Simulating the Mechanics of Protein-Induced DNA Looping and Protein-Constrained DNA Buckling

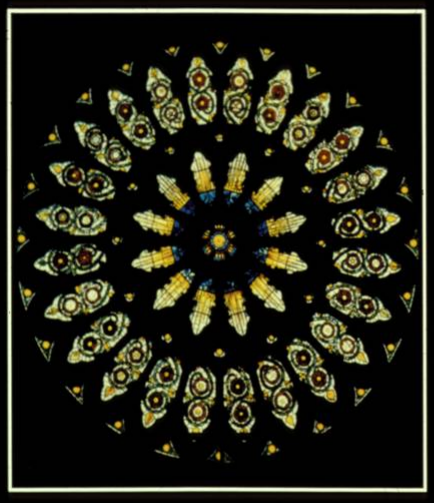
by

Andrew D. Hirsh

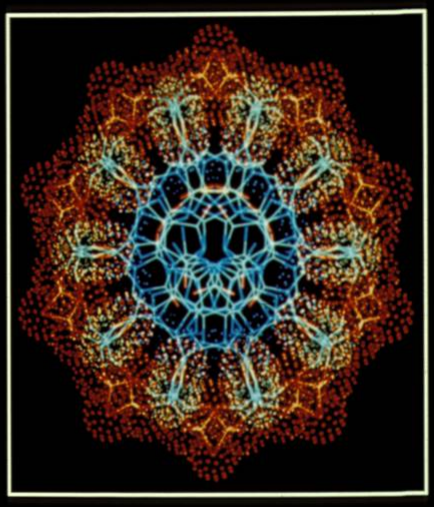
A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mechanical Engineering)
in The University of Michigan
2013

Doctoral Committee:

Professor Noel C. Perkins, Co-Chair
Professor Edgar Meyhöfer, Co-Chair
Associate Professor Ioan Andricioaei, University of California - Irvine
Professor Charles L. Brooks III
Associate Professor Bogdan Epureanu



Rose Window, York Minster
(Photograph by C.J. Bassham)



View Along the Axis
of the β DNA Double Helix
(Courtesy of Dr. R. Langridge)

To my parents Jack and Amy, my wife Jess, and our son Walter.

Acknowledgments

I am very grateful for the opportunity to work with my advisor, Prof. Noel Perkins. Under his patient guidance, I was afforded a great deal of independence to pursue my own research interests. As a mentor, he has provided me a wonderful example to emulate in my professional career.

I would also like to thank the other members of my committee: Prof. Edgar Meyhöfer (Mechanical Engineering, University of Michigan), Prof. Charles Brooks (Chemistry and Biophysics, University of Michigan), Prof. Bogdan Epureanu (Mechanical Engineering, University of Michigan), and Prof. Ioan Andricioaei (Chemistry, University of California - Irvine). Conversations with Prof. Meyhöfer broadened my understanding of molecular biology and challenged me to better frame my computations. I appreciate Prof. Epureanu's enthusiasm for our work and his insight into the relationship between biology and mechanics. I enjoyed conversations with Prof. Brooks about bacteriophage biology and his input helped strengthen the work in Chapter 4. I am also thankful for an ongoing collaboration with Prof. Andricioaei and his graduate student, Maryna Taranova, with whom we collaborated on Chapter 4. Conversations with them have given me a deeper understanding and appreciation for simulating DNA at the atomic level.

I am especially indebted to Prof. Todd Lillian (now in Mechanical Engineering, Texas Tech University) for his patient mentoring and enthusiasm in the lab as he helped me learn about DNA mechanics. I inherited his computational rod model for DNA.

I am also thankful for collaborations with, and guidance from, David Wilson (Postdoc, University of Michigan) and Troy Lionberger (Postdoc, University of California - Berkeley). Dr. Wilson helped me better understand Lac repressor looping and the experiments we studied in Chapter 2. Dr. Lionberger provided insight into work undertaken in Chapter 4 and helped me develop my writing skills. In addition, his participation in the Prison University Project teaching inmates in San Quentin State Prison has deeply inspired me.

During my PhD studies, I have had the pleasure of sharing ideas and collaborating with many bright science and engineering students and faculty including: Rob Brown, Stephen Cain, Sachin Goyal, Aaron Haeusler, Eric Hammond, Charles Chang Jiang, Dave Johnson,

Suyi Li, Ryan McGinnis, Joel Revalee, and Diane Wiener.

Finally, I would like to thank my wife who has been an unwavering source of support. Your faith and love have been a constant encouragement. Thanks for believing in me. Without you, this accomplishment would feel hollow.

I am grateful for the financial support that I have received over the course of my PhD studies. The Mechanical Engineering Department at the University of Michigan provided me with a First Year Department Fellowship. Under my advisor, I have also received continued support from NSF grants CMMI-0825488 and CMMI-0941470.

Table of Contents

List of Tables

Table

List of Figures

List of Appendices

Appendix

Abstract

The bending and twisting mechanics of DNA are known to play a crucial role in many biological processes, yet fundamental details, even in relatively simple systems, remain unclear. The overall objective of this dissertation is to advance this knowledge in the context of two important systems including 1) the role of protein flexibility in an elementary gene regulatory protein that loops DNA, and 2) the structure and biological function(s) of DNA buckling during packaging and ejection in bacterial viruses. To address this objective, we contribute new modeling techniques by extending an elastic rod model for DNA. For protein-induced DNA looping, we use the model to reinterpret two seminal studies on Lac repressor looped DNA including the stability and topology of loops. For viral packaging, we contribute a model for mechanical contact between DNA and a cavity of arbitrary geometry, and also the first multi-scale model for DNA under extreme compression, to examine a buckled DNA toroid recently discovered in a protein cavity below the bacteriophage $\phi 29$ viral capsid. This example further motivates an extensive study of the mechanics of constrained DNA buckles inside the portal cavities of a family of viruses.

The theoretical results for looped DNA successfully predict experimental observations and reveal that an extended protein conformation was active, yet overlooked, in classic experiments by the Müller-Hill laboratory. A detailed analysis of $\phi 29$ reveals that a DNA toroid can form under biologically-relevant force levels supplied by a packing motor. Computed DNA density maps compare favorably with the experimental data. Upon simulating the dynamic ejection of the toroid from the cavity, we reveal that large reaction forces/torques develop at the portal that could be used to signal genome release. Using Greenhill's equation, we show that DNA buckling is also feasible in a number of other bacteriophages including T7 and P22 that contain large portal cavities. Simulating DNA buckles in these cavities shows that large reaction forces develop on the portal walls that could signal a motor to terminate packing. Despite differences in size and shape, the cavities possess the same energy density.

Chapter 1

Introduction

The double helical structure of DNA, first postulated by Watson and Crick [146] in 1953, remains one of the most significant discoveries in modern science. They proposed that DNA was composed of two long helical chains coiled around the same axis to form the iconic double helix. Each polymer chain consists of a sequence of four types of nucleotide subunits: adenine (A), cytosine (C), guanine (G), and thymine (T) linked together by a sugar-phosphate backbone. Hydrogen bonds connect the bases on one chain to those on the complementary chain using the pairing rules that A binds with T and G binds with C. The complementary nature of DNA immediately suggested to them that it was responsible for transferring genetic information, which is carried in the linear sequence of the nucleotide bases. Already, one can observe how the structure of a DNA molecule enables its biological function on the short basepair-level length scale.

Since its discovery 60 years ago, DNA and the famous double helix have continued to make headlines and fascinated seasoned scientists and school children alike. High profile studies have revealed that, remarkably, only about 2% of the human genome's 3.2 billion basepairs (*bp*) code for proteins [62] while the remaining 98% is noncoding DNA. Very recently, and following in the wake of the *Human Genome Project*, the *Encyclopedia of DNA Elements* (ENCODE) *Project* published its initial findings (September 2012) which sought to fully describe all functioning elements in the human genome. The study attributed a biological function to as much as 80% of the genome (see [29] for summary) including noncoding DNA's role as "genetic switches" that regulate expression of the coding DNA. This dispelled the outdated view that the non-coding 98% is merely "junk" DNA and answered the question of why the cell expends the energy to faithfully copy all of its DNA. It is essential to understand how gene regulation operates via these genetic switches, even in the context of simpler prokaryotes, to understand many cellular functions. Such understanding may ultimately lead to developing better disease treatments. This need motivates

the first portion of this dissertation which investigates the mechanics of DNA bound by a gene regulatory protein found in *E. coli.* bacteria that 'loops' approximately 150 bp of highly-bent DNA to regulate expression of downstream genes.

In the lifecycle of the cell, it is also essential that DNA permit genome compaction, a process that strongly depends on the mechanics of the molecule. For example, DNA in a fully-elongated human chromosome measures 3 cm long but is packaged (alongside several other chromosomes) into a cell nucleus only approximately 5 μm in diameter resulting in compaction over four orders of magnitude [11]. Despite compaction, the cell must also preserve accessibility of this genetic information. Another system which highlights DNA compaction is the packaging and ejection of DNA into/from viral capsids. For example, bacteriophage $\phi 29$ packages its micron length genome into a capsid only tens of nanometers wide to near crystalline density using the strongest known molecular motor. The energy stored in the highly-bent DNA is then available to drive the genome into the host during infection. To understand this process, it is again imperative to understand the role of DNA mechanics and how DNA interacts with packing motor proteins. Such knowledge may ultimately foster the development of engineered systems that encapsulate DNA for targeted gene therapy. The second portion of this dissertation investigates the mechanics and function of highly-bent DNA under compression in bacteriophages including $\phi 29$ as well as T7 and P22.

Further motivation for this research on DNA mechanics draws from emerging technologies that include DNA computing [1], DNA origami [115], and affordable data storage [39]. As an example of the latter, scientists encoded all 154 of Shakespeare's sonnets, an audio recording of Martin Luther King's "I Have a Dream" speech, and a JPEG image in DNA and recovered the data with 100% accuracy. It was reported that all text ever written by humans (approximately 50×10^9 megabytes of data) could be stored in a DNA structure with a mass of only tens of grams. These exciting fields are in their infancy but pose significant challenges to engineers. Clearly, understanding how nature exploits the mechanics of DNA molecules will enable engineers to better design these nanoscale devices. These needs underscore the motivation for the studies contained in this dissertation that advance our understanding of the mechanics of DNA.

1.1 DNA Looping

Protein-mediated DNA looping is a well-known gene regulation mechanism. A canonical example of looping occurs in *E. coli.* bacteria to inhibit expression of genes that code for

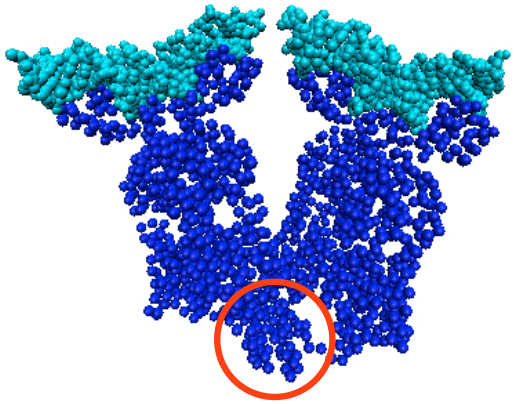


Figure 1.1 Co-crystal structure of the Lac repressor protein bound with DNA operator fragments derived from [78]. Dark blue spheres represents C-α atoms while operator fragments are indicated by cyan. The red circle at the vertex of the 'V' region indicates an area of suspected flexibility. Illustration created using software of [60].

proteins required to metabolize lactose. In the absence of lactose, the Lac repressor protein (LacI) ensures that the Lac operon is 'switched off' by simultaneously binding two distant operator sites upstream and causing the intervening DNA to form a loop which blocks RNA polymerase from the downstream genes in the operon [28, 120]. The loop is subject to considerable bending and twisting. Conversely, DNA looping can also activate transcription by causing NtrC to directly contact RNA polymerase.

1.1.1 Experimental Studies of Lac Repressor Looped DNA

Key to understanding LacI-induced DNA looping and stability is the structure of Lac repressor protein as revealed by the X-ray cocrystal structure of the repressor bound to ideal, palindromic, oligonucleotide operators [33, 78] as shown in Fig. 1.1. This static image of LacI is a dimer of dimers, where the N-terminal headpiece forms a DNA-binding domain and the C-terminal tetramerization domain is composed of a possibly flexible four-helix bundle (red circle in Fig. 1.1). Microscopy at lower resolution of the LacI-DNA complex points to LacI undergoing conformational changes [74, 116, 143].

While LacI has been determined at crystallographic precision, the structure and dynamics of looping remain elusive. A major roadblock to understanding looping is our inability to visualize DNA loops in their native environment. The 2 *nm* diameter of DNA is well below the wavelength of light ($\sim 10^2$ *nm*) rendering visualization in a light microscope impossible. Alternative techniques including traditional electron microscopy and

atomic force microscopy (AFM) are capable of resolving DNA. These techniques, however, require that DNA be removed from its biological conditions or flattened to mount on a substrate. The final structures may likely be tainted and instead reflect the imposed mechanical manipulations. Furthermore, these static images say little about looping dynamics.

An alternative technique which overcomes these limitations is cryo electron microscopy (cryo-EM). It eliminates the need for DNA to be attached to a substrate by flash-freezing it in a thin film (approximately 40 *nm* thick) of its physiological solution [27]. Two orthogonal views of the same image produce a three dimensional helical axis of DNA to within 1 *nm*, but single nucleotides cannot be resolved [64]. To resolve a specific nucleotide sequence, additional factors (with unknown influence) are necessary such as a binding protein or flanking DNA segment to register successive images. The flash freezing process also occurs on fast timescales which does not allow DNA to relax to its final structure; essentially trapping DNA at an instant in time. A single image may therefore provide a snapshot of DNA as thermal perturbations cause it to oscillate about a minimum energy conformation. One is then tasked with gathering and aligning an ensemble of images to obtain a complete understanding of looping. Finally, in these methods observers remain blind to the torsional strain in the molecule.

Extensive experimental studies have revealed fundamental mechanisms controlling protein-mediated DNA looping. One well established behavior is the dependence of loop stability (as measured by repression level [7, 8, 94]) on the interoperator DNA length. Loop stability was found to oscillate with a period of about a helical turn [7, 8, 74, 94]. This periodicity arises simply because the relative angular alignment ('phasing') of the operators that LacI must bind to form a loop. The helical nature of DNA dictates that inserting a basepair into the interoperator DNA alters the phasing by $\sim 34^\circ$ and requires the loop conformation accommodate additional twist to preserve proper alignment of the DNA operators with respect to LacI binding sites. Recent single molecule studies [51, 135] also confirmed this periodic dependence on operator phasing.

Several *in vitro* studies, including seminal contributions from the Müller-Hill lab [73, 74], have aimed to characterize looping using a variety of biochemical techniques in a controlled environment. Specifically, they evaluated DNA loop formation over a broad range of interoperator lengths (153 to 168 bp) using electron microscopy, non-denaturing polyacrylamide gel electrophoresis, and DNase I protection experiments [74]. This battery of experiments provided needed insight into the effects of operator phasing on loop size, stability, energetics, topology (as measured by linking number), degree of substrate supercoiling, and resistance to enzyme degradation. Moreover, [74] identified the optimal interoperator length/phasing that maximizes loop stability. Following these experiments on linear DNA, the Müller-Hill

lab also investigated the effects of supercoiling on DNA minicircles on loop stability and topology [73]. Loop stability was probed by adding Lac inducer (IPTG) or competitor DNA. Results show that the degree of substrate supercoiling alters the interoperator length for optimal loop stability. The topology (i.e. linking number) of the DNA loops were probed using topoisomerase I relaxation assays. The authors conclude that LacI-induced looping can absorb twist up to one superhelical turn.

Despite these and other observations of LacI-induced DNA looping, many unanswered questions remain including the role of protein flexibility. There is mounting evidence (both experimental and theoretical) to suggest that LacI can adopt an "open" or "extended" conformation by pivoting the two dimers about the four-helix bundle. Both electron microscopy and X-ray scattering [116, 132] reveal extended conformations in conjunction with the crystallographic 'V' structure cited above. The extended conformation was proposed to explain how short ($\sim$150 bp) stable loops form with LacI even when the operators are out of phase [152]. Compelling evidence also arises from single molecule tethered particle motion (TPM) experiments which reveal a significant conformational change in LacI [96, 152]. These studies demonstrate that flexibility of the tetramerization domain (at the vertex of the 'V', see Fig. 1.1) is responsible for the observed results as opposed to the two hinge domains (at the top of the 'V') which connect to the DNA-binding heads. Other studies that alter the mechanical stiffness of the protein by introducing hinge mutants [140] and cross linking the protein to prevent opening of the tetramerization domain [117] reveal that looping is completely prevented. Evidence for LacI conformational changes also come from experiments using TPM [117] and AFM [152] that suggest the two looped states interconvert (change binding topologies) without first unlooping. Bulk FRET experiments suggest that the LacI-DNA complex exists in two states (possibly 'V'/extended or parallel/antiparallel) by examining different DNA constructs in which Lac operators bracket a sequence directed bend [30]. Additional studies of other looping proteins demonstrate that loop geometry and energy depend on the geometry of the protein [3, 37, 137].

Given mounting evidence for an extended LacI conformation, it is natural to hypothesize that this structure was active yet overlooked in earlier experiments. A partial answer is found in the study of Swigon et al. [127] that employs a discrete model for DNA to compute the free energy of looping as a function of the interoperator DNA length as in the *in vitro* looping experiments of [73, 74]. Computations by Swigon et al. demonstrate that an extended conformation best agrees with DNase I cutting patterns and the interoperator lengths yielding optimal loop stability [74]. A detailed analysis of the supercoiled DNA minicircle experiments in [73], however, has not been conducted.

1.1.2 Models of the Lac Repressor Looped DNA Complex

Various models for DNA looping have been developed that span multiple levels of resolution and computation effort. Providing the highest level of detail, molecular dynamics (MD) simulation calculates the position of individual atoms in DNA and its surrounding buffer as a function of time. However, the number of atoms composing DNA, the surrounding fluid, and the protein represent an impractically large system to analyze by MD alone - and particularly on biologically-relevant (i.e. pico- to micro-second) time scales [119]. To track the fast vibrations of the atoms, MD requires exceedingly small time steps which limits overall simulation time of significantly sized systems to the pico- to nano-second range. To circumvent this limitation, coarse-grained DNA models have been developed drawing from statistical mechanics [5, 67, 70, 118, 121, 127, 150, 151, 157, 159] and continuum (rod) mechanics [99, 119] and are computationally efficient at the (long) length/time scales governing looping. 'Discrete' models typically treat each basepair as a rigid body elastically coupled to its nearest neighbors [22, 67, 100, 111, 127, 157, 159]. However, their ability to resolve basepair-level detail is contingent upon basepair-level determination of elastic properties. Continuum rod models have been used to simulate DNA looping [6, 44, 82, 135] and have successfully predicted the aforementioned oscillatory behavior of looping free energy [111]. A rod model requires elastic stiffness properties averaged over at least a helical turn and has proven capable of incorporating the sequence-dependent curvature and stiffness properties [44, 46, 82].

Modeling the LacI-DNA complex clearly requires knowledge of the protein structure when bound to DNA. Early modeling efforts have assumed that the protein remains inflexible during looping and rely solely on the known crystal structure [78] to derive boundary conditions. For example, the rod model derives the position and orientation at its boundaries by aligning them with the bound oligo DNA operators in the crystal structure. As mentioned above, several studies have probed the role of protein flexibility and prompted models to consider deviations from the 'V' structure in Fig. 1.1. For example, [67, 118, 127, 157, 158] investigated flexibility of tetramerization domain (at the vertex of the 'V') by introducing approximations for an extended LacI as advised by experiments. Others have accounted for protein flexibility using continuum [79] and MD [141, 142] models of LacI. In the latter, an elastic rod was coupled to an all-atom protein model. Their results implicated head domain flexibility and suggested the protein is stiffer regarding 'V' deformations than previously thought. However, these models necessarily prescribed artificially high reaction forces from DNA on the protein required for a tractable MD simulation.

1.2 DNA Packaging and Ejection in Bacterial Viruses

Bacteriophages, or phages for short, are viruses that infect bacteria. Tailed bacteriophages constitute over 90% of all known prokaryote viruses and are estimated to be the most abundant organisms on the planet. For example, the population is estimated at 10^{31} and the number of infections occurring per second is on the order of Avogadro's number [52]. To put it in perspective, every phage laid end to end could span the distance between the Earth and Sun 10^{13} times. Despite their staggering numbers and frequency of infection, relatively little is known about the mechanistic details underlying phage assembly and infection.

From an engineering perspective, double-stranded DNA (dsDNA) bacteriophages are amazingly sophisticated biochemical machines. Greatly simplified, phages typically package their micron-length genome into a preassembled protein capsid shell only tens of nanometers wide. The well-studied bacteriophage $\phi 29$ is one of the smallest dsDNA bacteriophages and it has long served as a model system to study virus structure, assembly, and ejection [90]. Its relatively small genome of $\sim$19.3 kbp encodes for only $\sim$20 genes. Despite the short 6.6 μm length, the DNA is packed to near crystalline density within the confines of the $\sim$50 nm capsid by the most powerful ATP-driven molecular motor known to approximately 60 atm [123]. Rotary motors typically constitute part of an elaborate protein portal system which is embedded at a single location in the capsid. These remarkable viral DNA packing motors overcome considerable energetic barriers of compaction (due to DNA bending/twisting, electrostatic self-repulsion, and entropy) by generating forces around 100 pN [114, 123]. Skeletal muscle myosin II, by comparison, is only capable of generating 2-3 pN. If viewed as a thin-walled pressure vessel, the capsid would require tensile strength on par with a bulk aluminum alloy. While the nano-chemical motor is very small, its power density is estimated at 5,000 kW/m^3 or about twice that of a sports car engine [16]. The high pressure is thought to be the driving force behind genome ejection. To infect a host cell, the phage binds to host cell receptors and delivers its genome through a syringe-like tail appendage. Clearly, the mechanics of DNA bending and twisting play a role in these systems.

Single molecule optical traps have revealed many subtle but important details employed by phage packing motors to accomplish the seemingly straightforward task of packing DNA into a capsid. Experiments on $\phi 29$ [15, 36, 114, 123], T4 [34], and λ [35] have characterized the kinematics and kinetics of packaging and uncovered surprising behavior. For example, they revealed that packaging is interrupted by pauses and slips where the motor disengages and reengages DNA at a rate that is proportional to the amount of DNA packaged and varies for individual systems. Packaging forces/rates were also shown to depend strongly on ionic concentration as ions tend to screen a portion of the electrostatic repulsion of DNA [36].

These findings are consistent with experiments that alter capsid pressure by modulating parameters including ionic conditions, temperature, binding proteins, and DNA length to induce ejection and/or speed the dynamics of genome release [18, 40, 47, 72, 84, 86, 97, 153]. Recent rotor bead assays spotlight the role of motor rotation during packaging by examining the firing sequence of the $\phi 29$ motor [156]. They discovered that the motor rotates DNA in an untwisting (i.e. left handed) sense and to degrees ranging from approximately $4.8^{\circ}/bp$ [156] to upwards of $10^{\circ}/bp$ near the end of packaging (C. L. Hetherington, Bustamante Lab, personal communication, 2013). T4 and $\phi 29$ motors are even robust enough *in vitro* to package nicked DNA substrate although gapped and mismatched bases impede translocation [93, 102].

A wide variety of models have advanced our understanding of the underlying physics of genome packaging and ejection. These models include Brownian dynamics [124], Monte Carlo [23], elastic rod [104, 109, 110], beaded chain [75, 89, 105, 106], MD [4], and statistical mechanics [61]. Typically models parametrically vary temperature, ionic concentration, and twist during packaging and observe their effects on genome organization and the forces/energetics of packaging. Despite coarse-graining DNA to varying degrees, low-resolution models confirm the observed experimental behaviors.

Another established class of experiments use cryo-EM to obtain rich structural detail of bacteriophages, at different stages over the bacteriophage lifecycle including prohead (i.e. unfilled capsid), mature phage, and emptied. Reconstructions exist for a plethora of phages including $\phi 29$ [129, 131, 154], P22 [14, 17, 76, 128, 130, 160], T4 [32, 77], $\varepsilon 15$ [65], SPP1 [103], P-SSP7 [83], and T7 [2] that have been determined to varying degrees of resolution (approximately 25 Å to 7 Å) to date. Several cryo-EM reconstructions, most notably $\phi 29$ [129, 154] and P22 [128], focus on the portal and couple reconstructions to known high resolution portal crystal structures ($\phi 29$ [122], P22 [98]). The goal is to piece together the underlying mechanisms causing conformational changes and deviations from the crystal structure when embedded in its native system [128, 129, 154]. Nearly all cited reconstructions indicate that the phage portal undergoes at least one conformational change. Unfortunately, many of the studies cited above either reconstruct removed components or do not resolve all lifecycle stages using the same experimental techniques which are necessary for critical comparisons of conformational changes.

The role DNA plays in signaling portal conformational changes remains a fascinating topic of current debate. The first reconstruction of P22 [76] resolved a DNA 'belt' that wrapped around the exterior of the portal and 'squeezed' the connector shut under the capsid pressure and was believed to transmit the 'head-full' signal that terminates motor packaging. Upon discovery of the P22 portal crystal structure [98] and higher resolution cryo-EM [128],

the head-full signal was augmented to also include the pressure sensing 'barrel' atop the portal that resembles a toy Chinese finger trap.

1.2.1 DNA Buckling in $\phi 29$

At the heart of the packing motor is the $\phi 29$ connector, a 12-mer ring of gene product 10 (gp10) proteins [122], which acts as a portal through which DNA is threaded. The connector actively participates in almost every stage of the phage lifecycle. Bacteriophage $\phi 29$, unlike many other tailed DNA bacteriophages, packages a unit length genome, capped with terminal protein (gp3) covalently linked to the 5' ends of the genome. After packing the DNA-gp3 complex, the motor assembly is broken down by dissociation of the ATPase (gp16) and pRNA. The connector is then left available as the binding site for the tail assembly including the lower collar/tail tube (gp11), tail knob (gp9), and appendages (gp12) [154]; see Fig. 1.2. This process may be initiated through interactions between the motor and gp3 on the terminal end of the genome [49]. However, the connector must somehow prevent the highly pressurized DNA from leaking from the prohead after motor disassembly but before tail assembly. Recently, Grimes et al. [49] implicated the connector channel loops, structures which protrude into the center of the connector and contact DNA, as the mechanism responsible for genome retention. In these experiments, mutated connectors with altered channel loops did not significantly affect DNA translocation but significantly hindered phage maturation. Additionally, Geng et al. [31, 38, 66] revealed that the connector exhibits genome gating behavior and undergoes a conformational change through a sequence of steps, possibly triggered by interactions between DNA and the C-terminal flexible domain. These studies showcase the sophisticated reconfigurability of the portal and the sequence of conformational changes required to bring the phage to maturity.

