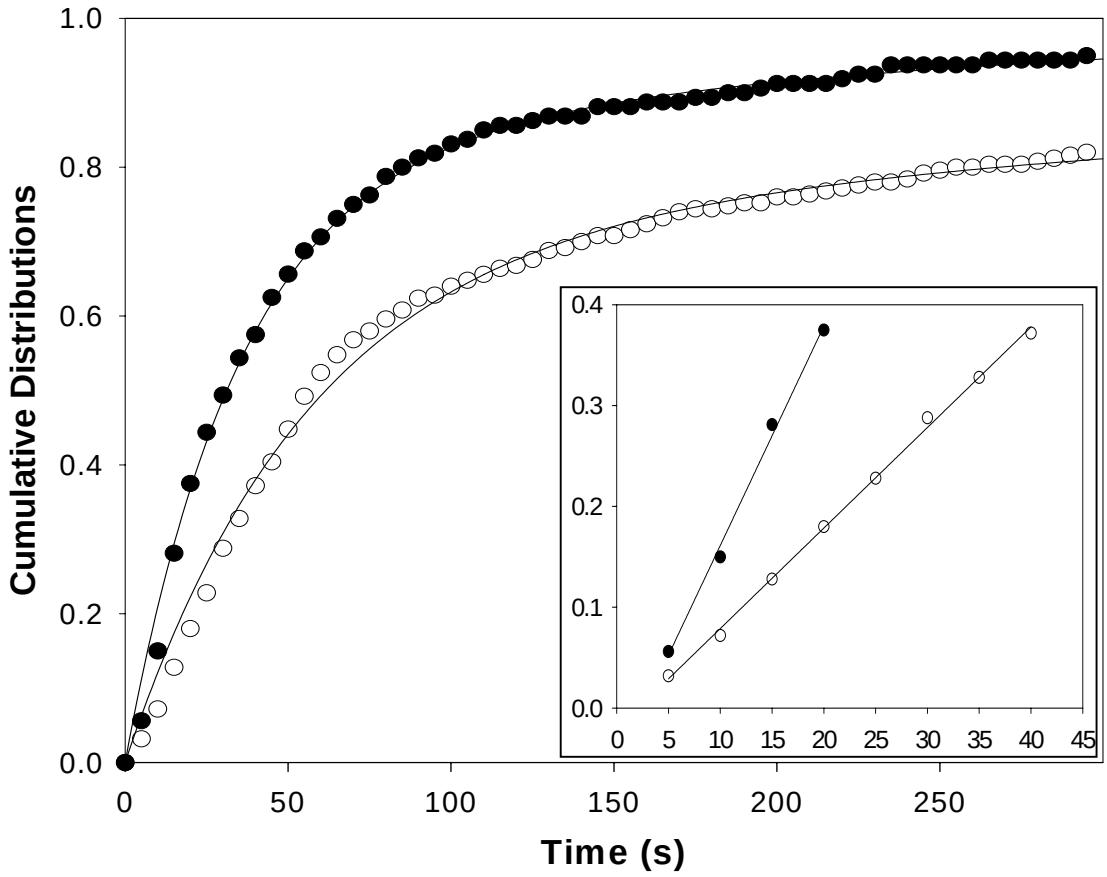

Fig. 5.6. *Sau3AI* Cutting as a Function of Time. The filled circles represent the cumulative distribution for 1X, the empty circles represent the cumulative distribution for ¼X. Both cumulative distributions use the right vertical axis. The cumulative distributions were fit with the double saturating exponential fits (Eq. 5.4). (Inset). The cumulative distributions for low times. The axes are the same as those in the larger plot. The filled (empty) circles represent the data and fits to the 1X (¼X) distribution, the empty circles the 1X. The linear fits do not include the (0,0) point. The number of points included in the low time distribution was determined subjectively, but included as many points as maintained linearity.

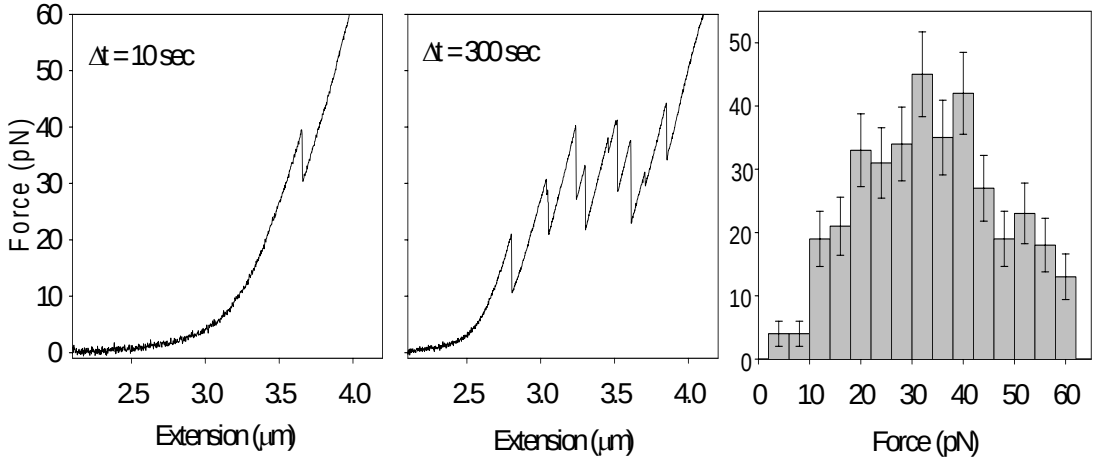

Fig. 5.7. Example Force-Extension Curves for Different Incubation Times. (Left) A pull after a 10s incubation. (Center) A pull after 300s. Note the large number of peaks for the 300s curve relative to the 10s. Each peak represents a loop being pulled apart and has an increase in tether length associated with it, ΔL. Events with very small ΔL, such as the one at ~ 3 μm on the 300s curve, were sometimes difficult to discern; this introduces a small amount of uncertainty into N for each curve. Due to differences in tether location on the pipette bead, the identical DNA molecules have different effective L_0 values (~3.3 μm for the 10s curve, ~3.5 μm for the 300s curve), as mentioned in the text. The maximal contour length of the DNA template is 3.69 μm. For each molecule, the fractional extension at which the DNA was incubated in the enzyme was $x/L_0 = 0.5$. The DNA-enzyme complexes were pulled to 60 pN in order to obtain the original L_0. (Right). A distribution of loop disruption forces. The vertical axis is number of observed events.

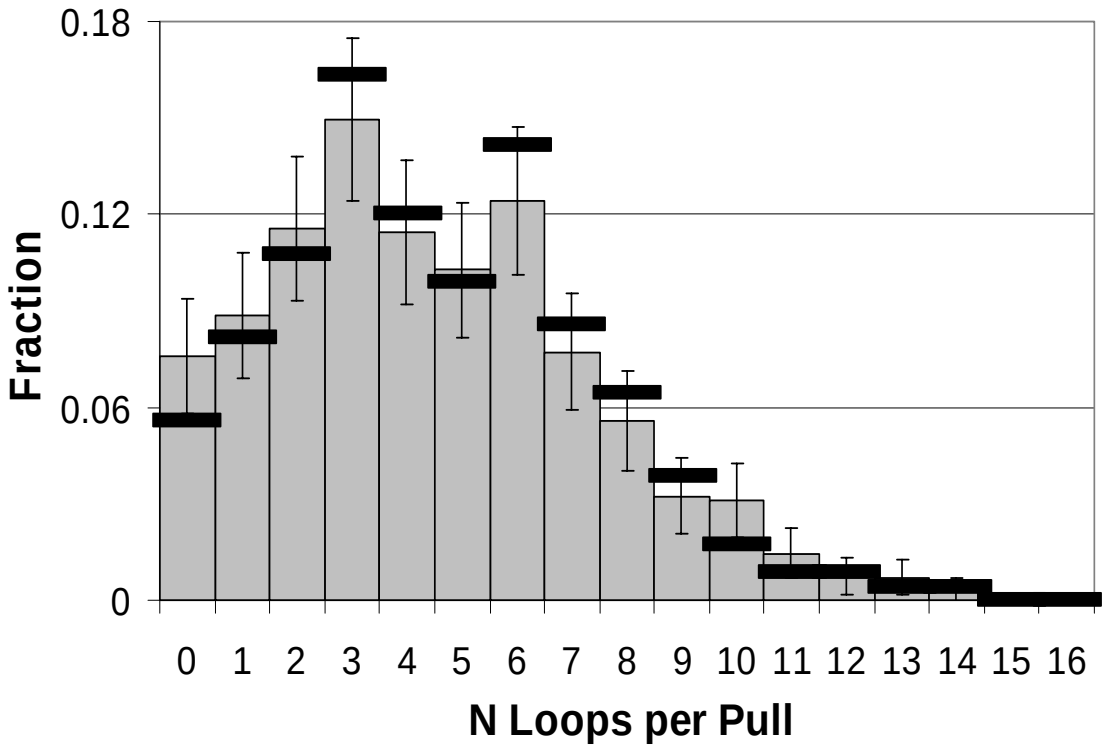

Fig. 5.8. Distribution of N Loops per Pull for 120s. The heavy black bars represent the uncorrected distribution, whereas the gray bars represent the distribution corrected according to Eq. 5.5. The error bars on the corrected distribution were calculated as the (Corrected Distribution(N)/Total Number of Pulls)^{½}. Because there can only be false events (and not the false lack of events), the distribution is generally shifted to lower N.

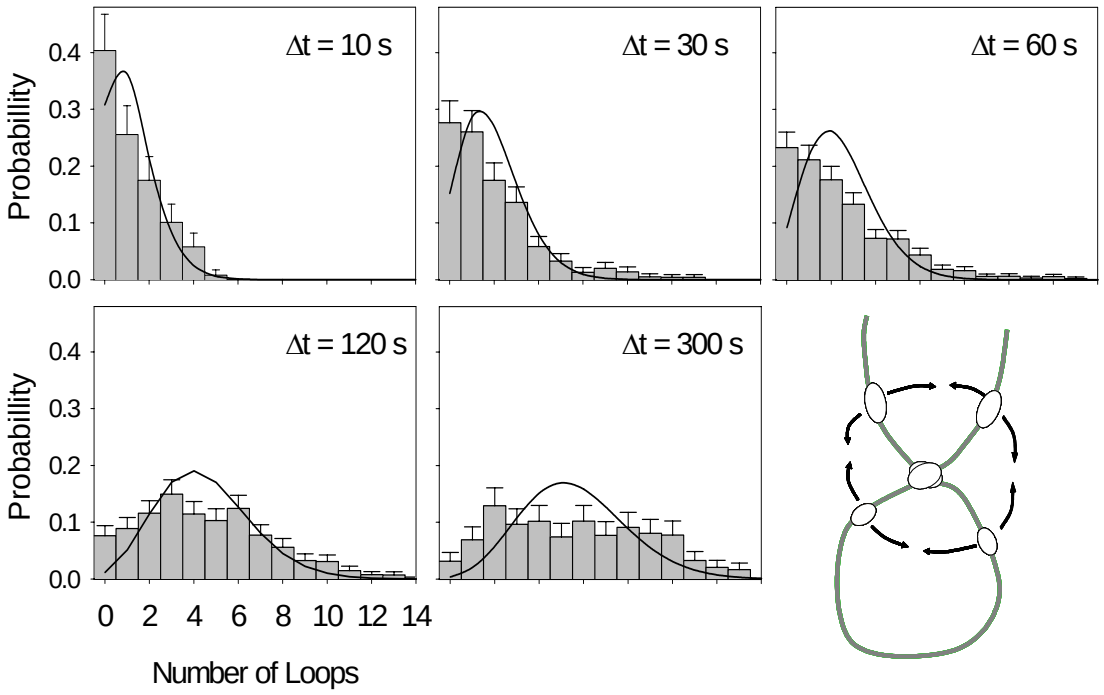

Fig. 5.9. N Distributions and the theoretical Poisson curves for the incubation time data sets with t_{inc} denoted in each. The lines in each represent the Poisson curves, $y(j) = e^{-<N>}<N>^{-j}/(j!)$, which were calculated using the measured expectation values of $<N> = 1.18$, 1.88, 2.38, 4.46, and 5.60, respectively. The error bars were calculated according to $(N_{pulls}(i))^{½}/ N_{Total}$. The cartoon in the lower right demonstrates cooperativity in loop formation. The gray line represents the DNA, the white ovals indicate *Sau3AI* monomers, and the black arrows represent the possible facilitated loop formations. The way in which the enzyme dimer is drawn is not intended to reflect the geometry with which *Sau3AI* actually dimerizes.

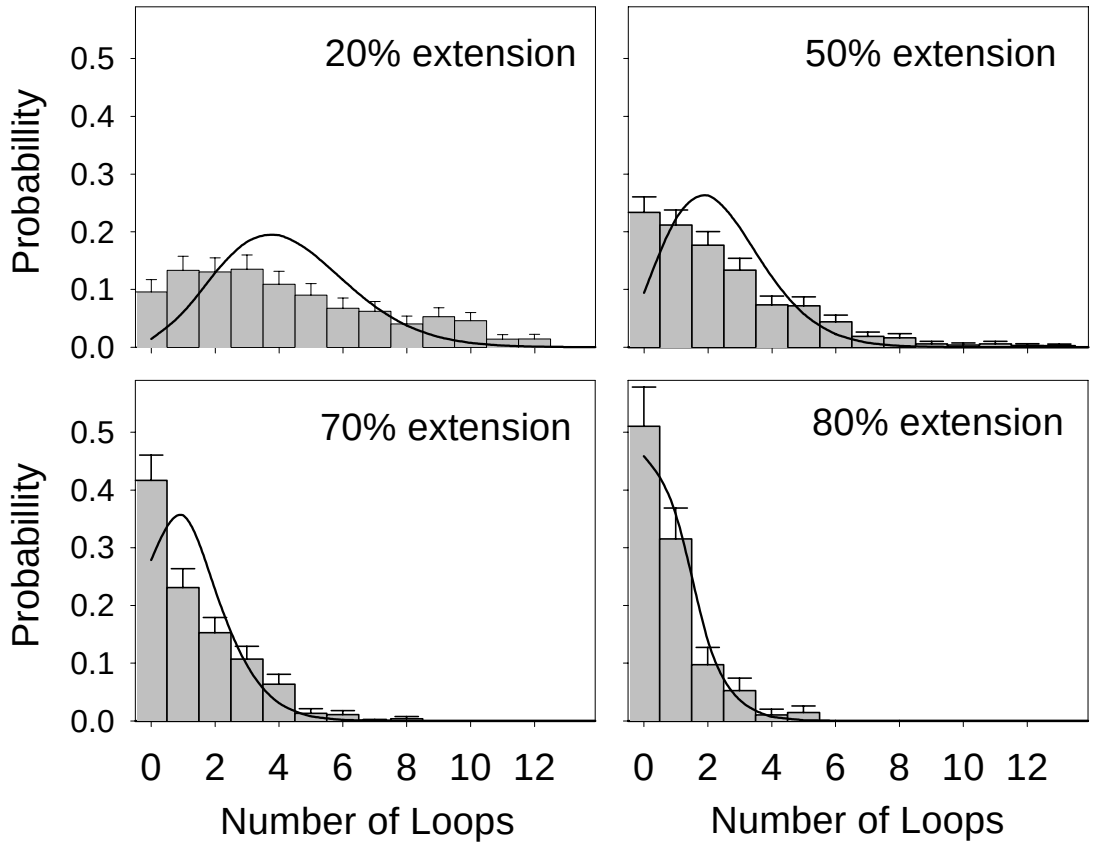

Fig. 5.10. N Distributions for the Fractional Extension data. In each graph the horizontal axis is the number of loops per pull and the vertical axis is the fraction of the total number of pulls at each experimental condition. The upper left is the N distribution for $x/L_0 = 0.2$, the upper right for 0.5, the lower left for the 0.7, and the lower right for 0.8. The smooth black lines represent the Poisson curves which were calculated using the calculated expectation values of $\langle N \rangle = 4.24$, 2.36, 1.28, and 0.78, respectively. The error bars were calculated according to $(N_{pulls}(i))^{1/2} / N_{Total}$.

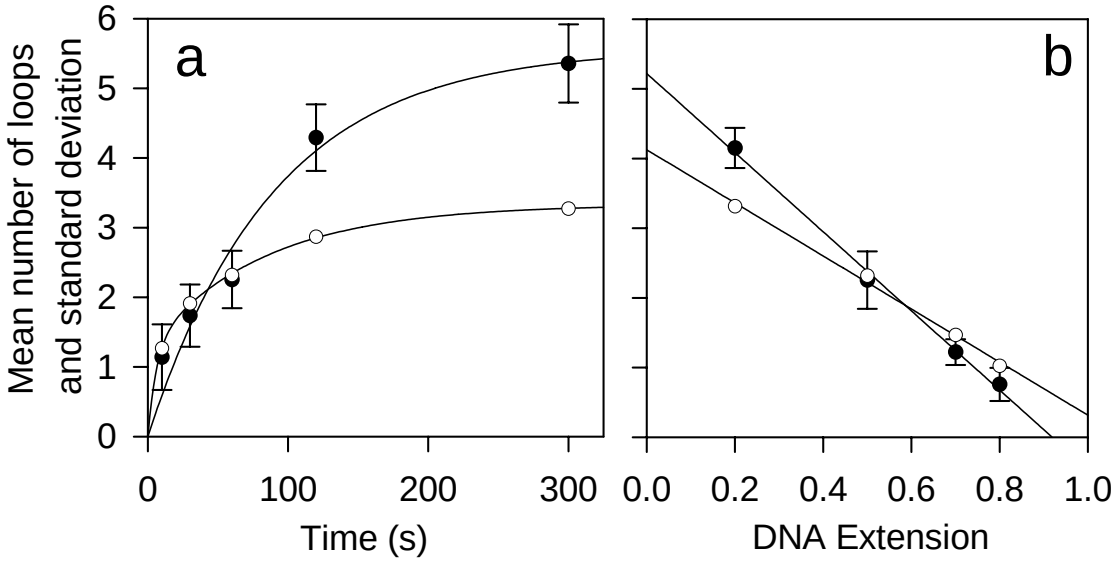

Fig. 5.11. (a) The N distribution statistics versus incubation time. The filled circles represent the expectation value for number of loops, <N>, and the empty circles represent the second moment of the N distributions, σ_N. The uncertainties on <N> were calculated as the square root of the sum of the squares of the uncertainties from the N distribution, or $\sigma_{<N>} = (\Sigma\sigma_{P(i)}^2)^{1/2}$. The line represents a fit to the <N> data of the form $y = a(1-e^{-kt})$. The σ_N data were fit with a saturating double exponential. (b) The N distribution statistics versus fractional extension. The filled circles represent the expectation value for number of loops, <N>, and the empty circles represent the second moment of the N distributions. The <N> uncertainties were calculated in the same way as for the time data sets. The fits to the data are both linear, and the parameters are discussed in the text.

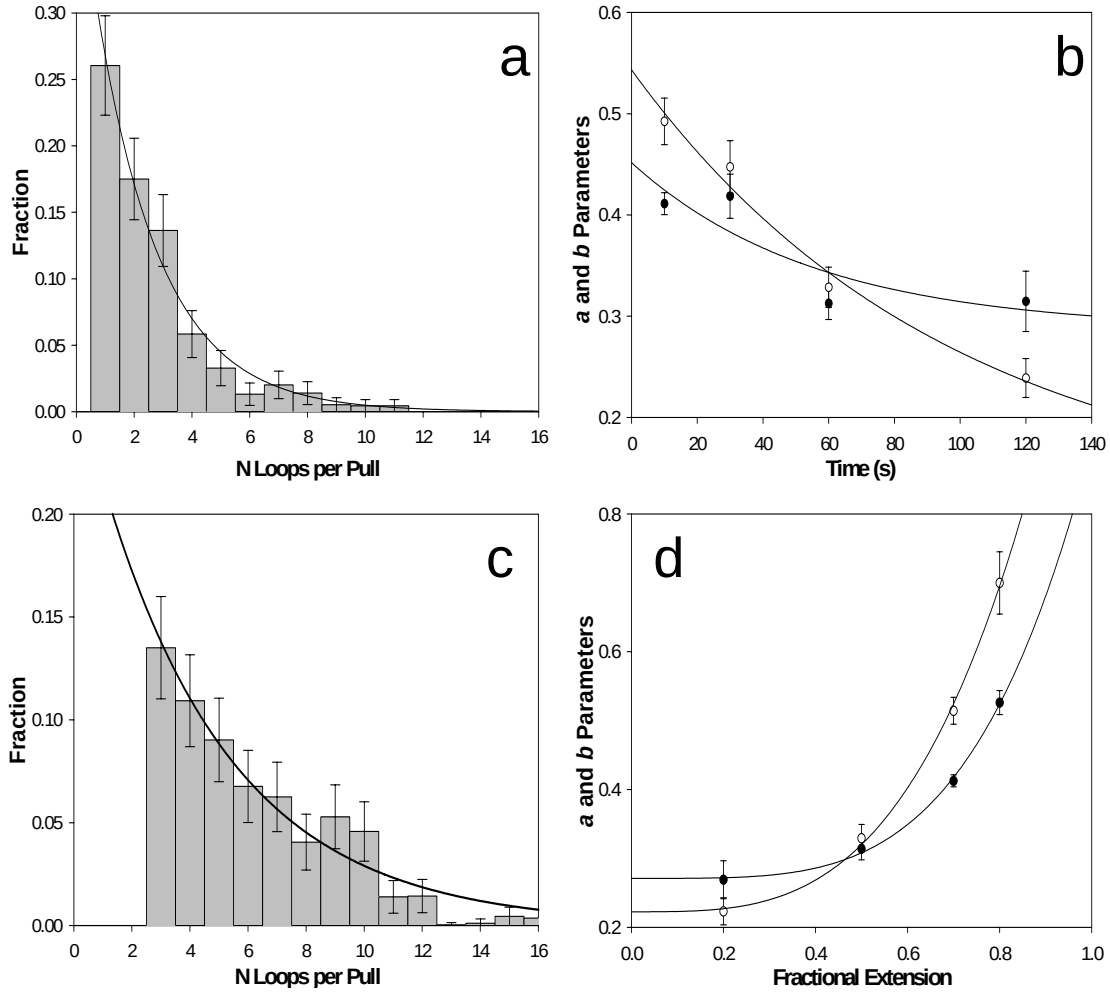

Fig. 5.12. (a) The tail section (N ≥ 1) for the t_{inc} = 30s N distribution with a decaying exponential fit, $y = ae^{-bN}$. (b) The exponential fit parameters as a function of t_{inc}. The black circles represent the *a* parameter, while the empty circles represent the *b* parameter. The respective line fits, solid line for *a* and dotted line for *b*, are just visual aids intended only to capture the general trend of the data, especially as $t_{inc} \rightarrow 0$ or $t_{inc} \rightarrow \infty$. As discussed in the text, there was no discernable well fit tail section to the 300s distribution. (c). The $x/L_0 = 0.2$ N distribution with the tail section (N ≥ 3) fit with an exponential, $y = ae^{-bN}$. The fit parameters for this example are a = 2.69e-1 ± 2.76e-2 and b = 2.23e-1 ± 1.97e-2. The selected tail ranges are discussed in the text. (d). The exponential fit parameters versus x/L_0. The black circles represent the *a* parameter, while the empty circles represent the *b* parameter. Both the black and dotted lines are of the form $y = y_0 + c(x/L_0)^d$. For the *a* parameter (solid black line), these fit values are $y_0 = 0.27 \pm 0.01$, $c = 0.63 \pm 0.07$, and $d = 4.09 \pm 0.48$. For the *b* parameter, $y_0 = 0.22 \pm 0.02$, $c = 1.00 \pm 0.09$, and $d = 3.36 \pm 0.38$.

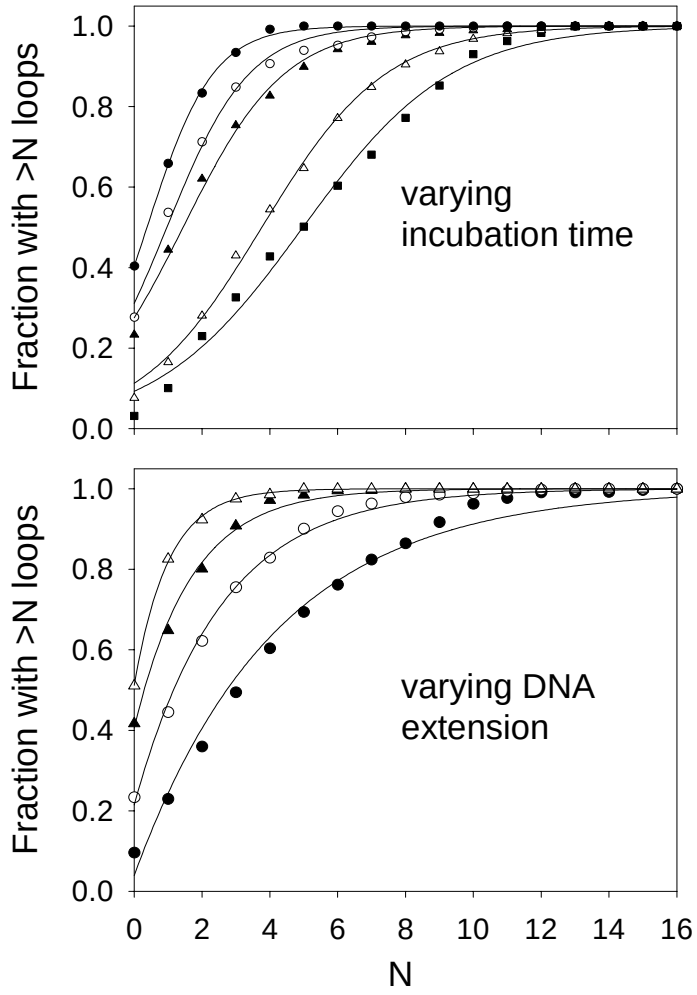

Fig. 5.13. Cumulative N Distributions. (Top) The t_{inc} = 10s, 30s, 60s, 120s, and 300s data sets are represented by the filled circles, empty circles, filled triangles, empty triangles, and filled squares, respectively. (Bottom) The x/L_0 = 0.2, 0.5, 0.7, and 0.8 data sets are represented by the filled circles, empty circles, filled triangles, and empty triangles, respectively. The lines in both represent the fits according to Eq.5.6.

