

© 2022

Robert T. Young

ALL RIGHTS RESERVED



STRUCTURAL VARIATION IN CIRCULAR DNA INFLUENCED BY  
NUCLEOTIDE SEQUENCE AND ROTATIONAL RIGID BODY PARAMETERS

By

ROBERT T. YOUNG

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Physics & Astronomy

Written under the direction of

Dr. Wilma K. Olson

And approved by

---

---

---

---

New Brunswick, New Jersey

January 2022

# **ABSTRACT OF THE THESIS**

## **STRUCTURAL VARIATION IN CIRCULAR DNA INFLUENCED BY NUCLEOTIDE SEQUENCE AND ROTATIONAL RIGID BODY PARAMETERS**

by ROBERT T. YOUNG

Thesis Director:

Dr. Wilma K. Olson

The physical configuration of double-stranded DNA plays a significant role in the protection and expression of genetic material. Studies of biologically relevant DNA configurations show the importance of covalently-closed, circular structures with lengths less than one thousand base pairs. Smaller circular forms must contend with thermodynamic barriers, such as differences between bending and torsional properties of DNA, that may prevent them from reaching an energetically stable, low-energy state. An elastic energy minimization study was conducted using a base-pair step harmonic potential function to observe deformations of a set of DNA minicircles. The topoisomers, of length 150 base pairs, vary in both linking number and the nucleotide coding sequence, comprised of equal numbers of the purine nucleotides adenine and guanine. The initial portion of the study focuses on differences between optimized states found with two homopolymeric models that represent B-form DNA. Additional complexity was introduced by using sequence-specific intrinsic states, incorporating differences in the twist and the roll component of the intrinsic bend at AA, AG, GA, and GG dimer steps. Results show connections between optimized energetic states and the uptake of twist of DNA as well as

changes in local bend and twist, producing high-energy, out-of-plane deformations of the global structures of circles comprised of specific repeated tracts of adenine and guanine. The structural distortions occur in  $A_4G_4$  and  $A_5G_5$  sequence motifs in the form of either kinked polygonal optimized states or states of out-of-plane bending. The connection between global distortions and sequence found here may be related to effects sequences seen in naturally occurring DNA circles. Further studies of chains made up of other sequence motifs or incorporating more realistic modeling of DNA, such as the known elastic coupling between bend and twist, are needed to establish firm connections between sequence and global DNA structure.

## ACKNOWLEDGEMENTS & DEDICATION

The work presented in this thesis could not have been done if it were not for the support and encouragement of my colleagues in the Olson Biopolymer Structures Laboratory. I want to thank Dr. Wilma Olson for her mentorship, support, and encouragement in my graduate career. Thank you, also, to Dr. Pamela Perez and Dr. Stefford Todolli for their guidance, training, and camaraderie through this study and with my dissertation work. An additional thank you to Dr. Charlotte Baker for her support both on this project and throughout my academic career. Finally, a big thank you to my parents, Gary and Sandra Young.

Dedicated to first generation college students everywhere.

Yes, you *can* do it.

# TABLE OF CONTENTS

<b>ABSTRACT OF THE THESIS</b>	<b>ii</b>
<b>ACKNOWLEDGEMENTS &amp; DEDICATION</b>	<b>iv</b>
<b>TABLE OF CONTENTS</b>	<b>v</b>
<b>LIST OF TABLES</b>	<b>vii</b>
<b>LIST OF FIGURES</b>	<b>viii</b>
<b>LIST OF DEFINITIONS</b>	<b>xi</b>
<b>INTRODUCTION</b>	<b>1</b>
<i>Beyond the Chromosome: Circular Forms of DNA</i>	4
<i>Use of in Silico Methods to Observe Sequence-Dependent DNA Behavior</i>	8
<b>METHODOLOGY &amp; CALCULATIONS</b>	<b>13</b>
<i>Initial Circular Models</i>	13
<i>Elastic Energetic Optimization with emDNA</i>	17
<b>RESULTS &amp; DISCUSSION</b>	<b>23</b>
<i>Elastic Energy Minimization Using Homopolymeric Forcefields</i>	24
<i>Incorporating Sequence-dependent Intrinsic Twist and Roll</i>	28
I. Optimization with the Trifonov Rest-State Model	30
II. Optimization with the Zhurkin Rest-State Model	33
III. Optimization with the Combined Zhurkin/Trifonov Model	40
<b>CONCLUSION</b>	<b>44</b>

<b>APPENDIX</b>	<b>47</b>
<i>A.I. DNA: Building from the Base Up</i>	47
<i>A.II. The Base-Pair Step</i>	51
<i>A.III. Elastic Energy Minimization Calculations</i>	56
<i>A.IV. Supplemental Results</i>	59
<b>BIBLIOGRAPHY</b>	<b>70</b>



## LIST OF TABLES

TABLE 1. THE BASE-PAIR STEP PARAMETERS ASSOCIATED WITH THE RIGHT-HANDED A-, B-, AND C-DNA HELICES. THESE PARAMETER VALUES ARE FROM CANONICAL FIBER MODELS AND CAN BE READILY FOUND AT THE WEB 3DNA SERVER (WEB.X3DNA.ORG) <sup>16,17</sup> .	10
TABLE 2. THE INTRINSIC BASE-PAIR STEP PARAMETERS OF THE TWO HOMOPOLYMERIC FORCEFIELDS IDEALDNA AND BDNA98.	20
TABLE 3. THE INTRINSIC BASE-PAIR STEP PARAMETERS USED IN THE TRIFONOV, ZHURKIN, AND THE COMBINED ZHURKIN/TRIFONOV SEQUENCE-DEPENDENT FORCEFIELDS.	20
TABLE 4. SUMMARY OF THE INITIAL CIRCULAR STRUCTURAL DETAILS SEPARATED BY STRUCTURAL CONSTRUCTION AND SEQUENCE COMPOSITION. FOR ADDITIONAL DETAILS, VISIT THE METHODS “INITIAL CIRCULAR MODELS” SECTION.	24
TABLE 5. RELATING DIMER TWIST ANGLES TO THE BP-STEP TWIST ENERGY, $\Psi T_{wopt}$ , FROM THE OPTIMIZED STATES USING THE TRIFONOV TWIST-ONLY FORCEFIELD.	32
TABLE 6 SUPERHELICAL PARAMETERS OF THE A <sub>4</sub> G <sub>4</sub> AND A <sub>5</sub> G <sub>5</sub> LINEAR MODELS OPTIMIZED WITH THE ZHURKIN ROLL-SPECIFIC FORCEFIELD USING THE MIYAZAWA POLYMER CALCULATIONS.	40
TABLE A.IV. 1 THE MEAN ELASTIC ENERGY/STEP OF THE 150-BP MINICIRCLES OPTIMIZED WITH THE ZHURKIN ROLL-SPECIFIC FORCEFIELD. THE TABLE INCLUDES THE TOTAL, TWIST, AND BEND ENERGY/STEP WITH EACH ASSOCIATED STANDARD DEVIATION.	63
TABLE A.IV. 2. SEQUENCE ALIGNMENT SCORES* OF THE A <sub>4</sub> G <sub>4</sub> AND A <sub>5</sub> G <sub>5</sub> LINEAR MODELS SEEN IN SUPPLEMENTAL FIGURE A.IV.7.	67
TABLE A.IV. 3. THE MEAN ELASTIC ENERGY/STEP OF THE 150-BP MINICIRCLES OPTIMIZED WITH THE COMBINED ZHURKIN/TRIFONOV FORCEFIELD. THE TABLE INCLUDES THE TOTAL, THE TWIST, AND THE BEND ENERGY/STEP WITH EACH ASSOCIATED STANDARD DEVIATION.	69

## LIST OF FIGURES

FIGURE 1. THREE RIGHT-HANDED FORMS OF DNA. _____	3
FIGURE 2. TOPOLOGY OF DNA CIRCLES. _____	6
FIGURE 3. THE 150-BP INITIAL CIRCULAR STATES OF DNA. _____	15
FIGURE 4. SEQUENCE COMPOSITION OF THE 150-BP DNA CIRCLES. _____	17
FIGURE 5. STRUCTURAL AND ENERGETIC DATA OF THE 150-BP CIRCLES OPTIMIZED WITH THE HOMOPOLYMERIC FORCEFIELDS. _____	25
FIGURE 6. OPTIMIZED ROLL AND TWIST PARAMETERS WITH RESPECT TO THE $\frac{1}{2}$ - $K_B T$ ENERGY BARRIER OF THE 150-BP MINICIRCLES. _____	27
FIGURE 7. THE OPTIMIZED ELASTIC ENERGY VALUES (UNITS OF $K_B T$ ) OF THE LK 14 (GREEN) AND LK 15 (YELLOW) 150-BP MINICIRCLES USING SEQUENCE-DEPENDENT FORCEFIELDS WHOSE ELASTIC FORCE CONSTANTS WERE IDENTICAL. _____	29
FIGURE 8 THE OPTIMIZED STATES OF THE 150-BP MINICIRCLES USING THE TWIST-SPECIFIC FORCEFIELD. _____	31
FIGURE 9. THE CENTERLINE CURVES OF THE OPTIMIZED 150-BP MINICIRCLES USING THE ZHURKIN MODEL FORCEFIELD RECONSTRUCTED FROM A 3-D PRINCIPAL COMPONENT ANALYSIS (PCA) OF THE BASE PAIR ORIGINS. _____	35
FIGURE 10. OPTIMIZED ELASTIC ENERGY PROFILES OF THE 150-BP MINICIRCLES WITH THE A <sub>4</sub> G <sub>4</sub> SEQUENCE OBTAINED USING THE ZHURKIN ROLL-SPECIFIC FORCEFIELD. ____	36
FIGURE 11. COMPARISON OF THE OPTIMIZED CONFIGURATION OF THE 150-BP LK 14 A <sub>5</sub> G <sub>5</sub> MINICIRCLE FROM THE ZHURKIN ROLL-SPECIFIC FORCEFIELD WITH BOTH BEND AND TWIST PARAMETERS AND ENERGY VALUES. _____	38
FIGURE 12. FRAGMENT ANALYSIS OF AN A <sub>4</sub> G <sub>4</sub> LINEAR MODEL OPTIMIZED WITH THE ZHURKIN ROLL-SPECIFIC FORCEFIELD. _____	39
FIGURE 13. THE OPTIMIZED BP-STEP ENERGIES FOR THE THREE SEQUENCE-SPECIFIC FORCEFIELDS SEPARATED BY BEND AND TWIST PARAMETERS. _____	41

FIGURE 14. RELATIONSHIP BETWEEN THE OPTIMIZED ROLL AND TWIST STEP PARAMETERS FROM THE THREE SEQUENCE-DEPENDENT FORCEFIELD OPTIMIZATIONS OF THE A <sub>5</sub> G <sub>5</sub> SEQUENCE WITH RESPECT TO THE ½-K <sub>B</sub> T ENERGY BARRIER.	43
FIGURE A.I. 1. THE DNA NUCLEOTIDE BASE.	48
FIGURE A.I. 2. DNA BASE EDGES.	49
FIGURE A.I. 3. THE MAJOR AND MINOR GROOVES OF DNA.	50
FIGURE A.II. 1. DNA MODELING AT THE BASE PAIR LEVEL.	52
FIGURE A.II. 2. THE BASE-PAIR STEP.	53
FIGURE A.II. 3. THE SIX BASE-PAIR STEP PARAMETERS.	54
FIGURE A.IV. 1. STRUCTURAL ANALYSIS OF THE 150-BP OPTIMIZED MINICIRCLES USING HOMOPOLYMERIC FORCEFIELDS.	59
FIGURE A.IV. 2. THE BP-STEP ROLL AND TILT PARAMETERS WITH RESPECT TO THE ½-K <sub>B</sub> T ENERGY BARRIER OF THE 150-BP MINICIRCLES.	60
FIGURE A.IV. 3. THE OPTIMIZED ELASTIC ENERGY PER BASE-PAIR STEP ALONG THE 150-BP MINICIRCLE SEQUENCE OBTAINED USING THE TRIFONOV TWIST-ONLY FORCEFIELD.	61
FIGURE A.IV. 4. THE OPTIMIZED ELASTIC ENERGY PER BASE-PAIR STEP ALONG THE 150-BP MINICIRCLE SEQUENCE OBTAINED USING THE ZHURKIN ROLL-SPECIFIC FORCEFIELD.	62
FIGURE A.IV. 5. RELATIONSHIP OF THE OPTIMIZED CONFIGURATIONS OF THE A <sub>4</sub> G <sub>4</sub> MINICIRCLES OF BOTH LK 14 (LEFT) AND LK 15 (RIGHT) WITH CHANGES IN DIMERIC BEND AND TWIST.	64
FIGURE A.IV. 6. LINEAR MODELS OF A <sub>4</sub> G <sub>4</sub> AND A <sub>5</sub> G <sub>5</sub> SEQUENCES OF LENGTH 150-BP OPTIMIZED WITH THE ZHURKIN ROLL-SPECIFIC FORCEFIELD.	65

FIGURE A.IV. 7. LINEAR MODELS OF THE  $A_4G_4$  (TOP) AND  $A_5G_5$  SEQUENCES (BOTTOM) OF LENGTH 150-BP OPTIMIZED WITH ALL THREE SEQUENCE-SPECIFIC FORCEFIELDS. \_\_ 66

FIGURE A.IV. 8. THE OPTIMIZED ELASTIC ENERGY PER BASE-PAIR STEP ALONG THE 150-BP MINICIRCLE SEQUENCE OBTAINED USING THE COMBINED ZHURKIN/TRIFONOV FORCEFIELD. \_\_\_\_\_ 68

## LIST OF DEFINITIONS

nt	nucleotide; monomer unit of a DNA base pair that contains a phosphate group, a five-carbon sugar, and a nitrogen-containing base
dA	adenine (DNA nitrogenous base); also written simply as A
dG	guanine (DNA nitrogenous base); also written simply as G
dT	thymine (DNA nitrogenous base); also written simply as T
dC	cytosine (DNA nitrogenous base); also written simply as C
R	purine; an aromatic, dual-ringed structure with four nitrogen atoms
Y	pyrimidine; an aromatic, single-ringed structure with two nitrogen atoms
bp	base pair; commonly used as a unit of DNA length
WC	“Watson-Crick” base pairing motif of A-T and G-C bases
$\hat{d}_i$	unit vectors of the standard DNA reference frame; either the short axis ( $i=1$ ), the long axis ( $i=2$ ), or the base-pair normal to the $\hat{d}_1\hat{d}_2$ plane ( $i=3$ )
$\theta_i$	rotational base-pair step parameters: tilt ( $\theta_1$ ), roll ( $\theta_2$ ), and twist ( $\theta_3$ )
$\rho_i$	translational base-pair step parameters: shift ( $\rho_1$ ), slide ( $\rho_2$ ), and rise ( $\rho_3$ )
$\vec{p}^n$	$6 \times 1$ parameter vector of the $n^{\text{th}}$ step; $\vec{p}^n = (\theta_1^n, \theta_2^n, \theta_3^n, \rho_1^n, \rho_2^n, \rho_3^n)$
$\vec{p}^0$	rest-state parameter vector of elastic energy 0; $\vec{p}^0 = (\theta_1^0, \theta_2^0, \theta_3^0, \rho_1^0, \rho_2^0, \rho_3^0)$
$\hat{e}_i$	principal component analysis basis vectors
Lk	Linking number; topologically-invariant integer value representing the times a curve, C, intersects a surface, S, for which the curve is bounded
Wr	Writhe; deformation of the DNA centerline curve
$k_B T$	Boltzmann energetic factor; $1 k_B T = 4.11 \times 10^{-21} \text{ J} = 4.114 \text{ pN} \cdot \text{nm}$ (at 298K)
$\text{\AA}$	Angstrom; $1 \text{\AA} = 0.10 \text{ nm} = 10^{-10} \text{ m}$

## INTRODUCTION

The scientific community has experienced sweeping successes over the last several decades in the mapping and manipulating of genomes in viruses, archaea, and more complex multicellular organisms, through the prolific Human Genome Project and CRISPR gene editing technologies<sup>1,2</sup>. The intricacies of the genetic code—from nucleotide base composition, sequence mutability, and replication from parent to daughter cells—tells only part of the story of DNA. The other critical component lies in the myriad arrangements of the polymer that arise from its unique biophysical characteristics. Breakthrough findings in structural biophysics from a set of *Nature* papers in 1953<sup>†</sup> suggested that the DNA polymer is arranged in a double-helical configuration with the chain of paired nucleotides situated between anti-parallel sugar-phosphate backbones. This physical composition makes it difficult to access the nucleotide code without external mechanisms<sup>3–6</sup>. Using genetic material through transcription<sup>7–9</sup>, replication<sup>10</sup>, or recombination<sup>11,12</sup> relies on protein binding interactions along small tracts of the macromolecule, which introduces various degrees of deformation along the chain. This deformability is essential in the organization of DNA into densely packed regions, mediated with and without protein interactions<sup>7</sup>.

The ease at which DNA can deform depends on a plethora of factors both internal (e.g. the number of nucleotides in the polymer; the interactions between bases that connect

---

<sup>†</sup> The publications in *Nature* 1953, Volume 171, pages 737-741, are in the following order: Watson, J.D. and Crick, F. H. “A structure for deoxyribose nucleic acid”; Wilkins, M.H.F, Stokes, A.R., and Wilson, H.R. “Molecular structure of deoxypentose nucleic acids”; Franklin, R.E., and Gosling, R.G. “Molecular configuration in sodium thymonucleate”.

the backbone chains<sup>‡</sup>) and external (the surrounding solution conditions). The environment surrounding DNA is dynamic, varying in localized solute and ionic concentrations which can perturb the polymer into different forms<sup>13</sup>. While there are a number of forms that the polymer can take, we focus here on three closely related helical structures, the A- B-, and C-DNA forms seen in Figure 1.A<sup>§</sup>, with a right-handed base pair rotation along the 5'-3' direction used in reading the nucleotide sequence<sup>14-17</sup>. These forms differ primarily in base stacking and twisting of successive base-pairs. B-DNA, the dominant form under physiological conditions, has a regular twist of 10.5 base pairs (bp) per helical turn in solution and a regular stacking where the helical axis closely coincides with the path of the base-pair centers. This contrasts with underwound A-DNA, a form that contains an 11-bp/turn twist, a high inclination angle of base pairs with respect to the helical axis, and a noticeable displacement between the path of the base-pair centers and the central helical axis (Figure 1.A, left)<sup>18-20</sup>. The overwound C-DNA form (Figure 1.A, right) contains a 9.4-bp/turn twist, a base-pair inclination of orientation opposite to that of A-DNA, and a smaller non-zero displacement of base pairs from the helical central axis<sup>16,21</sup>.

These relationships at the base pair level influence the overall configuration of DNA (Figure 1.B) and the groove pattern along the polymer. These grooves are channels that occur as the result of the stacking and twisting of repetitive asymmetric base pairs: one of the long sides of the base pair contains both sugar rings that connects a base to the phosphate backbone, resulting in a closer proximity of the backbones and, thus, the minor groove of DNA (Supplemental Figure A.1.3). The major groove, then, are channels made

---

<sup>‡</sup> For a primer on the chemical structure and physical properties of the DNA base pair, see Appendix A.I-DNA: Building from the Base Up

<sup>§</sup> All 3-D models throughout this text were made using the PyMOL molecular modeling software<sup>102</sup>.

