COMPUTATIONAL MODELING OF DNA ELASTIC ENERGY TO PREDICT STRUCTURE AND TOPOLOGY OF PROTEIN MEDIATED DNA LOOPS

By

PAMELA JOAN PEREZ

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Computational Biology and Molecular Biophysics

Written under the direction of Dr. Wilma K. Olson

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October 2017

ABSTRACT OF THE DISSERTATION

COMPUTATIONAL MODELING OF DNA ELASTIC ENERGY TO PREDICT STRUCTURE AND TOPOLOGY OF PROTEIN MEDIATED DNA LOOPS

by Pamela Joan Perez

Dissertation Director:

Dr. Wilma K. Olson

In addition to the genetic message, DNA base sequence carries a multitude of structural and energetic signals important to the packaging and processing of the genetic material. One way in which these signals enter is through the looping of DNA, mediated by proteins that attach to specific, widely separated base-pair elements along the chain molecule. Here I explore the influence of local sequence-dependent features of DNA on the ease of looping between the binding headpieces of the Lac repressor protein. I then consider the role that conformational flexibility of the Lac repressor plays on the conformation of the intervening DNA. I also provide insight into genome architecture by modeling nucleoprotein systems of protein partitioned-minicircles with two topologically independent domains. I identify the energetically preferred spatial pathways of short, protein-anchored fragments of ideal DNA and show that the energies capture the looping propensities and modes of chain attachment found by direct computer sampling. I examine the effects of the helical repeat, mode/range of local deformations, and intrinsic curvature on overall energy and chain configuration. I discuss the findings in the context of the effects of nucleotide sequence seen in recent studies of Lac repressor-mediated loops, including the looping topologies dictated by the settings of a naturally curved DNA insert and the looping propensities of A·T- vs. G·C-rich DNA. I describe the effects of fluctuations in protein conformation on looping likelihood.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Wilma K. Olson. Thank you for your support, encouragement and especially patience. You helped me establish the direction of my thesis, answered many of my questions but more importantly, guided me to answer many of my own questions.

Thank you also to my dissertation committee Dr. David A. Case, Dr. Darrin York, Dr. Alexander M. Schliep for the time and input for my research and to Dr. Alexandre V. Morozov for graciously agreeing to join my committee when Dr. Schliep moved on.

Thank you to the members of the Olson group both past and present who made my time as a graduate student at Rutgers enjoyable and inspiring. Thank you Andrew Colasanti for showing the ropes on being a graduate student, and Stefjord Todolli for rescuing me with the use of your laptop when mine failed.

To my colleagues at the BioMaPS Institute, especially Liyang, Iwen, Guatam, Michael, Ken, Tahir and Robb whose friendship I will forever cherish.

To my children Gabriela, Geoffrey, David and Marissa; I can't express enough how honored I am to be your mother. You supported me and sacrificed for me and I love you for it.

To my amazing husband Daniel; You believed in me, encouraged me, put up with me, picked up the slack for me. The achievement is as much yours as it is mine.

Dedication

To my husband, *Daniel*, your love, support and sacrifice made it possible for me to complete this work. You believed in me, encouraged me and put up with me. You truly are the wind beneath my wings.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction and Background

The Human Genome project succeeded in identifying just under 20 thousand genes in the human genome. It was believed that knowledge of sequence would be the key to treating or curing most human disease. However, it turns out that there are many fewer genes than were originally anticipated. Less than 2% of our DNA codes for proteins and if you consider non-coding RNA and other regulatory elements it still only accounts for ~8% of our DNA.

Mutations in genes are the most obvious cause of disease. However, variation in the DNA sequence in non-coding regions plays a less obvious role. The resultant structural differences don't change the proteins produced by the gene but may instead change the way the gene behaves.

We are only just beginning to understand what role structure and topology play in cellular processes like DNA packing and gene expression. What we do know is that DNA is a very long, thin, moderately stiff polymer. It not only must be compacted into a very small volume in the nucleus of the cell, it must also remain accessible to participate in cellular processes. We know that it is not a passive molecule waiting to be grabbed by a protein. Rather it is a dynamic molecule twisting and looping to adopt structural motifs that facilitate replication, repair and modulate gene expression. The formation of loops by constraining distant sites on the DNA is one of the tricks employed. It is the aim of this thesis to advance our understanding of the factors that govern the dynamics of DNA by modeling the structural properties of looped DNA.

1.1 DNA Looping

DNA loops are ubiquitous in cellular processes as a regulatory tool and as a genome organizational mechanism. Many crucial processes including transcription, replication, and recombination, employ DNA looping as an essential component.[1] DNA looping allows distant sites on the genome to come in close proximity and facilitates long-range communication. In some examples of gene expression, looping is a requirement for the initiation of transcription. In other cases, DNA loops enhance the regulation properties of some processes, either increasing repression levels or amplifying expression. DNA loops can act to partition the genome into topologically independent domains.[2] The ability of DNA to form loops is affected by the length and the sequence of the DNA molecule as well as by mediation through protein interactions. Some proteins or protein complexes can bind to two spatially distant binding regions on the DNA molecule, enforcing constraints on the ends of the intervening strand and thus forming a loop. The creation of longer DNA loops is mitigated by the movement of the DNA within the cell such that the ends of the loop entropically come into close enough proximity to form the protein-DNA loop complex. In contrast, the formation of shorter DNA loops, <150 base pairs, require severe deformation of the otherwise stiff, straight double-helical molecule. The bending deformation necessary for loop formation may be provided by additional architectural proteins, which bind within the loop.

1.2 The Lac Operon

DNA looping mediated by the Lac repressor protein is a classic exemplar for understanding protein-DNA interactions and allosteric mechanisms in gene regulation. The *lac* operon is a collection of three linked genes under common coordinate control.[3–5] It codes for three proteins involved in the metabolism of lactose. A fourth gene, located upstream from the metabolizing protein genes, codes for the regulatory Lac repressor protein (LacR). In order to effectively regulate the production of these proteins the *lac* operon uses a two-part

coordinated mechanism whereby transcription only occurs in the presence of lactose and the absence of the preferred energy source glucose.

Two key components of this system involved in transcription repression in the absence of lactose are the Lac repressor protein LacR and the DNA loop generated by the binding of the protein to two sequentially distant positions on the DNA. In the presence of allolactose, a metabolite that signals the presence of lactose, the LacR protein undergoes a conformational change that reduces its DNA binding affinity. Without the LacR-bound loop transcription of the genes necessary for lactose metabolism is induced.

Structurally the LacR protein complex is a homo-tetramer composed of four identical polypeptide chains (Figure 1.1). Each chain includes a globular DNA binding domain or headpiece joined by a helix-hinge linker to the two globular sub domains of the core or regulatory domain and ending with an α-helix tetramerization region. Each of these chains partners with another chain to form one of the two dimeric "arms". An α-helix bundle formed by the coiling of helices in the tetramerization region of each monomer joins the two arms into the complete V-shaped protein assembly.

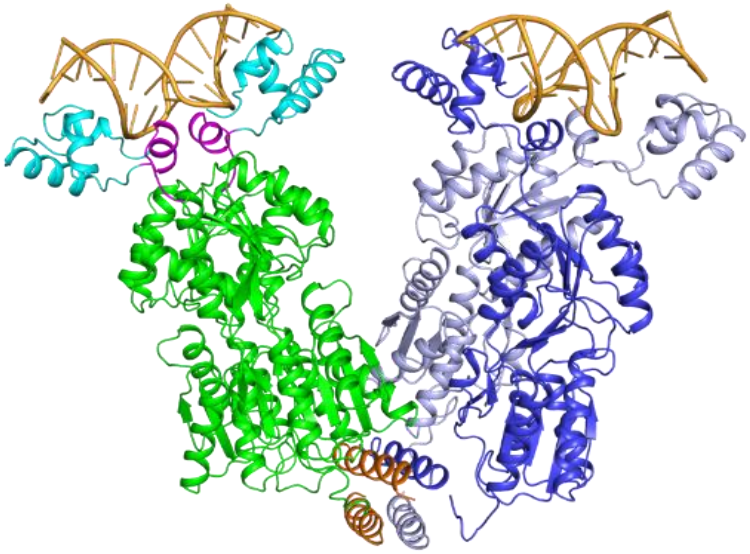

Figure 1.1 LacR in complex with DNA: The right arm shows each of the monomeric chains, one in light blue and the other in dark blue. The left arm indicates the tetramerization region in orange, the regulatory domain in green and the DNA binding domain in cyan with the helix hinge linker in magenta

The DNA loop is formed when the repressor protein binds to the primary operator site O_1 and one of two secondary operators O_2 or O_3.[4] The O_1 site is located within the region of the promoter of the genes that code for the proteins involved in lactose metabolism. The O_2 and O_3 sites are located 92 base pairs (bp) upstream and 401 bp downstream, respectively, from the primary operator. In the absence of cellular lactose, the O_1 operator binds to one arm of the protein. A more stable binding and greater repression is achieved when one of the secondary sites is bound to the other arm of the repressor protein. This coordinated binding necessitates a tightly bound DNA loop. Each operator can bind in one of two orientations, with the 5' direction of the DNA pointing either toward or away from the center of the protein, giving way to four possible loop types A1, A2, P1, P2 (Figure 1.2).[6–8] Loops where the

operators are bound in approximately the same orientation or in opposite orientations are referred to as “parallel” or “anti-parallel”, respectively.[9]

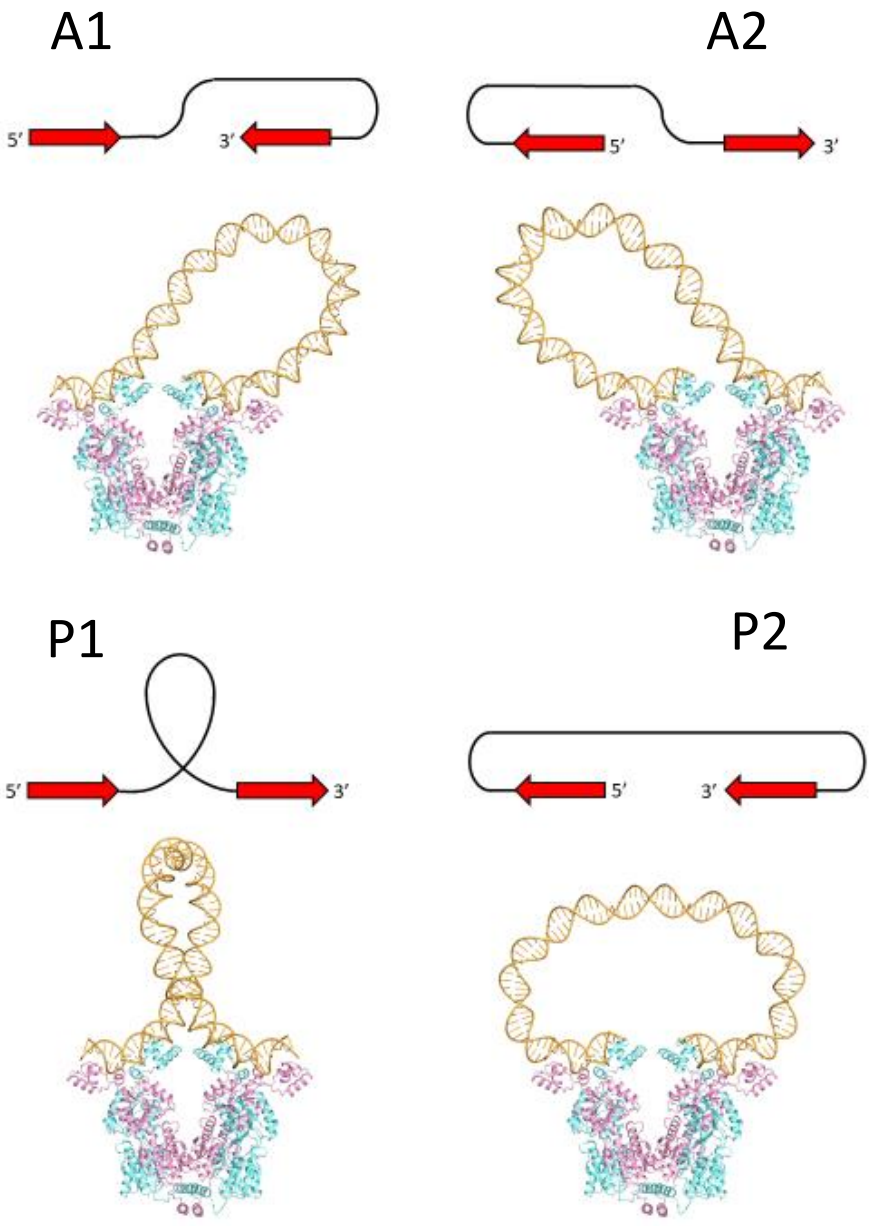

Figure 1.2 Four DNA loop orientations

The control of loop formation is mitigated by the presence of allolactose, a byproduct of lactose metabolism. It binds to LacR in a pocket between the two globular sub-domains of

the core region of the protein, allosterically effecting a conformational change in the binding headpiece, releasing the bound DNA, and inducing transcription. Although the *lac* operon is one of the most heavily studied gene regulatory systems very little is known about the actual mechanism of the DNA-protein interactions involved. Likewise, the structure and dynamics of the DNA loops formed are also poorly understood.

From the time Jacob and Monod originally proposed the *lac* operon model,[10] we have continued to increase our understanding of the *lac* operon and its components. Early microcrystal structure studies of the Lac repressor protein suggested that the arms of the protein can be oriented in an extended conformation[11]. That is, the angle between the two arms is nearly 180° and the headpieces are not in close proximity to one another. Later small-angle X-ray scattering (SAXS) studies supported this view,[12] Crystal structures of the entire protein and bound operator, however, indicate the molecule adopts a more rigid V-shaped closed configuration with very little variability in the opening angle. The crystal structures also show a small molecular interface between the two arms suggesting a low energy cost to opening the arms of the LacR. The distinct conformations identified support a two-state model of protein architecture. Transmission electron microscopy (TEM) images indicate a roughly 60:40 mix of open and closed proteins,[13] again further supported by SAXS studies of protein-DNA complexes.[12] Förster resonance energy transfer (FRET), footprinting and gel studies lend further support to a two-state model.[13–17] Nuclear magnetic resonance (NMR) studies have revealed that the headpiece only appears when the protein is bound to the operator; otherwise that region is disordered.[18–20] The vast amount of single-molecule imaging studies have provided confirmation of several regions of LacR flexibility. The helix-hinge linker, which binds the DNA binding headpiece between the two globular sub-domains, and a rotation about the four-helix bundle are implicated in protein deformation.

Additional characteristics of DNA looping are provided by experiments like tethered particle motion (TPM) studies, FRET, electron microscopy (EM) and atomic force microscopy (AFM). TPM studies elucidate the lifetime of loops and when the protein is in an extended configuration or adopts a closed "V" configuration and the DNA loops in different orientations.[7,8,21–23] TPM experiments also suggest that there may be two types of A1 loops, those that wrap around the protein and those that wrap away from the protein. FRET studies have added information about helical phasing of curved inserts. AFM images provide clues to other features that influence the formation of DNA loops such as DNA length, helical phasing of the operator sites and other sequence-dependent features of the DNA. Gene expression studies indicate how the presence of architectural proteins like HU and ligands like IPTG, Isopropel β-D-1-thiogalactopyranoside, a molecular mimic of allolactose, affect DNA looping. Unlike allolactose, IPTG is not broken down, thus its concentration remains constant making it a useful inducer in experimental systems. Finally, our knowledge base is complimented by numerous simulation and modeling methods, which continue to expand our understanding of the role that DNA looping plays in gene regulation in the *lac* operon.[24–26]

1.3 Overview of Thesis

To describe precise conformations of protein-bound DNA loops like those created in the *lac* operon we must employ computational models. In the following chapters I will present the computational methods I used to model short energetic DNA loops and the protein that ties them. In Chapter 2, I describe the components of the model used to identify the energetically preferred configuration of protein-constrained DNA loops as well as background on experiments used for comparison. The remaining chapters present applications of this model to investigate features of DNA loops. Chapter 3 describes modeling the DNA sequence-dependent properties of protein-DNA loops. Chapter 4 presents modeled

alternate DNA loop-end constraints resulting from simulated protein deformation. Chapter 5, which was previously published[27], discusses the implications of sequence length and protein deformation on topological domains of protein segmented minicircles.

Chapter 2 Methods

Because the *lac* operon is such a well-studied system there is a plethora of experimental data to facilitate the development and validation of computational models to describe the structure and behavior of protein bound DNA. These models not only predict characteristics but also may provide an understanding of structure that is not available from experiment. There are limitations to extracting details from experimental data. Some techniques (X-ray, EM, NMR) make it possible to get very a detailed atomic-level picture of small segments of DNA but provide little insight into details of overall structure. Other techniques, (TPM, FRET, expression studies) can provide details of the existence and some measure of the stability of DNA loops. Still other experimental techniques like AFM and SAXS provide coarse two dimensional information about DNA loops. At this time there are no available experimental methods to describe high-resolution details about the spatial features of even a relatively small DNA loop. We are forced to compromise between a detailed high-resolution picture of small segments of DNA and an imprecise description of larger multi-component protein-DNA constructs.

2.1 Model Components*:*

2.1.1 Model Protein

The Lac repressor assembly is represented by a rigid, V-shaped model, constructed by superposition of the structures of two well-resolved macromolecular complexes[9,25] specifically the 2.6-Å resolution structure of the Lac protein dimer bound to a self-complimentary DNA operator[4] and the 2.7-Å resolution structure of the tetrameric form of the protein without DNA-binding headpieces.[28] The modeled protein-bound DNA is assigned the rigid-body parameters adopted in the former complex and the full repressor is included as a 'side group' of the DNA, *i.e.*, the atoms of protein are expressed in the reference frame of one of the bases in the full assembly. The reported loop lengths correspond to the number of

base-pair steps between the centers of the bound DNA operators, *i.e.*, seven of the 13 steps attached to each arm of the modeled repressor-operator assembly plus the DNA steps subjected to configurational changes. Thus, the reported lengths of the loops are 14 bp steps longer than the lengths of the optimized, ligand-free DNA segments, corresponding to the conventional operator center-to-center measure of DNA loop length.

In the absence of knowledge of the directions in which the DNA attaches to the arms of the repressor, each operator is placed in two orientations on the protein-binding headpieces and four different types of loops are generated.[29] I.e., antiparallel (A1, A2) and parallel (P1, P2) loops discussed above (Figure 1.2). The numerals 1 and 2 in this shorthand distinguish whether the initial direction of the loop points respectively toward the inside or outside of the repressor assembly[30].

2.1.2 DNA Model

Relatively short DNA loops like the 92 bp $O_1 - O_3$ bound loop in the *lac* operon require extreme distortion from the naturally straight, moderately stiff polymer. By evaluating the energy required to deform the DNA into these tightly bent loops we can determine the likely spatial pathways of the associated looped DNA.

The configuration of DNA loops is described by a collection of DNA base pairs and monitored by the displacement and orientation of rigid body reference frames embedded on each pair. The spatial relationship between successive base pairs or base pair steps is then described by the 'step' parameters that specify the translation and rotation from one base pair to the next along each of the three Cartesian coordinate axes. The parameters $(\theta_1^i, \theta_2^i, \theta_3^i)$ also called tilt, roll, twist, describe the angular rotation from one base pair to the next and the parameters $(\rho_1^i, \rho_2^i, \rho_3^i)$, or shift, slide, rise, describe the displacement (Figure 2.1). The parameters for the i-th step, denoted $\underline{p}^i = (\theta_1^i, \theta_2^i, \theta_3^i, \rho_1^i, \rho_2^i, \rho_3^i)$, specify the location of base

pair *i+1* with respect to base pair *i*. The entire strand can thus be described by the set of the step parameters for all of the steps.

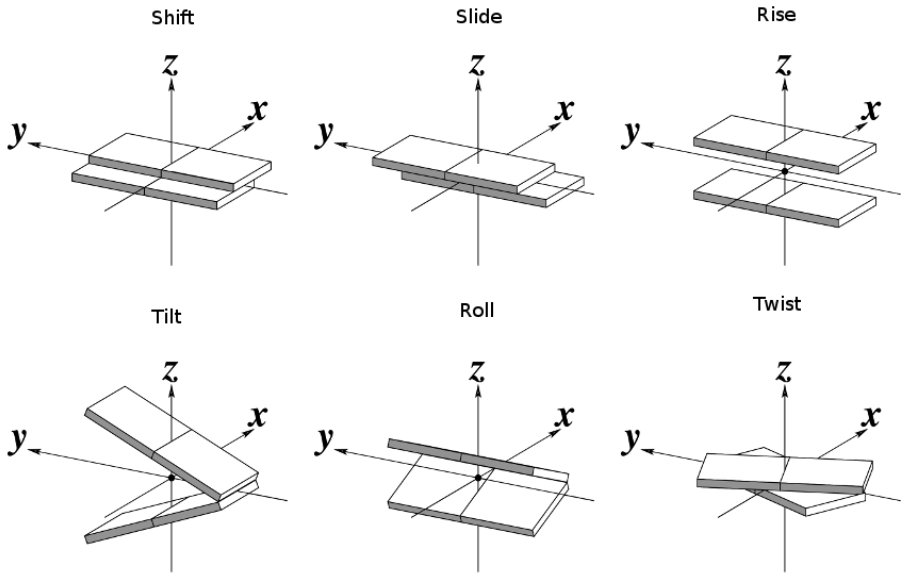

Figure 2.1 Base Pair Step Parameters

The energy of the DNA loop is described relative to that of canonical B-DNA with base pairs perfectly parallel to one another and symmetrically rotated about the perpendicular to the two base-pair planes to form a double helix.

The canonical B-DNA model is thus described by zero tilt, roll, shift and slide, a twist of about 36° and a rise of 3.4Å. X-ray structures show a range of variation in the twist with an average of about 34.29° or 10.5 bp/turn. In this model, referred to as ideal DNA throughout this dissertation, the parameters for the relaxed state of the naturally straight DNA are assumed thus $\underline{p}^0 = (0, 0, 34.29, 0, 0, 3.4)$. The treatment allows for other models of DNA elasticity which are described in Chapter 3.

The global twisting of the looped DNA is described in terms of a discrete ribbon constructed from the origins and reference frames of four successive base pairs.[31,32] In

contrast to the twist angle included in the six rigid-body parameters used to specify the relative spatial arrangements of successive base pairs (see below), the values of twist reported here, the so-called twist of supercoiling,[31] can be combined with the writhing number of a closed structure to obtain the correct linking number, an integer if the edges of the DNA ribbon are connected and the surface bounded by the ribbon is free of self-contact[33]. The writhing number is a standard measure of the global configuration of a closed space curve and the total twist quantifies the rotation of the DNA ribbon about its centerline. Note that the 'closed' DNA considered here includes both the constrained loop and a virtual connection through the protein assembly, here taken, for simplicity, as the vector connecting the centers of terminal base pairs (located respectively at the 3′- and 5′-ends of the constrained DNA segment). The torsionally relaxed state corresponds to the loop where the overall twist deformation is minimal.

2.1.3 Optimization

The foundation of my research relies on a novel optimization method which allows the optimization of the potential energy of deformation for the collection of base pairs described. The energy of deformation for a single step is measured according to the following equation, where $\Delta p_i = p_i - p_i^0$.

$$\psi = \frac{k_b T}{2} \sum_{i=1}^{6} \sum_{j=1}^{6} F_{ij} \Delta p_i \Delta p_j \tag{2.1}$$

The potential energy for the n^{th} step can be expressed by the quadratic

$$\psi^n = \frac{1}{2} (\Delta \underline{p}^n)^T F^n \, (\Delta \underline{p}^n) \tag{2.2}$$

$\Delta\underline{p}^n$is the deviation of the step parameter vector $\underline{p}^n$ from $\underline{p}^0$, the intrinsic step parameter vector that describes the step parameters of a base pair step of zero energy or the rest state, $\Delta\underline{p}^n = \underline{p}^n - \underline{p}^0$, and F^n is a 6x6 matrix containing the elastic moduli associated with each mode of deformation. Thus the total elastic energy of deformation is determined as

$$\Psi = \sum_{n=1}^{N} \psi^n \tag{2.3}$$

To determine the optimal configuration of a protein-bound DNA loop we must maintain the arrangement of the first and last base pairs and optimize the energy function Ψ for the remaining bases. We are thus presented with a complex constrained optimization problem.

The approach employed introduces an alternate representation of the base pair step parameters as $\underline{\phi}^n = \left(\theta_1^n, \theta_2^n, \theta_3^n, r_1^n, r_2^n, r_3^{in}\right)$. The angular components are the same as those described above; however the displacement parameters (r_1^n, r_2^n, r_3^n) are now defined relative to a global reference frame and the energy function is updated to reflect this. Now to describe the bound ends of the DNA loop we only need to ensure that the end-to-end vector through the global displacement parameters does not change nor does the total rotation, thus converting the problem from a constrained optimization to a more robust unconstrained optimization problem. The equation is then minimized using numerical optimization methods.

By treating each step as six independent degrees of freedom, a sequence of *N* steps can be modeled with 6*N* degrees of freedom, referred to as step degrees of freedom. In addition, the imposed end-to-end vector and end-to-end rotational constraints can be expressed such that the step degrees of freedom for the last base-pair step are functions of

the degrees of freedom of all the other steps. In other words, we reduce the dimensionality of the problem by taking into account the boundary conditions thus the dimension is reduced from $6N$ to $6(N-1)$. The gradient of the potential energy of elastic deformation for the collection of base pairs and the Hessian matrix are also obtained analytically.

The approach differs from the optimization of DNA elastic energy used in earlier work[34], which requires explicit specification of the forces and moments acting on the constrained base pairs (including an educated first guess of these values), or methods used by others[35] that take the boundary conditions into account through Lagrange multipliers. The optimization allows us to investigate the contributions of protein and DNA to loop formation at levels unattainable with random sampling techniques. Thus we can investigate rare supercoiled states and precise anchoring conditions, as opposed to the approximately closed structures of arbitrary topology and substantially higher energy captured with Monte Carlo methods[36].

2.2 Force Field Components

2.2.1 Intrinsic Step Parameters

The ideal DNA model (Table B-1) resembles the Shimada-Yamakawa twisted wormlike chain representation of DNA[37] which allows for fluctuations of the angular variables from the rest state, $(\theta_1^0, \theta_2^0, \theta_3^0) = (0,0,360/10.5)$, but fixes the translational parameters at values characteristic of B DNA, *i.e.*, $(\rho_1^0, \rho_2^0, \rho_3^0) = (0,0,3.4\text{Å})$. That is, a large force constant is introduced to put a very high cost to being more or less than 3.4 Å from the previous base pair and not allowing a base pair to shift or slide.

2.2.1 Force Constants

The values assigned to the angular force constants are determined by the known persistence length of mixed sequence DNA and the standard relationship between the bending and twisting constants of DNA. The tilt and roll force constants, F_{11} and F_{22}, can be

combined to give the bending component $\boldsymbol{f}$, which is related to the persistence length a and the base-pair spacing Δs through the expression $\frac{a}{\Delta s}$ i.e., $\boldsymbol{f} = F_{11} = F_{22} = \frac{\alpha}{\Delta s}$. When a is set to 476 Å (the observed value at ambient salt conditions[38]), Δs to 3.4 Å (the characteristic spacing of DNA base pairs) as in canonical B-DNA F_{11} and F_{22}, expressed in units of degrees, are equal to $4.84^{-2} deg^{-2}$. In other words, deformations of tilt or roll of 4.84° raises the energy $\frac{1}{2}\kappa T$ The value assigned to F_{33} is based on the standard relationship between bending and twisting constants of an elastic rod such that $F_{33} = 2\boldsymbol{f}/\gamma$. The value of γ=1.4 is compatible with measures of equilibrium topoisomer distributions of DNA minicircles[39] and the fluorescence depolarization of anisotropy of ethidium bromide intercalated in mincircles[40]. The result is $F_{33} = 4.05^{-2} deg^{-2}$. (Table B-1) Simply put, DNA is less likely to twist than bend.

2.3 Background on Comparison Data*:*

The validation of these models is assessed by comparison to results of several experimental and computational techniques. These include the J factor measurements based on the FRET experiments[16], TPM experiments[8,41,42] and the predicted ratio of J obtained in Monte Carlo simulations[43]. These studies are described in further detail below.

2.3.1 J Factor

The J-factor, a conventional measure of DNA flexibility, is used in much of this research to correlate the results of my optimizations to other techniques. The Jacobson-Stockmayer theory[44] is commonly used to describe ring closure probabilities from cyclization experiments involving linear double-stranded DNA flanked by complementary single-stranded 'sticky ends'. The ligation of the sticky ends produces either *(i)* a closed ring, when the ends of the same monomer are joined or *(ii)* a dimer when the ends of different monomers are joined. The cyclization equilibrium constant, or J-factor, is the ratio of the equilibrium constants for the cyclization reaction to the dimer reaction and describes DNA flexibility. The use of the J factor has been extended to Monte Carlo simulations of cyclical DNA as the ratio

of DNA configurations whose ends coincide to the entire sample size.[45] The theory can thus be further extended to Monte Carlo simulations of constrained DNA loops where J can be thought of as the effective concentration of DNA chain ends in the vicinity of the end conditions imposed on the loop.

2.3.2 FRET Experiments

Much of the results of my optimization is validated by comparison to experimental results from the J. Kahn lab. Experiments done in the Kahn lab provide insight into the role of sequence-dependent curvature on loop formation. The experiments describe the type of Lac repressor loops formed by 148-164 bp sequences all of which contain an 80 bp naturally curved sequence inserted between the two operator sequences. Fluorescent dyes are placed in different positions flanking the two operator DNA locations with a donor fluorophore on the first operator and an acceptor fluorophore on second. When both operators of the DNA are bound to the Lac repressor to form a loop the fluorophores are in close enough proximity such that the donor excites the acceptor and a FRET signal is detected. The fluorophores' sensitivity to distance makes FRET a valuable technique for determining structural characteristics of molecules. The location of the dyes, either two base pairs before or two base pairs after the operator sequence, results in four fluorophore position variants (FPV) designed to distinguish different types of loops. Like the DNA loop, the two variants (internal or external) of each of two operators (donor or acceptor) yield four combinations of donor acceptor configurations (D_IA_I, D_IA_E, D_EA_I, D_EA_E) designed to distinguish (P1, A1, A2, P2) loops, respectively. The nomenclature internal (I) and external (E) respectively refers to the link positioned within or outside the loop constrained by the operators. If the rigid V-shape protein is assumed the fluorophores will be in close enough proximity such that at FRET signal can be measured for each of the loop types in at least one FPV.

Crothers described intrinsically curved repeated 5-6 base pair stretches of consecutive A·T base pairs(so called A-tract),[46] which when included in a DNA minicircle stabilize the minicircle by reducing the free energy cost for bringing the operators together. This is the sequence that is included in the Kahn sequences. The phasing of the inserted sequence should alter the direction of curvature; either it will "potentiate or prevent" the formation of different loop configurations. That is, if the inserted curvature is consistent with an ideal B-DNA loop of a particular type it will facilitate the formation of that loop topology. By shifting the curved segment along the double helix, the direction of the curvature will now differ from the ideal B-DNA loop topology where the ends of the DNA fragment are in perfect register to the binding headpieces of the Lac repressor. The location of the location of the curved insert within the DNA may hinder the formation of that loop. The shifting of the curved sequence is achieved by varying the length of adapter sequences between the operators and the 80 bp curved A-tract segment. The Kahn group chose to vary the adaptor between the first operator and the curved segment between 5-13 base pairs in 2 bp increments. The second adaptor, after the curved segment, was varied by 10-18 base pairs, also in 2 bp increments, leading to 25 variations in the sequence. The combinations, denoted by yCx, correspond to constructs where y is the length of the first adaptor and x is the length of the second adaptor. This variety provides data on the effects of the location of the curved insert independently from the effects of the length of the sequence. For example; the 7C16, 9C14 and 11C12 have the same overall length but the curved insert is successively relocated 2 bps away from the previous insert, in effect rotating the direction of the curvature (Figure 2.2, Appendix A)

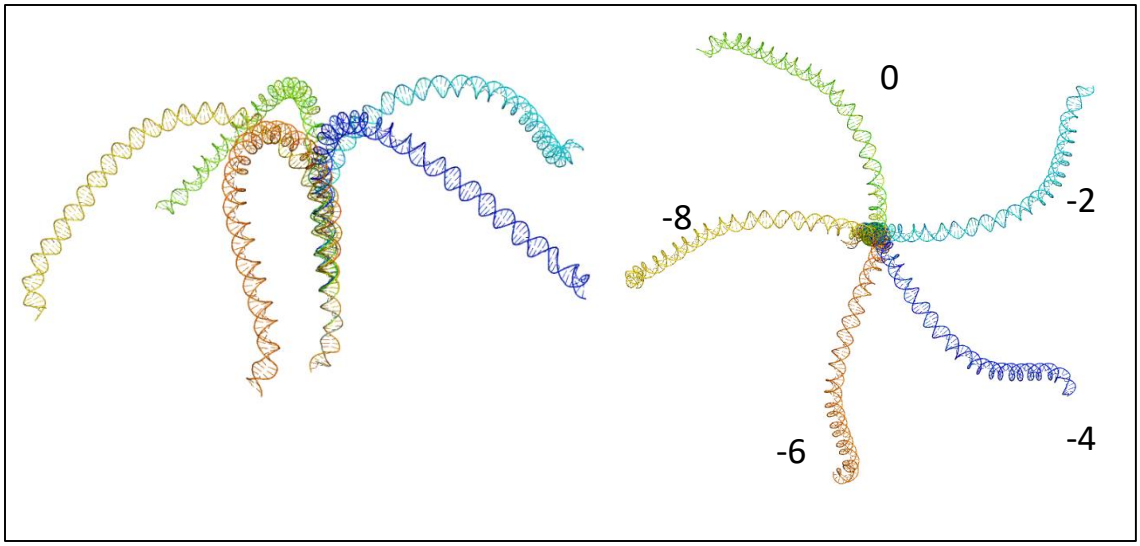

Figure 2.2 Molecular images illustrating the influence of the placement of a common, quasi planar 80 bp curved insert on the unconstrained, equilibrium rest states of the designed DNA constructs listed in Appendix A. Models superimposed in a common reference frame on the 5′-terminal base pair and viewed from the side (left) and top (right) to illustrate the common curvature and the differing rotational settings of the constructs. The structures are color-coded clockwise from green to yellow, corresponding to the displacement of the insert, at –2 bp increments, from the midpoint (denoted by 0) toward the 5' end of the loop.

2.3.3 Monte Carlo Results

Monte Carlo simulations done by Czapla et. al[43]. generated ensembles of configurational states satisfying a Boltzmann distribution subject to the same potential energy function used in our ideal B-DNA. Likewise, the rigid, V-shaped representation of the Lac repressor assembly used in these simulations is the same as that described in our model above.

A sampled chain is accepted as a closed loop if the end conditions coincide approximately with the end conditions of the modeled protein DNA complex. The assessment of loop formation is determined by aligning the first base pair of the sample with one end of the model operator. A phantom base pair N+1 is added with a reference frame of the end of the second operator. The step pair parameters between the last base pair of the sample N and the phantom base pair are determined. If the two base pairs coincide exactly all six step parameters would be zero. An acceptable approximation for coincidence allows for variation

in three components of the phantom step (Figure 2.3); (i) the magnitude of the vector $|r_{n:n-1}| < r_\varepsilon = 15$ Å, (ii) $\cos(\gamma) \geq 1 - \nu_\varepsilon$ where γ is the angle between the normals of the two base pairs and $\nu_\varepsilon = 0.02$ and (iii) $\cos(\theta) \geq 1 - \tau_\varepsilon = 0.98$, where θ is the twist angle between the two base pairs and $\tau_\varepsilon = 0.02$.

The results of the simulations are presented in terms of J factors determined by:

$$J = \frac{M_c}{QM} \tag{2.4}$$

Here M_c is the number of closed loop samples, M is the total number of samples and is normalized by the volume of the enclosed phase space with $Q = 4\pi N_A {r_\varepsilon}^3 \nu_\varepsilon \tau_e / 3$

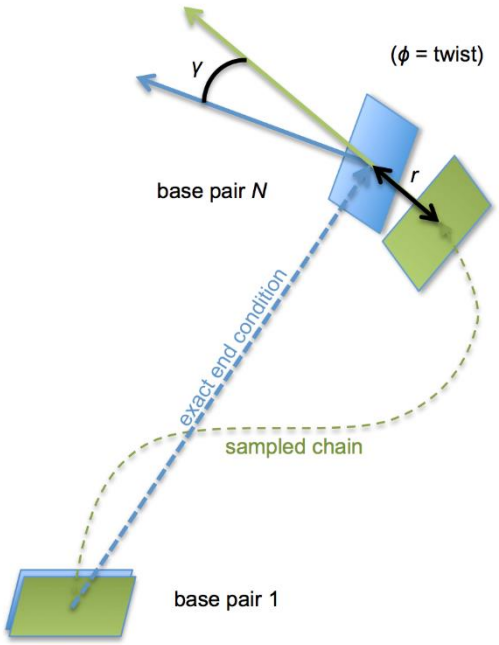

Figure 2.3 Schematic of the geometric constraints used to determine whether a linear DNA segment meets a specific end-to-end arrangement. Here a sampled chain of N base pairs (green blocks) adopts a configuration that approaches the desired geometry (blue blocks). The end-to-end vector r (thick black arrow) joins the Nth base pair of the simulated chain to that in the perfectly configured chain. Precise chain alignment requires that the net bend angle γ between base normals, the end-to-end Twist φ (defined by the long axes and normals of the same base pairs), and the components of r are null. Figure courtesy of Luke Czapla. [43]

We approximate the *J* factors of looping from the statistical weights, *i.e.*, the Boltzmann factors, of the minimum-energy structures. In contrast to Monte Carlo simulation the optimization of potential energy makes it possible to investigate rare events and/or specific molecular events, e.g., deformations to the LacR protein.

2.3.4 TPM

Tethered particle motion (TPM) is a simple, single-molecule experimental technique for studying polymers like DNA and their interactions with entities like proteins. TPM experiments[8,41,42,47,48] have examined the looping propensities of Lac repressor-bound loops like those that I am modeling. These experiments provide additional measures to corroborate the accuracy of our model.

TPM experiments are performed by attaching one end of a strand of DNA to a surface. A gold reflective bead is attached to the other end of the strand and allowed to move freely. An optical microscope and camera are used to track the motion of the reported bead. An analysis of the Brownian motion of the bead is used to determine the distance between the end of the DNA tether and the surface. The random walk is monitored over a sliding window of 4 seconds. In the presence of the Lac repressor, a window reporting a shorter distance between the bead and the surface indicates a looping event. Two types of loops are implicated by two distinct different shortened distances.

Variations in the components of the TPM experiments have included operator binding affinity, sequence length, sequence curvature and Lac repressor concentration. The binding affinity is altered using different operator sequences (namely O_{sym}, O_1 and O_2). The O_{sym} is a pseudosymmetric operator designed with a stronger binding than the naturally occurring O_1. The weakest of the three operators used is the naturally occurring O_2. The third naturally occurring O_3 sequence, which has even a weaker affinity than O_2, has not been used in the experiments. DNA strands with varying length between the operators were used. Two classes

of nucleotide composition were used, one that is a mixed sequence and one that includes part of the 601 nucleosome positioning sequence that has a higher cyclization J-factor. Varying the concentration of the repressor affects the ease of looping. As would be expected at low concentration the occurrence of a loop is low. Similarly, at high concentration the occurrence of looping is also low as it is more likely that each of the operators is occupied by a different repressor. It is at intermediate concentrations that loops are most easily formed. A series of experiments were done to fit the parameters to the following probability function:

$$p_{loop}([R]) = \frac{\frac{1}{2}\frac{[R]J_{loop}}{K_i K_{ii}}}{1 + \frac{[R]}{K_i} + \frac{[R]}{K_{ii}} + \frac{[R]^2}{K_i K_{ii}} + \frac{1}{2}\frac{[R]J_{loop}}{K_i K_{ii}}} \quad (2.5)$$

Where K_i and K_{ii} are the dissociation constants for each of the operators and $[R]$ is the Lac repressor concentration. Of particular interest are the observed effects of sequence and length on looping probability which can be directly compared to the results of my models.