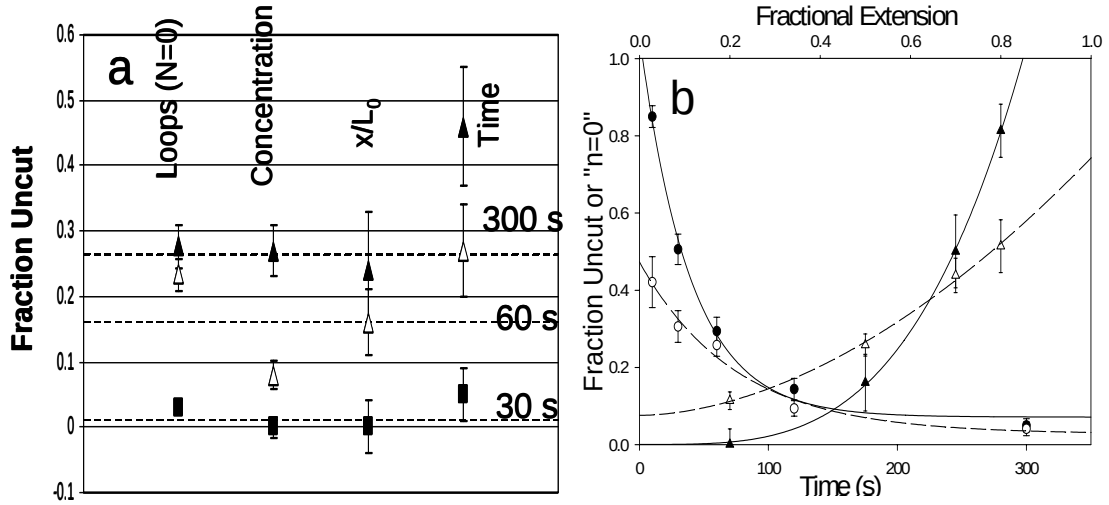

Fig. 5.14. Comparison of Cutting Data to Peaks Data. (a) The results by experiment, from left to right, for the looping done at $x/L_0 = 0.5$ and 1 min (N=0), and the fraction uncut at the same conditions for the enzyme concentration, x/L_0, and timed-pulsing cutting experiments. The filled triangles represent the fraction with N = 0 or uncut after 30s, the empty triangles after 60s, and the filled squares after 300s. The uncertainty on each cutting fraction is described previously; the uncertainty on the N = 0 fraction is that of the particular distribution presented in Figs. 5.9 and 5.10. The respective lines represent the average for all but the timed pulsed cutting data points, which are systematically different. Those averages are 0.262 ± 0.029 (30s), 0.158 ± 0.024 (60s), and 0.010 ± 0.014 (300s), where the uncertainties are the standard deviation of the three points entering the average. (b) Comparison of the uncut (filled circles, solid line) and the N = 0 (empty circles, dashed line) for the time experiments, which use the lower horizontal axis. The comparison of the uncut (filled triangles, solid lines) and the N = 0 (empty triangles, dashed lines) for the fractional extension experiments, which use the upper horizontal axis. The lines are visual aids only.

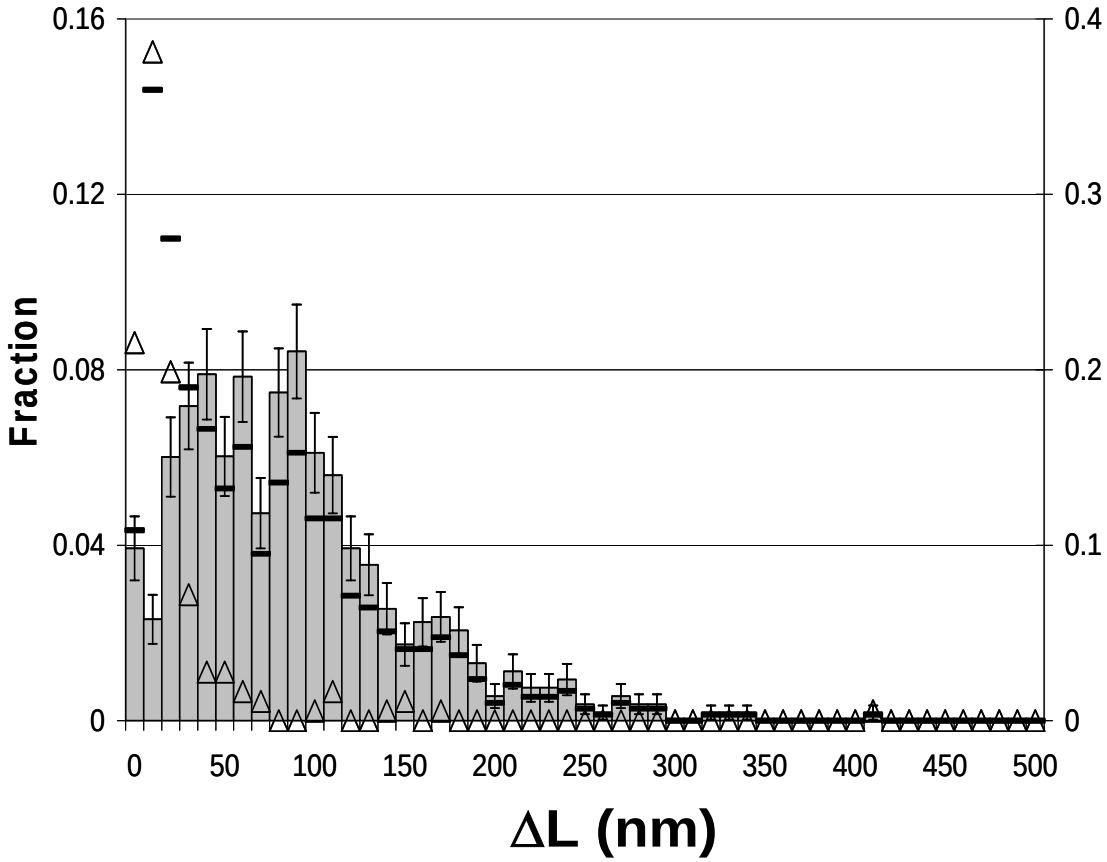

Fig. 5.15. Normalized ΔL Distribution for the 300s data set. The heavy black lines represent the measured ΔL distribution (10 nm bins) with no correction for false events. The empty triangles represent the normalized ΔL distribution for the false events and utilize the secondary y-axis on the right side of the plot (e.g., ~ 38% of runs have one false event, not ~15%). The gray bars represent the corrected ΔL distribution using the false ΔL distribution, the N distribution of false events, and Eq. 5.7. The error bars are calculated by as $(P(\Delta L)/N_{\text{Total Events}})^{1/2}$.

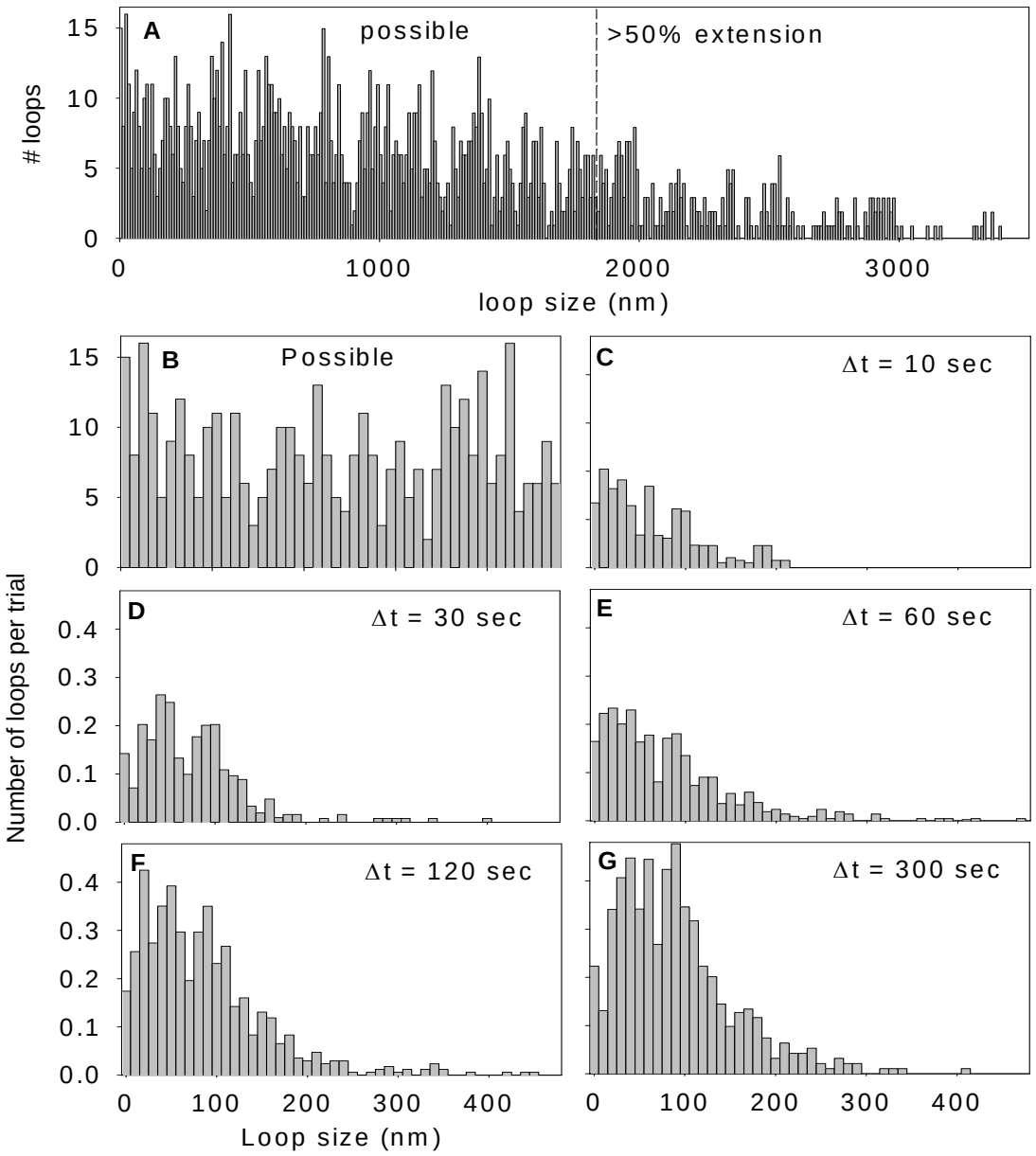

Fig. 5.16. Distribution of Number of Loops as a Function of Loop Length for the Incubation Time Data Sets. (a) Distribution of all possible loop lengths for the DNA template with 55 sites. The binning is 10 nm, as is the case for the experimental data. The dashed line represents 50% of the nominal contour length of the DNA, a length above which loops are geometrically impossible since we hold the DNA at x/L0 = 0.5 for the time data sets. (b) The distribution of possible loop lengths up to 480 nm, which encompasses the longest loop we observed. (c) through (g). The number of loops per length distributions for the 10s, 30s, 60s, 120s, and 300s data sets, each normalized as described in the text. Plot (c) uses the same vertical axis of plots (d) through (g).

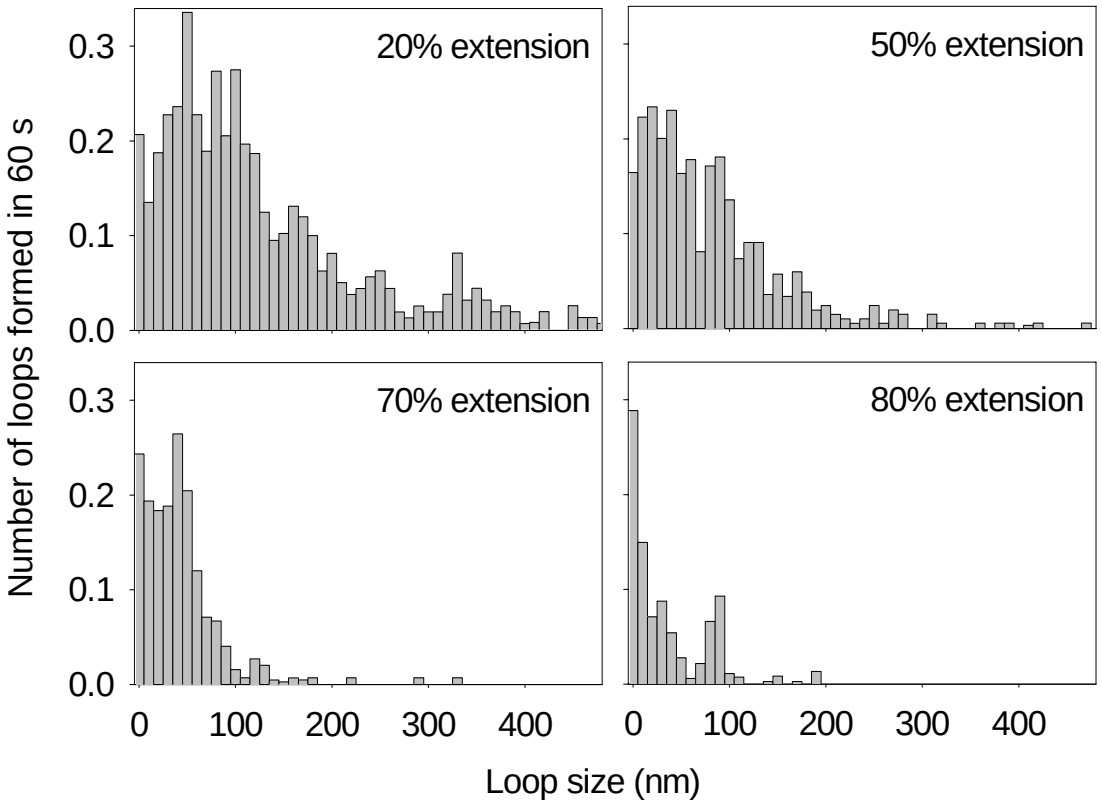

Fig. 5.17. Distribution of Number of Loops as a Function of Loop Length for the fractional extension data sets. Each molecule was held for 1 min at the given fractional extension. As with the time data sets, each distribution is normalized by procedure described in the text.

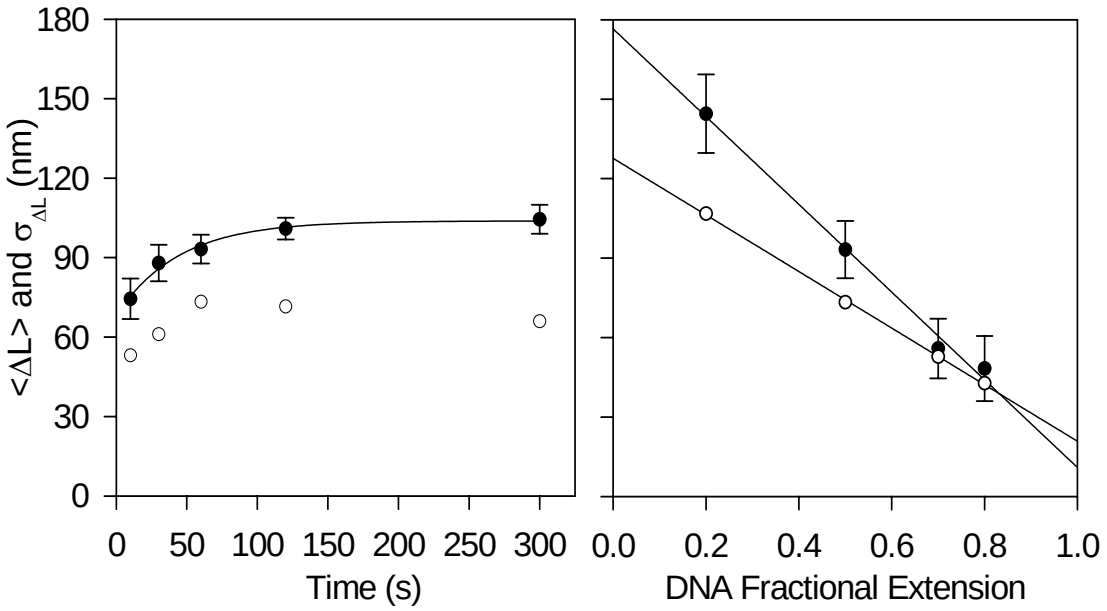

Fig. 5.18. (a) The expectation value <ΔL> for each distribution as a function of incubation time (black circles), with the uncertainties given by $(\Sigma\ (\Delta L \sigma(\Delta L))^2)^{1/2}$. The empty circles represent the second moments of the distributions, as discussed in the text. The line represents an offset saturating exponential fit to the <ΔL> values. (b) The expectation values <ΔL> for each distribution as a function of fractional extension are represented by filled circles, the second moments by empty circles. The lines represent linear fits, and the parameters thereof are discussed in the text.

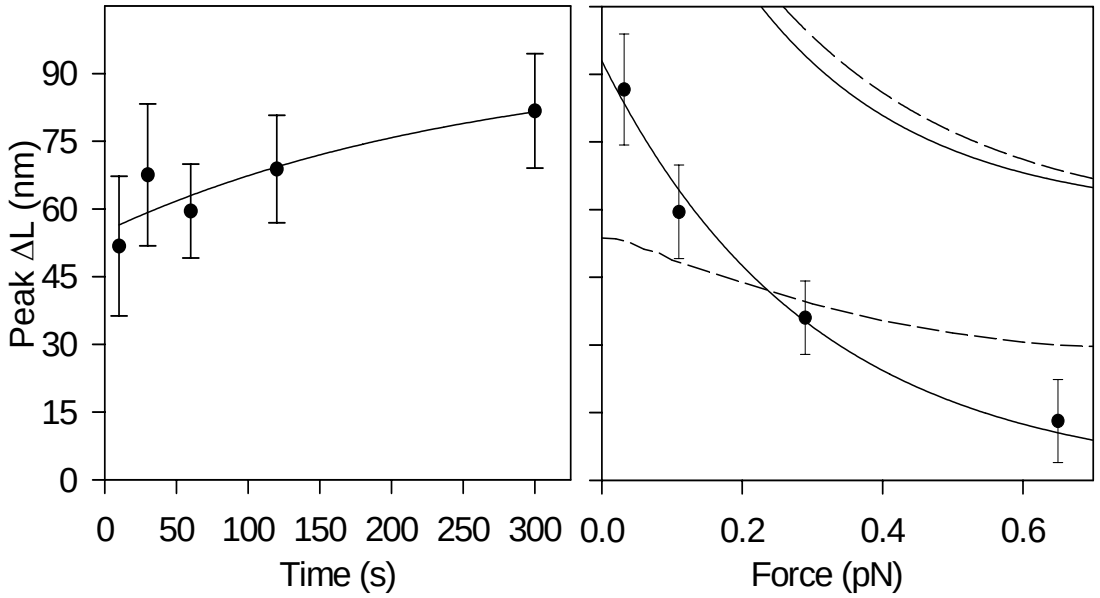

Fig. 5.19. The peaks of the ΔL distributions as a function of incubation time (a) and entropic tension (b). Error bars are the reported uncertainties from the peak fit parameters. The line on the time graph represents the fit of the saturating exponential function. The solid line through the fractional extension data represents the fit of the decaying exponential function. The dashed line near the data represents the theoretical results for the π/2-kink calculations inferred from Figure 8 of Ref. 39. The solid line in the upper right corner represents the theoretical prediction for most probable loop size as a function of force with no defects from Blumberg, *et al*. (41), whereas the dashed line in the upper right corner represents the comparable prediction of S. Sankararaman and J.F. Marko (39). The graph is cast in terms of entropic tension instead of fractional extension (the experimentally controlled variable) for easier comparison with those calculations for the most probable loop size.

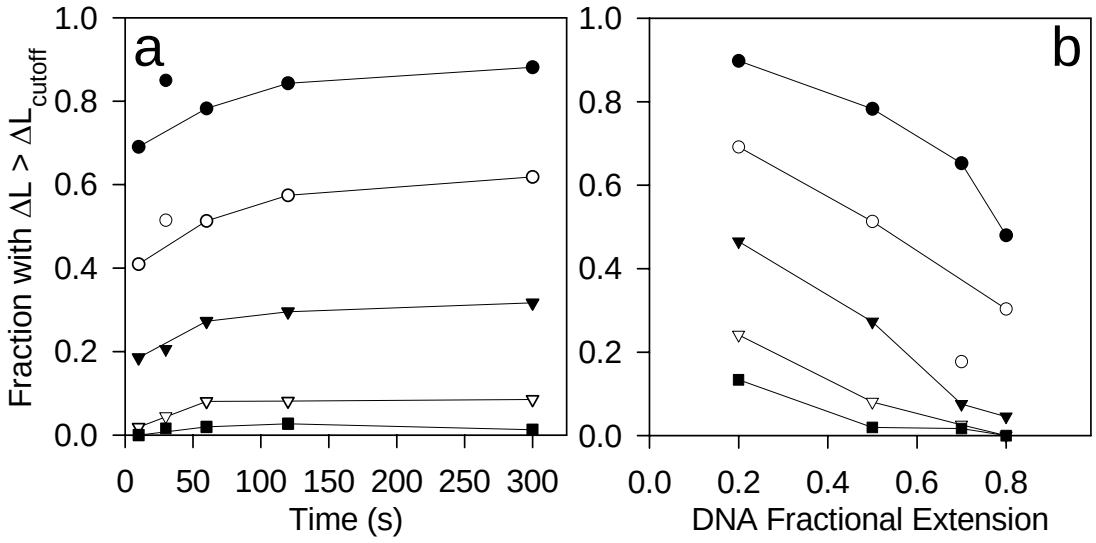

Fig. 5.20. (a) As a function of incubation time, the fraction of observed loops with ΔL > 40 (P), 80 (2P), 120 (3P), 200 (5P) and 300 (7.5P) nm (filled circles, empty circles, filled triangles, empty triangles, and squares, respectively). The lines are intended as visual aids and do not pass through the 30s points for every ΔL_{cutoff}, as they lie well off the trend for each ΔL_{cutoff}. (b) As a function of fractional extension, the fraction of events with ΔL > 40, 80, 120, 200, and 300 nm. The 0.7 point for the ΔL > 80 nm data is not included in the visual aid line.

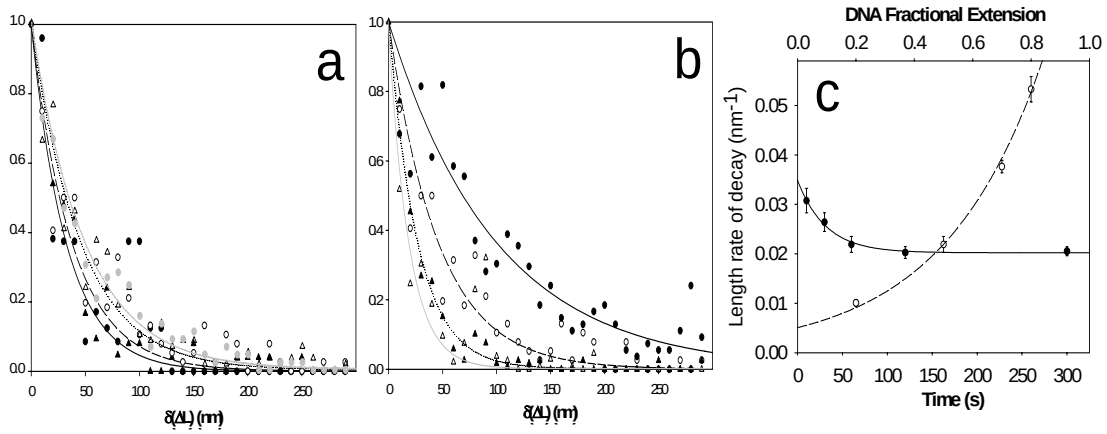

Fig. 5.21. Normalized and Shifted ΔL Tails with decaying exponential fits. (a) The time distributions and their associated fits are plotted as 10s (black circles, solid black line), 30s (black triangles, dashed black line), 60s (empty circles, dotted black line), 120s (empty triangles, solid gray line), and 300s (gray circles, dotted gray line). (b) The fractional extension distributions and their associated fits are plotted as 0.2 (black circles, solid black line), 0.5 (empty circles, dashed black line), 0.7 (black triangles, dotted black line), and 0.8 (empty triangles, solid gray line). (c) The *b* parameter (nm^{-1}) of the fits $y = e^{-b\delta(\Delta L)}$. The parameters from the time distributions are solid black circles and use the time axis (lower), those of the fractional extension fits are the empty circles and use the upper horizontal axis. The lines through the parameters are intended as visual aids, and the zero-force theory prediction for a π/2-kinked loop is 0.02 nm^{-1}, as noted in the text.

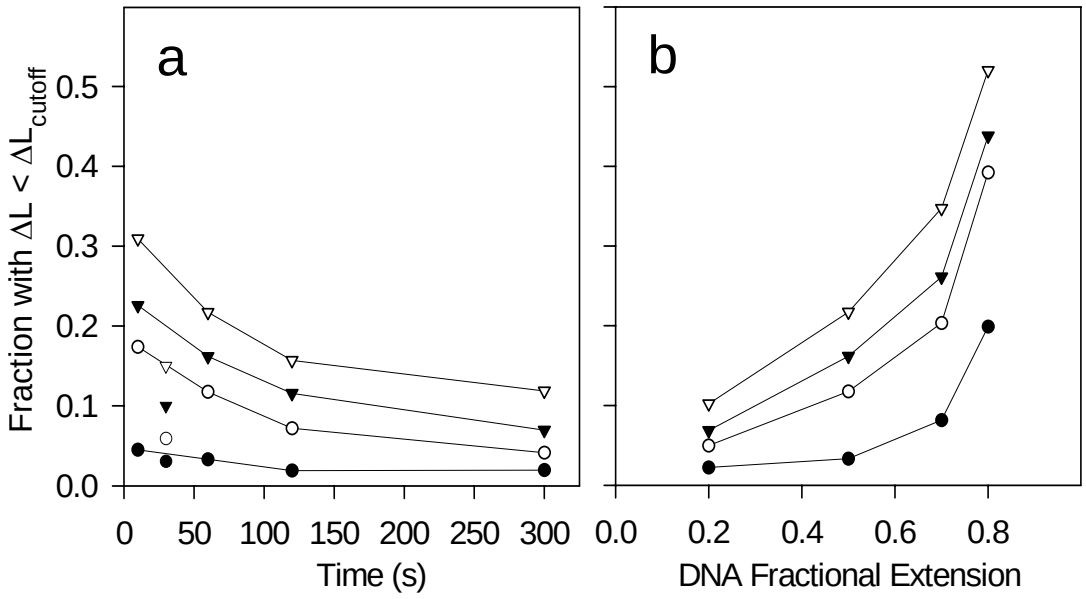

Fig. 5.22. (a) As a function of incubation time, the fraction of observed loops with ΔL < 10, 20, 30, and 40 nm (filled circles, empty circles, filled triangles, and empty triangles, respectively). The lines are intended as visual aids and do not pass through the 30s points, as they lie well off the trend for each ΔL. (b) As a function of fractional extension, the fraction of events with ΔL < 10, 20, 30, and 40 nm.