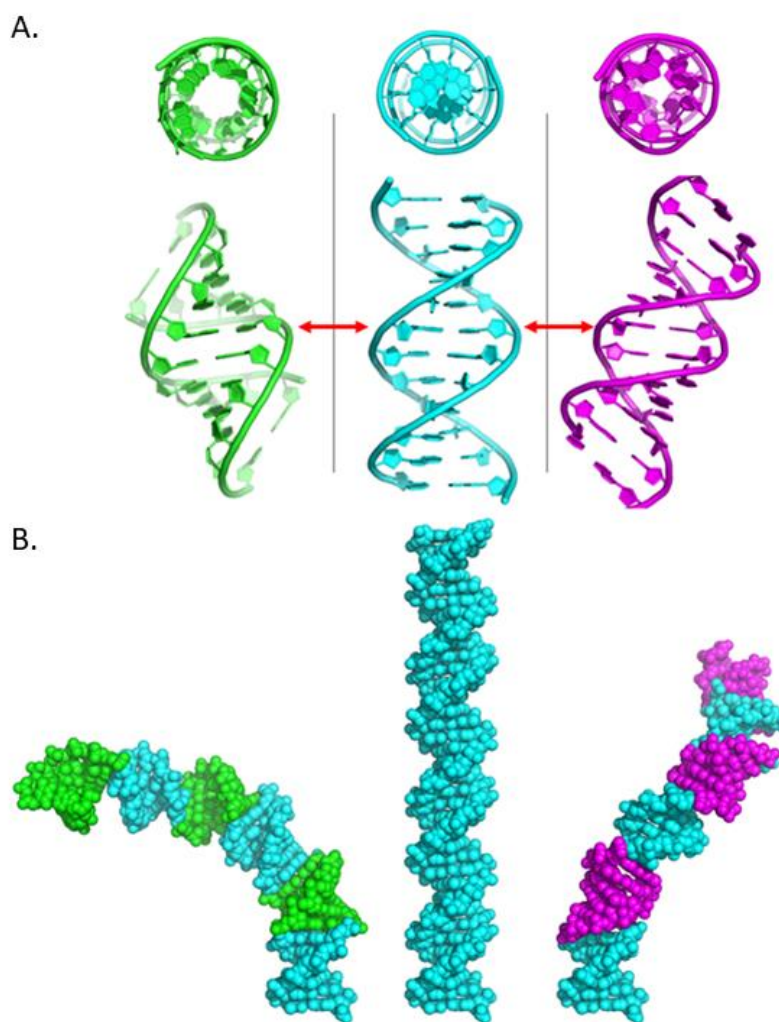


Figure 1. Three right-handed forms of DNA. (A) Two views of 12-bp molecular models of A-DNA (left, green), B-DNA (center, cyan), and C-DNA (right, magenta). The top images are views along the helical axis to observe differences in base-pair stacking. Below is a profile view positioned such that the 7th bp of each model is similarly oriented in space (red arrows). This is done to highlight locations and differences among each major groove (top left of every profile above the red arrow) and minor groove (top right above arrow). (B) A set of 36-bp space-filled DNA models where every 6-bp tract is of a different form. Each model is orientated such that the bottom B-DNA tracts (cyan) are of identical spatial arrangements. The models to the left and right of the linear all-B-DNA models incorporate A-DNA tracts ("B-A" models, cyan-green) and C-DNA tracts ("B-C" models, cyan-magenta), respectively, resulting in curved strands of different paths. Of the three strands, the B-A model has the shortest end-displacement (82.6 Å) and greatest angle between the first and last bp normal vectors (76°) compared to the all-B (118.1 Å, 5.5°) and B-C (100.7 Å, 18.6°) models.

from the more open sides of the base pairs. This may suggest the major grooves are more exposed to the solvent environment as opposed to the more sterically restricted side, but there exist different forms of DNA where the minor grooves are more exposed. Sequence studies used to understand DNA forms show that while any sequence can form B-DNA,



tracts of high adenine (dA) content favor B-DNA while tracts with high guanine | cytosine (dG | dC) content favor A-DNA<sup>22,23</sup>. However, environmental effects may supersede sequence such as the case with low humidity conditions inducing the A form or the presence of lithium salts, forcing the polymer to adopt the C form<sup>21,22</sup>. With the numerous differences in sequence and local external environment within a cell, a long chain of DNA can be made up of successive tracts with different helical forms, culminating in a configuration with some degree of bend (Figure 1.B). Given the correct conditions, DNA can take forms that appear coiled or circular with relative ease.

### *Beyond the Chromosome: Circular Forms of DNA*

The ease at which the DNA polymer can deform due to bend and twist variability appears biologically critical. Methods used to inhibit the transcription of genomic material have been linked to unfavorable changes in the twisting of DNA base pairs, thus imposing some degree of torsional stress which can disrupt the binding of protein machinery.<sup>24,25</sup> Genomic organization within cellular nuclei and viruses relies on the bendability of the polymer<sup>7,26</sup>. The importance of both bend and twist deformations can be seen in studies of circular forms of DNA— closed fragments of DNA whose ends are covalently bonded together. Circular forms of naked DNA exist biologically and are more prevalent than previously thought. Cytoplasmic or extrachromosomal DNA (eccDNA) circles commonly range from 200-400 bp in length, but can get as large as 20,000-bp, and often contain repetitive coding and non-coding sequences that can be found in the cellular genome<sup>27-30</sup>. Circular DNA is essential to the genomic organization in the mitochondria organelle found in the cells of most organisms as it consists of a network of interlocking circles either of

various sizes such as those found in certain parasites or in more evenly distributed sizes in some mammals<sup>31–33</sup>. Circular DNAs are a common tool in molecular biology experimentation (commonly referred to as plasmids) and have been used as an improved gene therapeutic due to the ease at which they can be produced and stored<sup>34–36</sup>.

Use of the moniker “circular DNA” may give the impression that these forms must be uniformly curved and flat in some 2-D plane, but this organization contains a multitude of topology states. A single length and composition of DNA may have several topoisomers, varying in three-dimensional complexity, superhelicity, and compaction. Topoisomers that vary by degree of superhelicity, or supercoiling, are denoted by their linking numbers (Lk), a topologically invariant integer value that depends on the number of times the two DNA backbones wind about each other, thus appearing as chain links between strand crossings if the paths are projected onto a 2D plane (Figure 2.A)<sup>14,35,37–39</sup>. As they are invariant, topoisomers can only change their linking numbers by external means. This can be done using a topoisomerase enzyme which nicks either one or both of the backbone chains, allowing the double helix to relax by unwinding if some internal stress resulted in a supercoiled configuration<sup>35,37,40</sup>. The value associated with the linking number depends on how the two strands appear to intertwine, where a right-handed crossing denotes a positive crossing and the sum of those crossings results in an integer value<sup>38,39</sup>. Mathematical work has shown that Lk can be partitioned into two global properties of the polymer: twist (Tw) and writhe (Wr)<sup>39,41–44</sup>. The twist or twist density is how the curve made by one of the backbone chain twists about the helical (centerline) axis and is related to how many base pairs are in a helical turn. The writhe of DNA is associated with how the path of the centerline curve deviates from an ideally planar path and is related to the average number

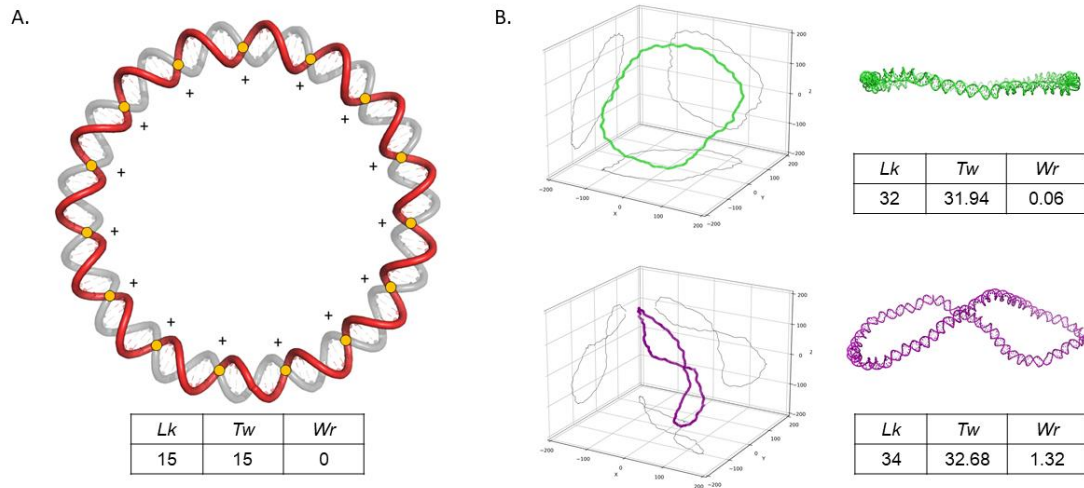


Figure 2. Topology of DNA circles. (A) Visualization of the linking number of a planar circle with a twist of 10.0-bp/turn. The red and light grey ribbons are the backbone chain curves of a 150-bp DNA model. Every yellow circle denotes positive backbone crossings of the DNA backbones of right-handed directionality after one full turn along the double-helix. The sum of the positive crossings results in the linking number of this structure, 15. As this is a planar circle with a uniform twist, writhe ( $Wr$ ) is 0 and, thus,  $Lk$  is made up exclusively of twist ( $Tw$ ). (B) Differences in global configuration between two topoisomers of length 336 bp from molecular dynamics simulations. At the top (green) is the topoisomer for an open state,  $Lk$  32, while the bottom (purple) is that of the supercoiled  $Lk$  34 state. The left section shows the 3D path of the base pair origins with 2D planar projections in grey. (right) The molecular models are in a profile view to show the degree of writhe of each structure. The  $Lk$  34 supercoiled state has a  $Wr$  value associated with the average number of superhelical crossings found on the three 2-D plane projections.

of superhelical crossings made by the polymer chain when viewed in all possible directions. For more planar configurations of DNA, the superhelical writhe is close to zero and, thus, the  $Lk$  closely resembles  $Tw$  (Figure 2.B, top). However, if a structure has 2-D projections that show superhelical crossings (Figure 2.B, bottom), it will have a non-zero writhe, meaning that the overall structure must be over- or under-twisted if  $Lk$  is to be preserved. Additionally, DNA can have a relaxed linking number,  $Lk_0$ , derived from a relationship between sequence length and the number of base pairs per helical turn,  $N/\gamma$ , and rarely an integer value. The linking number difference,  $\Delta Lk$ , between the two values is often used in descriptions of circular superhelicity, where a difference close to zero often represents some open conformation<sup>45,46</sup>. The naturally occurring, closed-form DNA tends to be unwound, having more base-pairs per turn relative to B-DNA with compensatory

effects along the helical axis that generate a highly writhed, negatively supercoiled state<sup>38,47</sup>.

The linking number, and thus the overall circular structure, is influenced by the deformability of both the duplex twist and bending of the centerline axis<sup>48</sup>. This interplay is described in the twistable worm-like chain (TWLC) model<sup>49</sup> of polymers, which describes how physical stresses can elicit some elastic response with respect to bend and twist stiffness<sup>50</sup>. These stiffness values are directly related to the persistence length, the length limit that dictates the rigidity/deformability of a polymer in solution and the energy costs required to do so on the Boltzmann energy scale (units of  $k_B T$ )<sup>7,14</sup>. Once the length of the polymer extends beyond this limit, the global structure will adopt a configuration that is more easily influenced by entropically-driven thermal fluctuations<sup>51-54</sup>. Depending on solution conditions, the bending persistence length is often cited as ~50nm while values of torsional persistence range from 25-100nm<sup>45,55</sup>. As the vertical displacement between bases in B-DNA is 0.34nm, the lengths can be re-expressed as ~150-bp for the bending persistence length and 70-300bp for the torsional length. Both the TWLC and the subsequent Marko-Siggia (MS) model reduce double-stranded DNA to a twistable, continuous polymer rod with no directional preferences in chain bending, known as the elastic rod model<sup>56-59</sup>. Use of this model means that any subsequent deformational energy will be independent of direction so long as DNA is treated as a thin, inextensible homogenous rod; however, work with the TWLC and MS models can allow for coupling between bend and twist<sup>58,60-62</sup>. By reducing DNA to such a simplified model, relationships between sequence length and deformability are reduced to a first-order approximation where the equilibrium state depends on the configuration of the helical axis<sup>57</sup>.

### *Use of in Silico Methods to Observe Sequence-Dependent DNA Behavior*

Exploring the physics of macromolecules has expanded with advancements in experimental methods used to determine structural details (e.g. gel electrophoresis, cryo-electron microscopy, X-ray scattering) and mechanistic properties (e.g. magnetic tweezers). These improvements, in turn, have led to advancements in *in silico* simulation methods. Experimental work with circular DNA has led to ring-closure, or cyclization, studies using Monte Carlo calculations to determine the probability of circular DNA formation in terms of the Jacobson-Stockmayer, or J, factor relating the rate constants of cyclization and bimolecular association of the linear ends<sup>63,64</sup>. Evidence has shown that cyclization depends on DNA length typically beyond the persistence length, yet additional results have shown that cyclization of fragments around length 100-bp occur more efficiently than expected by the TWLC model<sup>49,65</sup>. All-atom molecular dynamics (MD) simulations have been used to resolve discrepancies between single-molecule pulling experiments with various supercoiled forms of circular and looped DNA<sup>54,59,66</sup>. While these models and simulations are useful in understanding behavior on both coarse-grained and atomic scales, there are obstacles such as the loss of information when producing models that may be too coarse-grained or the computational expense involved in an MD simulation that must account for numerous atomic positions and interactions. A balance in studying structural differences without loss of information has led to the development and use of models that describe DNA at the base-pair step level. This framework allows for new insights into DNA bend and twist behavior, especially how local deformations influence global configurations of circular DNA<sup>67</sup>.

The nomenclature and calculation schematic for base-pair steps is based on the positional and angular relationships between successive base pairs<sup>68,69</sup>. Use of six rigid-body parameters from Euler-angle calculations—the rotational tilt, roll, and twist and translational shift, slide, and rise parameters<sup>\*\*</sup> — can effectively generate 3-D models, as seen in Figure 1.A, where the three forms of DNA have been constructed using the parameter values in Table 1. Using base-pair step parameters to model DNA can be especially helpful in understanding the role nucleotide sequence plays in DNA configurations as structural deformations and equilibrium states have been shown to depend on nucleotide sequence. The bp-step is commonly defined as a dimer step of two successive base pairs with  $4^2$  of 16 difference sequence combinations. Longer chain fragments, such as trimer (3 bp) and tetramer (4 bp) steps can also be used in modeling and are frequently treated in terms of the rigid-body parameters describing the constituent dimer steps<sup>70,71</sup>. Experimental X-ray crystallography studies of a dodecamer sequence spanning a full right-handed turn of B-DNA provided the first clear indication of non-uniform base-pair step parameters along the path of DNA<sup>72</sup>. In addition, gel electrophoresis studies of DNA fragments with different adenine repeat (“A-tracts”) lengths showed that DNA could be naturally curved, particularly in the fragments with A-tract length ranging from four to six base pairs in length<sup>73,74</sup>. These differences in global compaction are related to difference in localized bending between the A-tract regions and neighboring DNA, with the bp-step bend related to the tilt and roll angles. Different theoretical models have been developed to account for this behavior, including a nearest-neighbor wedge model based on sequence differences and a long-ranged junction bending model associated with the

---

<sup>\*\*</sup> The explanations, calculations, and useful figures of the bp-step parameters are found in Appendix A.II. The Base-Pair Step

Table 1. The base-pair step parameters associated with the right-handed A-, B-, and C-DNA helices. These parameter values are from canonical fiber models and can be readily found at the Web 3DNA server ([web.x3dna.org](http://web.x3dna.org))<sup>16,17</sup>.

	<b>Tilt (<math>\theta_1</math>)</b>	<b>Roll (<math>\theta_2</math>)</b>	<b>Twist (<math>\theta_3</math>)</b>	<b>Shift (<math>\rho_1</math>)</b>	<b>Slide (<math>\rho_2</math>)</b>	<b>Rise (<math>\rho_3</math>)</b>
<b>A-DNA</b>	0.0°	12.4°	30.3°	0.00 Å	-1.39 Å	3.30 Å
<b>B-DNA</b>	0.0°	1.7°	36.0°	0.00 Å	0.45 Å	3.36 Å
<b>C-DNA</b>	0.0°	-11.5°	38.4°	0.00 Å	1.75 Å	3.96 Å

boundaries between A-tracts and neighboring B-DNA regions (best seen in Figure 1.B, left)<sup>75–79</sup>.

Sequence-dependent studies of DNA have clarified structural details and revealed a series of unique structural steps. The 16 possible dimer step sequences reduce to ten unique sequences due to self-complementary that arises from the antiparallel nature of DNA. Closer inspection of the dimer steps shows that AA, AG, GA, and GG steps have similar parametric values as TT, TC, CT, and CC steps, respectively. This complementary relation is also seen between the CA and TG steps and the AC and GT step, thus accounting for the lower number. The dimer steps can also be separated into three standard chemical classes based on coding strand ring-types: pyrimidine-purine (YR), purine-purine (RR), and purine-pyrimidine (RY) steps, with the pyrimidine-pyrimidine (YY) behaving similarly to RR steps. Compiled structural data surrounding YR bp-steps indicates that these steps are the easiest to deform owing to the weak interactions associated with lesser base stacking<sup>80</sup>. Analysis of the roll-twist values in these three dimer groups has shown an interesting split within RR steps. The dispersion of the AG and GG step twist values is greater than that of the AA and GA steps, which have a greater dispersion in roll values and, thus, bend values as the dispersions across tilt are relatively similar<sup>80</sup>. This difference in bending and twisting supports observations from solution studies of A-tract bending, particularly in how under-twisted GG steps favor bending into the major groove, resulting

in a (-) roll angle, while over-twisted AA steps bend into the minor groove and a (+) roll<sup>68,81,82</sup>. As more structures of either naked or protein-bound DNA are solved at improved resolutions with improved biophysical technologies, continual use of databases, such as the Protein Data Bank (<https://rcsb.org><sup>83,84</sup>) and the Nucleic Acid Database (<http://ndbserver.rutgers.edu/><sup>85</sup>), will allow for improved modeling of the ten dimer steps and, within each step, the relationships among the rigid-body parameters<sup>80,81</sup>.

The effectiveness of base-pair step parameters extends beyond constructing 3-D models DNA as these parameters can be used in determining the elastic energy of a collection of bases with respect to known rest-states and elastic stiffness constants. For idealized models meant to represent B-form DNA, the elastic stiffness constants are used in determining persistence lengths from observational data, while models with increased sequence-dependence uses the covariance between the rigid-body parameters from known structures. A 2003 computational study of the effect of base-pair sequence on the equilibrium configurations of DNA described the total energy of a 150-bp kinked polygonal form of circular DNA as the sum of elastic energies at each step<sup>86</sup>. These kinked polygons were constructed from repeats of alternating 5-bp tracts of naturally straight steps with 5-bp tracts of locally bent steps. However, this study lacked a way to differentiate the equilibrium states that reflect different sequences. Inspired by the A-tract rigid bending observed in gel electrophoresis studies, the study conducted here investigates the role that nucleotide sequence plays in elastically-optimized configurations of 150-bp circular topoisomers with sequences of varied A and G tract lengths. This study utilizes a range of rest-state models that increase in complexity from homopolymeric models of linear B-DNA to sequence-dependent rest-state models with dimer-specific differences in bend and



twist parameters. This study uses the same set of elastic force constants for each step thereby allowing all differences in optimized states to be the result of the unique rest state of each dimer step<sup>46,87</sup>.