Despite this body of work, many details of basic phage biology remain unknown, particularly regarding how DNA interacts with portal proteins and its ability to modulate conformational changes. For example, many of the cited reconstructions ($\phi 29$, P22, $\varepsilon 15$, T7, and P-SSP7) locate DNA in or near the portal seemingly "poised" for ejection [129]. A phage N4 reconstruction appears to show no density in its tail [19]. Interestingly, Tang et al. [129], in a high resolution (7.8Å) of $\phi 29$, has exposed a previously undetected feature: an intriguing toroidal structure lodged in the cavity formed between the connector (gp10) and lower collar (gp11) [129] (see Fig. 1.2). While a previous study [154] reports that the terminal protein gp3 occupies the cavity, Tang et al. [129] report densities consistent with DNA inside the connector-lower collar cavity and gp3 instead resides in the center of the tail channel below the cavity. The DNA density reveals a toroid believed to consist of 30-40 bp

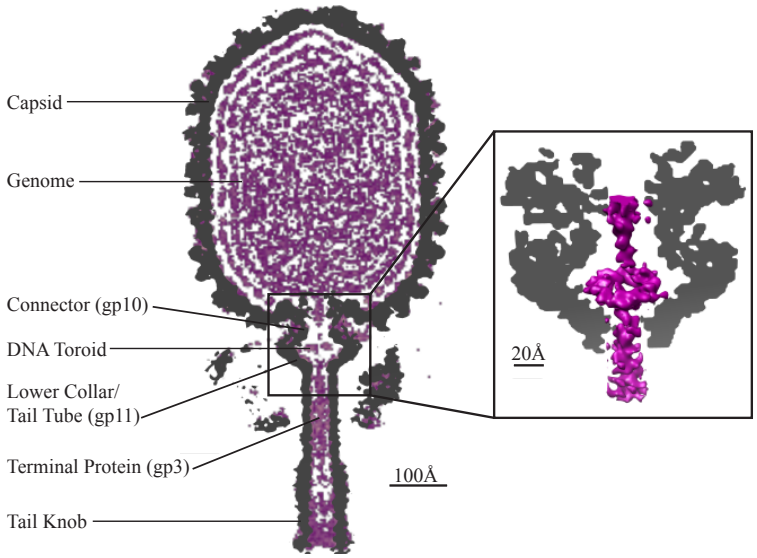


Figure 1.2 Cross section of the 3-dimensional cryo-EM $\phi 29$ reconstruction [129]. The virion map (purple) is overlaid with the emptied map to highlight the genome. Inset shows the region to be simulated; the toroid constrained in the connector/lower collar cavity. Density maps generated with UCSF Chimera [107].

of highly bent dsDNA surrounding a central void [129]. Although the mechanism needed to form such a highly bent DNA toroid remains unclear, we hypothesize that after DNA packing is completed and the tail is assembled, the connector undergoes a conformational change which opens the channel and allows the capsid pressure to drive DNA out of the capsid. The capsid pressure, in turn, fills the tail with DNA and forms the toroid. Prior to the connector opening, the terminal protein gp3 may likely be constrained inside the cavity/tail. Our hypothesis may explain how gp3, with its smallest dimension larger than the diameter of the tail channel [68, 154], is compressed and driven to the distal end of the tail. Indeed, the connector has been observed to undergo conformational changes between prohead and maturity stages [129, 154]. Thus, the $\phi 29$ cavity not only permits dsDNA to pass through but it is also large enough to allow DNA to bulge outward in forming the observed toroidal supercoil.

Should $\phi 29$ compress DNA so tightly that it forms a 30-40 bp toroid of DNA, it would present one of the most extreme case of DNA bending known to occur in nature. It would also present a unique model system for understanding DNA under extreme compression/bending, a stress state of DNA that has yet to be examined. DNA is known to exhibit higher flexibility on shorter length scales than is predicted by a worm-like chain with persistence-length elasticity properties [149] as demonstrated recently using cyclization assays [95, 136]. Close

examination of the experimental density map of [129] reveals a 20 Å thick toroid with outer and inner radii of 29 Å and 9 Å, respectively. This sub-persistence length toroid shares the same extreme curvature as a 30-40 bp dsDNA minicircle. By contrast, experiments [26] and atomistic simulations [91] of larger unconstrained 65 bp minicircles reveal a host of non-canonical DNA structures including bubbles, kinks, and wrinkles that form under high bending and torsional stress. Integration host factor (IHF) has proven capable of wrapping 35 bp of DNA around itself to form a 'U-turn' bend angle greater than 160° enabled by two kinked regions [113]. Thus, the ϕ29 protein cavity may similarly preserve dsDNA despite the sharp bending.

1.2.2 DNA Buckling in a Family of Phages

Given the convincing evidence in [129] that DNA is buckled into a toroid in ϕ29, it is natural to hypothesize that buckling could be a useful mechanism exploited by other bacteriophages as well. Closer examination of the cryo-EM data shows that indeed several other phages including T7 [2] and P22 [128] have cavities larger than ϕ29's in their portal assemblies. Contrary to ϕ29, the cavities are formed in the prohead stage prior to packaging. Their existence is particularly puzzling given that, during translocation under high force, a portal motor should seek to maximize DNA contact with its surface. We briefly list biologically relevant conditions present in a family of phages that favor DNA buckling:

1. Cavities in other phages, e.g. T7 (50 Å tall × 100 Å wide) and P22 (60 Å tall × 75 Å wide) are taller and wider than ϕ29 (55 Å tall × 67 Å wide).
2. Cavities are located downstream of the packing motor and therefore subject DNA to large axial compressive forces with no protein tube to support it.
3. Phage T7 is packed to a higher density than ϕ29 and theoretically predicted to require larger packing forces [109].
4. Electrostatic isosurfaces derived from crystal structure data of ϕ29 and P22 show that the entry and exit of the cavities carries a negative charge (presumably to center the DNA in the portal during translocation) while the central part carries a positive charge (which attracts rather than repels DNA) [24].
5. The ϕ29 motor (and possibly others) rotates DNA during packaging [156]. The amount of rotation per basepair increases markedly ($\sim$10°/bp in ϕ29) near the end of packaging.
6. Phages which lack large cavities still have regions in the portal/tail that widen to $\sim$35-45 Å or over twice the width of dsDNA. These regions could still accommodate

a slender DNA buckle.

Considered collectively, these factors create conditions ripe for buckling DNA over short length scales (on the order of the DNA persistence length) and warrant a wider investigation. Bear in mind that a buckling event could also lead to further structural damage to the molecule. In P-SSP7, for example, a well-resolved feature is a dsDNA strand separation that occurs near the center of the mature phage portal [83] analogous to 'bird-caging' in the cable mechanics literature. What function this highly compressed (and possibly twisted) DNA serves remains a mystery, although it was postulated that flipped basepairs interact specifically with the portal [83]. It is known that motors can impose considerable strain on the substrate DNA. Recent bulk experiments on T4 placed FRET markers on DNA and observed that motor force can locally compress B-form DNA by as much as 22-24% [112]. Clearly, we need to obtain a better understanding of how a DNA molecule responds to compression-induced strains.

1.2.3 Models for Constrained DNA Buckling

Relatively little (if any) theoretical attention has been given to modeling short lengths of DNA under compression in phage portals/tails. While the models cited above have generated some understanding of the global packaging process, they all focus on the capsid DNA and overlook the short DNA in the portal.

Constrained rod buckling problems remain of considerable interest to theorists and in the context of applications well outside the field of molecular biology. Many equilibrium models have described buckling constrained, for example, between rigid parallel plates [25, 57, 58], in a potential field [87], and on [138, 139] and in [133, 148] cylinders. However, these models are limited to static problems and relatively simple surfaces. Herein we employ an elasto-dynamic rod model to evaluate the buckling and dynamic ejection of short 50-300 bp DNA lengths constrained in complex viral cavities.

1.3 Research Objective

Computational models for DNA provide needed insight into biological systems that fundamentally depend on the bending and twisting mechanics of DNA. The overall objective of this dissertation is to advance this knowledge in the context of two important systems. The first goal of this research is to determine the role of LacI protein flexibility by reinterpreting two seminal studies conducted over 25 years ago that explored the stability and topology of

looping straight and minicircle DNA. The second goal is to investigate the role of DNA mechanics during packaging and ejection in bacteriophages. Specifically, we model a purported highly-bent, buckled DNA toroid recently found in the bacteriophage $\phi 29$ portal cavity and further predict DNA buckling in a family of phages. To this end, we advance the elastic rod model for DNA to account for contact between buckled DNA and the portal protein. In addition, we develop a multi-scale rod-MD model for DNA to investigate the sharp bending in the toroid. The multi-scale model represents a collaboration with the Andricioaei lab at the University of California - Irvine.

1.4 Scope of Research

We contribute novel models and analyses for two systems whose biological functions fundamentally rely on the bending and torsional mechanics of DNA including: 1) DNA looping by the Lac Repressor protein, and 2) the compression and ejection of highly-bent DNA inside bacteriophage cavities. These studies are organized into four chapters as summarized below.

1.4.1 Chapter 2: Reinvestigating Classic Lac Repressor-DNA Looping Experiments

In this chapter, we reinterpret the classic *in vitro* studies [73, 74] conducted by the Müller-Hill lab as reviewed above. We do so by employing a computational rod model for the DNA-LacI complex that considers both the 'V' conformation of the LacI cocrystal structure and an extended conformation. Our early work confirms results of Swigon et al. [127] for linear DNA which we then extend by adding a thorough analysis of loop stability and topology of the supercoiled DNA minicircles [73].

1.4.2 Chapter 3: A Model for Highly Strained DNA Compressed Inside a Protein Cavity

In this chapter, we develop the first model that explores short-length (sub-persistence length) DNA under compression and within a cavity. Specifically, the elastic rod model is extended to include mechanical contact with a protein cavity. The model is used to evaluate the sensitivity of solutions to computational parameters including boundary conditions and cavity discretization. We then exercise the model to simulate the packaging and ejection of

DNA from a small volume. This new model serves as the basis for the studies in Chapter 4 and Chapter 5.

1.4.3 Chapter 4: The Structure and Dynamics of the $\phi 29$ DNA Toroid

Although the biological function of the sub-persistence length $\phi 29$ toroidal DNA supercoil remains unknown [12], the mechanical energy stored within this distinctive structure suggests several possible biological functions. In this study, we address two major issues. The first is whether dsDNA within the $\phi 29$ cavity can form a 30-40 bp toroid (as proposed by [129]) under mechanical compression alone. That is, can the virus provide adequate forces to compress DNA into the toroid proposed by Tang et al. [129]? If such a toroid forms, does the DNA remain intact? The second issue addresses the possible biological functions of this highly energetic DNA structure. Hypotheses we explore include the possible role of the toroid in 1) signaling genome release, 2) stabilizing the pressurized virion, and 3) initiating host infection. To address these issues, we develop a multi-scale model for DNA that couples our coarse-grain elastic rod model and an all-atom MD model developed by the Andricioaei lab at the University of California - Irvine.

1.4.4 Chapter 5: The Existence and Function of DNA Buckling in the Portal System of a Family of Bacteriophages

Motivated by an example of DNA buckling in one type of virus ($\phi 29$), we explore the possibility of this same phenomenon in other phages. Specifically, we investigate the mechanics of DNA buckling during packaging in bacteriophage T7 and P22 owing to their large cavities and the abundance of cryo-EM data for these phages. We also consider the effects of motor rotation on the computed structures and speculate on the biological function of a DNA buckle in these systems.

Chapter 2

Reinvestigating Classic Lac Repressor-DNA Looping Experiments

In this chapter, we exercise a computational rod model to explore the effects of interoperator DNA length on looping behavior and reinterpret two classic *in vitro* studies [73, 74] conducted by the Müller-Hill lab as reviewed in Chapter 1. We do so by employing a computational rod model for the LacI-DNA complex that initially considers the crystallographic 'V' structure of the LacI cocrystal structure. We then in introduce an "extended" conformation which largely resolves discrepancies with experimental results. First, we begin by reviewing the elasto-dynamic rod model which is common to all chapters. Significant passages in this chapter are drawn from [54].

2.1 Review of Elastic Rod Model

We review the fundamentals of the rod model to enable the reader to understand the contribution made in this and the following chapters and direct the interested reader to [41, 44, 46, 55, 82] for a complete development of our modeling approach. We now draw significantly from [55] which summarizes details from [44, 81].

The rod model is fundamentally a coarse-grain approximation for DNA which sacrifices atomic detail in favor of describing the long length/time scales necessary to capture the three-dimensional bending and twisting of the DNA helical axis. It is a numerical extension of (3D) elastica theory and captures geometrically large deformations/rotation. The rod model incorporates elastic properties that have been determined by experiment, see for example, [10, 119, 126] and by molecular dynamics (MD) simulation [9, 125, 147]. The resolution of the rod model approach is approximately 1 helical turn of the DNA double helix, about 3.5 *nm*.

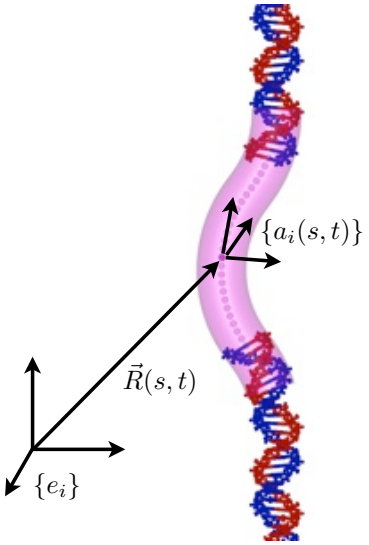


Figure 2.1 The atomic structure of DNA as modeled by an elastic rod with equivalent averaged stiffness properties. $\vec{R}(s,t)$ tracks the position of the helical axis as a function of contour length s and time t with respect to the inertial frame $\{e_i\}$. We also define a body-fixed reference frame $\{a_i(s,t)\}$ which is also a function of s and t.

Figure 2.1 depicts the atomic structure of DNA superimposed with an elastic rod possessing equivalent averaged elastic properties. The helical axis of the molecule is parameterized by the centerline curve, $\vec{R}(s,t)$, and a body-fixed cross sectional reference frame $\{a_i(s,t)\}$ where the independent variables, s and t, denote contour length (as measured from one rod boundary) and time, respectively. The rod is also described by the curvature/twist vector $\vec{\kappa}(s,t)$ which is defined as the spatial rotation rate of the body fixed frame [44].

The dynamical rod formulation begins with the Newton-Euler equations of motion for an infinitesimal rod element. Deformations to the molecule result in a net internal force $\vec{f}(s,t)$ and internal moment $\vec{q}(s,t)$. The dynamics of the rod are described by the translational velocity $\vec{v}(s,t)$ and angular velocity $\vec{\omega}(s,t)$ of the cross section. The kinetic and kinematic quantities are related through the following four vector equations of rod theory [44]:

$$\frac{\partial \vec{f}}{\partial s} + \vec{\kappa} \times \vec{f} = m\left(\frac{\partial \vec{v}}{\partial t} + \vec{\omega} \times \vec{v}\right) - \vec{F}_{body}, \tag{2.1}$$

$$\frac{\partial \vec{q}}{\partial s} + \vec{\kappa} \times \vec{q} = \mathbf{I}\frac{\partial \vec{\omega}}{\partial t} + \vec{\omega} \times \mathbf{I}\vec{\omega} + \vec{f} \times \hat{t} - \vec{Q}_{body}, \tag{2.2}$$

$$\frac{\partial \vec{v}}{\partial s} + \vec{\kappa} \times \vec{v} = \vec{\omega} \times \hat{t}, \tag{2.3}$$

$$\frac{\partial \vec{\omega}}{\partial s} + \vec{\kappa} \times \vec{\omega} = \frac{\partial \vec{\kappa}}{\partial t}. \tag{2.4}$$

The partial differential equations are expressed in the body fixed frame $\{a_i(s,t)\}$. Equations (2.1) and (2.2) represent the balance laws for linear and angular momentum, respectively. Equation (2.3) represents a constraint on the (assumed) inextensibility and unshearability of the rod. Equation (2.4) is a constraint on the curvature/twist and angular velocity vectors that ensures continuity of the cross section orientation. In these equations, $m(s)$ denotes the mass per unit contour length, $\mathbf{I}(s)$ denotes the tensor of principal mass moments of inertia per unit contour length, $\vec{F}_{body}(s,t)$ and $\vec{Q}_{body}(s,t)$ denote any externally distributed body forces or moments per unit contour length, respectively, and $\hat{t}(s,t)$ denotes the helical axis unit tangent vector. In this chapter we consider DNA in the absence of external forces and moments, such as those arising from thermal motion, and therefore neglect $\vec{F}_{body}(s,t)$ and $\vec{Q}_{body}(s,t)$. In Chapter 3, we will develop terms that describe protein surface contact and hydrodynamic drag that will be incorporated into the governing equations through these terms.

The internal moment $\vec{q}(s,t)$ results from the linear elastic constitutive law which relates the bending and torsional stiffness in a diagonal tensor, $\mathbf{B}$, to the local curvature $\vec{\kappa}(s,t)$. The constitutive law is

$$\vec{q}(s,t) = \mathbf{B}(s)\left(\vec{\kappa}(s,t) - \vec{\kappa}_o(s)\right). \tag{2.5}$$

The inter-atomic interactions of a DNA molecule combine to yield an averaged long length scale stiffness based on the persistence length of DNA. DNA is frequently assumed to be linearly elastic as adopted in, for example [6, 88]. The diagonal tensor $\mathbf{B}$ contains the commonly accepted values of the bending and torsional persistence lengths of 50 *nm* and 75 *nm*, respectively; see [50, 119]. The shape of an unstressed DNA molecule is not generally straight but depends on the basepair sequence. This 'intrinsic curvature' is represented by $\vec{\kappa}_o(s)$ and subsequent deformations from this unstressed state result in internal forces and moments. Although sequence-dependent intrinsic curvature and sequence-dependent stiffness have been incorporated in previous work [43, 45, 46, 82], these parameters are not well-defined in the systems considered herein and may have only modest influence on predicted behaviors. Therefore, we model DNA as an intrinsically straight, homogeneous isotropic elastic rod which results in the following elastic energy functional

$$E(t) = \int_0^L \frac{1}{2}[\{\vec{\kappa}(s,t)\}^T \mathbf{B}\{\vec{\kappa}(s,t)\}]ds, \tag{2.6}$$

where $E(t)$ is the elastic strain energy in units of kT.

The partial differential field equations (2.1)-(2.4) are numerically integrated upon defining boundary conditions which correspond to kinematic constraints (e.g. by LacI binding [46, 82] as described below). The equations are discretized using finite differencing and integrated using the generalized-alpha method in space and time [20, 44]. In Chapters 2 and 5 the dynamic rod formulation is used as a numerical means to arrive at the final equilibrium DNA conformation (i.e. when $\vec{v} = \vec{\omega} = 0$) whereas in Chapters 3 and 4 we rely on the full dynamic model to simulate the dynamics of viral DNA ejection.

2.2 Methods and Model Simplifications

The above formulation is completed upon specifying boundary conditions that describe the binding of DNA to LacI. These boundary conditions are derived by aligning the DNA operator sites, representing the ends of the elastic rod model, with the operator-bound DNA in the LacI crystal structure; refer to PDB ID: 1LBG [78] in Fig. 1.1. Following [46], we align rod-fixed reference frames [101] at the rod ends to corresponding basepair-fixed reference frames of the bound operator DNA (and at the site of the third basepair into the operators). Therefore, the protein crystal structure determines the position and orientation of the rod at its terminal ends. The computation begins from an initial condition with the DNA unbound and in a stress-free (straight) conformation and, following dynamic relaxation, concludes with equilibrated DNA bound to LacI and satisfying the above boundary conditions. Analysis of the equilibrium conformation yields the enthalpic (elastic energy) cost of looping per (2.6) and the complete three-dimensional equilibrium geometry of the interoperator DNA.

The symmetric or ideal operators, found in the crystal structure, permit the protein to bind to the substrate DNA multiple ways yielding multiple possible binding topologies [46, 101]. To reduce the computational effort, we follow [82] and assume that LacI is symmetric about its dyadic axis, which necessitates very small position and orientation corrections to the above boundary conditions from the original (slightly asymmetric) crystal structure. These approximations lead to three possible binding topologies for the closed LacI protein, denoted **P1**, **P2**, and **A1** in Fig. 2.2(a). Several topoisomers may also exist for each binding topology, often including two lowest-energy topoisomers that do not induce self contact of the intervening DNA, one over-twisted and one under-twisted. Ignoring electrostatics (except in the neighborhood of near self-contact) remains a reasonable assumption [5] and therefore we analyze looped conformations without self contact recognizing their lower energetic cost (hence greater likelihood) of formation. Therefore, our analysis

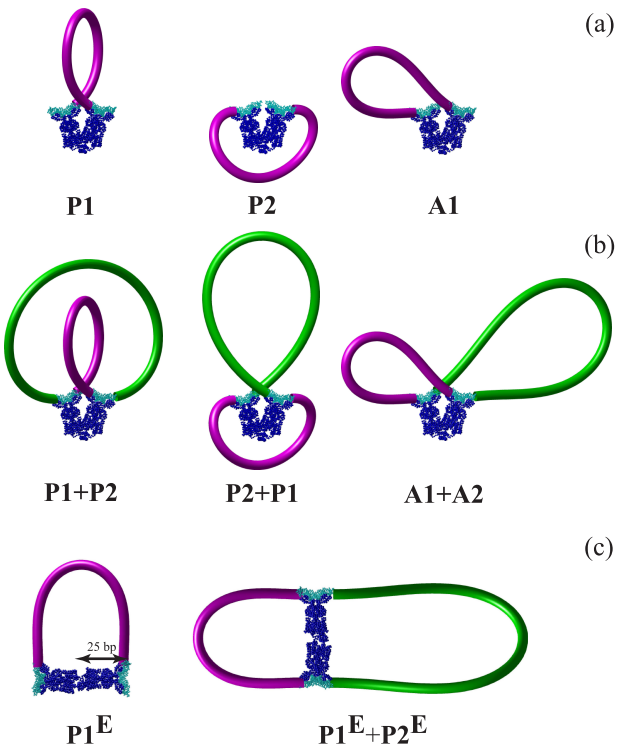


Figure 2.2 (a) Three possible binding topologies for the closed protein for looping experiments of linear DNA. (Due to the assumed rotational symmetry of the protein, the **A1** binding topology remains equivalent to the **A2** binding topology.) (b) Three possible binding topologies for the closed protein for looping experiments of supercoiled DNA minicircles. The binding topologies follow from pairing a secondary loop (green) to the primary interoperator DNA loop (purple) using the only available binding topology remaining following binding of the primary loop. (Due to protein symmetry, the primary + secondary combination is identical for the combinations **A1**+**A2** (=**A2**+**A1**)). The operator DNA is considered fixed by the protein and fills the gap between the two rod domains. (c) Approximation to the extended protein conformation (as introduced in the Discussion). Due to the assumed rotational symmetry of the protein, the $\mathbf{P1^E}$ primary loop is equivalent to the $\mathbf{P2^E}$ primary loop and the primary + secondary combination is identical for the cases $\mathbf{P1^E}$+$\mathbf{P2^E}$ (=$\mathbf{P2^E}$+$\mathbf{P1^E}$).

of a single molecule (i.e. one interoperator length) requires computation of six looped conformations distinguished by three possible binding topologies each having under-twisted and over-twisted topoisomers.

We simulated the LacI-DNA looping experiments of [74] on free (unsupercoiled) DNA having interoperator DNA lengths ranging from 153 to 168 *bp*. Adopting the convention used in [74], we define the interoperator DNA length as the length in basepairs between the centers of symmetry of the two ideal Lac operators. (In [46, 82] we defined the interoperator length as the length in basepairs from the third basepair inside the DNA operators.) We assume that all sequences have a helical pitch of 10.5 *bp* and ignore any influence of the

unlooped "tail domains" outside the interoperator DNA.

We also simulate the looping experiments of [73] on supercoiled 452 *bp* DNA minicircles. LacI binding in this instance generates two looped domains which are modeled independently. One of the DNA loops, termed the 'primary' loop, ranges in length from 153 to 168 *bp* whereas the 'secondary' loop forms the remainder of the DNA minicircle. For instance, the 153 and 168 *bp* primary loops generate companion 299 and 284 *bp* secondary loops, respectively. Primary loops formed with the **P1**, **P2**, and **A1** binding topologies necessarily form secondary loops with the **P2**, **P1**, and **A2** binding topologies, respectively; see Fig. 2.2(b). Additionally, each loop composing the minicircle possesses over- and under-twisted topoisomers. Therefore, computation of twelve conformations (three binding topologies $\times$ two loops $\times$ two topoisomers) are required for each length of the primary interoperator loop. Furthermore, the minicircles of [73] have an initial ΔLk of -2, -1, 0, and +1. Here, ΔLk is relative to the Lk of the topoisomer with the least superhelical stress, the $\Delta Lk = 0$ topoisomer. ΔLk is the sum of ΔTw (the relative amount of twist with respect to the relaxed state) and Wr (writhe as calculated following [71]). As with computations for linear DNA, we ignore (higher energy) inter- as well as intra-domain DNA self contact and protein contact. Consequently, we do not prescribe ΔLk of the minicircle but calculate it from the computed equilibrium conformation.

To further interrogate the topology of LacI-DNA minicircle complexes, Krämer et al. [73] employed topoisomerase I enzymes. Specifically, topoisomerase I was used to relax some of the superhelical stress stored in the 452 *bp* DNA minicircle-loop complexes with initial ΔLk of -2, -1, 0, and +1. Gel migration assays were used to measure the ΔLk of the 'relaxed' DNA minicircle-loop complexes. The degree of superhelical stress remaining after relaxation is an artifact of the topology of the DNA minicircle-loop complex before relaxation.

To simulate the topoisomerase I assays using the computational model, we employ several assumptions: 1) topoisomerase I enzymes relax the DNA with LacI remaining bound (i.e. binding topology remains fixed), 2) the relaxed state is independent of the dynamic relaxation process (i.e. we do not model the actual dynamic pathway leading to the final equilibrium conformation), and 3) topoisomerase I allows a complex to relax to a lower energy state, which may also increase ΔLk. Consequently, we consider all possible loops (**P1**+**P2**, **P2**+**P1**, **A1**+**A2**) with fixed ΔLk as a population of 'reactants.' To facilitate direct comparison with experimental results, we present our data in the form of computationally-predicted

gel band distributions by employing the relation

$$P_i^x = \frac{e^{-E_i/kT}}{\sum_{j=1}^{N^x} e^{-E_j/kT}}. \tag{2.7}$$

Here N^x is the number of loop conformations possible in a given ensemble of reactants or products (which we denote by replacing x with either r or p, respectively), E_i is the computed elastic energy of the i^{th} loop conformation, and P_i^x is the proportion of molecules forming a loop with the i^{th} conformation. To calculate the band grayscale intensity, we first calculate the proportion, P_i^r, of each of the reactants (loop conformations) using (2.7) where the ensemble is composed of N^r conformations possessing a given ΔLk. The ensemble of products for an individual reactant is composed of all N^p conformations with the same binding topology as the reactant. The product $P_i^r P_j^p$ for each reactant-product pair constitutes the proportion of loops that relax to form the j^{th} conformation of the i^{th} reactant. Next we sort the products by ΔLk, considering all reactants. Finally, to calculate the grayscale intensity for a prescribed ΔLk, we sum the proportions of all the reactant-product pairs that result in the given ΔLk.

The overall free energy budget for looping includes contributions from: 1) the elastic energy of deforming the DNA into the looped conformation, 2) the entropic energy cost, 3) the electrostatic energy due to possible DNA self-interaction, 4) the free energy of binding to the operator sites, and 5) the free energy of deforming the protein into the looped state. In this study, we assume that the binding free energy and entropic cost (though non-zero) remain essentially constant from one loop to the next. Further, we assume that the free energy change of LacI upon looping is constant. The electrostatic contribution, which is expected to remain modest and relatively constant (except possibly when very close contact is approached [5]), is neglected. Thus, among these contributions to the free energy, it is the elastic energy (2.6) reported herein that most likely discriminates looped conformations.

2.3 Results

We open this section by focusing on calculations that assume the closed ('V'-shaped) protein conformation. Comparing these theoretical results to experimental findings reveals significant shortcomings. In the next section, we then introduce results that assume the open protein conformation that largely overcome these shortcomings.

Consider first predictions for the looping of linear DNA paralleling the experiments of [74]. We compute the elastic energy of LacI-DNA complexes as a function of the interopera-

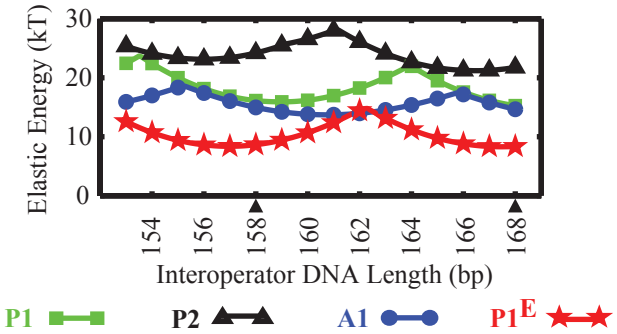


Figure 2.3 Elastic energy E (kT) as a function of interoperator DNA length (bp). The reported elastic energy is the minimum energy of the over- and under-twisted topoisomers. Shown are results for the **A1** binding topology (circles/blue), the **P1** binding topology (squares/green), and the **P2** binding topology (triangles/black). The curve with stars, (red) corresponds to the **P1**$^{\mathbf{E}}$ binding topology with $E_{LacI} = 0\ kT$ as described in the Discussion. The reported energy is the elastic energy necessary to deform DNA into a loop with a given binding topology. For the extended conformation, the energy penalty associated with opening the protein is added to the elastic energy as a means to compare binding topologies. As E_{LacI} is increased, the energy required to form a loop in the extended conformation increases and simply shifts the curve upwards. The triangles below the axis indicate locations where the most stable loops form in the experiments. These results reproduce those of Swigon et al. [127] obtained using a different modeling approach.

tor length over the experimental range of 153-168 bp. Figure 2.3 presents the elastic energy required to form a loop for each of the three possible binding topologies which reproduces known results from Swigon et al. [127] obtained using a different modeling approach. We report the minimum energy of the over- and under-twisted topoisomers. The elastic energy for each binding topology is periodic with an interoperator length of approximately the DNA helical repeat. Note that the **A1** binding topology (circles/blue) is energetically preferred for all interoperator lengths and has an energetic minimum at 161 bp. The next highest energy binding topology is the **P1** (squares/green) with a minimum at 159 bp followed by the highest energy **P2** binding topology (triangles/black) with minima at 156 and 167 bp. In the original gel data, sharp, highly-resolved bands occur at 158 and 168 bp and indicate that stable loops form at those locations.