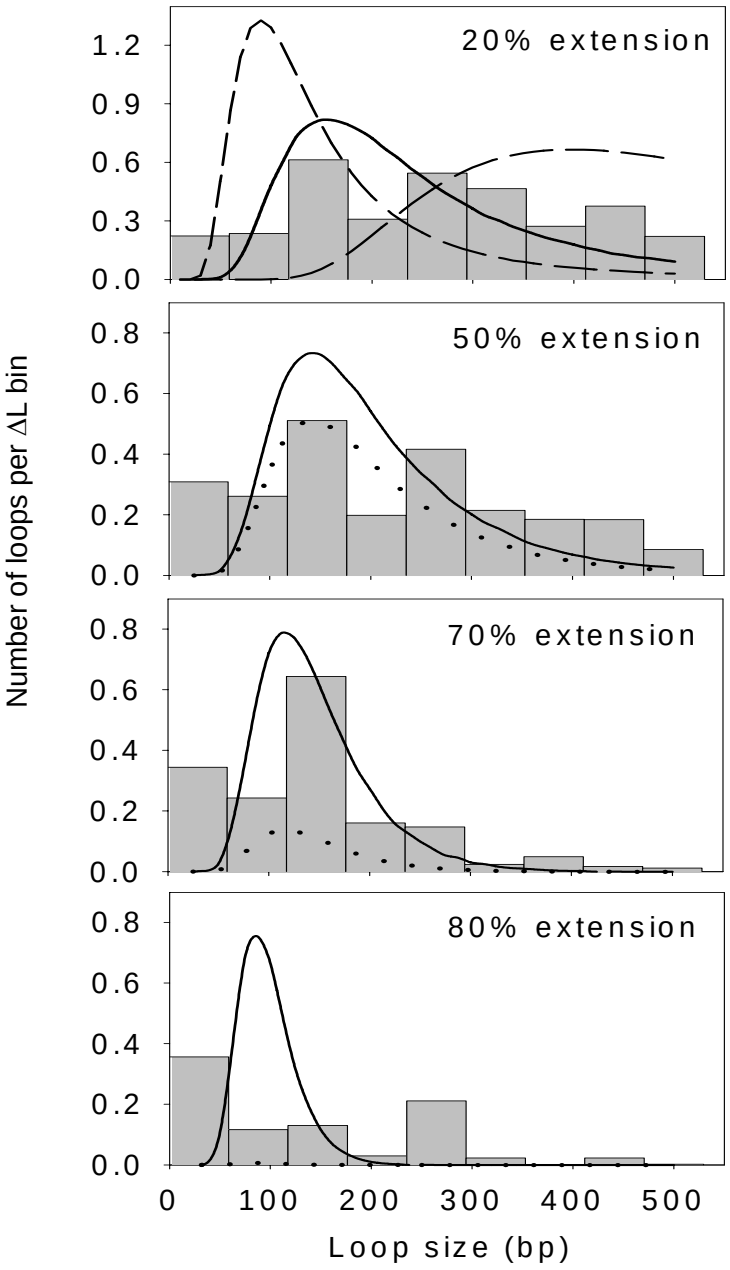

Fig. 5.23. The distribution of number of loops as a function of ΔL for the DNA fractional extension data sets. The solid line in each is the probability distribution at the closest force represented in Fig. 7 of Ref. 39 (0.04, 0.15, 0.3, and 0.7 pN). Each curve is scaled to have the same mean as the observed distribution. The dotted lines are the same curves, but scaled to maintain their original size relative to that of the solid line in the top panel. The dashed lines in the top panel are the force free probability distributions calculated from Eq. 8 from Ref. 39 for the π/2-kinked loop (left) and teardrop loop (right) with P = 40 nm.

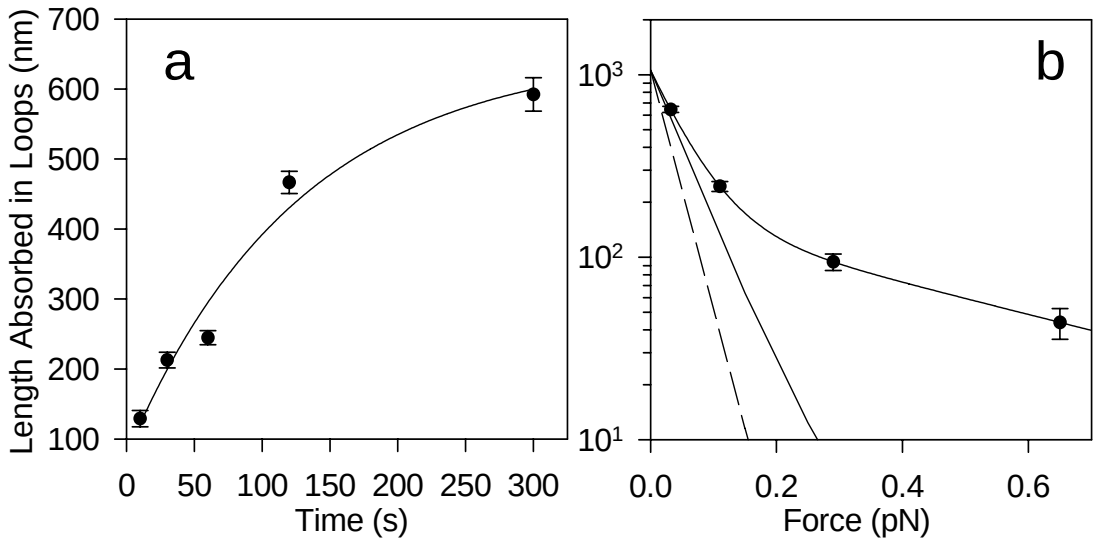

Fig. 5.24. The loop length absorption for the (a) time and (b) fractional extension data sets. The loop length absorption is the expected amount of DNA incorporated into loops. These loops per ΔL distributions were formed from the normalized ΔL distributions by the scaling described in the text. On the left, the line through the data is intended as a visual aid. In (b), the horizontal axis is expressed in terms of entropic tension (pN) and the vertical axis is logarithmic base-10 for more direct comparison with Fig. 10 of Ref. 39. The line through the data (a decaying double exponential) implies two force scales, in accord with theoretical predictions; the solid line is the theoretical prediction for a kinked loop, and the dashed line is the theoretical prediction for a teardrop loop. The theoretical lines are normalized so that the f = 0 projection of each equals the f = 0 projection of our data.

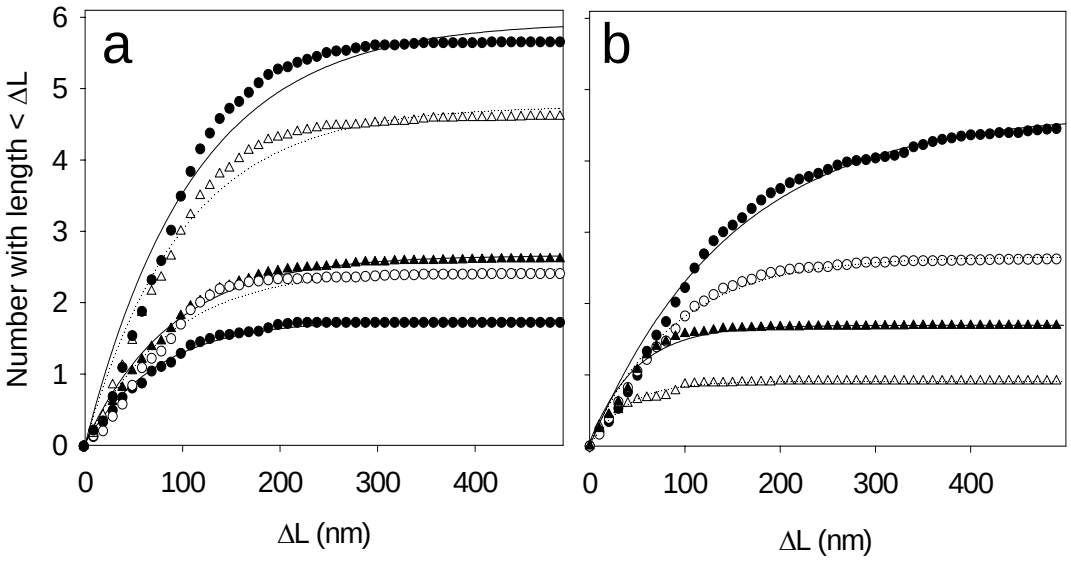

Fig. 5.25. The cumulative N Loops per ΔL vs ΔL distributions for (a) the time and (b) fractional extension data sets, shown with saturating exponential fits. In (a), the symbol line representations are circles-solid lines (10s), empty circles-dotted line (30s), triangles -solid line (60s), empty triangles-dotted line (120s), and squares-solid line (300s). In (b), the symbol line representations are circles-solid lines (0.2), empty circles-dotted line (0.5), triangles-solid line (0.7), and empty triangles-dotted line (0.8). Note that increasing the time results primarily in an increase in the number of loops formed, such that the curves can be considered vertical scalings of each other. Decreasing the fractional extension shifts the distribution towards higher ΔL.

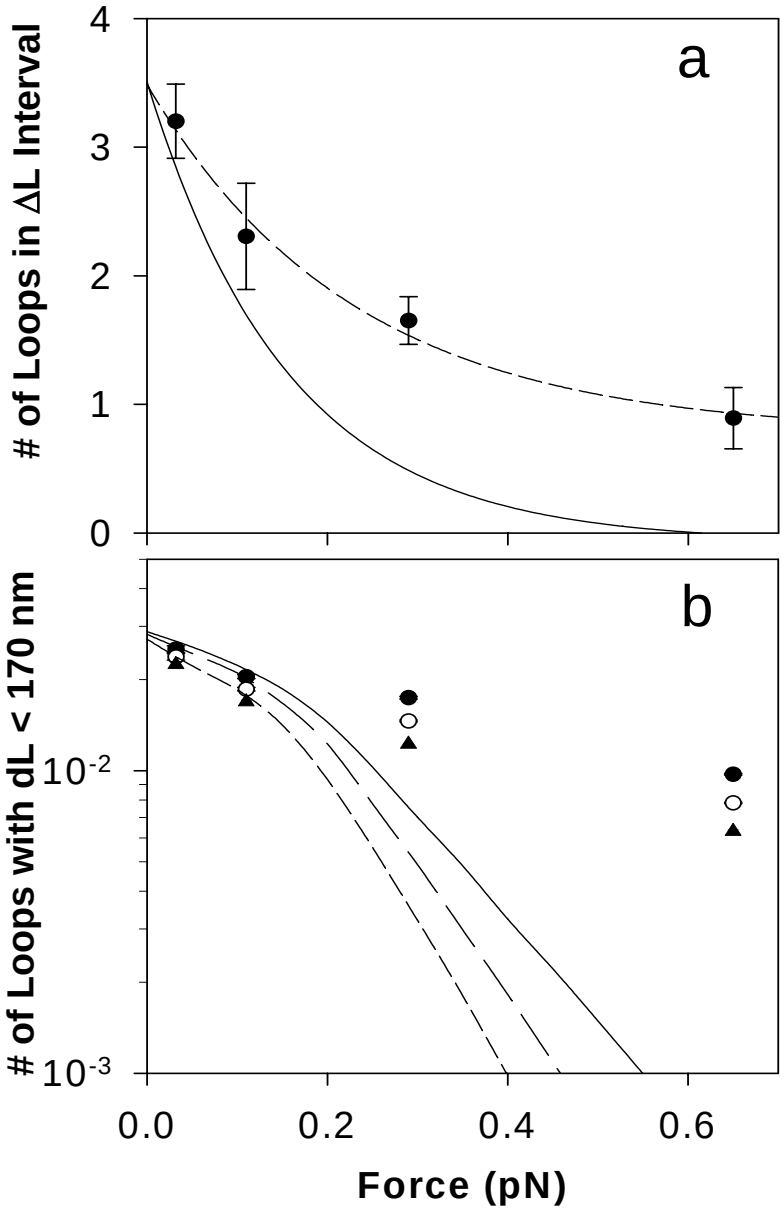

Fig. 5.26. (a) The number of loops (ΔL < 170 nm) as a function of force from the cumulative loops per ΔL distributions in Fig. 5.25. For simplicity, the error bars are the same as those for <N> in Fig. 5.11. The dashed line is a visual aid and has an f = 0 projection of ~ 3.5 loops. The solid line is the interpolated trend of the areas under the probability curves for a π/2-kinked loop in Fig. 7 of Ref. 39. The line is scaled to have the same f = 0 projection as our data. (b) The loops per ΔL distribution values as a function of force for 100 bp (filled circles), 125 bp (empty circles), and 150 bp (triangles). The solid, long dashed, and short dashed lines are the theoretical predictions for formation probability for a π/2-kinked loop of 100, 125, and 150 bp, respectively. Each line is scaled to have the same f = 0 projection as our data.

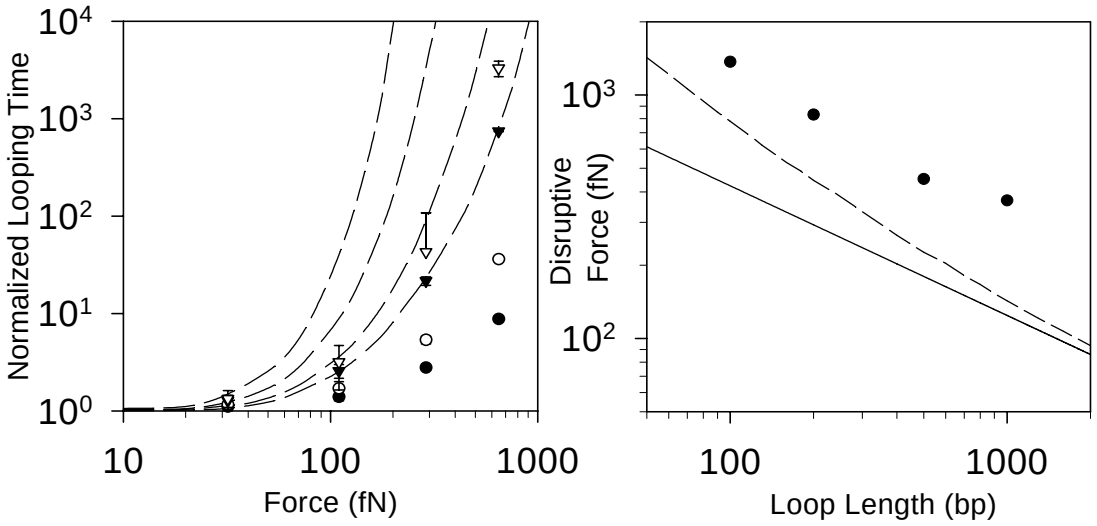

Fig. 5.27. (a) The normalized looping time as a function of DNA tension. Normalized looping time is defined so as to be unity at zero force. The filled circles represent the data for a 100 bp loop, the empty circles represent the data for a 200 bp loop, the filled triangles represent the data for a 500 bp loop, and the empty triangles represent the data for a 1 kbp loop. These values were determined as described in the text. The dashed lines represent the calculated normalized looping times for a hairpin loop of, left to right, a 1 kbp, 500 bp, 200 bp, and 100 bp (41). (b) Disruptive tension as a function of loop length. The solid line represents the disruptive tension calculated for a hairpin loop, the dashed line for a circular loop (41). The data points were determined from extrapolations of the data points in the left graph.

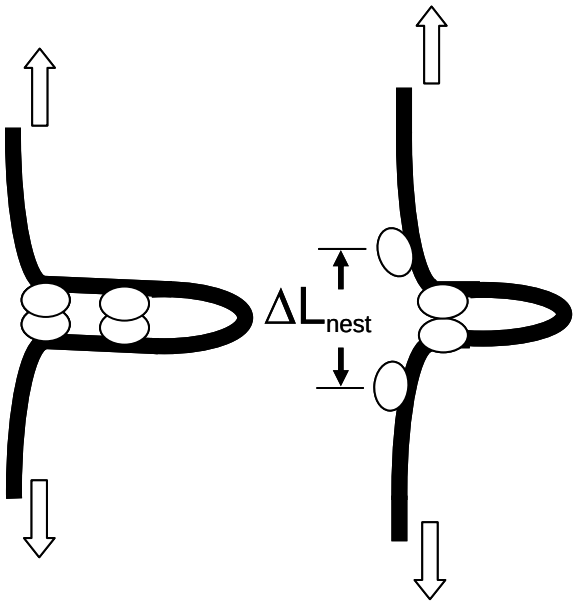

Fig. 5.28. Nested loops before (left) and after (right) being pulled apart. The empty arrows represent the applied disruptive force to the DNA by the optical tweezers, while the *Sau3AI* monomers are represented by the ovals. Upon being pulled apart, the measured released distance, ΔL_{nest}, represents the sum of the distances between the outer monomers and the inner dimer, and not the true distance between the sites. This diagram is not to imply that the dimers are necessarily pulled apart as such; the dimer may stay intact but be forced from the DNA under tension.

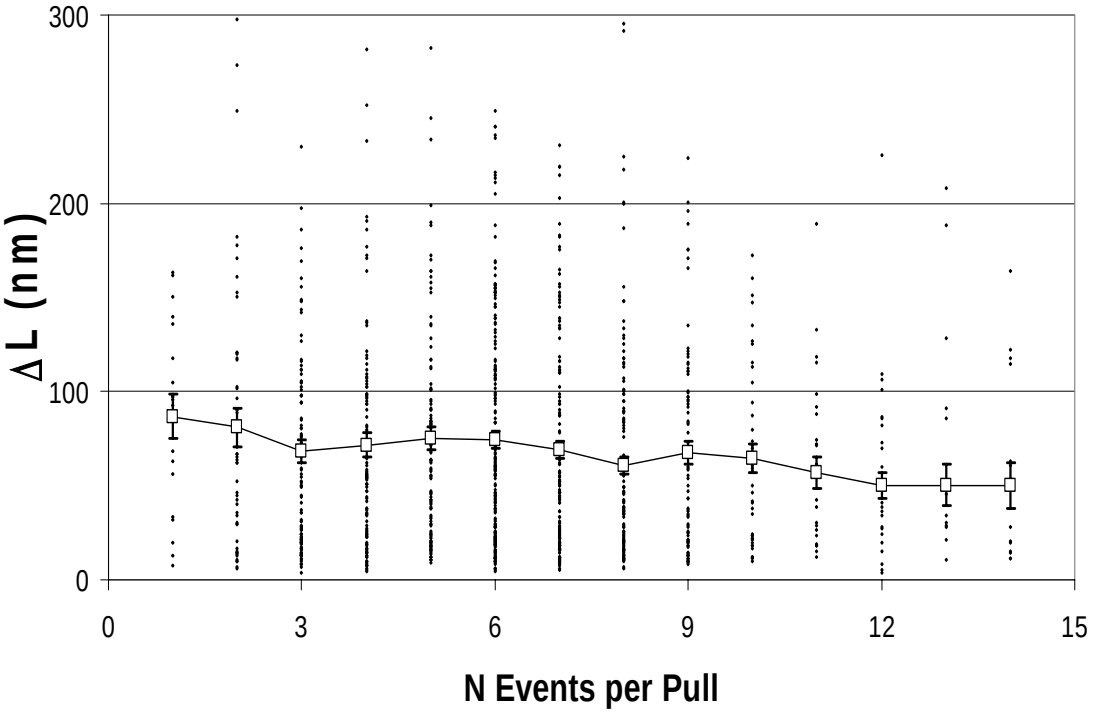

Fig. 5.29. The ΔL values plotted against their respective N Events per Pull for the 120s data set. The ΔL axis is shown only to 300 nm; there were a fair number of events longer than this. The small black points are the data pairs; the large empty squares are the averages with the outliers removed, and the error bars on them are the resulting standard errors. The solid line joining the points is a visual aid. The dotted line represents the <ΔL> for the entire 120s data set, 87 ± 4 nm. With the outliers removed, the linear correlation coefficient is $\rho \sim -0.10$. It should be noted that the aforementioned error correction for false events, which applies to the distributions and not individual runs, has not been applied to the groupings by N Events per Pull. There were not enough events at any particular N in the control set to form useful false ΔL sub-distributions.

References for Chapter 5

1. Schleif, R. (1992) DNA Looping. *Annu. Rev. Biochem*. **61**: 199-223

2. Page, M.I. & Jencks, W.P. (1971) Entropic Contributions to Rate Accelerations in Enzymic and Intramolecular Reactions and the Chelate Effect. *PNAS* **68**: 1678 – 1683

3. Vilar, J.M.G. & Leibler, S. (2003) DNA Looping and Physical Constraints on Transcription Regulation. *J. Mol. Biol.* **331**: 981-989

4. Halford, S.E. (2001) Hopping, Jumping, and Looping by Restriction Enzymes. *Biochem. Soc. Trans.* **29**: 363-73

5. Halford, S.E., Welsh, A.J., & Szczelkun, M.D. (2004) Enzyme-Mediated DNA Looping. *Annu. Rev. Biophys. Struct*. **33**: 1-24

6. Chen, Y. & Rice, P. (2003) New Insight into Site-Specific Recombination from FLP Recombinase-DNA Structures. *Annu Rev Biophys Biomol Struct.* **32**:135-159

7. Weiner, B.M. & Kleckner, N. (1994) Chromosome Pairing via Multiple Interstitial Interactions Before and During Meiosis in Yeast. *Cell* **77**: 977- 991

8. Su, W., Middleton, T., Sugden, B. & Echols, H. (1991) DNA Looping between the Origin of Replication of Epstein-Barr Virus and its Enhancer Site: Stabilization of an Origin Complex with Epstein-Barr Nuclear Antigen 1. *PNAS* **88**: 10870-10874

9. Mukherjee, S., Erickson, H., & Bastia, D. (1988) Enhancer-Origin Interaction in Plasmid R6K Involves a DNA Loop Mediated by Initiator Protein. *Cell* **52**: 375-383

10. Dunn, T.M., Hahn, S., Ogden, S., & Schleif, R. (1984) An Operator at -280 Base Pairs that is Required for Repression of araBAD Operon Promoter: Addition of DNA Helical Turns between the Operator and Promoter Cyclically Hinders Repression. *PNAS* **84**: 5017-5020

11. Oehler, S., Amouyal, M., Kolkhof, P., Wilcken-Bergmann, B., & Müller-Hill, B. (1994) Quality and position of the three *lac* Operators of *E. coli* Define Efficiency of Repression. *EMBO. J.* **13**: 3348 – 3355

12. Law, S., Bellomy, G., Schlax, P., & Record, M.T. (1993) *In vivo* Thermodynamic Analysis of Repression with and without Looping in *lac* Constructs: Estimates of Free and Local *lac* Repressor Concentrations and of

Physical Properties of a Region of Supercoiled DNA *in vivo*. *J. Mol. Biol*. **230**: 161-173

13. Blackwood, E.M. & Kadonaga, J.T. (1998) Going the Distance: A Current View of Enhancer Action. *Science* **281**: 60-63

14. Geanacopoulos, M., Vasmatzis, G., Zhurkin, V.B., & Adhya, S. (2001) *Gal* Repressosome Contains an Antiparallel DNA Loop. *Nat. Struct. Biol*. **8**: 432-436

15. Plumbridge, J. & Kolb, A. (1998) DNA Bending and Expression of the Divergent nagE-B Operons. *Nuc. Acids Res.* **26**: 1254-1260

16. Müller, J., Oehler, S., & Müller-Hill, B. (1996) Repression of *lac* Promotor as a Function of Distance, Phase, and Quality of an Auxiliary *lac* Operator. *J. Mol. Biol*. **257**: 21-29

17. Embleton, M.L., Siksnys, V., & Halford, S.E. (2001) DNA Cleavage Reactions by Type II Restriction Enzymes that Require Two Copies of Their Recognition Sites. *J. Mol. Biol*. **311**: 503-514

18. Milsom, S.E., Halford, S.E., Embleton, M.L., & Szczelkun, M.D. (2001) Analysis of DNA Looping Interactions by Type II Restriction Enzymes that Require Two Copies of Their Recognition Sites. *J. Mol. Biol.* **311**: 515-527

19. Mücke, M., Lurz, R., Mackeldanz, P., Behlke, J., Krüger, D.H., & Reuter, M. (2000) Imaging DNA Loops Induced by Restriction Endonuclease *EcoRII*. *J. Biol. Chem*. **275**: 30631-30637

20. Gormley, N.A., Hillberg, A.L., & Halford, S.E. (2002) The Type IIs Restriction Endonuclease *BspMI* is a Tetramer that Acts Concertedly at Two Copies of an Asymmetric DNA Sequence. *J. Biol. Chem*. **277**: 4034-4041

21. Friedhoff, P., Lurz, R., Lüder, G., & Pingoud, A. (2001) *Sau3AI*, a Monomeric Type II Restriction Endonuclease that Dimerizes on the DNA and Thereby Induces DNA Loops. *J. Biol. Chem*. **276**: 23581-23588

22. Shore, D., Langowski, J., & Baldwin, R.L. (1981) DNA Flexibility Studied by Covalent Closure of Short Fragments into Circles. *PNAS* **78**: 4833-4837

23. Crothers, D., Drak, J., Kahn, J., & Levene, S. (1992) DNA Bending, Flexibility, and Helical Repeat by Cyclization Kinetics. *Methods Enzymol*. **212**: 3-29

24. Krämer, H., Amouyal, M., Nordheim, A., & Müller-Hill, B. (1988) DNA Supercoiling Changes the Spacing Requirement of Two *lac* Operators for DNA Loop Formation with *lac* Repressor. *EMBO J*. **7**: 547-556

25. Hochschild, A., & Ptashne, M. (1986) Cooperative Binding of λ Repressors to Sites Separated by Integral Turns of the DNA Helix. *Cell* **44**: 681-687

26. Krämer, H., Niemöller, M., Amouyal, M., Revet, B., von Wilcken-Bergmann, B., & Müller-Hill, B. (1987) *Lac* Repressor Forms Loops with Linear DNA Carrying Two Suitably-Spaced *lac* Operators. *EMBO J.* **6**: 1481-1491

27. Mossing, M.C., & Record, M.T., Jr. (1985) Thermodynamic Origins of Specificity in the *lac* Repressor-Operator Interaction: Adaptability in the Recognition of Mutant Operator Sites. *J. Mol. Biol*. **186**: 295-305

28. Bustamante, C., Marko, J.F., Siggia, E.D. & Smith, S. (1994) Entropic Elasticity of lambda-phage DNA. *Science* **265**: 1599-1600

29. Smith, S., Finzi, L. & Bustamante, C. (1992) Direct Mechanical Measurements of the Elasticity of Single DNA Molecules by Using Magnetic Beads. *Science* **258**: 1122-1126

30. Hanke, A. & Metzler, R. (2003) Entropy Loss in Long-Distance DNA Looping. *Biophys. J.* **85**: 167-173

31. Shimada, J. & Yamakawa, H. (1984) Ring Closure Probabilities for Twisted Wormlike Chains: Application to DNA. *Macromolecules* **17**: 689-698

32. Rippe, K., von Hippel, P.H., & Langowski, J. (1995) Action at a Distance: DNA-Looping and Initiation of Transcription. *Trends Biochem. Sci*. **20**: 500-506

33. Wang, M.D., Yin, H., Landick, R., Gelles, J., & Block, S.M. (1997) Stretching DNA with Optical Tweezers. *Biophys. J.* **72**: 1335-1346

34. Podtelezhnikov, A.A. & Vologodskii, A.V. (2000) Dynamics of Small Loops in DNA Molecules. *Macromolecules* **33**: 2767-2771

35. Cloutier, T.E. & Widom, J. (2004) Spontaneous Sharp Bending of Double-Stranded DNA. *Mol. Cell*. **14**: 355-362

36. Du, Q., Smith, C., Shiffeldrim, N., Vologodskaia, M., & Vologodskii, A. (2005) Cyclization of Short DNA Fragments and Bending of the Double Helix. *PNAS* **102**: 5397-5402

37. Yan, J. & Marko, J.F. (2004) Localized Single-Stranded Bubble Mechanism for Cyclization of Short Double Helix DNA. *Phys*. *Rev*. *Lett*. **93**: 108108

38. Wiggins, P.A., Phillips, R., & Nelson, P.C. (2005) Exact Theory of Kinkable Elastic Polymers. *Phys*. *Rev*. *E*. **71**: 021909

39. Sankararaman, S. & Marko, J.F. (2005) Formation of Loops in DNA Under Tension. *Phys. Rev. E*. **71**: 021911

40. Yan, J., Kawamura, R., & Marko, J. F. (2005) Statistics of Loop Formation Along Double Helix DNAs. *Phys. Rev. E*. **71**: 061905

41. Blumberg, S., Tkachenko, A., & Meiners, J. (2005) Disruption of Protein-Mediated DNA Looping by Tension in the Substrate DNA. *Biophys J.* **88**:1692-1701

42. Katiliene, Z., Katilius, E., & Woodbury, N. (2003) Single Molecule Detection of DNA Looping by NgoMIV Restriction Endonuclease. *Biophys. J.* **84**: 4053 - 4061

43. Finzi, L. & Gelles, J. (1995) Measurement of Lactose Repressor-Mediated Loop Formation and Breakdown in Single DNA Molecules. *Science* **267**: 378-380

44. Lia, G., Bensimon, D., Croquette, V., Allemand, J., Dunlap, D., Lewis, D., Adhya, S., & Finzi, L. (2003) Supercoiling and Denaturation in *gal* Repressor /Heat Unstable Nucleoid Protein (HU)-Mediated DNA Looping. *PNAS* **100**: 11373-11377

45. Sakata-Sogawa, K., Kurachi, M., Sogawa, K., Fujii-Kuriyama, Y., & Tashiro, H. (1998) Direct Measurement of DNA Molecular Length in Solution Using Optical Tweezers: Detection of Looping Due to Binding Protein Interactions. *Eur. Biophys. J.* **27**: 55-61

46. Yan, J., Skoko, D., & Marko, J.F. (2004) Near Field Magnetic Tweezer Manipulation of Single DNA Molecules. *Phys. Rev. E.* **70**: 011905

47. Gemmen, G.J., Millin, R., & Smith, D.E. (2006). DNA Looping by Two-site Restriction Endonucleases: Heterogeneous Distributions for Loop Size and Unbinding Force. *Submitted to Nucleic Acids Research*.