## METHODOLOGY & CALCULATIONS

As advancements continue in biophysical experimentation and structural characterization, so do *in silico* research methods, particularly in optimization and dynamical studies of DNA configurations. *In silico* methods require a balance between expense and detail: a 100-bp chain of DNA contains nearly three thousand atomic coordinates, each used within a fully detailed simulation to describe numerous angular relationships, bonding arrangements, and non-bonded interatomic distances. While an all-atom approach in studying a configuration of DNA may provide highly detailed structural information, such a study can be computationally expensive to perform and manage. A coarse-grained treatment may produce a less expensive model, but the quality of the model is tied to the size with which one approximates the repeating unit of the DNA chain. In the past, coarse-grained models have approximated a number of successive bases by a single sphere, resulting in models similar to beads on a string<sup>63,88,89</sup>. While models were readily generated, the large size of the base leads to a model of less detail, making it difficult to understand the possible deformations along the chain. The base-pair-level approach used here allows one to model the ~3000 atomic coordinates in a 100-bp chain as a series of 99 base-pair steps, which can also be used in the minimization of the elastic energy.

### *Initial Circular Models*

Three initial circular states of length  $N=150$ -bp were used for this study, with different bp-step parameters and coding sequences of dA and dG residues. Each starting state was constructed from a collection of  $N$  bp-step vectors,  $\vec{P}$ , made up of the six

translational and rotational bp-step parameters<sup>††</sup>. The 6x1 vector at the  $n^{\text{th}}$  bp-step has the form:

$$\vec{p}^n = (\theta_1^n, \theta_2^n, \theta_3^n, \rho_1^n, \rho_2^n, \rho_3^n) \quad [1]$$

These parameter vectors can be further converted into a collection of  $N+1$  base pair positions and reference frames using either the Web 3DNA server (<http://web.x3dna.org/>) or the 3DNA command line software package<sup>17,90</sup>. To accurately model circular DNA, an additional condition must be applied so that the first and last base pairs overlap at the final step connecting the ends. This is done by adding a virtual base pair that has the same base pair position  $\vec{o}^{N+1}$  and orthonormal reference frame  $\mathbf{B}^{N+1}$  as the first base pair<sup>‡‡</sup>:

$$\vec{o}^1 = \vec{o}^{N+1}, \mathbf{B}^1 = \mathbf{B}^{N+1}. \quad [2]$$

Two of the initial circles are analytically-constructed planar circles using intrinsic twist angles associated with either 10.5- or 10.0-bp/turn ( $34.3^\circ$  and  $36.0^\circ$ , respectively). These two twist angles produce relaxed linking numbers,  $Lk_o$ , of 14.3 and 15.0, thus naming these initial circles as P-14 and P-15 (Figure 3 in green and yellow, respectively). The initial rotational parameters at the  $n^{\text{th}}$  step can be calculated from the formulas:

$$[\theta_1]^n = \Gamma \sin(n\theta_3) \quad [3]$$

$$[\theta_2]^n = \Gamma \cos(n\theta_3) \quad [4]$$

$$[\theta_3]^n = 360^\circ \left( \frac{Lk}{N} \right). \quad [5]$$

---

<sup>††</sup> Explanations and calculations of the bp-step parameters are found in Appendix A.II. The Base-Pair Step

<sup>‡‡</sup> A note on mathematical representation:

A vector will be italicized and represented with an arrow above variable (as  $\vec{v}$ ); a unit vector will be italicized and represented with a hat above variable (as  $\hat{v}$ ); a matrix will be bold-faced without symbols (as  $\mathbf{M}$ ) while an element within a matrix will be denoted using square brackets and subscript index notation of the  $i^{\text{th}}$  row,  $j^{\text{th}}$  column order (as  $[\mathbf{M}]_{ij}$ ); for basis representation, curly braces will be used to contain basis unit vectors, such as the commonly used Cartesian coordinate system  $\{\hat{x}, \hat{y}, \hat{z}\}$ .

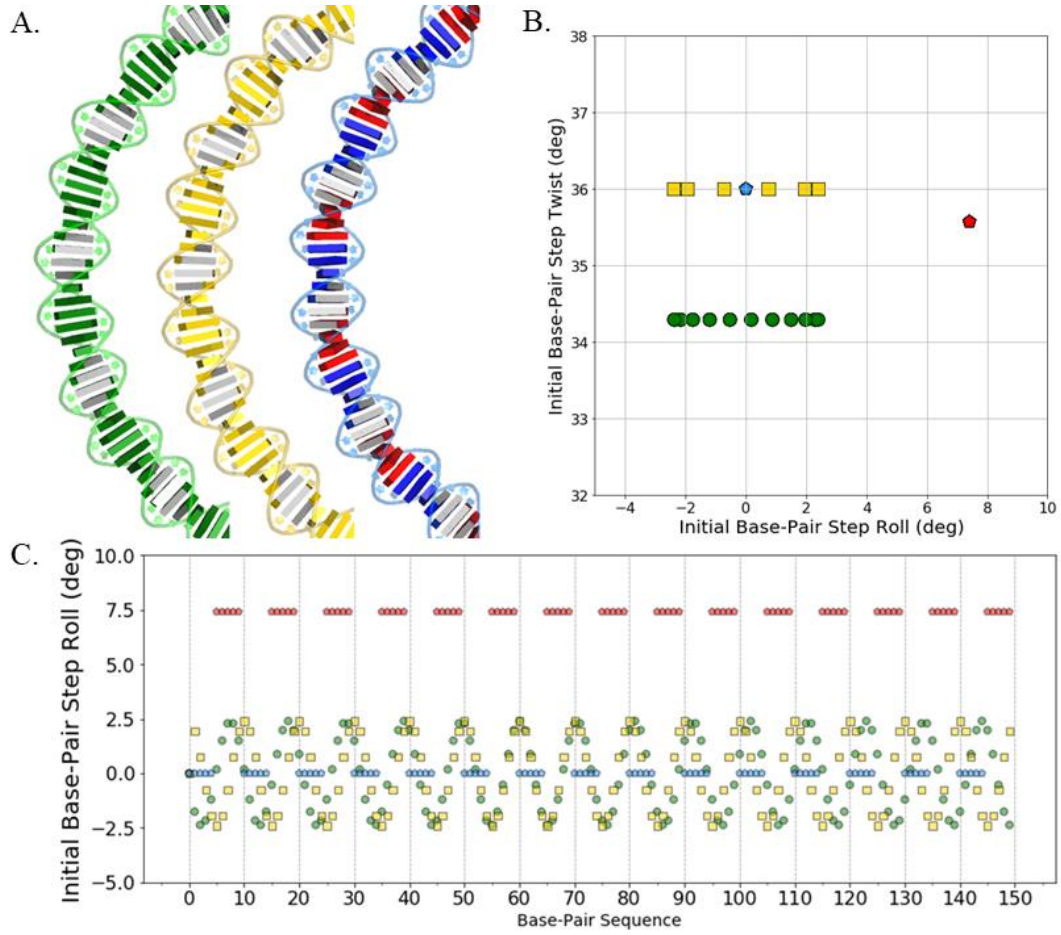


Figure 3. The 150-bp initial circular states of DNA. (A) Curved segments from the molecular models of the three initial configuration states. The centerline paths of the planar circles of Lk 14 (P-14, green) and Lk 15 (P-15, gold) may be identical but differences in Lk means the orientation of the minor grooves (silver faces) will be different due to dimer twist. This difference in minor groove locations is also seen in the o-ring construct of Lk 15 (R-15) with its characteristic 10-bp repeating motif: a 5-bp tract of naturally straight DNA (blue) and a 5-bp tract of bent DNA (red). (B) The relationship between the initial roll and twist parameters of the P-14 (green circles), P-15 (gold squares), and R-15 (blue and red hexagons) initial states. The values for the P-14 and P-15 initial states are found in Equations 4 and 5, respectively, while the R-15 states are in Equation 7. (C) A closer look at the dimer roll along the 150-bp circle (color and symbols match those in (B)).

The use of the sine and cosine functions generates a smooth planar circle of zero writhe and constant radius, as seen in Figure 3. The tilt and roll angles equations described in [3] and [4] ( $\theta_1$  and  $\theta_2$ , respectively) are related to the bending angle  $\Gamma$  at each of a bp-step given by the quotient of  $360^\circ$  and the circular length in bp,

$$\Gamma = \frac{360^\circ}{N} = \sqrt{\theta_1^2 + \theta_2^2}; \quad [6]$$

thus, for  $N=150$ -bp,  $\Gamma=2.4^\circ$ . The initial twist,  $\theta_3$ , depends on the length of the circle and the initial linking number of the structure. It is important to note that the relaxed linking number can be a non-integer value. If the 150-bp circle is constructed using the integer linking number of 14, the initial twist is  $33.6^\circ$ , or 10.7-bp/turn. If constructed with Lk 15, the initial twist is  $36^\circ$ , or 10 bp/turn. The third topoisomer is a kinked polygon referred to as an “o-ring” of Lk $^\circ$  15 (R-15, Figure 3 in alternating red and blue). This o-ring is constructed by repeating a 10-bp structural motif whose first five bp-steps are naturally straight followed by five steps that are locally bent while maintaining the same isotropic translational parameters<sup>86</sup>:

$$\begin{aligned}\overrightarrow{[p]}^{1-5} &= (0^\circ, 0^\circ, 36.00^\circ, 0 \text{ \AA}, 0 \text{ \AA}, 3.4 \text{ \AA}), \\ \overrightarrow{[p]}^{6-10} &= (0^\circ, 7.413^\circ, 35.568^\circ, 0 \text{ \AA}, 0 \text{ \AA}, 3.4 \text{ \AA})\end{aligned}\quad [7]$$

This parametric repeat, with its jumps in bend and differences in twisting, is somewhat analogous to having regions of B-DNA next to regions of A-DNA. This linear/bent repetition means the o-ring is essentially constructed with two roll values and no tilt as opposed to the isotropic bending in the analytical constructs. For all initial circles, the translational parameters used were those of ideal B-DNA and are the same throughout the initial states:

$$[\rho_1, \rho_2, \rho_3]^n = [0 \text{ \AA}, 0 \text{ \AA}, 3.4 \text{ \AA}]. \quad [8]$$

To study the influence of nucleotide sequence, each initial circle is set with a series of specific nucleotide sequences along the coding strand as seen in Figure 4. Previous works surrounding the mechanism behind A-tract bending used various poly(dA)-poly(dT) tracts that varied in tract lengths and compositions<sup>77,91,92</sup>. Here, the sequences are built from an  $A_XG_X$  sequence repeat, where X represents the tract length of either base in the DNA

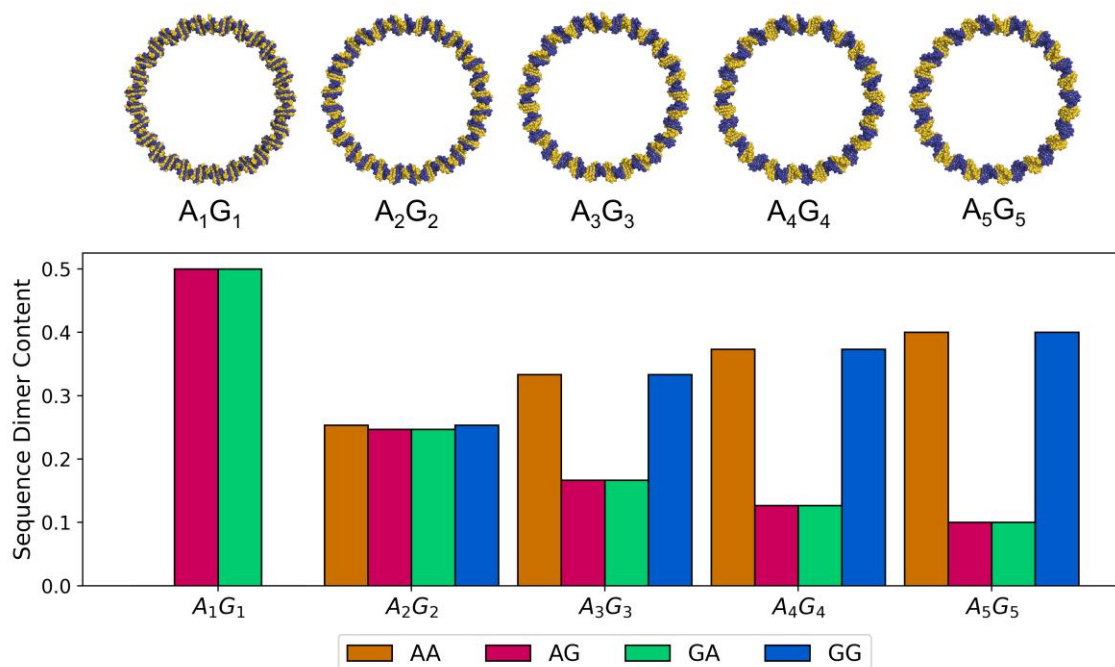


Figure 4. Sequence composition of the 150-bp DNA circles. (Top) Molecular models of a 150-bp minicircle with the five  $A_XG_X$  sequences, with X from 1 to 5, dA bases in gold, and dG bases in blue. For each circle, the number of dA and dG bases are equal (75 bp each) and are evenly dispersed. (Bottom) A column graph showing the sequence dimer content within each sequence motif. An increase in the X-value in the sequence motif results in increases of the homodimers AA and GG and a decrease in the AG and GA heterodimers.

sequence (from 1 to 5), to narrow the focus to a single nucleotide base type (the dual-ringed purine). Use of this sequence repeat maintains the same number of dA and dG bases used in each circle but changes the dimer content: as X increases, the number of homodimeric steps (AA, GG) will increase (Figure 4). In the cases of the X=2 and X=4, a starting cap of GGG and an ending cap of AAA has been used to ensure a 150-bp sequence.

### *Elastic Energetic Optimization with emDNA*

Determination of optimized states using the three sets of initial circular structures was conducted using the emDNA software package<sup>46</sup> conducted either locally or using the Office of Advanced Research Computing (OARC) Amarel computing cluster at Rutgers, the State University of New Jersey (<https://oarc.rutgers.edu>). The emDNA toolkit utilizes elastic energy minimization calculations in bp-step parameter space to build molecular

models of DNA. The optimization calculations center around the quadratic expression for the elastic energy at the  $n^{\text{th}}$  base-pair step,  $\psi^n$ , such that:

$$\psi^n = \frac{1}{2} (\vec{p}^n - \vec{p}^o)^T (\mathbf{E}_{\text{STEP}})^n (\vec{p}^n - \vec{p}^o) \quad [9]$$

In the quadratic expression, the  $6 \times 1$  vector  $\vec{p}^n$  is taken from the collection of base-pair steps for a particular starting configuration state  $\vec{P}$ ,  $\vec{p}^o$  is the parameter vector of a zero-energy intrinsic state that may depend on bp-step sequence, and  $\mathbf{E}_{\text{STEP}}$  is a  $6 \times 6$  elastic stiffness force constant matrix that may also depend on step sequence. The overall optimization calculation utilizes a gradient descent method with respect to step parameters,  $\frac{\partial \psi^i}{\partial p^i}$  §§. For this study, the bp-step size is a dimeric step (denoted with an XY sequence) but can be potentially expanded to a tetrameric step.

The dimer-specific rest-states and elastic constants used in Equation 9 can be sourced from a wide range of options that include experimental data from solved, three-dimensional biophysical structures or from predictive modeling using atomic-level molecular dynamics simulations. Surveys of high-resolution crystal structures deposited in the Protein Data Bank have provided critical insights into relationships between nucleotide sequence and base-pair step parameters, and as the quality of biophysical methods increase so too will the quality of dimer and tetramer step parameters<sup>80</sup>. Structural studies have already shown that the polymer tends to bend more easily into the grooves than along the phosphate chains, hence changes in bend are more commonly associated with roll values than tilt<sup>93,94</sup>. There is also, on average, a slightly greater variation in bend than twist. Similarly, changes in translational slide along the long axis of the reference frame are easier

---

§§ A detailed explanation of the optimization calculations for emDNA is found in Appendix A.III. Elastic Energy Minimization Calculations

than changes in shift along the short axis. In comparison to these shear components, changes to the vertical rise are much smaller due to base-stacking interactions.

Although a plethora of sequence-specific structural data continue to be solved, there is enough evidence that supports the use of a simple physical model to best understand the elastic behavior of a chain of DNA. As there is a natural propensity for unbound DNA to be in the B-form, an isotropic, inextensible rod model of DNA mimicking the twistable worm-like chain model that is independent of sequence is employed here, referred to as the IdealDNA model (Table 2). Additional simplistic models which do not account for sequence composition include a set of uncultured B-DNA parameters from a 1998 survey of the Protein Data Bank that represented a mixed assortment of sequences in ligand-free crystals (referred to here as the BDNA98 model)<sup>80</sup>. This set of parameters has a non-zero intrinsic roll and slide, parameters defined relative to the long  $\hat{d}_2$  axis, as well as a reduced rise, or vertical displacement. However, as the dispersion for each of these parameters is high, BDNA98 model is considered low-quality based on early limited data.

These simple homopolymeric models provide insights into deformation, but models can increase in sophistication by taking consideration of either the intrinsic parameters or elastic stiffness parameters based on the step sequence. For this study, the changes in models associated with step sequences are done for dimer steps and are only applied to the intrinsic step parameters (Table 3). Naturally curved regions of DNA from A-tracts have been found to be the result of cooperativity between roll and twist<sup>57,82</sup>. Analysis of curved



Table 2. The intrinsic base-pair step parameters of the two homopolymeric forcefields IdealDNA and BDNA98.

	<b>Tilt (<math>\theta_1</math>)</b>	<b>Roll (<math>\theta_2</math>)</b>	<b>Twist (<math>\theta_3</math>)</b>	<b>Shift (<math>\rho_1</math>)</b>	<b>Slide (<math>\rho_2</math>)</b>	<b>Rise (<math>\rho_3</math>)</b>
<b>IdealDNA</b>	0°	0°	34.3°	0Å	0Å	3.4Å
<b>BDNA98</b>	0°	1.4°	35.4°	0Å	0.35Å	3.32Å

Table 3. The intrinsic base-pair step parameters used in the Trifonov, Zhurkin, and the combined Zhurkin/Trifonov sequence-dependent forcefields.

	<b>Sequence Dimer</b>	<b>Tilt (<math>\theta_1</math>)</b>	<b>Roll (<math>\theta_2</math>)</b>	<b>Twist (<math>\theta_3</math>)</b>	<b>Shift (<math>\rho_1</math>)</b>	<b>Slide (<math>\rho_2</math>)</b>	<b>Rise (<math>\rho_3</math>)</b>
<b>Trifonov</b>	AA	0°	0°	35.6°	0Å	0Å	3.4Å
	AG	0°	0°	27.7°	0Å	0Å	3.4Å
	GA	0°	0°	36.9°	0Å	0Å	3.4Å
	GG	0°	0°	33.7°	0Å	0Å	3.4Å
<b>Zhurkin</b>	AA	0°	-2.0°	34.3°	0Å	0Å	3.4Å
	AG GA GG	0°	3.0°	34.3°	0Å	0Å	3.4Å
<b>Zhurkin / Trifonov (Combined)</b>	AA	0°	-2.0°	35.6°	0Å	0Å	3.4Å
	AG	0°	3.0°	27.7°	0Å	0Å	3.4Å
	GA	0°	3.0°	36.9°	0Å	0Å	3.4Å
	GG	0°	3.0°	33.7°	0Å	0Å	3.4Å

DNA showed differences in roll at different dimers, particularly a decrease of  $\sim 5^\circ$  in the roll of the AA dimer compared to the three other RR steps (AG, GA, GG) (Table 3). These RR-step roll values are combined here with the same elastic stiffness constants as the isotropic rod model to produce what is termed the Zhurkin three-state model, but without the  $+5^\circ$  increase in roll as that applies to YR dimer steps CA and TG<sup>82</sup>. Separate analysis of dimeric twist from electrophoretic, crystallographic, and nuclease digestion data reported ten sequence-dependent twist values based on the ten unique dimer sequence steps, referred to here as the Trifonov ten-state model<sup>95</sup>. Thus, each of the four RR steps in the current study will have its own twist value, with the AG step twist as the most under-twisted of the four and the GG dimer, at 10.69-bp/turn twist, at the closest to the isotropic

rod model of 10.5 bp/turn (Table 3). These two models can be combined for a more robust analysis of sequence dependence.