Consider next computational results that parallel the experimental LacI-induced looping of DNA minicircles [73]. Specifically, we consider supercoiled DNA minicircles that retain the previous interoperator lengths of 153-168 bp.

For each of the four minicircle topoisomers ($\Delta Lk = \{-2, -1, 0, +1\}$), we report in Fig. 2.4 the change in energy (ΔE) as a function of interoperator DNA length. The computed ΔE represents the difference in elastic energies of the minimum energy LacI-looped complex

and the minimum energy of the unbound minicircle possessing the given ΔLk. Although the entropic contribution may play a role, its contribution should be similar in the LacI-DNA complex and in the unbound minicircle because they are the same overall size. For each ΔLk topoisomer, we report the change in energy for the minimum energy complex which may adopt different binding topologies as the interoperator length changes. Thus, the dashed lines indicate ΔE for all looped complexes among all three possible (closed protein) binding topologies in Fig. 2.2(b). The apparent discontinuities arise due to minimum energy complexes possessing different binding topologies. Minimum energy looped complexes possessing the **P1**+**P2** binding topology are denoted by the dashed green lines while those possessing the **A1**+**A2** binding topologies are denoted by the dashed blue lines. Looped complexes possessing the **P2**+**P1** binding topology are never energetically preferred due to the high energetic cost of the primary **P2** loop in the **P2**+**P1** complex.

The energy of the unbound minicircle template serves as an energy 'datum' for each topoisomer. Computation of the energy datum values follow a separate calculation in which a straight isotropic rod is bent into a circle and then twisted to achieve the requisite ΔLk. It therefore follows that the ΔLk = +1 and ΔLk = -1 topoisomers require equal energy (16.5 kT) to form while the ΔLk = 0 topoisomer requires only the energy (6.6 kT) to bend the rod into a circle. Computation of the ΔLk = -2 topoisomer (29.4 kT) requires two turns of twist which causes the circle to deform out of plane into a figure-8 shape. This conformation induces self contact at the junction of the figure-8. For this special instance, we employ a previous form of our model that explicitly accounts for the electrostatic self-repulsion of DNA and which prevents the 'cut through' in the figure-8; see [80]. However, the change in energy from this datum only dictates the offset on the y-axes between each topoisomer in Fig. 2.4 and does not affect any other results in this chapter. Finally, for the ΔLk = -2 topoisomer, self-contact exists for several interoperator lengths illustrated by the shaded regions. While self-contact is neglected, equilibria in these regions require substantially higher formation energy due to added electrostatic energy.

To further probe loop topology, we investigate the topoisomerase I relaxation experiments of [73]. The computed relaxation products for the topoisomerase I relaxation experiments of [73] are illustrated as a distribution of bands in the predicted gel of Fig. 2.5($_1$) as described in Methods. Shown are the relaxation products as a function of interoperator length for each of the four topoisomers. The predicted band distributions do not estimate migration velocity and are constructed to mimic the appearance of experimental gel data.

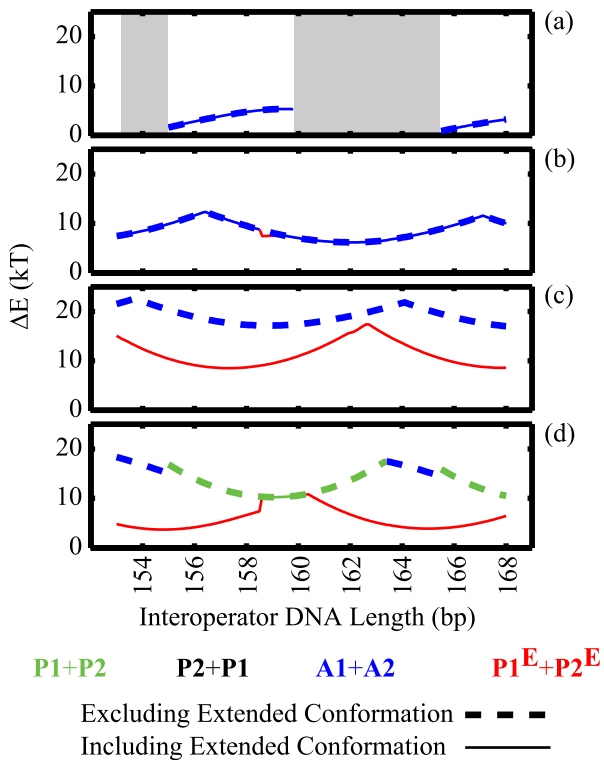


Figure 2.4 ΔE (kT) of the minimum energy loop as a function of interoperator DNA length (bp) for (a) ΔLk = -2 , (b) ΔLk = -1, (c) ΔLk = 0, and (d) ΔLk = +1 topoisomers. **P1+P2** = green, **P2+P1** = black, **A1+A2** = blue, $\mathbf{P1^E}$+$\mathbf{P2^E}$ = red. The shaded regions in the ΔLk = -2 topoisomer indicate regions where self contact occurs. The dashed lines represent minimum energy solutions considering only the closed-protein binding topologies while the solid lines represent minimum energy solutions considering both the closed- and extended-protein binding topologies with $E_{LacI} = 0\ kT$. The extended LacI conformation is described in the Discussion. The **P2+P1** binding topology is never energetically preferred due to the high energetic cost of forming the primary **P2** loop in the **P2+P1** complex.

2.4 Discussion

We open by comparing the predicted results for looping of linear DNA with the experimental results of [74]. While the binding topologies are not observable from the experiments, the rod model predicts the preferred binding topology/topologies. We assume that the minimum energy conformation considering all three possible binding topologies in Fig. 2.2(a) is the complex most likely formed in the experiments. From Fig. 2.3, the **A1** binding topology is energetically preferred irrespective of interoperator length. Moreover, the **A1** binding topology has an energetic minimum, and therefore most stable loops, at 161 bp. In the experiments however, the most stable loops, indicated by sharp bands in electrophoretic gel, occur at 158 and 168 bp. In our interpretation, we believe the theory should predict

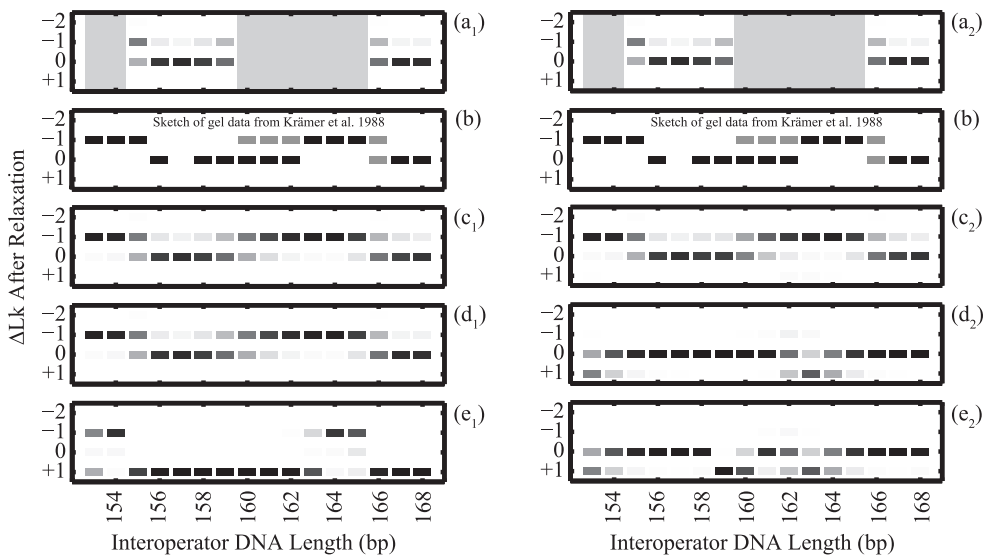


Figure 2.5 ΔLk after relaxation as a function of interoperator DNA length (bp) for (a) ΔLk = -2, (c) ΔLk = -1, (d) ΔLk = 0, and (e) ΔLk = +1 topoisomers. Subscript 1: Predicted band distributions assuming a closed protein. Subscript 2: Predicted band distributions now including the extended conformation $\mathbf{P1^E}$ with $E_{LacI} = 0\ kT$ as described in the Discussion. The band intensity is weighted on a grayscale based on the energy between the lowest energy solution and possible higher energy solutions (2.7). The shaded regions in the ΔLk = -2 topoisomer indicate regions where self contact occurs. (b) A sketch of the original gel data for the ΔLk = -1 topoisomer from Krämer et al. 1988 [73]. Highly-resolved/sharp bands are shown in black while faint bands are shown in gray.

the optimal interoperator lengths to within about a basepair. A single basepair difference may arise due to differences in our assumed helical repeat (10.5 bp) versus the actual helical repeat. By this criterion, the theory does not predict the location of the most stable loops.

Similarly, we compare the most stable loops for supercoiled minicircles [73] to the energetic minima from computations. The results of Fig. 2.4 clearly demonstrate that the degree of supercoiling strongly influences the location of the energetic minima (and therefore the interoperator length yielding the most stable looped complexes); see dashed lines.

A major conclusion from the experiments is that supercoiling changes the interoperator length requirement for stable loop formation. Our computations agree by predicting stable loops at 159 bp for the ΔLk = 0 topoisomer (notably, near the 158 bp interoperator length for linear DNA). If negative supercoiling is added, the location of the most stable loop increases by 4 bp. The model captures this trend and predicts stable loops at 162 bp for the ΔLk = -1 topoisomer, in accord with Swigon et al. [127]. Experimentally, the location of the most stable loop for the ΔLk = -2 topoisomer lies at 163 bp. We report two minima near 155 and 166 bp (at the edge of the shaded region indicating contact) and note that our computed minima, albeit 3 bp away from the experimental minimum, likely remain

even if the self contact (incurring higher energetic cost) were incorporated in the model. Regardless, the shaded regions in Fig. 2.4(a) indicate that self contact likely plays a role for the ΔLk = -2 topoisomer. Conversely, adding positive supercoiling to the substrate DNA was experimentally shown to decrease the length of the interoperator DNA for the most stable loops by 3-4 bp. However, our computations for the ΔLk = +1 topoisomer predict an interoperator length of 159 bp, which is 5 bp away from the experimental value.

As with linear DNA and supercoiled minicircles, the computations for the topoisomerase I assays yield mixed agreement with the experimental results. In agreement with a major conclusion from [73], we predict that up to one superhelical turn can be constrained by LacI looping (i.e. we predict relaxation products having non-zero ΔLk in Fig. 2.5($_1$)). Gel data at every interoperator length is available for the ΔLk = -1 topoisomer (as sketched in Fig. 2.5(b)) and reveals single bands centered about 158 and 168 bp (indicating the ΔLk = 0 topoisomer) and centered about 153 and 163 bp (indicating the ΔLk = -1 topoisomer). However, between these regions double bands exist; see Fig. 2.5(b). The computations correctly predict the ΔLk = 0 topoisomer at 158 and 168 bp, the ΔLk = -1 topoisomer at 153 and 163 bp, and the step-wise transition with multiple products (gray bands) in between these limits; see, for example, the region near 160 bp. While we are able to compare with gel data for the ΔLk = -1 topoisomer at every interoperator length, original gel data for each topoisomer is available only for interoperator lengths of 158 and 163 bp. Results at these interoperator lengths are summarized in Table 2.1 and show disagreement with the ΔLk = +1 topoisomer. Complete gel data at all interoperator lengths for the ΔLk = -2, 0 and +1 topoisomers would, of course, enable a comprehensive comparison.

	Relaxation Products			
ΔLk	158 Expt.	158 Comp.	163 Expt.	163 Comp.
-2	0	0 (**0**)	-1	N/A
-1	0	0 (**0**)	-1	-1 ($-$**1**)
0	0	0 (**0**)	0/-1	-1 ($+$**1**)
+1	0	+1 (**0**)	0/-1	+1 ($+$**1**)

Table 2.1 Relaxation products for all topoisomers at 158 and 163 bp showing experimental and computed values. A forward slash separating values indicates multiple populations present in the experiments. N/A indicates that the model does not predict solutions at this location. Results including the extended conformation are shown in bold in parenthesis.

Despite the successful prediction for the ΔLk = -1 topoisomer relaxation, the predictions for the linear DNA and supercoiled minicircles do not capture the major experimental findings. We propose that including LacI protein flexibility will resolve these disagreements for

both linear DNA and the $\Delta Lk = +1$ minicircle. Allowing an extended protein conformation will likely yield lower-energy loops thus strongly affecting our results.

Following [127, 135, 158], we assume that the protein can rotate about the tetramerization domain and adopt an extended conformation in which the angle formed between the two dimers is approximately 180°. Specifically, we assume that the protein is bistable and rigid in both the closed and extended conformations. As estimated by Swigon et al., the change from the closed to extended conformation could involve a free energy penalty of 1.8 to 9.4 kT (though likely closer to 1.8 kT) [127]. Let the extended conformation be denoted by $\mathbf{P1^E}$ with free energy penalty E_{LacI}.

We now return to the linear DNA experiments and augment the boundary conditions previously associated with the closed protein conformation with those defined by the extended conformation; see Fig. 2.2(c). In addition, we include an estimate of the cost of protein opening (E_{LacI}) in arriving at the energetic cost of looping for the $\mathbf{P1^E}$ binding topology. We treat E_{LacI} as an unknown parameter and vary it over an estimated range. In Fig. 2.3, the curve with stars (red) represents the new results for the limiting case $E_{LacI} = 0\ kT$. Increasing E_{LacI} simply shifts this curve (the energetic cost of looping) upwards relative to the previous results for the closed protein conformation. For low values of E_{LacI} (under about 4 kT), the $\mathbf{P1^E}$ binding topology is energetically preferred over all previously considered binding topologies and for nearly all interoperator lengths. Moreover, its energetic minima occur at 157 and 168 bp which agree well with the most stable loops (resolved/tight bands) at 158 and 168 bp observable in the gel data [74]. By contrast, recall that the **A1** binding topology is energetically favorable for the closed protein conformation but achieves a minimum at 161 bp which is significantly different from the experimental observations. The facts that the $\mathbf{P1^E}$ binding topology is 1) energetically preferred, and 2) possesses the correct minima for optimal loop stability provide compelling evidence that LacI adopts an extended conformation in these early experiments.

The looped complexes in the experiments exhibited noticeably different gel migration behavior [74] and this provides an opportunity for comparison with the model. In Appendix A, we compute the radius of gyration of the LacI-(linear)DNA complex (Fig. A.1) as a means of comparing size to migration velocity which suggests that the $\mathbf{P1^E}$ and not **A1** binding topology is present.

Return now to the predicted looping of the DNA minicircles upon adding the extended protein conformation; refer to the red lines in Fig. 2.4 representing new results for the $\mathbf{P1^E}$+$\mathbf{P2^E}$ binding topology for the limiting case $E_{LacI} = 0\ kT$. The solid lines arise from considering solutions from all closed protein binding topologies and the extended conformation in arriving at the overall energetic minimum. Including the extended LacI conformation

yields new energetically favorable loops at specific interoperator lengths for the ΔLk = +1, 0, and -1 topoisomers. More importantly, for the ΔLk = +1 topoisomer, the energetic minima now occur at 155 and 165 bp when including $\mathbf{P1^E}$+$\mathbf{P2^E}$ in full agreement with experimental observations. The minima for the cases ΔLk = -2 and -1 remain largely unchanged whereas there is a modest change (from 159 to 157 bp) for the case ΔLk = 0 which is also more closely aligned with experiment. The overall results, summarized in Table 2.2, clearly suggest that calculations that incorporate the extended protein conformation yield superior agreement with experimental observations. The energetic minima reported in Tab. 2.2 for $E_{LacI} = 0\ kT$ remain unchanged over the entire range $E_{LacI} \leq 6\ kT$. (As noted above, increasing E_{LacI} simply increases the energetic cost of looping and only extreme values $E_{LacI} > 13\ kT$ render the $\mathbf{P1^E}$+$\mathbf{P2^E}$ energetically unfavorable relative to the solutions for the closed LacI conformation for all interoperator lengths.

ΔLk	-2	-1	0	+1
Experiment	163	161/162	157/158	154,165
Comp. excluding $\mathbf{P1^E}$+$\mathbf{P2^E}$	155,166	162	159	159
Comp. including $\mathbf{P1^E}$+$\mathbf{P2^E}$	155,166	162	157,168	155,165

Table 2.2 Interoperator DNA length (bp) for most stable loops excluding and including the extended LacI conformation. A forward slash separating values indicates two neighboring stable complexes found in the experiments.

Interestingly, the computed results for the ΔLk = -2 and -1 topoisomers exhibit little to no sensitivity to changes in E_{LacI} in contrast with the ΔLk = 0 and +1 topoisomers which exhibit significant sensitivity. Thus, it appears that each topoisomer favors a specific binding topology. In particular, the ΔLk = -2 and -1 topoisomers largely prefer the **A1**+**A2** binding topology of the closed protein. Conversely, the ΔLk = 0 and +1 topoisomers largely prefer the $\mathbf{P1^E}$+$\mathbf{P2^E}$ topology (with some dependence on E_{LacI}).

A major contribution of the above analysis is the new light it sheds on the energetically preferred binding topology for the looping of both linear and supercoiled (minicircle) DNA. In the original work [73], it appears that changes in supercoiling were thought to be partitioned mainly into Tw (and not Wr) which, when absorbed by the minicircle, alters the helical repeat of the intervening DNA. Our data dispels the notion that the addition of basepairs to the intervening DNA sequence is simply a small perturbation of the same structure (Fig. 2.3, Fig. 2.4). By contrast, the computations strongly suggest that the loops analyzed in the electrophoretic gels likely represent a variety of binding topologies that are dependent on the length of the intervening DNA. The insertion of a single basepair alters the phasing

to such a degree that the loop may bind in a completely different orientation. Moreover, in the interactions between LacI and DNA minicircles, the length of the intervening DNA not only shifts the locations of the most stable loops but dictates which binding topology the complex prefers.

In light of the above, we now also revisit the topoisomerase I relaxation experiments. Topoisomers in the extended conformation are now added in the thermal ensemble (see (2.7)) with those of the closed protein in arriving at computationally-predicted band distributions. Figure 2.5($_2$) presents the relaxation products as a function of interoperator length and Tab. 2.1 (in bold) summarizes the relaxation products at 158 and 163 bp upon including the extended conformation.

The computed relaxation products for the ΔLk = -2 and -1 topoisomers show little to no sensitivity to the inclusion of the extended conformation in contrast to the significant sensitivity exhibited by the ΔLk = 0 and +1 topoisomers; compare windows in Fig. 2.5($_1$)/($_2$). These trends are consistent with the above discussion of minicircle looping. In the original experiments, data at every interoperator length exists only for the ΔLk = -1 topoisomer, which we show has little sensitivity to including the extended conformation. At 158 bp, the model now agrees with the experiments by predicting that all topoisomers relax to the ΔLk = 0 topoisomer. At 163 bp, however, the model predicts the ΔLk = 0 and +1 topoisomers mainly remain at ΔLk = +1 which is still inconsistent with experimental results; see Tab. 2.1. However, the sensitivity to the inclusion of the extended conformation in the ΔLk = 0 and +1 topoisomers provides a clear means by which to determine if the extended conformation is present. Should similar experiments be carried out at every interoperator length for the ΔLk = 0 and + 1 topoisomers, the model predictions could be directly tested to ascertain whether or not the extended conformation plays a role.

2.5 Conclusions

This study extends a computational rod model for DNA [46, 82] to reinterpret classical experiments on LacI-induced DNA looping in linear DNA [74] and supercoiled DNA minicircles [73]. The model quantitatively predicts the structural features of DNA-LacI interactions including distinguishing the possible binding topologies. Computed equilibrium conformations of the DNA-LacI complex replicate the experimental observations and provide detailed structural insights that extend well beyond what can be interpreted from the gel data alone. Upon comparing theory to experiment, substantial evidence reveals that LacI adopts an extended conformation and that this extended conformation was unknowingly present in the

Müller-Hill experiments conducted over 25 years ago.

The two experimental studies explore the effects of operator phasing on loop stability by varying the length of interoperator DNA from 153 to 168 bp. The computed optimal interoperator lengths for maximum loop stability align strikingly well with the experimental evidence. In addition, the calculations for linear DNA reveal that the most stable loops for linear DNA occur when LacI adopts the extended conformation. For supercoiled DNA minicircles, the calculations again predict the location of the most stable complexes for the ΔLk = -2, -1, 0, and +1 topoisomers. Upon accounting for relaxation by topoisomerase I, the calculations confirm that one superhelical turn can be absorbed by LacI looping.

Chapter 3

A Model for Highly Strained DNA Compressed Inside a Protein Cavity

In this chapter, we extend the rod model reviewed in Chapter 2 to account for mechanical contact with a protein cavity surface. Doing so then enables the studies in Chapters 4 and 5 where this extended model predicts DNA buckling and toroid formation under high compression in ϕ29, T7, and P22. In particular, the extended model in this chapter will enable us to simulate toroid formation within a protein cavity and the subsequent dynamic collapse upon genome release. For these purposes, we approximate the dimensions of the ϕ29 connector-lower collar cavity and determine the sensitivity to model parameters (e.g. contact forces and cavity discretization) later employed in Chapter 4. Significant passages in this chapter are drawn from [55].

3.1 Methods

In the following sections we introduce a model for protein-DNA mechanical contact and a model for the large viscous drag forces which dominate the dynamics of toroid ejection.

3.1.1 Incorporating Cavity Surface Contact

To incorporate mechanical contact between DNA and a protein cavity, the surface is discretized and interacts point-wise with our discretized rod model. The cavity dimensions are estimated from [129] and are shown in Fig. 3.1. Although we have approximated the cavity geometry, it is likely that the major dimensions of overall width and height of the cavity most influence the shape of the final DNA conformation. In this approximation, the cavity

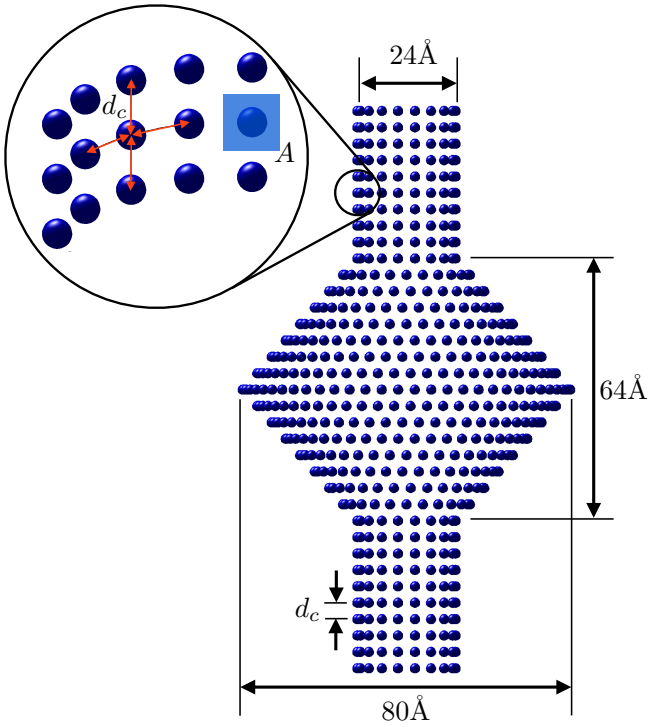


Figure 3.1 Side view of the 3-D cavity structure estimated from the connector and lower collar geometry in the cryo-EM images [129] with relevant dimensions labeled. We assume the cavity is symmetric about the vertical axis. Cavity grid points are spaced in stacked rings of points separated by d_c above and below one another and d_c along the circumference of each ring (see red arrows). Here, we have set d_c to 4 Å to reduce the number of grid points by an order of magnitude (over 1 Å separation) and thereby gain computational efficiency. The shaded blue box indicates the approximate area $A = {d_c}^2$ surrounding each grid point.

geometry has top-to-bottom symmetry (as well as symmetry about the vertical axis). This approximation is later relaxed in Chapter 4 where we model the significant details of $\phi 29$.

While a molecule of DNA carries negative charges distributed over the sugar-phosphate backbone of the double helix, it interacts with a protein surface with, in the case of $\phi 29$, unknown charge magnitude and distribution. Although it has been proposed that the negatively charged connector channel interior is responsible for centering DNA in the portal during packaging [122], the charge on the lower collar/tail remains unknown. This information is required in order to fully model the physics of DNA interacting with this specific protein cavity. Although this information is missing, a reasonable model for these interactions should at least describe mechanical contact between DNA and the cavity. Under ideal mechanical contact, DNA would experience no repulsion until the distance between DNA and the surface reaches zero, at which point it would immediately experience large repulsive

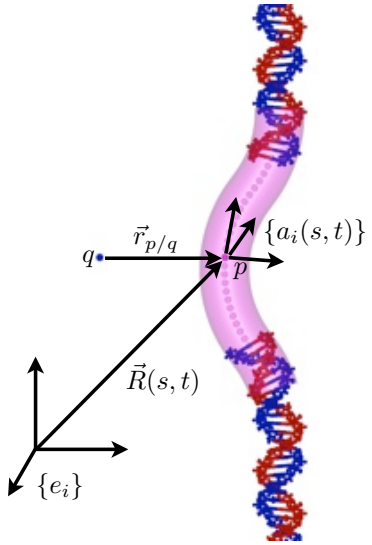


Figure 3.2 The interaction forces are dependent on all pairwise vectors between rod grid points p and all points q representing the cavity surface which are fixed in space.

contact forces. Therefore, we introduce an interaction force $\vec{F}_{p,q}$ between a DNA grid point p and a cavity grid point q (see Fig. 3.2) which approximates mechanical contact with an inverse power law as given by

$$\vec{F}_{p,q} = \frac{\rho_{\text{DNA}}\, l_o\, \rho_{\text{cavity}}\, A\, C}{(r_{p,q} - r_o)^{\beta}}\, \frac{\vec{r}_{p,q}}{r_{p,q}}. \tag{3.1}$$

In this inverse-power law, the distance between a pair of rod (p) and cavity (q) grid points is given by $r_{p,q}$ (Å), ρ_{DNA} is the 'charge' density of DNA, l_o is the length of DNA between adjacent point charges, and ρ_{cavity} is the 'charge' density per unit area surrounding a cavity grid point. For convenience in our derivation as in [80], point charges are assigned to the spatial grid points used to discretize the elastic rod such that l_o is equal to the discretization length. For simulations in this chapter, we approximate 90 bp of DNA with 144 grid points, yielding an l_o of 2.125 Å. For simplicity, ρ_{DNA} and ρ_{cavity} are set equal to one to indicate one charge per rod grid point and cavity point area, respectively. We choose r_o = 10 Å to account for the physical radius of a DNA molecule which extends from the centerline axis of the DNA described by the rod model. The order of the inverse-power law is given by β and essentially tunes the 'contact stiffness' of the cavity walls. For numerical convenience, we set $\beta = 6$. A is the approximate surface area surrounding each point on the cavity surface; see Fig. 3.1 (inset). The cavity grid points should be spaced closely enough to prevent the rod from penetrating (escaping) the cavity. Initially, we spaced the cavity grid points

$d_c \approx 1$ Å apart (see Fig. 3.1) which yielded a large number of pairwise interactions and significantly increased computational time. To increase computational efficiency, we set $d_c \approx 4$ Å (and $A = {d_c}^2 = 16$ Å^2) which reduces the number of cavity points by an order of magnitude. Ideally, solutions should be identical regardless of the number of points used to discretize the cavity and so A accounts for a coarser mesh by effectively 'amplifiying' the remaining grid points. C is a tunable constant set equal to 0.1 with units of $pN{\cdot}$Å^6 for dimensional consistency. These parameters combine to yield the approximation for mechanical contact. As the rod approaches the surface, forces remain essentially zero until in the neighborhood of the singularity (at $r_{p,q} = r_o$) at which point the repulsive forces increase rapidly to prevent DNA from penetrating the cavity.