Chapter 3 DNA Sequence-Dependent Properties Of Protein-DNA Loops

The energy optimization model described in the previous chapter allows us to investigate features of DNA loops that would be impossible to achieve through experiment. By adjusting the parameters in the model we can simulate many of the features that affect the shape and stability of loops. In this chapter I describe the sequence-dependent effects on Lac repressor bound loops. Nucleotide sequence influences the looping of DNA in subtle ways, determining the lengths and types of loops formed with greatest ease in model systems. Moreover, certain sequence-dependent features of DNA, such as the curvature of DNA associated with runs of four to six recurring A·T base pairs, the A-tracts,[49–51] described in Chapter 2, stabilize protein-mediated DNA loops *in vitro*. The curved inserts within the Kahn group's designed loop constructs,[14,16] however, contain as many nucleotides as the wild-type DNA loops associated with the *Escherichia coli* Lac repressor protein *in vivo*. Changes in the nucleotide composition of shorter DNA detected through TPM measurements loops have less profound effects on looping stability. For example, the flexibility attributed to the 601 nucleosome-positioning sequence enhances Lac repressor-mediated loop formation, albeit to a much lesser extent than anticipated[41], and the stiffness underlying the resistance of long stretches of A·T base pairs to nucleosome formation appears to play no part in the context of looping.[47] The addition of architectural proteins, such as the histone-like E. coli protein HU, overwhelms the subtle sequence-dependent differences in looping detected *in vitro*.[52] It is not clear whether the influence of such proteins outweighs that of DNA sequence in the context of gene expression and looping *in vivo*. On the other hand, the arrangement of DNA regulatory elements within short Lac-repressor mediated loops modulates the activity of RNA polymerase.[53] Loops with the promoter elements recognized by the enzyme centered on their outer face are an order of magnitude more easily processed than those with the elements displaced by a half-helical turn to the inside.

Nucleotide sequence controls the helical repeat as well as the intrinsic folding and deformability of DNA. Changes in nucleotide composition perturb the classic right-handed B-DNA double helix with ~10.5 bp per turn to over- and underwound structures of the same helical sense with respectively fewer or more base pairs per turn. For example, whereas a homopolymer comprised exclusively of A·T base pairs adopts a roughly tenfold helix, the corresponding duplex with G·C pairs has a 10.6-10.7 bp repeat.[54–56] Repeated stretches of G·C base pairs also undergo transitions more easily from the B to the A form of DNA, a more compact helix with ~11 bp/turn,[57,58] than do stretches of A·T pairs[59]. These solvent-induced changes in helical repeat are coupled to changes in the inclination and displacement of base pairs, which are linked, in turn, to changes in the positions and orientations of consecutive base pairs[60]. Whereas the planes of base pairs stack roughly perpendicular to the global helical axis of B DNA, they deviate by 10-20° from this alignment when the helix is over- or undertwisted. Inclinations of this magnitude introduce 'kinks' at sites where different DNA helical structures, and hence different sequences, meet[61]. Such kinks and the underlying changes in local base-pair geometry provide a structural rationale for the curvature that accumulates in repeated DNA A-tracts,[51,62] and a framework for interpretation of various sequence-dependent features of DNA.[29,63–65]

Here we focus on sequence-dependent features of DNA that contribute to the formation of Lac repressor-mediated DNA loops. We first examine how changes in torsional stress associated with the variation of chain length and intrinsic double-helical twist, *i.e.*, how differences in the imposed twisting of DNA compared to that of a relaxed, spatially unconstrained molecule, contribute to the configurations of ideal DNA loops. We also determine the changes in looping associated with variation in the imposed end conditions, namely the rotation of one of the ends of a protein-anchored chain.

We consider the extent to which helical twist and other spatial and energetic features encoded in the DNA sequence contribute to looping. We focus on seven sequences with naturally curved inserts at locations introduced in the Kahn group's design of 156-160 bp Lac repressor-mediated loops with particular topologies,[16] *i.e.*, with the bound DNA operators attached to the protein assembly in one of four possible orientations.[6] We compare the energetically preferred orientations of the differently modeled loops with the populations of DNA loops extracted from the interactions of fluorescent dyes linked to the same sequences (See Chapter 2 for a more thorough description of the Kahn group experiment).

We first ignore the large-scale opening of the Lac repressor detected in low-resolution structural studies.[12,13,66] That is, we treat the protein as a rigid scaffold. The effect of Lac repressor opening in discussed in Chapter 4. We also omit consideration of the effects on looping of non-specific architectural proteins, such as HU, that distort the structure of DNA.[67,68] We further ignore fluctuations in the attachments of DNA operator sequences to the repressor headpieces.[69]

We find that the minimum elastic energies of short, protein-anchored fragments of ideal, naturally straight DNA mirror the simulated looping propensities and modes of chain attachment previously found by direct Monte Carlo sampling[43] and thereby capture the chain-length-dependent variation in loop formation deduced from experimental studies. The optimizations reveal two families of closed DNA structures and allow us to examine the competing influences of torsional and bending stress on the relative ease of formation of different closed spatial forms. We follow the changes of DNA energy and chain configuration, including the changes at individual base-pair steps, with variation in loop length, helical twist, and imposed anchoring conditions. We find that more realistic treatments of DNA, which incorporate the preferential bending of the double helix into the major and minor grooves and the coupling of this bending to the local twist, enhance the computed looping

propensities, particularly in torsionally stressed chains. The looping enhancement is even more pronounced in chains with a naturally curved insert. Moreover, the predicted modes of looping, dictated by the setting and superhelicity of the insert, capture the looping topologies deduced from fluorescence resonance energy signals on designed DNA sequences attached to the Lac repressor assembly. The intrinsic structure of the curved insert overwhelms imposed sequence-dependent variations in local chain deformability.

The energies and structures of the optimized loops are compared with the probabilities and configurations of loops determined previously by Czapla through direct Monte Carlo treatment of the same DNA-protein system,[43] namely the ideal DNA model and rigid, V-shaped Lac repressor structure described in Chapter 2. The optimization allows us to investigate the contributions of protein and DNA to loop formation at levels unattainable with random sampling techniques. Thus we can investigate rare supercoiled states and precise anchoring conditions, as opposed to the approximately closed structures of arbitrary topology and substantially higher energy captured with Monte Carlo methods.[36] Here, for simplicity, we focus on the energies and topological properties of the optimized loops and omit consideration of the thermal fluctuations of the closed structures, which can be estimated, for example, by normal-mode analysis.[70,71] We thus approximate the *J* factors of looping from the statistical weights, *i.e.*, the Boltzmann factors, of the minimum-energy structures.

3.1 Model Refinements

The various parameters in the DNA-protein model provide the ability to isolate a variety of properties that may affect loop formation. As a first step in addressing the sequence-dependent properties of looped DNA, I consider different force fields that take into account DNA sequence while keeping the rigid "V" shaped protein described Chapter 2. Thus the loop end constraints are kept constant. The features investigated include molecular

factors that influence torsional stress, a more realistic treatment of the DNA strand than ideal B DNA, and selected-nucleotide specific properties.

3.1.1 Torsional Stress

Torsional stress refers to change in twist with respect to the torsionally relaxed or rest state. Factors like sequence length, intrinsic twist and orientation of end constraints play a role. By increasing the length of the chain one base pair at a time we can examine the effect of sequence length on torsional stress. With each incremental change the twist energy redistributes along the chain. For example, in the case of a closed circle the average twist is equal to the number of helical turns times the number of base pairs per turn, i.e., (#turns * 360°)/#bp. As the chain increases or decreases in length the change in total twist must be distributed uniformly along the chain. As to be expected there is a phase-modulated quality to the energy distribution.

Torsional stress can also be altered by a change in the helical repeat. This may arise from a change to the intrinsic twist (relaxed state) as might occur with an environmental change like an increase in salt content or a change in temperature. It can also be effected by a change in constraints as might occur as a result of protein binding. To model the first effect I adjusted the model by changing the value of θ_3^0 in the intrinsic step parameters of the force field. (Table B-2) To model the second, or change in end constraints, I rotated the coordinates of the last base pair. Recall that in the optimization the first and the last base pair of the input configuration are held fixed and the potential energy for each base pair step is minimized.

3.1.2 Realistic Treatment of DNA

3.1.2.1 Hinge bending

To model the observed tendency of DNA to bend into the major or minor groove rather than along the backbone we can employ the features of the Schellman hinge-bending model of DNA,[72,73] which describes the DNA bending exclusively into the major or minor

groove. This representation can be achieved by setting the force constant for *tilt* to a large value so as to inhibit deformation into the backbone (about the short axis) and reducing the force constant for *roll* to retain the persistence length feature of the model.

3.1.2.2 Roll-twist coupling

A survey of high-resolution crystal structures reveals the interdependence of DNA step parameters. In particular, the untwisting of adjacent residues usually corresponds to an increase in roll and a decrease in slide. Similarly, the overtwisting of adjacent residues decreases roll and increases slide. Adding cross-terms to the force constant matrix for roll and twist (θ_1 and θ_3) allows the model to take into account the interdependence of these two parameters. (Table B-9) The treatment of the DNA as an inextensible elastic rod does not allow for the modeling of the correlation of the slide to twist and roll. A twist-roll modulus $f_{23} = f_{32} = 9.37^{-2}$ takes account of the correlation of twist and roll but is less strong than the observed coupling (Table 3-1).

3.1.3 Nucleotide Sequence Specific Characteristics

3.1.3.1 Pyrimidine-purine deformability

Studies in the bending behavior of the nucleosome show CA:TG steps to be most deformed. [74] The treatment of DNA at the level of the constituent pyrimidine and purine bases incorporates key sequence-dependent elastic properties of the double helix. The apparent ease of deforming pyrimidine-purine (YR) base-pair steps and the stiffness of purine-pyrimidine (RY) steps relative to purine-purine (RR) and pyrimidine-pyrimidine (YY) steps, found in well resolved protein-DNA crystal structures,[9,25]are approximated by a potential that reduces the elastic constants of the YR steps by a factor of 1–σ and raises those of the RY steps by a factor of 1+σ compared to those assigned RR and YY steps (Table 3-1). The persistence length of mixed-sequence DNA is accordingly preserved if the RR and YY steps take on the values assigned to the monomeric steps of the ideal elastic and hinge models. The persistence length of a homopolymer made up solely of YR-like steps, however,

is less than that of mixed-sequence DNA and the mean extension of a chain built only of RY-like steps is correspondingly greater (by –25% and +75%, respectively, when $\sigma=\frac{1}{4}$, Table 3-1)

	P+DNA structures*				DNA models§						
	YR	RR	RY	MN	Ideal	Hinge	Ro-Tw	RoTw′	YR	RR	RY
Angles (deg)											
$\langle\Delta\theta_{11}\rangle$	2.73	2.36	2.71	2.54	4.84	0	4.84	0	6.05	4.84	3.63
$\langle\Delta\theta_{22}\rangle$	4.72	4.01	3.20	3.98	4.84	6.85	4.84	6.85	6.05	4.84	3.63
$\langle\Delta\theta_{33}\rangle$	4.07	3.45	3.47	3.61	4.05	4.05	4.05	4.05	5.06	4.05	3.04
Displacements (Å)											
$\langle\Delta\theta_{44}\rangle$	0.57	0.47	0.54	0.50	0	0	0	0	0	0	0
$\langle\Delta\theta_{55}\rangle$	0.70	0.43	0.33	0.47	0	0	0	0	0	0	0
$\langle\Delta\theta_{66}\rangle$	0.21	0.17	0.18	0.18	0	0	0	0	0	0	0
Covariance (Å)											
$\langle\Delta\theta_{23}\rangle$	7.73	7.21	6.20	7.10	0	0	9.37	9.37	0	0	0
Persistence length† (Å)											
a	—	—	—	—	472	487	455	418	308	472	826

Table 3-1 Observed and Modeled Elastic Parameters of DNA Base-pair Step.

**Apparent deformational constants extracted from 2862 DNA base-pair steps in a database of 135 non-redundant, high-resolution protein-DNA structures;[65] values are averages of the quantities for all steps of a given type: YR – pyrimidine-purine; RR – purine-purine; RY – purine-pyrimidine; MN – generic, sequence-independent step with* $\Delta\theta_{ij}=\frac{1}{4}\left(\Delta\theta_{ij}^{YR}+2\Delta\theta_{ij}^{RR}+\Delta\theta_{ij}^{RY}\right)$*. The force constants impeding deformations of individual parameters (Eqn. 2.1) are the inverse squares of the listed values, i.e.,* $f_{ij}=\langle\Delta\theta_{ij}\rangle^{-2}$.

§Individual models that incorporate selected features of double-helical structure: ideal – isotropic bending and independent twisting consistent with macroscopic properties (see text); hinge – preferential bending into the grooves (via Roll); Ro-Tw – isotropic bending and coupled deformations of roll and twist; Ro-Tw′ – hinge bending and coupled deformations of roll and twist; YR – enhanced local deformability typical of pyrimidine-purine steps; RR – purine-purine steps equated to ideal DNA: RY – reduced local deformability typical of purine-pyrimidine steps.

†Persistence lengths of inextensible, naturally straight homopolymers with the specified elastic parameters. Values obtained, following Flory,[59] from the average projection of the end-to-end vector on the initial direction of the chain in the limit of infinite length, here taken as 1,000 bp. The computed distances converge to stable average values in Monte-Carlo simulated samples of ~25,000 configurations.

3.1.3.2 Sequence Dependent curvature

The curvature of DNA introduced in certain Lac repressor-mediated loops by the insertion of 80 bp A-tracts[14,16] is treated with a simple qualitative model of DNA bending[62] based on the sequence-dependent roll angles of stretches of A·T base pairs in known high-

resolution structures. The model, which accounts for the magnitude and directionality of A-tract bending observed in gel and cyclization experiments,[75,76] describes a quasi planar pathway in the equilibrium rest state and includes three important features of local DNA structure: (i) on average, the roll of AA steps is less than that of all other steps; (ii) there is significant positive roll in the CA·TG steps at the 5′-termini of the A-tracts (or the TG·CA steps at the 3′-termini of the T-tracts in the complementary strand); (iii) the average roll of base-pair steps is ~3°. Thus the intrinsic roll (θ_2^0) of CA·TG steps is taken to be 5° more positive and that of AA·TT steps 5° more negative than the 3° roll assigned to all other steps. (Table 3-2) By contrast, the model of DNA bending, based on DNase cutting patterns and nucleotide sequence preferences in nucleosomal DNA[64] and used as the reference state in previous studies of DNA loops stabilized by A-tracts,[77] incorporates features indicative of an overwound helix, namely large negative values of roll and twist angles greater than those of canonical B DNA, and produces a superhelical curve in the equilibrium rest state. The sequence-dependent model used here includes the set of twist angles[63] deduced over 30 years ago from the first B-DNA crystal structure[78] and contemporaneous studies of DNA electrophoretic migration[54,55] and nuclease cutting patterns[56] and subsequently verified in the accumulating database of protein-DNA crystal structures.[65,79,80] The average twist angles in the latter data sets correspond to helices with ~10.6 bp/turn.

Step	Twist* (1144 steps)[79]	Roll*	Twist* (2862 steps)[65]	Roll*	Twist‡	Roll§	Twist†	Roll†
CG	36.1	5.4	34.4	5.5	29.8	3	34.3	–11.6
CA·TG	37.3	4.7	35.0	5.1	34.5	8	35.3	–12.4
TA	37.8	3.3	37.4	2.5	36.0	3	36.2	–11.3
AG·CT	31.9	4.5	32.5	4.1	27.7	3	35.0	–12.1
GG·CC	32.9	3.6	33.3	5.0	33.7	3	36.0	–11.2
AA·TT	35.1	0.7	35.1	0.7	35.6	–2	36.0	–8.1
GA·TC	36.3	1.9	35.5	1.9	36.9	3	37.0	–11.7
AT	29.3	1.1	29.8	1.0	31.5	3	35.8	–9.4
AC·GT	31.5	0.7	31.7	1.6	34.4	3	36.7	–10.9
GC	33.6	0.3	33.7	1.2	40.0	3	37.7	–14.1
MN∞	34.2	2.6	33.8	2.9	33.8	3	36.0	–11.2

Table 3-2 Observed and Modeled Sequence-dependent Rigid-body Parameters of DNA Base-pair Steps

**Mean values characterizing the base-pair steps in two sets of high-resolution protein-DNA crystal structures.[65,79] While the earlier dataset includes structural redundancies among the 93 selected complexes, there are no redundancies among the135 structures in the latter dataset.*

‡Values extracted from a linear fit of early measurements of DNA twist.[63]

§Simplified representation of the sequence-dependent roll angles detected in solution and solid-state studies of DNA A-tracts.[62]

†Average rigid-body parameters of base-pair steps in models of linear DNA with the designed A-tracts listed in Appendix A. The models, which are based on the DNase cutting patterns and nucleotide sequence preferences in nucleosomal DNA,[64] were generated using the bend.it webserver[81] and analyzed with the 3DNA suite of programs.[60,82] Note the higher twist and more negative roll angles compared to other experimentally derived datasets.

∞Average parameters for a generic MpN step based on equal weighting of the average parameters of the 16 common dimers, i.e., AA and TT, AG and CT, etc., have identical averages except for different signs of tilt.[83]

3.2 Results

3.2.1 Looping Profiles of Ideal DNA

The sums of the statistical weights of energy-optimized loops of ideal DNA mirror the *J* factors captured in previous Monte Carlo simulations of the looping of the same type of chain, a naturally straight, inextensible molecule with 10.5 bp per helical turn subject to isotropic bending and independent fluctuations of twist, against the V-shaped Lac repressor

assembly (Figure 3.1).[43] The agreement is best at shorter chain lengths, including the lengths of wildtype loops mediated by the Lac repressor protein with center-to-center operator spacing of 92 bp steps. The estimated *J* factors exceed the directly simulated *J* factors at longer chain lengths, where the sums of the statistical weights of the optimized loop configurations deviate from the directly computed values of *J* by as much as 27%. The orientational preferences of the optimized loops, nevertheless, reproduce trends found through Monte Carlo sampling; that is, the more easily formed loops anchor against the protein in antiparallel orientations and the less easily formed loops in both antiparallel and parallel orientations. Moreover, the predicted occurrences of different loop orientations from the two approaches agree surprisingly well: the mean absolute differences in the fractional populations of 73-143 bp loops fall in the range 0.01-0.06, with a few differences of 0.2 or more for high-energy loops with limited Monte Carlo examples (Figure 3.2). Notably, the build-up of low-energy loops with parallel attachments at longer chain lengths follows the simulated increases in population, including the increasingly preferential formation of loops in P2 rather than P1 orientations ,where the DNA follows a semicircular as opposed to an ℓ-shaped pathway when viewed in a direction perpendicular to the "V" formed by the arms of the protein assembly (Figure 3.6).

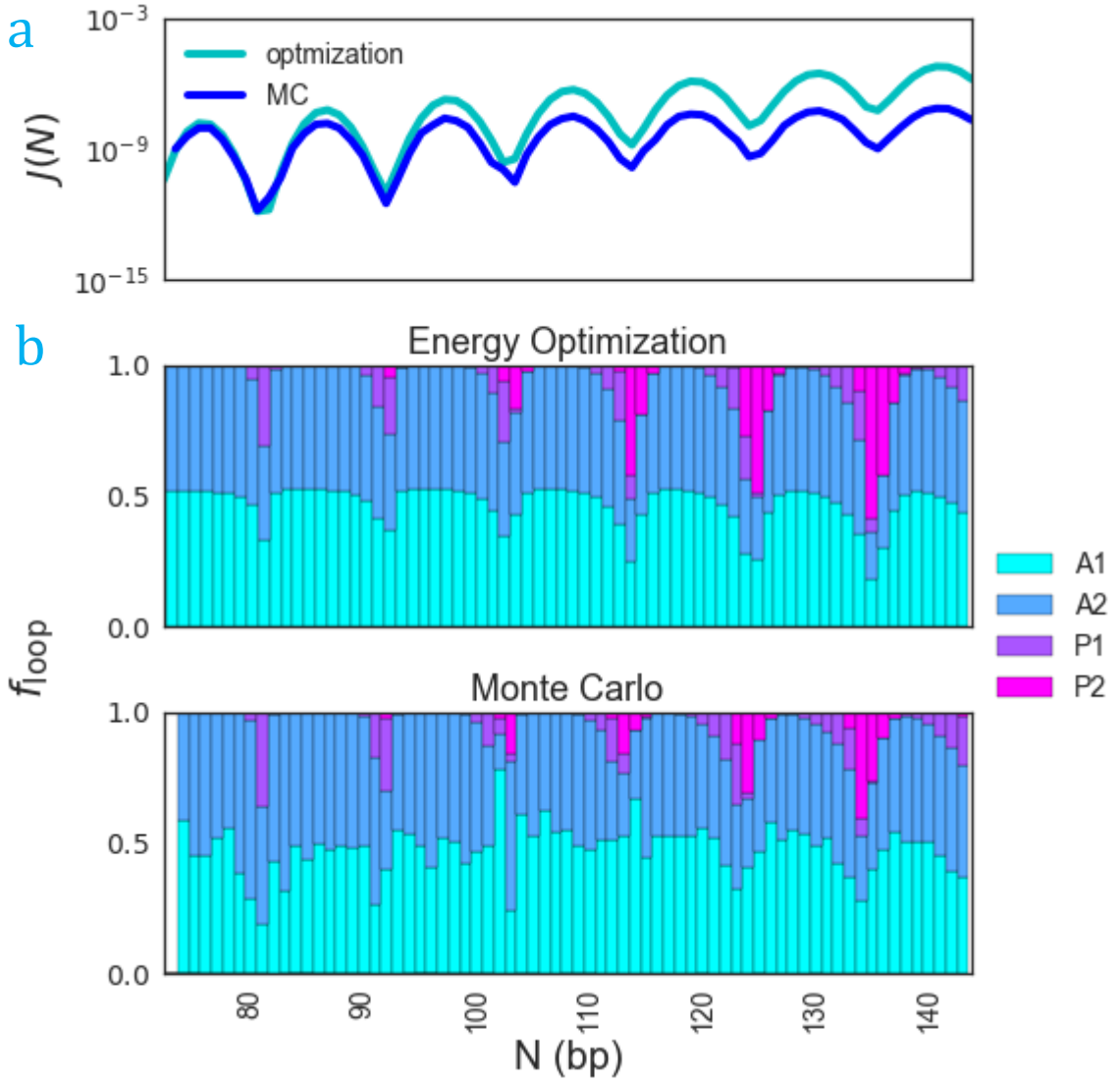

Figure 3.1 The propensities and orientations of Lac repressor-mediated DNA loops determined through energy optimization mirror the chain-length dependent behavior found from Monte Carlo sampling. (a) Values of the J factor estimated from the sums of the statistical weights of the minimum-energy configurations of DNA loops with center-to-center operator spacing N (cyan curve) compared with the proportion of simulated configurations[43] in ensembles of 10^{12}-10^{16} fluctuating chains of the same length with terminal base pairs in the vicinity of the requisite closure conditions (blue curve). (b) Corresponding fractions of loops f_{loop} with DNA attached in one of the four specified orientations against the V-shaped protein assembly: predictions from minimization (upper plot) and simulation[43] (lower plot). The 5'-3' directions of the DNA operators in each loop orientation is as depicted Figure 1.2. DNA modeled as an ideal, inextensible isotropic chain with an intrinsic twist of 10.5 bp/turn.

The chain-length-dependent looping profile associated with each orientational setting arises from two families of structures, revealed through the energy optimization (Figure 3.3, Figure 3.4). The energy and twist of the loops in the two families, here termed F1 and F2, exhibit similar ~21 bp periodic dependencies on chain length. The oscillations of energy *U*, with chain length *N*, however, differ in phase by a helical turn (Figure 3.3); that is, the valleys in the energy profile of one family of looped structures coincide with the peaks in the other and *vice versa*. The energy minima occur at chain lengths where the uptake of twist is minimal and the ends of the DNA fall in nearly perfect register with the binding sites on the anchoring protein.

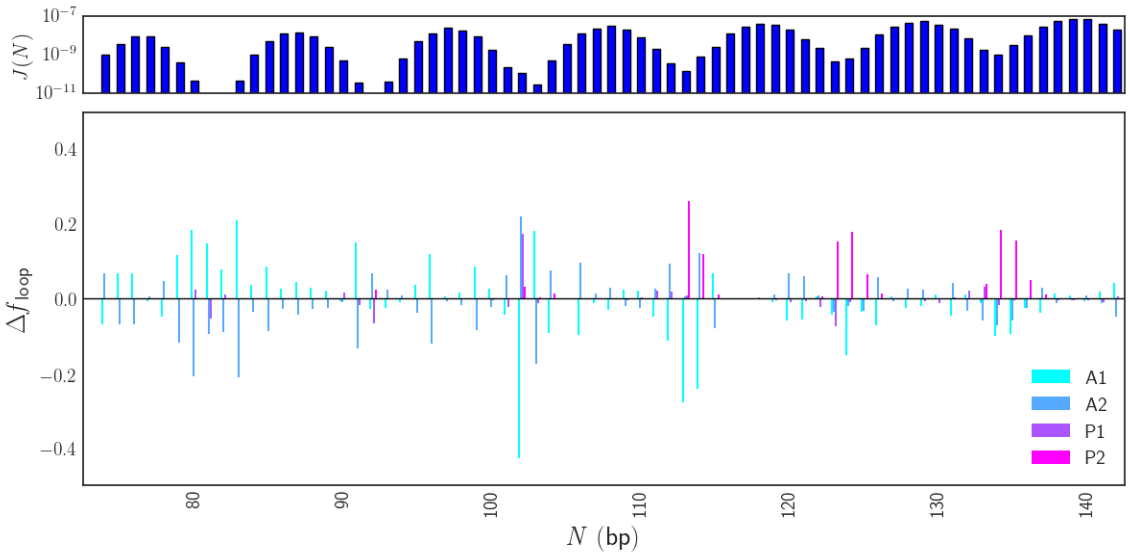

Figure 3.2 Differences in the fractional populations of Lac repressor-mediated loops of ideal DNA predicted through energy optimization vs. Monte Carlo sampling[43]. Data reported for loops of 73-143 bp anchored in antiparallel (A1, A2) and parallel (P1, P2) orientations on the Lac repressor assembly. The largest discrepancies occur at chain lengths where the looping propensity (upper histogram) is low.

The addition or removal of base pairs has little effect on the overall twist of the closed loops until the length of the constrained DNA differs by roughly a full helical turn from that of the nearest minimum in the associated energy profile. The energy resulting from the

accumulated deficit or surplus in twist peaks in value at this point and the total twist rises or falls precipitously (Figure 3.3). The under- or overwound loop subsequently converts to a comparably over- or underwound loop differing in chain length by a single base pair. The sudden change in twist is much sharper for loops anchored in parallel than in antiparallel orientations and for parallel loops anchored in P2 as opposed to P1 orientations. The jumps in twist give rise, in turn, to jumps in linking number of ±2 if the loops form within longer, covalently closed DNA molecules. Thus, one family of loops forms in closed DNA chains with even values of the linking number and the other in closed chains with odd values. The linking number jumps by ±1 at chain lengths where the energy profiles cross and the stability between the modes of looping switches between structural families.

Not surprisingly, the jumps in twist and the peaks in energy give rise to large changes in the folding of the repressor-bound loops (Figure 3.5, Figure 3.6). For example, the addition or removal of base pairs in the simple, planar U-turns of DNA — formed when the molecule assumes an antiparallel binding orientation and the ends of the chain lie in nearly perfect register with the protein headpieces — induces out-of-plane bending in the anchored duplex. Moreover, the extent of global distortion differs in the two structural families and alternates at the points of maximum energy where the twist jumps precipitously with further increase in loop length. In other words, antiparallel loops of increasing chain length adopt configurations that swing in and out of the plane that contains the axis of the DNA in the states of low energy. The direction of deformation depends upon loop orientation, *i.e.*, A1 vs. A2, but is independent of loop family, *i.e.*, A_1^{F1} vs. A_1^{F2} and A_2^{F1} vs. A_2^{F2}. The loops anchored in parallel orientations undergo different types of distortions and, unlike the antiparallel loops, show different behaviors at short and long chain lengths. The low-energy states of the parallel loops describe helical arcs of low pitch directed in roughly perpendicular directions with respect to one another, *i.e.*, with the global P1 axis approximately collinear with the bound DNA

operators and the shallow P2 pathway approximately in the plane of the operators. The addition or removal of base pairs makes the pathways more elliptical and at shorter chain lengths may lead to pronounced out-of-plane distortions of the loop as a whole. The ends of the DNA appear, from certain viewpoints, to cross and uncross as the loops rotate globally with increase or decrease of chain length

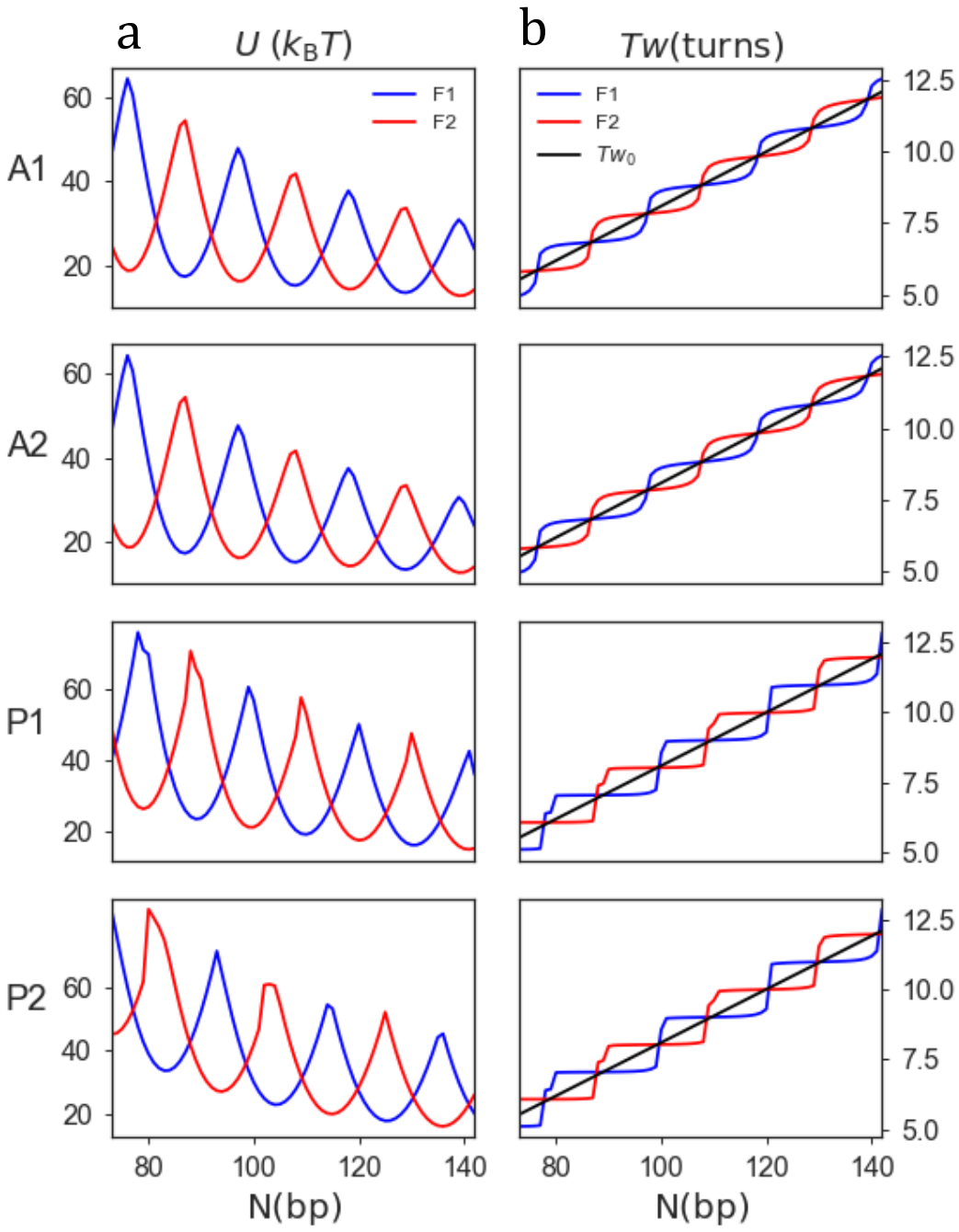

Figure 3.3 Energy and Helical Turns for competing families. DNA alternately attaches to protein in one of two families of looped structures with added turns of the double helix. Overlapping profiles, as a function of DNA length, of (a) *the minimum elastic energy U and (b) the total twist Tw of two families of Lac-repressor mediated loops anchored in antiparallel and parallel orientations. Note the jumps between energetically preferred structural forms when one of the loops is overtwisted and the other undertwisted, each by about half a helical turn, with respect to the total intrinsic twist Tw0 (black straight line in* b) *and the different locations and energies of 'jump' sites on the antiparallel and parallel loops.*

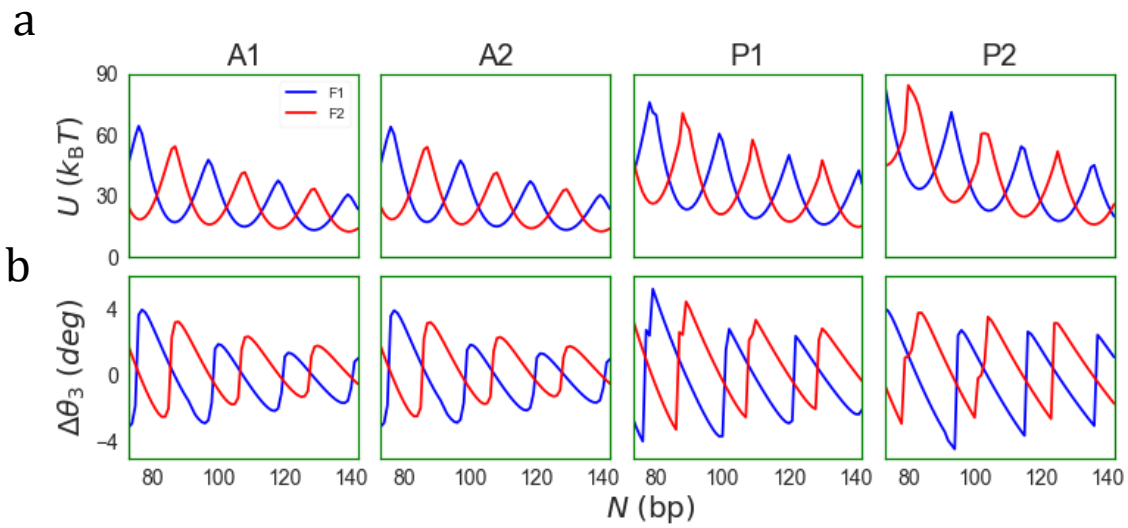

Figure 3.4 Overlapping profiles of (a) *the minimum elastic energy U, and* (b) *the average local torsional stress* $\langle \Delta\theta_3 \rangle = \langle \theta_3 \rangle - \theta_3^0$ *of two families of ideal, 73-143 bp DNA loops* ($\theta_3^0 = 360°/10.5$) *anchored in all possible orientations on the Lac-repressor. The plotted data in* (a) *are the same as in Figure 3.3 and included to facilitate comparison*

Not surprisingly, the jumps in twist and the peaks in energy give rise to large changes in the folding of the repressor-bound loops (Figure 3.5, Figure 3.6). For example, the addition or removal of base pairs in the simple, planar U-turns of DNA — formed when the molecule assumes an antiparallel binding orientation and the ends of the chain lie in nearly perfect register with the protein headpieces — induces out-of-plane bending in the anchored duplex. Moreover, the extent of global distortion differs in the two structural families and alternates at the points of maximum energy where the twist jumps precipitously with further increase in loop length. In other words, antiparallel loops of increasing chain length adopt configurations that swing in and out of the plane that contains the axis of the DNA in the states of low energy. The direction of deformation depends upon loop orientation, *i.e.*, A1 vs. A2, but is independent of loop family, *i.e.*, A_1^{F1} vs. A_1^{F2} and A_2^{F1} vs. A_2^{F2}. The loops anchored in parallel orientations undergo different types of distortions and, unlike the antiparallel loops, show

different behaviors at short and long chain lengths. The low-energy states of the parallel loops describe helical arcs of low pitch directed in roughly perpendicular directions with respect to one another, *i.e.*, with the global P1 axis approximately collinear with the bound DNA operators and the shallow P2 pathway approximately in the plane of the operators. The addition or removal of base pairs makes the pathways more elliptical and at shorter chain lengths may lead to pronounced out-of-plane distortions of the loop as a whole. The ends of the DNA appear, from certain viewpoints, to cross and uncross as the loops rotate globally with increase or decrease of chain length

3.2.2 Helical Constraints and DNA Loop Formation

Variation in the assumed twist of DNA introduces a phase shift in the looping profiles, with the maxima and minima in loop formation involving different numbers of base pairs. The statistical weights and relative populations of the loops constructed from DNA with different helical repeats, however, coincide when expressed in terms of the number of helical repeats (Figure 3.8). For example, the energies of loops attached to the repressor in antiparallel orientations are lowest and the *J* factors (statistical weights) highest when the center-to-center distance between operators is slightly more than an integral number of helical turns (10.5n+2-3 bp of DNA with a 10.5 bp helical repeat or 10n+3 bp of DNA with a 10 bp repeat, where n is the number of helical turns and even values of n arise from one family of looped structures and odd values from the other). Similarly, the energies of loops anchored in P1 orientations are lowest in loops of 10.5n+4-5 bp or 10n+5-6 bp and those closed in P2 orientations in loops of ~10.5n–1-2 bp or 10n+0 bp. As noted above, the P1 states become less dominant and the P2 states increasingly important with increase in chain length. That is, when sufficiently long, the energies of loops anchored in P2 orientations may become more favored than those of loops attached in P1 orientations and thereby dominate the configurational landscape at torsional settings where antiparallel loop formation is costly.

The cost of bending a chain in the semicircular arrangements characteristic of P2 looping approaches and at certain chain lengths falls below that of anchoring the DNA in either the ℓ -shaped pathways typical of P1 loops or the U-turns assumed by antiparallel loops (Figure 3.7).