48. Gemmen, G. J., Sim, R. , Haushalter, K.A., Ke, P.C., Kadonaga, J.T., & Smith, D.E. (2005) Forced Unraveling of Nucleosomes Assembled on Hetrogeneous DNA Using Core Histones, NAP-1, and ACF. *J. Mol. Biol*. **351**: 89-99

49. Bouchiat, C., Wang, M.D., Allemand, J.F., Strick, T., Block, S.M., & Croquette, V. (1999) Estimating the Persistence Length of a Worm-Like Chain Molecule from Force Extension Measurements. *Biophys. J.* **76**: 409-413

50. van den Broek, B., Noom, M.C., & Wuite, G.J.L. (2005) DNA-Tension Dependence of Restriction Enzyme Activity Reveals Mechanochemical Properties of the Reaction Pathway. *Nuc. Acids. Res*. **33**(8): 2676-2684

51. Yasmin, R., Yeung, K.T., Chung, R.H., Gaczynsak, M.E., Osmulski, P.A., & Noy, N. (2004) DNA-Looping by RXR Tetramers Permit Transcription Regulation "At a Distance". *J. Mol. Biol.* **343**: 327-338

52. Vipond, B., Baldwin, G., & Halford, S. (1995) Divalent Metal Ions at the Active Sites of the *EcoRV* and *EcoRI* Restriction Endonucleases. *Biochem.* **34**: 697-704

53. Vipond, B.I. & Halford, S.E. (1995) Specific DNA Recognition by *EcoRV* Restriction Endonuclease Induced by Calcium Ions. *Biochemistry* **34**: 1113-1119

54. Sankararaman, S. & Marko, J.F. (2005) Entropic Compression of Interacting DNA Loops *Phys. Rev. Lett*. **95**: 0708104

55. Kim, J.L., Nikilov, D.B., & Burley, S.K. (1993) Co-Crystal Structure of TBP Recognizing the Minor Groove of a TATA Element. *Nature* **365**: 521-527

56. Harpaz, Y., Gerstein, M., & Chothia, C. (1994) Volume Changes on Protein Folding. *Structure* **2:** 641-649

57. Wah, D.A., Bitinaite, J., Schildkraut, I., & Aggarwal, A.K. (1998) Structure of *FokI* has Implications for DNA Cleavage. *PNAS* **95**: 10564-10569

58. Soundararajan, M., Chang, Z., Morgan, R.D., Heslop, P., & Connolly, B.A. (2002) DNA Binding and Recognition by the IIs Restriction Endonuclease *MboII*. *J. Biol. Chem*. **277**: 887-895

59. Weber, I.T., McKay, D.B., & Steitz, T.A. (1982) Two Helix DNA Binding Motif of CAP Found in *lac* Repressor and *gal* Repressor. *Nuc Acids Res.* **10**: 5085-5102

60. Hodges-Garcia, Y., Hagerman, P.J., & Pettijohn, D.E. (1989) DNA Ring Closure Mediated by Protein HU. *J. Biol. Chem*. **264**: 14621-14623

61. Ringrose, L., Chabanis, S., Angrand, P.A., Woodroofe, C., & Stewart, A.F. (1999) Quantitative Comparison of DNA Looping *in vitro* and *in vivo*: Chromatin Increases Effective DNA Flexibility at Short Distances. *EMBO. J.* **18**: 6630-6641

62. Guéron, M. & Leroy, J. (1995) Studies of Base Pair Kinetics by NMR Measurement of Proton Exchange. *Methods Enzymol*. **261**: 383-413

63. Cloutier, T.E. & Widom, J. (2005) DNA Twisting Flexibility and the Formation of Sharply Looped Protein-DNA Complexes. *PNAS* **102**(10): 3645-3650

64. Lee, D.H. & Schlief, R. (1989) *In vivo* DNA Loops in *ara*CBAD: Size Limits and Helical Repeat. *PNAS* **86**: 476-480

65. Wang, M.D., Schnitzer, M.J., Yin, H., Landick, R., Gelles, J., & Block, S.M. (1998) Force and Velocity Measured for Single Molecules of RNA Polymerase. *Science* **282**: 920-907

66. Wuite, G.J., Smith, S.B., Young, M., Keller, D., & Bustamante, C. (2002) Single Molecule Studies of the Effect of Template Tension on T7 DNA Polymerase Activity. *Nature* **404**: 103-106

67. Maier, B., Bensimon, D., & Croquette, V. (2000) Replication by a Single DNA Polymerase of a Stretched Single-Stranded DNA. *PNAS* **97**: 12002-12007

68. Winkler, F.K, Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K., & Wislon, K.S. (1993) The Crystal Structure of EcoRV Endonuclease and of its Complexes with Cognate and Non-Cognate DNA Fragments. *EMBO. J.* **12**: 1781-1795

69. Rudnick, J. & Bruinsma, R. (1999) DNA-Protein Cooperative Binding through Variable-Range Elastic Coupling. *Biophys*. *J*. **76**: 1725-1733

70. Maniotis, A.J., Chen, C.S. & Ingber, D.E. (1997) Demonstration of Mechanical Connections between Integrins, Cytoskeletal Filaments, and Nucleoplasm that Stabilize Nuclear Structure. *PNAS* **94**: 849-854

Chapter 5 Acknowledgements

The text of Chapter 5, in part, appears in Gemmen, G.J., Millin, R. & Smith, D.E. (2006). Dynamics of Single DNA Looping and Cleavage by Restriction Endonuclease Sau3AI Measured with Optical Tweezers. Submitted to the *Biophysical Journal*. The dissertation author was the primary researcher and author for the research which forms the basis of this chapter.

Chapter 6. Forced Unraveling of Nucleosomes Assembled on Heterogeneous DNA Using Core Histones, NAP-1, and ACF

6.1. Abstract

Periodic arrays of nucleosomes were assembled on heterogeneous DNA using core histones, the histone chaperone NAP-1, and ATP-dependent chromatin assembly factor (ACF). These complexes were stretched with optical tweezers. Abrupt events releasing ~55–95 bp of DNA, attributable to the forced unraveling of individual nucleosomes, were frequently observed. This finding is comparable with a previous observation of 72 –80 bp unraveling events for nucleosomes assembled by salt dialysis on a repeating sea urchin 5 S RNA positioning element, but the unraveling force varied over a wider range (~5–65 pN, with the majority of events at lower force). Because ACF assembles nucleosomes uniformly on heterogeneous DNA sequences, as in native chromatin, we attribute this force variation to a dependence on DNA sequence. The mean force increased from 24 pN to 31 pN as NaCl was decreased from 100 mM to 5 mM. Spontaneous DNA re-wrapping events were occasionally observed in real time during force relaxation. The observed wide variations in the dynamic force needed to unravel individual nucleosomes and the occurrences of sudden DNA re-wrapping events may have an important regulatory influence on DNA directed nuclear processes, such as the binding of transcription factors and the movement of polymerase complexes on chromatin.

6.2. Introduction

In the nucleus of eukaryotic cells, DNA is assembled with proteins into a periodic, highly folded complex referred to as chromatin (1). This compaction allows

human cells to fit two meters of DNA into a nucleus that is only a few microns in diameter. The basic repeating unit of chromatin, the nucleosome, consists of 147 bp of DNA wrapped in ~1.7 superhelical turns around a protein octamer that consists of two copies each of histones H2A, H2B, H3, and H4. Fourteen points of contact between the DNA and octamer define the wrapping path (2). The length of DNA in the nucleosome core is conserved among eukaryotes, although nucleosomal repeat lengths vary from ~165 bp to 230 bp. The histone octamer is shaped as a cylinder of diameter ~11 nm and thickness ~8 nm containing many positively charged residues, whereas DNA is a negatively charged semi-flexible polymer of diameter ~2 nm and persistence length ~50 nm. Thus, the wrapping of DNA around the histone octamer involves competition between electrostatic adhesion and resistance to bending.

Chromatin is a dynamic structure that alters its properties and composition during the cell cycle and in response to external signals. Modulation of DNA–histone interactions allows cells to regulate processes such as replication, transcription, and repair through dynamic tuning of the structure and accessibility of chromatin (3). Access to genetic information can require partial or full unwrapping of DNA from the histone octamer, which may occur by a variety of mechanisms. Thermal fluctuations can transiently expose portions of the nucleosomal DNA (4), remodeling enzymes can actively reposition nucleosomes (5), and chemical modification of the N-terminal tails of histones can modulate DNA accessibility (6). Additionally, DNA sequence affects nucleosome stability (4). Investigating the forces needed to unravel chromatin complexes will provide insights into the remodeling, unwrapping, and rewrapping required by fundamental biochemical processes. For instance, DNA and RNA

polymerases have been shown capable of exerting transient forces up to ~40 pN (7–9), and it is important to know the effects that such forces can have on nucleosomes.

Several recent studies have employed single molecule manipulation techniques to probe chromatin structure. In pioneering work, Cui & Bustamante (10) stretched chromatin fibers extracted from chicken erythrocytes using optical tweezers. Their studies revealed a "decondensation" transition at a force of ~5 pN that was interpreted as disruption of inter-nucleosomal interactions. Above ~20 pN a second, irreversible transition was observed and interpreted as being due to disruption of nucleosomes, although discrete unraveling events were not observed. Subsequent studies have focused on *in vitro* assembly of histone–DNA complexes. Rapid compaction of DNA was observed upon flowing concentrated *Xenopus* egg extracts or protein solutions past single tethered DNA molecules (11–13). Upon stretching such complexes assembled *in situ*, Bennink, *et al*., observed abrupt lengthening events frequently releasing 382 bp and 191 bp of DNA, which were attributed to unraveling of nucleosomes (12). Subsequently, however, Brower-Toland, *et al*., obtained a different result when using complexes preassembled in bulk by salt dialysis using core histones (14). An engineered DNA template containing periodic repeats of the sea urchin 5 S RNA positioning sequence was used to create well defined arrays (15). When stretching these complexes, continuous lengthening was observed up to ~20 pN, followed by a discrete series of unraveling events each releasing ~80 bp of DNA. These events were proposed to be the disruption of strong interactions at ± 40 bp from the dyad axis in the nucleosome structure (2). It has

alternatively been proposed that such unraveling events could be explained on a physical basis without invoking specific interactions (16).

To shed further light on the mechanics of nucleosomes, we report here on the optical tweezers stretching of single nucleosomal arrays (Fig. 6.1a) assembled on a heterogeneous DNA sequence using core histones, nucleosome assembly protein 1 (NAP-1), and ATP-dependent chromatin assembly and remodeling factor (ACF) (17,18). In this assembly method, NAP-1 acts as a chaperone that delivers histones to the DNA template and prevents non-specific aggregation, while ACF couples the energy of ATP hydrolysis to the processive deposition of highly periodic arrays of nucleosomes (19). An advantage of this method is that it yields extended periodic arrays in non-repetitive DNA, as in native chromatin (20). The use of heterogeneous DNA and ACF-catalyzed chromatin assembly is expected to lead to differences in behavior compared with nucleosome reconstitution on tandemly repeated positioning elements by salt dialysis. ATP-dependent assembly factors are believed to assemble nucleosomes processively along any DNA template (21) without any apparent bias towards sequences having higher equilibrium affinities. For example, nucleosomes that are assembled onto a repeating 5 S rRNA positioning sequence with *Drosophila* S-190 extracts are not positioned in register with the 5 S rRNA sequences (M. Pazin & Kadonaga, J.T., unpublished observations).

Periodic assembly of nucleosomes on a wide variety of sequences is much closer to the actual situation in native chromatin than that obtained using salt dialysis and repeating positioning elements. Nucleosome assembly using a defined set of purified proteins also has advantages over assembly using cell extracts since the

protein composition is precisely controlled. For these reasons, this assembly method has proven useful for producing well characterized nucleosome arrays for *in vitro* studies of processes such as transcription, replication, and recombination (22–24).

6.3. Results

6.3.1. Characterization of Complexes

After assembling chromatin in bulk under standard conditions, we assessed the sample by partial micrococcal nuclease digestion (Fig. 6.1b). Digestion yielded at least four resolvable bands spaced ~168 bp apart, consistent with the expected nucleosome repeat length for arrays lacking the linker histone H1 (18). This ~168 bp repeat length includes 147 bp wrapped around the histone octomer core (the nucleosome core particle) and ~21 bp of linker DNA. Because the DNA is linear, it was not possible to use a supercoiling assay to quantify the extent of nucleosome assembly precisely. Chromatin that is assembled with linear DNA tends to exhibit less extensive nucleosomal periodicity than chromatin assembled with closed circular DNA (18). It is apparent, however, that these assembly reactions yield periodic nucleosome arrays that contain sufficient numbers of nucleosomes for our analysis. Also, as it is believed that ACF assembles nucleosomes processively, partially assembled arrays will contain sections of periodic nucleosomes similar to those present in a fully assembled array. Importantly, the nuclease digestion results presented here are for the very same sample used in the optical tweezers, insuring that the complexes being interrogated contain properly formed nucleosomes. Such

independent precharacterization of the same complexes was not possible in several previous studies where complexes were assembled *in situ* (11–13).

Analysis of complexes assembled using ACF and NAP-1 by sucrose gradient sedimentation shows that while core histones remain bound, NAP-1 dissociates after assembly (D. Fyodorov & Kadonaga, J.T., unpublished observations). As the beads are washed upon injection into the flow cell, NAP-1 would not be present during measurements. On the other hand, a significant proportion of ACF appears to remain bound to the newly assembled chromatin and may therefore still be present during our measurements. However, ACF is only present in catalytic quantities (~1 ACF molecule per 50 nucleosome lengths of DNA) and is not active for nucleosome assembly due to our exchange into an ATP-free buffer. Thus, we expect that ATP-dependent chromatin remodeling by ACF would not occur during our measurements.

6.3.2. Nucleosome Unraveling

Complexes were stretched by moving the micropipette in 0.75 nm steps at 100 Hz. This stretching caused the force applied across individual nucleosomes to ramp as high as ~65 pN on a time scale of ~10 seconds, which we have chosen to be similar to the time it takes the RNA polymerase II complex to transcribe through a nucleosome *in vivo* (25,26). Initially, we observed a nonlinear increase in force (Fig. 6.1c). Such behavior is characteristic of polymer elasticity in general (27). The most striking departure from this simple behavior was the occurrence of runs of small sawteeth in the force-extension data (Fig. 6.1d), corresponding to sudden lengthening events (Fig. 6.1e and 6.1f). As each event releases <147 bp of DNA and occurs in

<10 ms, which is one to two orders of magnitude less than the average time between events, we attribute these to the unraveling of individual nucleosomes. Variations in the events will be discussed below.

While unraveling events were identifiable throughout all data sets, significant heterogeneity in the behavior of different complexes was observed. They displayed different initial degrees of compaction and elastic behavior. Similar heterogeneity was reported in studies of native chromatin (10). We attribute these variations to differences in the degree of assembly of complexes, partial adsorption to the microspheres, partial unraveling prior to and during tethering, and lack of complete unraveling of some complexes at even the highest applied forces.

As seen in Fig. 6.1c and 6.1g, a force plateau appears at ~54 pN (in 5 mM NaCl) and ~63 pN (in 100 mM NaCl), characteristic of the "overstretch transition" of double-stranded (ds) DNA (28). The observation of this transition indicates that the elasticity in this high force regime is dominated by sections of naked DNA exposed by unraveled nucleosomes, plus any remaining linker DNA. As shown in Fig. 6.1h, nucleosome unraveling continued to occur into the overstretching transition, an effect not seen in previous studies (12,14). When complexes were partially overstretched and then relaxed, large hysteresis in force-extension was almost always observed (Fig. 6.1g), indicating that many nucleosomes were irreversibly unraveled.

6.3.3. Multiple Stretch–Relax Cycles

We were sometimes able to carry out multiple stretch–relax cycles before the tether broke (Fig. 6.2). Most complexes did not stretch out to the full extension

expected of the DNA template. Remarkably, unraveling events occurred even after multiple cycles that partially overstretched the DNA. Such high force events were not observed in previous studies, where nucleosomes seemed to be irreversibly dissociated upon reaching the overstretching transition (12,14). Our finding may explain the observation that the elasticity of native chromatin does not return to that of dsDNA even after a partial overstretch (10). Also, sections of the assembled arrays that were non-specifically stuck to the microspheres were liberated with each pull. The contour length of the tether often increased in large increments, which were followed by a series of nucleosomal length events (55 to 95 bp; see below) on the same or subsequent pulls. Such large increases are evident in Fig. 6.2. For this reason, as well as the relative rarity of tethers that survived more than one or two pulls, gathering meaningful statistics of the nucleosomal events as a function of the pull number was not possible.

6.3.4. Distribution of Unraveling Lengths and Forces

The peak force (F) and DNA length (ΔL) released were noted for each event. All resolvable events in all the data sets were included in constructing the ΔL distribution. Shown in Fig. 6.3a is the distribution of ΔL values for the 5 mM NaCl data; it was unchanged for the measurements in 20 mM, 50 mM, and 100 mM NaCl (Fig. 6.3b). A peak ranging from ~55 bp to 95 bp with a maximum at ~74 bp contained ~70% of all events. The events in the peak were deemed "nucleosomal". Of the rest, ~ 67% were >150 bp, too large to be nucleosomes. We attribute these to unbinding of complexes partially adsorbed to the beads. Notably, the value of 72–80

bp reported for positioned nucleosomes (14,29) falls within our peak, whereas the 382 and 191 bp values most frequently observed for complexes assembled *in situ* (12) fall well outside it (<5% of events were within 10 nm of 191 bp). However, more recent high resolution measurements report events of both ~176 bp and ~88 bp (30).

The distribution of nucleosome unraveling forces was broad. It extended from ~5 pN to 65 pN in 100 mM NaCl and shifted to higher force as the NaCl was decreased to 5 mM (Fig. 6.4a and 6.4b), consistent with the mechanical stability of nucleosomes increasing with decreasing monovalent salt. This effect had not been investigated in previous stretching experiments. The observed range of forces is significantly larger than the ~22–32 pN range reported for positioned nucleosomes (14). In both cases, the forces are higher than the value of ~2 pN predicted from equilibrium theory (31), suggesting that unraveling is a non-equilibrium process. The distribution of ΔL values for high force events ($\geq$42 pN) was the same as for all events (Fig. 6.3a), confirming that they are consistent with nucleosome unraveling.

6.3.5. Linear Correlation between F and ΔL

The linear correlation coefficient, ρ, between F and ΔL was calculated for both nucleosomal and non-nucleosomal events, and the results are shown in Fig. 6.5. Recall that nucleosomal events are those for which 55 bp < ΔL < 95 bp, representing ~ 70% of the total events, as in Fig. 6.3a. For the non-nucleosomal events, the correlation is roughly constant at about ρ = -0.1. While statistical tests indicate that this value is slightly different from zero, the fact that it does not change with salt reflects that the non-nucleosomal events are largely non-electrostatic. Instead, the

non-nucleosomal events are due to non-specific protein "stickiness" and the like. Intuitively, F and ΔL would not be strongly correlated for non-specific interactions.

For the nucleosomal events, all the correlations are negative, indicating that the nucleosomes with the most tightly wound inner turns of DNA (i.e., low ΔL) are the strongest. This is reasonable if we assume that the canonical nucleosome is a minimal configuration of DNA wrapping around the histones. Also, the correlation values decrease linearly with increasing salt. This reflects more than just the force distribution shifting to lower values at higher salts, because a simple translation of one (or both) of the variables leaves ρ unchanged. Instead, we feel that it is due to the fact that as the salt concentration is increased, those nucleosomes that are not in the minimal configuration, but close enough to fall within the ΔL range, are affected by the increased ionic screening more so than canonical nucleosomes. These so-called non-canonical nucleosomes will then be selectively weakened relative to those nucleosomes that have all of their contact points in place, causing the negative correlation between length and force to grow. Note that this does not necessarily imply that these so-called nucleosomes were assembled with "defects". Nucleosomes are constantly fluctuating thermally (4,32,33) and the effects of dilution may cause partial unraveling to occur. As alluded to in the text, the unraveling force of any particular nucleosome will depend strongly on what state we capture it in.

6.3.6. Number of Nucleosomal Events

In addition to the gel electrophoresis of the micrococcal nuclease digestion (Fig. 6.1b), the extent of nucleosome assembly of our samples can be assessed from

our tweezers measurements. Specifically, the number of nucleosomes observed per first pull reflects in part the extent of nucleosome assembly on our ~ 24 kbp templates (see Methods). Only first pulls that reached ≥ 40 pN were included so as to not low-bias the distribution; tethers that break prematurely assumedly did so with at least a portion of their nucleosomes unraveled. Likewise, only first pulls are included because of the assumption that subsequent pulls will systematically show fewer nucleosomes. As discussed in the section "Multiple Stretch–Relax Cycles", however, the removal of sections of the nucleosome arrays that were stuck to the beads occasionally led to subsequent pulls with many more nucleosomal events than previous pulls. Because of these effects, the distribution of number of nucleosomal events per first pull represents a *minimal* numerical extent of nucleosome assembly. This distribution is shown in Fig. 6.6. As with the ΔL distributions, no difference in this N distribution was observed in the different salt concentrations. Thus, all of them were considered together to improve statistics. For the purpose of a smoother distribution, the bin size was two events, and the count in each was normalized by the total number of events so that the error on each is $(N_{bin}/N_{total})^{1/2}$. The distribution for N ≥ 2 was fit with a decaying exponential ($y = ae^{-bN}$), yielding the parameters a = 0.23 ± 0.01 and b = 0.11 ± 0.01 (nucleosomes per pull)$^{-1}$. The mean of this distribution is thus ~ 9 nucleosomes per pull, which again represents the minimal assembly efficiency. Taking the repeat length to be 165 bp in the absence of linker histones, the maximum number of nucleosomes we expect on our DNA is ~ 150. In our estimation, however, such an efficiently packaged DNA molecule would have been nearly impossible to tether in the tweezers; efficient tethering required that the

biotin and digoxigenin "handles" be well exposed. Nonetheless, the large disparity between maximal packaging and the observed number emphasizes, one, the substantial under-assembly of our sample and, two, the inherent difficulties in fully unraveling our nucleosome arrays due to tethers breaking and non-specifically sticking to the beads.