The net cavity contact force acting on a rod grid point p, summed over all pairwise interactions with M cavity grid points, is given by

$$\vec{F}_{p,contact} = \sum_{q=1}^{M} \frac{\rho_{\text{DNA}}\, l_o\, \rho_{\text{cavity}}\, A\, C}{(r_{p,q} - r_o)^{\beta}} \frac{\vec{r}_{p,q}}{r_{p,q}}. \tag{3.2}$$

Although the potential is an inverse-power law chosen largely for numerical convenience, it can be tuned by altering C, r_o, or β (or even replaced entirely with a different potential such as a Lennard-Jones potential). We address the robustness of our solutions to changes in each of these parameters in the Results section.

The force due to cavity contact is inserted into the discretized equations of motion for the rod as an additional body force ($\vec{F}_{body}$) acting on each segment of the rod in (2.1). $\vec{F}_{p,contact}$ is a distributed force over a rod segment and therefore must be divided by the discretization length l_o when adding its contribution to $\vec{F}_{body}$. Given this rather general strategy for incorporating DNA-protein interactions, the model can be tailored to include alternative and potentially more accurate potentials should they become known. For instance, one may insert potentials obtained from approximating high-resolution molecular dynamic solutions.

Next, we discuss the numerical implementation of the contact forces in the rod model. As discussed in Chapter 2, (2.1)-(2.4) are discretized to obtain a system of equations at each timestep [44] that must be solved using Newton-Raphson iterations for the field variables ($\vec{v}$, $\vec{\omega}$, $\vec{\kappa}$, $\vec{f}$) at the next timestep. To do so, the system of equations is linearized and a Jacobian matrix is constructed. The Jacobian contains partial derivatives of each of the field variables with respect to the others. However, forces defined in (3.2) contain $\vec{r}_{p,q}$, which depends indirectly on $\vec{v}$ and $\vec{\omega}$. The steps necessary to include these terms and construct a proper Jacobian are discussed in detail in [80]. However, in these simulations, a full Jacobian has proven unnecessary. As pointed out in [80], the Jacobian need only point in the direction of

steepest descent. Computational efficiency is gained here by forming a reduced Jacobian at the expense of more Newton-Raphson iterations at each time step.

3.1.2 Incorporating Viscous Drag

In addition to modeling the equilibrium compression of DNA within a protein cavity, we are also interested in simulating the subsequent dynamic collapse/ejection of the highly-energetic toroid. DNA is likely to experience a myriad of interactions that hinder its translation through the tail tube and these interactions are largely unknown. Here, we approximate these interactions by introducing a large viscous drag force that likely dominates the dynamics of DNA (also implemented in [81, 161]) during ejection. Viscous drag coefficients are obtained following [59] for the linear and angular velocity components of a straight rod. To approximate drag per unit length acting on DNA, we first calculate the total drag force acting on a straight rod (with identical contour length as DNA) in uniform flow and compute the drag per unit length. This drag force is assumed to be evenly distributed along the length of the straight rod. We then assume that the drag coefficients per unit length of the straight rod are equivalent to the drag coefficients per unit length of DNA. Following these approximations, we arrive at drag coefficients per unit length for velocities perpendicular ($C_\perp$), parallel ($C_\parallel$, which acts along $\hat{t}$), and rotations about the helical axis (C_{axial}) given by:

$$C_\parallel = \frac{2\pi\mu}{\ln\left(\frac{L}{2r_o}\right) - 0.2}, \tag{3.3}$$

$$C_\perp = \frac{4\pi\mu}{\ln\left(\frac{L}{2r_o}\right) + 0.84}, \tag{3.4}$$

$$C_{axial} = \pi\mu(2r_o)^2. \tag{3.5}$$

Here the contour length of the molecule L = 306 Å (90 bp) and $\mu = 1.3\times10^{-3}$ $kg/(m{\cdot}s)$ is viscosity of water. The dissipation from these forces/moments enters into the governing equations as another contribution to $\vec{F}_{body}$ (along with the cavity forces) in (2.1) and $\vec{Q}_{body}$ in (2.2). These quantities are expressed in terms of the local $\{a_i\}$ fame in Fig. 3.2 as follows:

$$\vec{F}_{drag} = -\begin{bmatrix} C_\perp & 0 & 0 \\ 0 & C_\perp & 0 \\ 0 & 0 & C_\parallel \end{bmatrix} \vec{v} \tag{3.6}$$

$$\vec{Q}_{drag} = -\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & C_{axial} \end{bmatrix} \vec{\omega}. \tag{3.7}$$

Numerical implementation of these effects is straightforward, as the drag forces are scalars that multiply the field variables $\vec{v}$ and $\vec{\omega}$. However, the computed relaxation dynamics will likely provide an upper bound of the actual ejection speed because, as advised in [59], the fluid surrounding DNA is influenced by nearby surfaces which add additional drag effects compared to free DNA in solution.

3.2 Results and Discussion

We present and discuss three sets of results describing 1) toroid formation, 2) sensitivity to cavity discretization, and 3) dynamic ejection. We also briefly speculate on the implications of these results to the biology of $\phi 29$ which motivates the detailed study in Chapter 4.

3.2.1 Simulating Toroid Formation

We present a sample simulation of toroid formation and report the force and energy required to sustain it within the cavity. The rod models a portion of the viral genome that includes the toroid plus short straight DNA segments above and below the toroid. The simulation shown in Fig. 3.3 begins with the compression of a rod which approximates a 90 bp segment of DNA within the cavity. The upper boundary condition, which is fixed in space, allows rotation of the rod cross section about the tangent vector, $\hat{t}$ (vertical axis), but prevents rotation about any in-plane axis (i.e. otherwise clamped). The lower boundary is fully clamped and the rod is buckled by prescribing the position of the lower boundary with respect to the upper boundary. The initial condition corresponds to a slightly bent and twisted rod to bias the formation of a left handed supercoil. The buckled rod is pushed into the cavity and, after dynamic relaxation, forms the final toroid (Fig. 3.3(c)). We then compute a sequence of equilibrium solutions as the bottom boundary is moved downward. The equilibria ($v = \omega = 0$) are free of twist (consequence of the torque-free upper boundary) and therefore yield a lower bound estimate of the elastic energy required to form the toroid.

We report the internal (compressive) force at the top/bottom rod grid points and the elastic energy in Fig. 3.3(b) as functions of the shortening of the rod ($\delta = L - d$) where d represents the distance between the upper and lower boundaries. At $\delta = 10$ Å, the rod

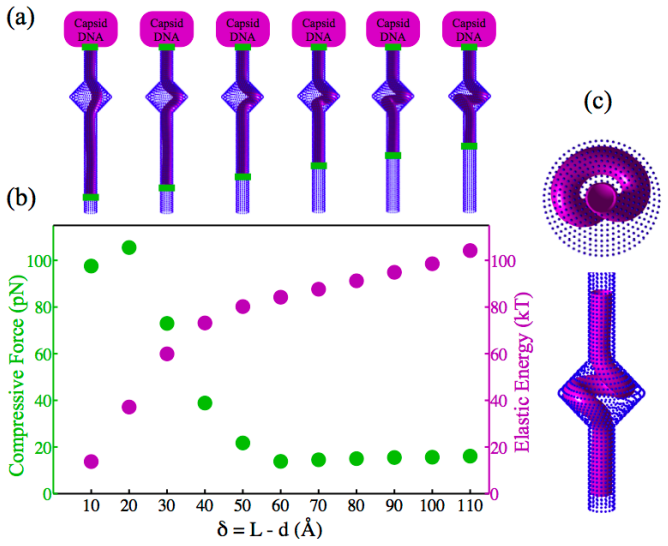


Figure 3.3 (a) Snapshots (aligned to the dots on the plot directly below) from simulating 90 *bp* of DNA compressed within the cavity. (b) Internal (compressive) force (*pN*) and elastic energy (*kT*) following (2.6) as functions of the shortening of the rod $\delta = L - d$ where L is the length of DNA (306 Å) and d is the distance between the upper and lower rod boundaries. The green box illustrates where we report internal force along the contour length. (c) Top and side view of the final toroidal structure.

is largely straight and requires a large compressive force, around 97 *pN*, to compress the stiff structure. The rod is a nearly planar (two dimensional) buckle that has not contacted the walls. At $\delta = 20$ Å, the rod contacts the cavity and the stiff planar buckle becomes even stiffer and now requires 105 *pN*. Beyond this value of δ, the rod is a 'softer' three dimensional buckle that requires considerably less force to contain it within the cavity; about 16 *pN* for the final toroid. The reported internal forces on the top and bottom grid points remain essentially equal. In these preliminary simulations, we neglect DNA-DNA self contact interactions as detailed in [80], recognizing the likely modest contribution of electrostatics compared to the high bending energy of the molecule. While the compressive forces vary significantly as δ is increased, elastic energy increases monotonically as more straight DNA becomes bent after being pushed into the cavity. The computed force/energy values remain the same upon shortening the simulated domain by 10 *bp* (5 *bp* on each end), therefore our choice of $L = 90$ *bp* is sufficiently insensitive to changes in length and boundary conditions.

Interestingly, the forces computed by the rod model are well within those measured in experiments. Early single molecule experiments used optical tweezers to pull on DNA during packing and reveal that the mechanochemical packing motor is able to generate a force of 57 *pN* [123]. This study was extended using improved experimental techniques

to reveal that the packing force exceeds 100 pN and rises sharply as the packing process nears completion [114]. This packing motor is the strongest molecular motor known to date. Again, the predicted compressive forces for formation of the toroid lie well within the range achievable by the packing motor.

The computed equilibrium shape of the toroid requires high elastic energy, over 100 kT, which is concentrated in the highly bent DNA in the cavity in the final conformation. This high elastic energy may be sufficient to denature DNA where the hydrogen bonds connecting the basepairs are broken and basepairs flip out. Modeling highly compressed and bent, sub-persistence length (and possibly denatured) DNA clearly necessitates a finer-grained modeling approach which motivates the atomistic modeling discussed in Chapter 4.

We now address the robustness of our solutions to changes in parameters in the interaction force relationship. We begin by analyzing the final equilibrium conformation, Fig. 3.3(c). By changing C by an order of magnitude (to 1.0 and 0.01 $pN{\cdot}\text{Å}^6$), the force and energy values change by less than $\sim 3\%$. The computed quantities also remain nearly identical upon changing β to 4 and 8. By changing r_o to 9 Å, we begin to see the artificially-thick DNA rod penetrating the cavity. Ultimately, we require that the DNA rod, as shown in Fig. 3.3(c), for example, remain completely inside the cavity grid points. In Chapter 4, the rod-computed equilibria are used as initial conditions to all-atom simulations. To be useful, this initial condition must not have DNA atoms that sterically interfere or overlap with the cavity.

3.2.2 Sensitivity to Cavity Discretization

We now explore the robustness of an equilibrium conformation (and computed quantities of force and energy) to changes in cavity discretization. The approximate distance between cavity gridpoints, d_c, remains an important computational parameter as a large number of pairwise interactions levy considerable computational cost. Clearly, the cavity points need to be spaced closely enough to capture the major dimensions and create a smooth circular cavity. It is also reasonable to predict that, below a certain value of d_c, solutions will become invariant to finer discretization.

To test this hypothesis, we select the final toroid (Fig. 3.3(c)) and alter d_c from 4 Å to 1, 2, and 8 Å. Following equilibration, the new rod conformations are shown in Fig. 3.4. All four conformations appear identical and possess equivalent elastic strain energy (102 kT) regardless of choice of d_c. Compressive forces remain essentially equal ($\approx$16 pN) for d_c = 1, 2, and 4 Å. However, for d_c = 8 Å, the compressive force on the bottom and top of the rod changes to -16.4 and -14.6 pN, respectively. The difference between these values (and also

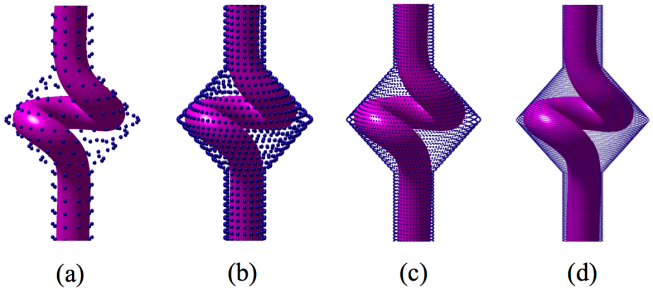


Figure 3.4 Equilibrium conformation from Fig. 3.3(c) upon altering the cavity grid spacing parameter d_c to (a) 8 Å, (b) 4 Å, (c) 2 Å, (d) 1 Å.

between those of the other cavity discretizations) could be attributed to a 'rough' cavity and too coarse a discretization. Therefore, we conclude that a d_c of 4 Å is sufficiently converged and a finer spacing is unnecessary. This spacing nets a 16 fold decrease in the number of grid points (compared to 1 Å spacing).

3.2.3 Simulating Ejection Dynamics

The ejection of DNA from viral capsids provides an exquisite example of the dynamic unbending and untwisting of the molecule from highly compacted states. While the equilibrium toroidal supercoil is an intriguing structure in its own right (and with yet-unknown function), it also serves as a highly energetic initial condition for the ejection process. To explore further, we simulate the fast (non-equilibrium) ejection dynamics beginning with the final conformation in Fig. 3.3(c) as an initial condition. To this end, the top end of the DNA remains clamped and stationary while the bottom end is permitted free rotation about the tangent vector (vertical axis) as well as free translation along the tangent. As stated previously, the model likely employs a lower bound estimate of the dissipation (friction) on the DNA in the cavity and the tail regions.

As the toroid collapses, the DNA follows the sequence of conformations illustrated in Fig. 3.5(a) with rapid dynamic transitions. In particular, an ejection time of a mere 30 *ns* is required for the toroid to propel the DNA out of the cavity. Thereafter, the molecule untwists on a slower time scale in achieving the final unstressed (straight) conformation; see Fig. 3.5. Simulating beyond 45 *ns* would show that the remaining torque continues to decay to zero.

These preliminary results also point to possible implications to the biology of $\phi 29$. We hypothesize that toroid relaxation could initiate the ejection process by signaling the connector to release its grip on the DNA or by disturbing the packed capsid DNA. In Fig. 3.5 we compute the internal force and internal torque at the top rod grid point. The compressive

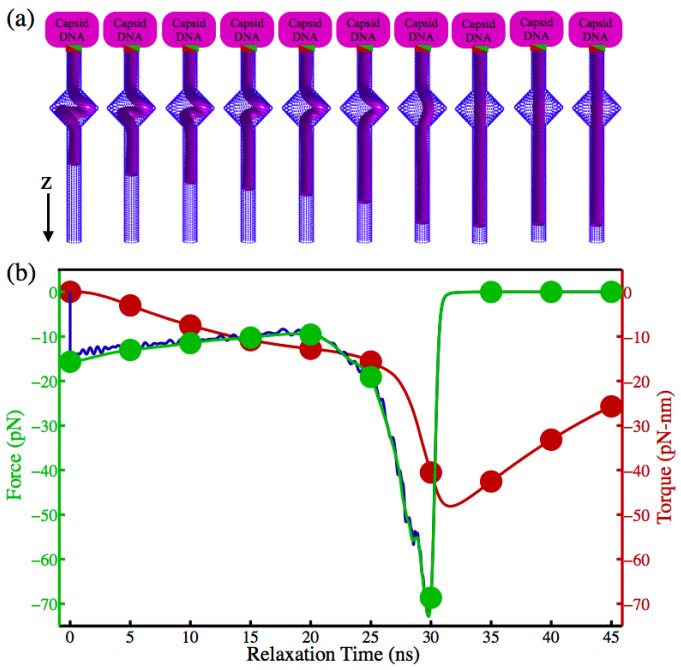


Figure 3.5 (a) Dynamic ejection and toroid collapse. Snapshots illustrate DNA conformation at each of the 5 *ns* increments denoted in the figure below. (b) DNA reaction force (solid green line and circles) and torque (solid red line and circles) on the remaining capsid/DNA (upper end) as a function of time. The solid blue line represents the component of the viscous drag force (integrated over the length of the rod) acting along the z-axis (vertical axis).

force acting on the DNA begins at around 16 pN and, as the buckle collapses, rises sharply to around 73 pN. Thus, the rapid dynamics generates over a four-fold increase in the force pushing against the connector/capsid DNA. The internal torque on the rod begins at zero and jumps to -48 $pN-nm$ as the toroid collapses. Interestingly, the negative sign on the internal torque corresponds to the left-handedness of the buckled supercoiled DNA. The torque would then be positive if the initial condition was as a right-handed supercoil. Thus, the reaction force and torque acting on the connector/capsid DNA from the toroidal DNA are quite large and may signal ejection.

The ejection dynamics are strongly influenced by viscous drag forces. To determine the effects of drag on ejection, we integrate $\vec{F}_{drag}$ over the length of the rod and plot the component of the drag force that acts along the z-axis (vertical axis) in Fig. 3.5(b) (blue line). Observe that the drag force begins at 0 pN (at equilibrium) and immediately becomes negative when ejection begins (i.e. there is a net upward force due to drag opposing the ejection of the rod). For the entirety of the ejection process, the drag closely matches the reaction force (i.e. the blue line lies behind green line) which means that the upwardly-directed drag force is almost completely responsible for the compressive force generated at the top of the

rod. Thus, as expected at this (sub-persistence) length scale, viscous drag dominates inertial effects (i.e. Brownian dynamics limit).

The dynamic simulation above captures only the very beginning of the ejection process. Modeling the entire ejection process requires specifying many uncertain features, including the role of the host cell. For $\phi 29$, the pressure within the capsid appears to drive only the first part of ejection and an unknown source in the host cytoplasm pulls the rest of the DNA into the cell [40]. However, the predicted relaxation time, though only on the order of nanoseconds, poses a significant challenge to MD simulation. Thus, the rod model is well positioned to evaluate this system.

3.3 Conclusions

We extend an elasto-dynamic rod approximation for DNA to include contact with surfaces required to simulate DNA compressed within a protein cavity. The model is used to investigate the compression and ejection of DNA within an approximated cavity in $\phi 29$. It predicts equilibrium conformations with biologically realistic forces and high elastic strain energy which may denature DNA. Our results warrant further investigation utilizing atomistic models and accurate cavity geometry as described in Chapter 4. The coarse-grain elastic rod is an efficient way to generate solutions that can function as initial conditions to subsequent all-atom (MD) simulations. Upon incorporating an approximation for hydrodynamic drag, we compute the fast dynamics and the large conformational changes of toroid collapse which occur on timescales on the order of nanoseconds. Furthermore, this study reveals appropriate values for the computational parameters (including contact forces and cavity discretization) that are employed in Chapter 4.

Chapter 4

The Structure and Dynamics of the $\phi 29$ DNA Toroid

In this chapter, we conduct a thorough analysis of the DNA toroid proposed by the Baker lab in Tang et al. [129] by implementing a multi-scale elastic rod/all-atom model for DNA together with a more accurate representation of the cavity geometry. Specifically, we use the models to determine 1) whether DNA can mechanically deform into a toroid under biologically relevant conditions, and 2) the biological function(s) the toroid could serve. This study is conducted as part of a joint collaboration with the Andricioaei lab at the University of California - Irvine where the all-atom computations were undertaken by Maryna Taranova. Significant passages in this chapter are drawn from [56] including specific contributions from the Andricioaei lab to provide a complete analysis for the reader.

4.1 Methods and Model Assumptions

4.1.1 Computing Cavity Geometry

The $\phi 29$ connector/lower collar cavity dictates the constraints for DNA buckling and eventual toroid formation. While a crystal structure of the monomer of the $\phi 29$ connector exists [122], the atomic resolution structure of the lower collar/tail tube is unfortunately unknown. Thus, we must necessarily make simplifying assumptions to approximate the cavity geometry. Here, we assume that the cavity is axisymmetric about the viral tail axis. To obtain a smooth cavity, we analyze the cryo-EM density map (see cross section in Fig. 4.1(a)) and employ averaging. Specifically, the density map is sampled at twelve azimuthal angles (each hour on a clock face). For example, the right and left side of the cavity in Fig. 4.1(a) are at

(a) (b)

Figure 4.1 (a) Cross section of the connector/lower collar cavity with toroidal DNA from [129]. (b) Pointwise-discretized constraint surface, derived from the protein cavity walls, specifies the location of computational grid points in the rod model or inert atoms in MD rigidly positioned on the cavity surface.

3 and 9 o'clock, respectively. At each azimuthal cross section, we identify the innermost points with non-zero density value. Each of these points is a known distance from the central axis and, upon collecting all such points, we construct a best-fit line through this collocation extending from top to bottom on the density map. The best fit line is used to define the constraint surface which is then point-wise discretized (see Fig. 4.1(b)). We retain the same spacing parameter $d_c = 4$ Å as determined in Chapter 3.

4.1.2 Overview of Multi-Scale Model

Simulating DNA buckling under biologically relevant forces in all-atom MD requires prohibitively large computational resources. One strategy to shorten computation time is to apply significantly higher forces on DNA than exists in the biological system. However, this approach greatly increases the likelihood of forcing the system away from the physical pathway and thereby introducing artifacts into the computed DNA structure. Alternatively, a coarse-grained rod model can efficiently simulate buckling but fails to capture the (atomistic) structure on length scales below a helical turn. Therefore, we exploit the coarse-grain equilibrium conformations predicted by the rod model as initial conditions for refinement in MD. Doing so significantly reduces computation time yet preserves biologically-relevant force levels.

The initial condition for atomistic simulation is obtained by fitting an all-atom representation of B-form DNA (using the "model.it" toolset [144]) around the DNA helical

axis predicted by the rod model. Because fitting involves placing pre-optimized base pair geometries perpendicular to the helical axis, we are left with a free parameter, referred to as the 'registry' parameter. Specifically, the DNA atoms can be 'barrel rolled' about the helical axis without changing the linking number. This registry parameter may play an important energetic role because it dictates how the major and minor grooves align at the entry and exit of the toroidal DNA, an effect not included in the rod model.

4.2 Results

4.2.1 Toroid Conformations Predicted by Rod Model

After the tail machinery (including the lower collar, tail tube, and tail knob) is installed, the connector presumably allows an amount of DNA to escape, which, driven by high capsid pressure, descends to fill the tail and cavity. The resulting equilibrium conformations of a short (75 bp) segment of DNA are computed from the rod model including DNA within the cavity and short sections in the tube above and below the cavity. In the simulations, DNA is compressed from a nearly straight conformation to a highly buckled toroidal conformation. The buckling is simulated by allowing one basepair at a time to descend from the top portion of the connector tube into the cavity until an excess of 33 bp fills the cavity, forming a buckled toroid. During this simulation, the lower boundary is held fixed against translation while the upper boundary undergoes prescribed translation and incremental equilibration. The excess DNA fills the cavity and increases the toroid size. In addition, the only rotation permitted at the lower boundary is rotation about the helical axis which relieves any additional DNA twist during buckling. No rotations are permitted at the upper boundary. The rod model retains computational parameters outlined in the previous chapter and refines the discretization length to l_o = 1.7 Å. Images of selected equilibrium conformations during buckling are shown in Fig. 4.2 and the associated terminal compressive force and elastic strain energy of the DNA are reported in Fig. 4.3. While all equilibrium conformations remain twist free, the initial condition is a nearly straight right-handed supercoil that biases the formation of a right-handed toroid thereafter. While it is unknown whether the actual toroid is right- or left-handed, Yu et al. [156] determined that the packing motor twists DNA in a left-handed sense during packaging. As DNA exits the capsid, we hypothesize that DNA will relax by twisting in the opposite direction, thereby biasing the formation of a right-handed toroid.

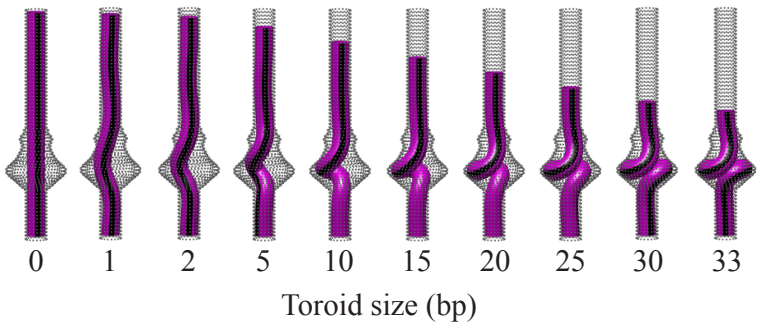


Figure 4.2 Rod model predicts toroid formation in the $\phi 29$ cavity. Snapshots of twist-free equilibrium conformations as a function of toroid size (bp). The black stripe aids in visualizing the twist state of DNA. During supercoil formation, twist is relieved through the lower boundary.

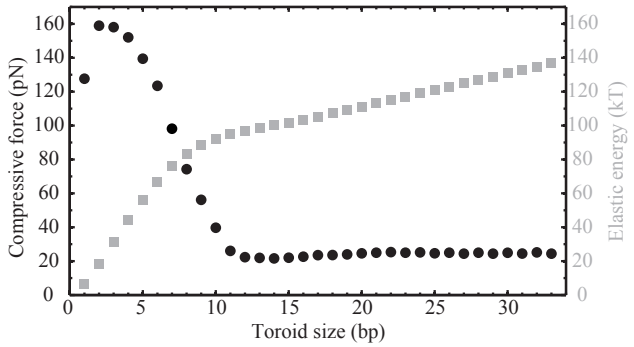


Figure 4.3 Compressive force required to form a toroid is well within packing motor capabilities (25 pN for toroids > 10 bp). Plotted are DNA internal (elastic strain) energy (gray squares) and compressive force (black circles) as a function of toroid size. DNA is initially nearly straight while the top boundary descends in the tube one basepair (3.4 Å) at a time.

4.2.2 Energy Minimization in All-Atom Force Field

We next perform all-atom energy minimizations of the DNA structures over a two parameter space defined by toroid size and registry. The result for right-handed toroids, Fig. 4.4, illustrates the total potential energy landscape and its constituent parts. Please see [56] for minimization protocol. The limiting 33 bp case represents the largest toroid contained in the cavity without nonphysical overlap of the sugar-phosphate backbone. For completeness, the energy landscape for left-handed toroids is shown in Fig. B.1 in Appendix B.

To overcome the absent atomic structures required for simulating the atomistic cavity constraints, we make the following simplifying approximations. We represent the cavity points as Lennard-Jones (L-J) atoms (with mass and van der Waals radius of a helium atom equal to 1.48 Å and fixed in Cartesian space at points shown in Fig. 4.1(b)) to enforce hard

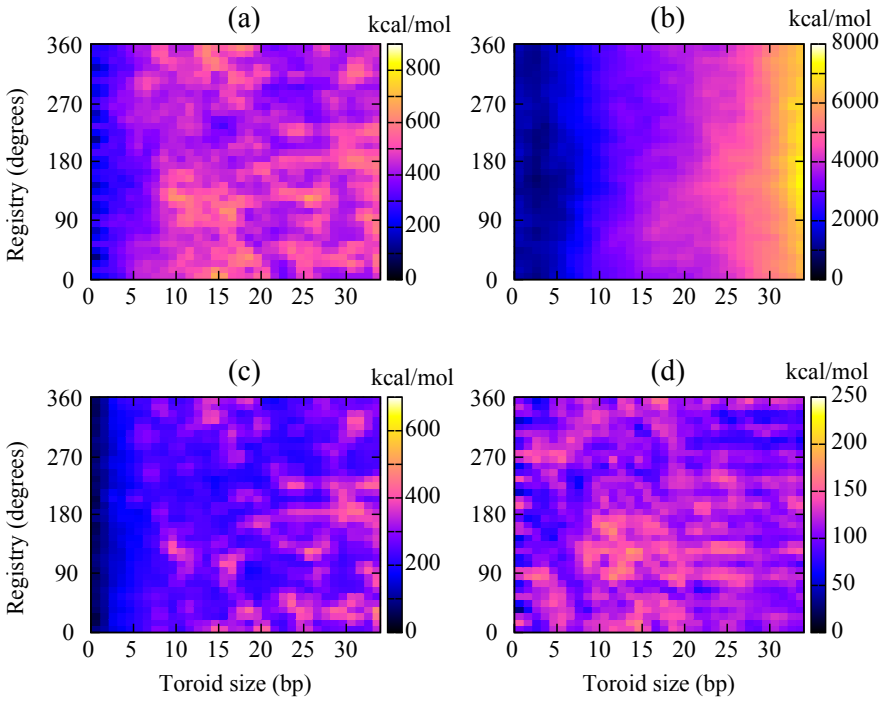


Figure 4.4 Results courtesy of M. Taranova at the Andricioaei lab. Energy landscape of all possible DNA conformations over the two parameter space of toroid size and registry. (a) Potential energy landscape and its (b) electrostatic, (c) bonded, and (d) Van der Waals components for right-handed toroids. To highlight energetic differences between conformations, energy values are scaled relative to the minimum value in the corresponding landscape.

contact between DNA and the cavity wall. This approach permits simulation of DNA within the constraints of the cavity and it can be refined to capture atomic details when additional crystal structure data is available.