The  $6 \times 6$  elastic force constant matrix in Equation 9, just like the intrinsic parameters, can be influenced by the step sequence in the model. Each value in this matrix represents a relationship between any pair of step parameters and can be broken up into four  $3 \times 3$  matrices,

$$\mathbf{E}_{XY} = \begin{bmatrix} \mathbf{F}_{XY} & \mathbf{G}_{XY} \\ \mathbf{G}_{XY} & \mathbf{H}_{XY} \end{bmatrix}, \quad [10]$$

organized with respect to either the rotational ( $\mathbf{F}_{XY}$ ), translational ( $\mathbf{H}_{XY}$ ), or rotational-translational ( $\mathbf{G}_{XY}$ ) parameters. For both the homopolymeric and sequence-dependent rest-state models, optimization will focus solely on changes with respect to the rotational parameters and will not take account of any effects of step sequence. Focusing on rotational parameters will result in a  $6 \times 6$  diagonal matrix after (i) setting  $\mathbf{G}_{XY} = \mathbf{0}$ , (ii) setting all cross-parameter interaction constants to zero ( $[\mathbf{E}_{XY}]_{ij} = 0$  for  $i \neq j$ ), and (iii) suppressing any changes of the translational parameters by setting the elastic stiffness terms for shift, slide, and rise to a high value  $(20.0 \frac{k_B T}{\text{\AA}^2})^{96}$ . The choice of bend and twist elastic stiffness constants for these rest-state models comes from the experimentally-determined bending persistence length and its relationship with stiffness due to twist<sup>65</sup>. Use of the 3.4-Å rise of the isotropic B-DNA model and the bending persistence of 47.77-nm result in angular fluctuations of  $\pm 4.84^\circ$  from the following modified equation<sup>65,97</sup>:

$$a = \frac{2\rho_3}{\langle \theta_1^2 \rangle + \langle \theta_2^2 \rangle}, \quad [11]$$

where  $a$  is the persistence length,  $\rho_3$  the rise, and  $\langle \theta_1^2 \rangle$  and  $\langle \theta_2^2 \rangle$  the root-mean-square fluctuations for tilt and roll, respectively. To preserve bending isotropy, both the tilt and

roll stiffness are assumed to be identical. Separate studies on DNA minicircles that explored equilibrium topoisomer distributions<sup>98</sup> and intercalating agent effects<sup>99</sup> led to the use of a twist fluctuation of  $\pm 4.09^\circ$ , corresponding to a torsional persistence length of 66.63-nm and a ratio between the stiffness constants of twist to bend of 1.4<sup>65</sup>. This produces the elastic stiffness matrix used for each of the dimer sequences given in Equation 12.

$$\mathbf{E}_{XY} = \begin{bmatrix} \frac{0.0427}{\text{deg}^2} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{0.0427}{\text{deg}^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{0.0598}{\text{deg}^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{20}{\text{\AA}^2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{20}{\text{\AA}^2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{20}{\text{\AA}^2} \end{bmatrix} k_B T \quad [12]$$

For every element of this matrix, the inverse-root of each non-zero element corresponds to the square of the parametric fluctuations. As with the intrinsic parameters, these elastic force constants can change with respect to different XY dimer sequences; however, this is the only place in the elastic energy calculation where parametric coupling, such as roll-twist, can be incorporated.

## RESULTS & DISCUSSION

The results discussed here are from elastic energy minimization calculations carried out on 150-bp minicircles that vary based on initial circular structure and nucleotide sequence, as summarized in Table 4. This study was conducted in two parts: the first focusing on the use of homopolymeric models—the isotropic rod model of B-form DNA (the IdealDNA model) and a set of B-form DNA parameters compiled from a 1998 Protein Data Bank survey (the BDNA98 model)—and the second focusing on sequence dependence. The latter portion of the study was conducted to understand the role nucleotide sequence plays in reaching low-energy stable states, primarily focusing on how base-pair step bend and twist influence the global structure. This is done using one of three forcefields, each of which differs in rest-state step parameters but maintain identical elastic force constants consistent with a bending persistence length of 47.77nm, torsional:bending stiffness ratio of 1.4, and no cross-parameter coupling. These forcefields include: the Trifonov model, which emphasizes variation in bp-step twist; the Zhurkin model, emphasizing bp-step roll; and a third model that brings together the two individual forcefields referred to as the Zhurkin-Trifonov model (Table 3 in Methodology). Here, variation in roll is used as the component of bend that describes the opening and closing of grooves and is more likely to change compared to the tilt parameter, whose values reflect asymmetric stretching and compression in the phosphate backbones. To ensure that the rotational step parameters are the main influences in elastic optimization, all changes to the translational shift, slide, and rise parameters are suppressed (Equation 12).

Table 4. Summary of the initial circular structural details separated by structural construction and sequence composition. For additional details, visit the Methods “Initial Circular Models” section.

Structural Details		Initial tilt, roll, twist ( $\theta_1, \theta_2, \theta_3$ )
<b>P-14</b>	Planar Lk <sub>o</sub> 14.3	$(2.4^\circ \sin(34.3^\circ n), 2.4^\circ \cos(34.3^\circ n), 34.3^\circ), n = [1,150]$
<b>P-15</b>	Planar Lk <sub>o</sub> 15.0	$(2.4^\circ \sin(36.0^\circ n), 2.4^\circ \cos(36.0^\circ n), 36.0^\circ), n = [1,150]$
<b>R-15</b>	O-ring Lk <sub>o</sub> 15.0	$\begin{cases} (0^\circ, 0^\circ, 36.0^\circ), \text{ every } 1^{\text{st}} - 5^{\text{th}} \text{ step} \\ (0^\circ, 7.413^\circ, 35.568^\circ), \text{ every } 6^{\text{th}} - 10^{\text{th}} \text{ step} \end{cases}$
A <sub>x</sub> G <sub>x</sub> Sequence Details		Sequence from central 20 base pairs (n=66-85)
<b>X = 1</b>	A <sub>1</sub> G <sub>1</sub>	...AGAGA GAGAGAGAGA GAGAG...
<b>X = 2</b>	A <sub>2</sub> G <sub>2</sub>	*...GGAAG GAAGGAAGGA AGGAA...*
<b>X = 3</b>	A <sub>3</sub> G <sub>3</sub>	...AGGGA AAGGGAAAGG GAAAAG...
<b>X = 4</b>	A <sub>4</sub> G <sub>4</sub>	*...GGAAA AGGGGAAAAG GGGAA...*
<b>X = 5</b>	A <sub>5</sub> G <sub>5</sub>	...AAAAA GGGGGAAAAA GGGGG...

### *Elastic Energy Minimization Using Homopolymeric Forcefields*

The planar P-14, P-15, and polygonal R-15 initial circles<sup>\*\*\*</sup> were subjected to homopolymeric optimization calculations to determine if the differences in topological linking number and/or initial bp-step parameters will result in different optimized energetic states. Molecular models of the resultant optimized configurations showed no locally bent regions and a regular twist of base pairs along the double helical axis (Figure 5.A and Supplemental Figure A.IV.1). Calculations of the root-mean-square deviation (RMSD) between the IdealDNA and BDNA98 configurations of each circular state resulted in values of less than 2-Å, with circles of the latter model having a slightly smaller mean

<sup>\*\*\*</sup> These three initial circular states are described in Methods: Initial Circular States. The number in the shortened names matches the initial linking number. Summaries of these circles are found in Table 4.

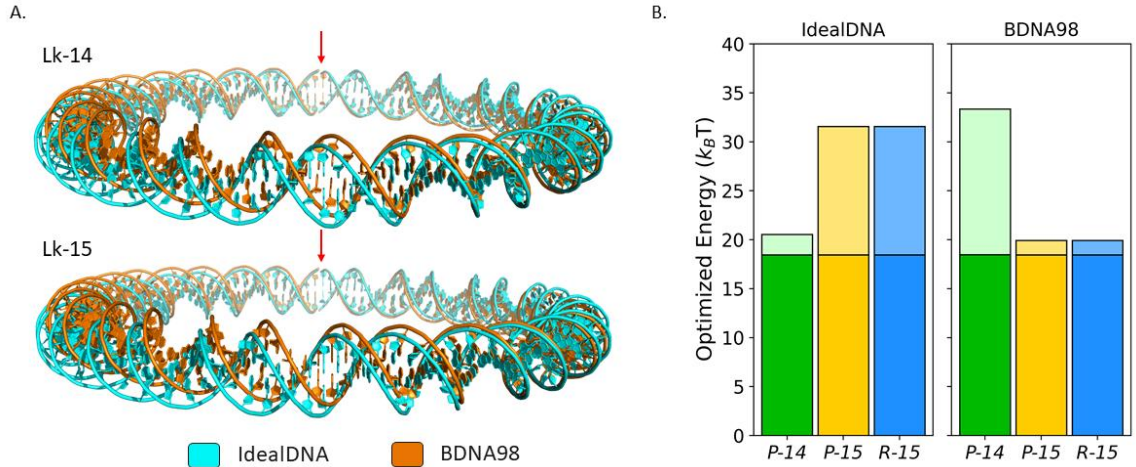


Figure 5. Structural and energetic data of the 150-bp circles optimized with the homopolymeric forcefields. (A) The pair of molecular models showing 150-bp minicircles separated by initial topology and colored with respect to homopolymer forcefield used. Both the Lk 14 (top) and Lk 15 (bottom) optimized states resulted in planar minicircles using either the IdealDNA (cyan) or BDNA98 (orange) forcefields. The models differ in the centerline curve made by the bp origins if each optimized state is superimposed at the first base pair (red arrow), resulting in similar RMSD values (1.72Å between the Lk 14 circles and 1.71 Å between the Lk 15 circles). While the P-15 and R-15 initial states were used, they resulted in identical optimized states. (B) The optimized elastic energy (units of  $k_B T$ ) of 150-bp minicircles determined with homopolymeric forcefields IdealDNA (left) and BDNA98 (right). Each column represents the three initial circular constructs used: planar circles of Lk 14 (P-14, green) and 15 (P-15, yellow), and the o-ring of Lk 15 (R-15, blue). In addition, each column is partitioned based on bp-step parameter contributions, where the bending energy (darker colors) is situated below the twist energy (light colors). For both forcefields, the bending elastic energy is 18.5  $k_B T$ , the value associated with a 150-bp planar circle with a uniform bend of  $2.4^\circ/\text{step}$ .

radius due to its lower intrinsic rise. This slight difference in atomic positions is evident once the optimized models are superimposed on the first base pair (red arrows in Figure 5.A), showing an apparent global tilt of both the BDNA98-optimized circular states of initial linking numbers 14 and 15 with respect to the isotropic IdealDNA circles. Use of 3-D principal component analysis (PCA) on the base-pair origins of each optimized state showed that optimization with the IdealDNA model resulted in perfectly planar circular structures while those optimized with the BDNA98 forcefield shows periodic deviations out of the  $\hat{e}_1\hat{e}_2$  principal axes plane (Supplemental Figure A.IV.1). Data in Table 2 showed that the intrinsic BDNA98 parameters include both a higher step roll and twist than IdealDNA, so one would anticipate an optimized model that has a consistent opening of the minor groove. These changes occur in concert with translation along the long axis

between successive base pairs, i.e. slide, and may work in concert to yield the overall change in optimized global structure. Additionally, while the P-15 and R-15 initial states differed in bp-step parameter values (Figure 3 in Methodology), optimization of both circles resulted in configurations with identical rigid-body parameters but out of phase base pair origin positions of the BDNA98 states (Supplemental Figure A.IV.1).

While PCA of the isotropically-optimized bp origins showed perfectly planar circles, the differences in linking number lead to differences in the twisting of base pairs about the double-helical axis, which plays a significant role in determining accessible energetic states. Each initial circle resulted in two different energy states based on the homopolymeric forcefield used, resulting in differences in magnitude by at least half (Figure 5.B). This difference appears to be tied to properties of the initial circle, with the isotropic rod model favoring the P-14 circle and the BDNA98 model favoring both the P-15 and R-15 circles. Because of the similarity between the optimized bp-step parameters of the P-15 and R-15 circles, these optimized energetic values are identical, as expected.

As there is no coupling between parameter terms in the forcefields, the total energy can be partitioned by bp-step parameter, highlighting how parameters contribute to the optimized state (Figure 5.B). In the case of these two simplified models, differences are seen in dimeric twist. Regardless of the intrinsic rest-state used, each optimized state has a total bending contribution of  $18.5-k_B T$  due to the uniform  $2.4^\circ/\text{bp}$ -step bend needed for a circle of length 150-bp to close. Comparison between the initial state and rest-state twist angles offers an explanation as to why P-14 has a much lower twist-contribution to the energy than P-15 for the IdealDNA model but higher for the BDNA98 model. The initial twist for P-14 is identical to the twist of IdealDNA ( $34.3^\circ$ ), while the initial twist for P-15

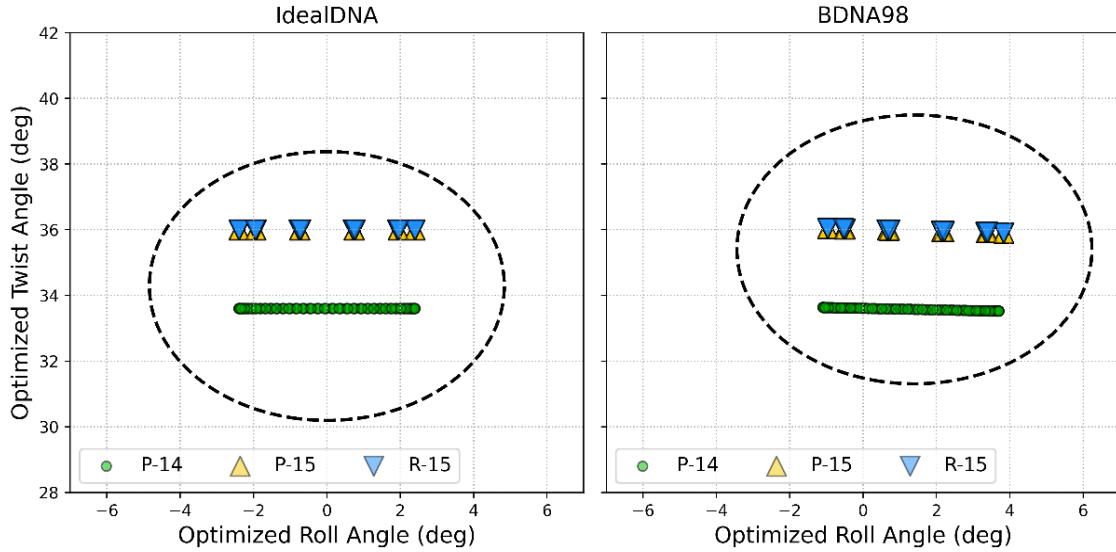


Figure 6. Optimized roll and twist parameters with respect to the  $\frac{1}{2}$ - $k_B T$  energy barrier of the 150-bp minicircles. (Left) Optimized bp-step roll and twist from IdealDNA optimizations of the planar circles of Lk 14 (P-14, green  $\circ$ ) and 15 (P-15, yellow  $\blacktriangle$ ), and the o-ring of Lk 15 (R-15, blue  $\blacktriangle$ ) within the  $\frac{1}{2}$ - $k_B T$  energy barrier (dashed ellipse), centered at  $(0^\circ, 34.3^\circ)$ . The optimized twist angles do not vary in any optimized state and are those specifically associated with planar circles of Lk 14 ( $\theta_3=33.6^\circ$ ) and Lk 15 ( $\theta_3=36.0^\circ$ ). The optimized roll angles range between  $\pm 2.4^\circ$  and when used with the optimized tilt angle result in a uniform bend of  $2.4^\circ$  (Supplemental Figure A.IV.2). (Right) Optimized bp-step roll and twist from the BDNA98 optimizations of the P-14, P-15, and R-15 initial minicircles within the energy barrier centered at  $(1.40^\circ, 35.4^\circ)$ . Both the Lk 14 and Lk 15 plots have a slightly negative slope, indicative of a slight variation in twist, with a repeat in the  $\pm 2.4^\circ$  range of roll values.

and R-15 are closer to the BDNA twist ( $35.4^\circ$ ). However, optimized twist angles show something slightly different, as exhibited in Figure 6, which highlights optimized roll and twist values with respect to a  $\frac{1}{2}$ - $k_B T$  energy barrier (dashed ellipse) whose center lies at the rest-state values. The optimized twist values associated with P-15 and R-15 are, again, consistent with the initial and BDNA98 intrinsic twists, resulting in low twist energy contributions with this forcefield and twist values relatively close to the center of the energy barrier. However, the P-14 optimized twist are slightly lower than both the initial and IdealDNA intrinsic twist. As stated earlier, the P-14 initial constructs were made with a  $34.3^\circ$  twist, in line with 10.5-bp/turn and a relaxed linking number of 14.3, which can be a non-integer value. However, the optimized twist angles as seen in Figure 6 are  $33.6^\circ$ , in line with an imposed linking number of exactly 14. As these minicircular models are



constructed with the first and last base pairs perfectly overlapped and optimization procedures were specific to keeping the last base pair frame rigid throughout the calculation, the optimized linking numbers must be integer values, so the emergence of the  $33.6^\circ$  optimized twist is expected. While the bp-step twist optimized to single values related to the linking number of the initial state, the bp-step roll values vary along the circle. The range of this spread is within  $2.4^\circ$  of the intrinsic roll value for each forcefield and is similar to what occurs with optimized tilt (Supplemental Figure A.IV.2). As the bend at each step is directly related to the tilt and roll of that step (Equation 6), this behavior is expected due to the assumed isotropic bending of the model.