6.3.7. Spontaneous Rewrapping

Although most nucleosomes unraveled irreversibly, abrupt shortening events were sometimes observed in real time during relaxation of the tethered complex (Fig. 6.7). Shortening events following overstretching of naked DNA have been observed previously and are due to reannealing of partly melted DNA (28). However, we observed shortening events after complexes were stretched to a force below the overstretch, where melting hysteresis is not observed. We therefore attribute these events to spontaneous rewrapping of DNA onto octamers that were not completely dissociated during stretching. Real time observation of re-wrapping events has not been reported previously although evidence of re-wrapping was obtained by observing unraveling following repeated stretching of fully unraveled complexes (extending to the naked DNA length) (14). Shortening events were rare compared with unraveling events. Only 73 were observed in 366 relaxation measurements. The mean decrease in tether length was 34 nm (σ =15 nm), not far from the ΔL value observed during unraveling. Event forces ranged from 7 pN to 63 pN (mean = 37 pN, σ = 15 pN), similar to those observed for unraveling. The mean waiting time for re-wrapping from the beginning of relaxation was 2.1 seconds (σ = 1.3 seconds),

although the distribution did not follow a single exponential decay as would be expected for a single step process in steady state. Rather, the distribution was peaked at ~1.7 seconds, consistent with re-wrapping kinetics being force-dependent.

6.3.8. Elasticity of Complexes

Sections of force-extension data without unraveling events were analyzed to determine the elastic properties. Native chromatin folds into a higher order structure in 100 mM NaCl (1). However, factors that promote folding, such as linker histones and divalent cations, were not included in our experiments. We found no evidence of the decondensation transition observed by Cui & Bustamante, suggesting that our complexes are unfolded. This observation agrees with the findings for positioned nucleosomes (14).

Good fits to the data were obtained using the extensible worm-like chain (WLC) model with persistence length (P), stretch modulus (S), and length (L_0) as parameters (34). Full packaging of DNA into H1-free nucleosome arrays is expected to result in a seven-fold compaction, but verification of this expectation was not possible due to incomplete assembly as well as partial adsorption of complexes to the beads. Here, we analyzed the elasticity of complexes exhibiting ~5% to 50% of the DNA length. Distributions of P and S are given in Fig. 6.8. In 5 mM NaCl, the mean P was 10 nm (σ = 11 nm, n = 359 fits), which is substantially smaller than the ~50–80 nm values reported for naked DNA (35–37) and 30 nm value reported for native chromatin (10). The mean S, on the other hand, was 840 pN (σ = 690 pN, n = 396), which is within the ~770–1000 pN range reported for naked DNA (35–37). It is quite

different, however, from the value of 5 pN reported for native chromatin (10). As the NaCl was increased from 5 mM to 100 mM, the mean value of P shifted up to 18 nm (σ =18 nm, n=146), implying higher extensibility at low force. The mean S dropped slightly to 770 pN (σ =370, n=95), which is slightly smaller than the reported values of 1000–1350 pN for naked DNA.

6.4. Discussion

6.4.1. Unraveling Length

The nucleosome core contains 147 bp of DNA wrapped ~1.7 turns around the core histone octamer, and hence the observed 55–95 bp events do not correspond to the entire nucleosome unraveling in a single step. This finding is consistent with the proposed mechanism that the outer turn peels off continuously at low force, whereas most of the inner turn (~80 bp) is released abruptly at higher force (14). More recent studies investigating the role of histone tails in forced nucleosome unraveling report a value of 72 (±4) bp for the inner turn (29). These values also agree with theoretical predictions of Kulíc & Schiessel for an elastic rod wrapped with uniform adhesion on a cylindrical spool. Their calculations show that the first turn can be peeled off easily, but removal of the final turn via stretching presents a significant kinetic barrier. This property is due to the spool having to tilt and the DNA having to bend before unraveling can occur (16).

In comparison to our data, events most frequently releasing 382 bp and 191 bp were observed for complexes assembled *in situ* using extracts (12). The 191 bp events were attributed to unraveling individual nucleosomes and the 382 bp events

were attributed to unresolved pairs of 191 bp events. One possible reason for this different finding is that the *Xenopus* extracts contain additional proteins, such as HMG and linker histones, that could increase the effective length of wrapped DNA. These complexes were also stretched ~13 times faster and the measurements also did not have high enough resolution to resolve smaller events. Indeed, more recent, higher resolution measurements indicate events of both ~176 bp and ~88 bp (30), the difference being attributed to nucleosomes either having or not having linker histone B4. Discrete unraveling events were not observed in the initial studies of native chromatin (10); however, stretching was done in ~150 bp steps, which were too large to resolve the unraveling of individual nucleosomes.

6.4.2. In Situ Assembly

A question that should be raised is whether the *in situ* assembly method used in several recent studies (11–13, 30) leads to proper nucleosome formation. Isolated DNA molecules, at essentially infinite dilution, are immobilized inside a chamber, and protein solution is flowed in continuously. These solutions contain ~1 to 100 mg/ml of histones and, in the case of extracts, many other proteins capable of binding to DNA. Therefore, the histone:DNA mass ratio is vastly higher than the ~1:1 level used in standard bulk assembly reactions and present in native chromatin. In our experience using core histones, NAP-1, and ACF, proper assembly *in vitro* only occurs in a narrow range around a ~1:1 histone to DNA mass ratio (17,18). While sample characterization assays were performed in these previous studies, these tests were done on different samples than those used in the single molecule experiments;

therefore we believe that they may not be correctly assembled. Assembly at high protein:DNA ratio tends to result in formation of non-specific protein–DNA aggregates rather than proper nucleosome arrays (17,18).

6.4.3. Variations in Unraveling Events

The most notable difference between our results and those obtained for positioned nucleosomes formed by salt dialysis (14) is that our unraveling events occurred over a wider force range. In 100 mM NaCl, our data revealed unraveling forces ranging from nearly fourfold lower to twofold higher than that observed for nucleosomes assembled on the 5 S RNA positioning element (~22–32 pN). The largest fraction of our events (52%) occurred at lower forces (<22 pN), whereas 28% fell in the 22–32 pN range and 20% occurred at >32 pN. In both studies the complexes contained only core histones, though different assembly methods and DNA templates were used. That the histones came from different organisms is not expected to cause differences because core histones are highly conserved across eukaryotes. Also, nuclease digestion patterns suggest that different assembly methods result in the same nucleosome core structure (1). While we used a higher stretching rate (75 versus 28 nm/s) this would not explain our finding a majority of events at lower force, because unraveling force increases with pulling rate (14,38).

Brower-Toland, *et al*., made measurements in 100 mM NaCl plus 1.5 mM $MgCl_2$, whereas we simply used NaCl ranging from 5 mM to 100 mM. We did not include divalent cations because complexes were extremely difficult to tether at higher overall ionic strengths in our experience, presumably because these conditions

promote condensation of the nucleosome arrays (1). Mg^{++} has been shown to cause increased screening of histone–DNA electrostatic interactions (39) potentially leading to lower unraveling forces. Therefore, some of our high force events could be interpreted as occurring due to our lack of Mg^{++} relative to that of the studies with positioning elements. However, we observe a majority of unraveling events at lower forces. This is the opposite trend of what would be expected if this shift were due to the absence of Mg^{++}. While bending of certain DNA sequence motifs may be stabilized by divalent cations, potentially favoring stability of some nucleosomes, recent crystallographic studies by Davey & Richmond suggest that "DNA conformation appears to dictate metal binding, as opposed to the converse" (40).

We therefore interpret the higher variability in unraveling force as being due to our use of a nonrepetitive, heterogeneous DNA sequence instead of the tandemly repeated 5 S RNA positioning sequence used by Brower-Toland, *et al.* (14). The relative affinities of nucleosomes on different DNA sequences have been previously studied in competition assays, revealing that different sequences can have up to ~4 kcal/mol differences in equilibrium stability (4,41). Relative to the 5 S RNA positioning element, random DNA has on average only a slightly (~0.5 kcal/mol) higher free energy, whereas physically selected sequences have been identified with ~3 kcal/mol lower free energy. This variability may arise from factors such as variations in inherent curvature and/or flexibility of different sequences (4). During mechanical unraveling, lower force events presumably correspond to lower affinity sequences and higher force events correspond to higher affinity sequences, although a quantitative relationship between these equilibrium and non-equilibrium properties

remains to be determined. That most events occur at lower force is consistent with λ DNA, which is non-eukaryotic and not naturally incorporated into nucleosomes, behaving largely as a random sequence. That some nucleosomes can unravel at forces as low as ~5 pN suggests that spontaneous unraveling may sometimes occur, further reconciling the need of cells for chromatin assembly and remodeling enzymes to maintain the integrity of chromatin.

Events that occurred at forces >32 pN also cannot fully be explained by our use of a higher stretching rate compared to the studies of positioned nucleosomes (14) because the expectation value of force is only a logarithmic function of the loading rate (38). Unraveling events also occurred after multiple cycles of stretching, as mentioned above. Thus, a small fraction of sequences in λ DNA appear to yield more stable nucleosomes than those probed in previous studies. Notably, enzymes such as RNA and DNA polymerases, which must partly unravel nucleosomes, are capable of exerting such high transient forces (7,8). Based on its high rate of ATP consumption (21), we suspect ACF is also capable of exerting comparable forces. Although two different sequences from λ DNA were used (see Methods), we did not observe a segregation of the data into two distinct groups with differing forces or unraveling lengths, as would be expected if the two DNA sequences were behaving differently. Our interpretation is that both molecules are behaving as random sequences from the point of view of nucleosome unraveling.

The increase in force as NaCl was decreased is consistent with reduced ionic screening, whereas the finding that ΔL is unchanged is consistent with the nucleosome maintaining its overall structure in this range of salt concentration (1).

Studies of endonuclease accessibility do indicate, however, that the nucleosome is a dynamic structure (4,32,33). Part of the width in our ΔL distribution may therefore be due to nucleosomes being in different structural states at different times. Wider variations in ΔL observed in the re-wrapping events may occur because the canonical nucleosome structure may not always properly reassemble after mechanical disruption, especially as assembly factors such as NAP-1 and ACF were not active.

6.4.4. Potential Effect of Dilution

The single molecule experiments reported here and by other groups usually involve a large dilution of complexes. One possible concern, therefore, is the phenomenon of dilution driven dissociation of nucleosomes (42–44). While previous studies suggest that a fraction of nucleosomes may dissociate upon dilution, we do not believe this alters our conclusions or biases our measured force distributions. Cotton & Hamkalo reported that dissociation of a fraction of nucleosomes "does not result from special properties of a subset of the nucleosomes" (42), as also noted by others (45). Therefore, dissociation of octamers would simply reduce our collection efficiency. Further, we use relatively low ionic strengths in our experiment. It is well established that high salt can dissociate nucleosomes by shielding the electrostatic attraction between histones and DNA. Indeed, most dilution driven dissociation experiments in the literature were carried out at [NaCl]>100 mM. Cotton & Hamkalo reported that dissociation only becomes significant at salt concentrations above 50 mM NaCl and decreases sharply as salt is decreased further. Also, only a fraction of nucleosomes dissociate and, even at 150 mM NaCl, this fraction is < 40%. In other

dissociation measurements Lilley, *et al*., used 150 mM ammonium sulfate and Thåström, *et al*., used 200 mM NaCl (44,46). In contrast, we made measurements spanning a range of lower concentrations (5, 10, 20, and 100 mM NaCl). Further, the literature indicates that the time scale of dilution driven dissociation is on the order of an hour (39,44,47,48), whereas we began some of our measurements within minutes after dilution. Being part of an extended array would also contribute to increased nucleosome stability. It has been reported that individual nucleosomes with DNA >155 bp dissociate more slowly than ones with only 147 bp (47). The inherent attraction between dissociated histones and the DNA molecule from which they came also creates an effectively increased histone concentration that mitigates the effects of dilution driven dissociation in extended arrays.

Another issue is whether dilution and/or mechanical stretching may lead to partial dissociation of the histone octamer, such as dissociation of the H2A/H2B heterodimer, which could provide an alternative explanation for the observation of approximately "half nucleosome sized" unraveling events. A recent study comparing the dissociation of radiolabeled histones from mononucleosomes after one hour of dilution and observed unraveling events in both dilute and stabilizing conditions suggest that a 25 nm unraveling length may be a consequence of H2A/H2B dissociation and that a 50 nm event corresponds to unraveling from the full histone octamer (49). However, histone octamer instability was also considered previously by Brower-Toland, *et al.*, who found that the 25 nm events are unchanged when histone octamers are crosslinked, implying that these events do correspond to the removal of DNA from whole octamers (14). Further studies of these events showed

measurable differences in unraveling force following cleavage of the tails of H2A/H2B (29), indicating that these histones were still present despite large dilutions. These findings are consistent with previous studies suggesting that dissociation of core histones from DNA in solutions below 0.75 M NaCl is highly cooperative and without intermediates (48). A different reason for two different unraveling lengths may be the presence or absence of linker histones, as suggested in a recent study in which events of both ~30 nm and ~60 nm were observed when stretching complexes assembled using *Xenopus* extracts containing the linker histone B4 (30).

6.4.5. Conformation of Complexes

The lack of a folding transition suggests that our complexes adopt the "beads-on-a-string" form observed for native chromatin in low salt (1). For lack of a specific model for chromatin elasticity, we fit our data using the extensible WLC model. A smoothly bending polymer is ascribed a persistence length that characterizes its entropic bending and a stretch modulus to describe enthalpic deformation. Our data fit quite well to this model, with persistence length, stretch modulus, and contour length as fitted parameters. These fits, however, do not necessarily imply that the picture of a smoothly bending polymer accurately portrays the microscopic conformation of the complexes. Regardless of its meaning in the WLC model, a fit persistence length of 10–20 nm implies that our complexes have a lower extensibility than naked DNA, particularly for fractional extensions <90%. This property could be due to the force needed to align the bent linker DNA and to peel DNA from the histone octamer. The observed increase in extensibility in 100 mM salt would then

be attributable to easier peeling of the DNA due to increased ionic screening. Our finding of a stretch modulus only ~34% smaller than that for naked DNA suggests that, by a fractional extension of 90%, most DNA straightening and peeling has occurred, and sections of linker DNA and unraveled DNA dominate the elasticity.

6.4.6. Dissociation of Histone Tails

Each histone consists of a precisely folded globular domain, and the octamer of these domains from the "spool" around which the DNA is coiled (2). Each histone also has a randomly oriented, N-terminal amino acid tail that extends from the nucleosome. Every histone in the octamer has one, though they are of different lengths. H2A histones have C-terminal tails as well (50). Like the globular domains, the histone tails are lysine rich and thus positively charged. It has been shown that in nucleosome core particles (NCPs), the tails interact electrostatically with nucleosomal DNA, stabilizing the nucleosome (1,51-57). Lilley, *et al*., showed that digestion the tails with trypsin destabilized NCPs in sedimentation experiments, indicating a partial unfolding of the nucleosome (55). In low salt, nucleosomal DNA will unwrap ~36 bp from the histone core for tailless histones (58,59).

In nucleosome arrays and chromatin, the tails play a role in higher order folding via interactions with other nucleosomes and linker DNA (1,56). They aid in the condensation of the nucleosome array into higher order structures (58, 60), as well as the formation of interchromatin fiber oligomers (61). The chemical modification of histone tails has been shown to modulate the accessibility of nucleosomal DNA (6). Transcription experiments have even indicated that the tails may act as a tether,

allowing the passing of the elongating polymerase without fully dissociating from the DNA (63). Therefore, understanding the dissociation of histone tails and the effect thereof on nucleosome stability would help elucidate the complicated yet essential relationship between chromatin structure and function.

X-ray scattering structural studies (62) indicate that as the salt concentration is increased from 10 to 50 mM, the maximal diameter (D_{max}) of nucleosome core particles (NCPs) increases by roughly twenty percent. Higher salt concentrations then cause D_{max} to level. They attribute this to the tails dissociating from the nucleosome cores "without significant change in the shape of the rest of the particle." Interestingly, they comment that the extension of the tails at higher salt (>50 mM) "mimics the effects induced by the acetylation of the histones." (62). Therefore, salt concentration provides a way of continuously modifying the dissociation of the histone tails *in situ* without utilizing the more drastic chemical modifications of the histones. Additionally, since the indicated dissociation is within our range of salts, one might ask to what extent is the dissociation of tails seen in our experiments.

In a recent paper, Brower-Toland, *et al*., showed that the both the complete removal and acetylation of histone tails lead to measurable differences in optical tweezers experiments of nucleosomal arrays. These differences appear in both the amount of DNA contained in the *outer* loop of DNA around the histones and the force needed to unravel the inner turn of DNA from the histones. The amount of DNA in the inner turn is conserved (29). In 100mM NaCl (1.5mM $MgCl_2$), they report a force difference from fully intact to fully digested histones of -2.9 ± 0.2 pN. Extensive acetylation of all the histone tails yielded a force drop of -1.8 ± 0.2 pN

(29). From 5 mM NaCl to 100 mM NaCl, a salt concentration at which we expect significant histone tail dissociation, we observe a mean force drop of -6.4 ± 0.7 pN. So, while it appears that the effects of increased salt are largely manifested in the screening the electrostatic attraction between histone core and DNA, the effects of tail modification and dissociation are indeed of the same order. Thus they should be measurable by our techniques, as discussed below.

The other effect of tail modification reported by Brower-Toland, *et al*., is the substantial decrease in the amount of DNA in the outer turn. A decrease of outer turn DNA upon pulling in the tweezers corresponds to an increase of DNA liberated from the nucleosome by the conditional change, whether it be chemical modification of the histone tails or salt-induced dissociation. From fully intact to fully digested tails, Brower-Toland, *et al*., reports a decrease in outer turn DNA of 37 bp (29), in agreement with the ~36 bp mentioned above and representing a change of ~60%. In our data, we did not measure the amount of DNA in the outer turn directly, as non-specific sticking of the complexes made determining lengths at low forces difficult and spurious. However, we were able to measure the persistence length of the complexes at low forces, the regime in which the outer turn of DNA peels off. From 5 mM NaCl to 100 mM NaCl, we observe an 80% increase in the measured persistence length towards the accepted value for naked DNA. Qualitatively, this indicates an increase in the amount of liberated DNA from the histone cores on the same order observed by others (29, 57). These numbers indicate that the increase in free DNA brought about by the salt-induced dissociation of the histone tails would likewise be measurable by our techniques.

To measure the salt-induced dissociation of the histone tails rigorously, one would assemble nucleosomes with unmodified histones (i.e., tails intact) on a repeating positioning element by salt dialysis (14), eliminating the effect of sequence on the measured forces. Pulling on these nucleosome arrays in a progression of salt concentrations would yield the effect of salt concentration on the force distributions. This effect, however, would be a convolution of the ionic screening of the electrostatic interaction between the histone tails and DNA, as well as the screening of the histone cores and the nucleosomal DNA. To isolate the effect of the tail dissociation, one would then want to repeat the experiment, but with tailless histones. The effect of salt on these force distributions would reflect only the ionic screening of the globular domains and the nucleosomal DNA. Deconvolving the latter from the former would yield the effect of the progressive tail dissociation with salt. If at any point the arrays with fully intact histones give essentially identical results to those of the arrays assembled with tailless histone, then it is an indication that, at that salt, the histone tails are indeed fully dissociated.

6.5. Conclusions

Our data provide direct evidence that there is significant variability in the non-equilibrium force needed to unravel individual nucleosomes assembled on heterogeneous DNA sequences. From a biological standpoint, this finding suggests that resistance to unraveling could be an important factor in the functioning of DNA-directed processes in the nucleus. Some nucleosomes will be easier to unwrap and reposition than others, and we speculate that this effect could depend on the

biological context. For example, in a promoter region of a gene where nucleosomes will be disrupted during binding of transcription factors, nucleosomes may present less resistance. On the other hand, in a region that needs to be transcriptionally repressed, nucleosomes may be more difficult to unravel. The movement of polymerases and helicases is also slowed by the presence of nucleosomes, which need to be partially or wholly unraveled to allow passage of these enzymes (63,64). The large variation in unraveling forces and occurrence of spontaneous rewrapping events are expected to have an important influence on these processes.

That approximately half of the DNA in the nucleosome core particle is resistant to unraveling provides a plausible explanation for how nucleosome stability is maintained while allowing access of DNA binding proteins. Brower-Toland, *et al.,* postulated that peeling of the outer turn would allow access to this portion, whereas the inner turn is protected by stronger interactions (14). In contrast, Kulíc and Schiessel (16) argued that no specific features in the nucleosome crystal structure or endonuclease accessibility data suggest a large energy barrier at a specific position. Their calculations suggest that the inner turn is actually bound with lower energy than the outer turn. During protein binding, electrostatic repulsion between the two DNA turns is envisioned to facilitate partial unwrapping of one turn, either inner or outer, while the other turn remains strongly bound to preserve overall stability.

The combination of *in vitro* chromatin assembly methods and optical tweezers manipulation is expected to open a number of future research directions. The dependence of nucleosome unraveling on DNA sequence, histone composition, and solution conditions may be studied. The effects of histone modifications such as

acetylation and phosphorylation, which correlate with changes in gene expression, are also of great interest. Very recently, optical tweezers have been used to measure differences in unraveling forces and lengths upon removal or acetylation of histone tails, indicating that the tails influence specific interactions within the nucleosome core (29). Other future directions include the analysis of factors governing higher order folding, which may depend on having long, highly periodic nucleosome arrays, and the study of chromosomal proteins such as the linker histone H1. Finally, mechanical manipulation of single chromatin fibers may allow the dynamics of assembly and remodeling enzymes such as ACF to be measured in real time.

6.6. Methods

6.6.1. Nucleosome Assembly

End-labeled DNA was prepared using *E. coli* DNA polymerase I Klenow fragment to fill in the ends of methyladenine-free λ DNA (NEB) with biotin-dATP and dCTP (Invitrogen). The DNA was then cut by XbaI and purified using a spin column (Promega Wizard DNA clean up kit). A second fill-in was done with DIG-labeled dUTP (Roche), resulting in two fragments of ~8.1 mm. Core histones were purified from *Drosophila* embryos and recombinant NAP-1 and ACF were purified following baculovirus expression in Sf9 cells as described (18). Nucleosome arrays were assembled exactly as described in the methods paper by Fyodorov, *et al.* (18).

6.6.2. Tethering of Complexes

Diluted complexes were incubated with 0.5% 2.2 mm streptavidin beads (Spherotech) in 20 mM Tris (pH 7.8), 5 mM NaCl, 1 mM EDTA, and 0.2 mg/ml BSA

for one hour at ~22 ºC. Microspheres were injected into the chamber containing 20 mM Tris (pH 7.8), 1 mM EDTA, and 5–100 mM NaCl, and brought in contact with a 2.8 mm protein G-antidigoxygenin coated microsphere to form a tether (Fig. 6.1a). It should be noted that the nature of the linkages used are such that the DNA is torsionally unconstrained. Measurements were done at ~22 °C.

6.6.3. Nucleosome Analysis Considerations

In this manuscript we have noted that pulls that did not reach 40 pN were discarded from the analysis as not to low-bias the force distribution. The DIG-αDIG bond often breaks by such forces, especially with as slow a pulling speed as used in these experiments. To investigate the effects of varying the cutoff force (called F_{max}), and to aid in selecting it for the analysis in this manuscript, we considered the force distributions of the nucleosomal events (55 to 95 bp) from the 5 mM NaCl data set, as it had the most data of all the salt concentrations. These distributions are shown in Fig. 6.9a, each one being normalized to the number of events in each. With increasing F_{max}, the force distribution shifted to higher force values, though the F_{max} = 0 pN and 40 pN curves were comparable in shape. The 40 pN curve represented only a ~16% rejection rate from the full distribution. Thus pushing the F_{max} as high as possible while maintaining as much of the character of the full distribution as possible made choosing F_{max} = 40 pN for the analysis a reasonable choice. For the 60 pN, a large secondary peak at higher forces developed relative to the F_{max} = 40 pN. We do not feel that this represented a population of intrinsically more stable nucleosomes. Instead, we feel that it represented a population of “normal” nucleosomes that never

experienced the range of lower forces because of sections of tethers non-specifically sticking to the microspheres. The higher the force, the more likely such sections became exposed, and the nucleosomes in them would have been disrupted at spuriously higher forces. In addition to this uncertainty, an $F_{max} = 60$ pN leads to a ~ 55% rejection rate, which we deemed too high.

The average force, <F>, and unraveling length, <ΔL>, of these distributions are shown as a function of F_{max} in Fig. 6.9b. The <F> climbed with increasing F_{max}, reflecting the spurious population of high force events. The <ΔL> also climbed a tiny bit with increasing F_{max}; the value across most of the F_{max} range was ~ 72 bp. Recall that the ~ 74 bp reported above represents the peak of the distribution of all events; the ~ 72 bp here represented the average <ΔL> of those values deemed nucleosomal, hence the minor discrepancy. It was evident that changing F_{max} did not seriously effect the ΔL distributions, but did effect the F distributions significantly.

Convoluted into the thousands of nucleosome events were thousands of events that represented either poorly formed nucleosomes or non-specific interactions between the fiber and a bead or between different parts of the fiber. Due to these factors, even the nucleosomal events were distributed in ΔL. The ~ 70% criteria mentioned above (55 to 95 bp) provides a statistically defensible selection range given the relatively Gaussian nature of the ΔL distribution near the peak. Windowing the peak more and more narrowly, however, focused the attention on those nucleosomes which were, in theory, more canonical, but at the expense of statistics. An increasingly wide window incorporated more events, but the average force of those events was expected to decrease as more and more non-canonical nucleosomes

and non-specific interactions were included. This characteristic is demonstrated by considering the 5 mM NaCl data but with an $F_{max} = 0$, allowing for the most data points to enter into the windowing consideration. From the peak at 25 nm, windows of 1 nm (~ 3bp), 3 nm (~ 10 bp), 10 nm (~ 34 bp), and all inclusive were selected. The resulting average forces of the events in that window were calculated and shown in Fig. 6.10, with the windowing axis on a log-10 scale. The all inclusive windowing was selected as 100 nm on the x-axis to be roughly commensurate with the scale of Fig. 6.3a. Because it included all of the events, it is really a window of "-25nm/+∞ nm" (i.e., there can be no negative dL unraveling events), but reflecting that on a graph was not possible. Therefore, a quantitative fit to the points is not meaningful. Nonetheless, the trend upwards in force as the definition of nucleosome becomes more stringent is obvious; an estimated "zero window" value is ~ 29 pN. We feel that the events in this narrow window represent the best formed nucleosomes. The trade off to this narrow windowing, of course, is the number of statistics. The window used in the analysis of throughout this manuscript encompasses ~70% of the events (with an $F_{max} = 40$ pN), whereas the narrowest range in Fig. 6.10 encompasses only ~10% of the total number of events without the F_{max} restriction. Again, we chose the 55 to 95 bp window to balance stringency with statistics.