A free parameter in the interaction potential with the cavity boundaries is the depth of the L-J potential well, ε, and we conduct identical MD simulations varying ε over three orders of magnitude (ε = 0.02, 0.2, 2.0, and 20 $kcal/mol$). The most stable DNA structures are observed for ε = 0.2 $kcal/mol$ and this value is adopted in all energy minimization runs (see Fig. B.2 and text in App. B for additional results and a discussion of sensitivity to ε).

The absolute energies of the all-atom force field are associated with a rapid increase in the repulsive electrostatic energy of the system as the highly charged DNA is being compressed within a confined volume. The hard (L-J) interaction potential between DNA and the cavity wall, whose rigidity is enforced in our model, confines this highly energetic DNA state and mimics the way the protein environment would confine it. While the absolute energy values are high, they are not higher than those expected in the case of a protein

viral capsid where DNA is highly pressurized to approximately 40 atm [53]. Additionally, measurements of Young's modulus using scanning force microscopy show that the protein shells of bacteriophages are extremely resilient and can withstand nanoNewton forces [63]. Moreover, the forces derived in our models are of the same order of magnitude as those reported in [123] and we therefore expect that the cavity can accommodate these force levels without significant deformation. While the formal possibility exists that the energy values associated with DNA compression inside the protein cavity can be smaller than those predicted with our approach (because of the simplifications in the potential), the model still provides important insights into the energetic dependence of various DNA toroidal shapes inside the confined volume of the exit cavity of $\phi 29$.

4.2.3 Structural Equilibration in MD

To determine the atomic equilibrium structure of the toroid, we choose a representative toroid from the above landscape and perform an equilibration in MD. The structure (33 *bp*, registry 280°) in Fig. 4.5(a, b) was chosen because of its large size and lower energy relative to neighbors on the landscape. Equilibration proceeds for 8 *ns* followed by 2 *ns* of production run (see [56] for simulation details). To enforce identical boundary conditions in MD, the terminal basepairs are harmonically restrained against translation along the tail tube but permitted to rotate about the tail axis. Figure 4.5(c, d) shows front and top views of the equilibrated structure, respectively.

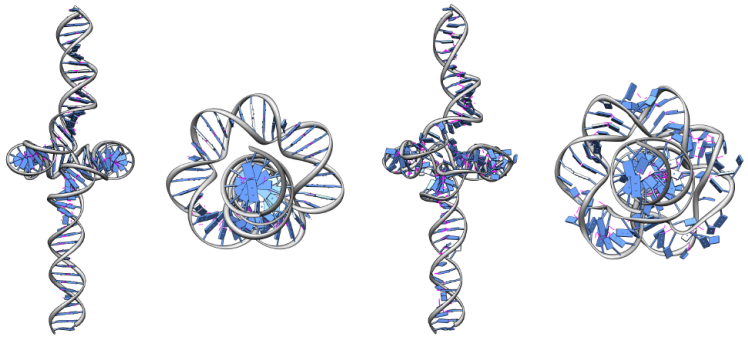


Figure 4.5 Results courtesy of M. Taranova at the Andricioaei lab. Molecular dynamics model of atomic DNA structure highly bent and under extreme compression. (a, b) Front and top views of all-atom initial condition of a 33 *bp* toroid with helical axis predicted by the rod model inside the constraints of the cavity. The DNA sequence is the first 75 basepairs from the $\phi 29$ genome. (c, d) Same views of final conformations following MD equilibration within the constraints of the cavity show that base-pairing is preserved. Hydrogen bonds shown in magenta. All atomistic DNA graphics generated with UCSF Chimera [107].

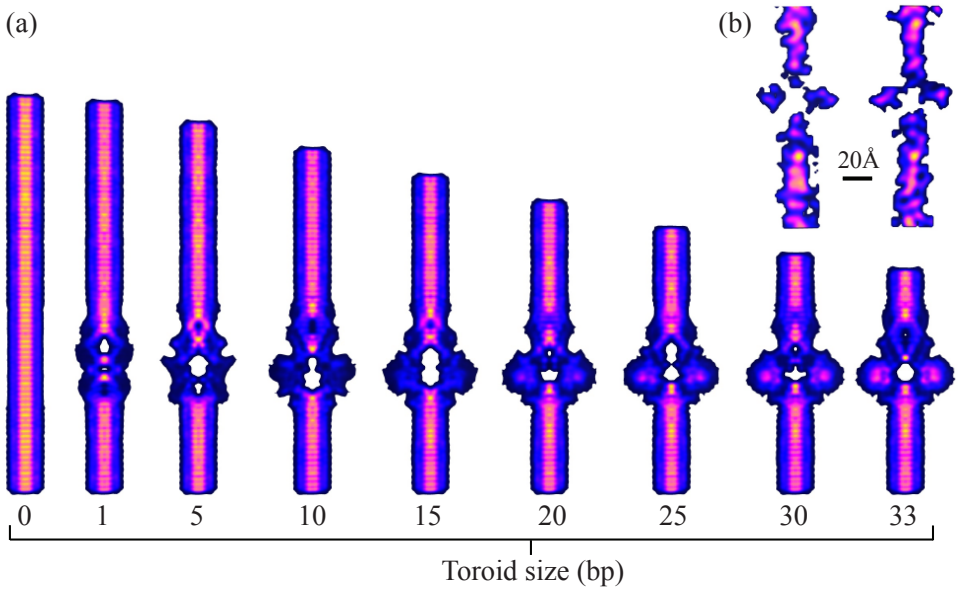


Figure 4.6 Predicted density maps compared side by side with experimental data. (a) Predicted maps for different sized toroids. (b) Two orthogonal cross sections of the experimental cryo-EM density map [129]. Yellow and dark blue indicate regions of highest and lowest intensity, respectively. Rotational averaging of the conformation was performed using UCSF Chimera [107]. Consistent with the experimental results, the color scale is determined by scaling the highest intensity bin relative to the lowest intensity bin (see App. B).

4.2.4 Comparison with Cryo-EM Density Map

The all-atom energy-minimized conformations are used to generate predicted density maps for direct comparison with the experimental cryo-EM density map (two orthogonal cross sections shown in Fig. 4.6(b)) [129]. In the experimental reconstruction, density attributed to DNA is obtained through subtraction of two maps: the virion reconstruction (EMD-1420) minus the post-ejection emptied reconstruction (EMD-1419) found in the EBI-MSD database. Recall that the capsid exhibits 5-fold symmetry while the connector/tail exhibit 12-fold symmetry. Although these components aid in aligning the exterior of the two dimensional particle images, the interior toroid may not possess the same azimuthal orientation from particle to particle. That is, if the buckled DNA is randomly oriented within the interior cavity then the resulting EM map will appear 'smeared out' upon rotational averaging of images about the tail axis. To mimic this process, we rotationally average conformations of increasing toroid size (while maintaining e.g. 280° registry) with the resulting cross sections of predicted density maps shown in Fig. 4.6(a) (see Fig. B.3 and text in App. B for additional maps with varying toroid registry and handedness).

4.2.5 Simulating Ejection Dynamics

During infection, the distal end of the tail knob opens to allow the genome to exit the virus [154]. Once opened, the highly-energetic toroid will undergo dynamic collapse and ejection from the cavity. We hypothesize that toroid collapse/ejection could play an important role in initiating infection, for example, by acting as a signaling mechanism. To address this hypothesis, we return to the rod model to simulate the ejection dynamics induced by toroid collapse and over long time scales that are practically inaccessible to MD.

The initial condition to the ejection simulation is adapted from the rod-model predicted equilibrium conformation (the 33 bp toroid in Fig. 4.2). To account for viscous drag effects, an additional 80 bp of DNA is added beneath the toroid to the opening of the tail tube. The viscous drag model for DNA [59] follows the numerical implementation of [55, 81] implemented in Chapter 3. At the opening of the tail tube (bottom), the rod remains restrained against rotation about the two transverse axes, but is now free to translate along and rotate about the tail (vertical) axis following the (assumed instantaneous) opening of the tail knob. The upper boundary remains clamped and fixed by the possible pinching of the connector and reacts to highly dynamic forces and torques induced during toroid collapse. The predicted axial reaction force and torque during this collapse are reported in Fig. 4.7 as functions of reduced time.

The reduced time scale in Fig. 4.7(b) is obtained by dividing simulation time (t) by the time required for the toroid to completely straighten/eject (τ). Thus Fig. 4.7(a) illustrates the conformation of the DNA toroid at times corresponding to 0%, 25%, 50%, 75%, and 100% of full ejection. While τ obviously depends on the selected friction model, the important findings illustrated in Fig. 4.7 do not. In particular, note the large reaction force and torque developed at the upper boundary right as the toroid straightens. These large reactions arise regardless of the assumed friction model which, for computational expediency, presently underestimates the total friction; it accounts for hydrodynamic drag and not the (unknown) friction of DNA within the tail tube of $\phi 29$. If one were to adopt the much larger (by 2-3 orders of magnitude) friction coefficients estimated from experiments on bacteriophage λ ejection [47] (0.17 to 2.4 $N{\cdot}s/m^2$), then the predicted ejection times would range from 6.8 $< \tau <$ 68 μs.

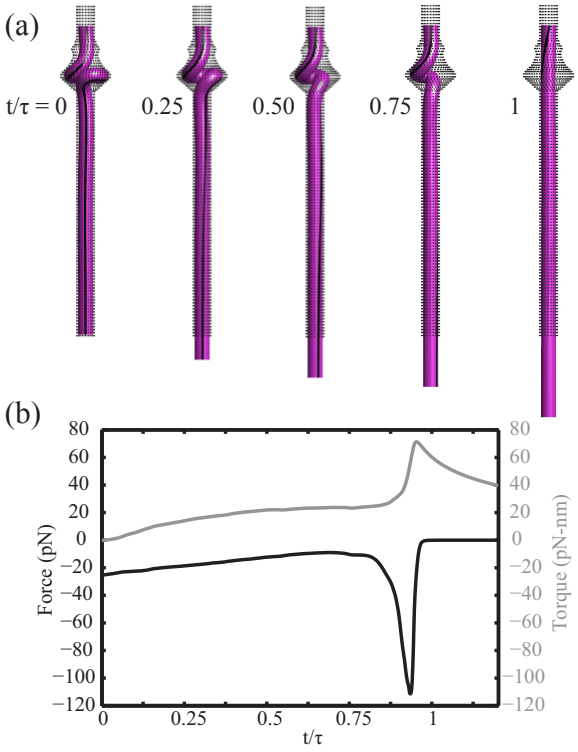


Figure 4.7 Rod model of toroid collapse predicts large reaction force/torque at the connector. (a) Snapshots show instantaneous conformations during ejection. (b) Axial reaction force and torque at the upper boundary as a function of reduced time t/τ, where τ is the time required for the toroid to fully straighten and can range from approximately $\tau = 6.8$ to 68 μs (see text for discussion).

4.3 Discussion

4.3.1 The Existence and Geometry of the Toroid

Both the rod and MD models corroborate the existence of a highly strained toroid within the small connector-lower collar cavity. Results from the rod model confirm that the capsid pressure is sufficient to mechanically compress and maintain DNA in a toroidal shape within the cavity. Inspection of Fig. 4.3 reveals that as the bucking process begins, the DNA deflects outward into a slender but stiff helical buckle which requires high compressive force (refer to Fig. 4.2). After 2 bp are inserted, DNA contacts the cavity wall, which further constrains the structure and results in maximum compression ($\sim$160 pN). As the upper boundary descends further, the DNA buckles dramatically to fill the cavity interior. This large scale buckling creates a substantially more compliant structure and significantly smaller compressive forces are therefore required for DNA equilibrium (refer to Fig. 4.3 for $bp > 10$ where the compressive force is reduced to a modest $\sim$25 pN). Therefore as

the toroid evolves, the compressive force remains nearly constant at $\sim$25 pN which, by comparison, is about 25% of the maximum force achievable by the packing motor.

During the initial buckling of DNA (Fig. 4.3), the rod model also predicts forces that exceed those achievable by the packing motor. However, these high initial forces result from the particular pathway to toroid formation chosen for this computation. The results illustrated in Fig. 4.3 follow from a near perfectly straight DNA segment that also remains torque-free, both limiting conditions that require the greatest compressive force to initiate buckling. The addition of any intrinsic DNA curvature, torque, or localized defects would significantly reduce the compressive force required to initiate buckling. For instance, the supercoiled capsid DNA would likely impart torque to the DNA as it descends through the tail tube, which in turn will reduce the compressive force required to initiate buckling [48]. Therefore, our results demonstrate that the capsid pressure provides sufficient driving force to buckle DNA into a toroid in this confined volume.

Upon refining the rod-predicted initial condition, the computed energy landscape in Fig. 4.4 illustrates energetic dependence on both toroid size and registry. The potential energy (Fig. 4.4(a)) and the elastic strain energy estimated with the rod model (Fig. 4.3) both increase with toroid size as expected. Furthermore, both models predict a rapid energy increase at the initial stage of compression (for toroids up to $\sim$10 bp). Inspection of Fig. 4.4(b) reveals that the electrostatic energy component monotonically increases as more DNA enters the cavity, an effect independent of registry as expected. By contrast, DNA bonded energy and the Van der Waals energy of the system exhibit dependence on size and registry. The variations in the potential energy (Fig. 4.4(a)), which largely follow that of the bonded and Van der Waals components, do not suggest a single favorable toroid size or registry. Instead, these results suggest that an ensemble of toroids may exist. Additionally, the energy landscape for left-handed toroids (Fig. B.1) remains energetically equivalent to the right-handed landscape suggesting no energetic preference exists regarding toroid handedness.

The predicted maps in Fig. 4.6(a) share many qualitative characteristics with the experimental map [129] and further argue for an ensemble of toroids. Note the appearance of a central void as soon as DNA buckles and the toroid intensity increases with toroid size as more DNA fills the cavity. In the straight tube sections above and below the toroid, observe extensive low (dark blue) and intermediate (light blue) intensity regions with patches of high intensity (magenta and yellow) near the center of the tail axis consistent with the experimental map. Also observe the low intensity region joining the toroid to the high intensity straight segments above and below it, as noted in Tang et al. [129]. All predicted maps for toroids larger than $\sim$20 bp similarly exhibit sharp (nearly 90°) density transitions

between straight and toroidal sections. For these larger toroids, note the dominant regions of intermediate (light blue) and high (magenta) intensities in the toroid which are consistent with the intensities reported in the experiments. Smaller toroids, by contrast, do not possess the high intensity regions observed in the experiments. Although difficult to discern exact toroid size and registry, we observe matching high (magenta) intensity in the toroid for sizes of $\sim$20 bp and larger. Thus, the experimental density map may reflect an averaged ensemble of toroids weighted toward larger ones (consistent with results in Fig. 4.4). Additionally, the shape of the void at the center of the toroid (as well as the existence and shape of the smaller voids above and below the toroid) changes with toroid size, registry, and handedness, but an ensemble average would also mask these dependencies (see Fig. B.3 in App. B for predicted density maps for varying handedness and registry).

Beyond the qualitative similarities noted above, the predicted density maps (Fig. 4.6(a)) also share remarkable quantitative similarities to the experimental map. Tang et al. [129] report a 20 Å thick toroid possessing an outer and inner diameter of 58 Å and 18 Å, respectively. Despite the small differences in the central void, the dimensions of the larger toroids remain remarkably consistent with these experimental measurements. In particular, the predicted toroid thickness and the inner diameter (distance between the high intensity regions) are $\sim$20 Å while the outer diameter is $\sim$60 Å (dictated of course by the diameter of the cavity).

Equilibration in all-atom MD simulations reveal that the initial toroid predicted by the rod model is largely preserved despite sharp bending and compression. Observe in Fig. 4.5 that the initial condition matches reasonably well with the MD-equilibrated structure and both confirm the existence of a central void. Aside from a few regions where we observe basepair flipping, the DNA remains double-stranded. The small denatured regions may also be an artifact arising from the assumed interaction potential with helium atoms which likely do not fully represent the biological system. To explore the implications of our assumptions, we have applied the model using several interaction potentials, characterized the sensitivity to this parameter, and speculate on the biological implications (see Fig. B.2 in App. B). From our systematic analysis emerges a model whereby dsDNA denaturation is limited in large part by the nonspecific interaction with the cavity.

Recall that the overall objective of this chapter is to explore the hypothesis put forth by Tang et al. [129] that DNA forms a toroidal structure. A coarse grain model, an MD model, and predicted cryo-EM density maps all provide consistent evidence for a toroid. However, for completeness we must also recognize that there are other structures that could yield the split density in the entry/exit cavity including DNA strand separation as observed in Podovirus P-SSP7 [83], flipped bases, a cruciform, etc.

4.3.2 Possible Functions of the Toroid

Recent experiments conducted after the reconstruction of [129] have increased our understanding of the sequence of conformational changes in the $\phi 29$ connector. In light of these discoveries, we explore several possible biological functions of the toroid including its ability to stabilize the mature virion and/or possibly trigger a conformational change.

1. *Toroid collapse triggers genome release.* The dynamic collapse of the toroid could function as a signalizing mechanism to initiate the release of the capsid DNA. Our simulations predict a rapid dynamic collapse of the toroid which would occur well before the remainder of the genome would be ejected from the capsid. During collapse, the compressive force on the connector increases four-fold from its value at equilibrium (see Fig. 4.7). In fact, this reaction force as well as the reaction torque achieve maxima right as the toroidal DNA straightens. It is impressive that rapid toroid collapse generates forces ($\sim$110 pN) on par with the strongest known molecular motors and without the need for additional ATP. This tandem reaction force/torque 'shock' might disturb the highly organized genome. Additionally, a recent study by Geng et al. [38] on the gating mechanism in the $\phi 29$ connector concluded that it could undergo stages of conformational change mediated by contact between DNA and the C-terminal flexible domain. If the C-terminal domain acts as the trigger to induce conformational changes, then it is possible that the large reaction forces/torques could cause DNA/trigger contact, initiating yet another conformational change. While the connector undergoes a conformational change upon DNA packing [49] to hold the pressurized genome inside the capsid, we hypothesized that it must also undergo a second conformational change to open the connector and allow DNA leakage back through the tail. Toroid collapse may further dilate the connector to allow unhindered genome escape. Additionally, it has been proposed that a triggered conformational change in bacteriophage SPP1 [108] and P22 [14] opens the connector and enables genome ejection. Finally, the large force/torque occurs right as DNA straightens and their magnitude is independent of toroid size. This size independence renders toroid collapse a robust signaling mechanism.

2. *Toroid stabilizes pressurized virion.* At first glance, the toroid resembles a single turn of a helical spring which may help stabilize the densely packed genome. Changes in environmental conditions (e.g. temperature or ionic conditions) over the life of $\phi 29$ cause fluctuations in the highly-pressurized capsid DNA. In fact, the dynamics of bacteriophage genome ejection have been probed by perturbing ionic conditions, which in turn affect capsid pressure [153]. The flexible toroid could serve as an essential source of mechanical compliance allowing the tightly packed capsid to remain stable despite environmentally-induced pressure fluctuations.

This possibility follows from the force-deflection curve of Fig. 4.3 which confirms that large change in toroid size produce little to no change in DNA compression (remains at $\sim$25 *pN* for toroids larger than 10 *bp*). By contrast, a straight column of DNA provides little compliance and potentially induces much larger stresses in the capsid protein shell in response to environmental disturbances.

3. *Toroid initiates host infection.* Upon opening the tail knob during infection, toroid collapse may rapidly push the terminal protein gp3 $\sim$30-40 *bp* into the host cell, a process requiring no aid from the capsid pressure or additional ATP. Hydrolytic enzymes (gp13) decorating the tip of the tail knob (gp9) help degrade the $\sim$250 Å host cell wall [21, 92, 155]. Thus, as the tail knob bores into the host, the toroid could apply pressure on gp3 and/or gp13 to aid in degradation or puncture of the cell membrane.

4.4 Conclusions

This chapter contributes a continuum-elastic rod and an all-atom MD model to simulate DNA buckling into a toroid and its subsequent ejection from the connector-lower collar cavity. The rod model predicts that an ensemble of toroids can mechanically form under biologically-relevant force levels. Equilibration in MD (Andricioaei lab) reveals that the dsDNA structure is largely preserved despite sharp bending through non-specific interactions with the cavity. Both the energy landscape (Andricioaei lab) and the predicted density maps support the likelihood of an ensemble of different sized toroids which combine to form the averaged experimental density map. Upon simulating the dynamic ejection of the toroid, we compute large reaction forces/torques at the connector which may signal an additional conformational change to allow the capsid DNA unhindered escape into the host cell. We hypothesize that the mechanics of DNA may not only help stabilize the phage, but also play an active role in governing genome release. Higher resolution reconstructions of other bacteriophages may reveal similar DNA structures in the phage portal/tail, a hypothesis explored next in Chapter 5.

Chapter 5

The Existence and Function of DNA Buckling in the Portal System of a Family of Bacteriophages

In this chapter we investigate the possibility and biological function(s) of DNA buckling and the mechanics of buckles in a family of bacteriophages. Specifically, we include bacteriophages P22 [76, 98, 128], T7 [2], ε15 [65], and T4 [77] by drawing from their recent cryo-EM reconstructions; see Fig. 5.1. In the cases of P22 and T7, we are privy to the structure of the pre-packaged prohead as revealed by cryo-EM (T7 [2]) or the crystallographic portal structure (P22 [98]). These structures show barrel shaped (T7) and funnel shaped (P22) cavities in which DNA might buckle under compression. As in the last chapter, we use the rod model to simulate the packaging of DNA into these cavities.

The reconstructions in Fig. 5.1, representing the highest resolutions to date, all reveal DNA near or inside the portal of the mature phage. While none (yet) report a toroid akin to ϕ29, all possess cavities large enough to allow DNA buckling as revealed by each reconstruction. The prohead of T7 has the most obvious cavities in the portal core, as indicated by three colored boxes measuring 115 Å tall $\times$ 35 Å wide (green), 50 Å tall $\times$ 110 Å wide (purple), and 100 Å tall $\times$ 35-45 Å wide (red). The mature T7 locates dsDNA inside the phage portal and shows the dramatic rearrangement that the portal proteins undergo upon filling. Specifically, the portal column collapses in height by 20 Å. The reconstruction does not specifically identify a toroid, and the resolution of 24 Å is considerably coarser than that of ϕ29 (7.8 Å). Although it is difficult to discern the boundaries between DNA and protein, the T7 tail shows regions of high intensity interrupted by breaks and regions of low intensity where a toroid might be seen if imaged at higher resolution. The portal of mature ε15 bears a strong resemblance to mature T7 and also locates dsDNA in the tail measuring $\sim$270 Å tall. Note that the massive cavity in T7 is only observed in the prohead

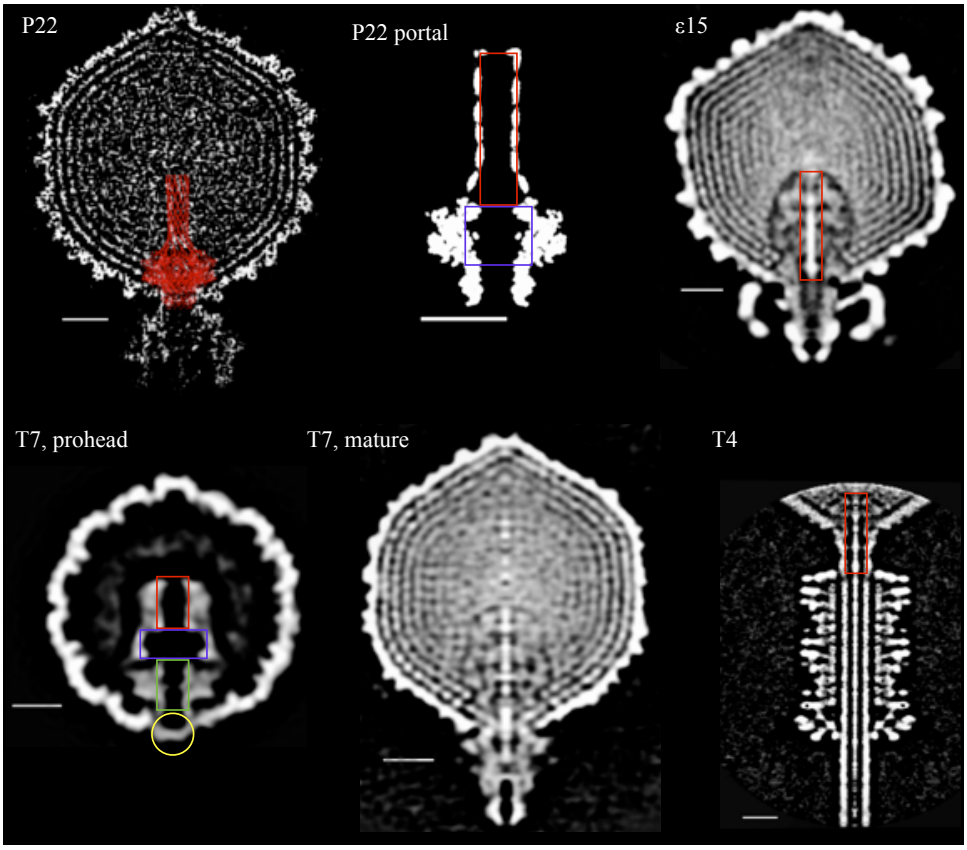


Figure 5.1 Cryo-EM reconstructions for a family of phages. White scale bars = 100 Å. **P22**: Mature phage reconstruction of [128] with portal crystal structure [98] (red) docked into reconstruction. **P22 portal**: Cryo-EM density map generated from crystal structure [98] reveals two cavities (colored boxes). **T7, prohead** [2]: Reconstruction of pre-filled head highlighting portal core with three distinct cavities (colored boxes) above the packing motor site (yellow circle). **T7, mature** [2]: Reconstruction of filled DNA head identifies DNA in the portal core. **ε15 [65]/T4 [77]**: Mature phage reconstructions where red boxes indicate dsDNA in the portal. T4 reconstruction shows the tail and the very bottom of the capsid.

reconstruction and therefore it seems possible that a reconstruction of the prohead state of ε15 might well reveal similar large cavities. The structure of the bacteriophage P22 portal has been identified by high resolution (7.8 Å) cryo-EM [128] and x-ray crystallography [98]. The lower cavity measures $\sim$60 Å tall $\times$ 75 Å wide (purple) and the barrel above measures $\sim$190 Å tall $\times$ 35 Å wide (red). While a toroid analogous to ϕ29 was not identified, it was noted that density approximately twice the width of dsDNA occupies the lower cavity. Although Tang et al. currently attributes the wide density to DNA, they could not rule out the possibility that it corresponds to protein. Finally, several bacteriophages including T4 (shown) and SPP1 (not shown), while not possessing large cavities, do include regions where the portal widens to $\sim$1.5-2$\times$ the width of dsDNA which would permit a slender

buckle to form under sufficiently large compression.

5.1 Results and Discussion

5.1.1 Predicting the Buckling Load

We begin by posing the question of what compressive load and/or twist might be required to buckle DNA within any of the cavities noted above. We address this fundamental question by employing a classical buckling analysis attributed to Greenhill [48] which, in this context of this research, predicts the onset of buckling as a function of cavity height and possible twist. The buckling load is given by the Greenhill equation

$$F_b = -\frac{\pi^2 EI}{L^2} + \frac{M^2}{4EI}, \tag{5.1}$$

where the first term represents the classic Euler buckling load for a pinned column of length L (in our case, the height of the cavities) and bending stiffness EI and the second term accounts for an applied torque developed by the packing motor; see [48] and [134] for a modern treatment. The applied torque is calculated based upon the torsional stiffness of DNA and the imposed twist of the packing motor. Figure 5.2 plots the buckling load from (5.1) over two parameters representing the cavity height and motor rotation during packaging.