#### *Incorporating Sequence-dependent Intrinsic Twist and Roll*

Use of the homopolymeric forcefields highlighted differences between the initial constructs mainly because of the topological linking numbers, where results showed differences in low-energy states that depend heavily on the relationship between the intrinsic and optimized twists. While optimized bend remained constant due to the length of the circles, further work can be done to determine the extent to which bend and twist influence elastic energy minimization at the bp-step level. This will be done by utilizing a set of rest state models that vary in the intrinsic twist and/or roll based on the dimer sequence. Optimization calculations were conducted on the three 150-bp initial states, each of which were set with one of five purine-rich sequence motifs whose dimer sequences were uniformly distributed (Table 4), resulting in a total of 15 initial states. While three initial configurations were used, the optimized states of both the planar P-15 and o-ring R-15 were again identical (as seen in the homopolymer optimization results). Therefore, for

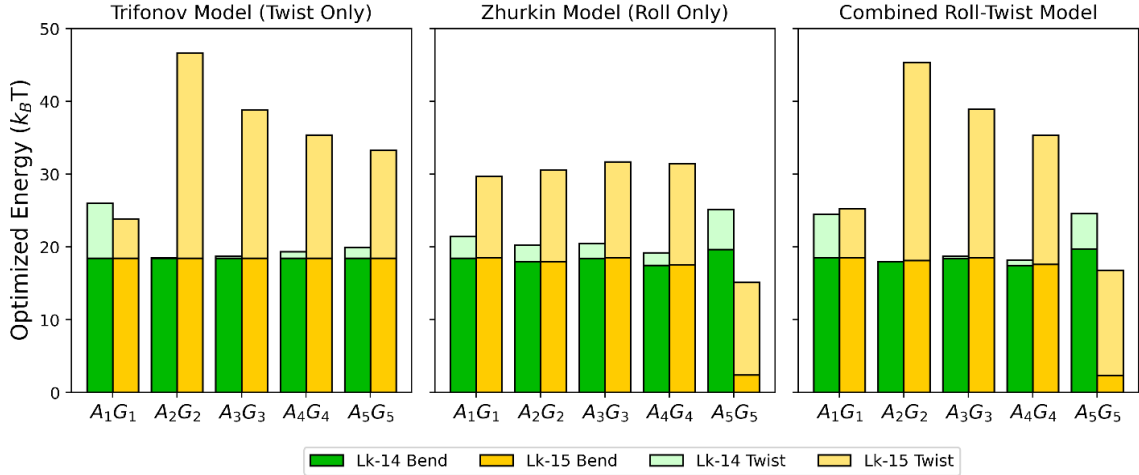


Figure 7. The optimized elastic energy values (units of  $k_B T$ ) of the Lk 14 (green) and Lk 15 (yellow) 150-bp minicircles using sequence-dependent forcefields whose elastic force constants were identical. Separate plots are from the three different intrinsic rest states used: the twist-specific Trifonov model (left), the roll-specific Zhurkin model (center), and the combined Zhurkin/Trifonov model (right). The labels along the x-axis are of the sequences used in the initial circular states (see Table 4 for additional details). Each column is partitioned based on bp-step parameter contributions, with bending contributions (darker colors) placed below twist contributions (light colors). (Left) Energetic values from optimization with the Trifonov model show that Lk 14 circles are the consistent low energy states once homodimers are incorporated into the sequence with little to no twist contribution to elastic energy. (Center) Energetic values from optimization with the Zhurkin model show that, again, the Lk 14 circles are the lowest energy states. However, this pattern changes for the A<sub>5</sub>G<sub>5</sub> sequence, where there is a major shift not only in the lowest energy state but also in the bending energy contributions. (Right) Energetic values from optimization with the combined Zhurkin-Trifonov model show similar energetic states between the A<sub>2</sub>G<sub>2</sub>-A<sub>4</sub>G<sub>4</sub> states here and the Trifonov results and between the A<sub>5</sub>G<sub>5</sub> states and the Zhurkin optimization.

the remainder of this study the results will be grouped by the linking numbers of the initial states: the planar P-14 circular data will now be labeled Lk 14 while both the planar P-15 and polygonal R-15 data will now be combined and labeled Lk 15.

The total elastic energy from each optimized set of minicircles is compiled in Figure 7, organized by rest state models, initial linking numbers, and nucleotide sequences. As the stiffness constants are uncoupled, the elastic energies can be partitioned between bend and twist contributions. A quick glance of Figure 7 shows that the majority of the low energy states involve the Lk 14 topoisomers and that, similar to what was observed in homopolymeric optimizations, both the optimized energies reflect the dimer twist while the bend contributions to elastic energy do not deviate significantly from the  $18.5-k_B T$

value. The only deviations that occur in bending contributions are found in circles with the  $A_5G_5$  sequence motif that are optimized with forcefields that have non-zero intrinsic roll values. These results show either somewhat higher (Lk 14) or substantially lower (Lk 15) energetic contribution from bend. Detailed analysis of results is organized with respect to the rest-state model used in the optimization.

### I. Optimization with the Trifonov Rest-State Model

Closer inspection of the energy results from the twist-specific Trifonov model in Figure 7 (left) shows a possible relationship between energetic state and homodimer sequence content. Results surrounding the  $A_1G_1$  sequence, made up of repeating AG and GA heterodimers, show that the elastic energy values for Lk 14 and Lk 15 are similar enough that the states are almost equally accessible. However, topological calculations showed that the linking number associated with the optimized Lk 15 starting circles changed to 13. A change in the linking number by 2 is common in bp-step optimization of supercoiled states if unrestricted by electrostatics or path collisions. A 150-bp circle of Lk 13 will be under-twisted at  $31.2^\circ$  (11.5 bp-turn), with a dimer twist that is within  $1^\circ$  of the average between the AG ( $27.7^\circ$ ) and GA ( $36.9^\circ$ ) Trifonov rest-state twist angles (Table 3).

Upon introduction of AA and GG homodimers, optimization results showed a clear separation in energetic states while staying in the same topological states. The next longer sequence motif order,  $A_2G_2$ , resulted in a decrease in energy of the Lk 14 state compared to that of  $A_1G_1$  and a marked increase in the Lk 15 states. As the sequence motifs lengthen to  $A_5G_5$ , the optimized energies stay relatively similar for Lk 14 while there is a difference in Lk 15 optimized energies. The increase in AA|GG dimers and the decrease in AG|GA

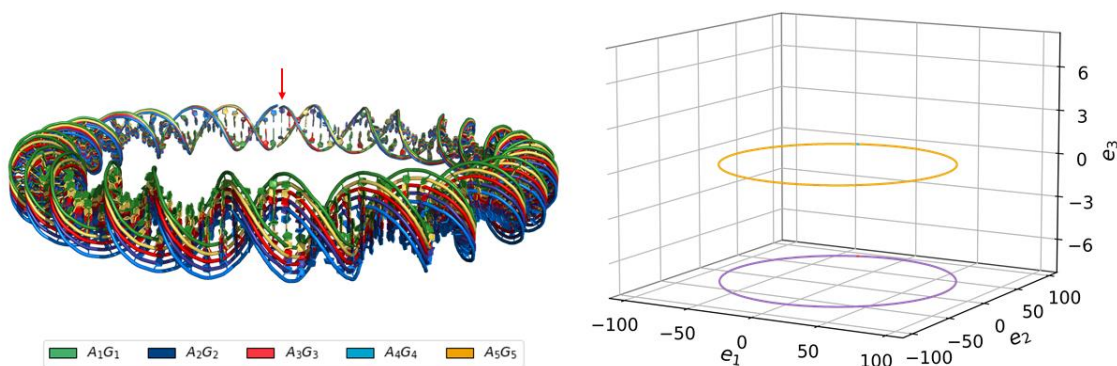


Figure 8 The optimized states of the 150-bp minicircles using the twist-specific forcefield. (Left) Molecular models of each optimized state superimposed on the first base pair of each state (red arrow) with the coding sequence depicted in the clockwise direction. Each sequence motif is colored as such: A<sub>1</sub>G<sub>1</sub> (green), A<sub>2</sub>G<sub>2</sub> (dark blue), A<sub>3</sub>G<sub>3</sub> (red), A<sub>4</sub>G<sub>4</sub> (blue), and A<sub>5</sub>G<sub>5</sub> (gold). (Right) The centerline curves of the optimized minicircles reconstructed from a 3-D principal component analysis of the base-pair origins. Both the 3-D path and the projection on the  $\hat{e}_1$ - $\hat{e}_2$  plane show smooth, planar curves for all five sequence motifs. This result is the same both when optimizing both the Lk 14 and the Lk 15 initial minicircles.

dimer lead to lower energies. As the Trifonov model pertains to changes in twist, each optimized state maintains the same bending energy contribution; thus the difference in optimized energy sits squarely in twist (Supplemental Figure A.IV.3). This recurrence in bending energy was observed above in the homopolymer optimizations and was associated with planar circles with no geometric writhe. Molecular models of the Trifonov-optimized circles, seen in Figure 8, show that the optimized states are planar and the paths of bp origins are only slightly different for the different sequences. When compared to the isotropic circles subject to the same initial conditions and sequences, the RMSD values for each circle was less than 0.40-Å, signifying little to no structural deviations from ideal with the exception of the A<sub>1</sub>G<sub>1</sub>-Lk 15 optimized state. This optimized state with a lower linking number and no out-of-plane bending has a higher RMSD of 9.36-Å from the difference in base pair twisting throughout the model.

Changes in optimized twist alone do not appear to influence the optimized structure but do influence the overall energetic state. Analysis of the bp-step data shows that as the number of homodimers increases (Figure 4), the mean optimized bp-step twist increases

Table 5. Relating dimer twist angles to the bp-step twist energy,  $\Psi_{Tw}^{opt}$ , from the optimized states using the Trifonov twist-only forcefield.

	$\langle \theta_{Tw} \rangle_{linear} (\sigma_{Tw})$	$[\Delta \theta_{Tw}]_{Lk\ 14}$	$[\Psi_{Tw}^{opt}]_{Lk\ 14}$	$[\Delta \theta_{Tw}]_{Lk\ 15}$	$[\Psi_{Tw}^{opt}]_{Lk\ 15}$
A <sub>1</sub> G <sub>1</sub>	32.05° (5.30°)	-1.55°	7.54 $k_B T$	-3.95° *	5.40 $k_B T$
A <sub>2</sub> G <sub>2</sub>	33.30° (4.43°)	-0.30°	0.06 $k_B T$	-2.70°	28.17 $k_B T$
A <sub>3</sub> G <sub>3</sub>	33.62° (4.07°)	0.02°	0.31 $k_B T$	-2.38°	20.38 $k_B T$
A <sub>4</sub> G <sub>4</sub>	33.87° (3.83°)	0.27°	0.91 $k_B T$	-2.13°	16.96 $k_B T$
A <sub>5</sub> G <sub>5</sub>	33.93° (3.69°)	0.33°	1.48 $k_B T$	-2.07°	14.85 $k_B T$

The change in bp-step twist,  $\Delta \theta_{Tw}$ , is calculated as the difference between the mean bp-step twist and the twist associated with Lk 14 (33.6°) and Lk 15 (36.0°) for circles of length 150-bp.

\* The optimized linking number associated with this state is 13, not 15, hence the large difference.

from 32.05° to 33.93° (Table 5). There also are differences between mean and expected twists of Lk 14 and Lk 15, which may add in explaining the different energetic patterns. While the twist differences for both circles increase with homodimer content, the Lk 14 differences start slightly below zero, becoming positive at the A<sub>3</sub>G<sub>3</sub> sequence, while Lk 15 differences begin at -2.7° for A<sub>2</sub>G<sub>2</sub> and increase to -2.1°. These discrepancies in magnitude appear connected to the magnitude of the bp-step twist contributions to elastic energy: greater twist contribution is the result of a greater difference in bp-step twist from zero. However, there does not appear to be a trend between increasing the length of homodimer tracts to differences in elastic energy as optimization of control sequences A<sub>150</sub> and G<sub>150</sub><sup>†††</sup> showed that the Lk 14 circles favored the dG-only sequence while Lk 15 circles favored the dA-only sequence, both with elastic energy values of 18.5- $k_B T$  due to the bending term as twist is close to 0 in the homopolymeric examples. This is expected due to the relations between the starting state linking number and the intrinsic twist of AA and GG

<sup>†††</sup> The three control sequences include an all-dA sequence (referred to as A<sub>150</sub>), an all-dG sequence (G<sub>150</sub>), and a 50:50 sequence of 75 dA bases followed by 75 dG (A<sub>75</sub>G<sub>75</sub>).

steps (Table 3). Additionally, the  $A_{75}G_{75}$  control sequence resulted in energies where the optimized states were separated by  $1-k_B T$ , as seen with the  $A_1G_1$  circles. However, unlike the  $A_1G_1$  circles, there were no changes in topological linking numbers, signifying the influence of heterodimer sequences. Here, the  $A_1G_1$  sequence repeated a constant change in intrinsic twist between  $27.7^\circ$  to  $36.9^\circ$ , yet the  $A_{75}G_{75}$  circle has one half at a constant  $35.6^\circ$  and its other half at  $33.7^\circ$ , thus a much less severe disruption to the intrinsic twist.

## II. Optimization with the Zhurkin Rest-State Model

Observations from minimization with respect to changes in the intrinsic bp-step twist showed a significant splitting of energy levels as the number of AA|GG homodimers increased. Use of the Zhurkin model forcefield—with values in intrinsic roll of  $-2^\circ$  for AA, which narrows the minor groove, versus  $+3^\circ$  for AG, GA, and GG dimers—did not reproduce the same sequence-specific pattern (Figure 7 center vs. left). The optimization results did maintain the Lk 14 states as the stable low-energy states for most of the sequences with the exceptions of the  $A_5G_5$  sequence circles. These  $A_5G_5$  circles not only show the Lk 15 circles optimized to the more favorable state but a significant change to parameter energy values was observed as the bending energies appear to drop dramatically from the  $18.5-k_B T$  values seen previously. In fact, the bending energies associated with the  $A_XG_X$  sequence motifs for  $X=1-4$  vary around  $18.5-k_B T$ . Analysis of these step-level energetics shows a variance in both the bend and twist contributions (Supplemental Figure A.IV.4). This unequal distribution in step-level energies, not seen with the Trifonov model, does not seem to steadily change with homodimer content although the standard deviations of the twist and bend energy is greater for  $A_4G_4$  than the  $A_1G_1$ - $A_3G_3$  sequences

(Supplemental Table A.IV.1). The A<sub>5</sub>G<sub>5</sub> sequence circles have either a substantial (Lk 14) or minimum (Lk 15) spread in energy while they both appear to have only one parametric change that affects its total energy—the Lk 14 twist energy is much higher than that of other four sequences while having a relatively similar bending energy. In comparison, the bend energy of the Lk 15 circle drops significantly while maintaining a similar twist contribution. This change in bend contribution is interesting as the number of base-pairs of the circle did not change, indicating that the drop in bending energy of the Lk 15/A<sub>5</sub>G<sub>5</sub> circle with respect to the X=1-4 sequence motifs to approximately  $3-k_B T$  is tied to having four homodimeric steps in a row (thus the same roll angle). For a tract of 10-bp in the A<sub>5</sub>G<sub>5</sub> circle, the four AA dimers have an intrinsic roll of  $-3^\circ$  and are then followed by six dimers with a roll of  $+2^\circ$  in the respective AG, GG, and GA dimers. With the Trifonov twist-specific model, the poly-A tracts are met with changes to twist but have a much higher difference between the AA and AG steps ( $-7.9^\circ$  between the AA and AG step) than at the GA and AA steps ( $-1.3^\circ$  between GA and AA).

Observations in energetic behaviors at the dimer level appear to agree with structural observations, particularly in the rise of out-of-plane bending with structures of higher total energy. Principal component representations using the bp origins of each of optimized state show some degree of variance in the short  $\hat{e}_3$  axis (Figure 9). For both Lk 14 and Lk 15 initial states, the A<sub>1</sub>G<sub>1</sub>-A<sub>3</sub>G<sub>3</sub> sequences optimized to relatively similar planar structures as expected due to the similarities in optimized energetic values. However, PCA projections of the A<sub>4</sub>G<sub>4</sub> optimized states, which have similar energy values as the previously-mentioned three, show regions of high variance by comparison, highlighted in the bottom portions of Figure 9 showing the  $\hat{e}_3$ -components along the circular sequence.

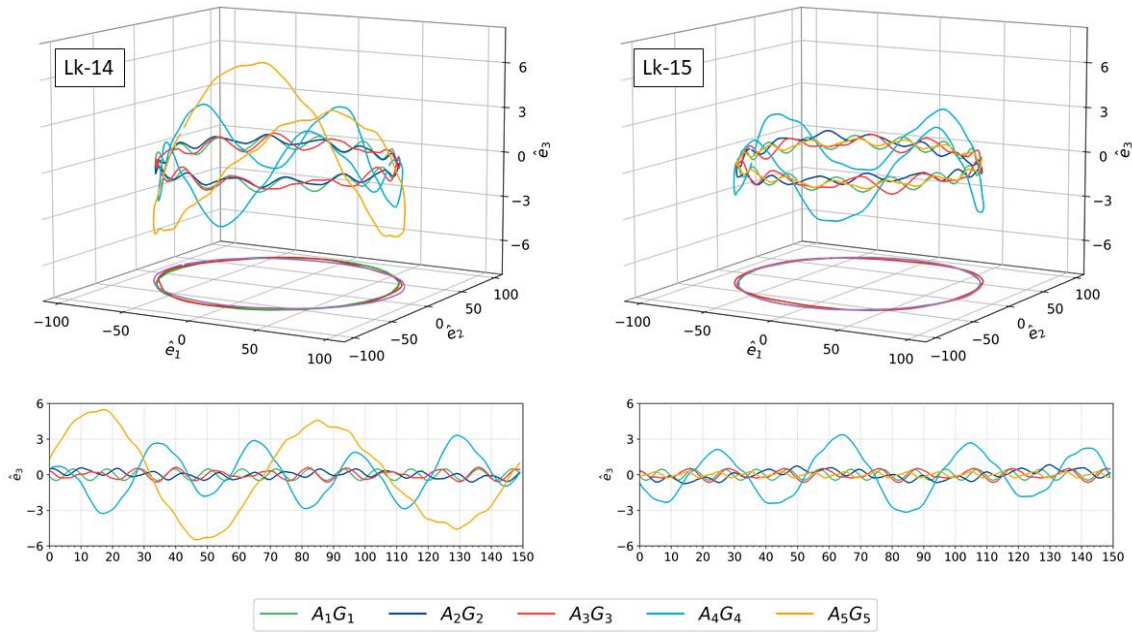


Figure 9. The centerline curves of the optimized 150-bp minicircles using the Zhurkin model forcefield reconstructed from a 3-D principal component analysis (PCA) of the base pair origins. (Top) The 3-D paths generated using PCA are centered at the origin and sit above the  $\hat{e}_1$ - $\hat{e}_2$  path projection. (Bottom) Plot of the  $\hat{e}_3$  component values along the circular sequence. For both the Lk 14 (left) and Lk 15 (right) minicircles, optimized structures for the  $A_1G_1$  (green),  $A_2G_2$  (dark blue), and  $A_3G_3$  (red) sequences are circular with only slight variation in the  $\hat{e}_1$ - $\hat{e}_3$  plane. This changes for the  $A_4G_4$  sequence circles (blue) and the Lk 14  $A_5G_5$  sequence (gold), which show periodic variation along the circle. For the  $A_4G_4$  sequence, the number of local maxima matches the number of structural bends in the optimized circular models.

The starkest difference lies in the different PCA projections for the  $A_5G_5$  sequences of either state. The Lk 14 optimized structure, which contains one region of high accumulated dimer energy and one of low, is a writhed structure unlike the planar Lk 15 structure of low bending energy, which is a planar circle with the least short axis variation.