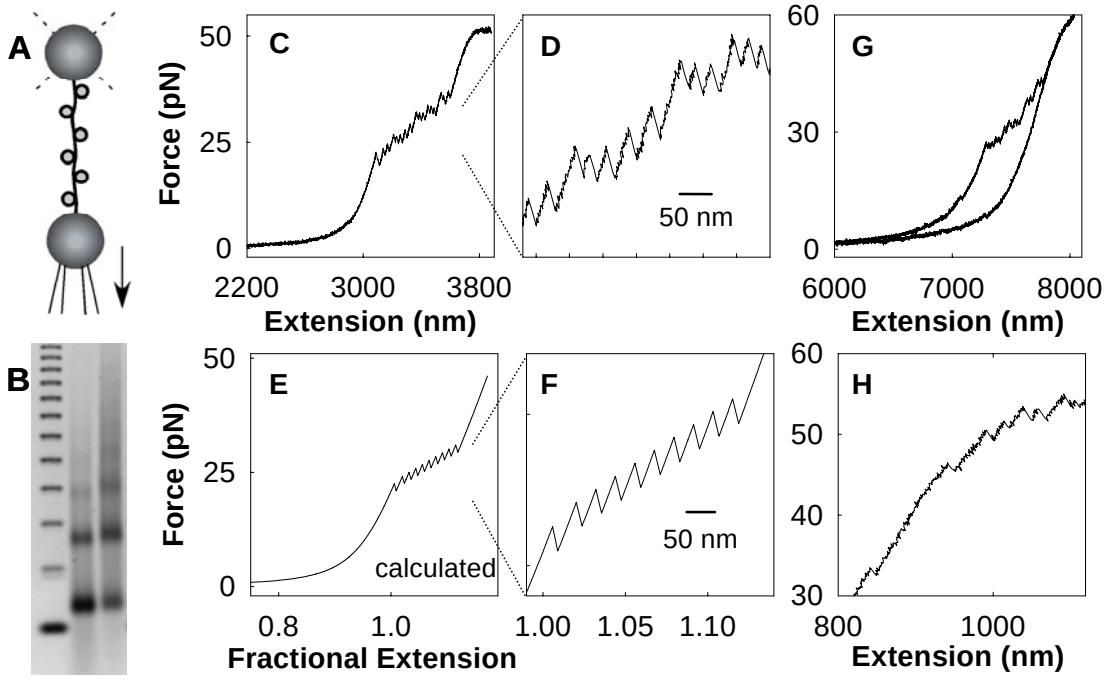

Fig. 6.1. (a) Schematic diagram of an individual complex stretched between a bead held by optical tweezers (top) and a bead held by a micropipette (bottom). (b) Agarose gel electrophoresis of samples partially digested by micrococcal nuclease. Lane 1 contains 123 bp ladder. Lanes 2 and 3 contain samples digested by 2X and 1X dilutions of micrococcal nuclease. (c) Force-extension measurement of a complex in 5 mM NaCl, showing a series of nucleosome unraveling events. Prior to reaching the full extension of ~8.1 mm, the DNA overstretches at ~54 pN, the expected value for 5 mM NaCl. Further stretching could not be done, since this complex detached from the beads. (d) Zoom of (c) showing the unraveling events. (e) Calculated force versus fractional extension for an extensible worm-like chain (WLC) in series with a trap of stiffness 0.17 pN/nm. Good agreement with the events in (c) was obtained by calculating events with 25 nm increases in the contour length (ten unraveling events spaced 40 nm apart are shown). A persistence length of 20 nm and a stretch modulus of 840 pN, average for these complexes, were used in these calculations. These calculated data were used to verify our method of determination of the length of DNA released in unraveling events. (f) Zoom of (e). (g) Data recorded in 100 mM NaCl, showing a stretching cycle, unraveling events, and a relaxation with hysteresis. In this example the fiber reaches nearly the expected naked DNA length. The DNA begins to overstretch at ~63 pN, as expected in 100 mM NaCl. (h) Another data set recorded in 5 mM NaCl showing nucleosome unraveling events occurring into the overstretching transition.

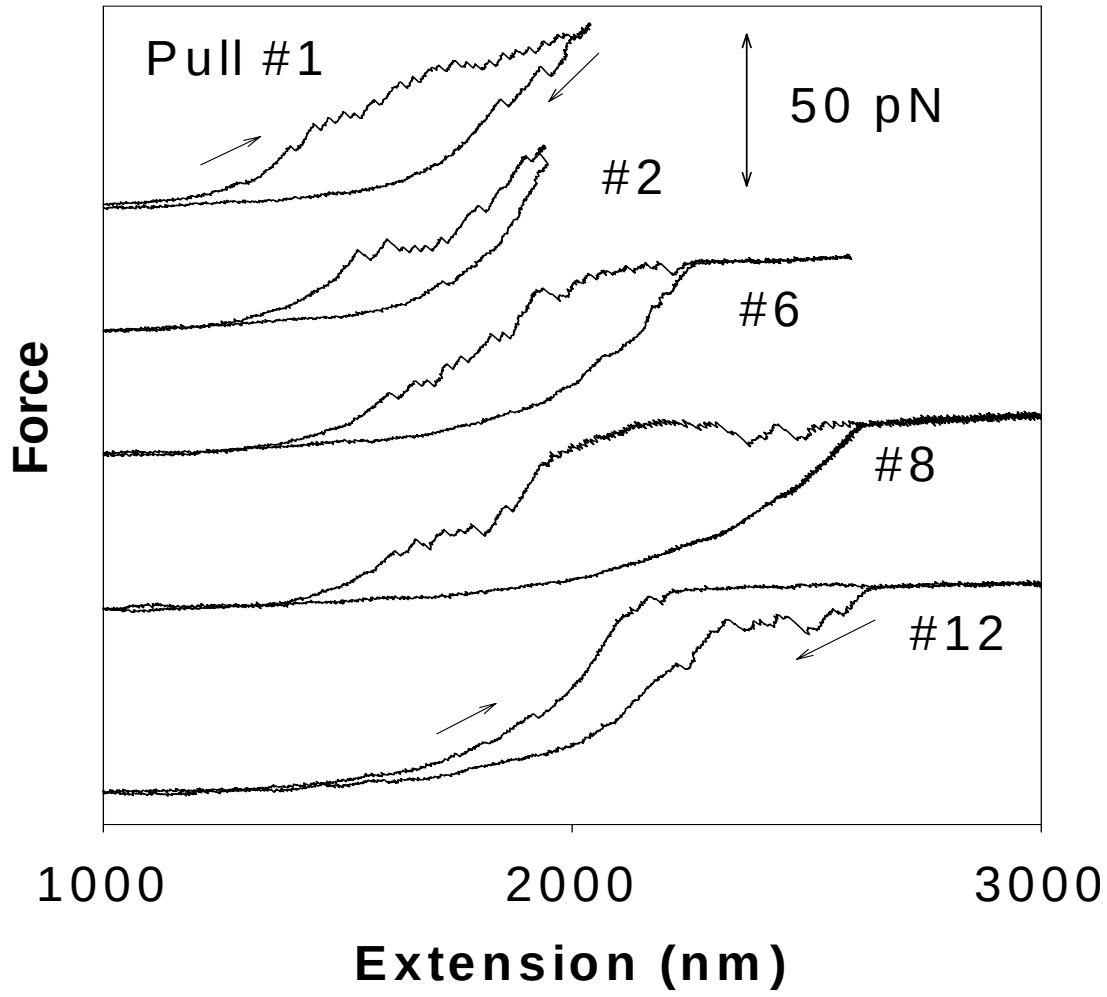

Fig. 6.2. Repeated stretching and relaxing of an initially compact fiber. It remained tethered during 12 stretch–relax cycles and displayed unraveling events and hysteresis even after multiple pulls into the DNA overstretching regime. Selected stretch–relax cycles are indicated by the numbers next to each plot. The plots have been shifted along the force axis for display purposes. The scale bar represents 50 pN. Although the frequency of unraveling events decreased after multiple stretches the tether had not unraveled to the expected DNA length (~8.1 μm).

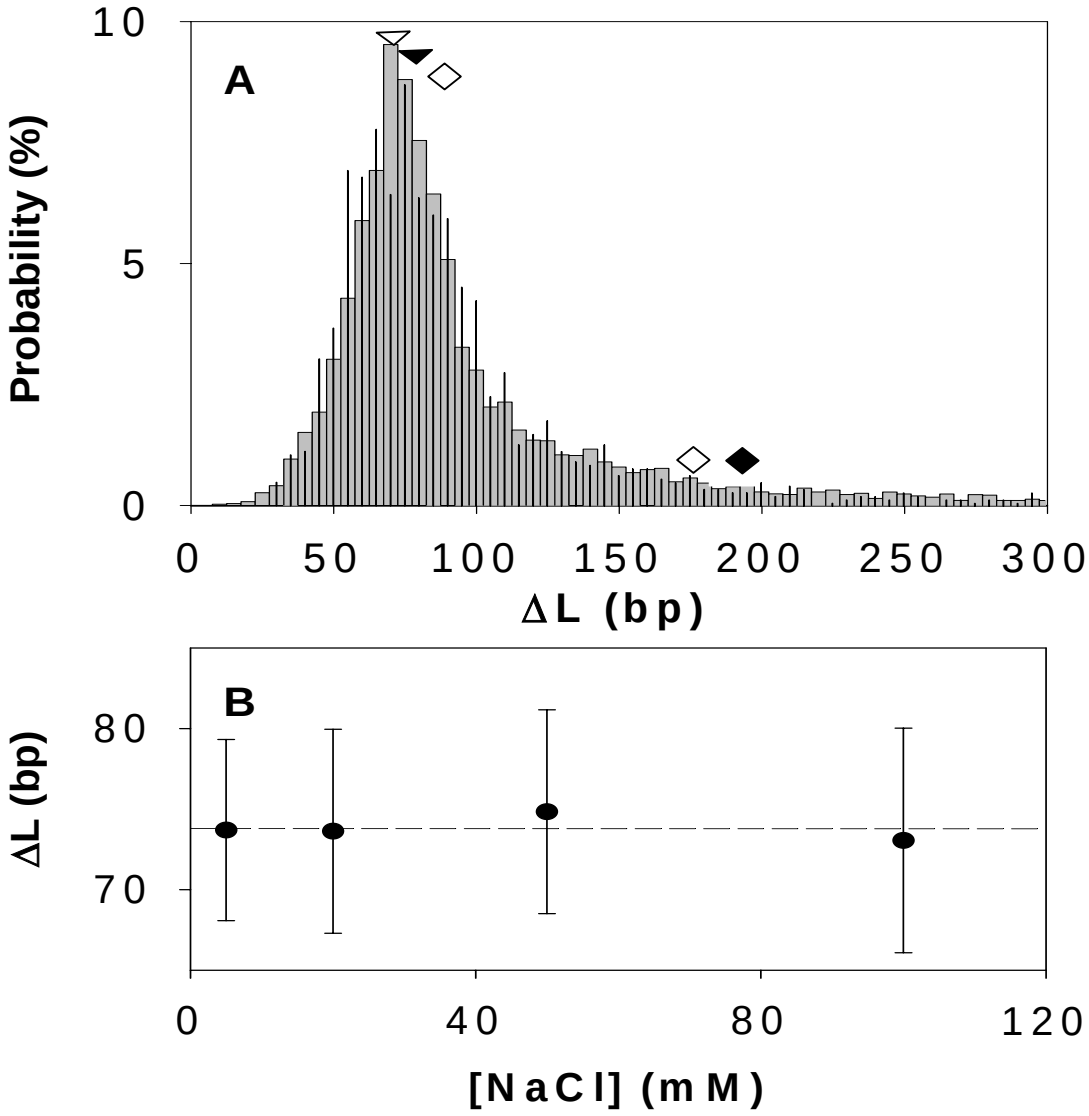

Fig. 6.3. (a) Distribution of DNA lengths (ΔL) released in unraveling events for complexes in 5 mM NaCl. ΔL was determined by measuring the distance along the extension axis from each peak to the following point of equal force. This procedure was validated on simulated data (Fig. 6.1f) calculated by using the extensible WLC model (34,65). Gray bars indicate entire set of results, while thin black needles represent only events that occurred at forces greater than 42 pN (and into overstretch). Previously reported values are indicated by the following symbols: open triangle (14); filled triangle (29); filled diamond (12); open diamonds (30). (b) Independence of the peak ΔL on NaCl. The ΔL values average to 74(±3) bp.

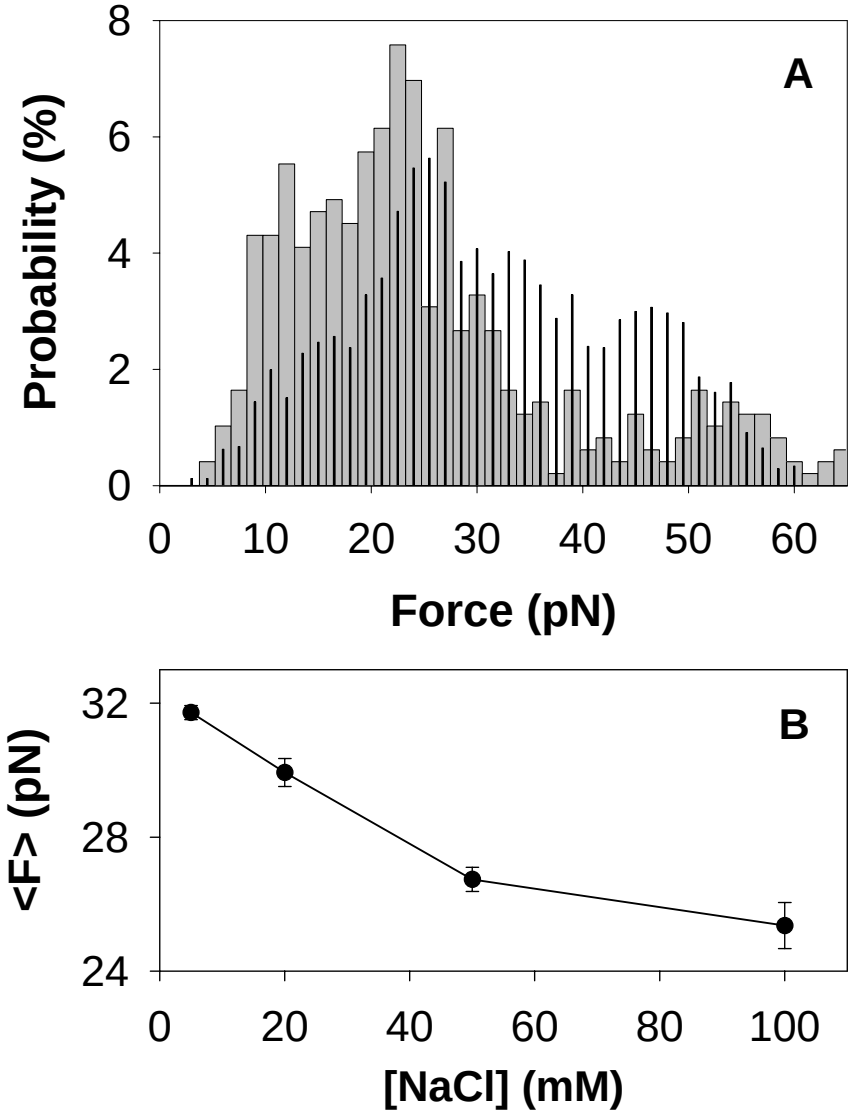

Fig. 6.4. (a) Distribution of unraveling forces in 100 mM NaCl (gray bars; mean = 25 pN, standard deviation σ =16 pN, N = 498 events) and 5 mM NaCl (black needles; mean = 32 pN, σ = 14 pN, N = 4305). In many trials the DNA tether broke before the measurement was complete, so in order to minimize a low force bias, only trials that reached F > 40 pN were included. The small peak in the distribution at high force is attributed to two effects. First, the force-extension curve begins to flatten ~5 pN before the overstretch plateau, resulting in a decreased loading rate that biases some events towards higher force. There is not a peak at the overstretch force plateau because the majority of complexes completely detach before the full overstretching plateau is reached. Second, at high force there is occasionally unbinding of portions of complexes partially adsorbed to the beads, also biasing some events towards higher force. (b) Average force versus salt.

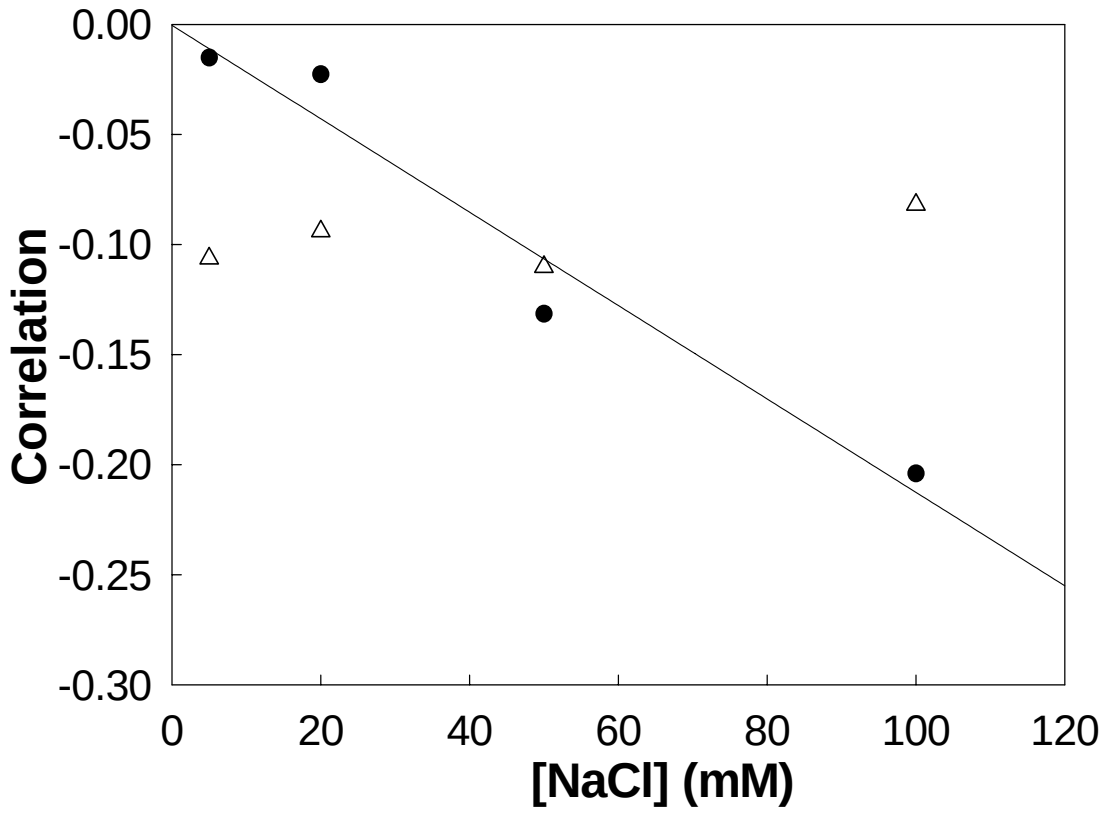

Fig. 6.5. The linear correlation coefficient, ρ, of F and ΔL versus salt concentration. Solid black circles represent the nucleosomal events (definition in the text); empty triangles represent non-nucleosomal events. The slope and intercept of the line are $-2.1\times10^{-3} \pm 3.3\times10^{-4}$ (mM^{-1}) and $4.2\times10^{-4} \pm 1.9\times10^{-2}$, respectively.

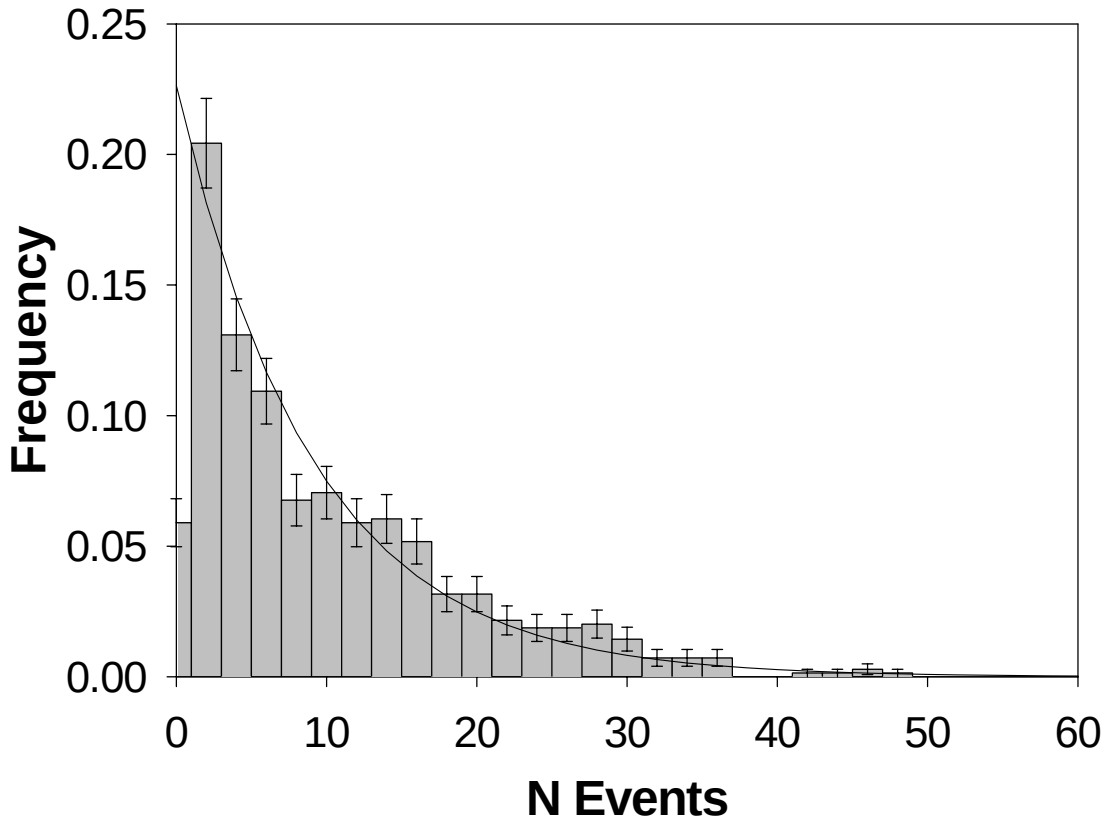

Fig. 6.6. The distribution of the number of observed nucleosomal events (55 bp < ΔL < 95 bp) on first pulls with a bin size of 2 events. Only those pulls that reached F_{max} > 40 pN were included, as not to low bias the distribution. Likewise, subsequent pulls of any particular fiber were not included. The individual distributions at each salt were statistically indistinguishable; therefore they were all included in one distribution to improve overall statistics. Since the population of fibers at each salt originally came from the same sample, this is reasonable. The exponential fit ($N \geq 2$) reflects that, as a whole, the nucleosome arrays were under assembled and nearly impossible to unravel fully, as discussed in the text.

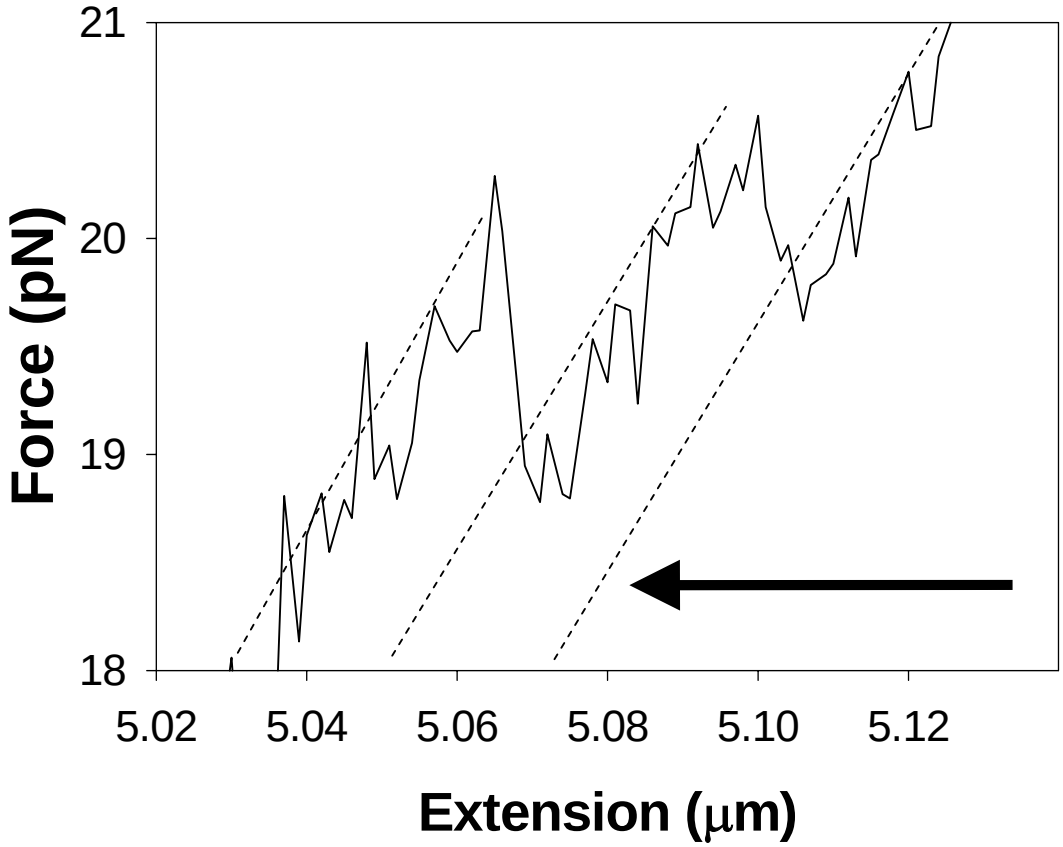

Fig. 6.7. An example of the spontaneous rewrapping of DNA around histone cores upon the relaxation of the tether. The black arrow indicates the direction of travel in time through the force-extension plot. The dotted lines are drawn in as a visual aid; they are ~ 25 nm apart, consistent with one turn of DNA rewrapping around the histone core. Observed rewrappings, however, were rare (see text). Upon rewrapping, their structures assumedly do not represent those found in canonical nucleosomes with specific contact points, because as ACF, as a remodeling factor, was inactive *in situ* (no ATP).