Figure 3.5 Molecular images of A1 and P1 loops . The torsional stress associated with the uptake of DNA base pairs perturbs the spatial arrangements of protein-anchored DNA loops. Molecular images illustrating the changes in overall folding of energy-optimized loops as 21 bp (two helical turns) of DNA are added to chain fragments anchored in antiparallel (upper rows) and parallel (lower rows) orientations on the Lac repressor. Examples correspond to loop lengths where the energy differences ΔU between configurations from two competing structural families (here A_1^{F1} vs. A_1^{F2} and P_1^{F1} vs. P_1^{F1}) are most and least extreme. States in the vicinity of transition points ($\Delta U \approx 0$) are denoted by a double arrow and configurations preferentially adopted at intermediate chain lengths ($\Delta U \gg 0$) by the head of a single arrow. Images rendered in PyMOL (www.pymol.org) with DNA backbones shown as gold tubes, DNA bases as gold sticks, and repressor chains as pink and cyan ribbons. Views looking down the axis perpendicular to the "V" formed by the arms of the protein assembly

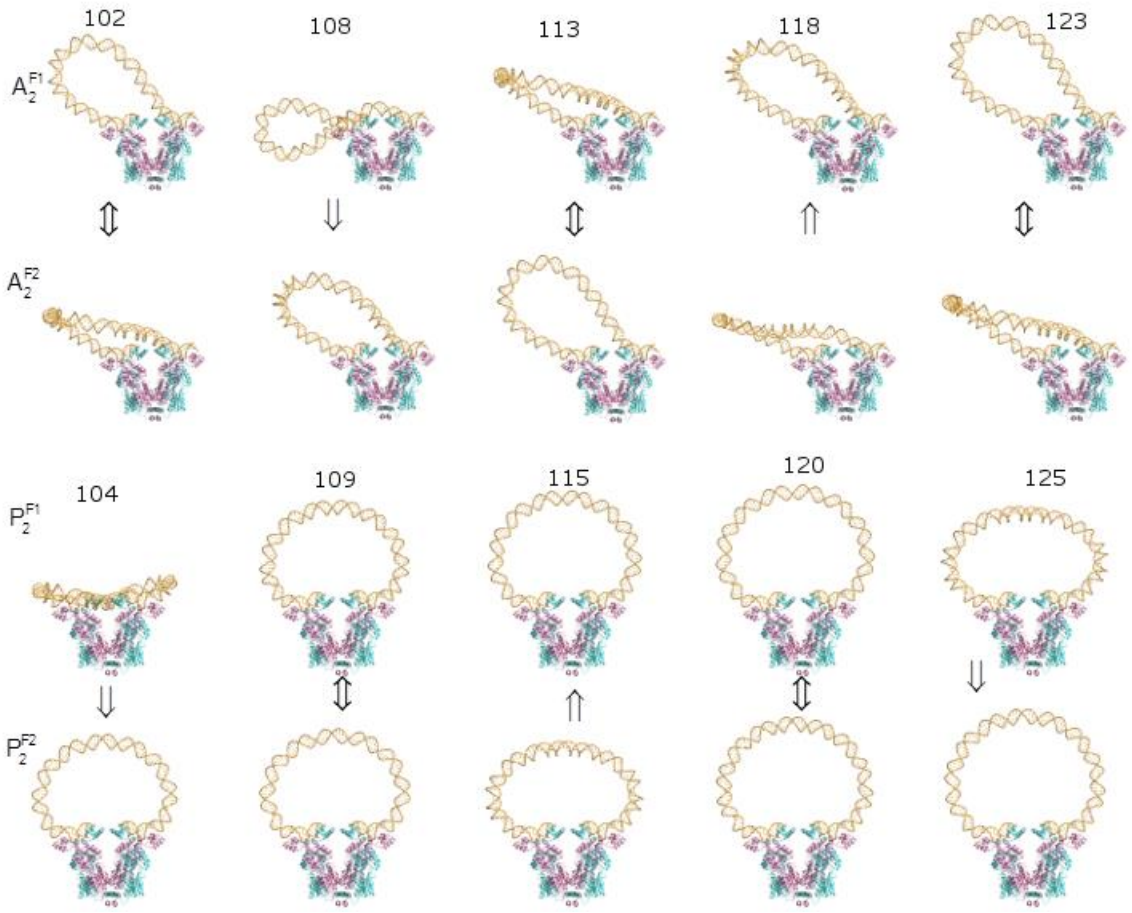

Figure 3.6 Molecular images of A2 and P2 loops, illustrating the changes in overall folding of energy-optimized loops as 21 bp (two helical turns) of DNA are added to chain fragments anchored in antiparallel and parallel orientations (here A_2^{F1} vs. A_2^{F2} and P_2^{F1} vs. P_2^{F2}) on the Lac repressor. For legend see Figure 3.5

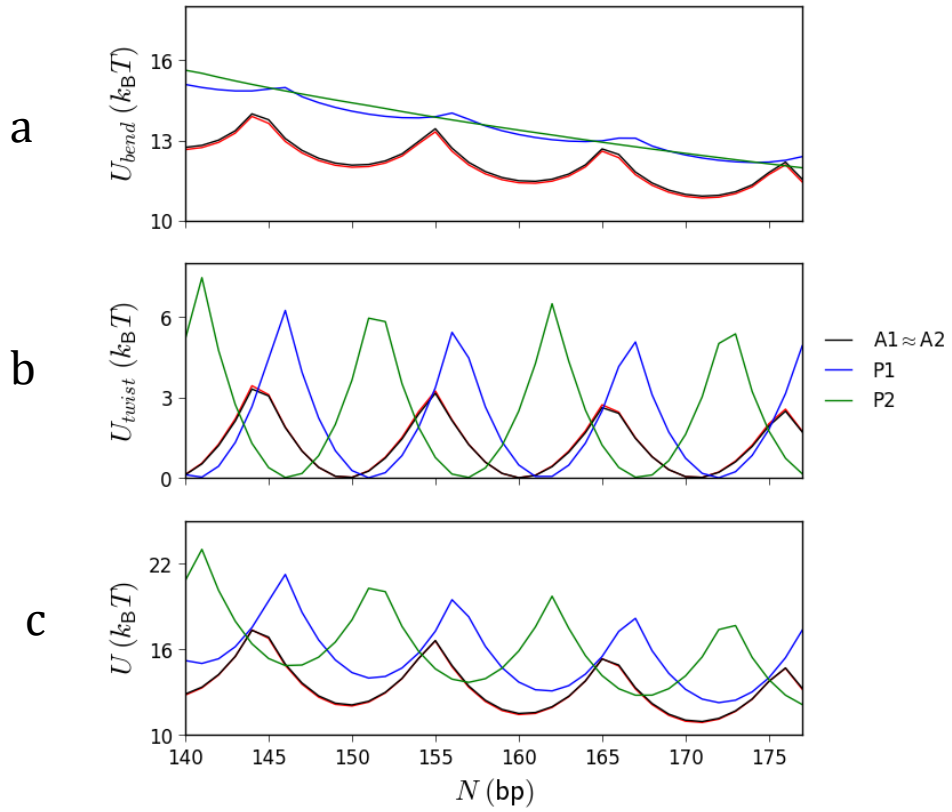

Figure 3.7 Chain-length-dependent variation in bending vs. twisting energy. The chain-length-dependent variation in (a) *the bending energy,* (b) *the twisting energy, and* (c) *the total elastic energy of 140-177 bp loops of ideal, naturally straight DNA attached in different orientations to the Lac-repressor assembly. The DNA helical repeat is set at 10.5 bp/turn and the energies are Boltzmann averages over the two families of looped structures associated with the specified binding orientations.*

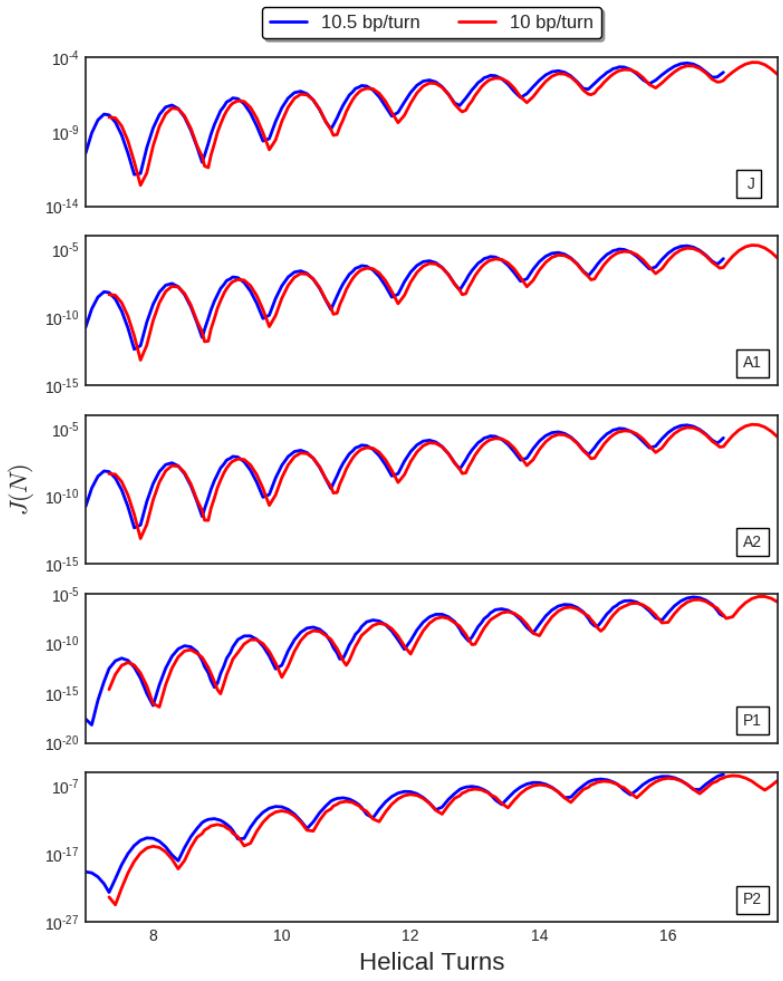

Figure 3.8 J factors of DNA chains with different intrinsic helical repeats. The looping propensities (J factors) of DNA chains with different intrinsic helical repeats coincide if expressed in terms of the number of helical turns. Probability of looping DNA between the headpieces of the Lac repressor assembly as a function of the number of turns of helices with 10 or 10.5 bp repeats (red, blue curves respectively). Note the differences in magnitude and phase associated with the formation of loops in different orientations on the protein and the increased likelihood (higher J factors) of parallel loops, particularly the P2 form, at longer chain lengths (≥11 DNA turns).

A change in DNA helical repeat simply alters the number and range of helical turns in the computed loops. For example, the 73-177 bp of DNA in the loops studied here comprise 7-17 turns of a 10.5 bp helix and 7.4-17.8 turns of a 10 bp helix. On the other hand, the types of loops predicted to occur at a given chain length differ substantially between DNA models with different helical repeats. For example, 92- and 113 bp loops of ideal DNA close more easily with much less torsional stress when the reference state is a 10-fold helix. Moreover, these

loops attach to the Lac repressor assembly almost exclusively in antiparallel orientations as opposed to the mix of antiparallel and parallel orientations assumed by the higher energy, more torsionally stressed loops of the same length with a 10.5 bp repeat (Figure 3.9). The trends switch if the DNA is a half helical turn longer or shorter. That is, loops of 97 and 108 bp form with greater ease and close preferentially in antiparallel arrangements if the DNA has a 10.5 bp rather than a 10 bp repeat. Chains with the 10 bp repeat then assume the mix of looped orientations characteristic of a torsionally stressed molecule. The build-up of torsional stress introduces a change in linking number as evidenced by the mixed topologies of the less easily formed loops (Figure 3.10).

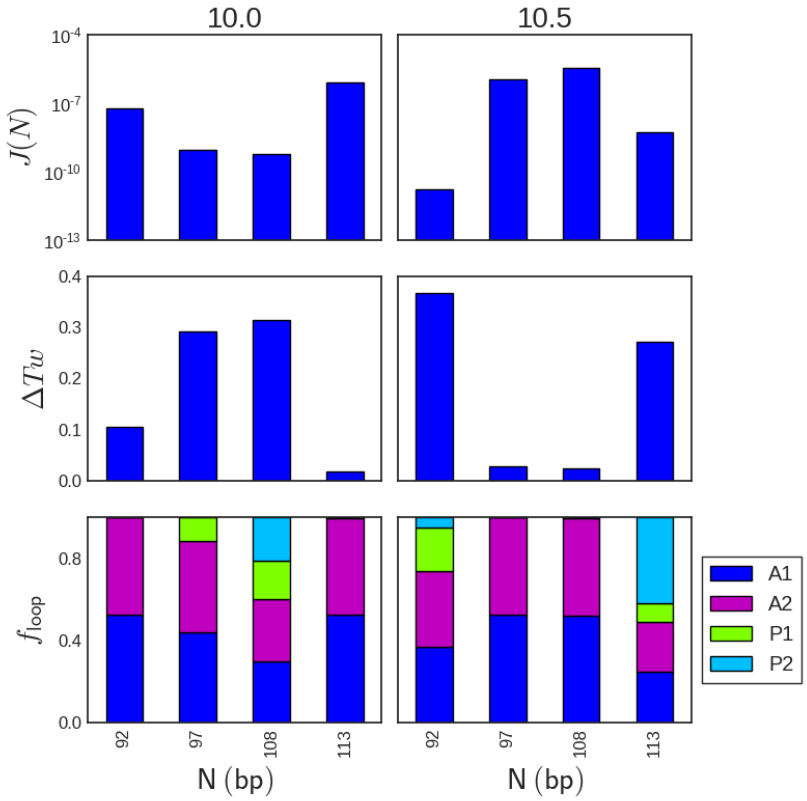

Figure 3.9 The mode of looping predicted to occur at a given chain length depends upon the assumed DNA helical repeat. Comparison of the J factors/statistical weights (top), mean torsional stress <ΔTw> (middle), and orientational frequencies (bottom) of optimized DNA loops of 92, 97, 108, and 113 bp, where the chain model has a helical repeat of (left) 10 or (right) 10.5 bp/turn and <ΔTw> is the number of excess helical turns averaged over all optimized loops of the given length and intrinsic twist

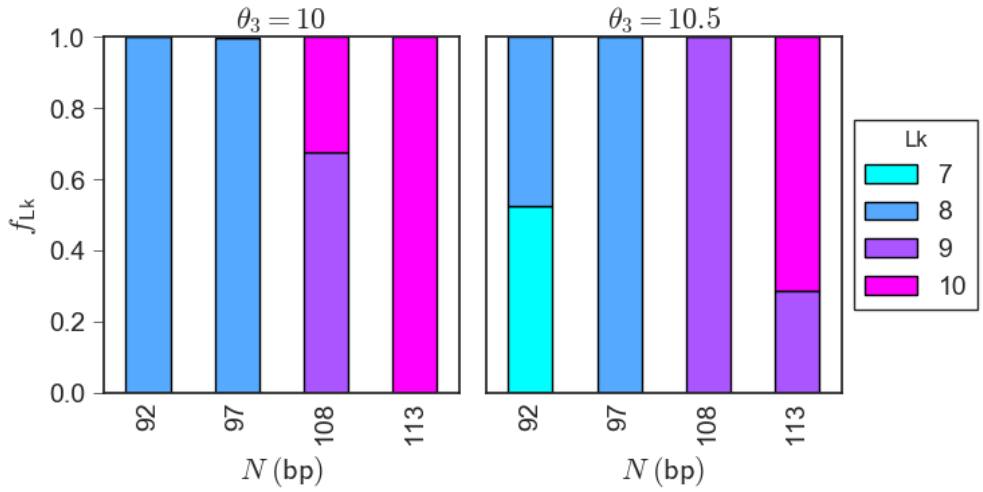

Figure 3.10 Comparison of the frequencies of linking numbers between 7 and 10 in energy-optimized 'closed' DNA loops of 92-113 bp, where the DNA helical repeat is 10 (left) or 10.5 (right) bp/turn repeats and the chain is closed by the vector connecting the centers of terminal base pairs

3.2.3 Orientational Constraints on Looping

The rotation of terminal base pairs, such as might be introduced by the deformation of a mediating protein, can produce the same net changes in overall DNA twist and looping propensity in a linear molecule as the variation of intrinsic helical twist. The change in twist and base-pair location arise in the former case from perturbations of the instantaneous values of the local base-pair twist and in the latter from the values associated with the minimum-energy reference state, *i.e.*, the θ_i vs. the θ_i^0 in (Eqn. 2.1). The optimized dinucleotide pathways of clamped, linear B-DNA chains of the same length — one with one of the two end base pairs rotated before clamping and the other with both ends clamped before changing the intrinsic twist — are identical if the rotation $\Delta\varphi$ imposed on the former chain is equal to the change in intrinsic twist ($N\Delta\theta_3^0$) introduced along the length of the latter. That is, the structural identity holds when $\Delta\varphi = N(360/\nu - 360/\nu_0)$ or conversely when $\nu = 360/(\Delta\varphi/N + 360/\nu_0)$, where N is the number of base-pair steps along the two DNA chains, ν is the helical repeat introduced in the chain with two fixed ends, and $\Delta\varphi$ is expressed in degrees. Here ν_0 is taken as 10.5 bp/turn.

On the other hand, although the variation in chain length imposes the same constraints on loop orientation as the intrinsic twist, the addition/removal of base pairs changes the magnitude of the elastic energy and the relative weights of Lac repressor-mediated looped states. For example, ±10 bp changes in loop length from the 92 bp wild-type spacing introduce a more asymmetric response in the elastic energy than the equivalent rotational changes relative to B DNA ($\Delta\varphi = \pm 346°$); the energy of the shortened chain is higher and that of the lengthened chain is lower than found for a chain of fixed length subjected to the same orientational constraints (Figure 3.11). Moreover, the differences in energy enhance the propensity for antiparallel loop attachments when the DNA is shortened and parallel attachments when the DNA is lengthened. Although the changes in energy follow the expected effects of chain length, *i.e.*, increases/decreases in value when local structural deformations are spread over fewer/more residues, the build-up of torsional stress differs in antiparallel vs. parallel loops. Whereas the energy terms contributing to the antiparallel loops reflect a mix of torsional and bending contributions, the bending term dominates the changes in energy of parallel loop formation with variation in chain length (Figure 3.12).

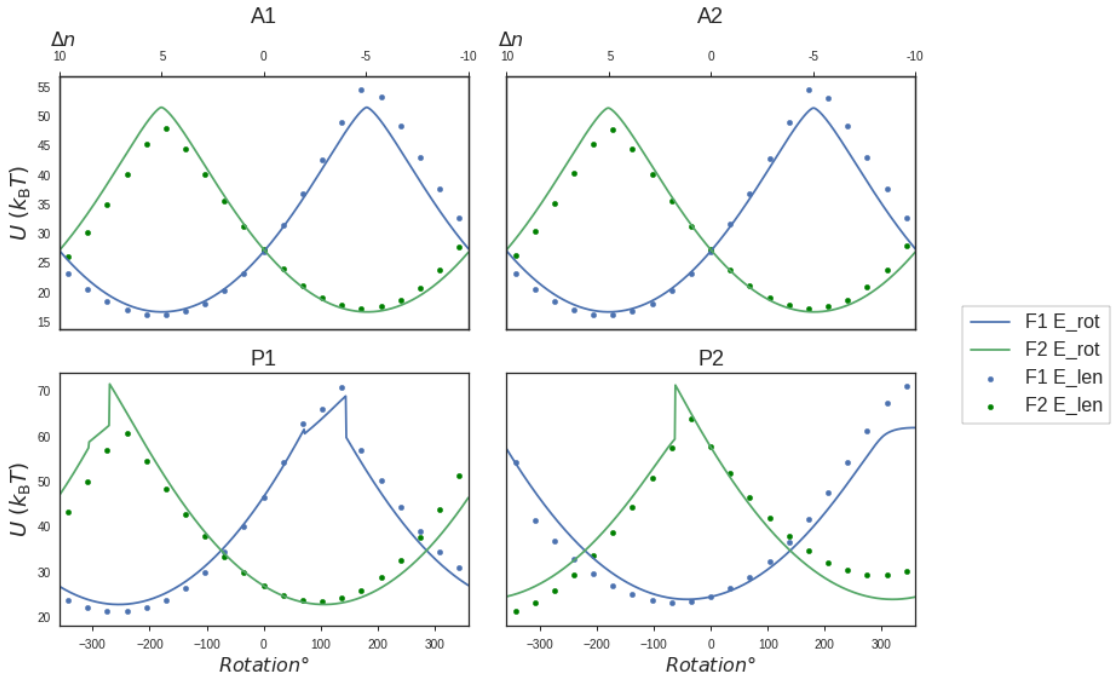

Figure 3.11 Comparison of elastic energy for changes in chain length vs. orientational constraints. Changes in loop length introduce an asymmetric response in elastic energy compared to changes in orientational constraints alone. Profiles, as a function of chain length (isolated points) and operator orientation (thin lines), of the minimum elastic energy of two families of Lac repressor-mediated loops anchored in antiparallel and parallel orientations. Note the differences in magnitude associated with the addition/removal of 10 bp compared to the reorientation of terminal residues.

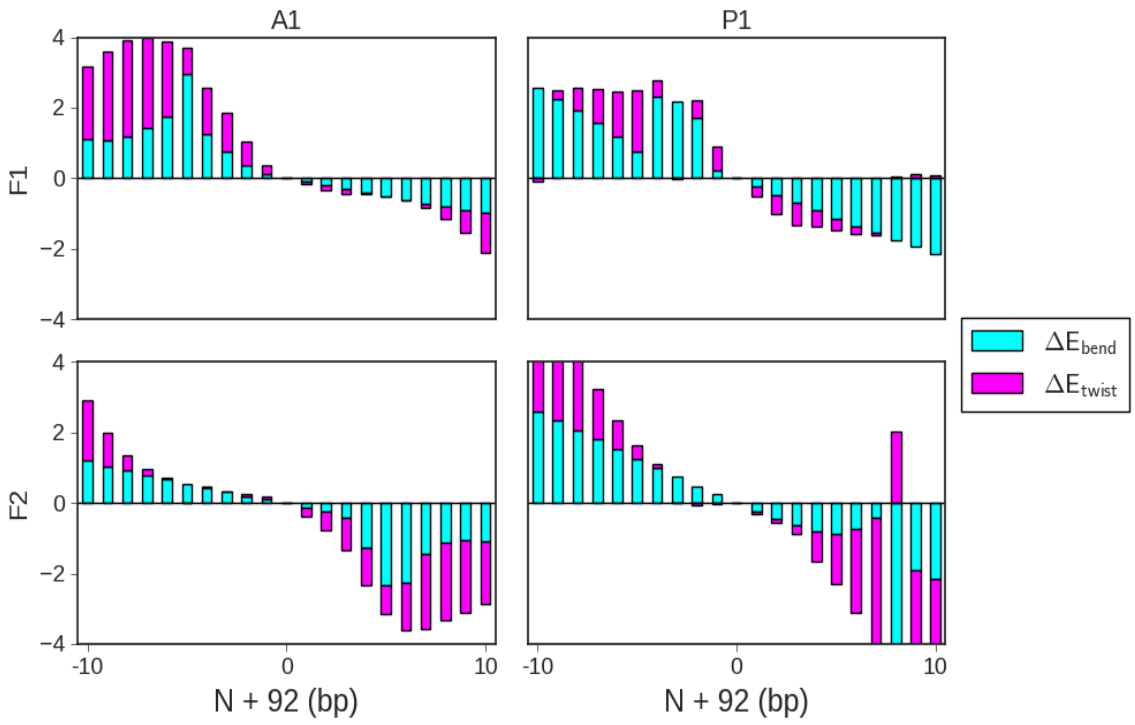

Figure 3.12 Differences in the bending and twisting energies of loops of ideal, naturally straight DNA subjected to the orientational constraints associated with ±10 bp changes in the length of 92 bp loops compared to the energies obtained with the corresponding addition/removal of base pairs. Note the different mix of torsional and bending contributions associated with antiparallel (A_1^{F1}/A_1^{F1}) and parallel (P_1^{F1}/P_1^{F2}) orientations.

3.2.4 Realistic Deformations of Looped DNA

More realistic treatment of DNA that takes account of the known tendency of base-pair steps to bend via roll (into the grooves) rather than tilt (toward one of the backbones) introduces small, chain-length-dependent variations in the predicted ease of looping. The optimized energies of closed molecules subject to such hinge bending differ only slightly from those of ideal, isotropic DNA loops over the range of chain lengths considered here. (Table B-7) The energies tend to be lower and the looping propensities higher than those of ideal, isotropically bendable DNA if the double helix is overtwisted upon attachment to the repressor and the energies higher and the looping propensities lower if the chain is undertwisted. On average, the mode of bending has only limited effects on loop formation as estimated from the statistical weights of the minimum-energy configurations. The ratio of the

derived *J* factors of chains subject to hinge vs. ideal, isotropic bending oscillates with diminishing amplitude about a mean value of unity as chain length increases (J_{hinge} curve in Figure 3.13).

By contrast, the introduction of roll-twist coupling enhances the looping propensity compared to that of a chain with independent angular fluctuations. As noted above, incorporation of the roll-twist modulus intensifies deformations of DNA away from the rest state and shifts the distribution of end-to-end distances in linear chains toward smaller values. The energies of closed chain configurations consequently become lower and the associated statistical weights higher. For example, the computed *J* factors (statistical weights) of 73-177 bp DNA chains subject to isotropic bending and roll-twist coupling range between 1.3 and 2.3 times those of ideal DNA loops of the same length and helical repeat. The enhancement is greater for the more difficult to form loops and smaller for the more easily closed chains. The interdependence of roll and twist accordingly dampens the oscillations in the ratio of *J* factors with chain length ($J_{\text{ideal+c}}$ curve in Figure 3.13).

The combination of hinge bending and roll-twist coupling has a more marked effect on loop formation. The energies of closed chains with these attributes are substantially lower than those of both ideal DNA and DNA capable of hinged, independent twisting and bending. Compared to the simpler models, the derived *J* factors increase by more than 20-fold in short, difficult-to-close chains and between three- and sixfold in more easily closed chains ($J_{\text{hinge+c}}$ curve in Figure 3.13). This more 'realistic' treatment of DNA makes it easier for the molecule to attach to the Lac repressor in less torsionally stressed, parallel orientations and thereby expands the variety of less easily closed loops. The large enhancement in looping stems from the more pronounced effect of roll-twist coupling on the range of angular deformations in the hinged model, *i.e.*, the likelihood of more extreme deformations is greater than that in a chain subject to isotropic bending or roll-twist coupling alone (Figure 3.14).

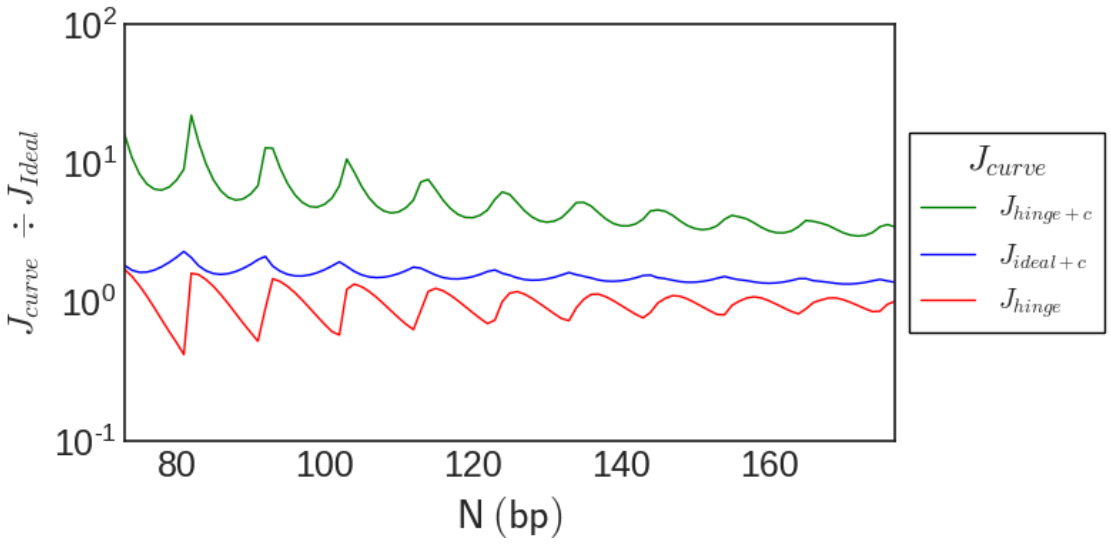

Figure 3.13 'Realistic' treatment of DNA enhances the formation of Lac repressor mediated loops. Relative J factors of DNA loops capable of hinge bending (via roll) and/or coupled (+c) deformations of roll and twist compared to those of an ideal DNA chain subject to isotropic bending and independent deformations of roll and twist.

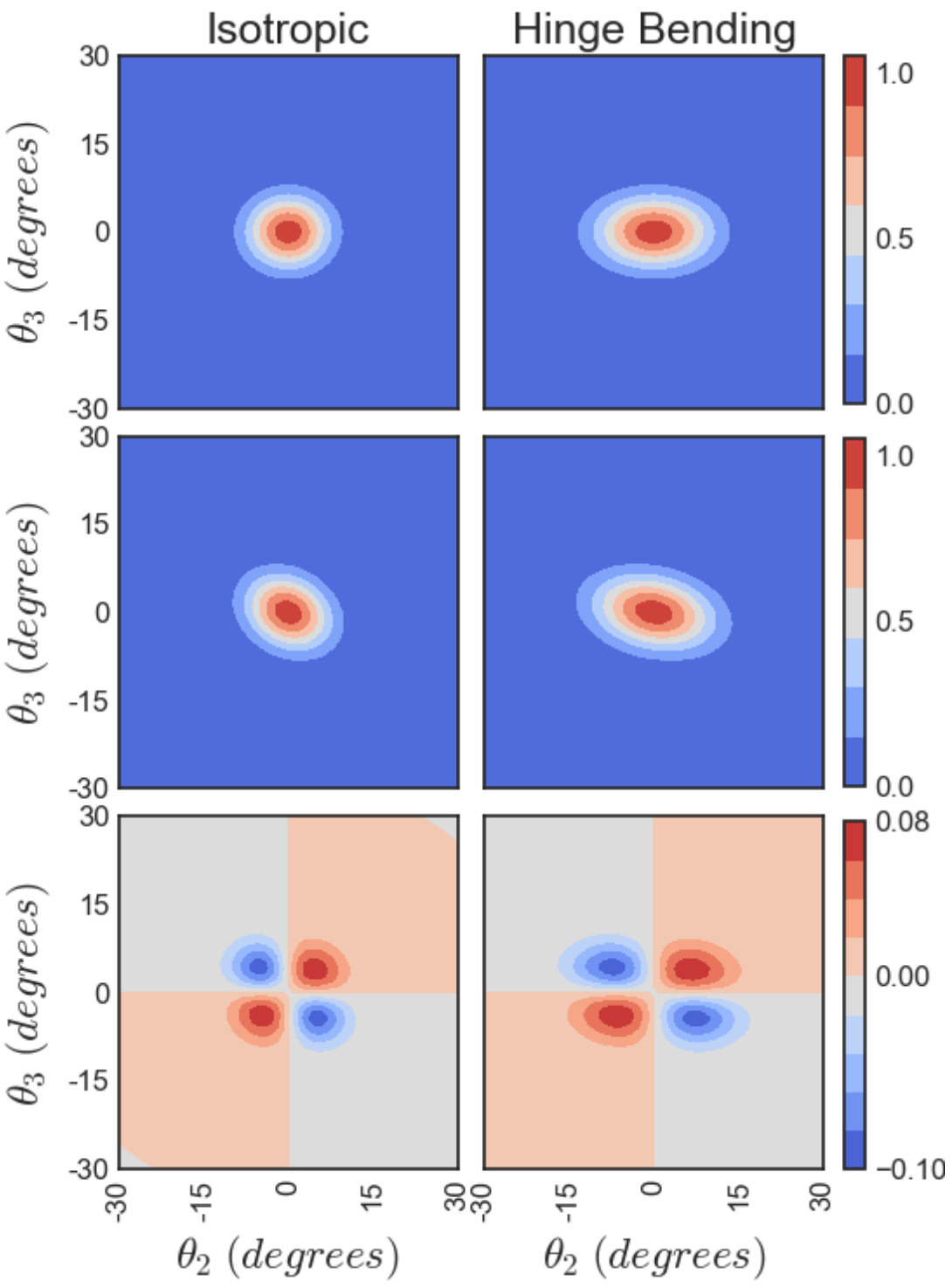

Figure 3.14 Probabilities of roll-twist deformation $w(\theta_2\theta_3)$ in DNA models subject to roll-twist coupling and isotropic (ideal) or hinge bending (middle surfaces) and the corresponding differences in probability Δw (lower surfaces) compared to chains of the same type free of coupling (upper surfaces). Probability values are normalized such that the total probabilities in the upper surfaces are unity and in the lower surfaces zero.

3.2.5 Bending Constraints and DNA Looping

The incorporation of sequence-dependent DNA curvature has especially striking effects on the computed ease of loop formation. The looping propensities estimated from the statistical weights of optimized 156 bp loops with different placements of the same 80 bp curved insert, exceed that found for an ideal, intrinsically straight DNA chain of the same length by 2-4 orders of magnitude, depending upon the location of the insert (Figure 3.15). Here the free insert, the sequence of A-tracts used in the design of stable Lac repressor-mediated DNA loops[14,16] adopts a roughly planar arc of ~150° based on a set of roll and twist angles, described in Chapter 2, that account for the experimentally observed magnitude and direction of A-tract bending.[62] The site of the insert, initially centered at the midpoint of the loop, here denoted by 0, and successively shifted by –2 bp increments toward the 5′-end of the DNA (Figure 3.17, Figure 3.18 and Appendix A), influences the predicted topology of looping, albeit somewhat differently from the populations of loops extracted from fluorescence resonance energy transfer studies.[53] The computed looping propensities and topologies remain nearly the same if the DNA adopts a uniform 10.5 bp/turn intrinsic helical twist instead of the set of sequence-dependent twist angles[63] incorporated in the curved-DNA model (Figure 3.13, Figure 3.15).[62]

The presence of the curved insert perturbs the configurations of protein-mediated loops compared to those adopted by the ideal chain (Figure 3.17, Figure 2.2). For example, the pathway connecting successive base pairs along the A_1^{F1} configuration favored by 156 bp loops with a central curved insert deviates from the corresponding pathway of ideal DNA by 22 Å on average. The optimized configurations of loops of the same construct closed in A_2^{F1}, P_1^{F1}, and P_2^{F1} orientations similarly differ from their ideal DNA counterparts by 11, 138, and 115 Å, respectively. The latter deviations reflect large global rearrangements of the P_1^{F1} and P_2^{F1} pathways induced by the insert rotations of the loops as a whole, by ~180° and ~90°,

respectively, relative to a loop of ideal DNA subject to the same anchoring conditions. In general, the predominant low-energy configurations of the insert-bearing loops resemble their ideal DNA counterparts more closely that do the higher energy forms, where the average displacement of base-pair centers is 30 Å or more from the pathway of the ideal loop. The curved inserts also tend to deviate less strongly from the imposed rest state than do the corresponding stretches in the ideal loops. Thus, the bending energies of the constructs are generally lower than those of ideal DNA under the assumption that the curved inserts are governed by the same force constants as naturally straight DNA.

The precise setting of the curved insert controls the proportion of loops anchored in specific orientations on the V-shaped assembly (Figure 3.15). Symmetric placement of the fragment introduces torsional and bending stress in the semicircular P_2^{F1} configuration preferentially adopted by an ideal DNA loop of the same length, thereby suppressing the occurrence of this closed form (Figure 3.16). The same positioning of the insert, however, lowers the cost of DNA attachment to protein in other orientations. Although the reduction in bending energy is greater for P_2^{F2} loop formation, the lower torsional stress associated with antiparallel loop formation leads to the predicted dominance of A_1^{F1} loops. The propensity for A_1^{F1} looping becomes even stronger with a –2 bp shift of the insert, which lowers both the bending and the twisting energies. Further displacement of the insert reduces the bending and torsional stress within the parallel loops, leading to mixes of loops dominated respectively by A_1^{F1} and P_2^{F1} forms and by A_2^{F1}, P_1^{F1}, and P_2^{F1} forms in the –4 and –6 bp shifted states. A –8 bp shift of the curved fragment favors almost exclusive occurrence of A_2^{F1} looping, where the costs of bending and twisting attain low values. In contrast to ideal DNA, where there is no preferential mode of antiparallel loop attachment, the placement of the curved insert dictates the (A_1 vs. A_2) binding orientation in the designed sequences.

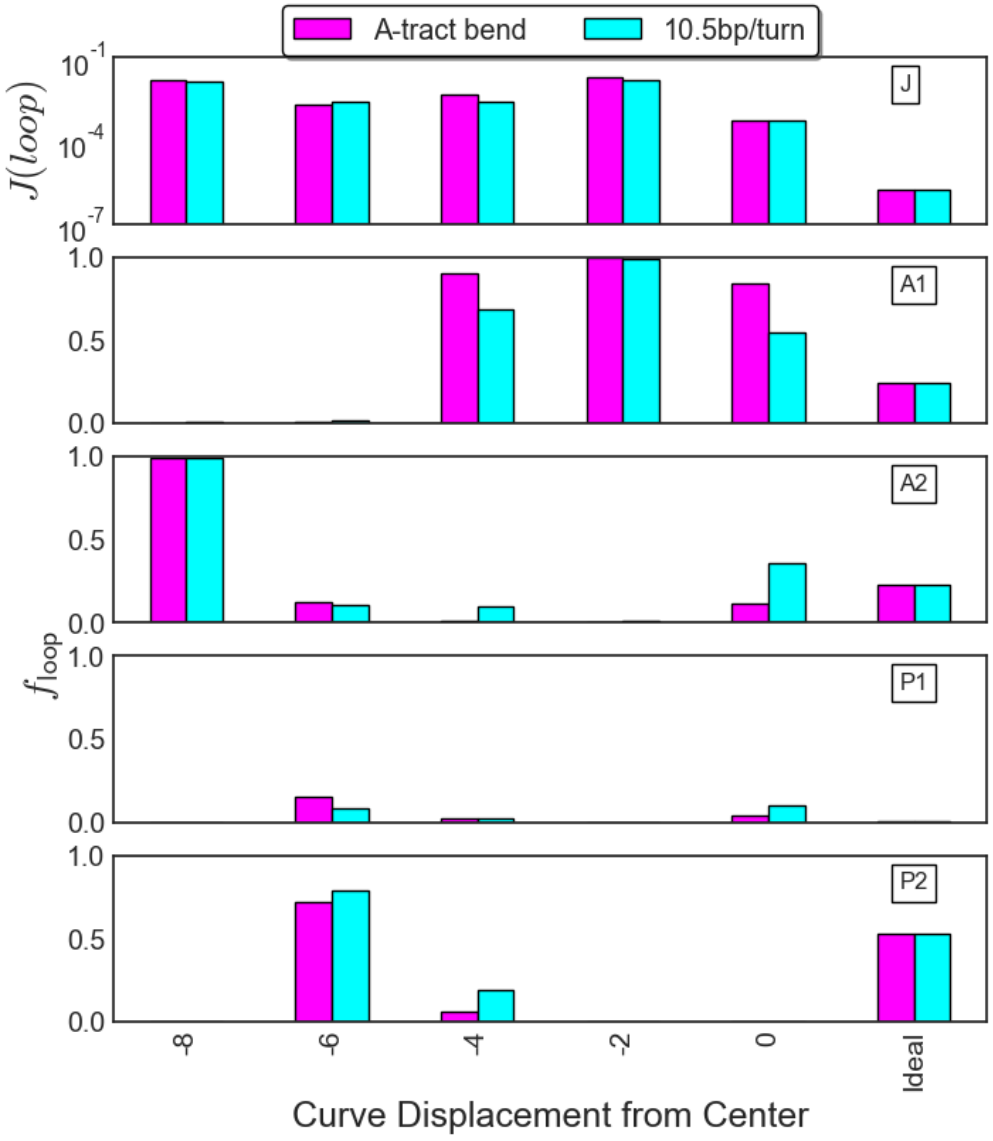

Figure 3.15 Looping propensities for curved inserts: Introduction of a curved insert perturbs the looping propensities and configurations of protein-mediated loops compared to those adopted by ideal DNA. Derived J factors (statistical weights) and orientational frequencies of 156 bp loops with different placements of a common 80 bp, planar curved insert compared to those of an ideal, naturally straight DNA helix with a 10.5 bp repeat. The location of the insert is described in terms of the displacement of its center from the midpoint of the loop (denoted by 0). The rest state of the curved fragment is described by a set of roll and twist angles (see text) that account for known features of A-tract bending (magenta bars) and by a model with the same values of roll and uniform 10.5 bp/turn intrinsic twisting (cyan bars).