The contour plot above reports a wide range of biologically attainable bucking forces. As expected, the height of the cavity is the dominant factor in determining the buckling load and end rotation modifies that load to some degree. Most notably, all forces shown reside within the capabilities of the $\phi 29$ packing motor. Initiating a buckle in the short cavities of T7, $\phi 29$, and P22 requires compressive forces of -82, -68, and -57 pN, respectively when twist is absent. For the tall, tube-shaped cavities of T7, P22, T4, and $\varepsilon 15$ the buckling load drops dramatically from -20 to -3 pN at L = 100 and 270 Å respectively. Interestingly, for cavities taller than about 90 Å, it is also possible for tension to develop under biologically relevant twist levels. That is, under significant rotation, tension is required to stave off buckling and this transition is demarcated with the red line in Fig. 5.2. A packing motor could detect the sudden transition from compression to tension and exploit such a mechanism as a signal to stop packaging. Even for phages such at T4, a tubular cavity twice the width of dsDNA could accommodate a slender helical buckle.

There are many unknown factors that could alter the above results. For example, Green-

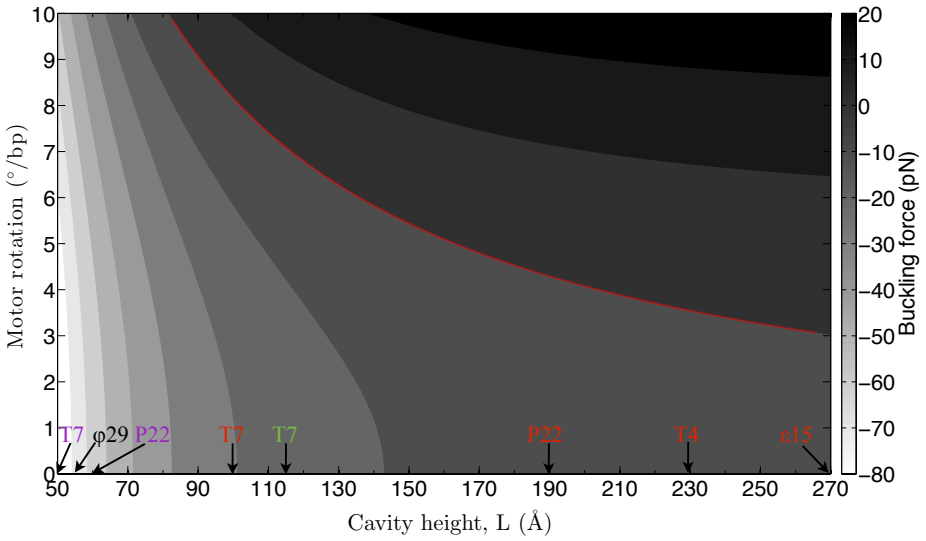


Figure 5.2 Predicted buckling force for a family of phages based on Greenhill's equation shows the buckling force as a function of cavity height (x-axis) and possible motor rotation (y-axis). The y-axis denotes lines of constant torque. The colored phage names correspond to the measured heights of the cavities (colored boxes) in Fig. 5.1. The negative sign indicates compression. The region above the red line indicates where tension is required to prevent buckling as the second term in Greenhill's equation exceeds the first. We consider up to $10°/bp$ rotations as has been demonstrated experimentally for $\phi 29$ (personal communication, C. L. Hetherington, Bustamante lab, 2013).

hill's equation neglects imperfections such as the inherent anisotropy of the chiral molecule and sequence-specific inhomogeneity, which would both reduce the buckling force. We also neglect electrostatic forces such as those between DNA and the portal or the negatively charged packaged genome which would have a restoring effect and increase the buckling load. Conversely, positively charged regions in the portal (e.g. in $\phi 29$/P22 [24]) would encourage buckling. We have also assumed pinned boundary conditions which yield a lower bound prediction of the buckling load. For completeness, we have solved the other extreme case of fixed boundary conditions in Appendix C which yield an upper bound prediction. Despite being an upper bound, the predicted buckling forces for the short cavities remain on the order of 10^2 pN. The 'real' boundary conditions reside somewhere in between the two cases.

5.1.2 Adding Effects of DNA Electrostatics and Improved DNA-Cavity Contact

In the following section we explore the mechanics of DNA buckles and their possible biological function(s) in bacteriophage T7 and P22 using the computational rod theory used in the prior chapters. In so doing, we also account for DNA electrostatics and an improved model for DNA-cavity contact.

Incorporating DNA Electrostatics

A DNA molecule carries negative charges distributed along the backbone of the double helix. As DNA is compressed into a cavity, distant sites along DNA are brought into close proximity which generates repulsive electrostatic forces. Various approximations to mechanical contact forces have been used to describe DNA self contact including the treatments in [42, 69]. Lillian et al. introduced an approximation to the true electrostatic interactions following [145]. We adopt this electrostatics model and summarize its salient features and implementation. The interested reader is directed to [80, 81] for a more extensive discussion.

The following Debye-Hückel potential describes the electrostatic interaction energies between charges along DNA.

$$E_{elec} = \sum_{p=1}^{N} \sum_{q=p+1+N_{ex}}^{N} \frac{\nu^2 l_o^2 e^{-\kappa_D r_{p,q}}}{D r_{p,q}}. \tag{5.2}$$

Here, the total energy is the sum of all pairwise interactions of N point charges at locations p and q along DNA. In (5.2), $r_{p,q}$ is the distance between point charges. An equivalent charge distribution for DNA in an ionic solution is obtained by altering the magnitude and density of the charges. As already mentioned, altering ionic conditions can affect electrostatic interactions as positive charges screen negatively charged DNA. Here, we set the charge density $\nu = 0.608\ e^-/\text{Å}$ for the assumed 0.1 M monovalent salt concentration; see [145]. The distance between point charges is $l_o = 0.34\ nm$, set here as the distance between neighboring basepairs. The Debye length $\kappa_D = 1.04\ 1/nm$ and $D = 4\pi\varepsilon$ where $\varepsilon = 80 \times 8.854 \times 10^{-12}$ F/m is the permittivity of water. We must also correct for nearest-neighbor interactions that would otherwise artificially increases the stiffness of the molecule above the bending persistence length of DNA we employ. For our simulations of DNA packed into a cavity, it is crucial that we capture self contact of tight supercoils and therefore we must also set a limit on excluding nearest neighbor interactions. To that end, we exclude $N_{ex} = 30$ nearest neighbor point charges (10.2 *nm* contour length) which is consistent with Vologodskii and Cozzarelli [145] which specifies a range between 10-20 *nm*.

The force acting on a point charge p due to an interaction with point charge q along

DNA is given by

$$\vec{F}_{p,q} = -\nabla E_{elec} = -\frac{\partial E_{elec}}{\partial \vec{r}_{p,q}}. \tag{5.3}$$

Upon substituting (5.2) into (5.3) we obtain

$$\vec{F}_{p,q} = \frac{\nu^2 l_o^2}{D} \frac{e^{-\kappa_D r_{p,q}}}{r_{p,q}^2} \left(\kappa_D + \frac{1}{r_{p,q}} \right) \vec{r}_{p,q}. \tag{5.4}$$

To account for all interactions between all point charges, we define $\vec{F}_{p,elec}$ as the net electrostatic force on point p due to all points q (excluding nearest neighbors) as

$$\vec{F}_{p,elec} = \sum_{p=1}^{N} \sum_{q=p+1+N_{ex}}^{N} \frac{\nu^2 l_o^2}{D} \frac{e^{-\kappa_D r_{p,q}}}{r_{p,q}^2} \left(\kappa_D + \frac{1}{r_{p,q}} \right) \vec{r}_{p,q}. \tag{5.5}$$

The electrostatic force $\vec{F}_{p,elec}$ is added as yet another contribution to the body force ($\vec{F}_{body}$) in (2.1) (upon converting it to a distributed force by dividing by the discretization length, l_o).

Improved DNA-Cavity Contact Model

In this section we improve upon methods from [55, 56] outlined in Chapter 3. Specifically, our algorithm is updated to efficiently simulate larger lengths of DNA and generalized to accommodate a cavity of arbitrary geometry. Figure 5.3 shows the barrel and funnel shaped cavities of T7 and P22, respectively.

The cavities are assumed symmetric about the viral tail axis and parameterized with cylindrical coordinates ρ, ϕ, and z. Rather than constructing a discretized cavity in advance, we instead employ an 'adaptive' scheme that dynamically locates the closest points on the cavity surface (q) to a rod centerline grid point (p). The number of interaction forces that must be computed scales efficiently with $\mathcal{O}(L)$. The azimuth is easily computed (i.e. $\phi_p = \phi_q$), and for square geometries such as T7 in Fig. 5.3, computing the nearest cavity point is straightforward. However, more complicated geometries composed of circular curves or tapered lines like P22 in Fig. 5.3 requires an additional step to identify the contact points. Essentially, we seek to minimize the distance between p and q and so define $d = \sqrt{\left(\rho_p - \rho_q\right)^2 + \left(z_p - z_q\right)^2}$. One need only specify $\rho(z)$ for the cavity. Substituting ρ_q as a function of z_q into d and forming $\partial d/\partial z_q = 0$ yields z_q as a function of ρ_p and z_p. Plugging z_q into $\rho(z)$ one can easily obtain ρ_q to locate the contact point. This simple distance minimization guarantees that the interaction forces are directed perpendicular to the cavity surface and thus no artificial 'friction' is introduced; see Fig. 5.3.

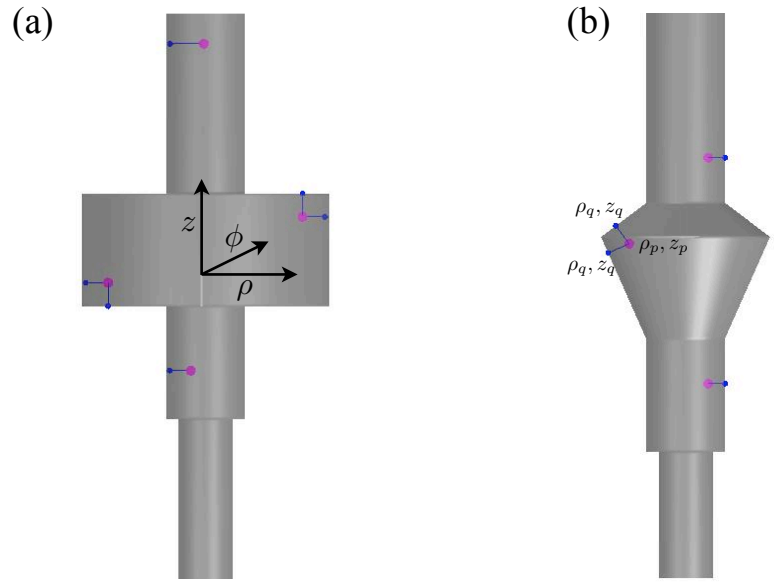


Figure 5.3 Approximated cavities for (a) T7 and (b) P22. Hypothetical rod grid points (pink) show their respective points of contact with the cavity (blue) identified by the adaptive algorithm.

5.1.3 The Mechanics of DNA Buckles

With the two improvements above, the prior computational rod model can now be employed to describe the nonlinear mechanics of DNA buckle formation and packing within the cavities. As in Chapter 4, the length of DNA is kept constant during the simulations (L_{P22} = 150 bp and L_{T7} = 300 bp). We prescribe the packaging of DNA through the shortening of the rod (δ) achieved by translating the lower boundary incrementally upward by one nanometer at a time. For each δ, the solution is equilibrated by dynamic relaxation. Snapshots of the equilibrium conformations are shown in Fig. 5.4.

We initially simulate twist free equilibria and therefore allow the upper boundary to freely rotate about the tail axis (but clamp it against rotations about the other two axes). The bottom boundary is fully clamped as indicated by the constant angular orientation of the black stripe at the lower boundary. A very small initial twist is prescribed to bias the formation of a right handed buckle. Recall that supercoil topology is measured by its linking number ΔLk which is the sum of twist (ΔTw) and writhe (Wr). As a buckle develops, it generates writhe which alters the linking number. The twist-free boundary conditions dictates $\Delta\cancel{Tw}^{0}$ so ΔLk changes equal and opposite to Wr. Thus, the superhelical state of a buckle can be quantified by monitoring the angular rotation of the upper boundary; note the black stripe.

DNA in both cavities adopts toroidal supercoils similar to those found in ϕ29; see

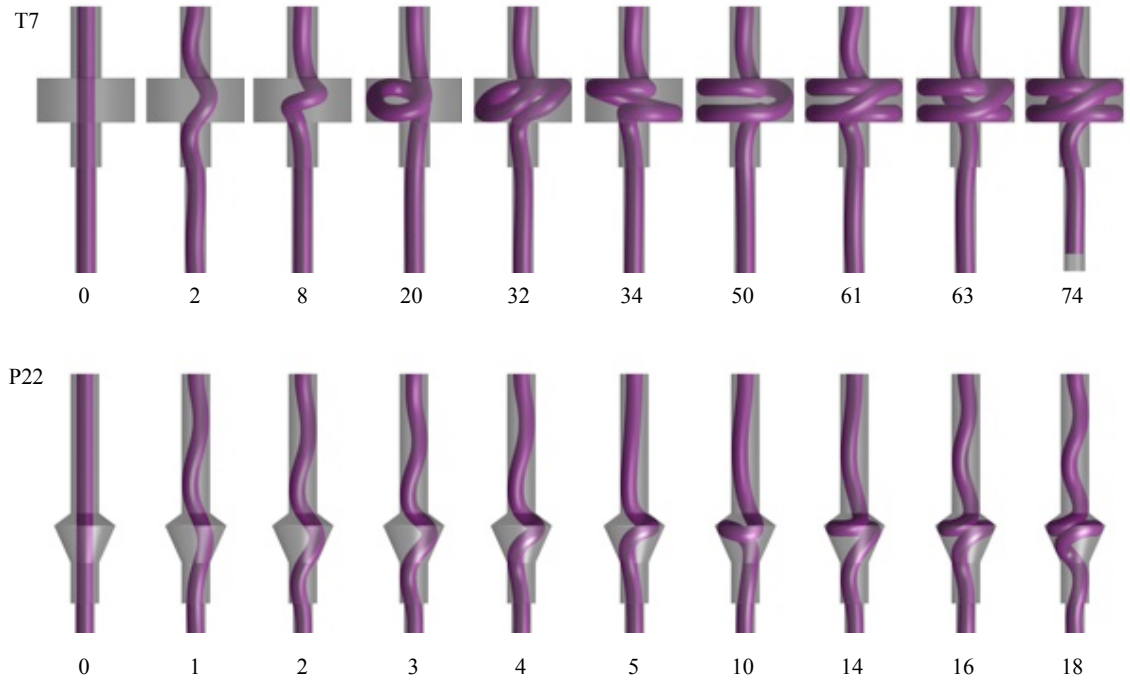


Figure 5.4 Snapshots of buckled equilibrium conformations in T7 and P22 at various values of δ (nm).

Fig. 5.4. The recent high-resolution P22 reconstruction of Tang et al. [128] did not reveal a toroid in the portal as we predict. However, they report densities about $2\times$ wider than dsDNA in the portal which they attributed to DNA. While the density could correspond to a tail protein, it could also correspond to denatured DNA resulting from specific interactions between straight or supercoiled DNA and the portal. Clearly, more detailed MD simulations are necessary to determine exactly how DNA interacts with the portal. In T7, the large toroidal supercoils could well be present but unresolvable given the low 24 Å resolution of the reconstruction.

Next, we compute a force-deflection curve to determine the forces required during buckling. Figure 5.5 reports the axial forces at the upper and lower boundaries as a function of packaging (δ).

The forces required for maintain the buckled conformations vary dramatically during the packaging process. Note the many force regimes evident in this figure. As in the case of $\phi 29$, the compressive force initially increases as the slender buckle contacts the tube regions. Upon inserting more DNA, a large, compliant buckle forms and the compressive force drops to $\sim$20 pN (P22) and $\sim$5 pN (T7). However, force levels in P22 increase

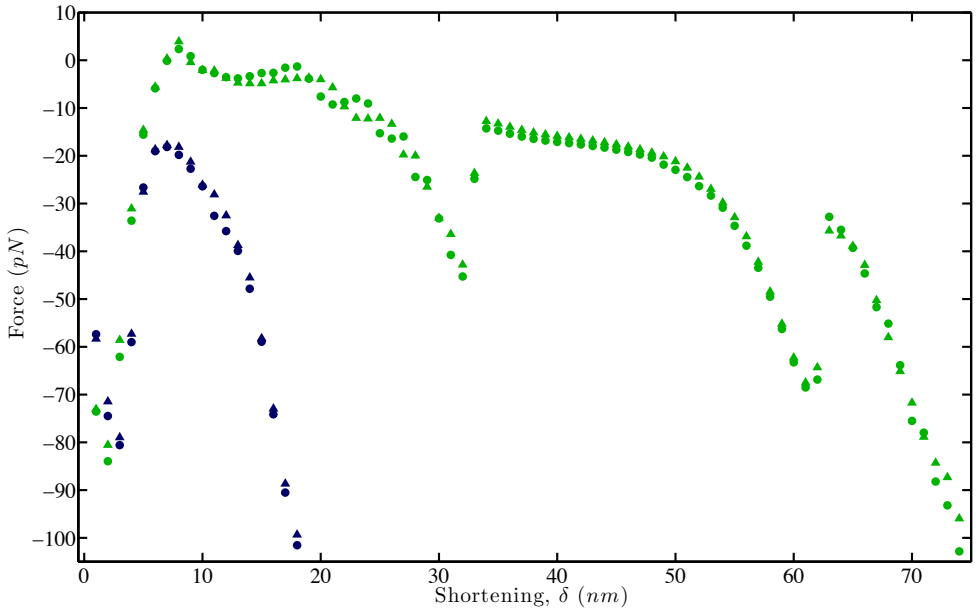


Figure 5.5 Axial force (*pN*) reported at the top (▲) and bottom (●) rod boundary as a function of δ (*nm*). T7 is shown in green and P22 in blue.

sharply as the cavity fills where electrostatic interactions resist further compression. We cease the simulations when force levels exceed reported packing motor limits ($\sim$100 *pN* [114, 123]). T7 exhibits additional interesting behavior. Beyond about 20 *nm*, force levels slowly rise until a transition occurs near 30 *nm* that once again reduces the compressive forces. Observe the large conformational change between 32 and 34 *nm* in Fig. 5.4. This 'saw tooth' behavior repeats again near 60 *nm*, and again observe the large conformational change between 61 and 63 *nm* in Fig. 5.4. Eventually, after wrapping two passes along the circumference of the cavity, DNA begins to fill the center. Akin to P22, electrostatics results in high forces that prevent further filling. In total, the P22/T7 cavities can accommodate $\sim$50/220 *bp*, respectively before 'lock up' occurs due to electrostatics.

Despite the different force profiles, both cavities exhibit similar behavior. First, both force profiles remain within the capabilities of the 100 *pN* ϕ29 packing motor. Both undergo oscillations in the forces that could serve as a signaling mechanism that a packing motor detects to stop packaging. Such a force sensing mechanism could even be employed by phages without large cavities such as P-SSP7 [83] as denatured DNA would also appear to become suddenly 'soft' to a motor like that of a buckle. Another interesting similarity is that the force at the top and bottom boundaries remain nearly equal (refer to triangle and circles) through filling despite major differences in the barrel and funnel shaped cavities. As

a benchmark, we simulated a fully symmetric T7 cavity with equal length tube section and observed equal forces. Thus, any slight differences arise purely from the geometry of the cavities.

Energy considerations provide additional insights regarding the mechanics of cavity filling. The elastic energy follows from (2.6) and electrostatic energy from (5.2). The contact energy is computed from

$$E_{contact} = \sum_{\#\ contact\ points} \frac{C_f \left(r_{p,q} - r_o\right)^{(1-\beta)}}{(1-\beta)}. \tag{5.6}$$

The contact energies are summed over all contact points identified in the adaptive algorithm and C_f is set to 0.11 $pN{\cdot}\text{Å}^6$. We plot the electrostatic, elastic, and contact energies as a function of packaging (δ) in Fig. 5.6.

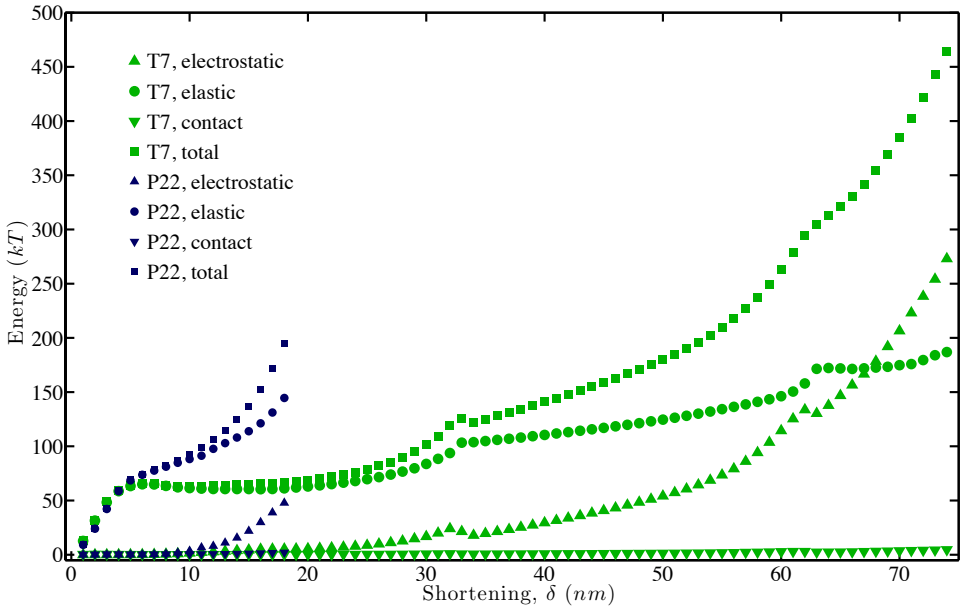


Figure 5.6 Componentwise energies including elastic, contact, and electrostatic (kT) as a function of δ (nm).

The elastic energy dominates the electrostatic energy and negligible (though non-zero) contact energy for most T7 conformations and all of the P22 conformations. The electrostatic energy in T7 finally overtakes the elastic energy $\delta = 68$ nm. While the elastic energy increases monotonically in both cases, the electrostatic energy in T7 is locally reduced at precisely the large DNA conformational transitions mentioned above. Interestingly when the cavities are full, P22's elastic energy is $3\times$ greater than its electrostatic energy while

T7's electrostatic is 1.5× its elastic. Although T7 clearly contains more energetic DNA, it is striking that both cavities exhibit similar volumetric energy density $\sim 1 \times 10^{-3}\ kT/\text{Å}^3$. Computing the energetic cost between subsequent steps in packaging provides a quantitative means to estimate the total energetic cost that a motor expends to package an additional nanometer of DNA. It is again striking that this energetic cost is equivalent for the two phages, and is equal to $\sim 22\ kT$.

We hypothesize that DNA buckling could serve one or more additional biological functions in addition to the three already proposed for $\phi 29$. They are:

1. *DNA buckling signals portal conformational changes during packaging.* A large buckle could function as a force transducer by converting the axial force applied by the packing motor to lateral forces at specific locations in the portal. Figure 5.7 shows the distribution of contact forces on the cavity walls for buckles subject to ~ 100 *pN* axial force. For T7, the forces vary between 5-30 *pN* on the outer rings to a maximum of 48 *pN* where the inner DNA contacts the top of the cavity. P22 achieves forces between 50-60 *pN* in locations near the top of the cavity. A DNA-triggered conformational change could provide the 'head-full' mechanism for the motor to cease packaging. Thus, a buckle could be thought of as a 'mechanical fuse' which prevents overfilling. Note that such a mechanism can be deployed without additional proteins and relies solely on the mechanics of compressed DNA.

The mechanism we propose adds yet another layer of sophistication to the head-full mechanisms proposed for P22. An early reconstruction of P22 detected a highly-resolved 'belt' of DNA that wraps around the outside of the lower section of the portal [76]; see Fig. 5.1. The authors hypothesized that, under pressure, the DNA belt 'pinched' the portal closed. A subsequent reconstruction observed conformational changes of the upper barrel section upon filling and branded the barrel as a sophisticated pressure sensor [128].

Contrary to these mechanisms that signal the portal from the outside, a DNA buckle would signal the portal from the inside. Moreover, DNA buckling would also be a reversible signal unlike the proposed DNA belt. If the portal is switched 'closed' by capsid pressure, then during ejection, DNA must somehow exit the capsid (with the portal closed) to relieve the pressure and reverse the signal. By contrast, a buckle is an energetic structure that is relieved the instant the tail is opened. The portal could detect when a buckle collapses and open fully for genome release.

2. *DNA buckles temporarily retain genome prior to tail assembly.* Agirrezabala et al. [2] show that the central cavity in T7 collapses from 50 Å to 30 Å and hypothesize that this portal contraction could be sufficient to temporarily prevent genome ejection. White tail-

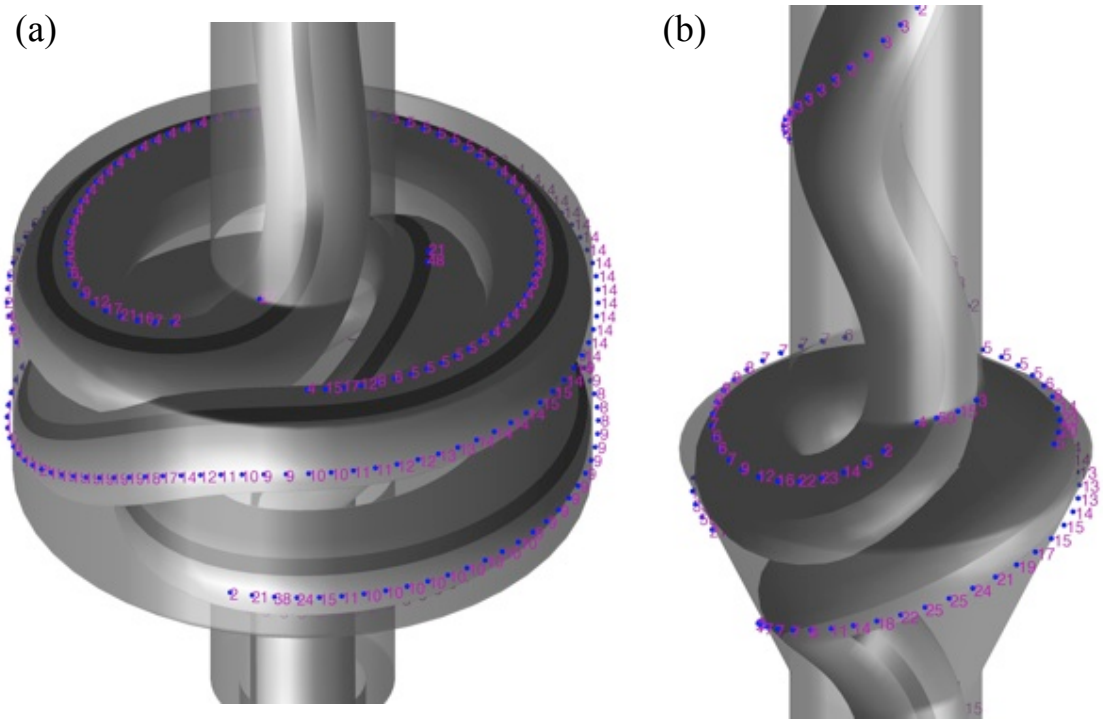


Figure 5.7 Distribution of contact forces for the fully packed (a) T7 and (b) P22 cavities. The adaptive algorithm identifies the contact points shown in blue and reports the compressive force (pN) shown in magenta.

deleted mutants will form filled capsids, they easily lose their genomes without a permanent plug [13]. A buckle could stabilize the genome by significantly increasing the DNA/protein surface contact area. The largest T7 buckle generates over an order of magnitude increase in DNA contour length contact versus a straight segment of DNA. In the case of $\phi 29$ and P22, positive (i.e. attractive to DNA) charges occupy the center of the portal [24] and may be 'sticky' to DNA and prevent sliding. The well-known capstan formula for belt friction relates the tension on two ends of a rope wound around a cylinder (capstan) with tension varying exponentially with wrap angle. A DNA buckle could be viewed as a three-dimensional analog of that simpler problem and even mild attraction to the cavity could prevent a buckle from collapsing.