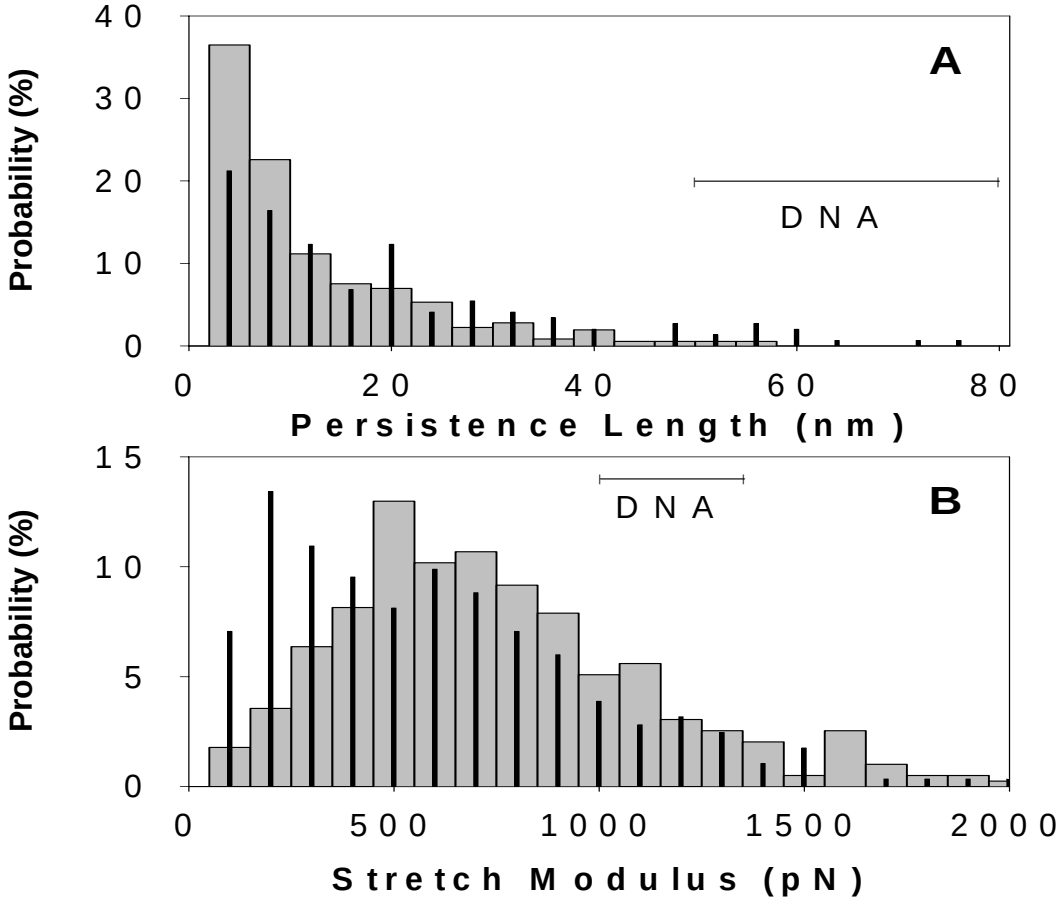

Fig. 6.8. Persistence length (*P*) and stretch modulus (*S*) were obtained by fitting the force-extension data to the extensible WLC model. The following expression, valid in the range of forces covered, was used: $x/L = 1 - \frac{1}{2}(kT/FP)^{1/2} + F/S$, where x/L is fractional extension, kT is the thermal energy, F is force (34). (a) Histogram of persistence length values for 100 mM (gray bars) and 5 mM NaCl (black needles). The scale bar indicates the range for naked DNA (36,37). As *P* measures extensibility at low force, only sections of data extending below 5 pN and spanning at least 3 pN (the average span was 14 pN) were fit (N = 359 fits). All sections of data satisfying these criteria, whether occurring before, between, or following observed unraveling events were included. Sections following unraveling events in one stretch cycle were included because multiple stretching cycles usually revealed many additional unraveling events, as shown in Fig. 6.2. (b) Histogram of stretch modulus (*S*) values for 100 mM (gray bars) and 5 mM NaCl (black needles). The scale bar indicates the range of reported values for naked DNA (36,37). As *S* measures extensibility at high force, only sections extending above 20 pN and spanning at least 5 pN (the average span was 18 pN) were fit (n = 396).

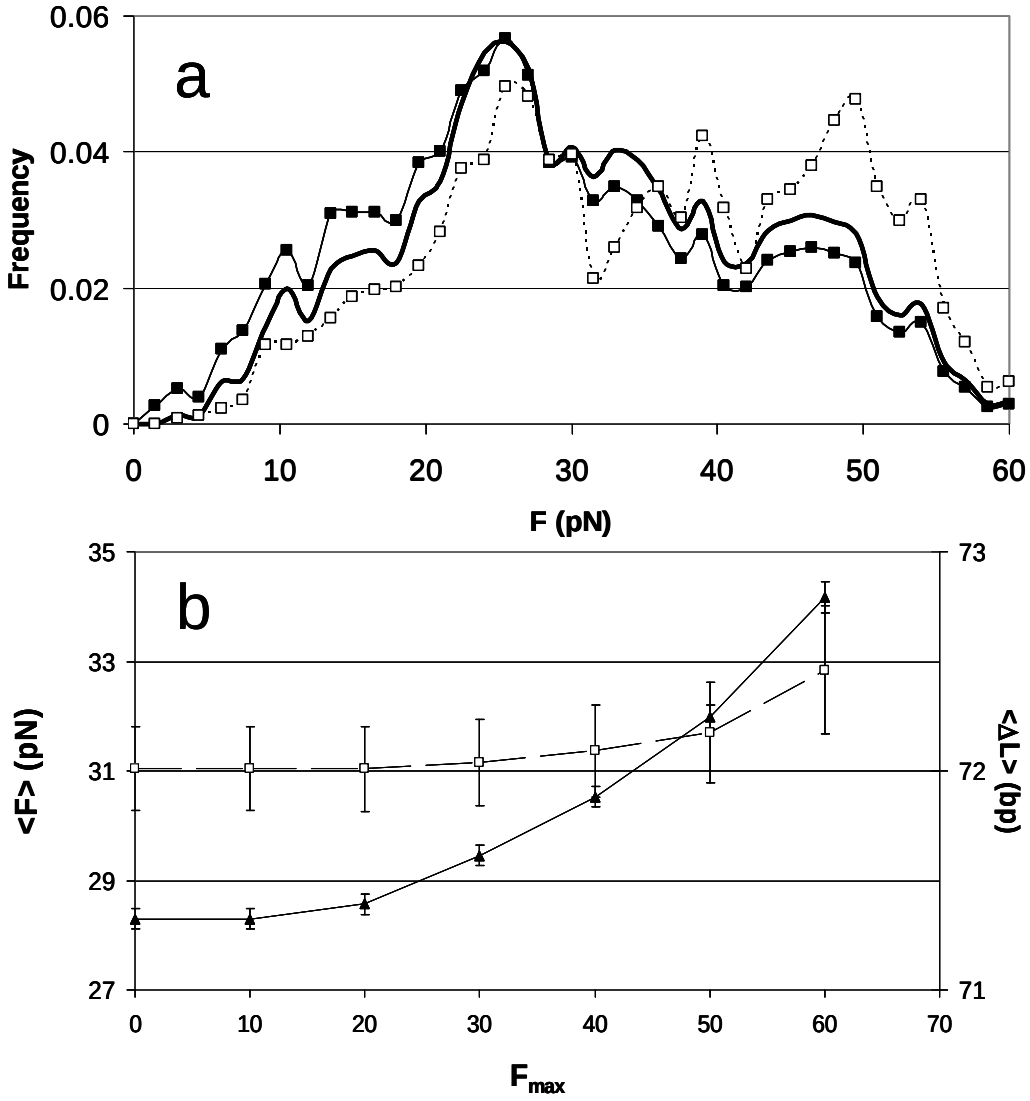

Fig. 6.9. Effects of F_{max} on the force and ΔL distributions of the 5 mM NaCl nucleosomes (55 bp to 95 bp). (a) The normalized force distributions for F_{max} = 0 pN (black squares and thin line), F_{max} = 40 pN (heavy black line), and F_{max} = 60 pN (empty squares and dotted line). (b). The <F> (black squares and solid line) and <ΔL> (empty squares and dashed line) as a function of F_{max}. Uncertainties for each case by standard error. The <ΔL> values use the right hand vertical axis.

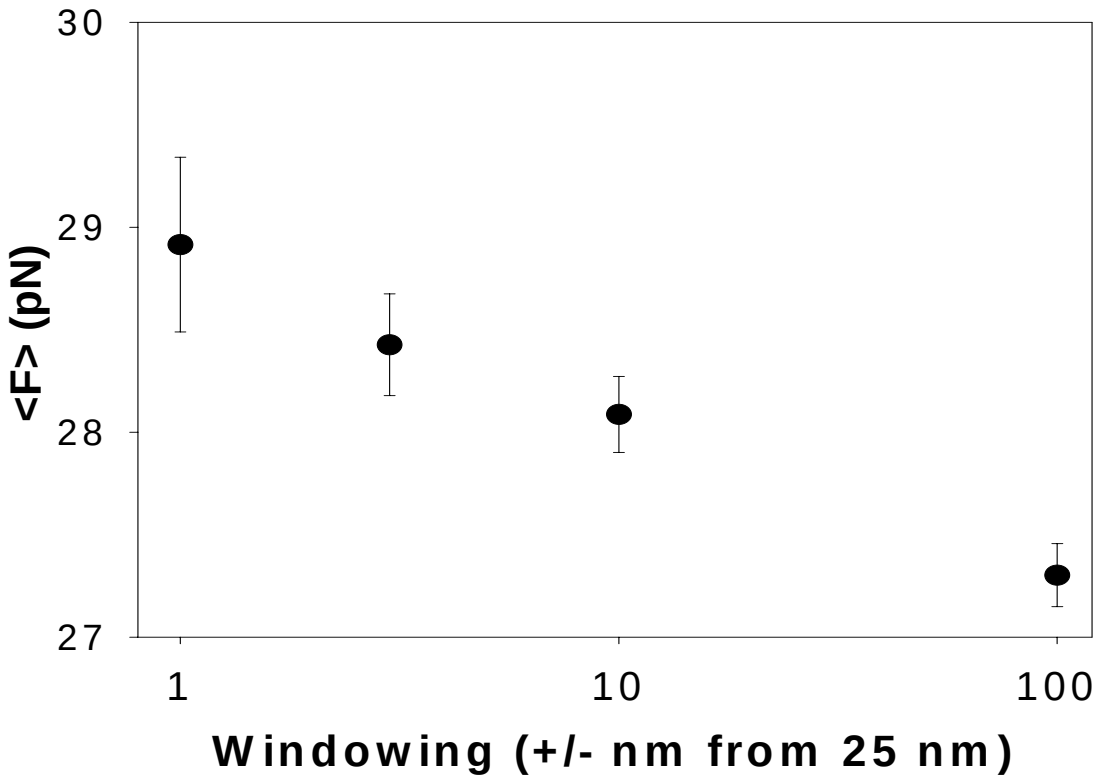

Fig. 6.10. Average unraveling force for nucleosomes as a function of the windowing about 25 nm. The horizontal axis is on a log-10 scale. As discussed in the text, the position of the widest window point on the horizontal axis is somewhat arbitrary.

References for Chapter 6

1. Wolffe, A. (1998). *Chromatin: Structure and Function*. 3rd edit., Academic Press, San Diego.

2. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A° resolution. *Nature*. **389**: 251–260.

3. Elgin, S. C. R. & Workman, J. L. (2000). *Chromatin Structure and Gene Expression*. 2nd edit., Oxford University Press, Oxford, NY.

4. Widom, J. (2001). Role of DNA sequence in nucleosome stability and dynamics. *Quart. Rev. Biophys*. **34**: 269–324.

5. Lusser, A. & Kadonaga, J. T. (2003). Chromatin remodeling by ATP-dependent molecular machines. *Bioessays.* **25**: 1192–1200.

6. Anderson, J. D., Lowary, P. T. & Widom, J. (2001). Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites. *J. Mol. Biol*. **307**: 977–985.

7. Wang, M. D., Schnitzer, M. J., Yin, H., Landick, R., Gelles, J. & Block, S. M. (1998). Force and velocity measured for single molecules of RNA polymerase. *Science*. **282**: 902–907.

8. Wuite, G. J., Smith, S. B., Young, M., Keller, D. & Bustamante, C. (2000). Single-molecule studies of the effect of template tension on T7 DNA polymerase activity. *Nature*. **404**: 103–106.

9. Maier, B., Bensimon, D. & Croquette, V. (2000). Replication by a single DNA polymerase of a stretched single-stranded DNA. *Proc. Natl Acad. Sci. USA*. **97**: 12002–12007.

10. Cui, Y. & Bustamante, C. (2000). Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. *Proc. Natl Acad. Sci. USA*. **97**: 127–132.

11. Ladoux, B., Quivy, J. P., Doyle, P., du Roure, O., Almouzni, G. & Viovy, J. L. (2000). Fast kinetics of chromatin assembly revealed by single-molecule video-microscopy and scanning force microscopy. *Proc. Natl Acad. Sci. USA*. **97**: 14251–14256.

12. Bennink, M. L., Leuba, S. H., Leno, G. H., Zlatanova, J., de Grooth, B. G. & Greve, J. (2001). Unfolding individual nucleosomes by stretching single chromatin fibers with optical tweezers. *Nature Struct. Biol*. **8**: 606–610.

13. Leuba, S. H., Karymov, M. A., Tomschik, M., Ramjit, R., Smith, P. & Zlatanova, J. (2003). Assembly of single chromatin fibers depends on the tension in the DNA molecule: magnetic tweezers study. *Proc. Natl Acad. Sci. USA*. **100**: 495–500.

14. Brower-Toland, B. D., Smith, C. L., Yeh, R. C., Lis, J. T., Peterson, C. L. & Wang, M. D. (2002). Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proc. Natl Acad. Sci. USA*. **99**: 1960–1965.

15. Simpson, R. T., Thoma, F. & Brubaker, J. M. (1985). Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: a model system for study of higher order structure. *Cell*. **42**: 799–808.

16. Kulic, I. M. & Schiessel, H. (2004). DNA spools under tension. *Phys. Rev. Lett*. **92**: 228101.

17. Ito, T., Bulger, M., Pazin, M. J., Kobayashi, R. & Kadonaga, J. T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell*. **90**: 145–155.

18. Fyodorov, D. V. & Kadonaga, J. T. (2003). Chromatin assembly in vitro with purified recombinant ACF and NAP-1. *Methods Enzymol*. **371**: 499–515.

19. Haushalter, K. A. & Kadonaga, J. T. (2003). Chromatin assembly by DNA translocating motors. *Nature Rev. Mol. Cell. Biol.* **4**: 613–620.

20. Lusser, A. & Kadonaga, J. T. (2004). Strategies for the reconstitution of chromatin. *Nature Methods*. **1**: 19–26.

21. Fyodorov, D. V. & Kadonaga, J. T. (2002). Dynamics of ATP-dependent chromatin assembly by ACF. *Nature*. **418**: 897–900.

22. Alexiadis, V. & Kadonaga, J. T. (2002). Strand pairing by Rad54 and Rad51 is enhanced by chromatin. *Genes Dev*. **16**: 2767–2771.

23. Jiang, W., Nordeen, S. K. & Kadonaga, J. T. (2000). Transcriptional analysis of chromatin assembled with purified ACF and dNAP1 reveals that acetyl-CoA is required for preinitiation complex assembly. *J. Biol. Chem*. **275**: 39819–39822.

24. Ura, K., Araki, M., Saeki, H., Masutani, C., Ito, T., Iwai, S., *et al*. (2001). ATP-dependent chromatin remodeling facilitates nucleotide excision repair of UV-induced DNA lesions in synthetic dinucleosomes. *EMBO J.* **20**: 2004–2014.

25. Thummel, C. S., Burtis, K. C. & Hogness, D. S. (1990). Spatial and temporal patterns of E74 transcription during Drosophila development. *Cell*. **61**: 101–111.

26. Ucker, D. S. & Yamamoto, K. R. (1984). Early events in the stimulation of mammary tumor virus RNA synthesis by glucocorticoids. Novel assays of transcription rates. *J. Biol. Chem*. **259:** 7416–7420.

27. Bustamante, C., Smith, S. B., Liphardt, J. & Smith, D. (2000). Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol*. **10**: 279–285.

28. Smith, S. B., Cui, Y. & Bustamante, C. (1996). Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*. **271**: 795–799.

29. Brower-Toland, B., Wacker, D. A., Fulbright, R. M., Lis, J. T., Kraus, W. L. & Wang, M. D. (2005). Specific contributions of histone tails and their acetylation to the mechanical stability of nucleosomes. *J. Mol. Biol*. **346**: 135–146.

30. Pope, L. H., Bennink, M. L., van Leijenhorst-Groener, K. A., Nikova, D., Greve, J., & Marko, J. (2005). Single chromatin fibre stretching reveals physically distinct populations of disassembly events. *Biophys. J.* **88**: 3572–3583.

31. Marko, J. F. & Siggia, E. D. (1997). Driving proteins off DNA using applied tension. *Biophys. J.* **73**: 2173–2178.

32. Polach, K.J., & Widom, J. (1999). Restriction enzymes as probes of nucleosome stability and dynamics. *Methods in Enzymology.* **304:** 278-298

33. Li, G., Levitus, M., Bustamante, C., & Widom, J. (2005). Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct. Mol. Biol.* **12**(1): 46-53.

34. Odijk, T. (1995). Stiff chains and filaments under tension. *Macromolecules*. **28**: 7016–7018.

35. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. (1994). Entropic elasticity of lambda-phage DNA. *Science*. **265**: 1599–1600.

36. Baumann, C. G., Smith, S. B., Bloomfield, V. A. & Bustamante, C. (1997). Ionic effects on the elasticity of single DNA molecules. *Proc. Natl Acad. Sci. USA*. **94**: 6185–6190.

37. Wang, M. D., Yin, H., Landick, R., Gelles, J. & Block, S. M. (1997). Stretching DNA with optical tweezers. *Biophys. J.* **72**: 1335–1346.

38. Evans, E. (2001). Probing the relation between force– lifetime–and chemistry in single molecular bonds. *Annu. Rev. Biophys. Biomol. Struct*. **30**: 105–128.

39. Godde, J. S. & Wolffe, A. P. (1995). Disruption of reconstituted nucleosomes: The effect of particle concentration, $MgCl_2$ and KCl concentration, the histone tails, and temperature. *J. Biol. Chem.* **270**: 27399–27402.

40. Davey, C. A.& Richmond, T. J. (2002).DNA-dependent divalent cation binding in the nucleosome core particle. *Proc. Natl Acad. Sci. USA*. **99**: 11169–11174.

41. Shrader, T. E. & Crothers, D. M. (1989). Artificial nucleosome positioning sequences. *Proc. Natl Acad. Sci. USA*. **86**: 7418–7422.

42. Cotton, R. W. & Hamkalo, B. A. (1981). Nucleosome dissociation at physiological ionic strengths. *Nucl. Acids Res.* **9**: 445–457.

43. Lilley, D. M. & Pardon, J. F. (1979). Structure and function of chromatin. *Annu. Rev. Genet.* **13**: 197–233.

44. Thastrom, A., Gottesfeld, J. M., Luger, K. & Widom, J. (2004). Histone-DNA binding free energy cannot be measured in dilution-driven dissociation experiments. *Biochemistry*. **43**: 736–741.

45. Widom, J. (1999). Equilibrium and dynamic nucleosome stability. *Methods Mol. Biol.* **119**: 61–77.

46. Lilley, D., Jacobs, M. & Houghton, M. (1979). The nature of the interaction of nucleosomes with a eukaryotic RNA polymerase II. *Nucl. Acids Res*. **7**: 377–399.

47. Yager, T. D. & van Holde, K. E. (1984). Dynamics and equilibria of nucleosomes at elevated ionic strength. *J. Biol. Chem.* **259**: 4212–4222.

48. Yager, T. D.,McMurray, C. T. & van Holde, K. E. (1989). Salt-induced release of DNA from nucleosome core particles. *Biochemistry.* **28**: 2271–2281.

49. Claudet, C., Angelov, D., Bouvet, P., Dimitrov, S. & Bednar, J. (2005). Histone octamer instability under single molecule experiment conditions. *J. Biol. Chem*. **280**: 19958–19965.

50. Hansen, J.C. (2002) Conformational Dynamics of the Chromatin Fiber in Solution: Determinants, Mechanisms, and Functions. *Annu. Rev. Biophys. Biomol. Struct*. **31**: 361-392.

51. Ausio, J., Dong, F., & van Holde, K.A. (1989) Use of selectively trysinized nucleosome core particles to analyze the role of the histone "tails" in the stabilization of the nucleosome. *J. Mol. Biol.* **206**: 451-463.

52. Baneres, J.L., Martin, A., & Parello, J. (1997) The N tails of histones H3 and H4 adopt a highly structured conformation in the nucleosome. *J. Mol. Biol*. **273**: 503-508.

53. Hayes, J.J. & Hansen, J.C. (2001) Nucleosomes and the chromatin fiber. *Curr. Opin. Genet. Dev.* **11**: 124-129.

54. Lee, K.M & Hayes, J.J. (1997) The N-terminal tail of histone H2A binds to two distinct sites within the nucleosome core. *Proc. Natl. Acad. Sci. USA*. **94**: 8959-8964.

55. Mutskov, V., Gerber, D., Angelov, D., Ausio, J., Workman, J., & Dimitrov, S. (1998) Persistent interaction of core histone tails with nucleosomal DNA following acetylation and transcription factor binding. *Mol. Cell. Biol.* **18**: 6293-6304.

56. van Holde, K.A. (1989) *Chromatin*. Springer-Verlag, New York.

57. Lilley, D.M.J., & Tatchell, K. (1977) Chromatin Core Particle Unfolding Induced by Tryptic Cleavage of Histones. *Nuc. Acids Res.* **4**(6): 2039-2055.

58. Gargia-Ramirez, M., Dong, F., & Ausio, J. (1992) Role of the Histone Tails in the Folding of Oligonucleosomes Depleted of Histones H1. *J. Biol. Chem*. **267**: 19587-19595.

59. Fletcher, T.M. & Hansen, J.C. (1995) Core Histone Tail Domains Mediate Oligonucleosome Folding and Nucleosomal DNA Organization through Distinct Molecular Mechanisms. *J. Biol. Chem*. **270:** 25359-25362.

60. Fletcher, T.M. & Hansen, J.C. (1996) The Nucleosomal Array: Structure/Function Relationships. *Crit. Rev. Eukaryot. Gene Expr.* **6:** 149-188.

61. Schwarz, P.M., Felthauser, A., Fletcher, T.M., & Hansen, J.C. (1996) Reversible oligonucleosome self-association: dependence on divalent cations and core histone tail domains. *Biochemistry.* **35**: 4009-4015.

62. Mangenot, S., Leforestier, A., Vachette, P., Durand, D., & Livolant, F. (2002) Salt-Induced Conformation and Interaction Changes of Nucleosome Core Particles. *Biophys. J.* **82**: 345-356.

63. Studitsky, V. M., Clark, D. J. & Felsenfeld, G. (1995). Overcoming a nucleosomal barrier to transcription. *Cell*. **83**: 19–27.

64. Eggleston, A. K., O'Neill, T. E., Bradbury, E. M. & Kowalczykowski, S. C. (1995). Unwinding of nucleosomal DNA by a DNA helicase. *J. Biol. Chem*. **270**: 2024–2031.

Chapter 6 Acknowledgement

The text of Chapter 6, in part, is a reprint of the material as it appears in Gemmen, G. J., Sim, R., Haushalter, K.A., Ke, P.C., Kadonaga, J.T., & Smith, D.E. (2005). Forced Unraveling of Nucleosomes Assembled on Heterogeneous DNA Using Core Histones, NAP-1, and ACF. *J. Mol. Biol.* **351**: 89-99. The dissertation author was the primary researcher and author for the research which forms the basis of this chapter.

Chapter 7. Conclusion

The previous five chapters detailed completed scientific projects, each of which has been submitted for publication to various journals. The relevance of each project is discussed in each chapter, as are potential future directions. However, the projects in the previous chapters were not the only undertakings in the laboratory. Several projects were pursued with different levels of success. Some were shelved because they increasingly seemed intractable; others were shelved because the projects of the previous chapters began working smoothly. The tweezers was used to assist in other people's work, while other projects "looked good on paper" but were never initiated experimentally. In this chapter I would like to give a brief overview of those endeavours; many of them may prove to be excellent projects on the new tweezers system in the group.

7.1. False Starts, Blind Alleys, and Preliminary Results

7.1.1. In Situ Nucleosome Assembly Experiments

Before pulling on the pre-assembled nucleosome arrays (Chapter 6), we pursued *in situ* assembly. We originally hoped to ascertain the real time dynamics of the ACF system, as discussed in Chapter 6. In preliminary work, we experimented with *in situ* assembly on single DNA molecules by flowing in core histones, NAP-1, and ACF at the concentration used in bulk assembly, as well as with a 1/10 dilution of core histones. In both cases we observed rapid compaction, similar to previous findings (e.g., Refs. 11–13 of Chapter 6). An example of such real time compaction is shown in Fig. 7.1. However, control experiments showed that the compaction was ATP-independent, whereas assembly in bulk is known to be ATP-dependent and

occur ~100 times more slowly (e.g., Ref. 21 of Chapter 6). Indeed, we observed similar compaction rates with only core histones. Leuba, *et al*., however, claimed to have observed nucleosome assembly *in situ* with only core histones and NAP-1. That such a wide variety of conditions, many of which should not yield proper nucleosomes, all gave similar compaction profiles illustrates the confusion and complications inherent in the *in situ* assembly problem. It was these complications, as well as the fact that there was no way of independently assessing the quality of *in situ* complexes, that ultimately led us to use pre-assembled nucleosome arrays.

A second approach to in situ assembly that we pursued was the flow-in of *drosophila* cell extract to the tethered DNA molecule. Cell extracts, which contain a host of assembly factors and other nucleic acids, are typically the most robust method of assembly in bulk. However, the many other architectural proteins make interpreting the force-extension curves from *in situ* assembly attempts very difficult. Other groups have claimed to have successfully assembled nucleosomes *in situ* by using *Xenopus* extracts (e.g., see Refs. 12 and 30 of Chapter 6). Our preliminary results with extract resulted in compaction rates comparable to those observed for the purified protein *in situ* assemblies. However, similar compaction rates were observed for cell extracts that were not competent for assembly in bulk. In other words, the compaction was simply due to the non-specific aggregation of protein on the tethered DNA. Additionally, no preponderance of events of the proper ΔL were observed when stretching these *in situ* extract assembled complexes. Finally, as shown in the inset of Fig. 7.1, the extracts were full of 'dirt' and protein 'goo' that fell into the optical trap and often broke the tether. Filtering the extract solutions did not help, as

the ‘goo’ seemed to recoagulate. Much of the debris would burn when hit the infrared laser, which was potentially very damaging to the flow cell making further measurements impractical. The other groups that utilized extracts alluded to this debris problem, but assumedly they did not suffer from it as badly as we did.