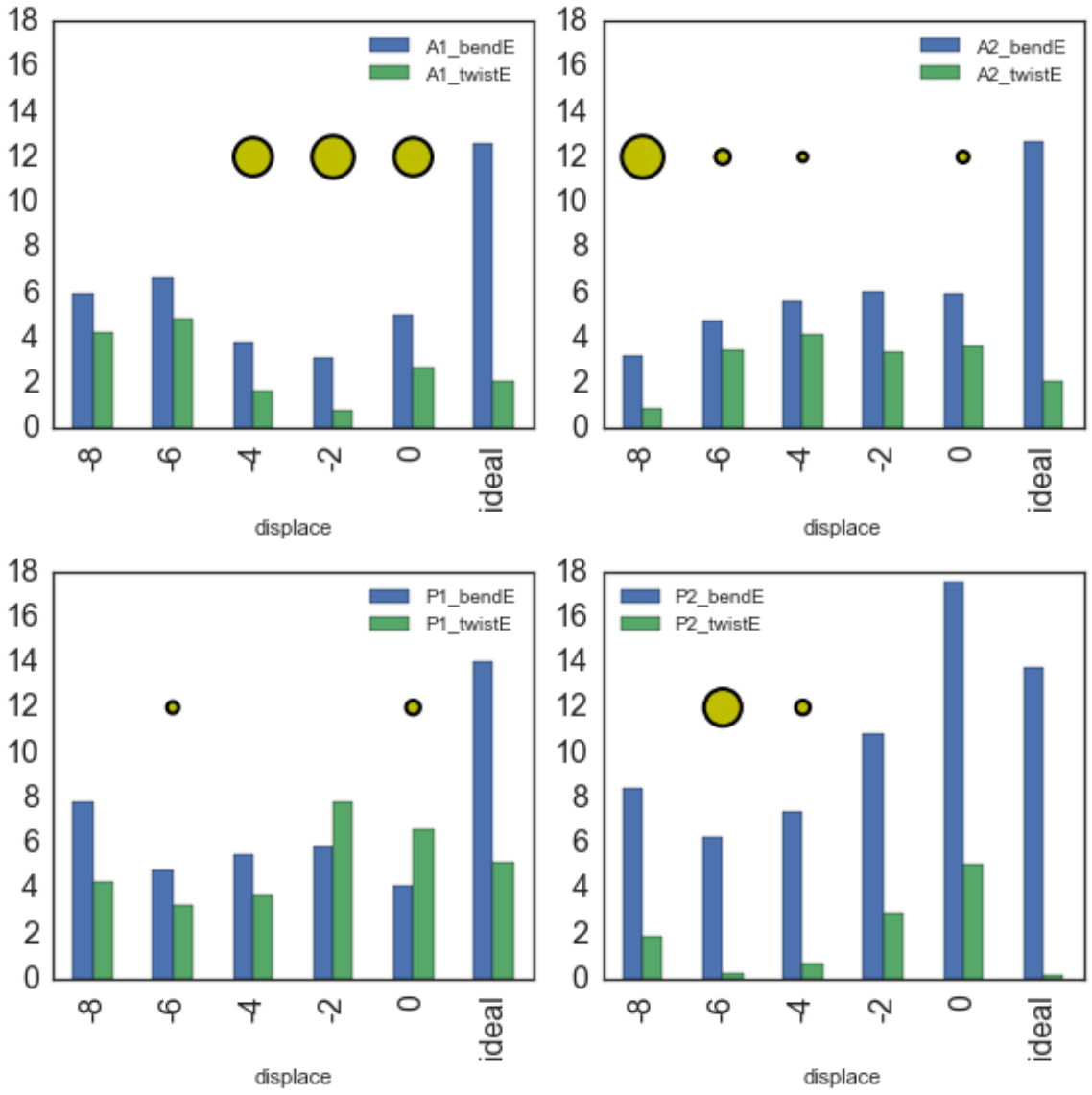

Figure 3.16 Contributions of bending and twisting to the total energy of 156 bp Lac repressor-mediated DNA loops with a common 80 bp, planar curved insert compared to those of an ideal, naturally straight DNA chain of the same length. The location of the insert is described in terms of the displacement of its center from the midpoint of the loop (denoted by 0). The dominant configurations associated with the different types of chains are highlighted by symbols proportional to their frequency of occurrence (see Figure 3.15). Data correspond to the loops of lowest total energy for each binding orientation of the given constructs.

The assumed helical twist alters the preferred orientations of the loops with curved inserts but does not necessarily bring the predicted populations of the looped constructs in line with those deduced from experiment.[16] The choice of helical rest state determines the pitch of the curved insert, *i.e.*, converting the nearly planar reference state formed with a 10.5

bp helical repeat to a superhelical pathway. Four of the seven constructs considered here, the five loops of 156 bp described above, plus loops of 158 and 160 bp with the same 80 bp curved insert respectively displaced by –5 and –2 bp from the DNA center, (Figure 3.17, Figure 3.18 and Appendix A) match the experimental populations very closely over a wide range of imposed values of intrinsic twist, including the 10.5 bp repeat characteristic of B DNA (Figure 3.20). Overtwisting the DNA model to 10.4 or fewer residues per helical turn improves the fit of two other constructs. The predicted build-up of antiparallel configurations of the DNA loop with the symmetrically positioned insert, however, exceeds that captured experimentally. The computations do not consider the attachment of the construct to open, extended forms of the repressor like those hypothesized to account for the observed (low) fluorescence signals.[9]

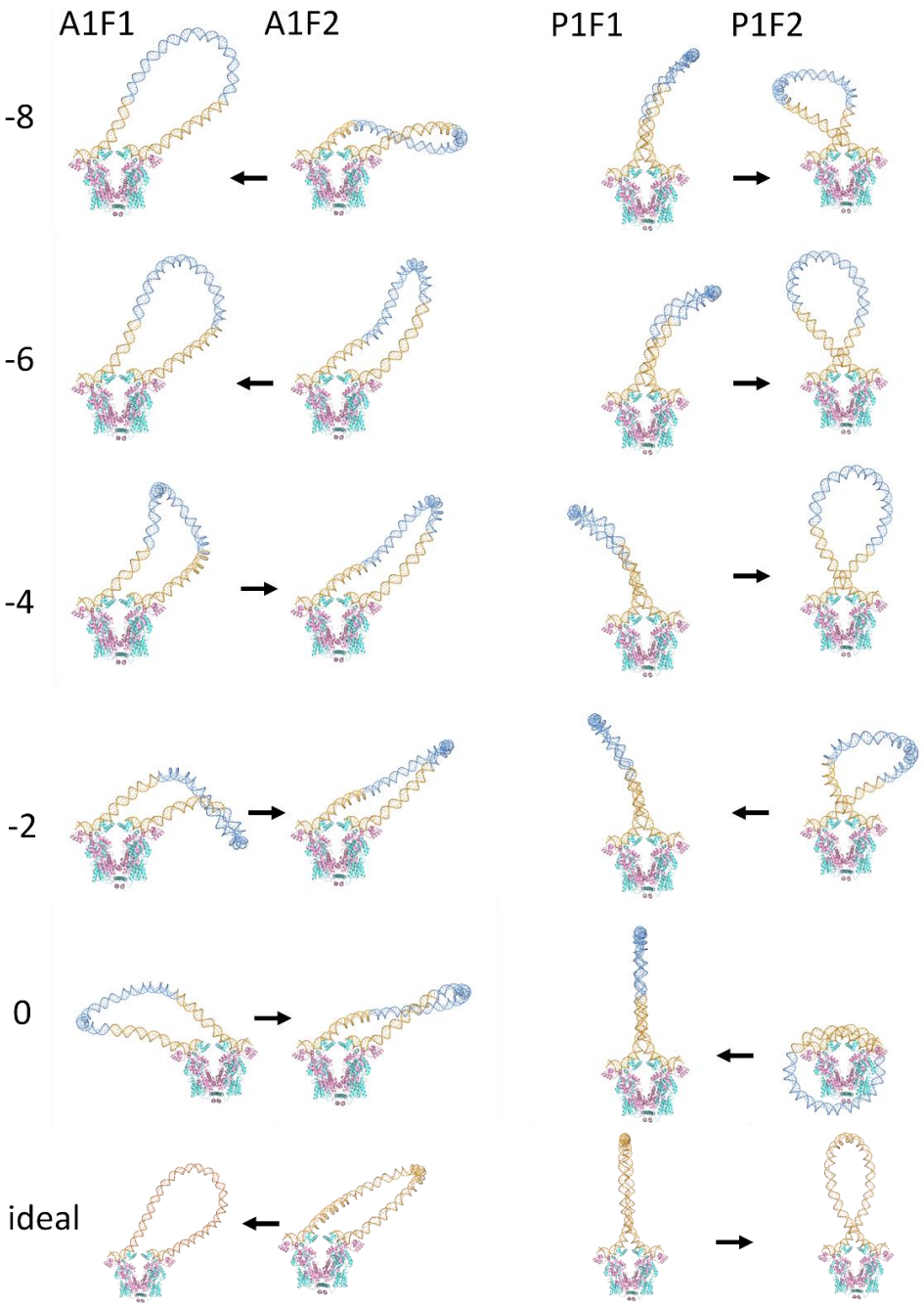

Figure 3.17 Molecular images of A1 and P1 curved insert: A curved insert perturbs the configurations of protein-mediated loops compared to those adopted by ideal DNA. Molecular images illustrating the changes in loop orientation associated with the displacement of the insert (dark blue), at 2 bp increments, from the midpoint (denoted by 0) toward the 5'-end of the loop, and the corresponding pathways of an ideal DNA chain of the same length and looping orientation (images on bottom row). Examples come from two competing structural families of antiparallel and parallel loops (here A_1^{F1} vs. A_1^{F2} and P_1^{F1} vs. P_1^{F2}). The preferred configurations of each orientational type are denoted by the head of an arrow. See legend to Figure 3.5.

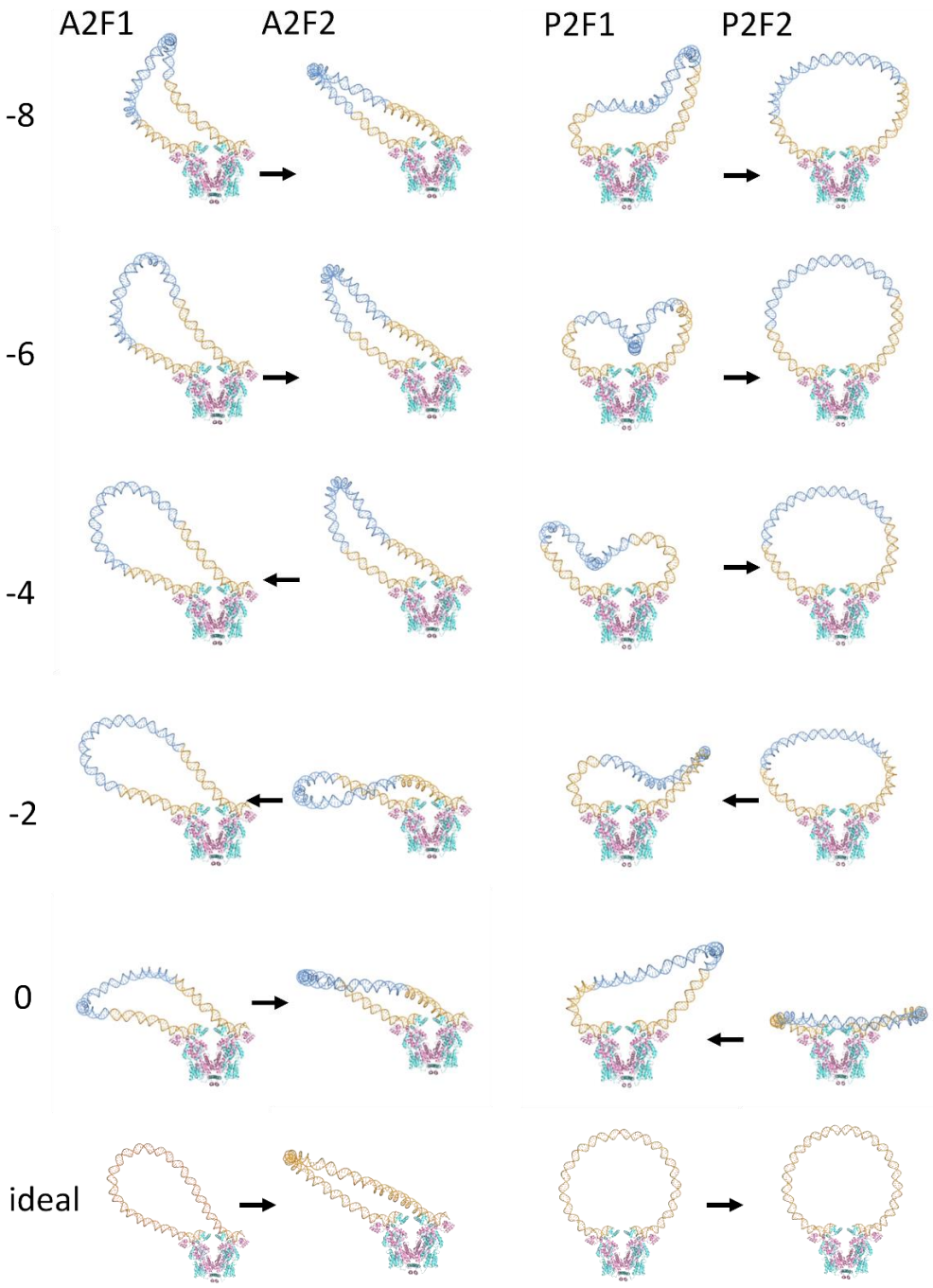

Figure 3.18 Molecular images of A2 and P2 curved insert: A curved insert perturbs the configurations of protein-mediated loops compared to those adopted by ideal DNA. Molecular images illustrating the changes in loop orientation associated with the displacement of the insert (dark blue), at 2 bp increments, from the midpoint (denoted by 0) toward the 5'-end of the loop, and the corresponding pathways of an ideal DNA chain of the same length and looping orientation (images on bottom row). Examples come from two competing structural families of antiparallel and parallel loops (here A_2^{F1} vs. A_2^{F2} and P_2^{F1} vs. P_2^{F2}). The preferred configurations of each orientational type are denoted by the head of an arrow. See legend to Figure 3.5

Curvature overwhelms other sequence-dependent features in the DNA model, including the elastic properties of different types of base-pair steps. Whereas the presence of a curved insert increases the estimated looping propensities of designed 156 bp fragments by several orders of magnitude over that of ideal DNA, added allowance for flexibility in pyrimidine-purine steps and stiffness in purine-pyrimidine steps (compared to purine-purine and pyrimidine-pyrimidine steps) changes the predicted *J* factors by a multiple of 1.5 or less, depending upon the site of the curved insert (Figure 3.19). The influence of the same type of sequence-dependent deformability on loop formation is slightly greater if the DNA is assumed to be naturally straight. Interestingly, the looping enhancement arising from sequence-dependent dimeric deformability diminishes in chains containing a modified $[T_6G_5T_6G_4]_3T_6G_5T_6$ insert with the same intrinsic curvature as the above constructs but with fewer and more widely spaced pyrimidine-purine and purine-pyrimidine steps. The modification introduces GG steps in place of the seven CG and seven GC steps incorporated in the designed inserts (Appendix A). Although these 'mutation' sites absorb the largest structural differences between the two looped sequences, the major differences in energy (and hence the underlying basis of the inferred looping propensities) sometimes accumulate at the ends of the anchored duplexes (Figure 3.21, Figure 3.22). The CG→GG change that precedes each tract of T·A base pairs consistently reduces the local energy and the GC→GG change that follows each tract increases it. The precise magnitude of the energy changes and the directions of the responsible structural deformations depend upon the site of the insert.

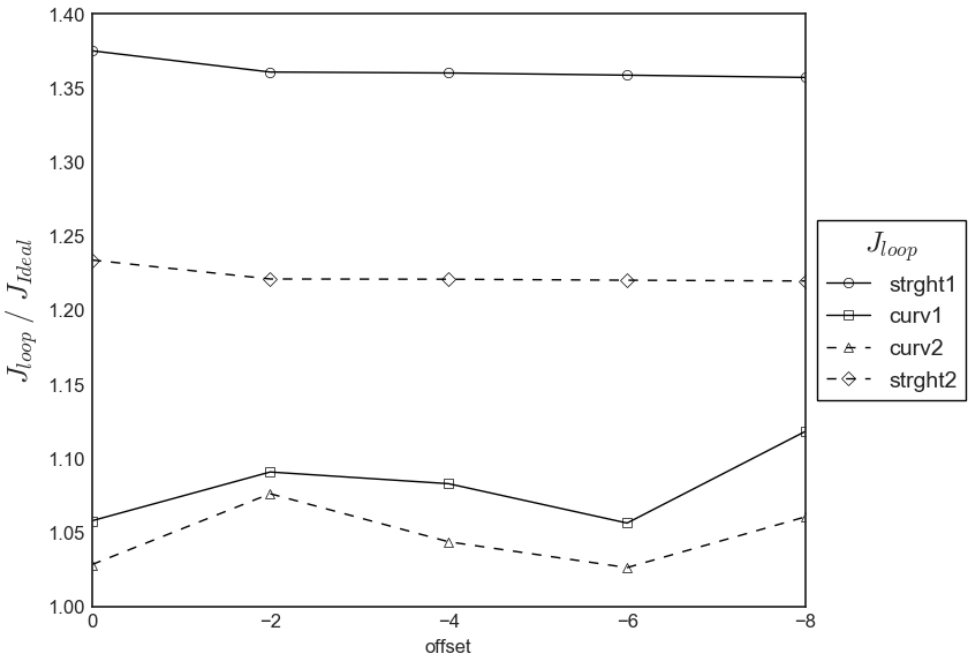

Figure 3.19 Allowance for sequence-dependent deformability has a limited effect on the looping propensities of designed loops . Relative enhancement of the J factors of 156 bp DNA loops associated with the natural flexibility of pyrimidine-purine (YR) steps and the stiffness in purine-pyrimidine (RY) steps compared to RR and YY steps in four series of modeled chain sequences: sequences with the curved insert shown in Figure 3.15, Figure 3.17 and Figure 3.18 to control loop topology (curv1); modified sequences with redesigned inserts of the same intrinsic curvature but containing fewer and more widely spaced YR and RY steps (curv2); naturally straight DNA models (strght1 and strght2) with YR, RY, RR, YY steps in the same positions as the respective designed and modified sequences. Values correspond to the ratios of the J factors (statistical weights) found in curved and straight chains with sequence-dependent deformability compared to chains of the same type with uniform deformability. Here the rest state of the curved fragment assumes the set of roll angles (see text) that account for known features of A-tract bending and the intrinsic twist is fixed at 10.5 bp/turn

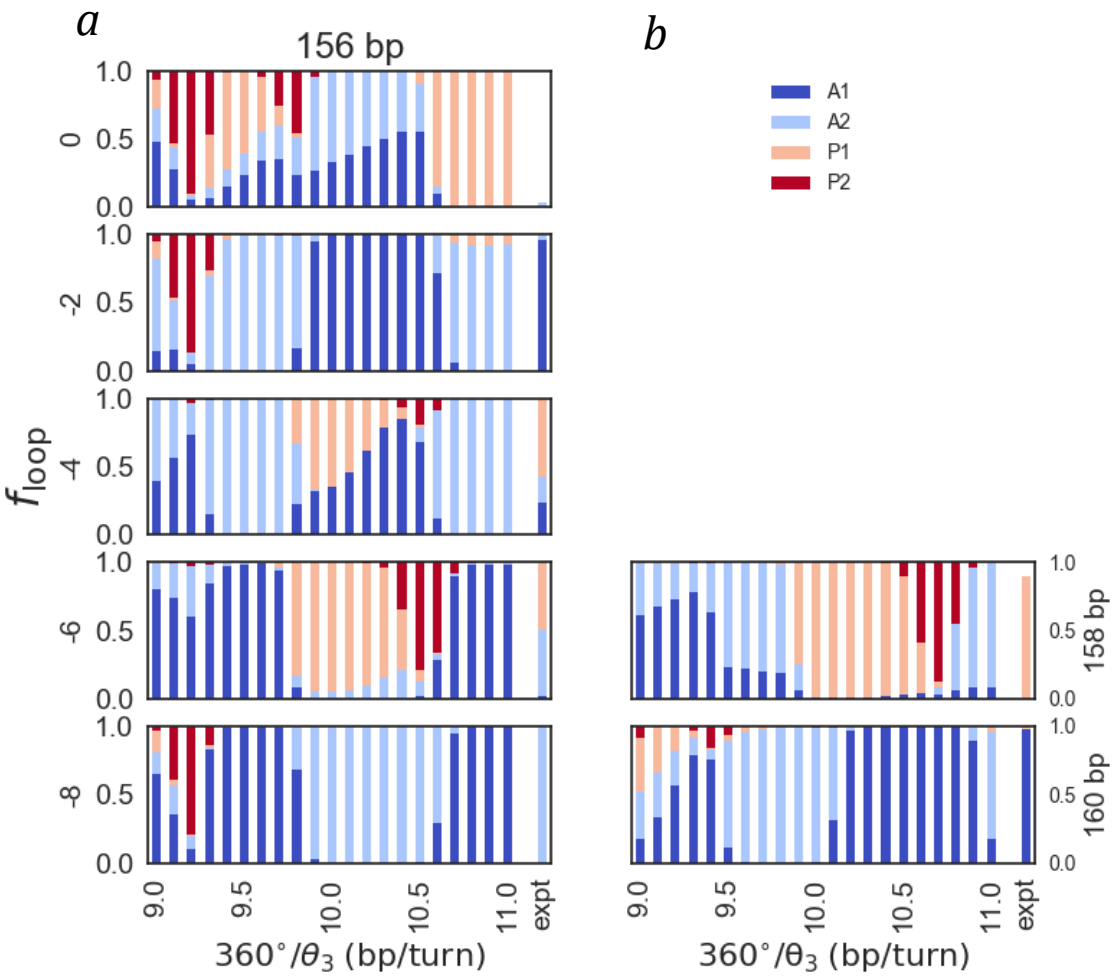

Figure 3.20 Predicted frequencies of curved inserts Comparison of the predicted orientational frequencies of Lac repressor-mediated loops with the observed configurations (expt) of designed DNA constructs bound to the protein assembly.[16] *Models allow for variation in the (sequence-independent) intrinsic DNA helical twist from 9-11 bp/turn and different placements of a common, 80 bp planar curved insert with respect to the centers of loops of different lengths. Insert sites (a) located in the middle and displaced by –2, –4, –6, –8 bp on a 156 bp loop; (b) displaced by –5 bp on a 158 bp loop, and by –2 bp on a 160 bp loop (see Appendix A).*

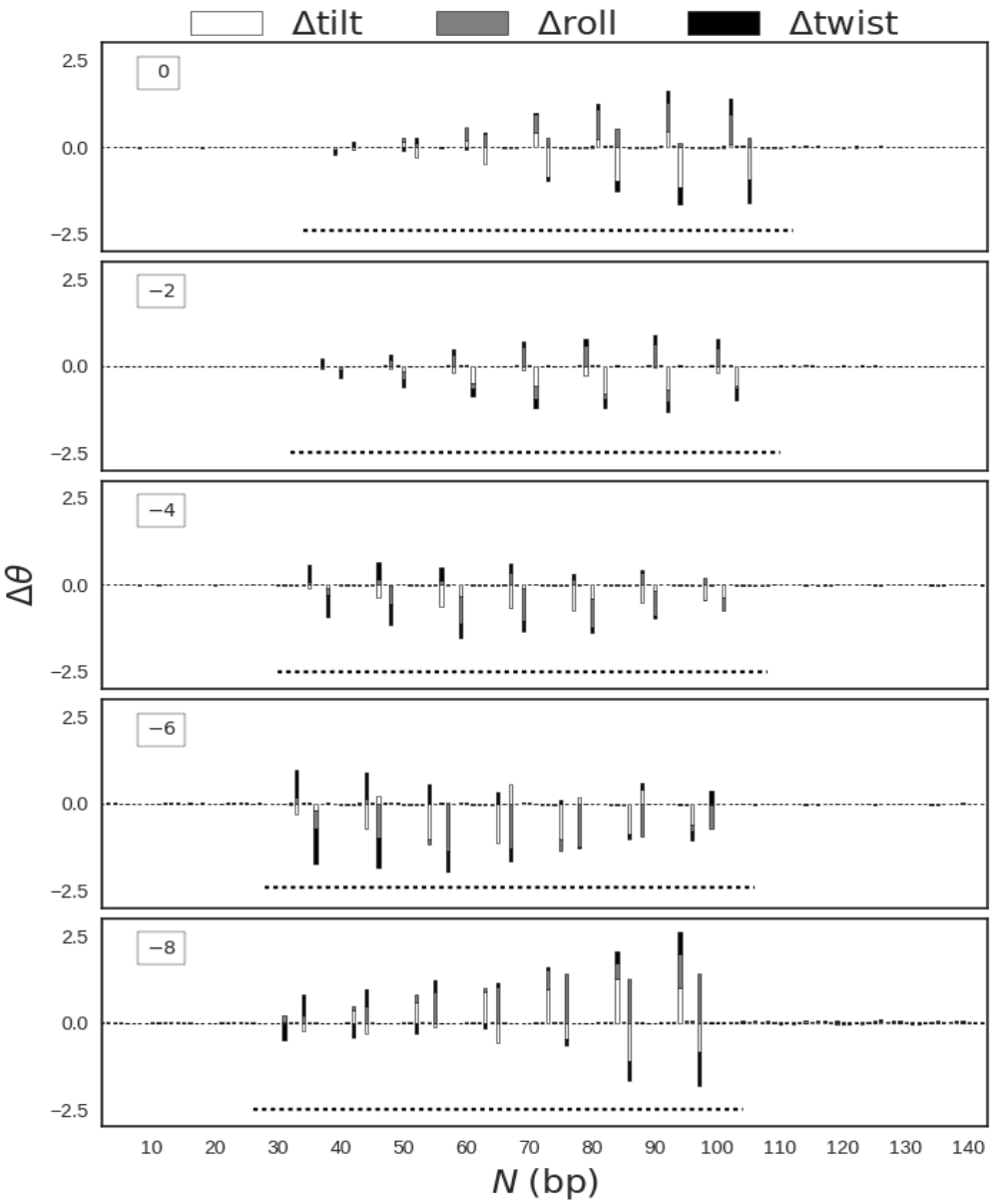

Figure 3.21 . Differences in the bending and twisting of individual base-pair steps in a series of antiparallel Lac repressor-mediated DNA loops bearing a designed, naturally curved insert (see Appendix A) and a 'mutated' $[T_6G_5T_6G_4]_3T_6G_5T_6$ insert with the same intrinsic curvature but with fewer and more widely spaced pyrimidine-purine and purine-pyrimidine steps. The plotted values depict the changes in orientation of successive base pairs in the modified sequence relative to that of the original. The dotted line indicates the position of the curved insert within the loop. The treatment takes account of DNA sequence-dependent deformability, with the pyrimidine-purine steps softened and the purine-pyrimidine steps stiffened relative to all other steps in the two sequences. Note the influence of the location of the insert described, as above, in terms of the displacement of its center from the midpoint of the loop on the magnitude and sites of largest structural change.

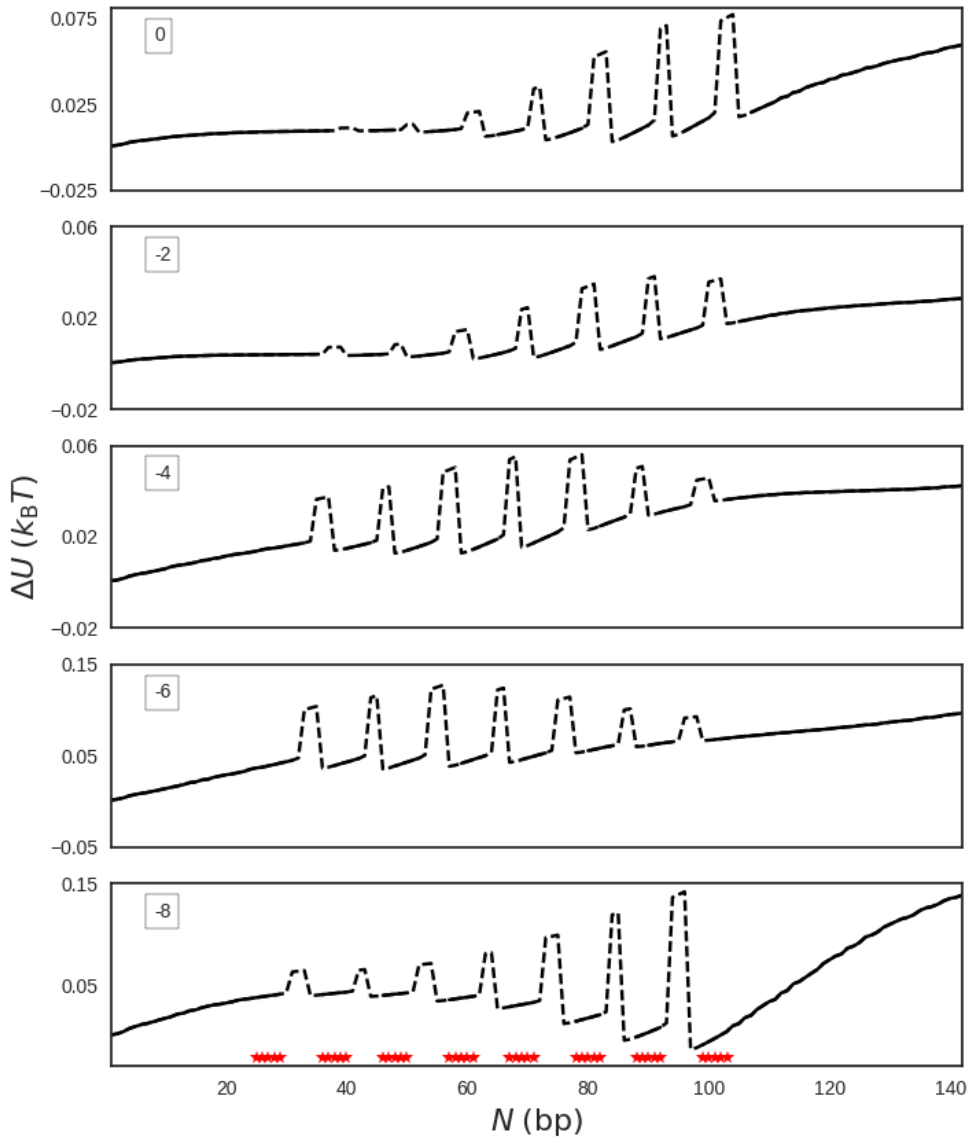

Figure 3.22 Build-up of the change in potential energy , $\Delta U = U^{mutated} - U^{original}$, at successive base-pair steps, associated with the 'mutations' of the looped DNA chains described in Figure 3.21 . Note the sharp rise in energy accompanying each GC→GG modification and the dip in energy accompanying each CG→GG modification. Red stars at the bottom of the lowest plot indicate location of AT base pairs.

3.2.6 Concluding Remarks

3.2.6.1 Bending vs. twisting.

Treatment of protein-mediated DNA looping as an energy-optimization problem makes it possible to decipher the competition among the many factors that contribute to the responses of short pieces of DNA to imposed spatial constraints. The major impedance to DNA loop formation lies in the large-scale bending of the double helix, the energy of which dominates the total elastic energy. The cost of bending depends upon the intrinsic structure of the DNA, the manner in which the DNA attaches to protein, and the length of the constrained insert. Short pieces of ideal, naturally straight DNA preferentially form gradual U-turns that anchor to the Lac repressor in an antiparallel orientation rather than close in the tight ℓ-shaped and semicircular configurations required of parallel attachments to the V-shaped assembly (Figure 3.1). The cost of the latter states, although roughly double those of the antiparallel attachments in very short loops, starts to match and fall below the energy of antiparallel loop formation when the loop spans about 140 bp of DNA and the magnitude of the bending energy has dropped by a factor of 2-3. Thus, at certain longer chain lengths, the DNA may preferentially anchor to protein in parallel orientations. As noted previously,[37] simulated loops of 401 bp, with the natural spacing of the *Escherichia coli* O_1 and O_2 *lac* operators, attach to the repressor with nearly equivalent ease in all four orientations. As evident from the differences between the *J* factors of DNA loops estimated from minimum-energy states and those determined by direct Monte-Carlo sampling (Figure 3.1), treatment of the looping propensities of long DNA chains requires consideration of large-scale motions, such as the fluctuations/normal modes of minimum-energy structures.

Although a much smaller component of the total energy, the twisting contribution determines the types of loops that dominate the configurational landscape at a given chain length. Torsional stress builds up more quickly with variation in chain length than does the

cost of DNA bending, and the differences in the twist energies between looped pathways at certain chain lengths rival the differences between the bending energies of parallel vs. antiparallel attachments to the Lac repressor. Thus loops anchored in more highly bent parallel orientations build up at chain lengths where antiparallel configurations are torsionally stressed (Figure 3.5, Figure 3.6). Torsional stress also builds up upon variation of the intrinsic DNA helical repeat in loops of fixed chain length and controls the balance of states anchored to protein in antiparallel vs. parallel orientations, *e.g*., a 73:27 mix if the DNA in a 92 bp Lac repressor-mediated loop has a 10.5 bp repeat and almost exclusive antiparallel looping (99.99%) if the repeat is 10 bp per helical turn (Figure 3.8, Figure 3.9). Variation in the end conditions imposed on the DNA loop introduces similar effects. The magnitude of the elastic energy and the relative proportions of looped states in torsionally stressed states, however, differ in chains of fixed length from those found in longer or shorter chains subject to the same orientational constraints (Figure 3.11). The different uptake of torsional stress in parallel vs. antiparallel loops contributes to the changing mix of looped states found in loops differing in length by multiples of a helical turn.

The influence of torsional stress on the mix of loops depends as well on the assumed ease of DNA bending vs. twisting. The current choice of force constants favors bending over twisting. That is, a deformation in bending gives rise to an increase in energy smaller than a change of the same magnitude in twist. Variation of the ratio of elastic force constants γ over the range of values suggested by assorted experiments[39,40] (0.7-1.7) changes the mix of loops at certain chain lengths, *e.g*., an 87:13 mix of antiparallel and parallel states in a 92 bp loop when γ is 0.7 and a 1:99 mix when γ is 1.7. The choice of γ, however, has little, if any, effect on the predominance of antiparallel protein attachments when the ends of the looped DNA lie in close register with the protein-binding headpieces.

The subtle influences of the DNA model on the predicted ease of loop formation hint of the way in which base-pair sequence might modulate the closure propensities of DNA. For example, the choice of intrinsic helical twist, one of the key sequence-dependent structural features of DNA,[80] introduces a phase shift in the looping profiles (Figure 3.9). That is, the looping tendencies peak at chain lengths differing by ~5 bp in models with the 10 bp/turn helical repeat characteristic of long stretches of A·T base pairs compared to models with the 10.5 bp/turn repeat closer to that of G·C stretches. The 1-2 bp phase shifts in loop formation detected in recent single-molecule studies of two sets of DNA sequences, one a series of molecules with closely spaced 4-9 bp stretches of A·T base pairs and the other a series of molecules of high (~60%) G+C content[41,47,52] are suggestive of such changes in helical twist, under the assumption that there are no accompanying sequence-dependent changes in DNA deformability and structure. The few experimentally characterized systems, however, precludes more detailed interpretation of the contributions of sequence to loop formation.

3.2.6.2 Looping transitions

DNA loops with the natural 92 bp center-to-center $O_3 \cdots O_1$ *lac* operator spacing lie at one of the critical points on the Lac repressor-mediated configurational landscape, where the accumulation of torsional stress results in a mix of closed structures and small variations in the DNA model lead to sharp changes in the spatial pathways. A 1 bp increment in loop length can shift the population of looped states to a subset of configurations as can a small reorientation of ends of the loop (Figure 3.11). The DNA is undertwisted relative to B DNA in roughly half of the 92 bp loops and overtwisted in the other, with the 78 bp protein-free DNA segments respectively taking up ~7 turns of helix with ~11 bp/turn or ~8 turns of helix with 10 bp/turn. An increase in loop length favors the overtwisted forms and a decrease the undertwisted forms. The accompanying reduction in torsional stress leads to the nearly exclusive uptake of a specific family of antiparallel loops on either side of the transition point.

More 'realistic' treatments of DNA reveal similar looping propensities. Allowance for hinge bending, *i.e*., preferential bending of the double helix into the major and minor grooves, and the coupling of this bending to the local twist (Table 3-1) have relatively limited effects on the optimized energies and derived *J* factors (statistical weights) of Lac repressor-mediated loops compared to those found in the absence of these contributions (Figure 3.13). The added terms, however, alter the populations of torsionally stressed loops and sharpen the transition between different families of looped configurations, particularly for short DNA. For example, hinge bending enhances the proportion of loops with the wildtype 92 bp spacing in overtwisted configurations and lowers the fraction in undertwisted states compared to the roughly even mix of under- and overtwisted looped forms adopted by an ideal DNA chain of the same length. The redistribution of states combined with their differences in energy leads to marked oscillations in the ratios of *J* factors with chain length (Figure 3.13). The contributions of coupling, although of much smaller magnitude than the bending and twisting terms, account for the reduction in the energies, relative to ideal DNA, of loops that incorporate this feature and the associated changes in the *J* factors. The introduction of sites of localized stiffness and flexibility also enhances the ease of looping ideal, naturally straight DNA, although limited here to selected sites of deformability within a 156 bp chain (Figure 3.19). The effect is less strong in a chain of the same makeup bearing a naturally curved insert. Investigation of the combined effects of sequence-dependent deformability, hinge bending, and roll-twist coupling on the looping propensities of specific DNA sequences is beyond the scope of the present work.

Earlier studies of optimum loop structures, including work from this group, have pointed to the coexistence of two configurations of different linking number that contribute with comparable likelihood to Lac repressor-mediated looping when the DNA is torsionally stressed and have illustrated the large-scale differences in looping topology found when one

of the forms dominates the configurational landscape.[9,25,77,84–86] The treatment of Zhang et al.[84,85] includes elastic terms that allow for 'wobbling' of the DNA binding headpieces and thereby perturb the angles of DNA loop exit and re-entry relative to the fixed values used in this work. The variation in anchoring conditions leads to differences in the phasing of the predicted looping pathways compared to those reported here. Other discrepancies between the currently predicted elastic energies and literature values for ideal DNA loops attached to the V-shaped Lac repressor assembly (Table 3-3) appear to stem from differences in methodology and/or modeled system. For example, the Perkins group[62,72] models DNA as a smooth space curve in the context of a dynamical rod theory as opposed to our treatment of the molecule at the base-pair level, and the elastic energies previously reported by our group are components of a potential that includes DNA electrostatic interactions.[9,25] The lower energies and the symmetry of under and overwound loops found here (*e.g.*, comparable magnitudes of writhing numbers) may also reflect the direct optimization procedure that we employ. In contrast to previous work, we examine the detailed pathways between configurational states and the influences of different types of torsional stress on chain folding. The studies from the Perkins group approximate the effects of chain length on loop energy in terms of the rotation of bound operators without consideration of the changes in DNA contour length.[77,86] Such treatments miss the build-up of different types of loops, *e.g.*, semicircular parallel (P2) attachments, with increase in chain length reported here.

	Elastic energy (k_BT)				Writhing number	
	This work	Ref. [77]	Ref. [9]	Ref. [25]	This work	Ref. [77]
A_1^{F1}	27.3	26.5	31.8	32.1	–0.08	–0.10
A_1^{F2}	26.9	34.0	38.9	39.1	0.08	0.23
A_2^{F1}	27.3	30.5	32.9	33.1	–0.09	–0.08
A_2^{F2}	27.0	30.0	–	41.6	0.08	0.09
P_1^{F1}	27.0	34.0	38.6	38.8	–0.05	–0.01
P_1^{F2}	46.6	38.0	–	62.5	0.01	0.25
P_2^{F1}	64.4	76.0	–	71.3	0.02	–0.20
P_2^{F2}	28.5	35.5	45.5	45.7	0.10	0.07

Table 3-3 Comparative Elastic Energy Values and Writhing Numbers of Optimized Wildtype 92 bp DNA Loops Anchored in Different Orientations to the Lac Repressor Assembly. Writhing number is associated with a closed pathway obtained adding a straight segment between the two ends of the protein-bound DNA.

3.2.6.3 Sequence dependent curvature.

The profound effect of sequence-dependent curvature on DNA looping arises from a steep reduction in the bending energy of the constrained molecule compared to that of an ideal, naturally straight, mixed sequence chain. The drop in bending energy accounts almost entirely for the predicted 2-4 order-of-magnitude enhancement of the *J* factors of designed 156 bp loops with different placements of a common 80 bp curved insert (Figure 3.15). Certain placements of the curved insert, however, increase the cost of forming certain types of loops compared to their ideal DNA counterparts, with the higher energy stemming in most cases from bending stress. Moreover, the added stress tends to localize at the ends of the less easily closed loops in contrast to the reduction in energy that accumulates preferentially along the curved inserts of the more easily formed loops.