Inspection of the  $A_4G_4$  molecular models shows slightly writhed circular structures with locally-bent regions whose global configurations resemble kinked polygons (Figure 10.A). The number of locally-bent regions found in the molecular models matches the number of local maxima observed in PCA projections (Lk 14 has five, Lk 15 four) as well as the number local maxima found in bp-step parameter profile (Supplemental Figure A.IV.5). These maxima for both the dimer bend and twist are coincident along the



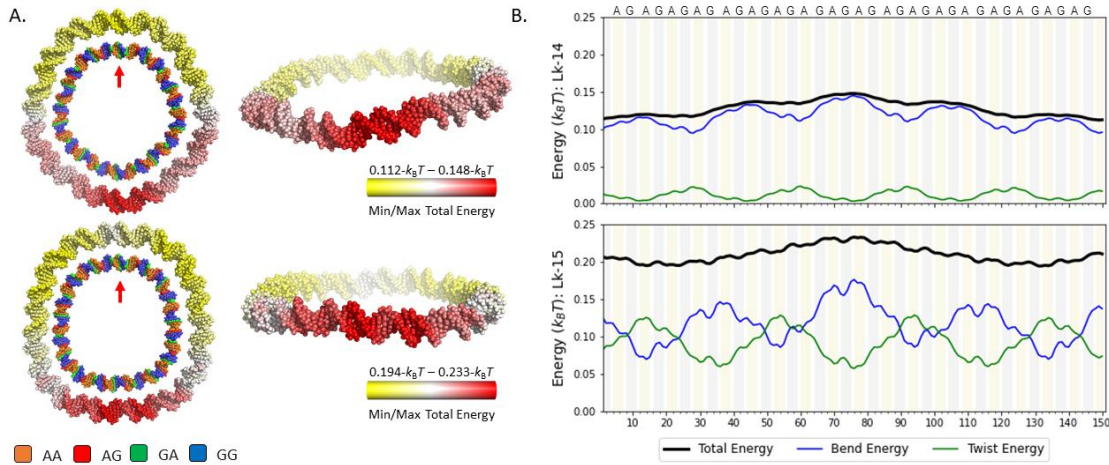


Figure 10. Optimized elastic energy profiles of the 150-bp minicircles with the A<sub>4</sub>G<sub>4</sub> sequence obtained using the Zhurkin roll-specific forcefield. (A) Molecular models of the Lk 14 (top) and Lk 15 (bottom) optimized configurations looking down the central axis (left) and rotated -80° about global x-axis (right). The exterior models are colored with respect to dimer step energy with a three-color min/max gradient (yellow for relative minimum energy, red for relative maximum). The interior models are colored with respect to the dimer step sequence (AA steps in orange, AG in red, GA in green, GG in blue). The red arrow indicates the first base pair position with the coding sequence moving in the clockwise direction. (B) The energy profiles of the Lk 14 (top) and Lk 15 (bottom) optimized states along the sequence. Black lines represent the total energy of the bp step while the blue and green lines are the bend and twist energy at the bp step, respectively. The background is colored to highlight stretches of homopolymer sequences with the top horizontal axis labeled to indicate poly-A (pale yellow) and poly-G (grey) tracts. To view the optimized bend and twist parameter relationships along the sequence of each optimized model, see Supplemental Figure A.IV.5.

sequence, indicating that for a localized bend to occur in these sequence-specific constructs there must be regions of high dimer bend with an overtwisted region. Unlike twist, the dimer bend is highly variable throughout the sequence, especially in regions of high twist, and the occurrence of structural bending may be the result of the placement of two dimer steps with high bend that are separated by 1 turn of DNA. A look at the variation of dimeric bend shows this spread in both Lk 14 and Lk 15 circles starts at an AG heterodimer and spans to the interior GG dimer in a poly-G tract 10 base pairs away (Figure 10.B). Figure 10.B also shows the decomposition of the dimer step energy profiles into bend and twist that highlights a different relationship: while bend and twist energy appear periodic, regions of energy maxima for one parameter coincide with minima of the other. This comparison also shows that the dispersion of energetic regions along the circular

configuration appears driven by bending energy, even though this trend is not observed in optimized dimer bend values.

Principal component analysis of the Zhurkin-optimized  $A_5G_5$  minicircles (Figure 9, gold lines) showed both the largest and smallest variations in the PCA axes for the Lk 14 and Lk 15 states, respectively, with the former having the greatest writhe and, thus, out-of-plane bending. Thus, detailed structural models were produced of the  $A_5G_5$ -Lk 15 state in Figure 11, where 11.A shows the spread of dimer-step energies and 11.B the dimer-step bend (exterior) and twist (interior). These models highlight observation in base-pair step analysis, detailing that for the first half of the sequence, the structure is highly energetic (A, exterior, red), under-twisted (B, interior, green) and accompanied by large differences in bend between immediate bp-steps. This changes for the last half, with an overtwisted region (Figure 11.B, central model, red region) with small changes to bend. Both changes in rotational parameters have a local and global periodicity, although the variation in twist are more apparent in the first-half of the sequence and accompanies regions with significant changes to bend. An interesting feature to the change in bend is that at the maxima associated with the change in twist, the changes in bend appear to change its periodic behavior either by flattening out (centered at  $n=33$ ) or forming a wave peak triplet (centered at  $n=108$ ). The changes to the  $A_5G_5$  optimized rotational parameters are also periodic with the Lk 15 states; however, unlike the changes found in the  $A_4G_4$  where the maxima overlapped, the changes in twist and bend are out of phase such that each parameter maximum aligns with the other parameter's minimum. This repetition was not demonstrated in the optimized circles using the Trifonov rest state, even though these states were all planar circles.

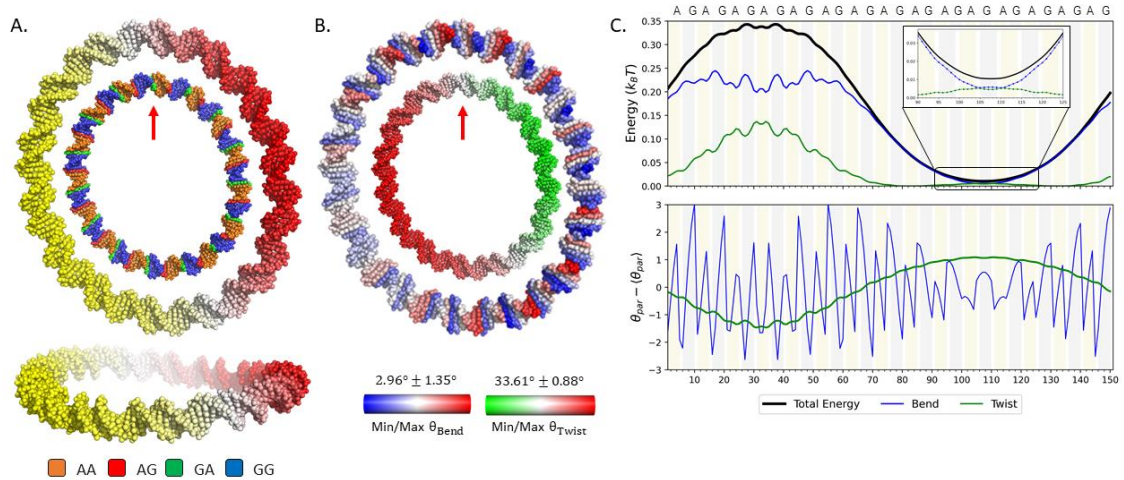


Figure 11. Comparison of the optimized configuration of the 150-bp Lk 14 A<sub>5</sub>G<sub>5</sub> minicircle from the Zhurkin roll-specific forcefield with both bend and twist parameters and energy values. (A) The molecular model with the view down the central axis (top) and rotated -80° about global x-axis (bottom) shows an overall bend of the structure. The exterior model is colored with respect to dimer step energy with a three-color min/max gradient (yellow for relative minimum energy, red for relative maximum). The interior models are colored with respect to the dimer step sequence (AA steps in orange, AG in red, GA in green, GG in blue). (B) The molecular model colored with respect the min-median-max of both optimized bend (exterior, in blue-white-red) and twist angles (interior, in green-white-red). (C) Profiles of elastic energy (top) and mean shifted optimized parameters (bottom) along the sequence. The total (black), bend (blue), and twist (green) energies show a region of high energy with a maximum at the 28<sup>th</sup> and 38<sup>th</sup> bp-steps and a region of low energy with a minimum at the 108<sup>th</sup> step. The parameter behavior, highlighted in (B), shows the high energy region consists of low dimer twist with great bend variance while the low energy region has higher twist and little bend variation. The backgrounds are colored to highlight stretches of homopolymer sequences with the top horizontal axis labeled to indicate poly-A (pale yellow) and poly-G (grey) tracts. The inset figure at the top shows the energy profile between bp-steps 90-125, a region of low elastic energy.

As there appeared to be periodic behaviors of the optimized bp-step parameters for both the A<sub>4</sub>G<sub>4</sub> and A<sub>5</sub>G<sub>5</sub> circles, linear models were constructed of each sequence using the intrinsic bp-step parameters of the Zhurkin model. Of the linear models constructed, the A<sub>4</sub>G<sub>4</sub> and A<sub>5</sub>G<sub>5</sub> models resulted in constructs with periodic helical curvature (Supplemental Figure A.IV.6). The A<sub>4</sub>G<sub>4</sub> linear model generated a spiral-like construct while the A<sub>5</sub>G<sub>5</sub> model curved in such a way that a 2-D projection is an approximate half-circle. Due to the regularity of the A<sub>4</sub>G<sub>4</sub> linear model, additional analysis was conducted to determine the superhelical character of the chain repeat to gain a better understanding of the kinked polygonal appearance of the optimized states (Figure 12). This was done using a method developed by Miyazawa by identifying virtual parameters of the polymer— bond length

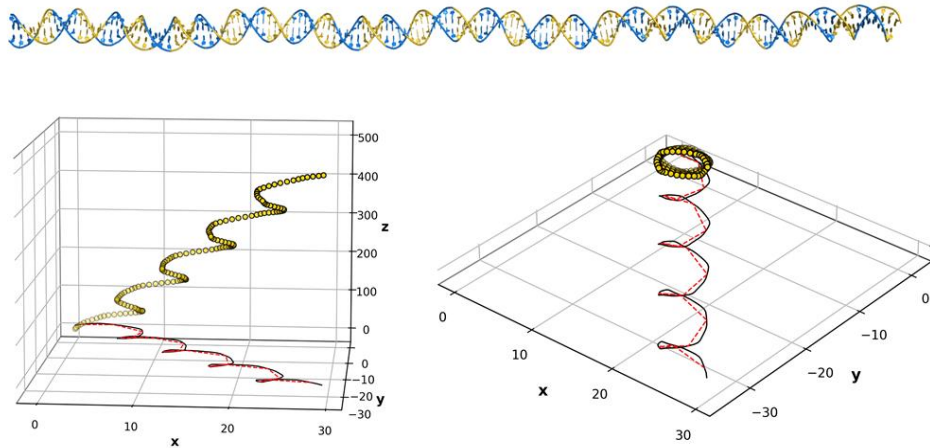


Figure 12. Fragment analysis of an A<sub>4</sub>G<sub>4</sub> linear model optimized with the Zhurkin roll-specific forcefield. (Top) The 150-bp linear molecular model of the A<sub>4</sub>G<sub>4</sub> sequence with DNA bases colored based on coding strand sequence (adenine (dA) in yellow, guanine (dG) in blue). (Bottom left) The path of the base pair origins of the A<sub>4</sub>G<sub>4</sub> linear model in 3-D space (yellow ○) with a projection of the path on the x-y plane (black line) and calculated fragments using the Miyazawa superhelical method of size  $n=8$  (dotted red lines). (Bottom right) Rotated view that looks down the A<sub>4</sub>G<sub>4</sub> superhelical axis.

( $\tau$ ), bond angle ( $\varphi$ ), and torsional angle ( $\tau$ ) between points along the polymer separated by a specific fragment size<sup>100</sup>. From these parameters the distance of each atom to the helical axis ( $\rho^{SH}$ ), the translation between two fragment points ( $d^{SH}$ ), and the rotation ( $\theta^{SH}$ ) of those points all with respect to the helical axis can be found from the following<sup>100</sup>:

$$\theta_i^{SH} = \arccos \left( 2 \cos \left( \frac{\tau_i}{2} \right) \sin \left( \frac{\varphi_i}{2} \right) \right) \quad [13]$$

$$d_i^{SH} = \frac{r_i \sin \left( \frac{\tau_i}{2} \right) \sin \left( \frac{\varphi_i}{2} \right)}{\sin \left( \frac{\theta_i}{2} \right)} \quad [14]$$

$$\rho_i^{SH} = \sqrt{\frac{r_i^2 - d_i^2}{2 (1 - \cos (\theta_i))}} \quad [15]$$

The results of these calculations for both A<sub>4</sub>G<sub>4</sub> and A<sub>5</sub>G<sub>5</sub> are found in Table 6. The fragment sizes from both sets of calculations matches the monomeric sequence repeat length. In addition, the location of the superhelical axis of the A<sub>4</sub>G<sub>4</sub> model is in much closer proximity to the path of the DNA chain than that of the A<sub>5</sub>G<sub>5</sub> model.

Table 6 Superhelical parameters of the A<sub>4</sub>G<sub>4</sub> and A<sub>5</sub>G<sub>5</sub> linear models optimized with the Zhurkin roll-specific forcefield using the Miyazawa polymer calculations.

	Fragment Size	$\rho^{\text{SH}}$ (Å)	$d^{\text{SH}}$ (Å)	$\theta^{\text{SH}}$ (deg)
A <sub>4</sub> G <sub>4</sub>	8	2.7	26.9	86.3°
A <sub>5</sub> G <sub>5</sub>	10	57.3	25.1	22.8°

### III. Optimization with the Combined Zhurkin/Trifonov Model

Previous elastic energy minimization calculations revolved around sequence-specific differences with respect to either the dimer twist or dimer roll, resulting in different parametric and structural data. Looking back, the Trifonov model had more variation in its intrinsic states than the Zhurkin model, as the former had four unique twist values (one for each dimer sequence) as opposed to only two unique roll values in the Zhurkin model. Even then, the Zhurkin model's variability hinged only on one homodimer, the AA sequence. These two sets of values for roll and twist were subsequently combined into a single rest-state to observe any differences that may result. However, construction of the 150-bp linear models with the A<sub>4</sub>G<sub>4</sub> and A<sub>5</sub>G<sub>5</sub> sequences shows that the initial states for the Zhurkin and Zhurkin/Trifonov combined models are very similar, suggesting that the optimized states between these two models shall also be similar (Supplemental Figure A.IV.7 and alignment data in Supplemental Table A.IV.2). Analysis of the optimized energetic data of the combined Zhurkin/Trifonov model seen in Figure 7 (right) shows that most optimized circular states are similar to the Trifonov energetic states except for the A<sub>5</sub>G<sub>5</sub> circular states.

Investigation of the mean bend and twist energies of all three sets of rest states used, organized in Figure 13, show that the twist energy profiles of the Trifonov and combined models are similar. However, the combined model (Figure 13.C) has a variance in the energy along the optimized states not seen in the Trifonov-optimized states (Figure 13.A).

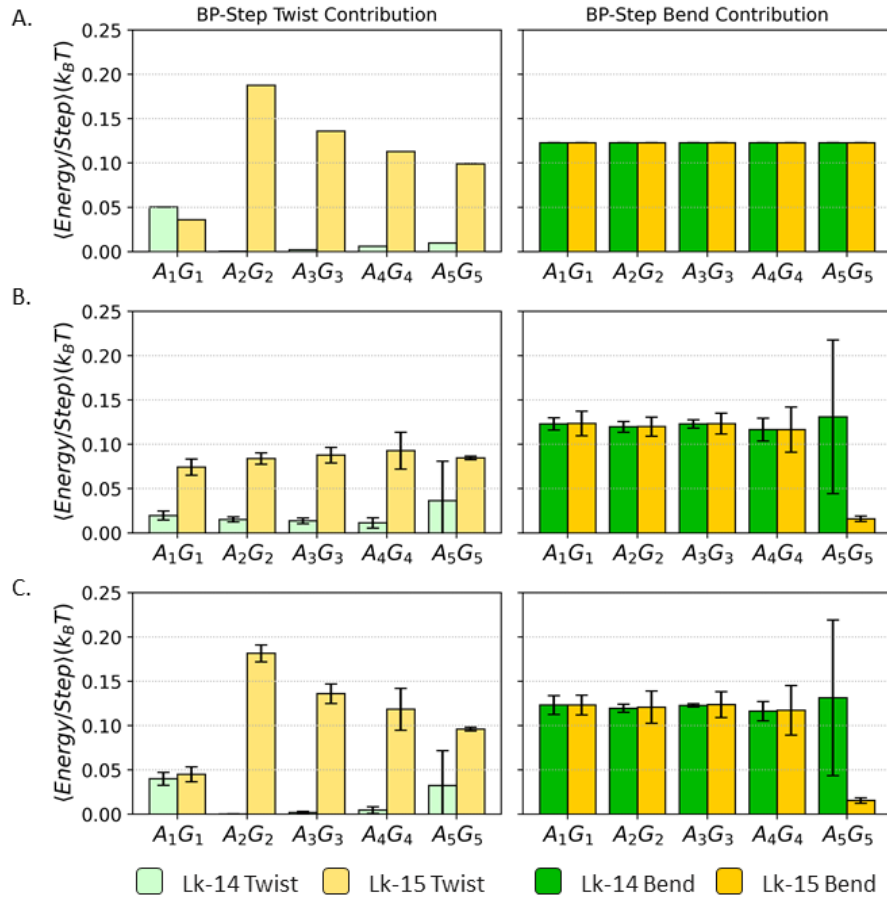


Figure 13. The optimized bp-step energies for the three sequence-specific forcefields separated by bend and twist parameters. This plot of mean dimer energy per step acts as a complement to Figure 7 with the addition of standard deviations. This figure is organized with dimeric twist on the left and dimer bend on the right. (A) Trifonov optimized twist and bend energy/step values. (B) Zhurkin optimized twist and bend energy/step values. (C) Combined Zhurkin/Trifonov optimized twist and bend energy/step values. The twist profile seen in (C) is similar to the Trifonov pattern in (A) for all homodimer-containing sequences except for the Lk 14 A<sub>5</sub>G<sub>5</sub> structure, which is like the Zhurkin optimized twist seen in (B). This comparison to the Zhurkin-optimized states continues with the bend contributions for all five sequences.

This emergence of energetic variation through the circular structure is also seen in the bending profile of the combined model, which resembles values from the Zhurkin model (Figure 13.B). This spread is also seen in the energetic profiles of the total, bend, and twist energies for each circle-sequence combination (Supplemental Figure A.IV.8 and Supplemental Table A.IV.3). These profiles, particularly those associated with the Lk 15 circular models, show general energy profiles dispersed like those from the Trifonov optimized results yet have different degrees of periodic behavior along the sequence, like

the Zhurkin profiles. The total energy profiles of these Lk 15 optimized states (Figure 13, yellow bars) show a low and high energy region with the A<sub>4</sub>G<sub>4</sub> sequence as observed with Zhurkin rest states but the emergence of a similar pattern with the A<sub>2</sub>G<sub>2</sub> sequence at roughly similar positions along the sequence (Supplemental Figure A.IV.8).

This spread of energetic values can be better understood upon closer inspection of the roll-twist parameter space with respect to  $\frac{1}{2}k_B T$  energetic barriers for, as an example, the different A<sub>5</sub>G<sub>5</sub> states (Figure 14). For the Trifonov model, all optimized twist values fell into four specific values while there was a spread of optimized roll; however, the tilt-roll parameter space showed uniform bend of 2.4° and, thus, no variance of twist, bend, and dimer energy. The Zhurkin model, with only two intrinsic roll values, resulted in a large spread of dimer twist and roll for Lk 14 circles or very little spread of either parameter in Lk 15 circles. The combined model appears to decouple the parameter spread seen in the Zhurkin model with respect to the intrinsic twist as the spread of heterodimers AG and GA appears to be the same for both Lk 14 and Lk 15 circular models.