While the force-extension curves of the *in situ* extract assembled complexes did not yield any promising results, some of the force-extension curves from the purified protein *in situ* assemblies demonstrated some nucleosomal traits similar to those reported elsewhere. An example is shown in Fig. 7.2. While macroscopically these curves looked no different from the aforementioned non-specific aggregation curves, the inset shows the presence of lengthening events of a very suggestive length. Additionally, these events occurred around a 20 pN plateau, as was observed for salt-dialyzed nucleosomes (Ref. 14 of Chapter 6). Lastly, the sloping region at low extensions may represent the unrolling of the outer turn of DNA from the nucleosomes, as suggested by Brower-Toland, *et al*. (Ref. 14 of Chapter 6). We did not see this region with the pre-assembled nucleosomes; perhaps putting the arrays onto beads caused their unwrapping before being pulled apart in the tweezers.

7.1.2. Systematic “Shortening” of the Unraveled Nucleosome

As discussed in Chapter 6, our observations of the forced unraveling of nucleosomes indicated a length of ~ 75 bp, as did the observations of other groups. This was attributed to the inner turn of DNA being popped off the histone core suddenly whereas the outer 0.7^{th} turn of DNA was slowly peeled off. One turn of DNA in the canonical nucleosome, however, represents 146 bp/1.7 turns ≈ 86 bp. Originally this discrepancy was attributed to a ~10 bp region about the dyad axis of

the nucleosome which was very strongly bound to the histone core and did not dissociate irreversibly. This lack of dissociation is what might have facilitated the observed spontaneous rewrapping of nucleosomes, a phenomenon confirmed by results from the Bustamante lab at UC Berkeley (Shirley Mihardja, private communication). This non-dissociated 10 bp spans about 40º, so that the hypothetical distance spanned in the pulling direction is ~ 9.8 bp. Thus after dissociation of this 10 bp region, the measured increase in tether length will be roughly 10 bp – 9.8 bp = 0.2 bp, a distance well below what even the best tweezers can measure in real time. Thus the discrepancy can not be due to the strongly bound region.

As we saw in Chapters 4 and 5 for DNA loops, however, the protein bridging can yield systematically shortened tether lengthening measurements. The same is true for unraveling nucleosomes. As shown schematically in Fig. 7.3, under tension the outer turn of the DNA will unpeel smoothly from the histone core and then turn to align best with the applied force. This turning was recognized by Kulíc and Schiessel (see Chapter 6 references). Once turned, however, the DNA entering and exiting the nucleosome is separated by a distance δ, as indicated in the figure. As a measured event in the tweezers is the length between the beads after the event less the length between the beads before the event, the measured unraveling will the one turn of DNA less δ. For a nucleosome, δ will be a minimum of the diameter of the DNA (~2 nm), which corresponds to a 6 bp distance, and a maximum of the thickness of the histone core, which is ~ 8 nm (25 bp). Therefore, the effective protein-bridging brought about the nucleosome structure might well explain this widely observed ~10 bp shortcoming. A theoretical concern for this model, however, is the kinking of the

DNA necessary (center panel of Fig. 7.3) under tension; it might well be that energy necessary to affect this kinking is more than the binding energy between the histone core and the entering/exiting DNA.

7.1.3. Homing Endonuclease Project

Another project which was not pursued past preliminary results was the measurement of the binding energy of the homing endonucleases (HEs). Unlike the real time chromatin assembly project, which was abandoned because of its intractability, the HEs project was shelved because the projects with the two site restriction enzymes began working successfully; the body of work derivative from the two site enzymes was extensive. Combined with the pre-assembled nucleosomes project becoming tractable, as well as the PCR labeling project, there was simply no time to pursue the HEs project.

The HEs project was originally intended to be an extension of the one site REase project, before it became an extension of the two site cleavage (Chapter 3). Before realizing that DNA bending by the enzyme was the critical factor in force inhibition of cleavage, we suspected that we could observe an effect due to the base pair stretching of DNA under tension. In a lock and key motif, the thought was that enzymes with longer recognition sequences would be more inhibited than would enzymes with shorter recognition sequences. The overwhelming majority of enzymes have 4–8 bp recognition sequences. The commercially available HEs, however, have extremely long recognition sequences (PI-SceI (39 bp), PI-PspI (30 bp), I-SceI (18 bp), I–CeuI (26 bp)), so the thought was that they would be especially susceptible to the effect of base pair spreading.

There are two prominent hurdles with studying HEs. First, because their recognition sequences are so long, the probability of getting site is low. Fortunately, the pBAC template discussed in previous chapters had a PI-SceI site; systematically studying these enzymes would have required an engineered, labeled template with all the sites. Second, HEs, in general, do not cleave the DNA and release the products like restriction enzymes. Instead, HEs cleave but do not dissociate from the DNA, holding the resulting pieces of DNA together as a "protein staple."

In bulk, this staple is typically undone by an SDS/proteinase K treatment, as shown schematically in the upper branch of Fig. 7.4. In the tweezers, however, we hoped to pull apart the protein staple under tension, as shown schematically in the lower branch of Fig. 7.4. If the forced removal of the staple was observable, utilizing the DFS mentioned in Chapter 2 would have allowed us to measure the equilibrium binding energy. Discerning the forced dissociation of the homing endonuclease (and the consequent disruption of the DNA tether) could have been discernable from simple tether dissociation by comparison with control data, because in all likelihood the template with sites for all four HEs would have been labeled with a single DIG molecule. Thus a complete study would have been very statistically intensive.

Preliminary work with PI-SceI and the stronger multiDIG pBAC template (which stayed tethered to greater forces and for longer times; see Chapter 5) gave mixed results. The bulk reactions behaved as expected, with the PI-SceI and the cleavage products remaining intact until the SDS/proteinase K treatment. Preloading the DNA with PI-SceI and then putting the DNA on the beads proved to be too much handling, as very few tethers were observed and those that were resembled naked

DNA. Assumedly the handling dislodged the PI-SceI, making subsequent tethering impossible. Loading the DNA onto the beads and then loading the PI-SceI resulted in many hookups, the majority of which looked like naked DNA. Assumedly the enzyme never got onto the DNA because of the beads (e.g., non-specific sticking of the enzyme to the beads). Flowing the PI-SceI into the flow cell with a DNA tethered proved to be difficult as well. With such a huge recognition sequence and only one site on the template, the PI-SceI reaction was rather slow. Even the bulk reaction needed to be run longer than typical restriction digests.

All that said, utilizing the first two schemes did yield a few instances of the tether breaking at ~10 pN, a much lower force than at which this tether typically broke. Definitively stating that these events were staples being pulled apart, however, would have required many controls as there could be many other reasons for a weak tether. Nonetheless, improvements in tethering strength and efficiency (e.g., the amino-carboxyl linkage discussed below) might allow this project to be revisited in the future, as it may aid in the study of sequence recognition.

7.1.4. Listeria Trapping

I also demonstrated that individual *listeria* bacterium could be optically trapped for Partho Ghosh (UCSD). His group hoped to study the forces involved as mammalian cells enveloped this rod-like bacterium, as they have isolated the surface protein involved. However, the geometry and force sensitivity of the optical tweezers was wrong for this project. The cells are put on microscope slides, meaning that the bacterium would have been pulled along the optical axis, a direction in which we could not measure forces. Additionally, the forces involved seemed to be rather low

for the optical tweezers to measure or exert reliably. To circumvent this, a magnetic tweezers was built by members of our group that was sensitive to a lower force range (as most magnetic tweezers are) and that measured forces in the right direction.

7.1.5. Different Motifs in Two Site Cleavage Experiments

An extension to the two site experiments of Chapter 3 that we pursued was the discernment between the two dimerization motifs; we were having limited success until the laser in our tweezers was needed elsewhere. The first motif, utilized by Sau3AI, is one in which each enzyme monomer binds to a different site and two monomers form a dimer. This active dimer then cleaves the DNA. This motif could also be applied to dimers forming tetramers. The second motif is one in which each enzyme unit (dimer or tetramer) has two binding clefts. Each cleft must be filled for the enzyme to be active; examples are EcoRII (dimer) and Cfr10I (tetramer).

With increasing enzyme concentration, the cleavage activity saturates in the former motif, as shown in Fig. 5.3. For the latter motif, too much enzyme saturates the sites, meaning that an unoccupied site can not fill the second binding cleft. Thus, one expects a decrease in activity eventually. We tried to demonstrate this in the tweezers with Cfr10I, the least expensive of the enzymes known to operate by this motif. As the effect had been demonstrated in bulk, however, we decided it was not crucial to finish the study, though it provides an extension of the work of Chapter 3.

7.1.6. Viral Packaging

The ϕ-29 bacteriophage packages DNA very powerfully and quickly compared to most other molecular motors, as was shown by Doug Smith and his

colleagues at UC Berkeley. Shortly after the tweezers were built, we demonstrated that we could observe this DNA packaging, as shown in Fig. 7.5. However, at the time the focus of the group (much smaller then) was the real time chromatin assembly, so the viral packaging was shelved. Additionally, the micropipette tweezers could not achieve the level of noise and drift reduction necessary to extend the packaging studies beyond what was already known. To address that, Peter Rickgauer and Al Schweitzer in our group have built a dual beam tweezers that replaces the pipette with a second laser, removing the thermomechanical coupling to the outside world. While the development of that apparatus has presented its own host of difficulties, I helped with it in whatever ways I could, namely by transferring bead preparation protocols, sharing DNA stocks, providing noise level references, and, ultimately, critical optical parts. Studies of the ϕ-29 bacteriophage are now in full swing on the dual beam tweezers by Peter Rickgauer and Derek Fuller.

7.1.7. Covalent Binding of DNA to Beads

As mentioned many times throughout this dissertation, a limiting factor of the optical tweezers experiments is the weakness of the DIG-αDIG bond; the biotin-streptavidin bond is much stronger. Even at very low forces, the DIG-αDIG bond has a finite lifetime that often interfered with our results, except for the bond strength study of Chapter 2 for which it was the result. For the one site restriction enzyme studies of Chapter 3 and the looping studies of Chapter 5, the multi-DIG template proved to be much stronger. However, the preparation of that DNA template was significantly more difficult and less efficient than the preparation of single-DIG

DNA. Also, the multi-DIG handle could unroll from the bead partially, adding uncertainty to the measurement of genuine tether lengthening events.

Therefore, the development of a protocol by which the DNA could be covalently attached to beads on one end would be valuable. The *in situ* attachment would be the biotin-streptavidin bond. Though none of the experiments in the previous chapters utilized such an attachment scheme, much of the testing of the tethers was done on my tweezers with my assistance. The first attempt at this linkage was started by Aurelie DuPont, who tried to attach an acrydite molecule to the DNA and then the DNA to a thiolated bead. She had only limited success, despite many hours spent on the optical tweezers. The second attempt at a covalent attachment scheme was started by Derek Fuller. He attached an amino-labeled oligo to carboxylated beads, and then ligated the rest of the DNA template (with biotin on the other end) to this oligo. He was ultimately successful, as described in Ref. 12 of Chapter 2.

7.2. Managerial Duties

As I move forward from this point in my career forward, the most valuable lessons from my time in the laboratory were not scientific lessons at all. From the time of building the lab through the many frustrated projects to the collection and analysis of the data to the writing of the papers, the managing of the time and efforts of the people in the laboratory, as well as the materials therein, proved to be the most relentless, challenging, and yet rewarding aspect of completing this dissertation. Ultimately and up until it was dissembled, nothing was done in the lab that was not under my direction. Consequently, it is difficult to quantify or even summarize what that fully entailed. Nonetheless, the constant maintenance of every aspect of the

optical tweezers, the allocation of bead and DNA stocks, the ordering of parts and enzymes, the construction of flow cells and pipettes, the preparation of buffers, and so many other recurrent tasks that simply made it possible to take data are actually what consumed most of time and efforts in the laboratory. Because so many of these resources were time sensitive (especially pertinent to the preparation of beads), I had to consistently make many difficult decisions regarding the priority or sufficiency of data collection. Even when the data for my projects was completely taken, keeping the tweezers afloat for other projects was my responsibility by default, for by then I was the only one who could do it. Perhaps unexpectedly, this responsibility was actually the most enjoyable aspect of working in the laboratory, as it afforded me so much control over the direction of the scientific projects.

The most important aspect of lab management with which I was constantly faced was the management of the time and efforts of other people. Data collection for all of the projects was aided by at least one other person; Rachel and I took much of the data in a day-night shift format. For the nucleosomes, however, the data was taken in a non-stop, two week push. Scheduling the time of people over whom I had no financial or academic sway was not easy, least of all for the midnight to morning shift. Keeping all aspects of the laboratory working smoothly enough to utilize their time and efforts kept me there through many of such shifts. As stated in the "Acknowledgements", however, I was fortunate to have worked with people who were willing to make the necessary sacrifices to finish the experiments.

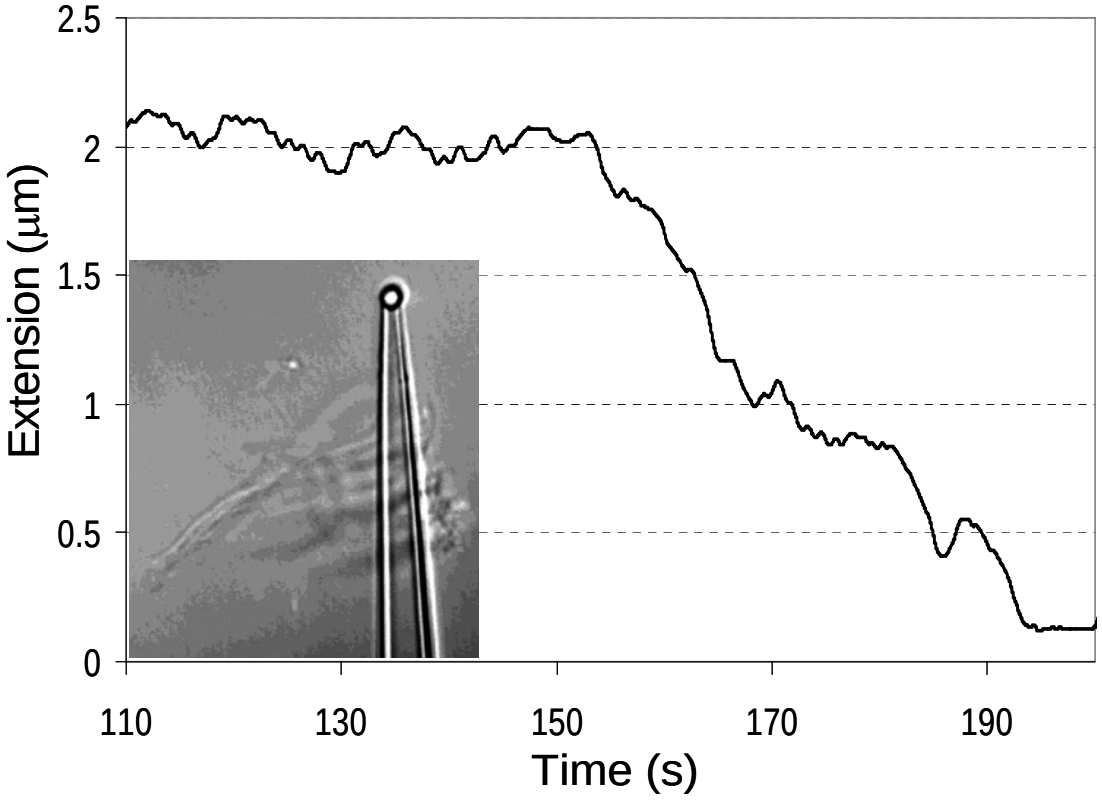

Fig. 7.1. Preliminary observations of chromatin assembly. Full mixture of core histones, ACF, NAP-1, ATP and buffer flowed in, with t = 0 being when the enzyme flow was initiated in the sample tube. The arrow represents the time at which the protein mixture arrived at the DNA molecule. Empirically, this represents a very slow enzyme sample flow. The position of pipette bead was changed 100 Hz in order to maintain a constant 4 pN tension on the DNA molecule (i.e., the force clamp of Chapter 1). Inset: Image grab of protein "goo" often associated with cell extract flow-in. This "goo" would often fall into the optical trap or destroy the tether, making such measurements very inefficient and frustrating.

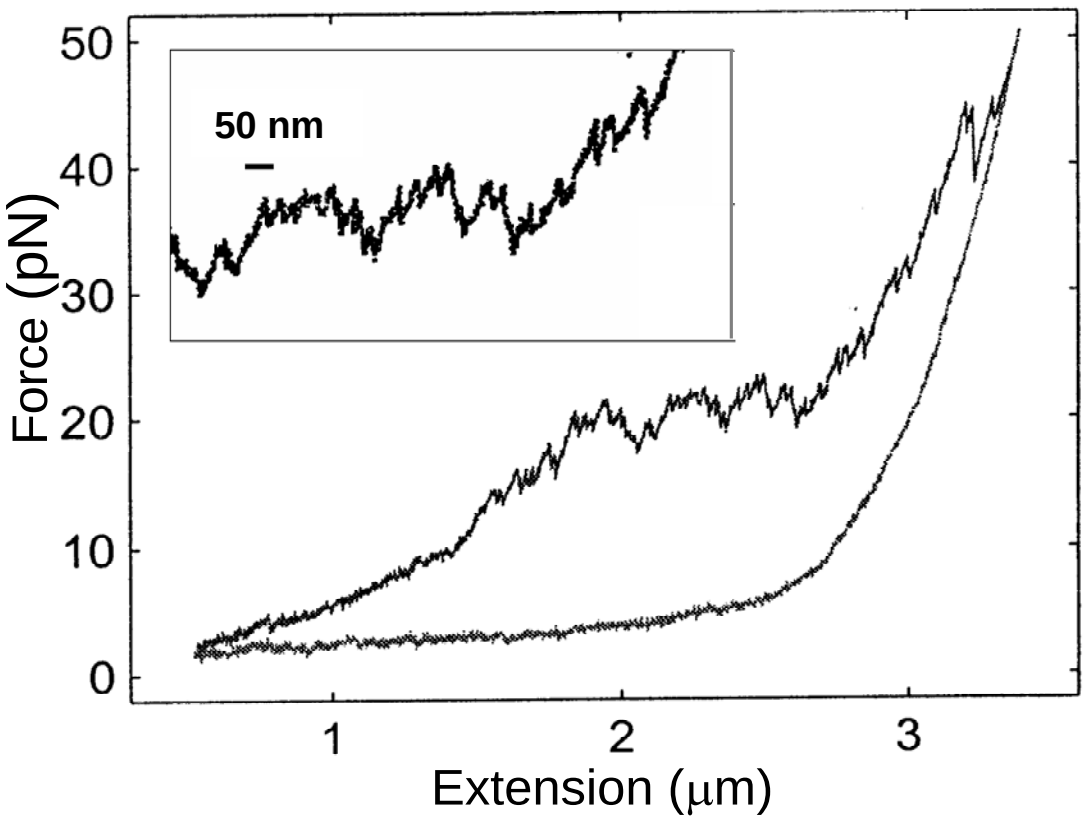

Fig. 7.2. Force-extension curve of a real time assembly run utilizing the full mix of proteins and ATP after the molecule (the single DIG pBAC of previous chapters) was held at 4 pN for several minutes. Note that the force-extension curves looked qualitatively similar to the force-extension curves that we observed for the pre-assembled nucleosome arrays (Chapter 6). Additionally, note the initial sloping of the force-extension curve, which others have interpreted as signature of the *outer* turn of DNA being peeled from the histone core. Also note the grouping of disruption forces about ~ 20 pN. Inset: The portion of the force-extension curve between one and two microns with a 50 nm scale bar. Note the number of events that are ~25 nm.

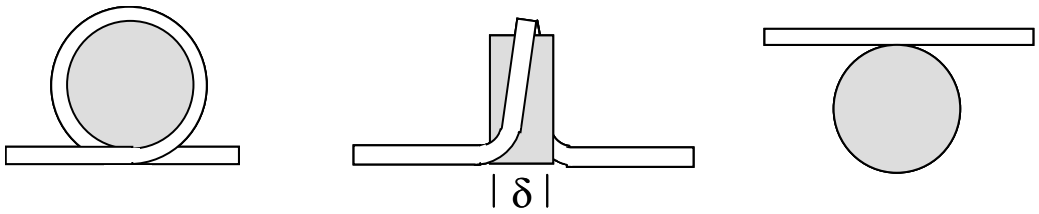

Fig. 7.3. Possible explanation for the systematic discrepancy between the measured increase in tether length for a nucleosome unraveling and the known amount of DNA in one turn about the nucleosome. (Left) A nucleosome experiences the applied tension and its outer turn peels off. (Center) The nucleosome rotates in response to the torque generated by the applied tension (i.e., the nucleosome rotates about a vertical axis in the plane of the figure). In the configuration in the center, the entering and exiting DNA is kinked, but the inner turn remains intact. The distance δ is the distance between the entering and exiting DNA. (Right) After the inner turn pops off suddenly, the measured increase in tether length is the amount of DNA in the inner turn less the initial distance between the entering and exiting DNA, δ. The distance d is at least the diameter of DNA, which is ~2 nm, corresponding to ~ 6 bp. The final panel is not meant to imply any particular geometry for the force-disrupted nucleosome.

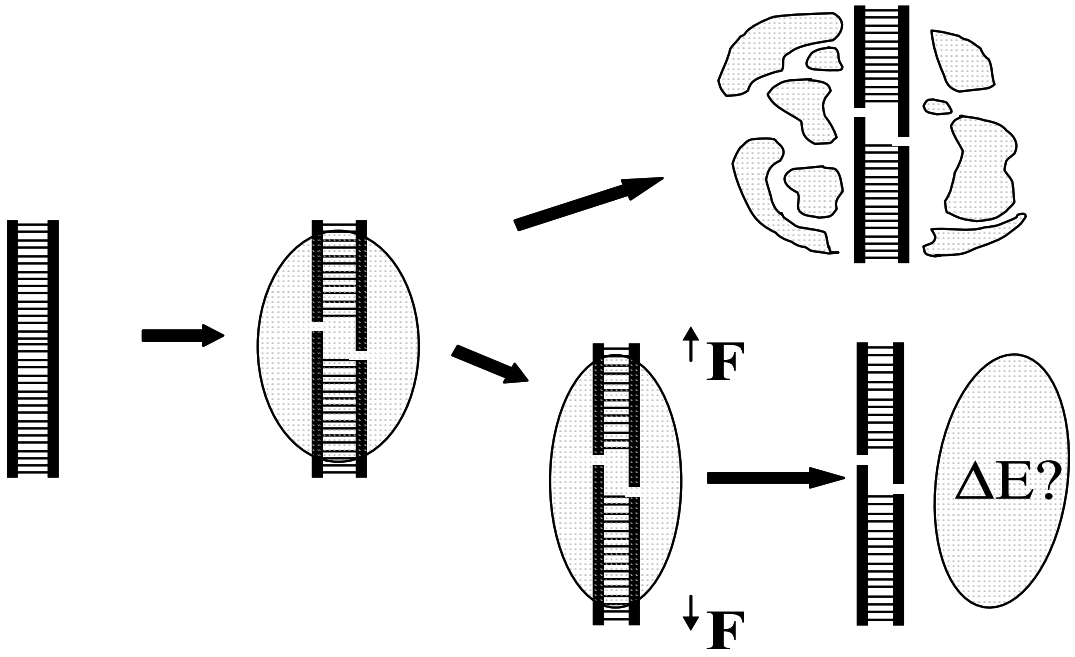

Fig. 7.4. Reaction scheme of the homing endonuclease in bulk (upper branch) and in the optical tweezers (lower branch). The first two images represent the binding of the enzyme to its DNA recognition site (39 bp for PI-SceI). In bulk the separation is done by denaturing the protein "staple" with an SDS/proteinase K treatment before running the gel. In the tweezers one might be able to pull apart the "staple" and, using the DFS mentioned in Chapter 2, ascertain the binding energy ΔE of the endonuclease.

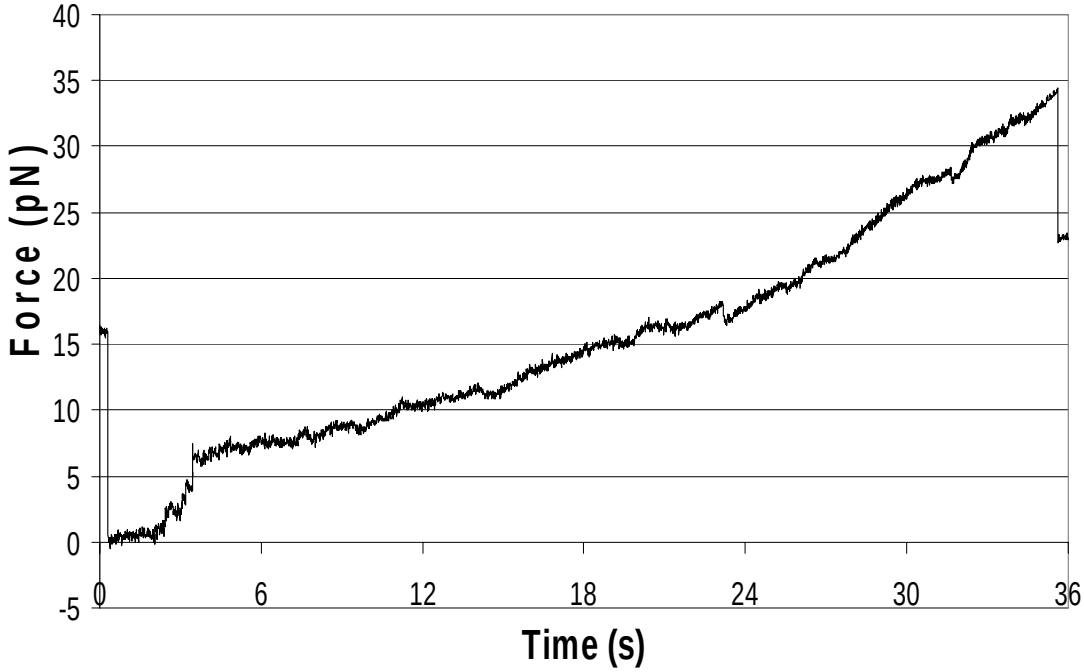

Fig. 7.5. Preliminary observation of ϕ-29 bacteriophage DNA packaging. Beads held at 1 μm apart (from shortly after t = 0) and force recorded as a function of time until the sharp drop off after t=35s. The sharp drop is most likely the DNA slipping back through the motor at high force, as the trace does not fall to zero force.