The 80 bp insert typically deforms beyond the assumed level of intrinsic bending within the optimized loops (Figure 3.17 and Figure 3.18). That is, the chain fragment must shorten relative to the imposed rest state in order to anchor between the headpieces of the Lac repressor assembly. The unconstrained insert forms a nearly planar bend of ~20° per helical repeat, consistent with the cyclization and topological properties of curved DNA.[4,25] The distance between the ends of the insert drops to 60-70% of the rest-state extension in the majority of closed loops, *i.e.*, end-to-end differences of 60 Å or more. By contrast, the naturally straight ends of the loops shorten by 90-95% (5-10 Å). The assumed level of intrinsic bending, however, coincides fairly closely with that found within the modeled P_2 loops, where the differences in insert extension are 10 Å or less. Systems that incorporate sharper localized bending of DNA, such as via ligand binding,[87] may suppress the closure of DNA in P2 orientations. In this regard, it is should be noted that the modeled non-planarity of the DNA insert may contribute to the predicted exclusion of P2 looping by the Perkins group.[77,86] The changes in double-helical twist found here to bring the predicted populations of looped constructs in line with experimental findings (Figure 3.17 and Figure 3.18) may also reflect accompanying changes in the imposed superhelicity of the curved insert. The insert tends to retain the imposed superhelicity in the lowest energy loops. On the other hand, 200 bp fragments of kinetoplast DNA bearing repeated A-tracts of the sort modeled here appear to close into nearly perfect circles in electron microscopic images.[88]

The computed closure propensities of the insert-bearing chains capture the looping topologies, *i.e.*, modes of Lac repressor attachment and overall DNA folds, deduced from fluorescence resonance energy studies[16] of six of seven designed sequences with naturally curved segments of the same length. The correspondence between theory and experiment in two of the six cases requires a 1-2° deg increase in the assumed helical twist of the DNA model (a series of local changes comparable to a 75-150° rotation of the free end of an anchored

piece of linear DNA). Our representation of the protein as a rigid V-shaped assembly, however, precludes treatment of a 156 bp loop with low fluorescence signals suggestive of large-scale opening of the repressor. We know from previous work that an ideal DNA chain of this length anchors preferentially to protein in a parallel orientation and assumes an extended, modestly curved configuration if the protein is allowed to undergo simple opening motions.[9,43] Furthermore, knowing that the uptake of architectural proteins on the repressor-DNA assembly limits the degree of repressor opening and changes the mix of looping topologies,[43,89] we expect that the introduction of a curved insert may have a similar influence on the extent and modes of protein deformation. In this regard, Chapter 4 describes methods we have developed to study the influence of specific motions of the Lac repressor on DNA looping and the interplay between the degree and handedness of DNA curved inserts and the level and direction of preferred protein opening.

Chapter 4 Modeling Alternate DNA Loop End Constraints Resulting from Simulated Protein Deformation.

Although the *lac* operon is one of the most heavily studied gene regulatory systems, very little is known about the DNA protein interaction involved. Among the missing links is a sense of the spatial arrangement of the combined DNA operator/Lac repressor complex. Most of the previous studies of this system have focused on either the properties of the DNA loop created by the protein-DNA interaction or sought to explain the possible conformational changes of the protein itself. Previous studies that investigated the DNA loops, including those in the previous chapter, assumed the protein to be rigid or considered only a few possible configurations, thus ignoring the role that protein dynamics may play. The conformational flexibility and the disordered regions of the LacR protein that are crucial to its function hinder the ability to ascertain comprehensive structural and dynamic details of the protein-DNA assembly. The approach presented here considers the combination of changes in the protein and its effects on the induced DNA loop.

4.1 Protein Deformation

The known structures of the Lac repressor and its complexes with small pieces of linear DNA, in both the solid state[4,5,11,28,90,91] and in solution[18,19,92,93], provide important clues to ways that the double helix can attach to the protein and how the protein might deform in response to the binding of small molecules, such as analogs of lactose that affect the affinity of the repressor for DNA.[94] The Lac repressor as discussed in Chapter 1 is comprised of four identical polypeptide chains. The N-terminus of each of the constituent polypeptide chains includes a globular headpiece that attaches to specific base pairs of operator DNA and the C-terminus comprises an α-helix that coils around the corresponding feature in the three

remaining chains. The intervening residues form two globular sub-domains that partner with the same sub-domains of a second chain and form a large 'arm'. The protein assembly thus consists of two dimeric arms linked at one end through a bundle of four coiled α-helices and attached at the other end to pairs of headpieces that bring distant parts of DNA into close proximity (Figure 4.1).

Early structural investigations of protein microcrystals suggested that the repressor might assemble in a fully extended, dumbbell-like arrangement[95]. Such an open form, later supported by small-angle X-ray scattering studies of the free tetramer in solution[66], would place the DNA-binding headpieces at the far ends of the assembly. The protein, however, adopts a V-shaped architecture in the crystalline state with limited variation in the angle of aperture between the dimeric halves of the structure[4,5,11,28,90,91]. On the other hand, the small molecular interface between the two arms of the crystallized protein hints of the potentially low energetic cost of repressor opening, and the long, unstructured peptide linkers that join the arms of the protein to the four-helix bundle suggest a locus of conformational interconversion[11,28].

The unstructured peptide residues in other parts of the Lac repressor structure, notably the hinge region connecting the DNA-binding headpieces to the core of each arm and the short linkers between the globular sub-domains within the two arms, point to other sites of potential protein deformability. The DNA-binding headpieces fail to adopt a globular structure in the absence of bound duplex.[11,28,90,91] The removal of DNA leads to an elongation of the dimer that can be detected in small-angle X-ray scattering measurements[12] and that supports the idea that the protein N-terminus unfolds upon DNA removal[18,96]. The subtle rearrangements of the core subdomains associated with the uptake of IPTG, a molecular mimic of the natural lactose metabolite that triggers transcription of the *lac* operon, have led to further hypotheses about the displacement and unfolding of the DNA headpieces upon the

binding of inducer molecules[90]. The small-angle scattering of operator-bound tetramers, however, provides no evidence of large displacements of the Lac headpieces upon IPTG binding[12].

Images of the Lac repressor captured by transmission electron microscopy reveal a roughly 60:40 mix of open and closed protein structures[13]. The same ratio of states persists in small angle X-ray scattering studies of operator-bound tetramers[12]. The footprinting patterns, electrophoretic mobilities, and fluorescence resonance energy transfer between dyes on designed, highly stable Lac repressor-mediated loops lend further support a two-state model of protein architecture[14–17,97]. The apparent mix of states depends upon the DNA chain length (148-164 bp), the presence of small molecules (such as IPTG), and the precise locations of the 80 bp curved DNA fragments used to enhance formation of these loops. The tethered particle motions of single DNA molecules capable of binding the Lac tetramer also produce signals that can be interpreted in terms of two kinds of loops anchored to open and closed repressor structures[7,8,22,23,48]. The variation in the signal over short times reveals direct interconversion between the two repressor-bound states, and chemical modifications designed to perturb the protein architecture affect the lifetimes and relative populations of the loops. The second looped state, however, occurs at only some chain lengths, e.g., in 158 bp but not 153 bp looped constructs[48]. Interpretation of these data depends upon the assumed orientation of the DNA loop on the repressor[22,48,98] , i.e., the approximate spatial arrangement, either parallel or antiparallel, of the 5′-3′ operator sequences on the protein headpieces[6], and whether or not the DNA wraps on the surface of the repressor[99]. Atomic force microscopy (AFM) images of the relatively long loops of DNA (197 and 402 bp) formed upon binding the Lac repressor to a DNA minicircle support the notion of DNA loops anchored in antiparallel orientations and directed away from a closed V-shaped tetramer assembly[100].

Simulation of the AFM images of smaller loops (153 and 158 bp), however, suggests that the DNA may wrap on the surface of the repressor[48].

Estimates of the free energy of DNA loop formation based on treatments of the double helix as an elastic rod point to conditions under which the opened form of the Lac repressor might persist[9,24,84]. For example, an ~80° increase in the angle of aperture between the arms of the V-shaped tetramer accounts for the DNase I cutting patterns of 52- and 74 bp Lac repressor-mediated DNA loops[9]. Allowance for the same type of protein deformations enhances the predicted likelihood of closure of short DNA loops, especially if the ends of the loop are misaligned, i.e., out of helical phase, with the bound operators[43,101]. Thus, the average degree of repressor opening and the associated mix of loop types are expected to vary with DNA chain length.

The displacement of the Lac repressor headpieces predicted in coarse-grained modeling of the protein dimer[26] surfaces as large-scale rearrangements of the DNA-bound headpieces in a multi-scale atomic-elastic treatment of the tetramer-loop assembly[102]. The simulated turning of the headpieces, ~90° about the long axis of each arm, is comparable to the large-scale reorientation of bound DNA found upon superposition[11] of the core of the dimeric purine repressor structure[103] on that of the Lac repressor.

Although these studies indicate several areas of flexibility, (Figure 4.1), I have focused on the flexible region around the 4 helix bundle tetramerzation region. The procedures developed here provide a framework for investigating other possible protein-deformed configurations of the DNA repressor assembly

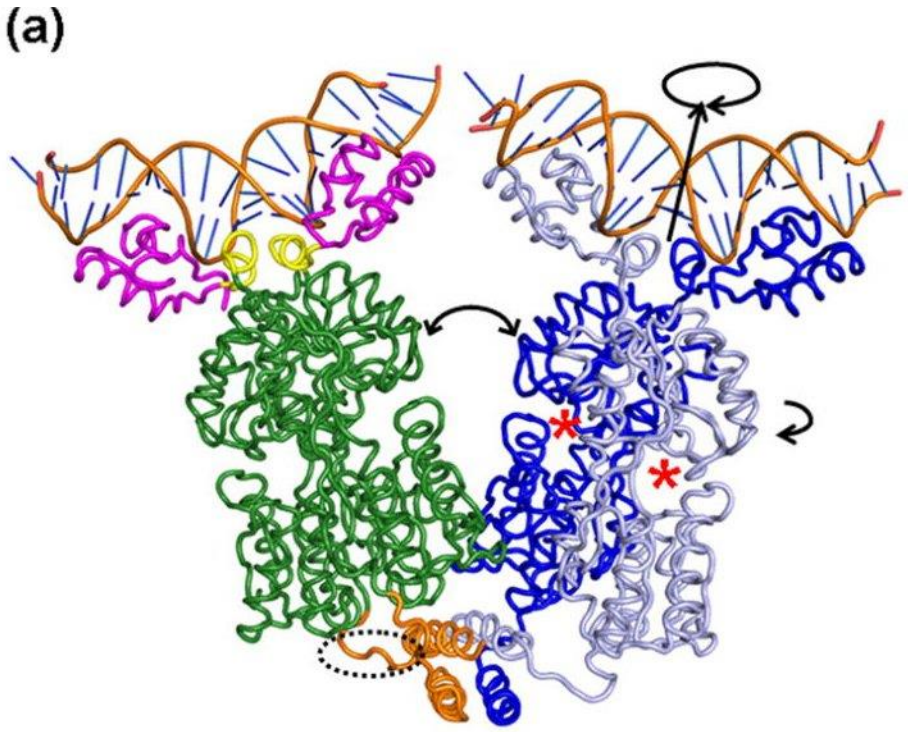

Figure 4.1 Molecular image of Lac repressor areas of interest. Curved black arrows indicate possible areas of protein deformation. These include the opening of the two dimeric arms (denoted by the curved arrows in the middle of the picture), rotation of the DNA binding domain (noted by the arrow on the top of the picture) and the bending of the subdomains of the core regulatory domain (shown by the arrow to the right of the structure). The dotted ellipse identifies an unstructured linker between the tetramerization region and the core domain which is believed to contribute to the deformability of the orientation of the two arms of the protein. The yellow hinge linker is implicated in the variability configuration of the DNA binding domain. The red stars indicate the ligand binding pocket where allolactose or a molecular mimic such as IPTG binds and leads to resulting in an allosteric change that reduce the DNA binding affinity of the protein.

4.2 Computational Approach

4.2.1 Modeling Protein Deformation

The 3D atomic model of the protein-DNA structure used as a starting structure was generated from a compilation of crystal structures of different elements of the entire complex. The common atoms are aligned to produce a complete complex as described in Chapter 2. In order to simulate the deformation of the protein we have identified two modes of flexibility. An opening of the angle between the two arms of the protein assembly and the rotation of one of the arms about the four-helix bundle tetramerization region. The movement of each of the dimeric arms of the Lac repressor is simulated by determining the

central axis of each arm. The axes are identified through by principal component analysis of the atomic coordinates of the atoms in the globular subdomain of the dimer. Since the axes do not intersect a point equidistant to the points on each axis where they come nearest to each other is chosen to rotate the axis. Using this definition of the axes, the crystallized LacR has an opening angle **α** of about 40° (Figure 4.2 and Figure 4.3). We consider values of α between 40° and 200° at 5° increments. The rotation of protein arms in a perpendicular direction is modeled by an angle **β** with values between 0° and 355° at 5° increments (Figure 4.2 and Figure 4.4). To maintain a common reference orientation, one arm of the protein is held fixed and the simulated opening and rotation are applied to the other arm. This provides 2592 combinations of opening and rotating arms configurations. Each configuration provides a new coordinate for the operator segment bound to the Lac repressor. The loop type (A1, A2, P1 or P2) determines which of the base pairs from each of the operators will define the constraints for DNA loop formation in the optimization. This method provides a rich landscape of end conditions at a granularity that cannot be achieved through experimental methods.

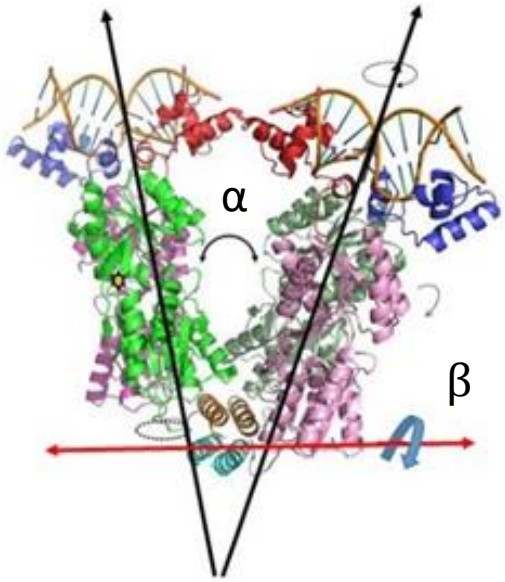

Figure 4.2 Axes of deformation. The long black arrows indicate the central axes of the core regulatory domain of each dimer. The angle α *between the two axes describes the opening of the dimeric arms. The red arrow indicates an axis parallel to the protein around which the central axis of one of the arms is rotated. The rotation is described by the angle* β.

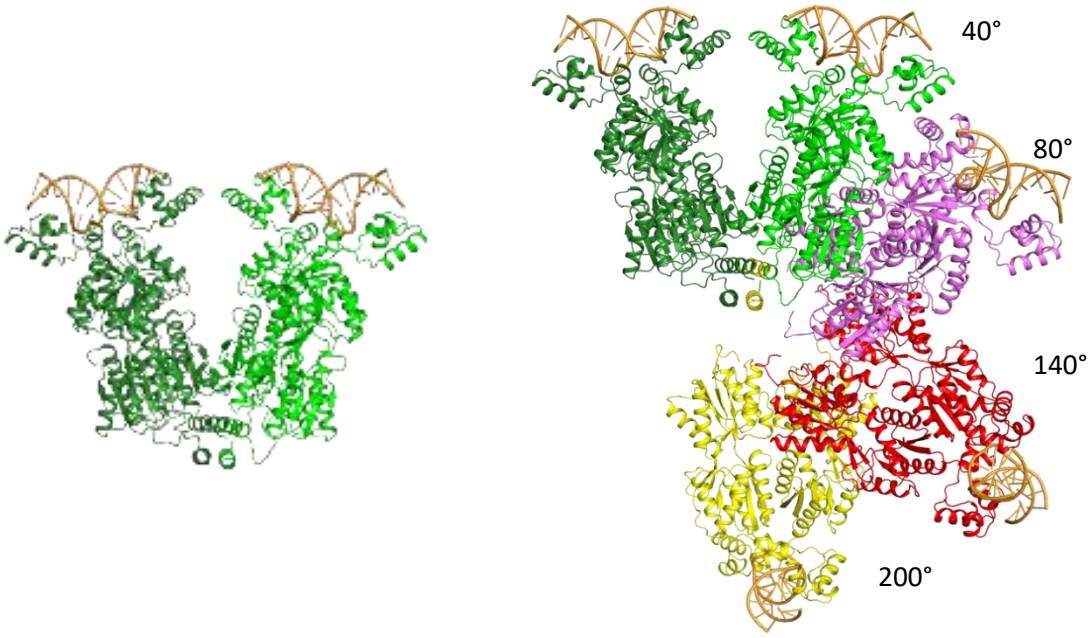

Figure 4.3 Molecular images of LacR opening. Gallery of images illustrating the opening α *between the two dimeric arms of the Lac repressor. The image on the left is the unmodified model with the typical viewpoint looking down the axis perpendicular to "V" formed by the arms of the protein assembly. The dark green arm is held fixed and the coordinates of the light green arm are rotated as described in the text. Examples of reoriented configurations share the dark green arm and the modified arms at three opened states are colored violet, red and yellow.*

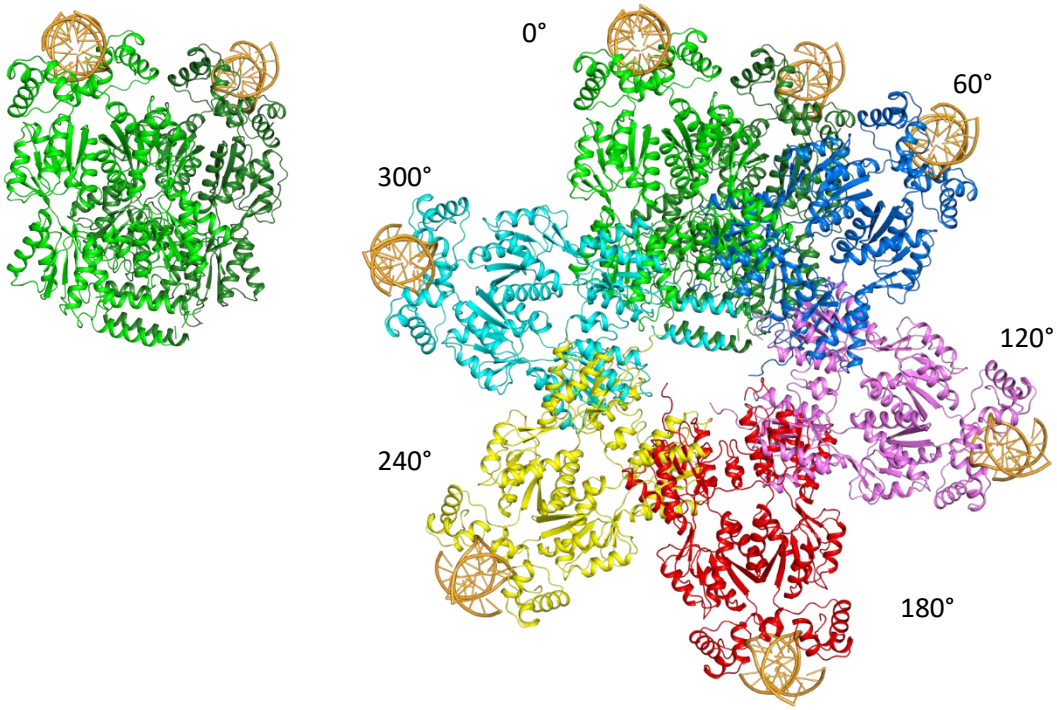

Figure 4.4 Molecular images of LacR rotation. Gallery of images illustrating a perpendicular rotation β *of one arm around the vertex of the "V" formed by the arms of the protein assembly. The image on the left is the unmodified model. For clarity, it is not the usual viewpoint as in Figure 4.3. Instead, the view is looking approximately at the side of the rotating arm in light green and the fixed arm behind in dark green. The dark green arm is held fixed and the coordinates of the light green arm are rotated as described in the text. Examples of reoriented configurations share the dark green arm and the rotated arms in five rotated states are colored blue, violet, red, yellow and cyan.*

4.2.2 DNA Loop

The manipulations to the DNA protein model provide a collection of DNA loop anchoring conditions to estimate the effects of different types of protein motion on DNA loop conformation. The likely pathway and probability of formation of DNA loops constrained by the modified protein model are obtained using the optimization method described in Chapter 2. The non-sequence specific ideal B-DNA force field, which treats the DNA as an inextensible elastic rod with isotropic bending independent of twisting, provides a reasonable first approximation of the optimal loop.

I investigated the effects of protein motion on loops of length 92 bp as in the native *Escherichia coli* DNA and with 5 additional base pairs in loops of 97 bp. The additional bases reorient the operator by one half a helical turn and make it possible investigate the effects of helical phasing. I also investigated the possible contribution of protein fluctuation on loops with the 156 bp length that was studied in the experiments from the Kahn lab. As noted in Chapter 3, the introduction of a curved insert in the center of the DNA construct appears to disrupt the fluorescent signal used by the group to characterize loop topology.

The four loop orientations, A1, A2, P1 and P2 (Figure 1.2) , as well as the two structural families identified in Chapter 3 are subject to optimization using the ideal B-DNA force field. Starting with the loop bound to the rigid "V" shaped protein configuration, gradual increases were introduced in the opening and rotation of the arms of the protein model. The end constraints of the successive states are replaced by those of the anchoring base pair on the arm that is opened or rotated 5° from the previous optimized configuration. The 3′-terminus of the constrained DNA fragment is thereby rotated through a series of intermediate states that mimic the reorientation of the protein-bound DNA operator by the protein deformation.

4.3 Results

The results from the optimizations provide insight into the role that protein deformation plays in protein-DNA loop formation. The statistical weight of the estimated free energy for each of the loop types describes a fractional population of that loop type. The sum of the statistical weights is used to approximate the J factor for each of the protein-DNA configurations.

4.3.1 Independent Opening (α) vs Rotation (β)

A slight opening of the angle α between the two dimeric arms reduces the elastic energy required to form loops for all of the lengths studied (Figure 4.5). At all lengths the J

factor initially increases upon opening the protein then quickly falls as opening is increased. The J factor for the 156 bp loop increases to a much greater extent but still falls off as the opening continues. The 92 bp loop configuration initially consists of a mixed population of loops but is quickly reduced to only P1 loops and finally to only A1 loops as the J factor drops upon further opening. The 97 bp length population gradually changes from a mix of A1 and A2 loops to only A1 loops and lower looping propensity as the protein opens. The 156 bp loops evaluated initially consist of a mixed population of A1, A2 and P2 loops. Opening of the protein shifts the P2 portion of the population to P1 and back to P2 as the protein becomes fully extended.

Rotation has the opposite effect on looping probability. The angular rotation β of one arm about the 4 helix bundle initially reduces the J factor then it begins to climb (Figure 4.6). The populations are more mixed when the probability of looping is small. The populations move from mostly mixed to a predominantly anti-parallel (A1 and A2) population as β increases. Rotation is more sensitive to sequence length with oscillating J factors.

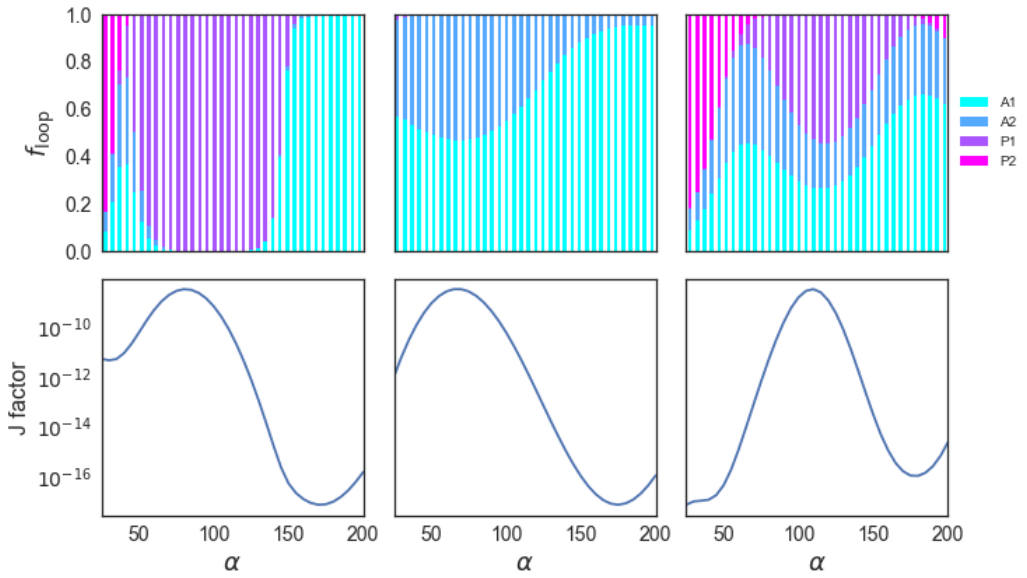

Figure 4.5 Fractional populations of loop types and J factors as the arms of the protein open, i.e., α varies and β is constant at 0° and in the

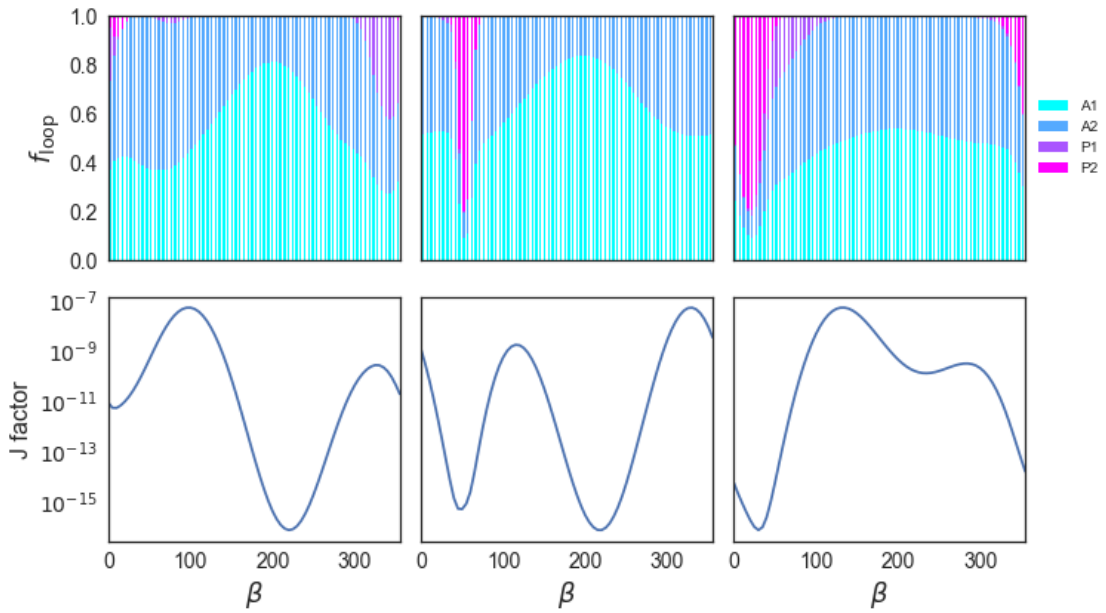

Figure 4.6 Fractional populations of loop types and J factors as the arm of the protein rotates.

4.3.2 Coupling of α and β

As the arms of the tetramer are relocated, either by opening or rotating, the distance between the two headpieces is increased (Figure 4.8). It is not surprising then that there is a correlation in the looping population generated by opened or rotated configurations of the protein. That is, an increment of 5° in the opening angle leads to a similar population with a comparable increment in the rotation.

The length of the loop determines which protein configuration is most likely to accommodate loops. The shortest, 92 bp loop, has the highest J factors for configurations where the protein is open between 80°-100° coupled with a rotation in the range 130°-160°. In this case, the population is a mix of orientations of A1, A2 and P1 loops, although there is only one structural family represented by each orientation. The 97 bp construct has a different sweet spot, with the protein fully extended with an opening between 165°-185° and rotated ¾ of the way around the vertex between 260°-280°. Here the population is dominated by the A1F1 structural family. Finally, the longest 156 bp loops seem to prefer

very little opening of the protein, in the area of 45°-60° and a rotation 130°-160°. The population is equally favored, with A1 and A2 loops equally favored, and, like the others only one structural family from each is represented. It should be noted, at 156 bp, the J factors for all of the protein configurations are higher than those for any of the configurations of the 92 bp loops.

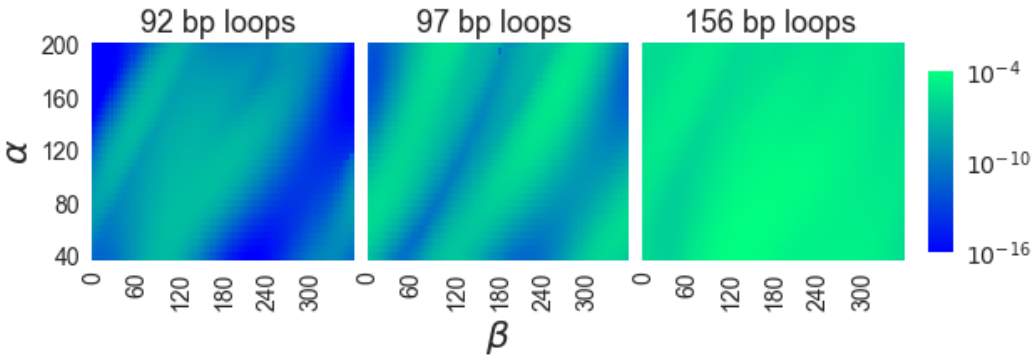

Figure 4.7 Estimated J factor for loops of several lengths on deformed proteins

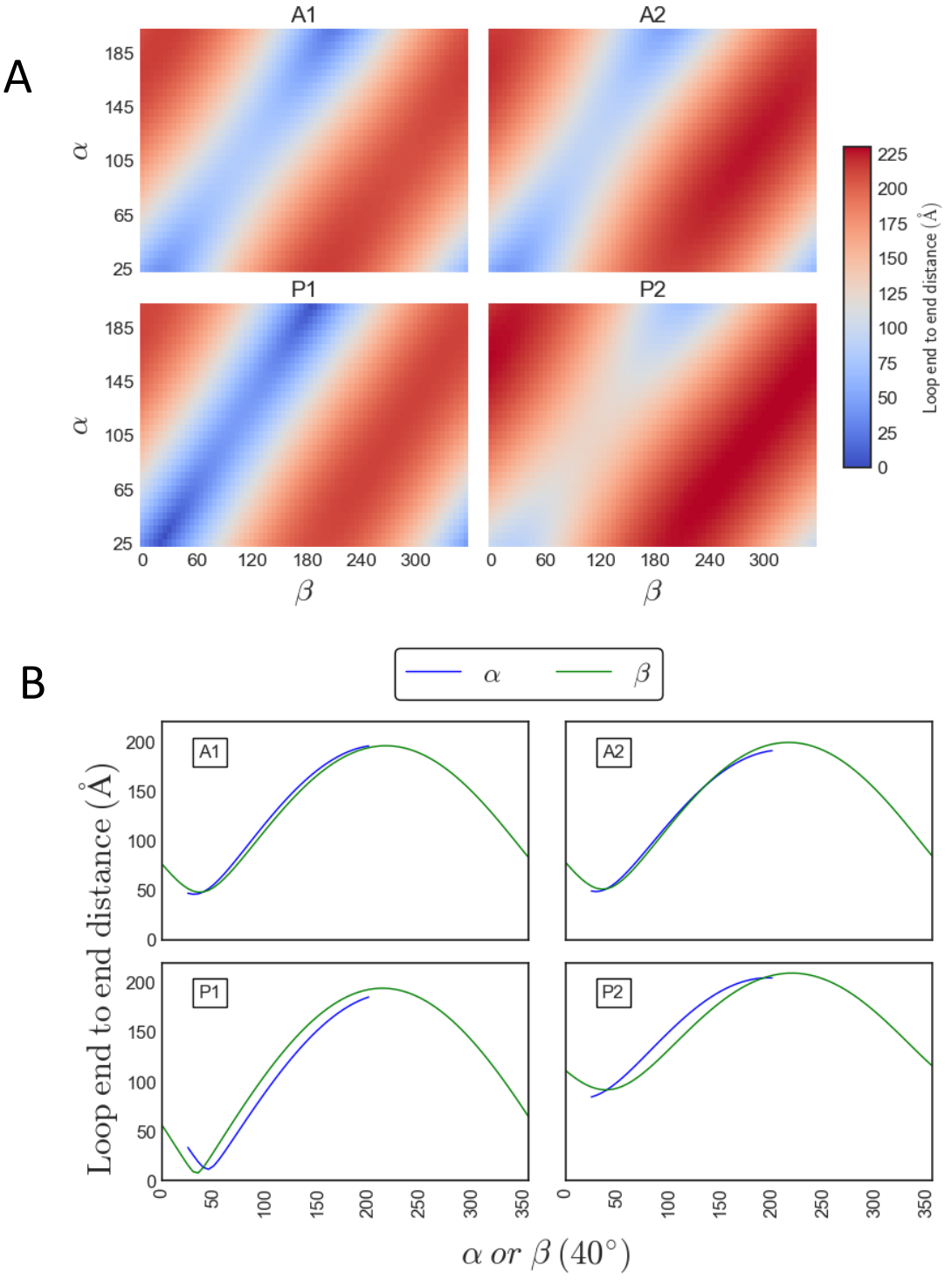

Figure 4.8 End to end distances between operators. (A) *Distance between the origins of base pair reference frames of the anchoring base pairs compared to estimated J factors. Distances between anchoring base pairs with one angle (α or β) set at 40° and the other varied i.e.,* $(\alpha, 40°) \approx (40°, \beta)$. *This trend is consistent for all values of α and β.*

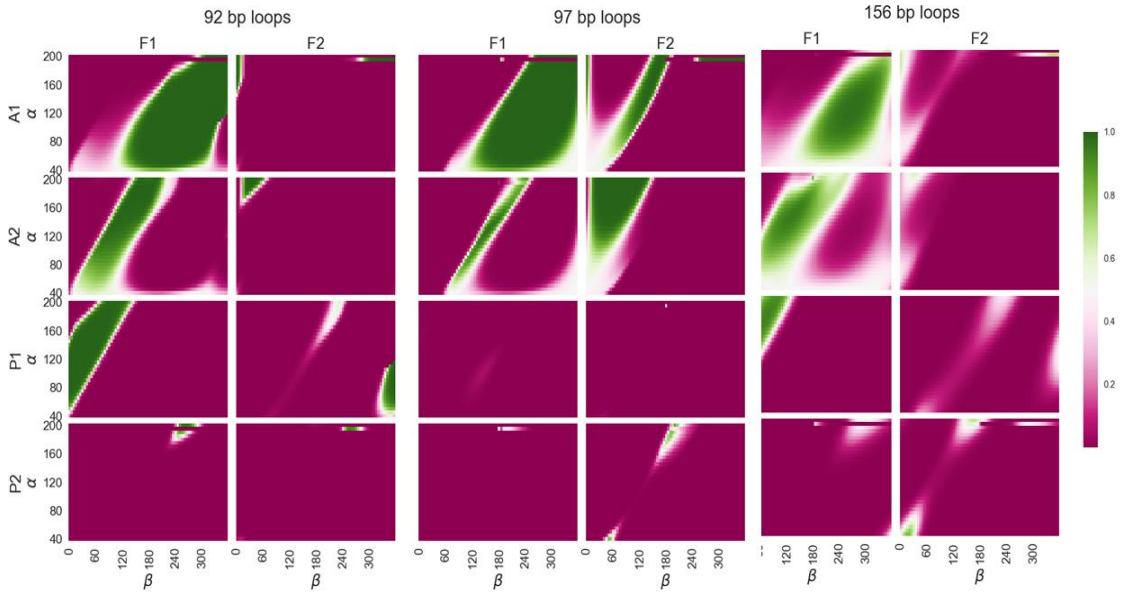

Figure 4.9 Fractional population of loops for different DNA loop lengths, structural families and operator orientations. All three lengths studied are included to facilitate comparison at the expense of resolution. For more detailed plots see Appendix C.

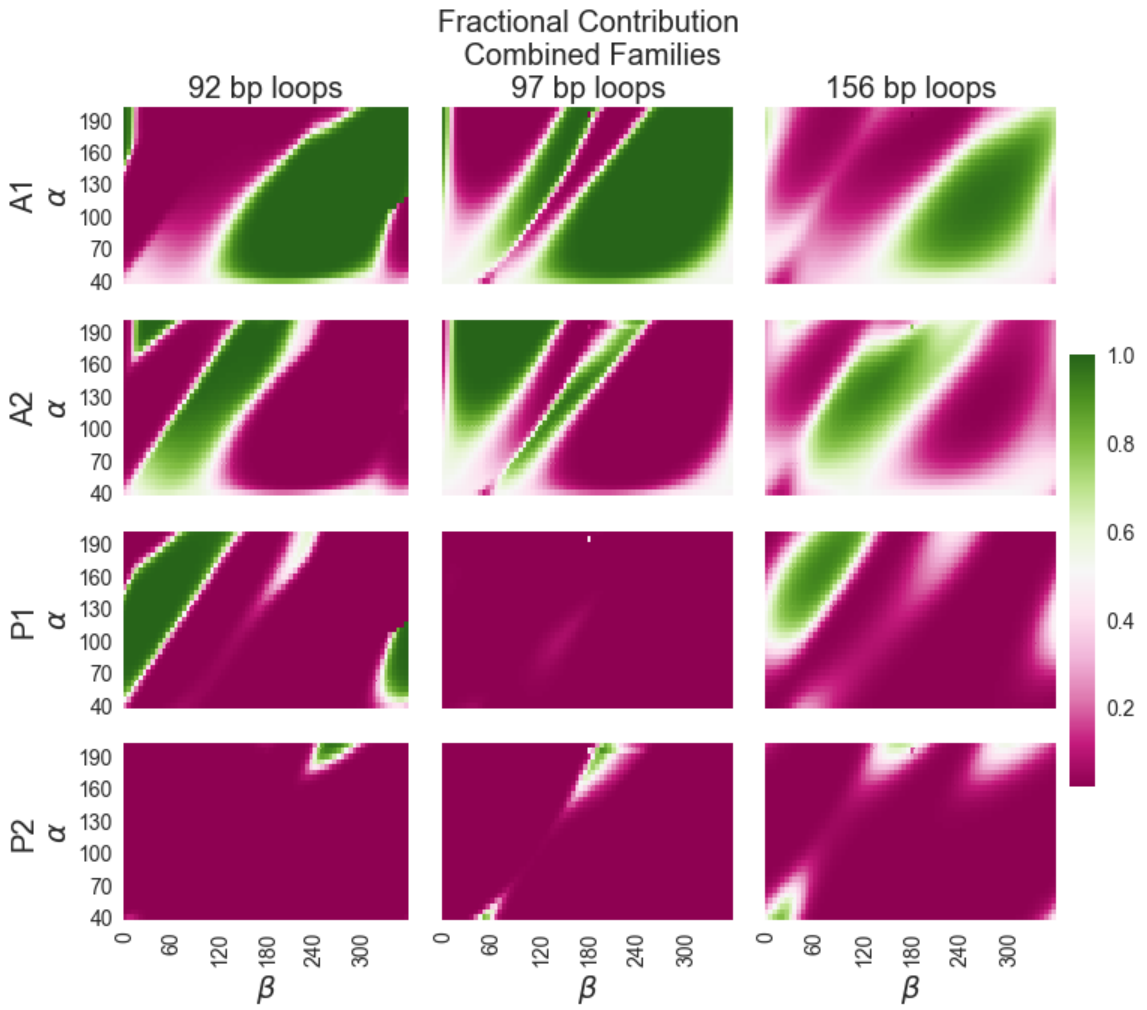

Figure 4.10 Fractional population of loops for different DNA loop lengths and operator orientation. Many experimental techniques do not distinguish between the two structural families. Here the populations are. combined by orientational loop type to give a single measure of loop type probability.