Although differences in optimized parameters are observed, they did not seem to alter the overall structures of the optimized states. Examination of the molecular models between the Zhurkin and combined Zhurkin/Trifonov optimized states showed very little structural differences, as initially speculated from similarities in linear models, with RMSD values from sequence-based structural alignments of less than 0.5-Å. There were subtle structural differences observed with the A<sub>2</sub>G<sub>2</sub>/Lk 15 optimized circle (RMSD 1.36-Å), which also showed a small difference in its energetic profile reminiscent of the A<sub>4</sub>G<sub>4</sub> profile (Supplemental Figure A.IV.8). The differences in energy profiles between A<sub>2</sub>G<sub>2</sub> and A<sub>4</sub>G<sub>4</sub> compared to the other sequences may be the result of the unique guanine and adenine trimer



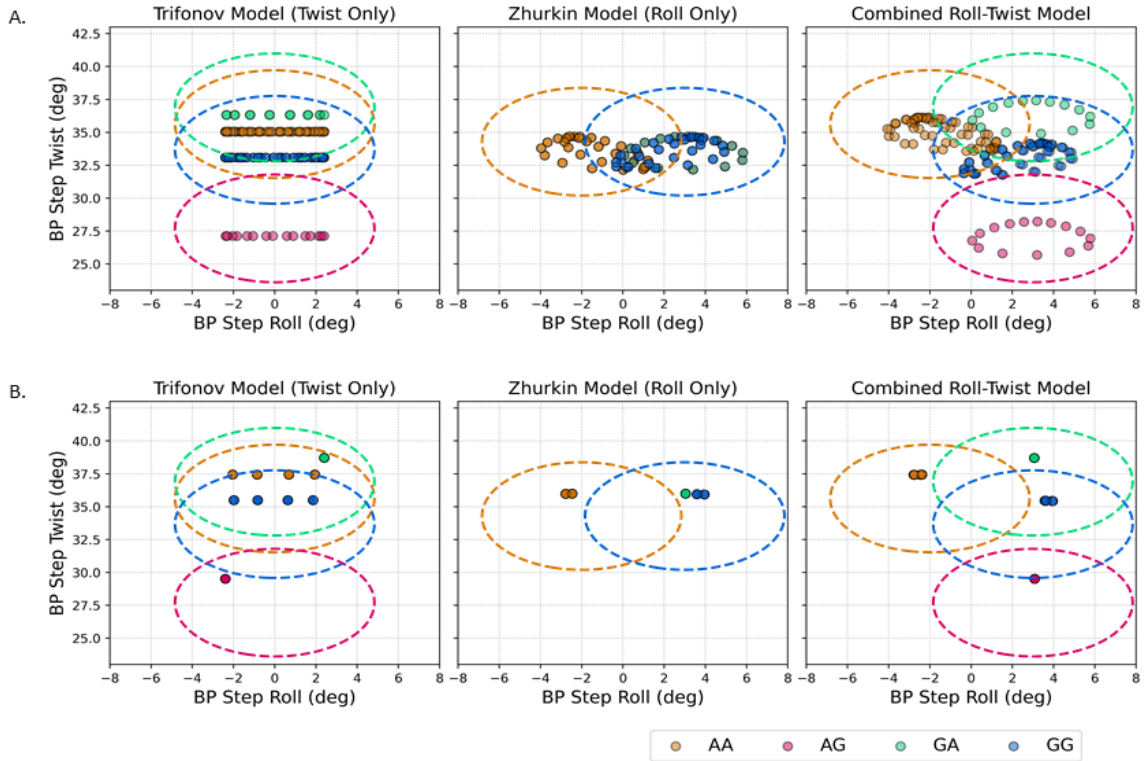


Figure 14. Relationship between the optimized roll and twist step parameters from the three sequence-dependent forcefield optimizations of the  $A_5G_5$  sequence with respect to the  $\frac{1}{2}k_B T$  energy barrier. (A) Optimized roll (x-axis) and twist (y-axis) parameters from the Lk 14  $A_5G_5$  minicircles. (B) Optimized roll and twist parameters from the Lk 15  $A_5G_5$  minicircles. The three plots in both (A) and (B) depict the three forcefields: Trifonov (twist-specific, left), Zhurkin (roll-specific, center), and the combined Zhurkin/Trifonov (right). Each energy barrier (dashed ellipse) is specific to the dimer-specific rest state values for each forcefield (AA yellow, AG red, GA green, and GG blue [see Table 3 for specific values]). Between the two sets of data, the Lk 15 optimized configuration, which is the high energy state from the Trifonov optimization and the lower-energy states for the other two, does not result in different roll/twist paired values as compared to the Lk 14 optimized states. This is especially true for the roll-influenced forcefields where the Lk 15 states appear to have two smaller regions of roll as opposed to the wide range of both roll and twist values in the Lk 14 optimized states. The combined forcefield appears to decouple the optimized twist and roll values while maintaining similar positions with respect to the elliptical centers.

sequence caps that were used in these circles to ensure a 150-bp length with as uniform a dispersion of the sequence repeat (see Table 4). However, these differences in  $A_2G_2$  optimized states of Lk 15 circles emerged only after use of the combined model. The linear models constructed with the  $A_2G_2$  sequence showed very little differences in local curvature.



## CONCLUSION

The global properties of mesoscale models of circular DNA are better understood when exploring the role of the constituent base-pair steps. This interplay between these levels of molecular structure can clarify observed behavior from previous experiments, including cyclization studies and how covalently closed forms of DNA can be formed, utilized, and protected in cellular environments. Developing a solid physical understanding that includes the influence nucleotide sequences play in DNA structure is a critical step forward in studying circular DNA and their role in cellular organization and implications to genetics and cancer research. Model development is especially useful regarding DNA chain sizes that are not easily deformable and requires a lot of energy to stay cyclized, even with multiple topoisomers that could exist for a given fragment size. The question of how differences in local level structure influences the global configuration of circular DNA remains for both experimental and computational research, especially once one includes protein-associated interactions and localized melting, bending, or kinking of base pairs that may result.

This study was designed to focus on a circular DNA size associated with the experimentally observed bending persistence length of 150-bp. In addition, two topoisomers closest to the relaxed linker number of 14.3 ( $150\text{-bp} \div 10.5\text{-bp/turn}$ ) was used as there will be a difference in the spatial disposition of successive base pairs that arises due to base pair twisting. Use of homopolymeric B-DNA models showed that twist influences which energy states are accessible and that structures of similar initial twist but different initial local-level bend will optimize to the same global configuration.

Incorporating sequence-dependence by using a repeated sequence motif of purine-purine steps provided local level control in this study, imparting rotational differences at the step level to see if they may lead to difference in global detail. Cases where only the local twist constantly changed did result in a topological change as optimization towards the lowest energy state occurred while produced global structures that maintained a planar structure with no discernable writhe. Studies where global writhe did occur were the result of incorporating changes to local bend through the opening and closing of the minor groove. Emergence of kinked polygonal shapes for A<sub>4</sub>G<sub>4</sub> sequence motif structures from roll-dependent optimization calculations led to a question of sequence composition as there was an inconsistent step pattern at the terminal ends of the sequence. Study of an A<sub>6</sub>G<sub>6</sub> monomer repeat (data not shown) which also had the same terminal end discrepancy resulted in highly-writhed, triangular shaped configurations that emerge from the same forcefields which is interesting due to previous studies of A-tract behavior and rigid curvature for spans of A<sub>5</sub> and A<sub>6</sub> within fragments<sup>74,79</sup>

One commonality across this study was the use of identical elastic force constants, something that works for idealized model use but does not work for study of a comprehensive sequence-dependent model. However, use of the optimization software allows for researchers to develop and implement forcefields that change the elastic force constants and incorporate parametric coupling of their choice. One such parametric coupling of interest has been the relationship between bend and twist. Recent physics experiments on single-molecule pulling and computational molecular dynamics simulations have been looking closely at the coupling between the roll and twist parameters<sup>54,59,62</sup>. Theories such as the Marko-Siggia and twistable worm-like chain

models incorporate the coupling of bend (both isotropic and anisotropic types) and twist. Preliminary findings from incorporating a roll-twist coupling term with the homopolymeric models with an elastic stiffness term of  $0.80 \cdot E_{\text{roll}}$  (data not shown) show a uniform decrease in total optimized elastic energy values but does show an increase in the variation of the optimized step parameters.

Results from this study did highlight how differences in local structure of topoisomers from nucleotide sequence resulted in changes to global structure. In the case of the  $A_5G_5$  sequence motif, the results did align with expectations from experimental findings pertaining to the rigid bending of A-tracts. By expanding this study to include the replacement of dG tracts with pyrimidines, increasing tract sizes beyond five, and incorporating elastic coupling terms, a physical model can be developed and used on structures of curved, looped, or circular forms of DNA found from experiment.

## APPENDIX

### *A.I. DNA: Building from the Base Up*

DNA is the most fundamental macromolecule in living cells and has a specific physical composition that allows for its characteristic double-helical shape. The helical backbones of DNA are made up of repeating nucleotide repeats, seen in Figure A.I.1 (left), containing an acidic phosphate group, a five-carbon ringed sugar, and a nitrogen-contained base. Along the backbone are phosphodiester linkages between the phosphate groups and deoxyribose sugars, where two oxygen atoms on one phosphate group bond to a 5' carbon of one sugar and a 3' carbon of a second<sup>15,101,102</sup>. Opposite the phosphodiester linkages on the sugar is the  $\beta$ -glycosidic bond that connects the sugar-phosphate backbone to a nitrogen atom on one of the four ringed, planar bases (Figure A.I.1 right)—adenine (dA), guanine (dG), thymine (dT), or cytosine (dC). The rotation angle of the glycosidic bond influences the orientation of the quasi-triangular base—the clockwise  $\alpha$ -face configuration as seen in Figure A.I.2 results in an arrangement of edges with various degrees of steric restriction that plays a role in binding activity, with the more active and open edge (the “Watson-Crick”, or “WC”, edge) pointing away from the phosphate backbone and the more restricted edge (the sugar edge) containing the glycosidic bond<sup>12,15,103–105</sup>. The Watson-Crick edges are the sites of the canonical WC  $R \bullet Y / Y \bullet R$  base-pairing, where hydrogen bonds form between one single ringed pyrimidine (Y) base with its complementary dual-ring purine (R) base as follows: the ‘AT’ and ‘TA’ base pairs ( $A \bullet T / T \bullet A$ ) and the ‘GC’ and ‘CG’ base-pairs ( $G \bullet C / C \bullet G$ ). An additional characteristic of canonical WC base-pairing is the arrangement of the sugar edges in that they occupy the same side of the base-pair (*cis*

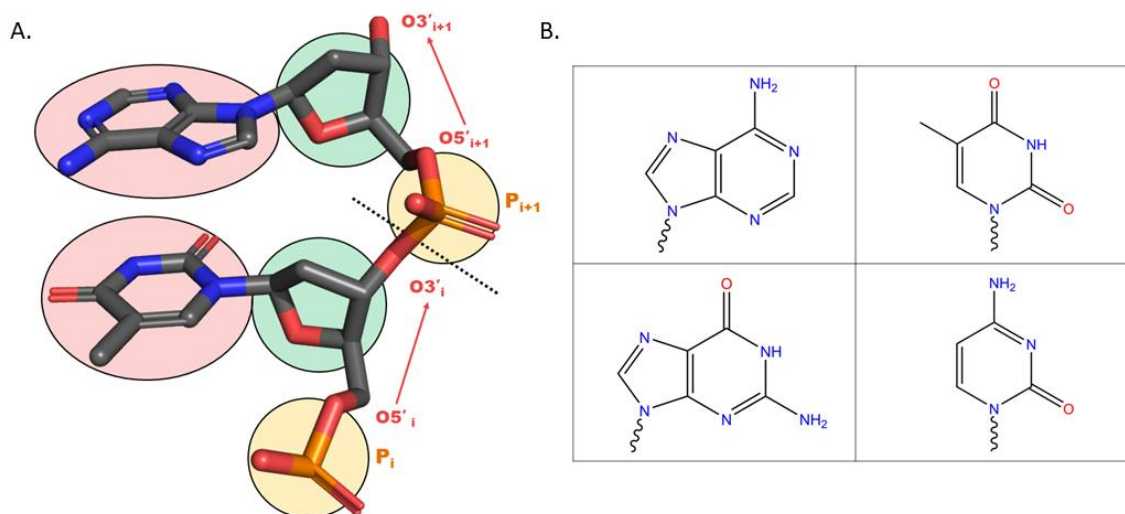


Figure A.I. 1. The DNA nucleotide base. (A) The three major components of the nucleotide monomer include: (i) an inorganic phosphate group (yellow circle), (ii) a deoxyribose sugar of variable puckering (green circle), and (iii) one of four nitrogenous bases (red circle). The phosphate group attaches to the sugar at its 5'-carbon atom while the 3' -carbon attaches to the phosphate group on the next monomer. (B) The chemical composition of the purine bases adenine (dA, top left) and guanine (dG, bottom left) and the pyrimidine bases thymine (dT, top right) and cytosine (dC, bottom right). The wavy line on the nitrogen atom represents the location of the glycosidic bond between the base and the sugar ring.

WC-WC interaction), causing the vertical axes of each  $\alpha$ -face to run in an antiparallel direction<sup>103,105,106</sup>. Experimental evidence has shown that additional base-pairing interactions can occur, particularly with the third edge of the purine bases (the “Hoogsteen” edge)<sup>††† 107–111</sup>. However, the isomorphous geometry of the canonical WC pairing allows for each possible base-pair combination to fit into the regular structural scaffold of the double-helical polymer<sup>15,105</sup>.

The establishment of base-pair (bp) level organization helps to explain certain details of the macromolecule assembly of DNA, especially details of its characteristic helix. The canonical Watson-Crick base-pairing leads to the anti-parallel arrangement of the helical backbone chains, solidified by the internal hydrogen bonding network which,

††† Two examples that incorporate the Hoogsteen edge are the Hoogsteen base-pair, where the Hoogsteen edge of a purine base interacts with the WC edge of its complementary pyrimidine base, and a base *triplet*, where a purine in an undisturbed WC base-pair will Hoogsteen pair with the WC edge of a neighboring third base (e.g. (C•G)\*G or an (T•A)\*T)

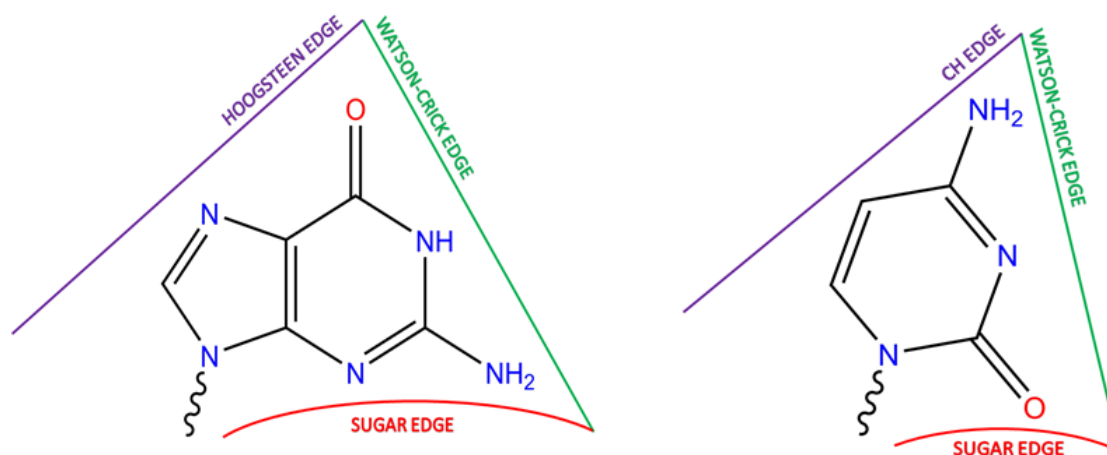


Figure A.I. 2. DNA base edges. Each ring type has three distinct edges<sup>104</sup>. Each base has a sugar edge (red edge) as the edge in closest proximity to the five-ringed deoxyribose sugar. The more chemically reactive edges that produce the biologically common forms of double helical DNA are referred to as the “Watson-Crick” (WC) edges (green edge) and are situated away from the glycosidic bond. The dual-ringed purine bases contain a Hoogsteen edge (purple edge) which is also chemically reactive and can base pair with the WC edge of a pyrimidine base given the correct conditions.

along with the pi stacking interactions between neighboring base pairs, leads to a highly stable structure<sup>104,112</sup>. As evidence from molecular biology shows proteins “read” the DNA from the 5'-3' direction along one of the polymer backbones, the antiparallel strands are subsequently labeled as the coding and complementary strands based solely on which strand is being read. This anti-parallel assembly has an added structural element in that both Hoogsteen edges in the WC base-pair will face the same direction and both sugar edges face the opposite direction, resulting in regions that are either more exposed to the external environment or more sterically crowded. This environmental exposure works for the hydrophilic sugar-phosphate backbone but not the hydrophobic bases whose edges are exposed, so a twist is imposed to best protect the internal genetic material, commonly in the right-handed orientation<sup>113</sup>. A grooving pattern emerges as a result (see Figure A.I.3), accentuated by the proximity of the sugar-phosphate backbones to one region of the macromolecule: a major groove, which contains the exposed Hoogsteen/CH edges, and a minor groove, housing the more restrictive sugar edges. This combination of stacking and

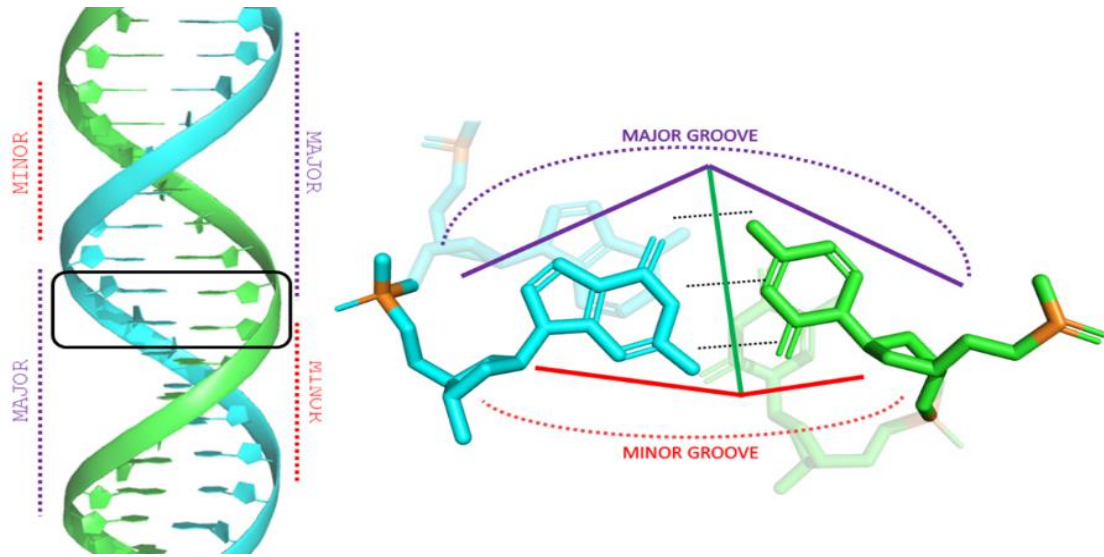


Figure A.I. 3. The major and minor grooves of DNA. A model of double-stranded B-form DNA (left) that shows the difference in widths of the major (purple dotted lines) and minor (red dotted lines) grooves. In this model the green strand is the coding strand moving in the 5'-3' direction (in this perspective, this direction is vertical). The black boxed outline highlights the dimer step on the right, which shows that the grooves depend on the edges of the base pairs. The base-pair edge exposed to the major groove consists of the Hoogsteen edge of a purine base and the CH edge of a pyrimidine base (purple bars) while the minor groove edge contains both sugar edges of the base pair (red bars).

twisting leads to a highly stable structure but does allow for a degree of variability due to environment that leads in differences in groove widths and grooving patterns, observed as far back as the initial structural findings from Franklin and Gosling<sup>3</sup>.

## A.II. The Base-Pair Step

To build the meso-scale model, the planar DNA base pairs will have assigned to them a standard right-handed orthonormal reference frame<sup>§§§</sup>,  $\mathbf{B} = \{\hat{d}_1, \hat{d}_2, \hat{d}_3\}$ , with axes that capture features of an ideal, planar, anti-parallel Watson-Crick (WC) base pair<sup>114</sup> (Figure AII.1). The  $\hat{d}_1$  axis of the plane runs parallel to the pseudo-dyad axis of an ideal base pair, pointing in the direction of the major groove. The  $\hat{d}_2$  axis runs parallel to the vector that connects the C1' atoms of the sugars in the base pair. The base-pair normal  $\hat{d}_3$  is the cross-product of  $\hat{d}_1$  and  $\hat{d}_2$  with the positive direction chosen to run parallel to the 5'-3' direction of the coding strand. The origin,  $\vec{o}$ , is chosen to minimize the displacements of the pyrimidine Y(C6) and purine R(C8) atoms along the  $\hat{d}_1$  axis<sup>114</sup>. The origins and local reference frames of a collection of DNA base pairs are used to construct the segmented pathway of the double-helical polymer.