5.2 Conclusions

In this chapter we explore the possibility of DNA buckling and mechanics of DNA buckles in a family of phages. Using Greenhill's equation, we predict that DNA can buckle under biologically relevant conditions derived from dimensions of cryo-EM reconstructions. Furthermore, we compute the mechanics of DNA buckling in bacteriophages T7 and P22 and discover that the forces required to initiate and grow a buckle are within packing motor capabilities. The resulting simulations provide evidence that a DNA buckle could perform several biological functions in addition to those we mention for $\phi 29$ in Chapter 4. Specifically, we demonstrate that a DNA buckle could provide large forces on the cavity interior to signal conformational changes. A DNA buckle may also help retain the genome prior to tail assembly by providing significant contact area with the portal.

It is important to emphasize that the predictions above are based on modeling compressed DNA in a cavity. The model clearly predicts that DNA buckles can form and exist under biologically relevant conditions. However, unlike $\phi 29$, no corroborating experimental evidence yet exists for DNA buckles in T7 and P22. The low 24 Å resolution of T7 limits our ability to make meaningful comparisons with the cryo-EM data but may well contain buckles that would become discernible at higher resolution. As discussed above, the P22 portal does contain an unknown structure with density about twice the width of dsDNA. The multi-scale model developed in Chapter 4 could again be employed but with greater accuracy as a portal crystal structure is available. Such modeling could predict the DNA structure and degree of denaturation responsible for the observed wide density.

Chapter 6

Summary of Major Contributions and Specific Findings

The goal of this dissertation research is to contribute novel modeling tools to provide needed insight into biological systems that fundamentally depend on the mechanics of DNA. We make significant strides toward understanding the role DNA mechanics plays in two systems including Lac repressor DNA looping and viral DNA packaging. In this chapter, we summarize major contributions and specific findings as found in Chapter 2 for LacI looping and Chapters 3 - 5 for viral DNA packaging.

LacI DNA Looping

Mounting theoretical and experimental evidence suggests that the Lac repressor protein adopts alternate conformations beyond what is reported in the crystallographic 'V'-shaped conformation. In Chapter 2, we explored the possibility of an "extended" protein conformation by reinterpreting a wealth of data from two sets of classic LacI looping experiments conducted by the Müller-Hill lab over 25 years ago. The rod model was exercised over a set of binding topologies and interoperator DNA lengths (153-168 bp) for linear and supercoiled minicircle DNA. For both DNA substrates, the computed energetic minima do not coincide with experimentally measured optimal interoperator lengths (i.e. operator phasing) for complex stability when we assume the 'V' conformation. However, the model successfully predicts the optimal interoperator lengths for stability upon introducing an approximation for an extended protein conformation. For linear DNA, the extended conformation possesses the lowest overall energy and correct phasing for complex stability. Specifically, computed energetic minima at 157 and 168 bp correlate with the most stable loops of 158 and 168 bp determined experimentally. For minicircle DNA, the model successfully predicts the optimal interoperator DNA lengths for minicircles with supercoiling of ΔLk = -1 (162 bp), 0 (157, 168 bp), and +1 (155, 165 bp). For minicircles, optimal stability may arise from either the

closed or the extended protein conformation depending on the degree of supercoiling and the interoperator length. For the topoisomerase I relaxation assays, we compute predicted band distributions (both with and without the extended conformation) for direct comparison with electrophoretic gel data. The predicted gel data aligns strikingly well with experimental measurements for the ΔLk = -1 minicircle. The remaining values of supercoiled minicircles ($\Delta Lk = \{+1, 0, -2\}$) unfortunately lack experimental counterparts in the Müller-Hill data to compare with our predictions. We predict that band distributions vary significantly whether including or excluding the extended conformation. Therefore, experiments for these minicircles (should they be carried out) could be used to determine the existence of the extended conformation.

DNA Packaging and Ejection in Bacteriophages

In Chapters 3 - 5, we develop and implement a model for highly-strained DNA buckled under compression in bacteriophage portals/tails. To our knowledge, we contribute the first model for DNA under significant compression at the resolution of a helical turn.

Chapter 3 extends the computational model used in Chapter 2 by accounting for DNA interactions with a surrounding protein cavity. The cavity selected is motivated by that appearing in the exit domain of $\phi 29$. In particular, we employ an inverse power law to describe the mechanical contact forces between DNA and the cavity. By considering order-of-magnitude changes to the associated potential, we observe negligible differences in the DNA conformation and compressive forces at the DNA boundaries at equilibrium. This result indicates that our predictions remain insensitive to the assumed (and unknown) model parameters we choose to represent DNA-protein contact. We also demonstrate that our predictions remain insensitive to the discretization scale chosen for the cavity wall, provided that the cavity wall grid point spacing remains smaller than $\sim$4 Å. The computational model that results is the first rod model that describes the nonlinear mechanics (including buckling) of DNA within a protein cavity. This model also represents a contribution to the general field of mechanics by extending well beyond the specialized theories for constrained buckling between parallel plated or within tubes.

As part of a collaboration with the Andricioaei lab at the University of California - Irvine, we conducted a thorough analysis of the DNA toroid in bacteriophage $\phi 29$ proposed by the Baker lab in Chapter 4. We simulate DNA buckling into a toroid and its subsequent ejection from the connector-lower collar cavity. The rod model predicts that an ensemble of toroids can mechanically form under biologically-relevant force levels ($\sim$25 pN). We also develop the first multi-scale model for DNA under compression by using the rod to generate a family of all-atom initial conditions for MD. Equilibration in MD (Andricioaei lab) reveals

that the dsDNA structure is largely preserved under sharp bending through non-specific interactions with the cavity. Both the computed energy landscape (Andricioaei lab) and the predicted cryo-EM density maps support the likelihood of an ensemble of different sized toroids which combine to form the averaged experimental density map. Simulating the dynamic ejection of the toroid using the rod model reveals large reaction forces/torques ($\sim$100 pN/70 $pN-nm$) at the connector. We postulate that these large reactions could signal an additional conformational change to allow the capsid DNA unhindered escape into the host cell. Thus, the mechanics of the DNA toroid may not only help stabilize the phage, but also play an active role in triggering genome release.

Given that DNA buckles into a toroid in ϕ29, we investigate the possibility and biological function(s) of buckling in a family of phages in Chapter 5. Greenhill's equation predicts that buckling is feasible under biologically relevant force and twist levels supplied by a packing motor. To initiate buckling in cavities of T7, ϕ29, and P22 requires compressive forces of -82, -68, and -57 pN, respectively. Both T7 and P22 possess large cavities in their portal systems which are formed prior to DNA packaging (unlike ϕ29) and downstream of the packing motor. To explore buckling in T7 and P22, we further advance the previous rod model by incorporating DNA self-repulsive electrostatics and an adaptive algorithm that efficiently identifies DNA/cavity contact. The adaptive algorithm significantly reduces computation time for modeling longer lengths of DNA in T7 (300 bp) and P22 (150 bp). Surprisingly, we find that the compressive forces on DNA, while within motor capabilities, vary dramatically as a function of the amount of DNA inserted. Such fluctuations could be interpreted by a motor as the signal to terminate packaging. Under $\sim$100 pN axial force on DNA, large reaction forces of $\sim$50 pN (T7) and $\sim$60 pN (P22) develop on the cavity walls which could be exploited to modulate conformational changes in the phage portals. In addition, a buckled DNA structure significantly increases the amount of DNA contour length in contact with the portal surface which may help temporarily retain the genome prior to tail assembly. For the largest T7 buckle, this surface contact increases by more than an order of magnitude relative to a straight segment of unbuckled DNA.

Appendices

Appendix A

Interpretation of gel data, Krämer et al. 1987 [74]

The looped complexes in the linear DNA experiments [74] exhibited noticeably different gel migration behavior and this provides an opportunity for comparison with the model. In particular, the migration velocity should correlate with the radius of gyration (R_g) of the LacI-DNA complex (i.e. larger complexes having slower migration velocities). In Fig. A.1 we report the computed radius of gyration of the minimum energy DNA-LacI complex as a function of interoperator length. The R_g is estimated from the computed equilibrium conformation as the root mean-square distance of each atom in the DNA-LacI complex from the average position of all the atoms. The atoms for each basepair are lumped into the associated spatial grid point of the discretized rod model; and the atoms of each amino acid are lumped into a single point located at the associated C-α atom in the crystal structure; only the C-α atoms are resolved in the crystal structure [78]. For more details, see [46, 82]. The **A1** binding topology (circles/blue) exhibits the largest R_g followed by smaller sized $\mathbf{P1^E}$ (stars/red), the **P1** (squares/green), and the smallest **P2** binding topology (triangles/black). Discontinuities, particularly observable for the **P2** binding topology, arise from the discontinuous switch in the lowest energy topoisomer (i.e. over- versus under-twisted).

A closer examination of the electrophoretic gel (Figure 2A in Krämer's original manuscript [74]) reveals a dependence of migration velocity on interoperator length. Specifically, resolved/tight bands (indicating stable complexes) appear centered about 158 bp, while faster diffuse/smeared bands appear centered about 163 bp. These diffuse bands are shown to migrate faster than the stable looped complexes (in gel lanes 156/158/159/160 bp), but slower than the singly-bound complex [74]. Since we predict the stable looped complexes possess $\mathbf{P1^E}$ binding topology, the diffuse bands may represent an additional binding topology present in the thermal ensemble. The likely candidate is the next-highest energy **A1** binding topology, which becomes energetically preferred in the region around 163 bp

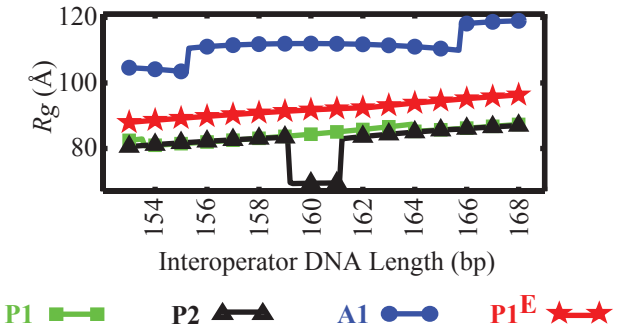


Figure A.1 R_g (Å) as a function of interoperator DNA length (*bp*). Results for the **A1** binding topology are indicated by circles (blue), the **P1** by squares (green), the **P2** by triangles (black), and the **P1**$^{\mathbf{E}}$ by stars (red).

if the energy penalty for **P1**$^{\mathbf{E}}$ increases. However, in Fig. A.1 we show the **A1** binding topology has a significantly larger R_g and is therefore predicted to migrate slower than **P1**$^{\mathbf{E}}$. This suggests that the diffuse bands near 163 *bp* do not include **A1** loops. A possible explanation, also proposed in [74], is that the faster moving diffuse bands could result from a "rapid equilibrium" between looped (**P1**$^{\mathbf{E}}$) and unlooped states. There are, however, limitations to relating R_g to migration velocity. Nevertheless, similar R_g calculations in our prior work [46, 82] demonstrated a correlation between computed R_g and experimental values of migration velocity.

Appendix B

Supporting Material for Chapter 4

This appendix contains supplementary data and methods for Chapter 4.

Potential Energy Landscape for Left-Handed Toroids

Currently it remains unknown whether the toroid is a right- or left-handed supercoil. While we believe the toroid is likely a right-handed helical supercoil, for completeness we include the 2-D energy landscape for left-handed toroids (see Fig. B.1). As with right-handed toroids, the potential energy for left-handed toroids exhibits large regions of low energy (blue/magenta) interrupted by patches of slightly higher energy (yellow) which vary with toroid size and registry. While the location of these patches differs slightly from the right-handed landscape, the left-handed toroids remain essentially energetically equivalent suggesting no energetic preference between a right- or left-handed configuration. Given that the right- and left-handed landscapes remain energetically equivalent, it is reasonable to postulate that the reconstruction of [129] may reflect an averaged ensemble of both types.

Sensitivity to L-J Potential Well Depth, ε

To determine the optimal interaction strength between DNA and the cavity, we present simulations which vary the strength of the Lennard-Jones (L-J) interaction potential, ε. Specifically, the system was equilibrated with five different values for the L-J potential depth (ε = 0, 0.02, 0.2, 2, 20 $kcal/mol$) following identical simulation protocol and monitoring the stability of the DNA double helix. In Fig. B.2, denatured regions (as indicated by large H-bond distance) were observed in the case of very weak interaction (ε = 0.02 $kcal/mol$) and near complete DNA destruction was observed for extremely strong interaction (ε = 20.0 $kcal/mol$). Minor deviations from the double-stranded structure were observed when epsilon was set to 0.2 $kcal/mol$ (see Fig. 4.5). This important observation supports the argument that the sharply bent DNA double helix in ϕ29 is sensitive to its environment. In the real system, the toroidal DNA motif inside the cavity is part of a complex interaction

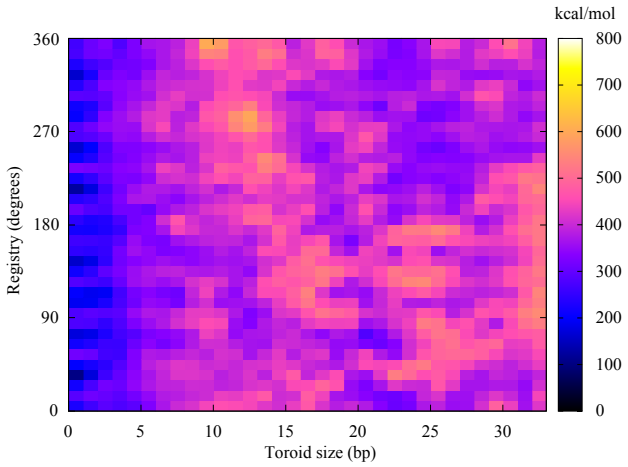


Figure B.1 Results courtesy of M. Taranova at the Andricioaei lab. Potential energy landscape for left-handed toroids of varying size (*bp*) and registry value (degrees). As in the right-handed landscape, energy values are scaled to the minimum value in the corresponding landscape.

network of the phage's molecular components. For instance, sequence-specific DNA-protein interactions and long-range electrostatic interactions with the capsid proteins and capsid DNA might impact the structure of the toroidal DNA.

Constructing Predicted Density Maps

UCSF Chimera [107] was used to generate the predicted density maps from all-atom energy-minimized toroids using the built in function *molmap*. Grid spacing is set to 1 Å and the width of the Gaussian distribution used to describe each atom, *sigmaFactor*, is set to 0.425. Rotational averaging is performed using *cyclic symmetry* by placing 100 equivalent copies around the vertical tail axis. A cross section is taken and 20 color bins are scaled between the highest volume data value (yellow) relative to the lowest (dark blue). The density maps remain unchanged upon adding more copies so long as the number of copies is sufficient. Thus, the maps are well converged using 100 copies. To smooth the color gradients, *per-pixel coloring* is selected. All other settings are default.

Additional Predicted Density Maps

We present additional predicted density maps to show how they change upon parametrically varying the registry for the right- and left-handed 33 *bp* toroid (Fig. B.3(a, b), respectively) which show slight changes in the size and shape of the central void but all possess toroid intensity which agrees with the experimental study [129].

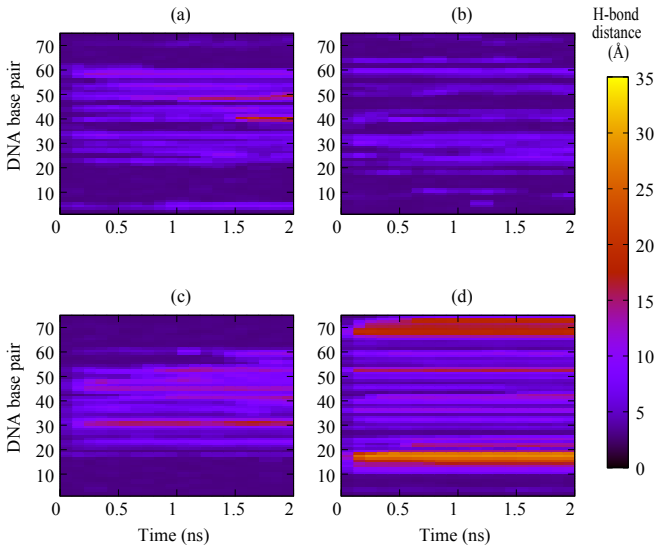


Figure B.2 Results courtesy of M. Taranova at the Andricioaei lab. Time evolution of the Watson-Crick base pair hydrogen bond distances between complementary bases during simulation. The simulated DNA is 75 bp long and contains the 33 bp toroid. Each plot represents a simulation with different values of the Lennard-Jones parameter ε = (a) 0.02, (b) 0.2, (c) 2, and (d) 20 $kcal/mol$. While some denatured regions (red colored regions) are observed in (a), severe denaturation (red and yellow colored regions) develops in (c) and (d). Plot (b) contains some regions, where Watson-Crick basepair hydrogen bond distances are slightly larger than 3-4 Å, but overall the structure does not exhibit any signature of denaturation.

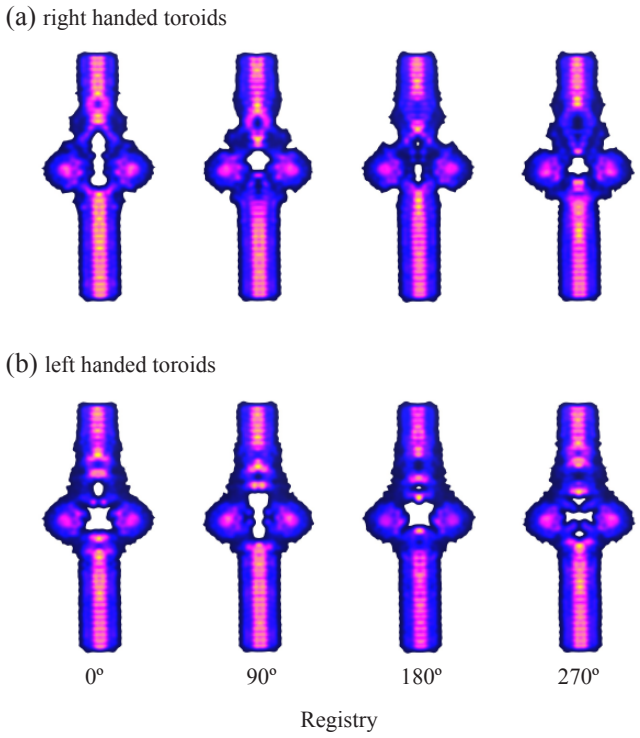


Figure B.3 Predicted density maps upon parametrically varying registry value (degrees) for the (a) right-handed and (b) left-handed 33 *bp* toroid. Note minor changes in the shape of the central void and smaller holes as registry and handedness change.

Appendix C

Greenhill Buckling Criteria for Fixed Ends

In this appendix, we compute the force required to buckle upon augmenting Greenhill's equation for fixed/clamped boundary conditions. Recall that we have reported the opposite extreme of pinned boundary conditions in Fig. 5.2.

Our analysis follows the formulation of Lu et al. [85] for pinned boundary conditions. Lu et al. re-cast the boundary value problem, initially expressed in the Serret-Frenet triad, in Cartesian coordinates to match Greenhill's analysis for buckling a straight rod. Upon ignoring distributed body forces, we obtain the linearized governing equations for transverse rod deflection given by

$$-x_2'''' - m_a x_3''' + f_a x_2'' = 0 \tag{C.1}$$

$$-x_3'''' + m_a x_2''' + f_a x_3'' = 0 \tag{C.2}$$

where f_a and m_a denote non-dimensional applied thrust and torque, respectively. Recall that for pinned ends, the boundary conditions are

$$x_2 = 0, \;\; x_3 = 0 \quad @s = 0 \tag{C.3}$$

$$x_2'' = 0, \;\; x_3'' = 0 \quad @s = 1 \tag{C.4}$$

The coupled equations (C.1) - (C.2) can be written as a single function of a complex variable. Substituting the complex displacement $z(s) = x_2(s) + i x_3(s)$ into (C.1) - (C.2) and

simplifying yields

$$-z'''' + im_a z''' + f_a z'' = 0 \tag{C.5}$$

with boundary conditions

$$z = 0, \;\; z'' = 0 \quad @s = 0 \tag{C.6}$$

$$z = 0, \;\; z'' = 0 \quad @s = 1 \tag{C.7}$$

The general solution to (C.5), provided $f_a \neq 0$, is

$$z(s) \;=\; h_1 e^{i\alpha_1 s} + h_2 e^{i\alpha_2 s} + h_3 s + h_4, \quad \textit{where } \alpha_{1,2} \;=\; \frac{m_a \pm \sqrt{{m_a}^2 - 4f_a}}{2} \tag{C.8}$$

Substituting (C.8) into (C.5) and applying the boundary conditions (C.6)/(C.7) simplifies to

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ -{\alpha_1}^2 & -{\alpha_2}^2 & 0 & 0 \\ e^{i\alpha_1} & e^{i\alpha_2} & 1 & 1 \\ -{\alpha_1}^2 e^{i\alpha_1} & -{\alpha_2}^2 e^{i\alpha_2} & 0 & 0 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{C.9}$$

The non-trivial solution to (C.9) exists when the determinant of the coefficient matrix equals zero. Upon solving for the vanishing determinant, we reproduce the well-known Greenhill equation in (5.1).

For the fixed case, the boundary conditions now become

$$z = 0, \;\; z' = 0 \quad @s = 0 \tag{C.10}$$

$$z = 0, \;\; z' = 0 \quad @s = 1 \tag{C.11}$$

Evaluating (C.8) at (C.10)/(C.11) and simplifying yields

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ i\alpha_1 & i\alpha_2 & 1 & 0 \\ e^{i\alpha_1} & e^{i\alpha_2} & 1 & 1 \\ i\alpha_1 e^{i\alpha_1} & i\alpha_2 e^{i\alpha_2} & 1 & 0 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{C.12}$$

To solve for the vanishing determinant in (C.12), we use Newton's Method and specify an

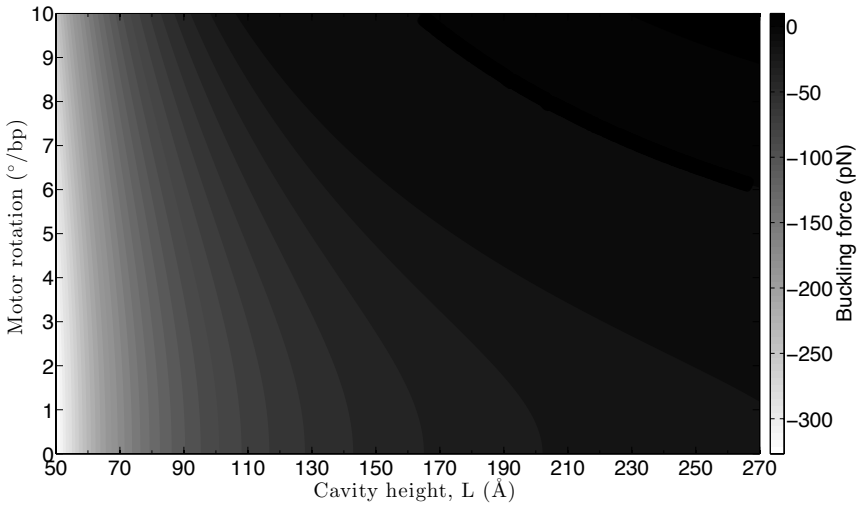


Figure C.1 Predicted buckling force for a rod upon augmenting Greenhill's equation for fixed boundaries. Buckling force (pN) plotted as a function of cavity height (Å, x-axis) and possible motor rotation ($^\circ/bp$, y-axis). The y-axis denotes lines of constant torque and the negative sign indicates compression. Again, we consider up to $10^\circ/bp$ rotations as has been demonstrated experimentally for $\phi 29$ (personal communication, C. L. Hetherington, Bustamante lab, 2013).

initial guess for the buckling force (f_a) equal to $-4\pi^2 EI/L^2 + M^2/4EI$ (i.e. the buckling load for a fixed beam plus the second term in Greenhill for the pinned case). We plot the updated buckling load for fixed boundary conditions in Fig. C.1 below.

The predicted buckling force for a rod with fixed boundaries increases from the pinned case as expected. For the short cavities in Fig. 5.1, the buckling forces increase in magnitude from -82 to -327 pN (T7), -68 to -270 pN ($\phi 29$), and -57 to -227 pN (P22) for the limiting case of zero torque. While the magnitude of the fixed-boundary buckling force is large, it is clearly an extreme upper limit and the largest forces are still within the same order of magnitude (10^2 pN) as those attainable by a packing motor. As already mentioned, the 'real' boundary conditions in the biological system fall in between the pinned or fixed extremes.

Bibliography

[1] Adleman, L. M. 1994. Molecular computation of solutions to combinatorial problems. *Science* 266:1021–4.

[2] Agirrezabala, X., J. Martín-Benito, J. R. Castón, R. Miranda, J. M. Valpuesta, and J. L. Carrascosa. 2005. Maturation of phage T7 involves structural modification of both shell and inner core components. *EMBO J.* 24:3820–29.

[3] Agrawal, N. J., R. Radhakrishnan, and P. K. Purohit. 2008. Geometry of mediating protein affects the probability of loop formation in DNA. *Biophys. J.* 94:3150–3158.

[4] Al Lawati, A., I. Ali, and M. Al Barwani. 2013. Effect of temperature and capsid tail on the packing and ejection of viral DNA. *PLoS One* 8:e52958.

[5] Balaeff, A, C. R. Koudella, L. Mahadevan, and K. Schulten. 2004. Modelling DNA loops using continuum and statistical mechanics. *Philos. T. Roy. Soc. A.* 362:1355–1371.

[6] Balaeff, A., L. Mahadevan, and K. Schulten. 2006. Modeling DNA loops using the theory of elasticity. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 73:031919.

[7] Becker, N. A., J. D. Kahn, L. J. Maher. 2005. Bacterial repression loops require enhanced DNA flexibility. *J. Mol. Biol.* 349:716–730.

[8] Bellomy, G. R., M. C. Mossing, and M. T. Record. 1988. Physical properties of DNA *in vivo* as probed by the length dependence of the lac operator looping process. *Biochemistry* 27:3900–3906.

[9] Beveridge, D. L., G. Barreiro, K. Suzie Byun, D. A. Case, T. E. Cheatham, S. B. Dixit, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai, and M. A. Young. 2004. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. research design and results on d(CpG) steps. *Biophys. J.* 87:3799–3813.

[10] Bustamante, C., Z. Bryant, and S. B. Smith. 2003. Ten years of tension: Single-molecule DNA mechanics. *Nature* 421:423–427.

[11] Calladine, C. R., H. R. Drew, B. F. Luisi, and A. A. Travers. 2004. Understanding DNA: the molecule and how it works. 3rd edition. Elsevier Academic Press.

[12] Casjens, S. R. and I. J. Molineux. 2012. Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. *Adv. Exp. Med. Biol.* 726:143–79.

[13] Cerritelli, M. E., N. Cheng, A. H. Rosenberg, C. E. McPherson, F. P. Booy, and A. C. Steven. 1997. Encapsidated conformation of bacteriophage T7 DNA. *Cell* 91:271–80.

[14] Chang, J., P. Weigele, J. King, W. Chiu, and W. Jiang. 2006. Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. *Structure* 14:1073–1082.

[15] Chemla, Y., K. Aathavan, J. Michaelis, S. Grimes, P. Jardine, D. Anderson, and C. Bustamante. 2005. Mechanism of force generation of a viral DNA packaging motor. *Cell* 122:683–692.

[16] Chemla, Y. R. and D. E. Smith. 2012. Single-molecule studies of viral DNA packaging. *Adv. Exp. Med. Biol.* 726:549–84.

[17] Chen, D.-H., M. L. Baker, C. F. Hryc, F. DiMaio, J. Jakana, W. Wu, M. Dougherty, C. Haase-Pettingell, M. F. Schmid, W. Jiang, D. Baker, J. A. King, and W. Chiu. 2011. Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *P. Natl. Acad. Sci. USA*. 108:1355–60.

[18] Chiaruttini, N., M. de Frutos, E. Augarde, P. Boulanger, L. Letellier, and V. Viasnoff. 2010. Is the *in vitro* ejection of bacteriophage DNA quasistatic? A bulk to single virus study. *Biophys. J.* 99:447–455.

[19] Choi, K. H., J. McPartland, I. Kaganman, V. D. Bowman, L. B. Rothman-Denes, and M. G. Rossmann. 2008. Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. *J. Mol. Biol.* 378:726–36.