4.3.3 Spatial Configurations

Protein-DNA complexes with the highest J factors for the lengths studied include highly deformed proteins (Figure 4.11-Figure 4.14). At all three lengths studied the configuration with the highest J factor has a greatly rotated arm. The magnitude of the opening is larger for the 92 and 97 bp loops and less significant at the longer 156 bp loop configuration. As shown in Chapter 3, P2 loops are seldom seen at lengths below 11 helical turns (~115 bp) when LacR is in the rigid "V" shape. Figure 4.13 is an example where P2 loops with α = 185° and β = 200° make up 55% of the population. The images show that in this very deformed

configuration the loops adopt a spatial arrangement similar to an antiparallel loop and the A1 and A2 loops adopt a shape more like P1 and P2 loops respectively. In the 97 bp configuration where the protein is fully extended and the rotation at its farthest from rigid model, i.e., α = 180° and β = 185°, the P2 loop actually dominates the population at 75%.

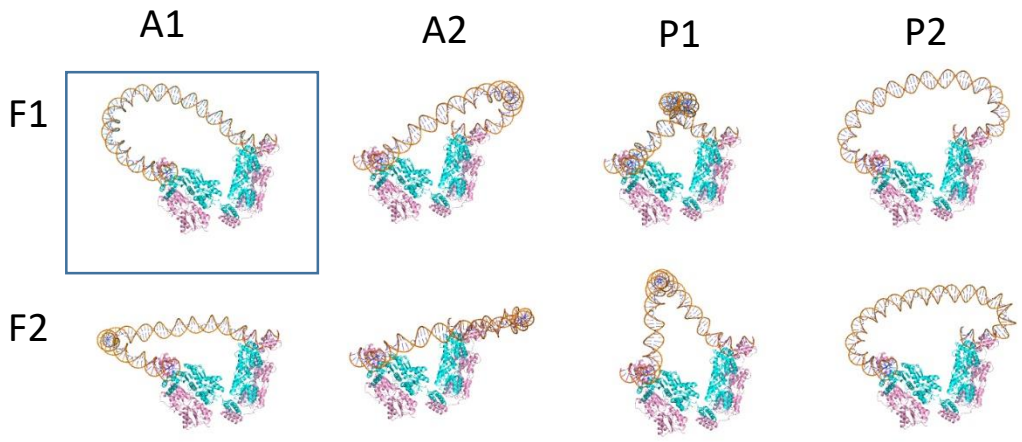

Figure 4.11 Molecular images of 92 bp loops at the configuration with the highest J factor. The deformation parameters are α = 95°, β = 160°. The blue box indicates the dominant loop type in the population.

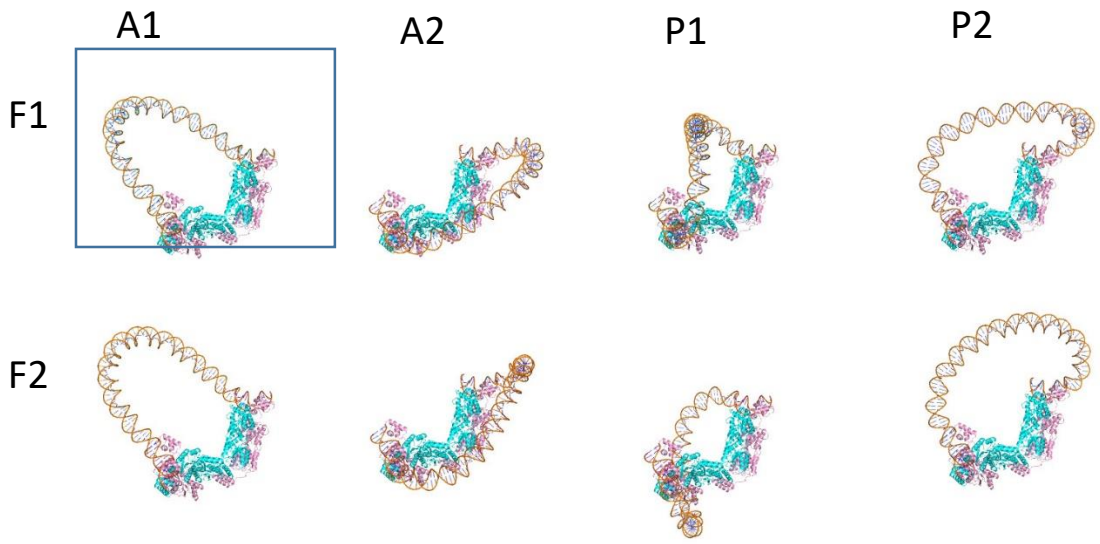

Figure 4.12 Molecular images of 97 bp loops at the configuration with the highest J factor. The deformation parameters are α = 170°, β = 265°. The blue box indicates the dominant loop type in the population.

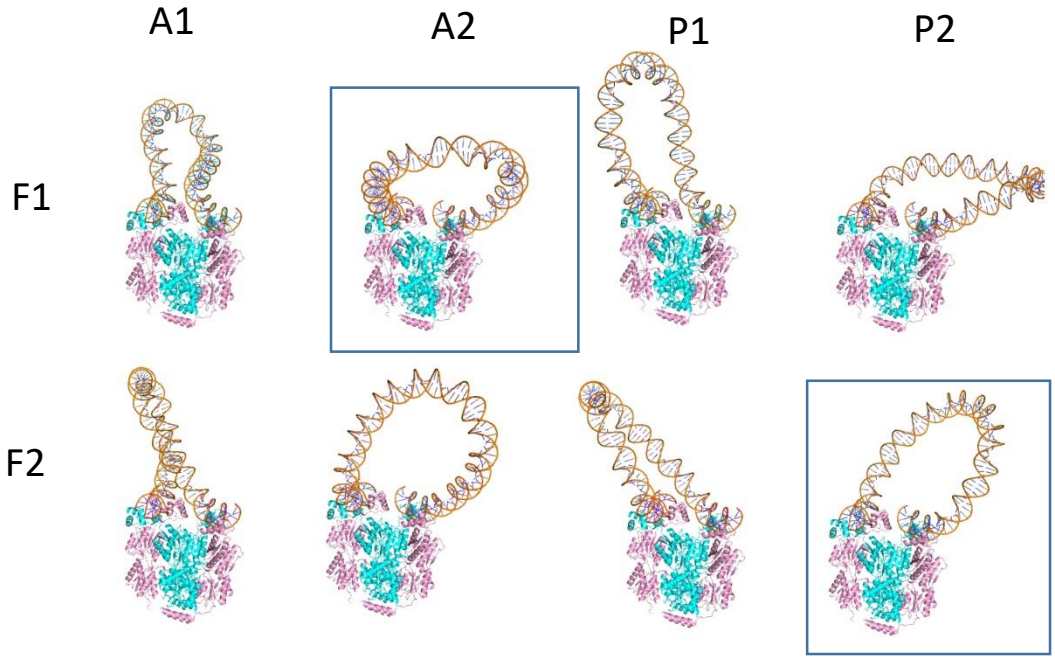

Figure 4.13 Molecular images of 97 bp loops at a configuration that includes a P2 loop. The deformation parameters are α = 185°, β = 200°. The blue box indicates the dominant loop types in the population.

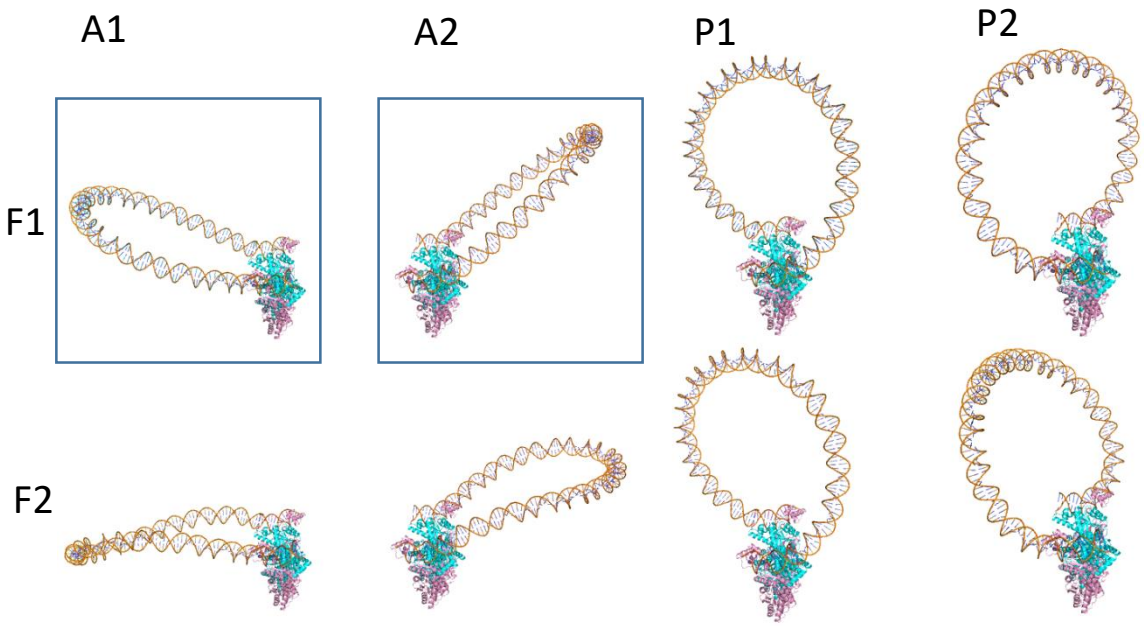

Figure 4.14 Molecular images of 156 bp loops at the configuration with the highest J factor. The deformation parameters are α = 55°, β = 145°. The blue boxes indicate the dominant loop types in the population.

Chapter 5 Implications of Sequence Length and Operator Displacement On Topological Domains of Protein-Bound Minicircles

Genomic DNA is vastly longer than the space allotted to it in a cell. The molecule must fold with a level of organization that satisfies the imposed spatial constraints as well as allows for the processing of genetic information. Key players in this organization include the negative supercoiling of DNA, which facilitates the unwinding of the double-helical molecule, and the associations of DNA with proteins, which partition the DNA into isolated loops, or domains. In order to gain insight into the principles of genome organization and to visualize the folding of spatially constrained DNA, we have developed new computational methods to identify the preferred three-dimensional pathways of protein-mediated DNA loops and to characterize the topological properties of these structures. Here we focus on the levels of supercoiling and the spatial arrangements of DNA in model nucleoprotein systems with two topological domains. We construct these systems by anchoring DNA loops in opposing orientations on a common protein-DNA assembly, namely the Lac repressor protein with two bound DNA operators. The linked pieces of DNA form a covalently closed circle such that the protein attaches to two widely spaced sites along the DNA. We examine the effects of operator spacing, loop orientation, and long-range contacts on overall chain configuration and topology and discuss our findings in the context of classic experiments on the effects of supercoiling and operator spacing on Lac repressor-mediated looping and recent work on the role of proteins as barriers that divide genomes into independent topological domains. We also discuss the effects of protein deformation on the effect of protein deformation on the formation of topological domains.

5.1 Introduction

The long strings of DNA that comprise the genomes of all organisms undergo several layers of compaction in order to fit and function inside the tight confines of a cell. A key feature in this organization is the formation of DNA loops held in place by proteins that are bound to widely spaced sites along the genome.[2,104–106] The proteins act as topological barriers that inhibit communication between elements on different loops and localize fragments of DNA with different levels of intertwining, or supercoiling, of the double-helical strands. Whereas the looser wrapping of strands characteristic of regions of negative supercoiling generally promotes genetic activities, the tighter wrapping found in regions of positive supercoiling tends to repress such functions.[107]

Insulator-like elements constructed by the binding of the Lac repressor protein to separated operator sites on DNA plasmids capture many of the features of DNA topological organization observed in vivo. The formation of a stable Lac repressor-operator complex introduces two protein-mediated loops on a covalently closed circular duplex (Figure 5.1). The recognition elements on the protein simultaneously contact two different sites on the DNA and constrain the pathways of the intervening pieces of the closed chain. The binding of the repressor to *distant* operator sites on plasmids of a few thousand base pairs (bp) thus inhibits communication between transcriptional elements located on the different halves of the DNA[108] and divides the DNA into two separate topological domains in which the level of supercoiling in one domain is independent of the other.[109] Conversely, the formation and sites of Lac repressor-mediated loops on DNA plasmids depend upon the level of supercoiling. Negative supercoiling enhances and stabilizes the association of the repressor with distant operators.[110–113] and the level of supercoiling dictates the distances between operators most readily incorporated in the paired loops.[112,113] Changes in supercoiling imposed through direct manipulation of single DNA molecules capture the interconversion of looped and

unlooped states implicated in the mix of repressor-operator complexes on circular DNA[8]. Moreover, the single-molecule studies reveal two distinct looped states that respectively build up and die off in phase with direct manipulation of one end of the DNA. Like the plasmids, the linear chains are spatially constrained, albeit in extended rather than closed configurations and with one rather than two protein-mediated loops.

Estimates of the free energies of DNA looping based on treatments of the double helix as an elastic rod offer insights into how a closed circular molecule may adapt to the constraints of protein binding. The ease of DNA looping depends upon the length of the molecule and the relative spatial disposition of its binding sites on protein[114]. The requirements of binding are best met if the ends of the DNA are appropriately phased with the ~10.5 bp DNA helical repeat.[9,43] Chains differing in length by 5-6 bp accordingly close with lesser ease against the given protein. The DNA chain length further determines the orientation of the loops anchored to the protein.[43] For example, when the DNA chain is relatively short, the most easily formed loops attach to the Lac repressor in antiparallel orientations and the less easily closed loops in parallel orientations (described below). Whereas the antiparallel loops follow U-shaped pathways with the ends of the molecule (*i.e.*, the protein-bound *lac* operator sequences) pointed in roughly opposing directions, the parallel loops describe helical arcs of low pitch with the DNA ends pointed in nearly the same direction. As each operator may bind to either of the repressor headpieces, the DNA loops can attach to the protein in four distinct ways[6] (Figure 1.2); an antiparallel form and a parallel form (respectively termed A_1 and P_1) that connect the first binding headpiece to the second and two opposing antiparallel and parallel forms (termed A_2 and P_2) that connect the second headpiece to the first. Moreover, pairs of opposing loops (A_1A_2 and P_1P_2) may simultaneously bind the same set of operators and thereby describe a continuous DNA chain. The Lac repressor complexes that partition closed circular duplexes form such pairs of loops (Figure

5.2). If the 5´-ends of the DNA are anchored to a common operator, the former pair of loops initially point toward the inside of the repressor assembly and the latter pair toward the outside.

The ease of dividing a closed DNA molecule into independent topological domains with an insulator protein thus depends upon the lengths and types of imposed loops. Here we estimate that cost from the relative elastic energies of short, Lac repressor-anchored fragments of ideal, naturally straight DNA. The optimized energies (Chapter 3) mirror the looping propensities and modes of chain attachment found in simulations of Lac repressor-mediated DNA chain closure[43] and capture chain-length-dependent variations in DNA loop formation determined experimentally[22,41]. We also take account of long-range contacts between looped segments that may influence formation of the protein-DNA assembly. We determine the preferred placement of operators at various positions along an ideal, 182 bp closed duplex and the likely configurations of the resulting loops. We identify conditions under which the modeled DNA may assume different levels of supercoiling and find that the binding of repressor to evenly spaced operator sites brings the energies of different topoisomers within close range of one another. We ignore the large-scale opening of the Lac repressor detected in low-resolution structural studies,[12,13,66] the effects of non-specific architectural proteins on looping,[67,115] the fluctuations in DNA operator sequences on the repressor headpieces,[36,69] and the local, sequence-dependent structural and energetic properties of DNA.[79] The work here thus serves as a starting point for further investigation of the contributions of repressor deformability, architectural proteins, minicircle chain length, and base-pair sequence to DNA topological organization.

5.2 Approach

Configurations of closed circular DNA molecules divided into separate topological domains by the binding of the Lac repressor were obtained by concatenating the structures of pairs of loops anchored in opposing orientations on the binding headpieces of the protein (Figure 5.2). The structures of the loops were identified with a new procedure that optimizes the energy of a collection of base pairs, in which the first and last pairs are held fixed.[32] The DNA is treated at the level of base-pair steps, using six rigid-body parameters to specify the arrangements of successive residues[60,82] and a potential that allows for elastic deformations of the parameters from their equilibrium rest states.[116] The base-pair steps are assigned the properties of an ideal, inextensible, naturally straight polymer, with isotropic bending deformations, independent fluctuations of twist, and a helical rest state with 10.5 bp per turn and successive residues displaced by 3.4 Å. The reported energies do not take account of the stabilizing interactions between protein and DNA, which are fixed in value by the assumed rigid repressor-operator model (Figure 5.1 and Figure 5.2).

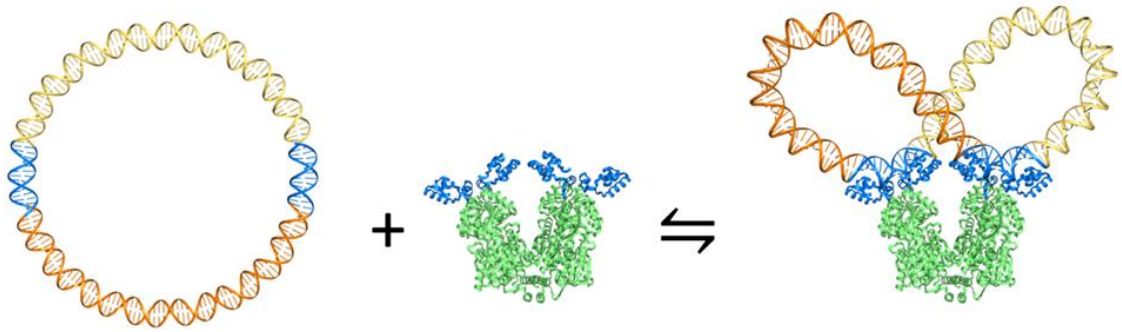

Figure 5.1 Molecular images illustrating the partition of a 182 bp DNA minicircle into equally sized, independent topological domains through the binding of the Lac repressor protein. Images rendered in PyMOL (www.pymol.org) with the DNA backbones depicted by tubes, the DNA bases by sticks, and the protein by ribbons. The DNA operators and the recognition headpieces on protein are highlighted in blue. The intervening segments of DNA are shown in different shades of gold and the remainder of the protein in green. Configurations were obtained by concatenating the structures of pairs of 92 bp energy-minimized loops anchored in opposing antiparallel orientations, here an A_1 loop in light gold and an A_2 loop in dark gold, on the binding headpieces of the protein.

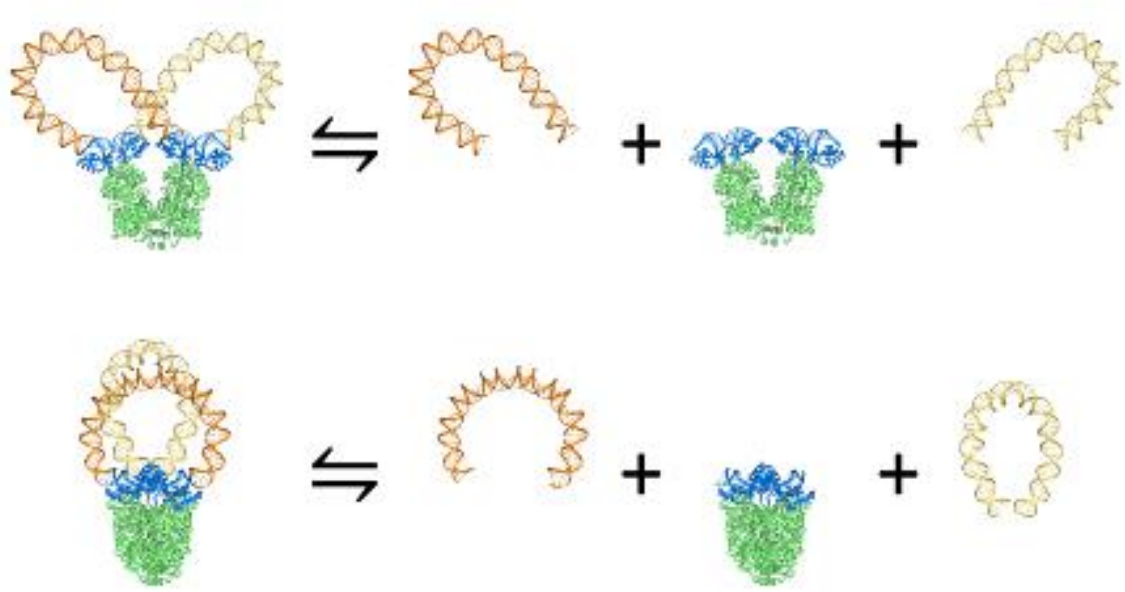

Figure 5.2. Configurations of Lac repressor-partitioned minicircles obtained by concatenating the structures of pairs of 92 bp energy-minimized loops, including the bound DNA fragments of the repressor–operator complex. The closed pathways constructed from antiparallel (A_1, A_2) loops (upper row) differ significantly from those formed from parallel (P_1, P_2) loops (lower row), illustrated here by the combination of A_1 or P_1 (light gold) and A_2 or P_2 (dark gold) loops. The molecular components of the former construct are oriented in the same viewpoint (down the axis perpendicular to the "V" formed by the arms of the protein assembly) as in Figure 5.1, and those in the latter from the vantage point obtained upon 45° rotation of the protein as a whole about the vertical axis through the center of the "V". See the legend to Figure 5.1. for molecular color-coding and styles.

The Lac repressor is represented by a rigid, V-shaped model, as described in Chapter 2. The reported loop lengths correspond to the number of base-pair steps between the centers of the bound DNA operators, namely seven of the 13 steps attached to each arm of the modeled repressor-operator assembly plus the DNA steps subjected to configurational variation. The lengths of the loops are accordingly 14 bp steps longer than the lengths of the optimized, ligand-free segments of DNA and the length of the circular DNA formed upon concatenation of two protein-mediated loops is 12 bp steps longer than the combined lengths of the two ligand-free segments. Concatenation of opposing loops entails removal of a common base pair from each of the bound operators, thereby reducing the number of base-

pair steps in the resultant minicircle by two compared to the total number of steps in the two loops.

The DNA loops include configurations from two competing families of Lac repressor-mediated structures associated with each orientational setting on the protein[36]. The energies of the loops in the two families, here termed F1 and F2, exhibit similar ~21 bp periodic dependencies on chain length. The oscillations of energy with chain length, however, differ in phase by a helical turn such that the valleys in the energy profile of one family of looped structures coincide with the peaks in the other and *vice versa* (Figure 3.3). The energy minima occur at chain lengths where the uptake of twist is minimal and the ends of the DNA fall in nearly perfect register with the binding sites on the repressor. Opposing pairs of antiparallel or parallel loops may accordingly attach in four distinct ways to a specified pair of operators on a closed duplex. That is, opposing antiparallel loops may combine as $A_1^{F1}A_2^{F1}$, $A_1^{F1}A_2^{F2}$, $A_1^{F2}A_2^{F1}$, or $A_1^{F2}A_2^{F2}$ pairs and opposing parallel loops as $P_1^{F1}P_2^{F1}$, $P_1^{F1}P_2^{F2}$, $P_1^{F2}P_2^{F1}$, or $P_1^{F2}P_2^{F2}$ pairs. The relative populations of the different forms depend upon the levels of torsional and bending stress associated with the assumed orientations and spacings of the DNA operators and the length of the DNA as a whole. Here we consider DNA minicircles of 182 bp and allow for all possible orientations of loops with center-to-center operator spacings between 73 and 111 bp. The loops that divide this minicircle in half contain the natural 92 bp center-to-center operator spacing found in *Escherichia coli.* The choice of loop lengths is based on the difficulty of anchoring very short (< 73 bp) loops to the assumed V-shaped Lac repressor model[43] and the restraints of the minicircle on maximum loop size, i.e., 182+2=73+111.

The topology of the repressor-partitioned minicircles, which follows from the combination of loop configurations, is described in terms of a discrete ribbon constructed from the origins and reference frames of four successive base pairs[31,117]. In contrast to the

twist angle included in the six rigid-body parameters used to specify the relative spatial arrangements of successive base pairs (mentioned above), the values of twist reported here, the so-called twist of supercoiling[31], can be combined with the writhing number of a closed structure to obtain the correct linking number, an integer if the edges of the DNA ribbon are connected and the surface bounded by the ribbon is free of self-contact[33,118]. The 'closed' DNA includes the two constrained loops and both protein-bound operators. The values of twist reveal the extent to which the protein contributes to DNA topology and how torsional stress is distributed along the constrained molecule.

5.3 DNA Looping On Closed Circular DNA

5.3.1 Lac Repressor-Partitioned Minicircles

Lac repressor-mediated loops with the natural 92 bp center-to-center operator spacing found in *Escherichia coli* adopt a mix of energy-optimized configurations with subtle differences in three-dimensional structure[27,36]. The antiparallel loops either under- or overwind, by ~2°/bp relative to ideal B DNA, to fit against the protein headpieces, and the DNA as a whole swings back and forth relative to the repressor structure. The loops anchored in some of the more highly bent parallel orientations are subject to less torsional stress and thus comparable in total elastic energy to the combined bending and twisting energies of the antiparallel loops.

Concatenation of the optimized 92 bp loops yields five major classes of Lac repressor-bound minicircles, four from the combination of antiparallel loops and one from the combination of parallel loops. The structures formed from the antiparallel loops follow similar spatial pathways with a narrow protein-bound 'waist' connecting two roughly U-shaped lobes (Figure 5.3). The lobes, however, adopt different configurations with respect to the protein and assume different levels of supercoiling. Both lobes may underwind, both

overwind, or one lobe underwind and the other overwind relative to B DNA. The base pairs along the two underwound loops, which are incorporated in the $A_1^{F1}A_2^{F1}$ protein-anchored construct, deviate as much as 60 Å from the corresponding residues along the two overwound loops included in the $A_1^{F2}A_2^{F2}$ construct. The latter minicircle is more extended and less contorted than the former with loop apices separated by intramolecular distances up to 240 Å vs. maximum distances of 213 Å in the more compact form. The respective writhing numbers of the two constructs, –0.5 and –0.2, reflect the closed and open arrangements of the DNA lobes relative to the protein assembly.

Although the Lac repressor unwinds the highly kinked CG base-pair steps at the centers of the DNA operators by more than 9° relative to B DNA, the protein headpiece introduces a net increase in twist over the bound operators (~5° over each 13-step fragment). The changes in twist along the protein-free DNA fragments are substantially larger. The ~150° net decrease in twist within each lobe of the $A_1^{F1}A_2^{F1}$ minicircle occurs in the context of a topoisomer with linking number Lk of 16 and the nearly comparable increase in twist within the arms of the $A_1^{F2}A_2^{F2}$ minicircle in the context of a topoisomer with Lk = 18. The protein-bound minicircles made up of one underwound and one overwound loop have the same linking number (Lk = 17) as a protein-free DNA minicircle of the same length. In contrast to the protein-free minicircle, where a ±1 change in Lk increases the energy substantially over that of the Lk = 17 topoisomer, the binding of the Lac repressor to operator sites spaced by 92 bp on the DNA equalizes the energies of these topoisomers. That is, the differences in elastic energy between the topoisomers decrease from tens to fractions of k_BT if the V-shaped protein attaches to operators with 92 bp center-to-center spacing. Although the local DNA deformations in the Lac-partitioned minicircle are much more costly than the bending and twisting of the free minicircle, the energies of the protein-mediated loops are not appreciably affected by the mode of protein attachment and associated levels of

supercoiling. It is worth noting that the Estimation of the twist from the rigid-body parameter of the same name exaggerates the unwinding at the CG step and incorrectly suggests that the repressor unwinds DNA by ~15°. Such treatment ignores the contribution to supercoiling from the large shear of base pairs at the CG step.[31]

The low-energy Lac repressor-bound construct formed from the pairwise combination of parallel (P_1^{F1}, P_2^{F1}) loops brings the centers of the two protein-free segments into close, potentially favorable contact (Figure 5.4). The outer P_1^{F1} loop runs roughly perpendicular to the inner P_2^{F2} loop with the edges of the grooves close enough to associate with a common ligand. The DNA as a whole follows a more highly folded but much less torsionally stressed pathway than the DNA attached to the repressor in the four antiparallel arrangements. The accumulated change in twist along each lobe — roughly –100° along the P_1^{F1} lobe and +40° along the P_2^{F2} lobe, relative to B DNA — is much lower than the corresponding changes in twist in the constructs with antiparallel attachments. The writhing number Wr of –1.2 and linking number Lk of 16 are consistent with the folded, plectonemic DNA pathway evident from certain views of the structure. The DNA follows similar folded pathways of comparable writhe but much greater torsional stress if bound to protein through other parallel attachments. For example, the accumulated changes in twist underwind a P_1^{F2} loop by nearly –300° and overwind a P_2^{F1} loop by almost 250°. Thus, the only feasible way to divide a 182 bp minicircle in half through parallel attachments to the modeled repressor is with the $P_1^{F1}P_2^{F2}$ construct.

5.3.2 Influence of Operator Spacing on Protein Uptake

Repositioning the *lac* operators on the 182 bp minicircle has a profound effect on the elastic energies and configurations of the repressor-bound system. A shift in operator spacing by a single base pair from the evenly spaced, or symmetric, attachment of protein considered

above reduces the mix of feasible repressor-bound configurations to a single state (Figure 5.5a, Figure 5.6). The changes in spacing alter the DNA torsional stress, with a 1 bp decrease in the separation of operators reducing the build-up of twist in A_1^{F1} and A_2^{F1} loops by ~25° and a 1 bp increase in spacing doing roughly the same in A_1^{F2} and A_2^{F2} loops. The net twist increases in magnitude by ~20° and the energy rises precipitously with changes in spacing of the opposite sense in the same loops, i.e., an increase in operator spacing along the first set of loops and a decrease along the second. Thus, a single configuration dominates a 182 bp minicircle unevenly divided by repressor attachments at sites 91 and 93 bp apart. The shorter lobe adopts an A_1^{F1} or A_2^{F1} configuration that is slightly less underwound and the longer lobe an A_2^{F2} or A_1^{F2} configuration that is slightly less overwound than the lobes formed upon symmetric protein attachment. The multiple configurations associated with the latter placement of protein accordingly serve as a 'transition state' between the $A_1^{F1}A_2^{F2}$ and $A_1^{F2}A_2^{F1}$ repressor-bound constructs that respectively dominate the configurational landscape when

the spacing between the first and second operators is smaller or larger than that between the second and first.

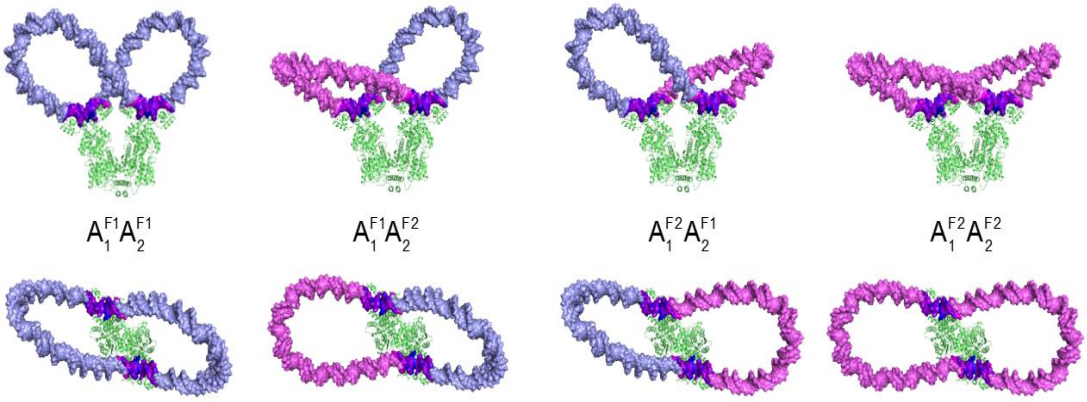

Figure 5.3 Molecular images showing the overall folding and torsional stress in a 182 bp DNA minicircle
Minicircles found upon attaching a pair of evenly spaced operators in all possible antiparallel orientations to the Lac repressor. Configurations are denoted by the settings and families of the energy-optimized loops anchored to the protein (see text). Views looking perpendicular (upper row) and down (lower row) the symmetry axis of the V shaped protein assembly. The torsional stress in expressed in terms of the net change in the twist of supercoiling relative to B-DNA, at individual base-pair steps along the closed molecule. DNA is color-coded such that the most underwound steps (located on the operators) in deep blue and the most overwound steps (also on the operators) in deep magenta, with intermediate regions of negative, null and positive deviations in twist varying respectively from blue to white to magenta. The uniform twist along the protein-free lobes reflects the treatment of DNA as an ideal elastic rod.

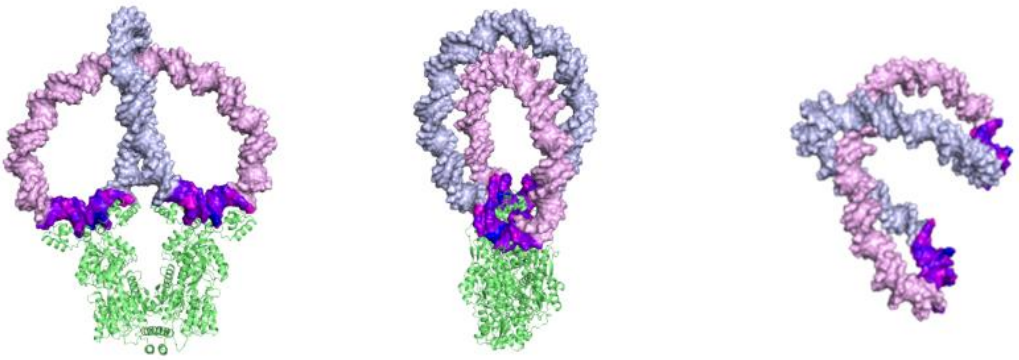

Figure 5.4 Plectonemic Folding of $P_1^{F1}P_2^{F2}$ Minicircles. Molecular images illustrating the plectonemic folding, torsional stress, and long-range contacts of a 182 bp DNA minicircle formed upon binding a pair of evenly spaced, operators in parallel ($P_1^{F1}P_2^{F2}$) orientations to the Lac repressor. Views of the protein-bound assembly looking perpendicular to (left) and down (center) the axis joining operator centers and from a perspective, in the absence of protein, that highlights the interwinding of the DNA as a whole (right). Torsional stress color-coded as Figure 5.3

The torsional stress on DNA nearly vanishes when the operator spacing changes by ±5 bp relative to the symmetric placement of repressor on the minicircle (Figure 5.6). The elastic energy attains its lowest values when the operator centers lie 87 and 97 bp apart and the recognition elements on DNA lie in almost perfect register with the repressor headpieces. The lobes of these structures fall in the same $A_1^{F1}A_2^{F2}$ and $A_1^{F2}A_2^{F1}$ conformational classes as those formed on repressor-bound minicircles with operators shifted ± 1 bp from their sites on the evenly divided minicircle. The Boltzmann factors of the ±5 bp shifted states are roughly eight orders of magnitude greater than the combined Boltzmann factors of the five types of Lac repressor-bound minicircles that comprise the set of evenly partitioned minicircles with 92 bp operator spacing (Figure 5.5b). The Boltzmann factors associated with two other protein-divided minicircles, formed when the operator spacing differs by ±16 bp relative to the evenly divided minicircle, are nearly as large. The lobes of these low-energy structures, however, fall into different conformational categories from those with ±5 bp operator displacements. The shorter lobes adopt A_1^{F2} or A_2^{F2} configurations and the longer lobes A_2^{F1} or A_1^{F1} configurations (Figure 5.5a, Figure 5.6). Transitions between the configurational families associated with the two low-energy forms occur at operator displacementsof ±10-11 bp. The mix of configurational states at these points (Figure 5.6) is less varied than that found for minicircles with repressor-mediated loops of the same size. One of the antiparallel loop combinations dominates the configurational landscape on either side of the transition points with a small proportion of minicircles divided by repressor through $A_1^{F1}A_2^{F1}$ and/or $A_1^{F2}A_2^{F2}$ attachments (Figure 5.5a). The single low elastic energy construct formed from the combination of unevenly sized parallel loops (a P_1^{F2} loop 20-24 bp shorter than its P_2^{F1}partner) self-intersects. The linking numbers of the configurations that comprise the 'transition states', nevertheless, span the same range of values (16-18) found in the evenly

divided minicircles (Figure 5.7). The Lk = 17 topoisomer dominates in all of the unevenly divided minicircles, including the 'transition states' with ±10-11 bp operator displacements.

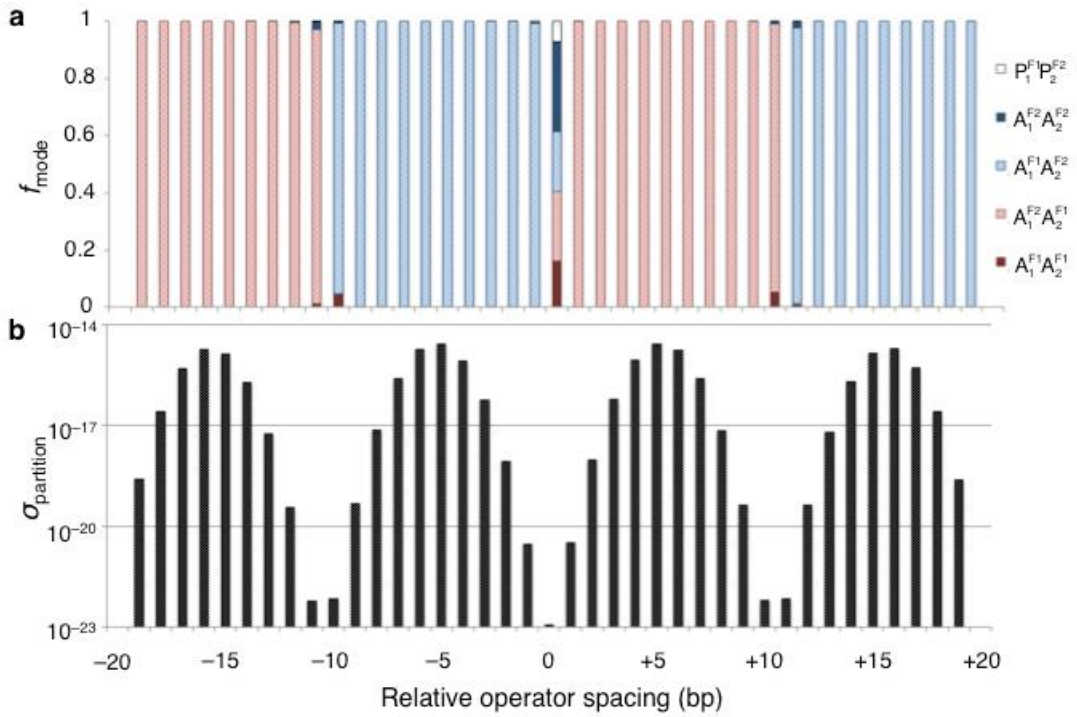

Figure 5.5 Profiles, as a function of operator spacing of (a) the fraction f_{mode} of 182 bp, Lac repressor-bound DNA minicircles in each of the energetically favored modes of loop attachment to the V-shaped protein assembly and (b) the relative ease of partitioning the DNA into two domains at these sites. Operator spacing is expressed in terms of the displacement of the recognition sequences, in base-pair steps, relative to the even positioning at 0 associated with the formation of equally sized loops having the 92 bp center-to-center spacing found in the Escherichia coli lac operon. That is, one of the protein-anchored loops is altered in size by the specified difference in spacing and the other by its negative. The ease of DNA partitioning $\sigma_{partition}$ is given by the sum of the Boltzmann factors of the elastic energies of the modeled Lac repressor-bound minicircles at each spacing. Note the much lower chances of dividing the minicircle in half compared to partitioning it at other sites

5.4 Concluding Remarks

Although this chapter focuses on the configurations of a single short DNA minicircle bound at widely separated operator sites to the Lac repressor protein, the patterns of looping in the optimized structures reveal principles that can be related to the general, protein-mediated topological organization of DNA. For example, the preferred mode of DNA attachment to the repressor, with operators pointed in opposing, roughly antiparallel directions against the protein binding headpieces, (Figure 5.3, Figure 5.6) may contribute to the low levels of communication found to occur between the two halves of a Lac repressor-partitioned plasmid[119]. The U-shaped pathways of the looped DNA segments lie on either side of the V-shaped protein, and unless appreciably long[120] cannot deform to the extent needed to generate sequentially distant, intersegmental contacts. Indeed, looping of this type provides a useful model of the compaction of chromatin during mitosis[121]. The generally more costly partitioning of closed circular DNA through the parallel alignment of operators on the repressor brings the two protein-free segments into close contact, especially if the chain is divided into comparably sized pieces (Figure 5.4). Most modes of such attachment, however, introduce sterically impossible DNA-DNA and/or protein-DNA contacts that rule out these structures (Figure 5.5a).

The changing mix of configurational states found upon repositioning *lac* operators on the modeled repressor-divided minicircles (Figure 5.5, Figure 5.6) offers a rationale for the known influence of DNA supercoiling on the spacing of operators optimal for loop formation[112]. The differences in torsional stress among the various modes of DNA attachment to protein give rise to a variety of topoisomers at each operator setting. The levels of torsional stress depend upon the operator spacings and attain their lowest values when the operators lie in close register with the repressor headpieces. A single low-energy topoisomer dominates

the configurational landscape if both lobes of the repressor-bound DNA are torsionally relaxed, such as the lobes with 87- and 97 bp spacing anchored in antiparallel settings on a 182 bp minicircle of linking number 17 (Figure 5.6, Figure 5.7a). The much higher energies of minicircles with different modes of loop attachment and DNA topologies, i.e., the Lk = 16 and 18 topoisomers of the 182-mer with the same 87-, 97 bp operator spacings, preclude these occurrences. Operator spacings that introduce comparable, non-zero levels of torsional stress in different types of loop attachments yield a wide range of repressor-bound topoisomers, such as the topoisomers of a 182 bp minicircle of linking numbers 16-18 formed from the combination of under- and overwound 92 bp loops (Figure 5.3, Figure 5.7a). The mix of topoisomers tends to be greatest when the spacing of operators differs by roughly a half helical turn (5-6 bp) from the operation separation distances of lowest torsional stress. The similar energies of these systems suggest that the repressor will bind to different topoisomers. The models further illustrate how different levels of torsional stress can build up in protein-divided minicircles[109] and how different looped states can build up and die off in phase with direct manipulation of one end of a tethered DNA molecule[8].

The measured effects of supercoiling on the binding of the Lac repressor to a larger repressor-partitioned minicircle[112] follow directly from these principles. The tetrameric protein assembly successfully divides a torsionally relaxed 452 bp closed circular DNA into lobes with 158 and 294 bp spacing but fails to bind to topoisomers of different linking number, a system analogous to the 182 bp minicircle with lobes of 74 and 84 bp that persists exclusively in the relaxed state. In addition, the protein shows minimal binding to a relaxed 452 bp minicircle with 163 and 289 bp operator spacings but associates readily with topoisomers of lesser linking number, a system analogous to the 182 bp minicircle divided into lobes of 79 bp with different topologies. The predicted distribution of topoisomers of a 318 bp ideal DNA minicircle partitioned by the V-shaped Lac repressor model into lobes

ranging in size between 153 and 168 bp captures the relative spacings of operators found to associate with different topoisomers of the 452 bp minicircle (Figure 5.7b). The modeled systems are constructed, as described above, by concatenation of the energy-optimized loops from competing families of Lac repressor-mediated structures of the selected lengths and the distribution of linking numbers is proportional to the sum of the Boltzmann factors of the variously combined structures. The total length of the minicircle is limited by the size of currently available energy-minimized loop structures. Our work to date has focused on the energies and topological properties of short DNA loops of 73-143 bp, where one can omit consideration of the thermal fluctuations of the closed structures.

The slight discrepancies in the looping propensities of linear DNA, optimal at operator spacings of 160 bp in the modeled system vs. 158 bp in the corresponding experimental construct[30], give rise to the offset in predicted vs. observed loop topologies at different operator spacings (displaced histogram at the top of Figure 5.7b).

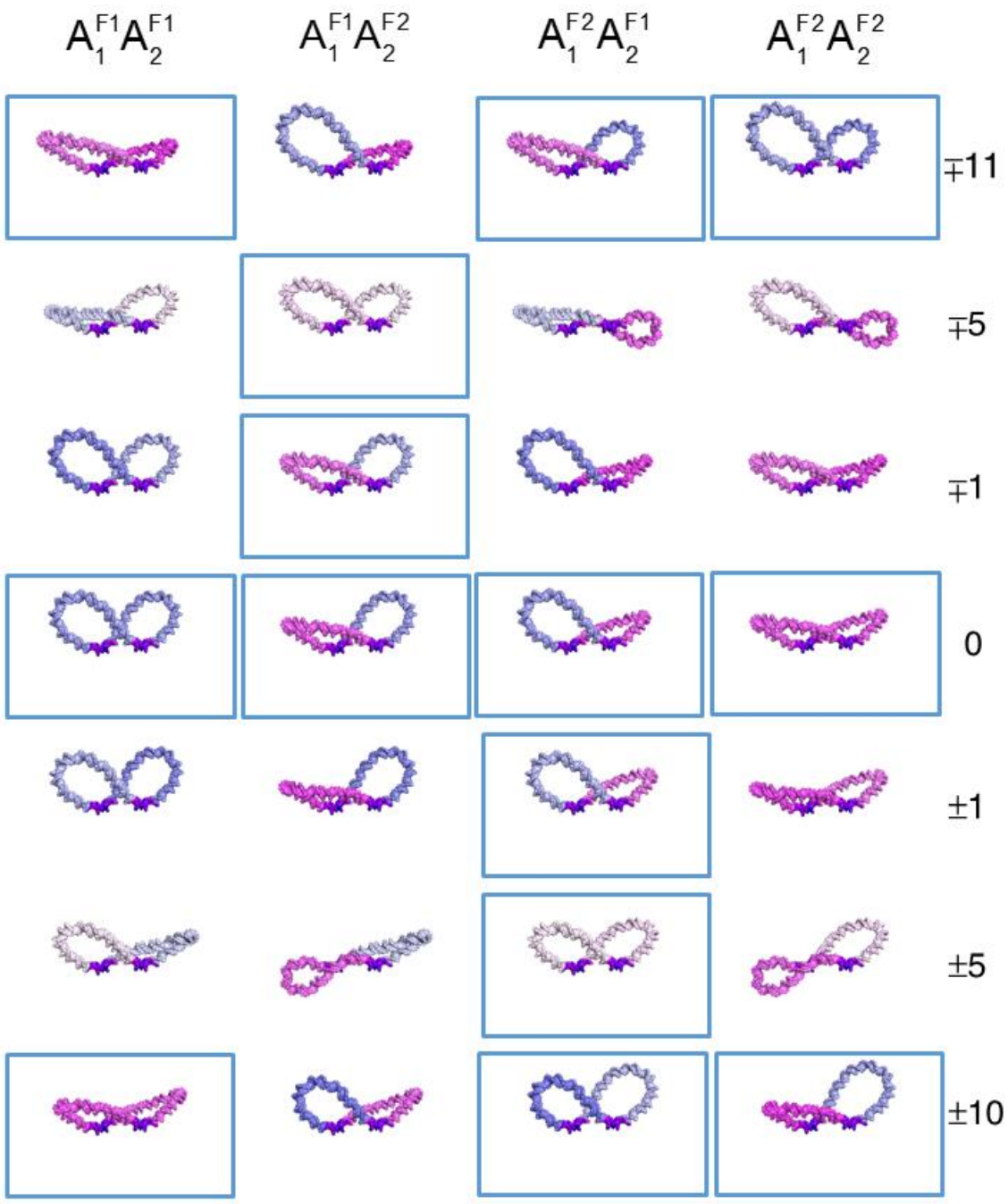

Figure 5.6 Gallery of molecular images illustrating the changes in overall folding and torsional stress in a 182 bp, Lac repressor-partitioned DNA minicircle upon selective displacement of operators relative to the even spacing entailed in the formation of two 92 bp loops. Configurations are labeled, color-coded, and oriented as in Figure 5.3, with the favored arrangements at the specified spacings outlined in blue. The upper and lower signs in the ∓ and ± symbols denote the respective changes in the lengths of the A_1 and A_2 loops. Structures are shown in the absence of protein to draw attention to the subtle differences among DNA structures

Slight changes in the DNA model, such as an increase in the assumed double-helical repeat from 10.5 to ~10.7 bp/turn, remove the difference in predicted vs. observed phasing of linking numbers, and as reported elsewhere[24], so does the opening of the V-shaped Lac repressor assembly. The DNA helical repeat reflects the underlying nucleotide sequence with values spanning a range, e.g., a tenfold repeat in the poly dA·poly dT homopolymer vs. a 10.6-10.7 bp repeat in poly dG·poly dC[54–56], much wider than the change noted above. On the other hand, whether the Lac repressor adopts an opened state in the 452 bp protein-divided minicircle remains an open question. Whereas the opening of the repressor is needed to mirror the DNAase I cutting patterns[30] of very short (52 and 74 bp) Lac repressor-mediated loops[9], it is not necessary to hypothesize such deformations[122] to account for the looping propensities deduced from tethered particle motions studies of longer (110-130 bp), Lac-repressor-mediated loops[22,41]. All other evidence supporting the opening of the repressor derives from studies of the isolated protein or complexes of the protein with the operator alone[12,13,66,95]. In this regard, we are investigating the effects of Lac repressor deformation on the preferred spatial pathways of the DNA looped between its binding headpieces and the implications of these changes on the partitioning of closed circular DNA by the protein assembly.

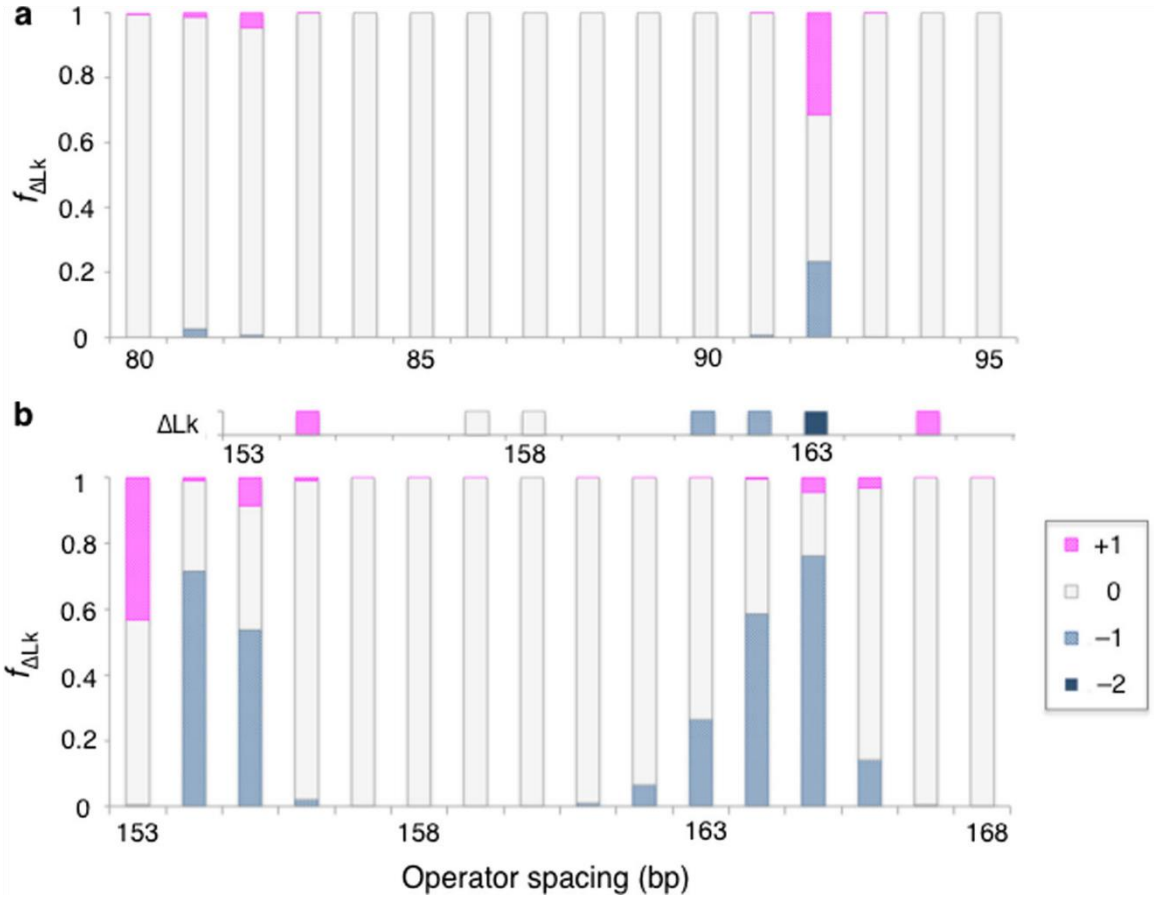

*Figure 5.7 Profiles, as a function of operator spacing, of the distribution of topoisomers in (*a*) 182 bp and (*b*) 318 bp Lac repressor-partitioned minicircles generated from eight modes of DNA loop attachment to the V-shaped protein assembly. Operator spacing is expressed in terms of the center-to-center distances of one of the two DNA lobes formed upon protein binding and DNA topology in terms of the differences in linking number $\Delta Lk = Lk - Lk_0$ of the topoisomers with respect to the relaxed state. Values of Lk_0 are the linking numbers of the constructs of lowest torsional stress, Lk = 17 in (a) and Lk = 30 in (b). The fraction $f_{\Delta Lk}$ of minicircles with a given value of ΔLk is proportional to the sum of the Boltzmann factors of those constructs. The predictions in (b) mirror the effects of supercoiling on the stability of loops of the same size formed in 452 bp Lac repressor-divided minicircles[112] if offset by 2 bp (note the shifted histogram of observed ΔLk values above the predicted data). The discrepancy in phase disappears upon a slight increase in the assumed DNA helical repeat (see text)*

Appendix A Kahn Sequence Data

Construct Label	**DNA Sequence (between EcoR V sites)**	Offset	Center-Center Operator Distance(bp)
5C18	ATCTGCAGGTCAGTCTAGGTAATTGTGAGC|GCTCACAATTAGATCTTCGTACGGAT CCGGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGT TTTTTGCCGTTTTTTGCCCGTTTTTTGCGCTGAACGCGTCCTAGACGCTATCGAAGC TAGCTAATTGTGAGC|GCTCACAATTCGTTGTGGTAAAGCTTTGAT	-8	156
7C16	ATCTGCAGGTCAGTCTAGGTAATTGTGAGC|GCTCACAATTAGATCTCTTCGTACGG ATCCGGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCC GTTTTTTGCCGTTTTTTGCCCGTTTTTTGCGCTGAACGCGTCCTAGACGATCGAAGC TAGCTAATTGTGAGC|GCTCACAATTCGTTGTGGTAAAGCTTTGAT	-6	156
9C14	ATCTGCAGGTCAGTCTAGGTAATTGTGAGC|GCTCACAATTAGATCTCAATTCGTAC GGATCCGGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTTTTGC CCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCGCTGAACGCGTCCTAGAATCGAAGC TAGCTAATTGTGAGC|GCTCACAATTCGTTGTGGTAAAGCTTTGAT	-4	156
11C12	ATCTGCAGGTCAGTCTAGGTAATTGTGAGC|GCTCACAATTAGATCTCAGATCTCGT ACGGATCCGGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTTTT GCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCGCTGAACGCGTCCTAATCGAAGCT AGCTAATTGTGAGC|GCTCACAATTCGTTGTGGTAAAGCTTTGAT	-2	156
13C10	ATCTGCAGGTCAGTCTAGGTAATTGTGAGC|GCTCACAATTAGATCTCAGATCTCGT CGACGGATCCGGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTT TTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCGCTGAACGCGTCCATCGAAGCT AGCTAATTGTGAGC|GCTCACAATTCGTTGTGGTAAAGCTTTGAT	0	156

Table A-1 156 bp DNA sequences used in the Kahn group FRET experiments. The curved phased A-tract sequence is highlighted in yellow. The symmetric operator regions are highlighted in green. The donor and acceptor adjusted linker sequence is highlighted in cyan and magenta respectively. The locations of the residue labeled with a donor fluorophore are indicated by bold blue text and acceptor fluorophore by bold red text.

Construct Label	**DNA Sequence (between EcoR V sites)**	Offset	Center-Center Operator Distance(bp)
9C16	ATCTGCAGGTCAGTCTAGGTAATTGTGAGC|GCTCACAATTAGATCTCAATTCGTACG GATCCGGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCC CGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCGCTGAACGCGTCCTAGACGATCGAAGC TAGCTAATTGTGAGC|GCTCACAATTCGTTGTGGTAAAGCTTTGAT	-5	158
13C14	ATCTGCAGGTCAGTCTAGGTAATTGTGAGC|GCTCACAATTAGATCTCAGATCTCGTC GACGGATCCGGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTTT TGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCGCTGAACGCGTCCTAGAATCGAA GCTAGCTAATTGTGAGC|GCTCACAATTCGTTGTGGTAAAGCTTTGAT	-2	160

Table A-2 158 and 160 bp DNA sequences used in the Kahn group FRET experiments See caption from Table A-1

Appendix B Force Field Constants

The following tables contain the force field constants used in to calculate the energy of deformation ψ for each base pair step according to equation 2.2 which can be rewritten as:

$$\psi^n = \frac{1}{2}(\underline{p}^n - p_0^n)^T F^n (\underline{p}^n - \underline{p}^0)$$

We then treat the DNA as inextensible elastic rod model such that, the angular, parameters tilt, roll and twist, are varied and the translational parameters, shift, slide and rise, are held constant. In the first six rows of each table describe the intrinsic step parameters, $\underline{p}_0$. The simplest model, described as Ideal DNA in Chapter 2 sets the zero energy, or rest state, as a straight rod with 3.4Å between each base pair and a helical repeat of 10.5 base pairs. The rest of the intrinsic step parameters are set to zero. The second section of each table describes the elastic moduli, F, associated with the deformation. The force constants are based on the persistence length $\boldsymbol{a}$ of DNA and base pair spacing $\Delta \boldsymbol{S}$ and are described in detail in Chapter 2. The values along the diagonal of the 6x6 force constants matrix F are for a single step parameter. The of diagonal values can be thought of as correlation values and represent a coupling of parameters. Except where noted the of diagonal are not considered in the model thus are set to zero and not included in the table. There are 10 unique step type columns in each table that represents an instance of force field parameters for a given step type. In force field where nucleotide sequence is not considered all of the columns will be identical.

Persistence length a = 476 bp

$\Delta s = 3.4\text{Å} = \rho_3^0$

Helical repeat $\nu = 10.5$

$\theta_3^0 = 360°/\nu$

$F_{11} = F_{22} = \frac{1}{\left(\sqrt{\Delta s/a} * 180/\pi\right)} = 0.0427$

$\gamma = 1.4, F_{33} = 2F_{11}/\gamma$

$\underline{p}_0$	AA/TT	AC/GT	AG/CT	AT	CA/TG	CC/GG	CG	GA/TC	GC	TA
θ_1 - Tilt	0	0	0	0	0	0	0	0	0	0
θ_2 - Roll	0	0	0	0	0	0	0	0	0	0
θ_3 - Twist	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857
ρ_1 - Shift	0	0	0	0	0	0	0	0	0	0
ρ_2 - Slide	0	0	0	0	0	0	0	0	0	0
ρ_3 - Rise	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
Force Constants F										
θ_1 - θ_1	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427
θ_2 - θ_2	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427
Θ_3 - θ_3	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597
ρ_1 - ρ_1	20	20	20	20	20	20	20	20	20	20
ρ_2 - ρ_2	20	20	20	20	20	20	20	20	20	20
ρ_3 - ρ_3	20	20	20	20	20	20	20	20	20	20

Table B-1 Ideal DNA force field.

The optimization model is adjusted to compare the effects of twist as a result of rotation of constraint vs. change in intrinsic twist on a constrained DNA loop. The rotation effect can be modeled by replacing the final reference frame R, a 3x3 matrix defining the constrained end, of the DNA loop with a rotated frame determined by:

$$R' = \left[R \cdot \begin{bmatrix} \cos\Delta\varphi & -\sin\Delta\varphi & 0 \\ \sin\Delta\varphi & \cos\Delta\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \right]^T$$

A rotation of $\Delta\varphi$ imposed on the end of a straight DNA chain of length N-bp, such that the twist is uniformly distributed along the chain, alters the helical repeat ν according to the following equation. $\nu = 360/(\Delta\varphi/N + 360/\nu_0)$. The change in intrinsic twist is modeled by adjusting θ_3 in the optimization force field. i.e., $f(\Delta\varphi, N) = 360/\nu$. In this work $\nu_0 = 10.5$ and $\Delta\varphi$ varied from -360° to 360° in 1° increments.

$\underline{p}_0$	AA/TT	AC/GT	AG/CT	AT	CA/TG	CC/GG	CG	GA/TC	GC	TA
θ_1 - Tilt	0	0	0	0	0	0	0	0	0	0
θ_2 - Roll	0	0	0	0	0	0	0	0	0	0
θ_3 - Twist	$f(\Delta\varphi, N)$	$f(\Delta\varphi, N)$	$f(\Delta\varphi, N)$	$f(\Delta\varphi, N)$	$f(\Delta\varphi, N)$	$f(\Delta\varphi, N)$	$f(\Delta\varphi, N)$	$f(\Delta\varphi, N)$	$f(\Delta\varphi, N)$	$f(\Delta\varphi, N)$
ρ_1 - Shift	0	0	0	0	0	0	0	0	0	0
ρ_2 - Slide	0	0	0	0	0	0	0	0	0	0
ρ_3 - Rise	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
Force Constants F										
θ_1 - θ_1	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427
θ_2 - θ_2	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427
Θ_3 - θ_3	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597
ρ_1 - ρ_1	20	20	20	20	20	20	20	20	20	20
ρ_2 - ρ_2	20	20	20	20	20	20	20	20	20	20
ρ_3 - ρ_3	20	20	20	20	20	20	20	20	20	20

Table B-2 Incremental twist.

Kahn sequences optimized for negative roll, θ_3, for A-tracts helical repeat ν from 9-11 bp/turn in increments of 0.1bp/turn. The force field adjusted $\theta_3 = 360°/\nu$.

$\underline{p}_0$	AA/TT	AC/GT	AG/CT	AT	CA/TG	CC/GG	CG	GA/TC	GC	TA
θ_1 - Tilt	0	0	0	0	0	0	0	0	0	0
θ_2 - Roll	-2	3	3	3	8	3	3	3	3	3
θ_3 - Twist	*f(ν)*	*f(ν)*	*f(ν)*	*f(ν)*	*f(ν)*	*f(ν)*	*f(ν)*	*f(ν)*	*f(ν)*	*f(ν)*
ρ_1 - Shift	0	0	0	0	0	0	0	0	0	0
ρ_2 - Slide	0	0	0	0	0	0	0	0	0	0
ρ_3 - Rise	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
Force Constants F										
θ_1 - θ_1	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427
θ_2 - θ_2	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427
Θ_3 - θ_3	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597
ρ_1 - ρ_1	20	20	20	20	20	20	20	20	20	20
ρ_2 - ρ_2	20	20	20	20	20	20	20	20	20	20
ρ_3 - ρ_3	20	20	20	20	20	20	20	20	20	20

Table B-3 Incremental twist as a function of helical repeat

$\underline{p}_0$	AA/TT	AC/GT	AG/CT	AT	CA/TG	CC/GG	CG	GA/TC	GC	TA
θ_1 - Tilt	0	0	0	0	0	0	0	0	0	0
θ_2 - Roll	-2	3	3	3	8	3	3	3	3	3
θ_3 - Twist	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857
ρ_1 - Shift	0	0	0	0	0	0	0	0	0	0
ρ_2 - Slide	0	0	0	0	0	0	0	0	0	0
ρ_3 - Rise	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
Force Constants F										
θ_1 - θ_1	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427
θ_2 - θ_2	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427
Θ_3 - θ_3	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597
ρ_1 - ρ_1	20	20	20	20	20	20	20	20	20	20
ρ_2 - ρ_2	20	20	20	20	20	20	20	20	20	20
ρ_3 - ρ_3	20	20	20	20	20	20	20	20	20	20

Table B-4 Curved DNA; $\theta_2 \pm 5$ constant twist

$\underline{p}_0$	AA/TT	AC/GT	AG/CT	AT	CA/TG	CC/GG	CG	GA/TC	GC	TA
θ_1 - Tilt	0	0	0	0	0	0	0	0	0	0
θ_2 - Roll	-2	3	3	3	8	3	3	3	3	3
θ_3 - Twist	35.62	34.4	27.7	31.5	34.5	33.67	29.8	36.9	40	36
ρ_1 - Shift	0	0	0	0	0	0	0	0	0	0
ρ_2 - Slide	0	0	0	0	0	0	0	0	0	0
ρ_3 - Rise	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
Force Constants F										
θ_1 - θ_1	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427
θ_2 - θ_2	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427	0.0427
Θ_3 - θ_3	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597	0.0597
ρ_1 - ρ_1	20	20	20	20	20	20	20	20	20	20
ρ_2 - ρ_2	20	20	20	20	20	20	20	20	20	20
ρ_3 - ρ_3	20	20	20	20	20	20	20	20	20	20

Table B-5 DNA curvature and twist $\theta_2 \pm 5$, $\theta_3 \pm$ determined by Kabsch et. Al. [63]

$\underline{p}_0$	AA/TT	AC/GT	AG/CT	AT	CA/TG	CC/GG	CG	GA/TC	GC	TA
θ_1 - Tilt	0	0	0	0	0	0	0	0	0	0
θ_2 - Roll	0	0	0	0	0	0	0	0	0	0
θ_3 - Twist	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857
ρ_1 - Shift	0	0	0	0	0	0	0	0	0	0
ρ_2 - Slide	0	0	0	0	0	0	0	0	0	0
ρ_3 - Rise	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
Force Constants F										
θ_1 - θ_1	0.042646	0.075816	0.042646	0.075816	0.027294	0.042646	0.027294	0.042646	0.075816	0.027294
θ_2 - θ_2	0.042646	0.075816	0.042646	0.075816	0.027294	0.042646	0.027294	0.042646	0.075816	0.027294
Θ_3 - θ_3	0.060923	0.108308	0.060923	0.108308	0.038991	0.060923	0.038991	0.060923	0.108308	0.038991
ρ_1 - ρ_1	20	20	20	20	20	20	20	20	20	20
ρ_2 - ρ_2	20	20	20	20	20	20	20	20	20	20
ρ_3 - ρ_3	20	20	20	20	20	20	20	20	20	20

Table B-6 Purine-Pyrimidine flexibility

$\underline{p}_0$	AA/TT	AC/GT	AG/CT	AT	CA/TG	CC/GG	CG	GA/TC	GC	TA
θ_1 - Tilt	0	0	0	0	0	0	0	0	0	0
θ_2 - Roll	0	0	0	0	0	0	0	0	0	0
θ_3 - Twist	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857
ρ_1 - Shift	0	0	0	0	0	0	0	0	0	0
ρ_2 - Slide	0	0	0	0	0	0	0	0	0	0
ρ_3 - Rise	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
Force Constants F										
θ_1 - θ_1	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000
θ_2 - θ_2	0.0213232	0.0213232	0.0213232	0.0213232	0.0213232	0.0213232	0.0213232	0.0213232	0.0213232	0.0213232
Θ_3 - θ_3	0.060923	0.060923	0.060923	0.060923	0.060923	0.060923	0.060923	0.060923	0.060923	0.060923
ρ_1 - ρ_1	20	20	20	20	20	20	20	20	20	20
ρ_2 - ρ_2	20	20	20	20	20	20	20	20	20	20
ρ_3 - ρ_3	20	20	20	20	20	20	20	20	20	20

Table B-7 Hinge bending

$\underline{p}_0$	AA/TT	AC/GT	AG/CT	AT	CA/TG	CC/GG	CG	GA/TC	GC	TA
θ_1 - Tilt	0	0	0	0	0	0	0	0	0	0
θ_2 - Roll	0	0	0	0	0	0	0	0	0	0
θ_3 - Twist	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857
ρ_1 - Shift	0	0	0	0	0	0	0	0	0	0
ρ_2 - Slide	0	0	0	0	0	0	0	0	0	0
ρ_3 - Rise	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
Force Constants F										
θ_1 - θ_1	10000	17777.78	10000	17777.78	6400	10000	6400	10000	17777.78	6400
θ_2 - θ_2	0.0213232	0.0379079	0.0213232	0.0379079	0.013647	0.0213232	0.013647	0.0213232	0.0379079	0.013647
Θ_3 - θ_3	0.0609235	0.1083084	0.0609235	0.1083084	0.038991	0.0609235	0.038991	0.0609235	0.1083084	0.038991
ρ_1 - ρ_1	20	20	20	20	20	20	20	20	20	20
ρ_2 - ρ_2	20	20	20	20	20	20	20	20	20	20
ρ_3 - ρ_3	20	20	20	20	20	20	20	20	20	20

Table B-8 Hinge bending + YR flexibility

$\underline{p}_0$	AA/TT	AC/GT	AG/CT	AT	CA/TG	CC/GG	CG	GA/TC	GC	TA
θ_1 - Tilt	0	0	0	0	0	0	0	0	0	0
θ_2 - Roll	0	0	0	0	0	0	0	0	0	0
θ_3 - Twist	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857	34.2857
ρ_1 - Shift	0	0	0	0	0	0	0	0	0	0
ρ_2 - Slide	0	0	0	0	0	0	0	0	0	0
ρ_3 - Rise	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
Force Constants F										
θ_1 - θ_1	10000	10000	10000	10000	10000	10000	10000	10000	10000	10000
θ_2 - θ_2	0.0213232	0.0213232	0.0213232	0.0213232	0.0213232	0.0213232	0.0213232	0.0213232	0.0213232	0.0213232
θ_2 - θ_3	0.0341669	0.0341669	0.0341669	0.0341669	0.0341669	0.0341669	0.0341669	0.0341669	0.0341669	0.0341669
Θ_3 - θ_2	0.0341669	0.0341669	0.0341669	0.0341669	0.0341669	0.0341669	0.0341669	0.0341669	0.0341669	0.0341669
Θ_3 - θ_3	0.0609235	0.0609235	0.0609235	0.0609235	0.0609235	0.0609235	0.0609235	0.0609235	0.0609235	0.0609235
ρ_1 - ρ_1	20	20	20	20	20	20	20	20	20	20
ρ_2 - ρ_2	20	20	20	20	20	20	20	20	20	20
ρ_3 - ρ_3	20	20	20	20	20	20	20	20	20	20

Table B-9 Hinge bending and roll/twist coupling

Appendix C Protein Deformation Supplemental data

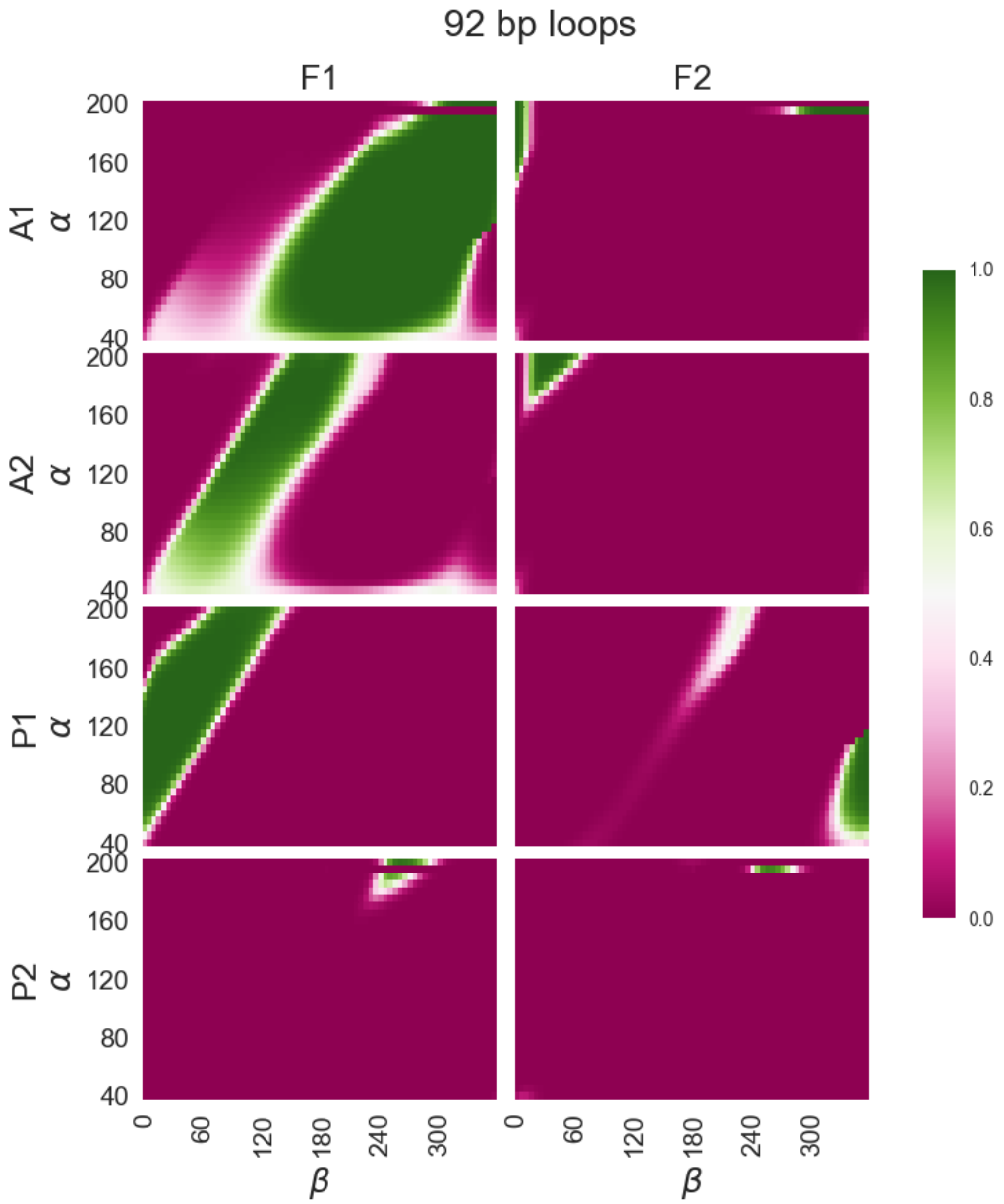

Figure C.1 Fractional populations for 92 bp loops for each structural family and operator orientation.

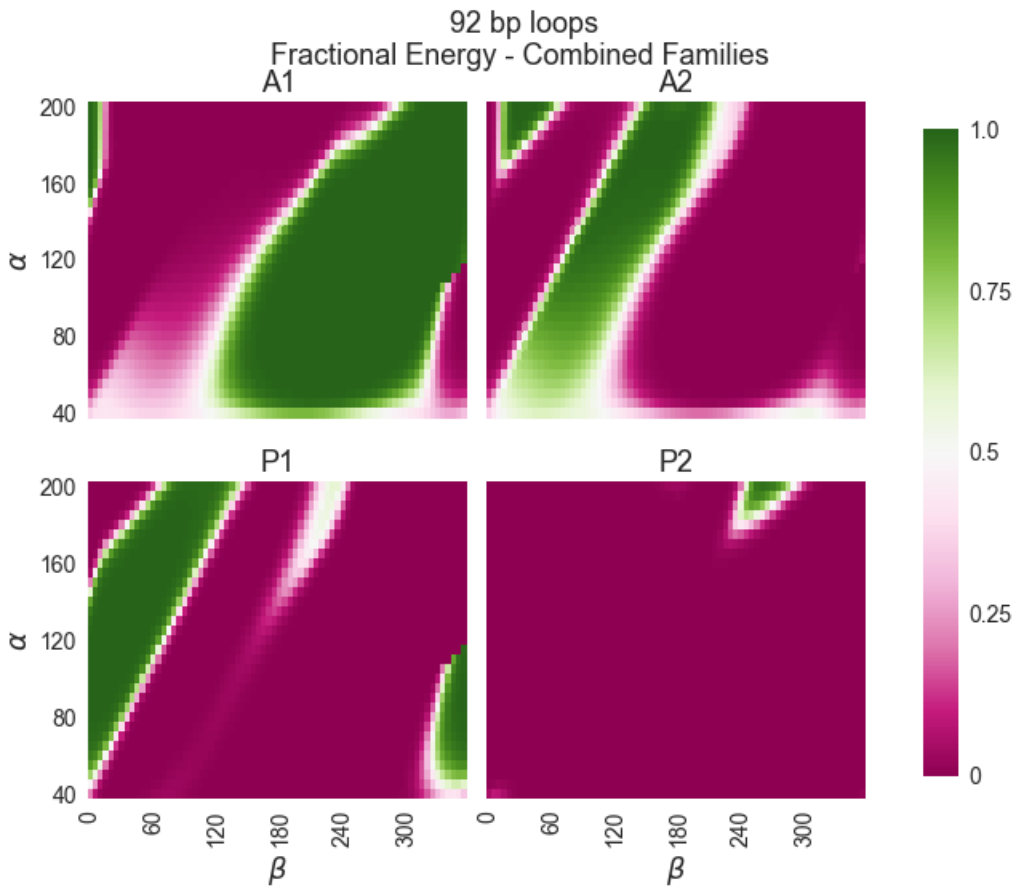

Figure C.2 Fractional Contribution by loop type 92 bp loop

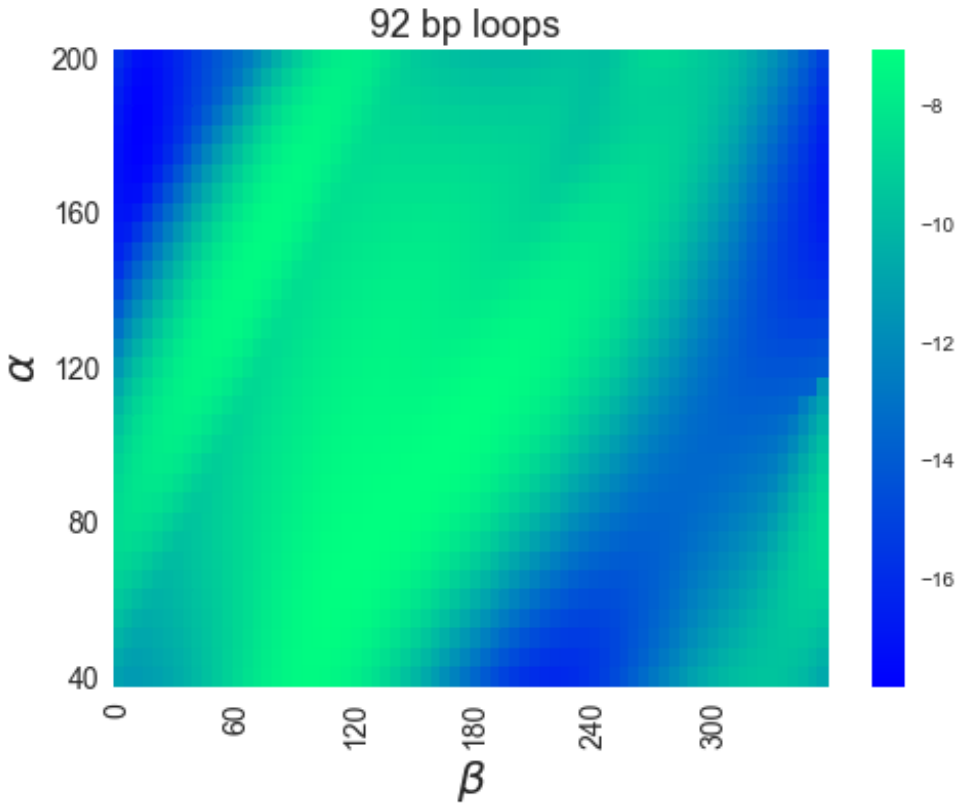

Figure C.3 92 bp loops J factors

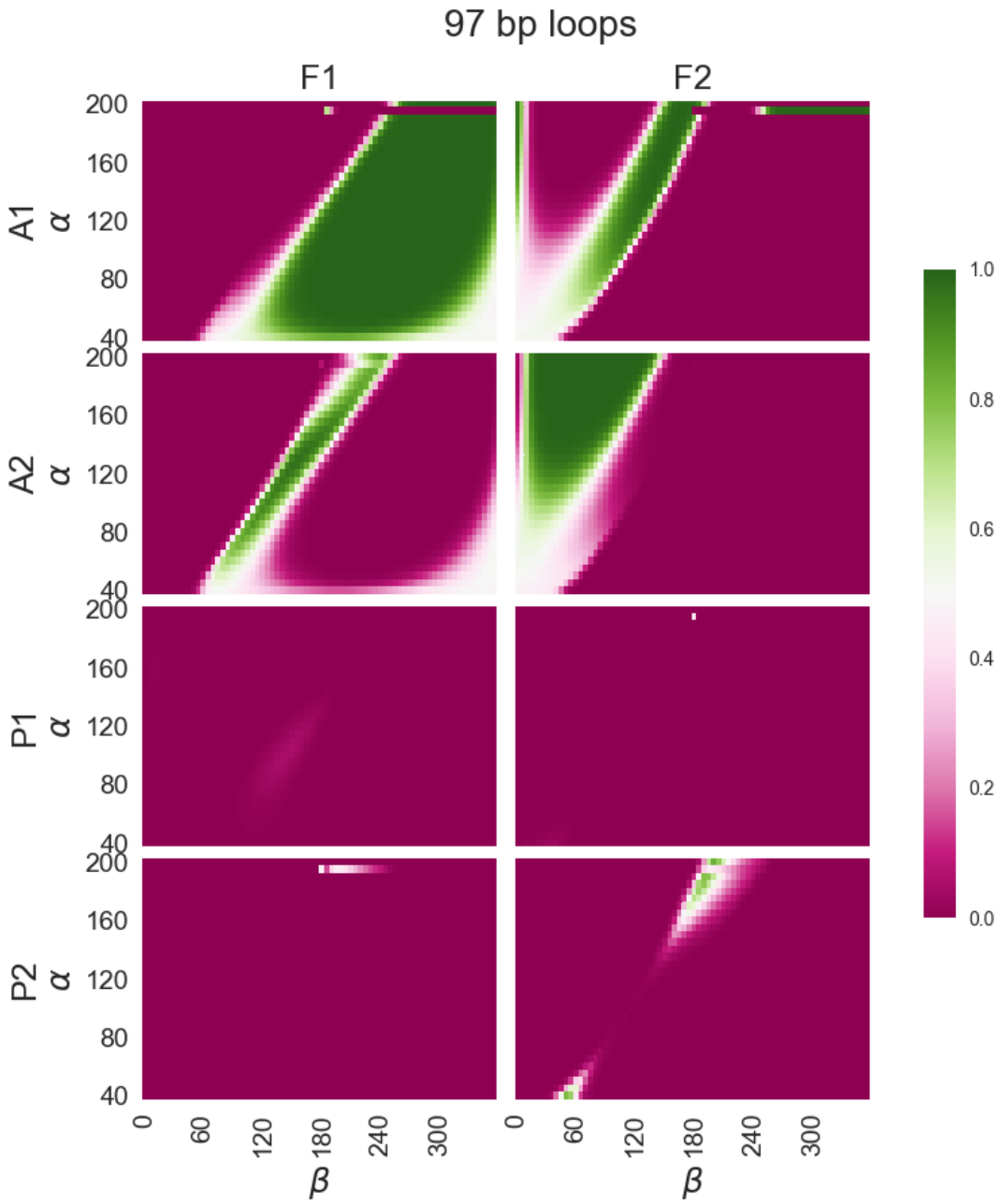

Figure C.4 Fractional populations for 97 bp loops for each structural family and operator orientation.