[20] Chung, J. and G. M. Hulbert. 1993. A time integration algorithm for structural dynamics with improved numerical dissipation: the generalized-α method. *J. Appl. Mech.* 60:371–375.

[21] Cohen, D. N., Y. Y. Sham, G. D. Haugstad, Y. Xiang, M. G. Rossmann, D. L. Anderson, and D. L. Popham. 2009. Shared catalysis in virus entry and bacterial cell wall depolymerization. *J. Mol. Biol.* 387:607–18.

[22] Coleman, B. D., W. K. Olson, and D. Swigon. 2003. Theory of sequence-dependent DNA elasticity. *J. Chem. Phys.* 118:7127–7140.

[23] Comolli, L. R., A. J. Spakowitz, C. E. Siegerist, P. J. Jardine, S. Grimes, D. L. Anderson, C. Bustamante, and K. H. Downing. 2008. Three-dimensional architecture of

the bacteriophage $\phi 29$ packaged genome and elucidation of its packaging process. *Virology* 371:267–277.

[24] Cuervo, A. and J. L. Carrascosa. 2012. Viral connectors for DNA encapsulation. *Curr. Opin. Biotech.* 23:529–36.

[25] Domokos, G., P. Holmes, and B. Royce. 1997. Constrained Euler buckling. *J. Nonlinear Sci.* 7:281–314.

[26] Du, Q., A. Kotlyar, and A. Vologodskii. 2008. Kinking the double helix by bending deformation. *Nucleic Acids Res.* 36:1120–8.

[27] Dubochet, J., M. Adrian, J.-J. Chang, J.-C. Homo, J. Lepault, A. W. McDowall, and P. Schultz. 1988. Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* 21:129–228.

[28] Echols, H. 1990. Nucleoprotein structures initiating DNA-replication, transcription, and site-specific recombination. *J. Biol. Chem.* 265:14697–14700.

[29] Ecker, J. R., W. A. Bickmore, I. Barroso, J. K. Pritchard, Y. Gilad, and E. Segal. 2012. Genomics: Encode explained. *Nature* 489:52–5.

[30] Edelman, L. M., R. Cheong, and J. D. Kahn. 2003. Fluorescence resonance energy transfer over $\sim$130 basepairs in hyperstable Lac repressor-DNA loops. *Biophys. J.* 84:1131–1145.

[31] Fang, H., P. Jing, F. Haque, and P. Guo. 2012. Role of channel lysines and the "push through a one-way valve" mechanism of the viral DNA packaging motor. *Biophys. J.* 102:127–35.

[32] Fokine, A., P. R. Chipman, P. G. Leiman, V. V. Mesyanzhinov, V. B. Rao, and M. G. Rossmann. 2004. Molecular architecture of the prolate head of bacteriophage T4. *P. Natl. Acad. Sci. USA.* 101:6003–8.

[33] Friedman, A. M., T. O. Fischmann, and T. A. Steitz. 1995. Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science* 268:1721–1727.

[34] Fuller, D. N., D. M. Raymer, V. I. Kottadiel, V. B. Rao, and D. E. Smith. 2007. Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *P. Natl. Acad. Sci. USA.* 104:16868–73.

[35] Fuller, D. N., D. M. Raymer, J. P. Rickgauer, R. M. Robertson, C. E. Catalano, D. L. Anderson, S. Grimes, and D. E. Smith. 2007. Measurements of single DNA molecule packaging dynamics in bacteriophage λ reveal high forces, high motor processivity, and capsid transformations. *J. Mol. Biol.* 373:1113–22.

[36] Fuller, D. N., J. P. Rickgauer, P. J. Jardine, S. Grimes, D. L. Anderson, and D. E. Smith. 2007. Ionic effects on viral DNA packaging and portal motor function in bacteriophage ϕ29. *P. Natl. Acad. Sci. USA.* 104:11245–11250.

[37] Gemmen, G. J., R. Millin, and D. E. Smith. 2006. DNA looping by two-site restriction endonucleases: heterogeneous probability size and unbinding force. *Nucleic Acids Res.* 34:2864–2877.

[38] Geng, J., H. Fang, F. Hague, L. Zhang, and P. Guo. 2011. Three reversible and controllable discrete steps of channel gating of a viral DNA packaging motor. *Biomaterials* 32:8234–8242.

[39] Goldman, N., P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, and E. Birney. 2013. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* 494:77–80.

[40] González-Huici, V., M. Salas, and J. M. Hermoso. 2004. The push-pull mechanism of bacteriophage ϕ29 DNA injection. *Mol. Microbiol.* 52:529–540.

[41] Goyal, S. 2006. A dynamic rod model to simulate mechanics of cables and DNA. Ph.D. thesis, University of Michigan.

[42] Goyal, S., N. C. Perkins, and C. L. Lee. 2008. Non-linear dynamic intertwining of rods with self-contact. *Int. J. Nonlinear Mech.* 43:65–73.

[43] Goyal, S. and N. C. Perkins. 2008. Looping mechanics of rods and DNA with non-homogeneous and discontinuous stiffness. *Int. J. Nonlinear Mech.* 43:1121–1129.

[44] Goyal, S., N. C. Perkins, and C. L. Lee. 2005. Nonlinear dynamics and loop formation in Kirchhoff rods with implications to the mechanics of DNA and cables. *J. Comput. Phys.* 209:371–389.

[45] Goyal, S., N. C. Perkins, and J. -C. Meiners. 2008. Resolving the sequence-dependent stiffness of DNA using cyclization experiments and a computational rod model. *J. Comput. Nonlin. Dyn.* 3:011003.

[46] Goyal, S., T. Lillian, S. Blumberg, J. -C. Meiners, E. Meyhöfer, and N. C. Perkins. 2007. Intrinsic curvature of DNA influences LacR-mediated looping. *Biophys. J.* 93:4342–4359.

[47] Grayson, P., L. Han, T. Winther, and R. Phillips. 2007. Real-time observations of single bacteriophage λ DNA ejections *in vitro*. *P. Natl. Acad. Sci. USA*. 104:14652–14657.

[48] Greenhill, A. G. 1883. On the strength of shafting when exposed both to torsion and to end thrust. *Proc. Inst. Mech. Eng.* 6:182–225.

[49] Grimes, S., S. Ma, J. Gao, R. Atz, and P. J. Jardine. 2011. Role of ϕ29 connector channel loops in late-stage DNA packaging. *J. Mol. Biol.* 410:50–59.

[50] Hagerman, P. J. 1988. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* 17:265–286.

[51] Han, L., H. G. Garcia, S. Blumberg, K. B. Towles, J. F. Beausang, P. C. Nelson, and R. Phillips. 2009. Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS ONE.* 4:e5621.

[52] Hendrix, R. W. 2003. Bacteriophage genomics. *Curr. Opin. Microbiol.* 6:506–11.

[53] Hernando-Pérez, M., R. Miranda, M. Aznar, J. L. Carrascosa, I. A. T. Schaap, D. Reguera, and P. J. de Pablo. 2012. Direct measurement of phage phi29 stiffness provides evidence of internal pressure. *Small* 8:2366–70.

[54] Hirsh, A. D., T. D. Lillian, T. A. Lionberger, and N. C. Perkins. 2011. DNA modeling reveals an extended Lac repressor conformation in classic *in vitro* binding assays. *Biophys. J.* 101:718–26.

[55] Hirsh, A. D., T. D. Lillian, T. A. Lionberger, M. Taranova, I. Andricioaei, and N. C. Perkins. 2013. A model for highly strained DNA compressed inside a protein cavity. *J. Comput. Nonlin. Dyn.* 8:031001.

[56] Hirsh, A. D., M. Taranova, T. A. Lionberger, T. D. Lillian, I. Andricioaei, and N. C. Perkins. 2013. Structural ensemble and dynamics of toroidal-like DNA shapes in bacteriophage ϕ29 exit cavity. *Biophys. J.* 104:2058–2067.

[57] Holmes, P., G. Domokos, and G. Hek. 2000. Euler buckling in a potential field. *J. Nonlinear Sci.* 10:477–505.

[58] Holmes, P., G. Domokos, J. Schmitt, and I. Szeberenyi. 1999. Constrained Euler buckling: An interplay of computation and analysis. *Comput. Method. Appl. M.* 170:175–207.

[59] Howard, J. 2001. Mechanics of motor proteins and the cytoskeleton. Sinauer Associates, Sunderland, MA.

[60] Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graphics* 14:33–38.

[61] Inamdar, M. M., W. M. Gelbart, and R. Phillips. 2006. Dynamics of DNA ejection from bacteriophage. *Biophys. J.* 91:411–20.

[62] International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:931–45.

[63] Ivanovska, I. L., P. J. de Pablo, B. Ibarra, G. Sgalari, F. C. MacKintosh, J. L. Carrascosa, C. F. Schmidt, and G. J. L. Wuite. 2004. Bacteriophage capsids: tough nanoshells with complex elastic properties. *P. Natl. Acad. Sci. USA.* 101:7600–5.

[64] Jacob, M., T. Blu, C. Vaillant, J. H. Maddocks, and M. Unser. 2006. 3-D shape estimation of DNA molecules from stereo cryo-electron micro-graphs using a projection-steerable snake. *IEEE T. Image Process* 15:214–27.

[65] Jiang, W., J. Chang, J. Jakana, P. Weigele, J. King, and W. Chiu. 2006. Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature* 439:612–6.

[66] Jing, P., F. Haque, D. Shu, C. Montemagno, and P. Guo. 2010. One-way traffic of a viral motor channel for double-stranded DNA translocation. *Nano Lett.* 10:3620–3627.

[67] Kahn, J. D., R. Cheong, R. A. Mehta, L. M. Edelman, and M. A. Morgan. 2006. Flexibility and control of protein-DNA loops. *Biophys. Rev. Lett.* 1:327–341.

[68] Kamtekar, S., A. J. Berman, J. Wang, J. M. Lázaro, M. de Vega, L. Blanco, M. Salas, and T. A. Steitz. 2006. The ϕ29 DNA polymerase: protein-primer structure suggests a model for the initiation to elongation transition. *EMBO J.* 25:1335–43.

[69] Klapper, I. 1996. Biological applications of the dynamics of twisted elastic rods. *J. Comput. Phys.* 125:325–337.

[70] Klenin, K., H. Merlitz, and J. Langowski. 1998. A Brownian dynamics program for the simulation of linear and circular DNA and other wormlike chain polyelectrolytes. *Biophys. J.* 74:780–788.

[71] Klenin, K. and J. Langowski. 2000. Computation of writhe in modeling of supercoiled DNA. *Biopolymers.* 54:307–317.

[72] Koester, S., A. Evilevitch, M. Jeembaeva, and D. A. Weitz. 2009. Influence of internal capsid pressure on viral infection by phage λ. *Biophys. J.* 97:1525–1529.

[73] Krämer, H., M. Amouyal, A. Nordheim, and B. Müller-Hill. 1988. DNA supercoiling changes the spacing requirement of two lac operators for DNA loop formation with lac repressor. *EMBO J.* 7:547–556.

[74] Krämer, H., M. Niemöller, M. Amouyal, B. Revet, B. von Wilcken-Bergmann, and B. Müller-Hill. 1987. Lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. *EMBO J.* 6:1481–1491.

[75] LaMarque, J. C., T. L. Le, and S. C. Harvey. 2004. Packaging double-helical DNA into viral capsids. *Biopolymers* 73:348–355.

[76] Lander, G., L. Tang, S. Casjens, E. Gilcrease, P. Prevelige, A. Poliakov, C. Potter, B. Carragher, and J. Johnson. 2006. The structure of an infectious P22 virion shows the signal for headful DNA packaging. *Science* 312:1791–1795.

[77] Leiman, P. G., P. R. Chipman, V. A. Kostyuchenko, V. V. Mesyanzhinov, and M. G. Rossmann. 2004. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* 118:419–29.

[78] Lewis, M., G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 271:1247–1254.

[79] Lillian, T. D. 2010. Mechanics and function of DNA looping and supercoiling. Ph.D. thesis, University of Michigan.

[80] Lillian, T. D. and N. C. Perkins. 2011. Electrostatics and self-contact in an elastic rod approximation for DNA. *J. Comput. Nonlin. Dyn.* 6:011008.

[81] Lillian, T. D., M. Taranova, J. Wereszczynski, I. Andricioaei, and N. C. Perkins. 2011. A multiscale dynamic model of DNA supercoil relaxation by topoisomerase IB. *Biophys. J.* 100:2016–2023.

[82] Lillian, T. D., S. Goyal, J. D. Kahn, E. Meyhöfer, and N. C. Perkins. 2008. Computational analysis of looping of a large family of highly bent DNA by LacI. *Biophys. J.* 95:5832–5842.

[83] Liu, X., Q. Zhang, K. Murata, M. L. Baker, M. B. Sullivan, C. Fu, M. T. Dougherty, M. F. Schmid, M. S. Osburne, S. W. Chisholm, and W. Chiu. 2010. Structural changes in a marine podovirus associated with release of its genome into Prochlorococcus. *Nat. Struct. Mol. Biol.* 17:830–836.

[84] Löf, D., K. Schillén, B. Jönsson, and A. Evilevitch. 2007. Forces controlling the rate of DNA ejection from phage λ. *J. Mol. Biol.* 368:55–65.

[85] Lu, C.-L. 1994. Three-dimensional flexural and torsional mechanics of low and high tension cables. Ph.D. thesis, University of Michigan.

[86] Mangenot, S., M. Hochrein, J. Radler, and L. Letellier. 2005. Real-time imaging of DNA ejection from single phage particles. *Curr. Biol.* 15:430–435.

[87] Manning, R. S. and G. B. Bulman. 2005. Stability of an elastic rod buckling into a soft wall. *P. Roy. Soc. A-Math. Phy.* 461:2423–2450.

[88] Manning, R. S., J. H. Maddocks, and J. D. Kahn. 1996. A continuum rod model of sequence-dependent DNA structure. *J. Chem. Phys.* 105:5626–5646.

[89] Marenduzzo, D., E. Orlandini, A. Stasiak, D. W. Sumners, L. Tubiana, and C. Micheletti. 2009. DNA-DNA interactions in bacteriophage capsids are responsible for the observed DNA knotting. *P. Natl. Acad. Sci. USA.* 106:22269–22274.

[90] Meijer, W., J. Horcajadas, and M. Salas. 2001. ϕ29 family of phages. *Microbiol. Mol. Biol. R.* 65:261–287.

[91] Mitchell, J. S., C. A. Laughton, and S. A. Harris. 2011. Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA. *Nucleic Acids Res.* 39:3928–38.

[92] Moak, M. and I. J. Molineux. 2004. Peptidoglycan hydrolytic activities associated with bacteriophage virions. *Mol. Microbiol.* 51:1169–83.

[93] Moll, W. D. and P. X. Guo. 2005. Translocation of nicked but not gapped DNA by the packaging motor of bacteriophage phi29. *J. Mol. Biol.* 351:100–107.

[94] Müller, J., S. Oehler, and B. Müller-Hill. 1996. Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. *J. Mol. Biol.* 257:21–29.

[95] Nelson, P. C. 2012. Spare the (elastic) rod. *Science* 337:1045–6.

[96] Normanno, D., F. Vanzi, and F. S. Pavone. 2008. Single-molecule manipulation reveals supercoiling-dependent modulation of lac repressor-mediated DNA looping. *Nucleic Acids Res.* 36:2505–2513.

[97] Nurmemmedov, E., M. Castelnovo, C. E. Catalano, and A. Evilevitch. 2007. Biophysics of viral infectivity: matching genome length with capsid size. *Q. Rev. Biophys.* 40:327–56.

[98] Olia, A. S., J. Prevelige, Peter E., J. E. Johnson, and G. Cingolani. 2011. Three-dimensional structure of a viral genome-delivery portal vertex. *Nat. Struct. Mol. Biol.* 18:597–604.

[99] Olson, W. K. 1996. Simulating DNA at low resolution. *Curr. Opin. Struct. Biol.* 6:242–256.

[100] Olson, W. K., D. Swigon, and B. D. Coleman. 2004. Implications of the dependence of the elastic properties of DNA on nucleotide sequence. *Philos. T. Roy. Soc. A.* 362:1403–1422.

[101] Olson, W. K., M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X. -J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C. -S. Tung, E. Westhof, C. Wolberger, and H. M. Berman. 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* 313:229–237.

[102] Oram, M., C. Sabanayagam, and L. W. Black. 2008. Modulation of the packaging reaction of bacteriophage T4 terminase by DNA structure. *J. Mol. Biol.* 381:61–72.

[103] Orlova, E., B. Gowen, A. Droge, A. Stiege, F. Weise, R. Lurz, M. van Heel, and P. Tavares. 2003. Structure of a viral DNA gatekeeper at 10 Å resolution by cryo-electron microscopy. *EMBO J.* 22:1255–1262.

[104] Panja, D. and I. J. Molineux. 2010. Dynamics of bacteriophage genome ejection *in vitro* and *in vivo*. *Phys. Biol.* 7:1–15.

[105] Petrov, A. S. and S. C. Harvey. 2007. Structural and thermodynamic principles of viral packaging. *Structure* 15:21–27.

[106] Petrov, A. S. and S. C. Harvey. 2008. Packaging double-helical DNA into viral capsids: structures, forces, and energetics. *Biophys. J.* 95:497–502.

[107] Pettersen, E., T. Goddard, C. Huang, G. Couch, D. Greenblatt, E. Meng, and T. Ferrin. 2004. UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605–1612.

[108] Plisson, C., H. E. White, I. Auzat, A. Zafarani, C. Sao-Jose, S. Lhuillier, P. Tavares, and E. V. Orlova. 2007. Structure of bacteriophage SPP1 tail reveals trigger for DNA ejection. *EMBO J.* 26:3720–3728.

[109] Purohit, P., M. Inamdar, P. Grayson, T. Squires, J. Kondev, and R. Phillips. 2005. Forces during bacteriophage DNA packaging and ejection. *Biophys. J.* 88:851–866.

[110] Purohit, P. K., J. Kondev, and R. Phillips. 2003. Mechanics of DNA packaging in viruses. *P. Natl. Acad. Sci. USA.* 100:3173–8.

[111] Purohit, P. K. and P. C. Nelson. 2006. Effect of supercoiling on formation of protein-mediated DNA loops. *Phys. Rev. E.* 74:061907.

[112] Ray, K., C. R. Sabanayagam, J. R. Lakowicz, and L. W. Black. 2010. DNA crunching by a viral packaging motor: compression of a procapsid-portal stalled Y-DNA substrate. *Virology* 398:224–32.

[113] Rice, P., S. Yang, K. Mizuuchi, and H. Nash. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87:1295–1306.

[114] Rickgauer, J. P., D. N. Fuller, S. Grimes, P. J. Jardine, D. L. Anderson, and D. E. Smith. 2008. Portal motor velocity and internal force resisting viral DNA packaging in bacteriophage $\phi 29$. *Biophys. J.* 94:159–167.

[115] Rothemund, P. W. K. 2006. Folding DNA to create nanoscale shapes and patterns. *Nature* 440:297–302.

[116] Ruben, G. C. and T. B. Roos. 1997. Conformation of Lac repressor tetramer in solution, bound and unbound to operator DNA. *Microsc. Res. Tech.* 36:400–416.

[117] Rutkauskas, D., H. Zhan, K. S. Matthews, F. S. Pavone, and F. Vanzi. 2009. Tetramer opening in LacI-mediated DNA looping. *P. Natl. Acad. Sci. USA.* 106:16627–16632.

[118] Saiz, L. and J. M. G. Vilar. 2007. Multilevel deconstruction of the *in vivo* behavior of looped DNA-protein complexes. *PLoS ONE.* 2:e355.

[119] Schlick, T. 1995. Modeling superhelical DNA: Recent analytical and dynamic approaches. *Curr. Opin. Struct. Biol.* 5:245–262.

[120] Semsey, S., K. Virnik, and S. Adhya. 2005. A gamut of loops: meandering DNA. *Trends Biochem. Sci.* 30:334–341.

[121] Shimada, J. and H. Yamakawa. 1984. Ring-closure probabilities for twisted wormlike chains - application to DNA. *Macromolecules* 17:689–698.

[122] Simpson, A., Y. Tao, P. Leiman, M. Badasso, Y. He, P. Jardine, N. Olson, M. Morais, S. Grimes, D. Anderson, T. Baker, and M. Rossmann. 2000. Structure of the bacteriophage ϕ29 DNA packaging motor. *Nature* 408:745–750.

[123] Smith, D., S. Tans, S. Smith, S. Grimes, D. Anderson, and C. Bustamante. 2001. The bacteriophage ϕ29 portal motor can package DNA against a large internal force. *Nature* 413:748–752.

[124] Spakowitz, A., and Z. Wang. 2005. DNA packaging in bacteriophage: is twist important? *Biophys. J.* 88:3912–3923.

[125] Spiriti, J., H. Kamberaj, A. M. R. de Graff, M. F. Thorpe, and A. van der Vaart. 2012. DNA bending through large angles is aided by ionic screening. *J. Chem. Theory Comput.* 8:2145–2156.

[126] Strick, T., J.-F. Allemand, V. Croquette, and D. Bensimon. 2000. Twisting and stretching single DNA molecules. *Prog. Biophys. Mol. Biol.* 74:115–140.

[127] Swigon, D., B. D. Coleman, and W. K. Olson. 2006. Modeling the Lac repressor-operator assembly: the influence of DNA looping on Lac repressor conformation. *P. Natl. Acad. Sci. USA.* 103:9879–9884.

[128] Tang, J., G. C. Lander, A. Olia, R. Li, S. Casjens, P. Prevelige, Jr., G. Cingolani, T. S. Baker, and J. E. Johnson. 2011. Peering down the barrel of a bacteriophage portal: the genome packaging and release valve in P22. *Structure* 19:496–502.

[129] Tang, J., N. Olson, P. J. Jardine, S. Girimes, D. L. Anderson, and T. S. Baker. 2008. DNA poised for release in bacteriophage ϕ29. *Structure* 16:935–943.

[130] Tang, L., W. R. Marion, G. Cingolani, P. E. Prevelige, and J. E. Johnson. 2005. Three-dimensional structure of the bacteriophage P22 tail machine. *EMBO J.* 24:2087–95.

[131] Tao, Y., N. Olson, W. Xu, D. Anderson, M. Rossmann, and T. Baker. 1998. Assembly of a tailed bacterial virus and its genome release studied in three dimensions. *Cell* 95:431–437.

[132] Taraban, M., H. Zhan, A. E. Whitten, D. B. Langley, K. S. Matthews, L. Swint-Kruse, and J. Trewhella. 2008. Ligand-induced conformational changes and conformational dynamics in the solution structure of the lactose repressor protein. *J. Mol. Biol.* 376:466–481.

[133] Thompson, J. M. T., M. Silveira, G. H. M. van der Heijden, and M. Wiercigroch. 2012. Helical post-buckling of a rod in a cylinder: with applications to drill-strings. *P. Roy. Soc. A-Math. Phy.* 468:1591–1614.

[134] Timoshenko, S. P and J. M. Gere. 1961. Theory of elastic stability. New York: McGraw-Hill.

[135] Towles, K. B., J. F. Beausang, H. G. Garcia, R. Phillips, and P. C. Nelson. 2009. First-principles calculation of DNA looping in tethered particle experiments. *Phys. Biol.* 6:025001.

[136] Vafabakhsh, R., and T. Ha. 2012. Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science* 337:1097–101.

[137] van den Broek, B., F. Vanzi, D. Normanno, F. S. Pavone, and G. J. L. Wuite. 2006. Real-time observation of DNA looping dynamics of type IIE restriction enzymes NaeI and NarI. *Nucleic Acids Res.* 34:167–174.

[138] van der Heijden, G. H. M., M. A. Peletier, and R. Planque. 2006. Self-contact for rods on cylinders. *Arch. Ration Mech. An.* 182:471–511.

[139] van der Heijden, G. H. M. 2001. The static deformation of a twisted elastic rod constrained to lie on a cylinder. *P. Roy. Soc. A-Math. Phy.* 457:695–715.

[140] Vanzi, F., C. Broggio, L. Sacconi, and F. S. Pavone. 2006. Lac repressor hinge flexibility and DNA looping: single molecule kinetics by tethered particle motion. *Nucleic Acids Res.* 34:3409–3420.

[141] Villa, E., A. Balaeff, L. Mahadevan, and K. Schulten. 2004. Multiscale method for simulating protein-DNA complexes. *Multiscale Model. Sim.* 2:527–553.

[142] Villa, E., A. Balaeff, and K. Schulten. 2005. Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. *P. Natl. Acad. Sci. USA*. 102:6783–6788.

[143] Virnik, K., Y. Lyubchenko, M. Karymov, P. Dahlgren, M. Tolstorukov, S. Semsey, V. Zhurkin, and S. Adhya. 2003. "Antiparallel" DNA loop in gal repressosome visualized by atomic force microscopy. *J. Mol. Biol.* 334:53–63.

[144] Vlahovicek, K., L. Kaján, and S. Pongor. 2003. DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucleic Acids Res.* 31:3686–7.

[145] Vologodskii, A. and N. Cozzarelli. 1995. Modeling of long-range electrostatic interactions in DNA. *Biopolymers* 35:289–296.

[146] Watson, J. D. and F. H. C. Crick. 1953. Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature* 171:737–738.

[147] Wereszczynski, J. and I. Andricioaei. 2006. On structural transitions, thermodynamic equilibrium, and the phase diagram of DNA and RNA duplexes under torque and tension. *P. Natl. Acad. Sci. USA*. 103:16200–16205.

[148] Wicks, N., B. L. Wardle, and D. Pafitis. 2008. Horizontal cylinder-in-cylinder buckling under compression and torsion: review and application to composite drill pipe. *Int. J. Mech. Sci.* 50:538 – 549.

[149] Wiggins, P. A., T. Van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson. 2006. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat. Nanotechnol.* 1:137–141.

[150] Wilson, D. P., T. Lillian, S. Goyal, A. V. Tkachenko, N. C. Perkins, and J.-C. Meiners. 2007. Understanding the role of thermal fluctuations in DNA looping. *In* Proc. SPIE, S. M. Bezrukov, editor, volume 6602. SPIE, 60208.

[151] Wilson, D. P., A. V. Tkachenko, and J. -C. Meiners. 2010. A generalized theory of DNA looping and cyclization. *EPL-Europhys. Lett.* 89:58005.

[152] Wong, O. K., M. Guthold, D. A. Erie, and J. Gelles. 2008. Interconvertible Lac repressor-DNA loops revealed by single-molecule experiments. *PLoS Biol.* 6:e232.

[153] Wu, D., D. Van Valen, Q. Hu, and R. Phillips. 2010. Ion-dependent dynamics of DNA ejections for bacteriophage λ. *Biophys. J.* 99:1101–1109.

[154] Xiang, Y., M. C. Morais, A. J. Battisti, S. Grimes, P. J. Jardine, D. L. Anderson, and M. G. Rossmann. 2006. Structural changes of bacteriophage ϕ29 upon DNA packaging and release. *EMBO J.* 25:5229–5239.

[155] Xiang, Y., M. C. Morais, D. N. Cohen, V. D. Bowman, D. L. Anderson, and M. G. Rossmann. 2008. Crystal and cryoEM structural studies of a cell wall degrading enzyme in the bacteriophage ϕ29 tail. *P. Natl. Acad. Sci. USA.* 105:9552–7.

[156] Yu, J., J. Moffitt, C. L. Hetherington, C. Bustamante, and G. Oster. 2010. Mechanochemistry of a viral DNA packaging motor. *J. Mol. Biol.* 400.

[157] Zhang, Y., A. E. McEwen, D. M. Crothers, and S. D. Levene. 2006. Statistical-mechanical theory of DNA looping. *Biophys. J.* 90:1903–1912.

[158] Zhang, Y., A. E. McEwen, D. M. Crothers, and S. D. Levene. 2006. Analysis of *in-vivo* LacR-mediated gene repression based on the mechanics of DNA looping. *PLoS ONE.* 1:e136.

[159] Zhang, Y. and D. M. Crothers. 2003. Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization. *Biophys. J.* 84:136–153.

[160] Zheng, H., A. S. Olia, M. Gonen, S. Andrews, G. Cingolani, and T. Gonen. 2008. A conformational switch in bacteriophage P22 portal protein primes genome injection. *Mol. Cell* 29:376–383.

[161] Zhu, Q., J. Zeng, M. Triantafyllou, and D. Yue. 2006. Direct numerical simulation of single-molecule DNA by cable dynamics. *J. Microelectromech. S.* 15:1078–1087.