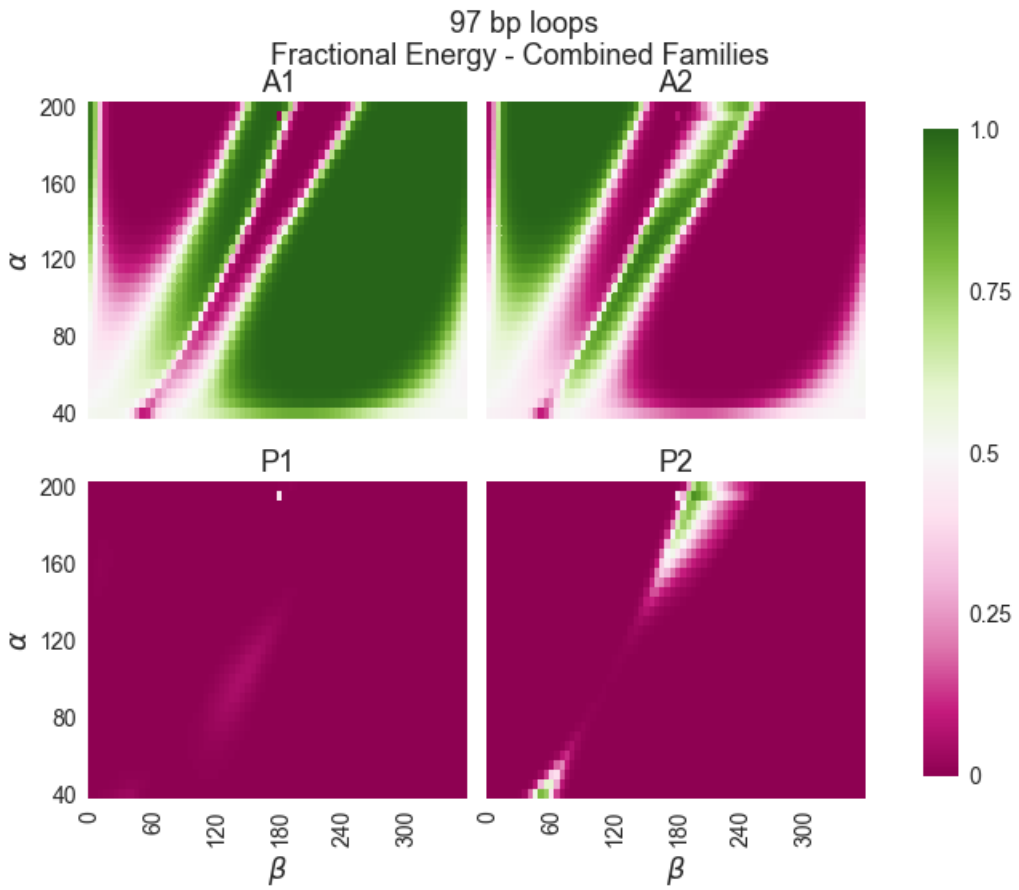

Figure C.5 97-bp loop fractional populations by loop type

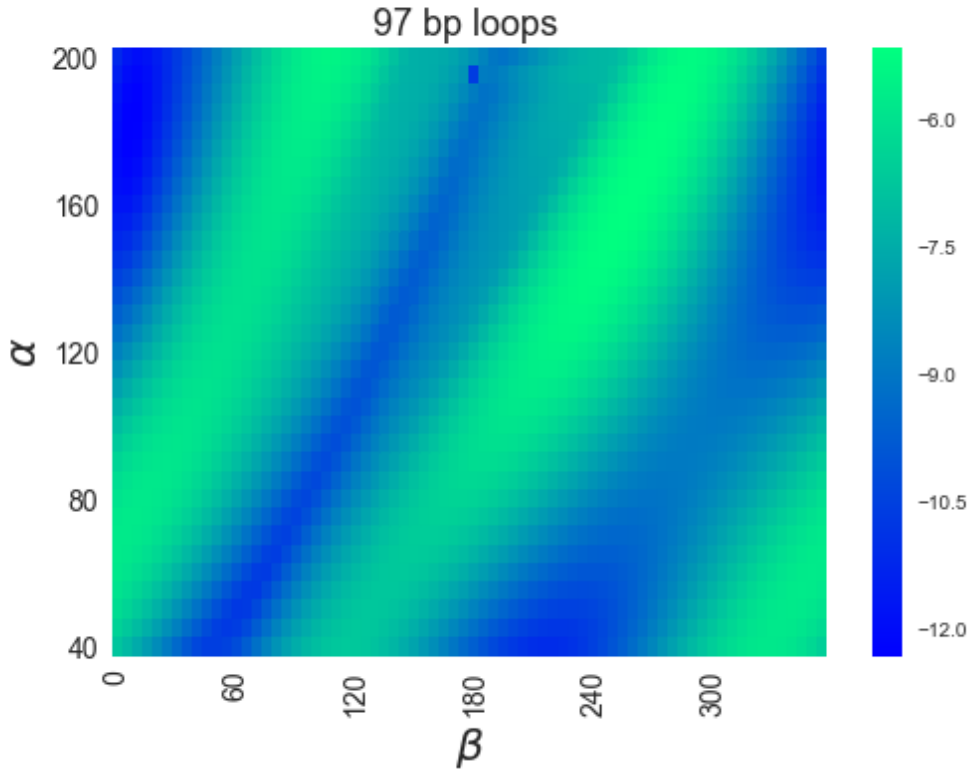

Figure C.6 97 bp loop J factors

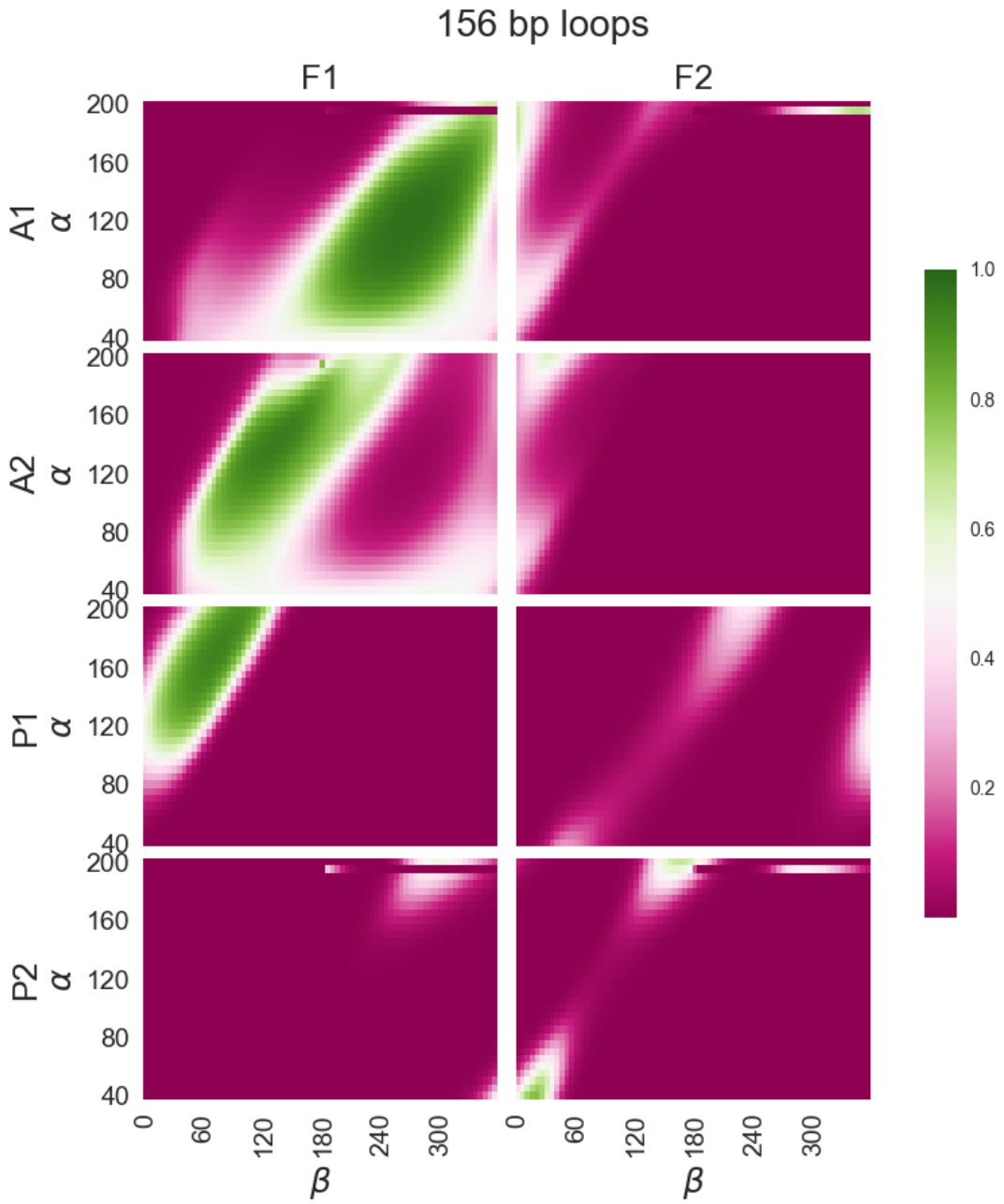

Figure C.7 Fractional populations for 156 bp loops for each structural family and operator orientation

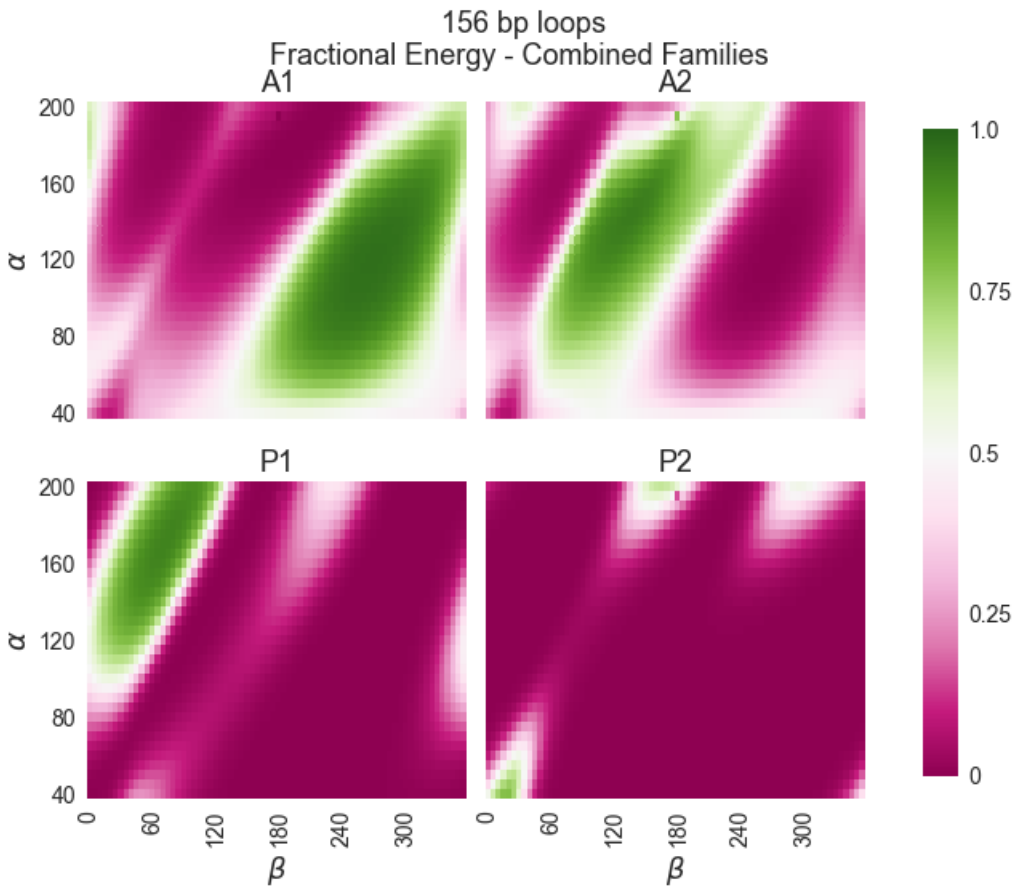

Figure C.8 156 bp loop fractional populations by loop type

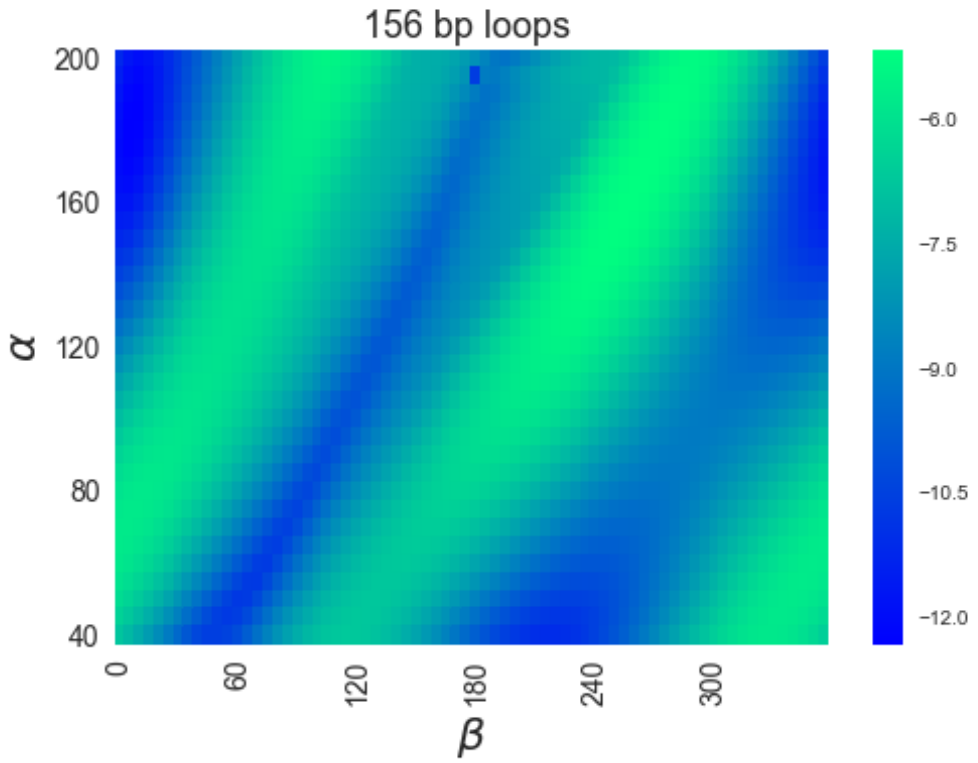

Figure C.9 156 bp loop J factors

References

1. Matthews, K. S. DNA looping. *Microbiol Rev* **56,** 123–136 (1992).

2. Postow, L., Hardy, C. D., Arsuaga, J. & Cozzarelli, N. R. Topological domain structure of the Escherichia coli chromosome. *Genes Dev.* **18,** 1766–79 (2004).

3. Müller-Hill, B. *The lac Operon; A Short History of a Genetic Paradigm*. (Walter de Gruyter, 1996).

4. Bell, C. E. & Lewis, M. A closer view of the conformation of the Lac repressor bound to operator. *Nat Struct Biol* **7,** 209–214 (2000).

5. Bell, C. E. & Lewis, M. Crystallographic analysis of Lac repressor bound to natural operator O1. *J Mol Biol* **312,** 921–926 (2001).

6. Geanacopoulos, M., Vasmatzis, G., Zhurkin, V. B. & Adhya, S. Gal repressosome contains an antiparallel DNA loop. *Nat Struct Biol* **8,** 432–436 (2001).

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

8. Normanno, D., Vanzi, F. & Pavone, F. S. Single-molecule manipulation reveals supercoiling-dependent modulation of lac repressor-mediated DNA looping. *Nucleic Acids Res* **36,** 2505–2513 (2008).

9. Swigon, D., Coleman, B. D. & Olson, W. K. Modeling the Lac repressor-operator assembly: the influence of DNA looping on Lac repressor conformation. *Proc Natl Acad Sci U S A* **103,** 9879–9884 (2006).

10. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3,** 318–356 (1961).

11. Friedman, A. M., Fischmann, T. 0 & Steitz, T. A. Crystal Structure of lac Repressor Core Tetramer and Its Implications for DNA Looping. *Science (80-.).* **268,** 1721–1727 (1995).

12. Taraban, M. *et al.* Ligand-induced conformational changes and conformational dynamics in the solution structure of the lactose repressor protein. *J Mol Biol* **376,** 466–481 (2008).

13. Ruben, G. C. & Roos, T. B. Conformation of Lac repressor tetramer in solution, bound and unbound to operator DNA. *Microsc. Res. Tech.* **36,** 400–416 (1997).

14. Mehta, R. A. & Kahn, J. D. Designed hyperstable lac Repressor·DNA loop topologies suggest alternative loop geometries. *J. Mol. Biol.* **294,** 67–77 (1999).

15. Edelman, L. M., Cheong, R. & Kahn, J. D. Fluorescence resonance energy transfer over approximately 130 basepairs in hyperstable lac repressor-DNA loops. *Biophys J* **84,** 1131–1145 (2003).

16. Haeusler, A. R. *et al.* FRET studies of a landscape of Lac repressor-mediated DNA loops. *Nucleic Acids Res* **40,** 4432–4445 (2012).

17. Goodson, K. A., Wang, Z., Haeusler, A. R., Kahn, J. D. & English, D. S. LacI-DNA-IPTG loops: equilibria among conformations by single-molecule FRET. *J Phys Chem B* **117,** 4713–4722 (2013).

18. Spronk, C. A. E. M., Slijper, M., van Boom, J. H., Kaptein, R. & Boelens, R. Formation of the hinge helix in the lac repressor is induced upon binding to the lac operator. *Nat Struct Biol* **3,** 916–919 (1996).

19. Spronk, C. A. E. M. *et al.* The solution structure of Lac repressor headpiece 62 complexed to a symmetrical lac operator. *Structure* **7,** 1483-S3 (1999).

20. Kalodimos, C. G. *et al.* Plasticity in protein-DNA recognition: lac repressor interacts with its natural operator 01 through alternative conformations of its DNA-binding domain. *EMBO J* **21,** 2866–2876 (2002).

21. Wong, O. K., Guthold, M., Erie, D. A., Gelles, J. & Cozzarelli, N. Interconvertible lac repressor-DNA loops revealed by single-molecule experiments. *PLoS Biol.* **6,** e232 (2008).

22. Han, L. *et al.* Concentration and length dependence of DNA looping in transcriptional regulation. *PLoS One* **4,** e5621 (2009).

23. Rutkauskas, D., Zhan, H., Matthews, K. S., Pavone, F. S. & Vanzi, F. Tetramer opening in LacI-mediated DNA looping. *Proc Natl Acad Sci U S A* **106,** 16627–16632 (2009).

24. Hirsh, A. D., Lillian, T. D., Lionberger, T. A. & Perkins, N. C. DNA modeling reveals an extended lac repressor conformation in classic in vitro binding assays. *Biophys J* **101,** 718–726 (2011).

25. Swigon, D. & Olson, W. K. Mesoscale modeling of multi-protein-DNA assemblies: the role of the catabolic activator protein in Lac-repressor-mediated looping. *Int. J. Non. Linear. Mech.* **43,** 1082–1093 (2008).

26. Hawkins, R. J. & Mcleish, T. C. B. Coarse-Grained Model of Entropic Allostery. *Phys Rev Lett* **93,** 98104 (2004).

27. Perez, P. J. & Olson, W. K. Insights into genome architecture deduced from the properties of short Lac repressor-mediated DNA loops. *Biophys. Rev.* **8,** 135–144 (2016).

28. Lewis, M. *et al.* Crystal Structure of the Lactose Operon Repressor and Its Complexes with DNA and Inducer. *Science (80-.).* **271,** 1247–1254 (1996).

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

30. Krämer, H. *et al.* lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. *EMBO J.* **6,** 1481–1491 (1987).

31. Britton, L. A., Olson, W. K. & Tobias, I. Two perspectives on the twist of DNA. *J Chem Phys* **131,** 245101 (2009).

32. Clauvelin, N. & Olson, W. K. The synergy between protein positioning and DNA elasticity: energy minimization of protein-decorated DNA minicircles. (2014).

33. White, J. H. & Bauer, W. R. Superhelical DNA with local substructures. A generalization of the topological constraint in terms of the intersection number and the ladder-like correspondence surface. *J Mol Biol* **195,** 205–213 (1987).

34. Coleman, B. D., Olson, W. K. & Swigon, D. Theory of sequence-dependent DNA elasticity. *J. Chem. Phys.* **118,** 7127–7140 (2003).

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

36. Colasanti, A. V *et al.* Weak operator binding enhances simulated lac repressor-mediated DNA looping. *Biopolymers* **99,** 1070–81 (2013).

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

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

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

40. Heath, P. J., Clendenning, J. B., Fujimoto, B. S. & Schurr, J. M. Effect of bending strain on the torsion elastic constant of DNA. *J Mol Biol* **260,** 718–730 (1996).

41. Johnson, S., Lindén, M. & Phillips, R. Sequence dependence of transcription factor-mediated DNA looping. *Nucleic Acids Res.* **40,** 7728–7738 (2012).

42. Vörös, Z., Yan, Y., Kovari, D. T., Finzi, L. & Dunlap, D. Proteins mediating DNA loops effectively block transcription. *Protein Sci.* (2017). doi:10.1002/pro.3156

43. Czapla, B. Y. L., Grosner, M. A., Swigon, D. & Olson, W. K. Interplay of Protein and DNA Structure Revealed in Simulations of the lac Operon. *PLoS One* **8,** e56548 (2013).

44. Jacobson, H. & Stockmayer, W. H. Intramolecular Reaction in Polycondensations. I. The Theory of Linear Systems. *J. Chem. Phys.* **18,** 1600 (1950).

45. Flory, P. J., Suter, U. W. & Mutter, M. Macrocyclization equilibriums. 1. Theory. *J. Am. Chem. Soc.* **98,** 5733–5739 (1976).

46. Crothers, D. M., Haran, T. E. & Nadeau, J. G. Intrinsically bent DNA. *J Biol Chem* **265,** 7093–7096 (1990).

47. Johnson, S., Chen, Y.-J. & Phillips, R. Poly(dA:dT)-rich DNAs are highly flexible in the context of DNA looping. *PLoS One* **8,** e75799 (2013).

48. Wong, O. K., Guthold, M., Erie, D. A., Gelles, J. & Cozzarelli, N. Interconvertible Lac Repressor–DNA Loops Revealed by Single-Molecule Experiments. *PLoS Biol.* **6,** e232 (2008).

49. Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. Bent helical structure in kinetoplast DNA. *Proc. Natl. Acad. Sci. U. S. A.* **79,** 7664–8 (1982).

50. Wu, H. M. & Crothers, D. M. The locus of sequence-directed and protein-induced DNA bending. *Nature* **308,** 509–13

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

52. Boedicker, J. Q., Garcia, H. G., Johnson, S. & Phillips, R. DNA sequence-dependent mechanics and protein-assisted bending in repressor-mediated loop formation. *Phys Biol* **10,** 66005 (2013).

53. Becker, N. A., Greiner, A. M., Peters, J. P. & Maher, J. L. I. Bacterial promoter repression by DNA looping without protein-protein binding competition. *Nucleic Acids Res* **42,** 5495–5504 (2014).

54. Wang, J. C. Helical repeat of DNA in solution (DNA structure/linking number/DNA topoisomerase/closed circular DNA). *Proc. Nati. Acad. Scd. USA* **76,** 200–203 (1979).

55. Peck, L. J. & Wang, J. C. Sequence dependence of the helical repeat of DNA in solution. *Nature* **292,** 375–378 (1981).

56. Rhodes, D. & Klug, A. Sequence-dependent helical periodicity of DNA. *Nature* **292,** 378–380 (1981).

57. Arnott, S. & Selsing, E. The structure of polydeoxyguanylic acid · polydeoxycytidylic acid. *J. Mol. Biol.* **88,** 551–IN53 (1974).

58. Arnott, S. & Selsing, E. Structures for the polynucleotide complexes poly(dA) · poly(dT) and poly(dT) · poly(dA) · poly(dT). *J. Mol. Biol.* **88,** 509–521 (1974).

59. Ivanov, V. I. *et al.* CRP-DNA Complexes: Inducing theA-likeForm in the Binding Sites with an Extended Central Spacer. *J. Mol. Biol.* **245,** 228–240 (1995).

60. Lu, X.-J. & Olson, W. K. 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31,** 5108–5121 (2003).

61. Selsing, E., Wells, R. D., Alden, C. J. & Arnott, S. Bent DNA: Visualization of a Base-paired and Stacked A-B Conformational Junction*. *J. Biol. Chem.* **254,** 5417–5422 (1979).

62. Zhurkin, V. B., Tolstorukov, M. Y., Xu, F., Colasanti, A. V & Olson, W. K. Sequence-Dependent Variability of B-DNA: An update on bending and curvature. *DNA Conform. Transcr.* 18–34 (2005). doi:10.1007/0-387-29148-2_2

63. Kabsch, W., Sander, C. & Trifonov, E. N. The ten helical twist angles of B-DNA. *Nucleic Acids Res.* **10,** 1097–1104 (1982).

64. Brukner, I., Sánchez, R., Suck, D. & Pongor, S. Trinucleotide models for DNA bending propensity: comparison of models based on DNaseI digestion and nucleosome packaging data. *J. Biomol. Struct. Dyn.* **13,** 309–17 (1995).

65. Balasubramanian, S., Xu, F. & Olson, W. K. DNA Sequence-Directed Organization of Chromatin: Structure-Based Computational Analysis of Nucleosome-Binding Sequences. *Biophysj* **96,** 2245–2260 (2009).

66. Mckay, D. B., Pickover, C. A. & Steitz, T. A. Escherichia coli lac repressor is elongated with its operator DNA binding domains located at both ends. *J. Mol. Biol.* **156,** 175–183 (1982).

67. Czapla, L., Peters, J. P., Rueter, E. M., Olson, W. K. & Maher, L. J. Understanding Apparent DNA Flexibility Enhancement by HU and HMGB Architectural Proteins. *J. Mol. Biol.* **409,** 278–289 (2011).

68. Swinger, K. K., Lemberg, K. M., Zhang, Y. & Rice, P. A. Flexible DNA bending in HU-DNA cocrystal structures. *EMBO J.* **22,** 3749–3760 (2003).

69. Spronk, C. A. E. M. *et al.* Hinge-helix formation and DNA bending in various lac repressor-operator complexes. *EMBO J.* **18,** 6472–80 (1999).

70. Matsumoto, A., Tobias, I. & Olson, W. K. Normal-Mode Analysis of Circular DNA at the Base-Pair Level. 1. Comparison of Computed Motions with the Predicted Behavior of an Ideal Elastic Rod. *J. Chem. Theory Comput.* **1,** 117–129 (2005).

71. Atsushi Matsumoto, †, ‡, Irwin Tobias, † and & Wilma K. Olson*, †. Normal-Mode Analysis of Circular DNA at the Base-Pair Level. 2. Large-Scale Configurational Transformation of a Naturally Curved Molecule. **1,** 130–142 (2005).

72. Schellman, J. A. Flexibility of DNA. *Biopolymers* **13,** 217–226 (1974).

73. Schellman, J. A. The flexibility of DNA: I. Thermal fluctuations. *Biophys. Chem.* **11,** 321–328 (1980).

74. Tolstorukov, M. Y., Colasanti, A. V, McCandlish, D. M., Olson, W. K. & Zhurkin, V. B. A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. *J Mol Biol* **371,** 725–738 (2007).

75. Koo, H. S., Drak, J., Rice, J. A. & Crothers, D. M. Determination of the extent of DNA bending by an adenine-thymine tract. *Biochemistry* **29,** 4227–34 (1990).

76. Tchernaenko, V. *et al.* Topological Measurement of an A-tract Bend Angle: Comparison of the Bent and Straightened States. *J. Mol. Biol.* **326,** 737–749 (2003).

77. Goyal, S. *et al.* Intrinsic curvature of DNA influences LacR-mediated looping. *Biophys J* **93,** 4342–4359 (2007).

78. Dickerson, R. E. & Drew, H. R. Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *J. Mol. Biol.* **149,** 761–86 (1981).

79. Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M. & Zhurkin, V. B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 11163–8 (1998).

80. Gorin, A. A., Zhurkin, V. B. & Olson, W. K. B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* **247,** 34–48 (1995).

81. Vlahoviček, K., Kaján, L. & Pongor, S. DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucleic Acids Res.* **31,** 3686–3687 (2003).

82. Lu, X.-J. & Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **3,** 1213–1227 (2008).

83. Dickerson, R. E. *et al.* Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.* **17,** 1797–1803 (1989).

84. Zhang, Y., McEwen, A. E., Crothers, D. M. & Levene, S. D. Statistical-mechanical theory of DNA looping. *Biophys J* **90,** 1903–1912 (2006).

85. Zhang, Y., McEwen, A. E., Crothers, D. M. & Levene, S. D. Analysis of in-vivo LacR-mediated gene repression based on the mechanics of DNA looping. *PLoS One* **1,** e136 (2006).

86. Lillian, T. D., Goyal, S., Kahn, J. D., Meyhöfer, E. & Perkins, N. C. Computational analysis of looping of a large family of highly bent DNA by LacI. *Biophys J* **95,** 5832–5842 (2008).

87. Becker, N. A. & Maher, J. L. I. High-resolution mapping of architectural DNA binding protein facilitation of a DNA repression loop in Escherichia coli. *Proc. Natl. Acad. Sci.* **112,** 7177–7182 (2015).

88. Griffith, J., Bleyman, M., Rauch, C. A., Kitchin, P. A. & Englund, P. T. Visualization of the bent helix in kinetoplast DNA by electron microscopy. *Cell* **46,** 717–24 (1986).

89. Wei, J., Czapla, B. Y. L., Grosner, M. A., Swigon, D. & Olson, W. K. DNA topology confers sequence specificity to nonspecific architectural proteins. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 16742–7 (2014).

90. Daber, R., Stayrook, S., Rosenberg, A. & Lewis, M. Structural analysis of lac repressor bound to allosteric effectors. *J Mol Biol* **370,** 609–619 (2007).

91. Stenberg, K. A. E. & Vihinen, M. Crystal structure of a 1.6-hexanediol bound tetrameric form of Escherichia coli lac-repressor refined to 2.1 A resolution. *Proteins* **75,** 748–759 (2009).

92. Kalodimos, C. G. *et al.* Structure and Flexibility Adaptation in Nonspecific and Specific Protein-DNA Complexes. *Science (80-.).* **305,** (2004).

93. Romanuka, J. *et al.* Specificity and affinity of Lac repressor for the auxiliary operators O2 and O3 are explained by the structures of their protein-DNA complexes. *J Mol Biol* **390,** 478–489 (2009).

94. Jobe, A. & Bourgeois, S. lac repressor-operator interaction. *J. Mol. Biol.* **69,** 397–408 (1972).

95. Steitz, T. A., Richmond, T. J., Wise, D. & Engelman, D. The lac Repressor Protein: Molecular Shape, Subunit Structure, and Proposed Model for Operator Interaction Based on Structural Studies of Microcrystals. *Proc. Natl. Acad. Sci.* **71,** 593–597 (1974).

96. Spolar, R. S. & Record, M. T. Coupling of local folding to site-specific binding of proteins to DNA. *Science (80-.).* **263,** 777–784 (1994).

97. Morgan, M. A., Okamoto, K., Kahn, J. D. & English, D. S. Single-molecule spectroscopic determination of lac repressor-DNA loop conformation. *Biophys J* **89,** 2588–2596 (2005).

98. Towles, K. B., Beausang, J. F., Garcia, H. G., Phillips, R. & Nelson, P. C. First-principles calculation of DNA looping in tethered particle experiments. *Phys Biol* **6,** 25001 (2009).

99. Tsodikov, O. V *et al.* Wrapping of flanking non-operator {DNA} in lac repressor-operator complexes: implications for {DNA} looping. *J. Mol. Biol.* **294,** 639–655 (1999).

100. Virnik, K. *et al.* Antiparallel DNA loop in gal repressosome visualized by atomic force microscopy. *J Mol Biol* **334,** 53–63 (2003).

101. Olson, W. K., Grosner, M. A., Czapla, B. Y. L. & Swigon, D. Structural insights into the role of architectural proteins in DNA looping deduced from computer simulations. *Biochem. Soc. Trans.* **41,** 559–64 (2013).

102. Villa, E., Balaeff, A. & Schulten, K. Structural dynamics of the lac repressor-DNA complex revealed by a multiscale simulation. *Proc Natl Acad Sci U S A* **102,** 6783–6788 (2005).

103. Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science* **266,** 763–70 (1994).

104. Dillon, S. C. & Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* **8,** 185–195 (2010).

105. Cavalli, G. & Misteli, T. Functional implications of genome topology. (2013). doi:10.1038/nsmb.2474

106. Dekker, J. & Misteli, T. Long-Range Chromatin Interactions. *Cold Spring Harb. Perspect. Biol.* **7,** a019356 (2015).

107. Liu, Z., Deibler, R. W., Chan, H. S. & Zechiedrich, L. The why and how of DNA unlinking. *Nucleic Acids Res.* **37,** 661–71 (2009).

108. Bondarenko, V. A., Liu, Y. V, Jiang, Y. I. & Studitsky, V. M. Communication over a large distance: enhancers and insulators. *Biochem. Cell Biol.* **81,** 241–251 (2003).

109. Leng, F., Chen, B. & Dunlap, D. D. Dividing a supercoiled DNA molecule into two independent topological domains. *Proc Natl Acad Sci U S A* **108,** 19973–19978 (2011).

110. Borowiec, J. A., Zhang, L., Sasse-Dwight, S. & Gralla, J. D. DNA supercoiling promotes formation of a bent repression loop in lac DNA. *J. Mol. Biol.* **196,** 101–111 (1987).

111. Whitson, P. A., Hsieh, W. T., Wells, R. D. & Matthews, K. S. Supercoiling facilitates lac operator-repressor-pseudooperator interactions. *J. Biol. Chem.* **262,** 4943–6 (1987).

112. Krämer, H., Amouyal, M., Nordheim, A. & Müller-Hill, B. DNA supercoiling changes the spacing requirement of two lac operators for DNA loop formation with lac repressor. *EMBO J.* **7,** 547–56 (1988).

113. Eismann, E. R. & Müller-Hill, B. lac Repressor forms stable loops in vitro with supercoiled wild-type lac DNA containing all three natural lac operators. *J. Mol. Biol.* **213,** 763–775 (1990).

114. Tobias, I., Coleman, B. D. & Olson, W. K. The dependence of DNA tertiary structure on end conditions: Theory and implications for topological transitions. *J. Chem. Phys.* **101,** 10990–10996 (1994).

115. Becker, N. A., Kahn, J. D. & Maher, J. L. I. Bacterial repression loops require enhanced DNA flexibility. *J Mol Biol* **349,** 716–730 (2005).

116. Czapla, B. Y. L., Swigon, D. & Olson, W. K. Sequence-dependent effects in the cyclization of short DNA. *J. Chem. Theory Comput.* **2,** 685–695 (2006).

117. Clauvelin, N., Olson, W. K. & Tobias, I. Characterization of the geometry and topology of DNA pictured as a discrete collection of atoms. *J Chem Theory Comput* **8,** 1092–1107 (2012).

118. White, J. H. Self-Linking and the Gauss Integral in Higher Dimensions. *Am. J. Math.* **91,** 693 (1969).

119. Bondarenko, V. A., Jiang, Y. I. & Studitsky, V. M. Rationally designed insulator-like elements can block enhancer action in vitro. *EMBO J* **22,** 4728–4737 (2003).

120. Huang, J., Schlick, T. & Vologodskii, A. V. Dynamics of site juxtaposition in supercoiled DNA. *Proc Natl Acad Sci U S A* **98,** 968–973 (2001).

121. Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164,** 1110–1121 (2016).

122. Perez, P. J., Clauvelin, N., Grosner, M. A., Colasanti, A. V & Olson, W. K. What controls DNA looping? *Int. J. Mol. Sci.* **15,** 15090–108 (2014).