The construction of a chain of  $N$  base pairs can be described in terms of the geometry of successive base pairs, i.e. each base-pair step as seen in Figure A.II.2. The origins  $\vec{o}^n$  and  $\vec{o}^{n+1}$  on base pairs  $n$  and  $n+1$  are used to calculate the distance vector of step  $n$ ,  $\vec{r}^n$ . The two orthogonal reference frames associated with these origins,  $\mathbf{B}^n$  and  $\mathbf{B}^{n+1}$ , can be described courtesy of a rotation matrix,  $\mathbf{D}^n$ , that relates these two rigid bodies with a unique rotation angle,  $\varphi^n$ , about some unit axis,  $\hat{u}^n$ . This rotation matrix is generated

---

§§§ A note on mathematical representation.

A vector will be italicized and represented with an arrow above variable (as  $\vec{v}$ ); a unit vector will be italicized and represented with a hat above variable (as  $\hat{v}$ ); a matrix will be bold-faced without symbols (as  $\mathbf{M}$ ) while an element within a matrix will be denoted using square brackets and subscript index notation of the  $i^{\text{th}}$  row,  $j^{\text{th}}$  column order (as  $[\mathbf{M}]_{ij}$ ); for basis representation, curly braces will be used to contain basis unit vectors, such as the commonly used Cartesian coordinate system  $\{\hat{x}, \hat{y}, \hat{z}\}$ .



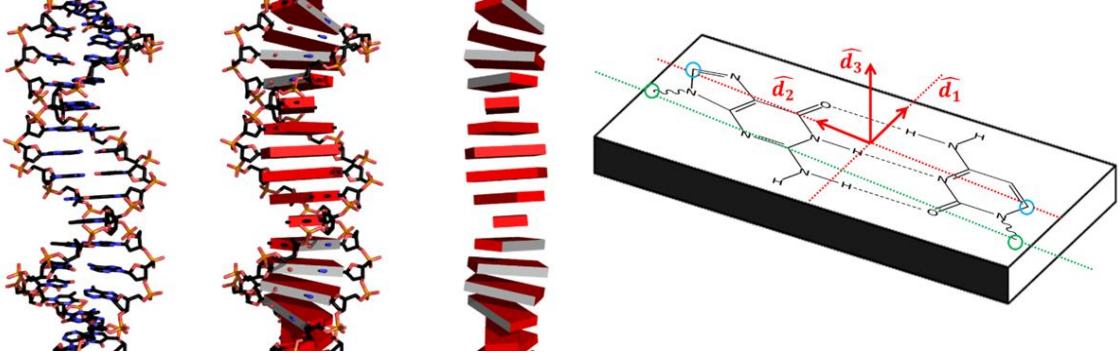


Figure A.II. 1. DNA modeling at the base pair level. (Left) Three molecular models of a full helical turn of DNA showing (left) an all-atom model, (right) a rectangular base-pair construct, and (center) an overlap of the two. (Right) A schematic showing how a base pair can be represented by a rectangle using basis  $\mathbf{B}$  (red dotted lines). The  $\hat{d}_1$  vector (the short axis) points toward the major groove edge of the base pair (minor groove edge is the black edge of the rectangle). The  $\hat{d}_2$  vector (the long axis) is set such that it runs parallel to the vector that connects the sugar atoms to each base through the glycosidic bond (green line and circles) and points toward the coding strand. The  $\hat{d}_3$  vector is the cross-product of  $\hat{d}_1$  and  $\hat{d}_2$  and points in the direction of the 5'-3' coding strand direction. The origin of  $\mathbf{B}$  sits along the WC-WC interface and is positioned to minimize displacements of  $\hat{d}_1$  with respect to the purine C8 and pyrimidine C6 atoms (blue circles).

from the multiplication of the reference frames and is related to the ZYZ Euler-angle rotation<sup>\*\*\*\*</sup>.

$$\mathbf{D}^n = (\mathbf{B}^n)^T \mathbf{B}^{n+1} = \mathbf{Z}(\zeta^n) \mathbf{Y}(\kappa^n) \mathbf{Z}(\eta^n) \quad [A1]$$

with angles  $\zeta^n$ ,  $\kappa^n$ , and  $\eta^n$ . The final  $\mathbf{ZYZ}$  rotation matrix can be used to extract the Euler angles from multiplying the reference frames bases as follows:

$$\kappa^n = \arccos([\mathbf{D}^n]_{33}) : \kappa^n \in [0, \pi] \quad [A2]$$

$$\zeta^n = \arctan2([\mathbf{D}^n]_{23}, [\mathbf{D}^n]_{13}) : \zeta^n \in [-\pi, \pi] \quad [A3]$$

$$\eta^n = \arctan2([\mathbf{D}^n]_{32}, -[\mathbf{D}^n]_{31}) : \eta^n \in [-\pi, \pi] \quad [A4]$$

For angles  $\zeta^n$  and  $\eta^n$  the inverse  $\arctan2(y, x)$  or  $\arctan2(\sin, \cos)$  function is used to account for different signs within the rotation matrix.

\*\*\*\* The classical definitions of axial rotation matrices for some rotation,  $\theta$ , are used here:

$$\mathbf{X}(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix}, \mathbf{Y}(\theta) = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}, \mathbf{Z}(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

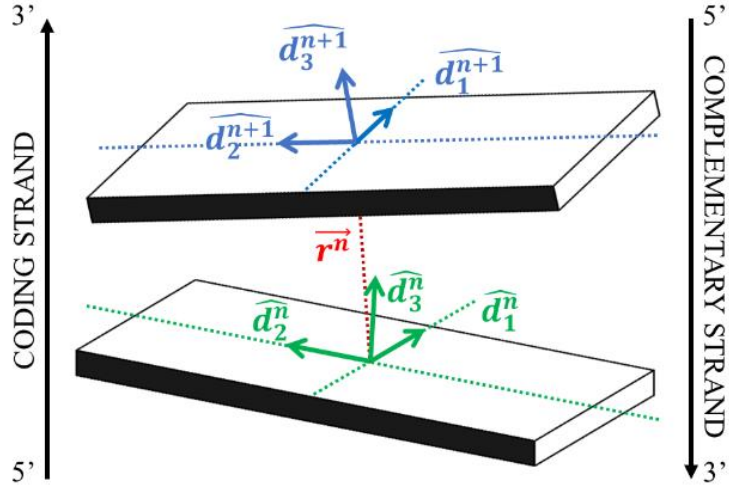


Figure A.II. 2. The base-pair step. The six base-pair step parameters are calculated with respect to bases  $\mathbf{B}^n$  (green) and  $\mathbf{B}^{n+1}$  (blue) and the displacement between their origins,  $\vec{r}^n$ . The bases are numbered with respect to the coding strand direction<sup>59</sup>. For each base pair, the black bar represents the minor-groove edge.

As such a rotation requires the two bases to be coincident to produce the axis-angle representation, a mid-step frame,  $\mathbf{s}^n$ , is generated at a position halfway between the respective origins where half the total rotation between  $\mathbf{B}^n$  and  $\mathbf{B}^{n+1}$  occurs. This frame requires a new angle,  $\gamma^n$ , with a separate **ZYZ** rotation, defined below.

$$\mathbf{s}^n = \mathbf{B}^n \mathbf{D}_s^n = \mathbf{B}^n \left( \mathbf{Z}(\zeta^n) \mathbf{Y}\left(\frac{\kappa^n}{2}\right) \mathbf{Z}(\gamma^n) \right) \quad [\text{A5}]$$

$$\gamma^n = \frac{1}{2}(\eta^n - \zeta^n) \quad [\text{A6}]$$

These calculations result in a six-dimensional vector,  $\vec{p}^n$ , that describes a base-pair step with three rotational parameters ( $\theta_1^n, \theta_2^n, \theta_3^n$ ; tilt, roll, and twist, respectively) and three translational parameters ( $\rho_1^n, \rho_2^n, \rho_3^n$ ; shift, slide, and rise, respectively):

$$\vec{p}^n = (\theta_1^n, \theta_2^n, \theta_3^n, \rho_1^n, \rho_2^n, \rho_3^n)$$

$$\vec{p}^n = (\kappa^n \sin(\gamma^n), \kappa^n \cos(\gamma^n), \zeta^n + \eta^n, \vec{r}^n \cdot [\mathbf{s}^n]_1, \vec{r}^n \cdot [\mathbf{s}^n]_2, \vec{r}^n \cdot [\mathbf{s}^n]_3) \quad [\text{A7}]$$

with  $[\mathbf{s}^n]_j$  representing the  $j^{\text{th}}$  column of the mid-frame basis. Figure AII.3 below shows a representative form of each of the six bp-step parameters<sup>86</sup>. The tilt and roll parameters

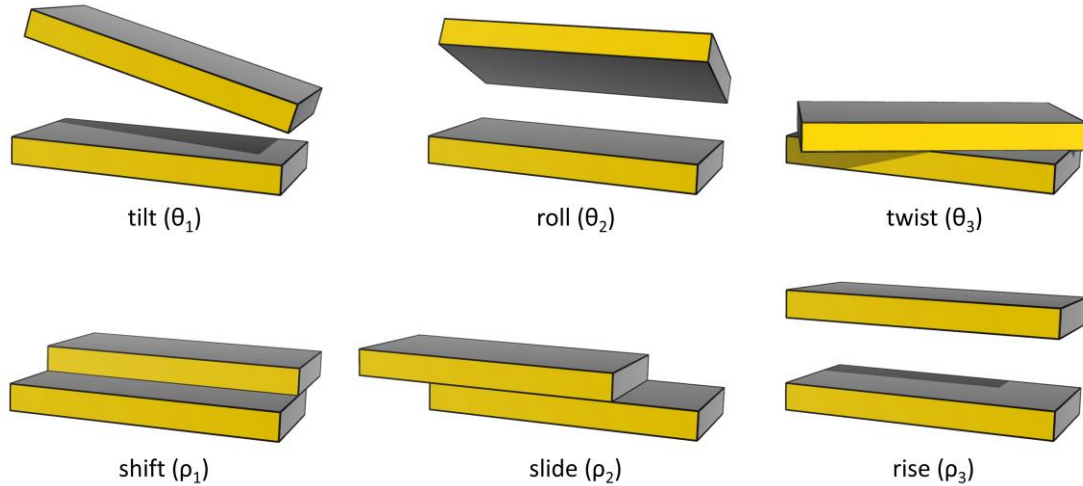


Figure A.II. 3. The six base-pair step parameters. (top) The three rotational parameters tilt, roll, and twist (top) and the three translational parameters shift, slide, and rise (bottom). The tilt and roll parameters make up the step bend,  $\Gamma$ , while the shift and slide make up the step shear. Each model is oriented such that the coding strand sits to the left of each base pair and the minor grooves face front (yellow). Each model represents the positive value associated with each parameter.

comprise the bending components of a base-pair step, while the shift and slide parameters comprise the shear components.

If a collection of step parameters is given and one wishes to derive the collection of base pair reference frames, start by obtaining the Euler angles for each step as follows:

$$\tan(\gamma^n) = \frac{\theta_1^n}{\theta_2^n} \quad [A8]$$

$$\kappa^n = \sqrt{(\theta_1^n)^2 + (\theta_2^n)^2} \quad [A9]$$

$$\zeta^n = \frac{1}{2}(\theta_3^n) - \gamma^n \quad [A10]$$

$$\eta^n = \frac{1}{2}(\theta_3^n) + \gamma^n \quad [A11]$$

The rotation matrices  $\mathbf{D}^n$  and  $\mathbf{D}_s^n$  can be made from these Euler angles. To generate the reference frame of the  $n+1^{\text{th}}$  base pair, the  $n^{\text{th}}$  base-pair reference frame is needed. If unknown, use the global reference frame of position (0, 0, 0) and the identity basis matrix.

The translational parameters  $\rho_1^n, \rho_2^n, \rho_3^n$  will be used to get the step displacement vector while the rotation matrix from the construct the reference frames:

$$\mathbf{B}^n \mathbf{D}^n = \mathbf{B}^{n+1} \quad [A12]$$

### A.III. Elastic Energy Minimization Calculations

Use of rigid-body parameters are effective in describing differences along the DNA influenced by the chirality of the polymer at any point along its contour length and the elastic stiffness through bending, torsional, shearing, and stretching deformations. These deformations arise from the relationships between sugar-phosphate backbones and base-pair stacking along the helical axis<sup>39,45</sup>. For simplicity, the elastic energy of a base-pair step  $\overline{p^i}$  can be calculated from a quadratic function that for every  $i^{\text{th}}$  base-pair step in a collection of  $N$  base-pairs,

$$\psi^i = \frac{1}{2} (\overline{p^i} - \overline{p_{XY}})^T (\mathbf{E}_{XY})^i (\overline{p^i} - \overline{p_{XY}}),$$

$$\mathbf{E}_{XY} = \begin{bmatrix} \mathbf{F}_{XY} & \mathbf{G}_{XY} \\ \mathbf{G}_{XY}^T & \mathbf{H}_{XY} \end{bmatrix} \quad [\text{A13}]$$

In the base-pair step energy expression,  $\overline{p_{XY}}$  denotes the 1x6 parameter vector from a stress-free intrinsic state for some dimer step sequence XY while  $\mathbf{E}_{XY}$  is a 6x6 elastic stiffness force constants matrix that can be broken up into multiple 3x3 matrices that describe rotation-rotation ( $\mathbf{F}_{XY}$ ), translation-translation ( $\mathbf{H}_{XY}$ ), and rotation-translation ( $\mathbf{G}_{XY}$ ) step parameter relationships<sup>96</sup>. The deformations that occur at this level depend on the defined rest state and not on neighboring base-pair steps, thus the total elastic energy is the sum of all step-level energies,

$$\Psi = \sum_{i=1}^{N-1} \psi^i \quad [\text{A14}]$$

Previously established energetic models work for a continuous tract of DNA and not for a base-pair step level collection. A predictive model developed by J.F Marko and

E.D. Siggia (the MS theory) related elasticity to the local deformation strains,  $\Omega$ , about a coordinate frame along the arc length,  $s$ , in the following form<sup>56</sup>:

$$E = \frac{k_B T}{2} \int_0^L (A_{11}\Omega_1^2 + A_{22}\Omega_2^2 + A_{33}\Omega_3^2 + 2A_{23}\Omega_2\Omega_3) ds \quad [A15]$$

Here, the  $A_{ij}$  are stiffness constants associated with the ease at which deformations occur over a given length. The MS theory can be further limited from this first-order approximation by stipulating that DNA is influenced only by a generalized bend ( $A_{11} = A_{22} = A$ ) and twist ( $A_{33} = C$ ) with no coupling ( $A_{23} = 0$ )<sup>49</sup>. This limitation is referred to as the twistable worm-like chain model (TWLC),

$$E = \frac{k_B T}{2} \int_0^L (A\Omega_{\text{bend}}^2 + C\Omega_{\text{twist}}^2) ds \quad [A16]$$

Stiffness along a polymer arises from angular correlations between neighboring bonds or segments and will decay after some distance limit, known as the persistence length<sup>50,115,116</sup>. Once a polymer extends beyond this size limit, its global configuration will be influenced by entropically-driven thermal motion and would be considered more flexible. In the case of the TWLC model, there are two persistence lengths and depending on solution conditions, the bending length is often cited at ~50nm (150bp) while values of twist persistence ranges from 25-100nm (70-300bp)<sup>45</sup>. Between the two, findings have pointed toward bending deformability as more influential in the ease or difficulty linear DNA can deform without loss of energy. While effective, the TWLC model does not account for a coupling between bend and twist which surveys of structural data suggest plays a role in configuration<sup>54,59</sup>.

The optimization of a collection of DNA base pairs starts with the gradient of the elastic energy of a base-pair step,

$$\frac{\partial \Psi^i}{\partial \vec{p}^i} = \frac{1}{2} \left( \mathbf{E}_{XY}^i + (\mathbf{E}_{XY}^i)^\top \right) (\vec{p}^i - \overline{\vec{p}_{XY}}) = (\mathbf{E}_{XY}^i)_s (\vec{p}^i - \overline{\vec{p}_{XY}}) \quad [A17]$$

where a symmetric version of the elastic stiffness constants matrix is used<sup>117</sup>. This form is used to determine the variation of the total elastic energy for the complete collection of DNA base-pairs in Equation [A10].

$$\delta \Psi = \left( \frac{\partial \Psi}{\partial \vec{P}} \right)^\top (\delta \vec{P}) = \sum_{i=1}^{N-1} \left( \left( (\mathbf{E}_{XY}^i)_s (\vec{p}^i - \overline{\vec{p}_{XY}}) \right)^\top (\delta \vec{p}^i) \right) \quad [A18]$$

While this works for the parameterization of base-pair steps from a set of  $\vec{p}^i$  vectors, there are dependencies on the base-pair origins and the base-pair frame axes within the rigid-body parameters that make the use of this gradient method difficult. Therefore, a vector with the step degrees-of-freedom (“step dofs”),  $\vec{\varphi}^i = (\alpha_1^i, \alpha_2^i, \alpha_3^i, r_1^i, r_2^i, r_3^i)$ , is constructed, where  $\alpha_1$ - $\alpha_3$  are identical to  $\theta_1$ - $\theta_3$  and  $r_1$ - $r_3$  are the components of the distance vector between step origins,  $\vec{r}^i$ . Like the collection of step parameters  $\vec{P}$ , the set of  $6(N-1)$  step dof vectors is defined as  $\vec{\Phi}$ . The variation in base-pair step parameters  $\delta \vec{P}$  is related to the variation in step dofs through a Jacobian matrix  $\mathbf{J}_{\vec{\Phi}}$  with respect to the step dofs.

$$\delta \vec{P} = \mathbf{J}_{\vec{\Phi}} \delta \vec{\Phi} = \left( \frac{\partial \vec{P}}{\partial \vec{\Phi}} \right) \delta \vec{\Phi} \quad [A19]$$

This leads to an expression of the elastic energy gradients between base-pair step parameters with the step dofs.

$$\frac{\partial \Psi}{\partial \vec{\Phi}} = (\mathbf{J}_{\vec{\Phi}})^\top \frac{\partial \Psi}{\partial \vec{P}} \quad [A20]$$

### A.IV. Supplemental Results

Figure A.IV. 1. Structural analysis of the 150-bp optimized minicircles using homopolymeric forcefields. The horizontal black line separates the data sets optimized with either the IdealDNA (top) and BDNA98 (bottom) homopolymeric forcefields. (Left) Molecular models of the optimized 150-bp minicircles. For each circle, rectangular blocks are used to represent base pairs connecting to a smooth sugar-phosphate backbone chain. Silver faces represent minor groove edges of the base pair. The three initial conditions P-14 (green), P-15 (yellow), and R-15 (blue) optimized with the IdealDNA forcefield (top) produced planar circles with uniform radius and no writhe. The three initial conditions optimized with the BDNA98 forcefield (bottom) resulted in circular structures with fluctuating radii and bp-step bend but with no geometric writhe. (Right) The 3-D curves of the optimized 150-bp minicircles reconstructed from principal component analysis of the base pair origins. Each 3-D principal component path has 2-D projections on the  $\hat{e}_1$ - $\hat{e}_2$  and  $\hat{e}_1$ - $\hat{e}_3$  planes to highlight any structural variations. Optimization with the IdealDNA forcefield (top) resulted in perfectly planar structures as seen by uniform centerline curves with no variation in the  $\hat{e}_1$ - $\hat{e}_3$  projection. Optimization using the BDNA98 forcefield (bottom) resulted in structures that, while producing a uniform circle in the  $\hat{e}_1$ - $\hat{e}_2$  plane, shows a degree of variance in the  $\hat{e}_1$ - $\hat{e}_3$  plane. Additionally, the P-15 and R-15 principal components appear to be out of phase yet has the same wavelength and frequency.

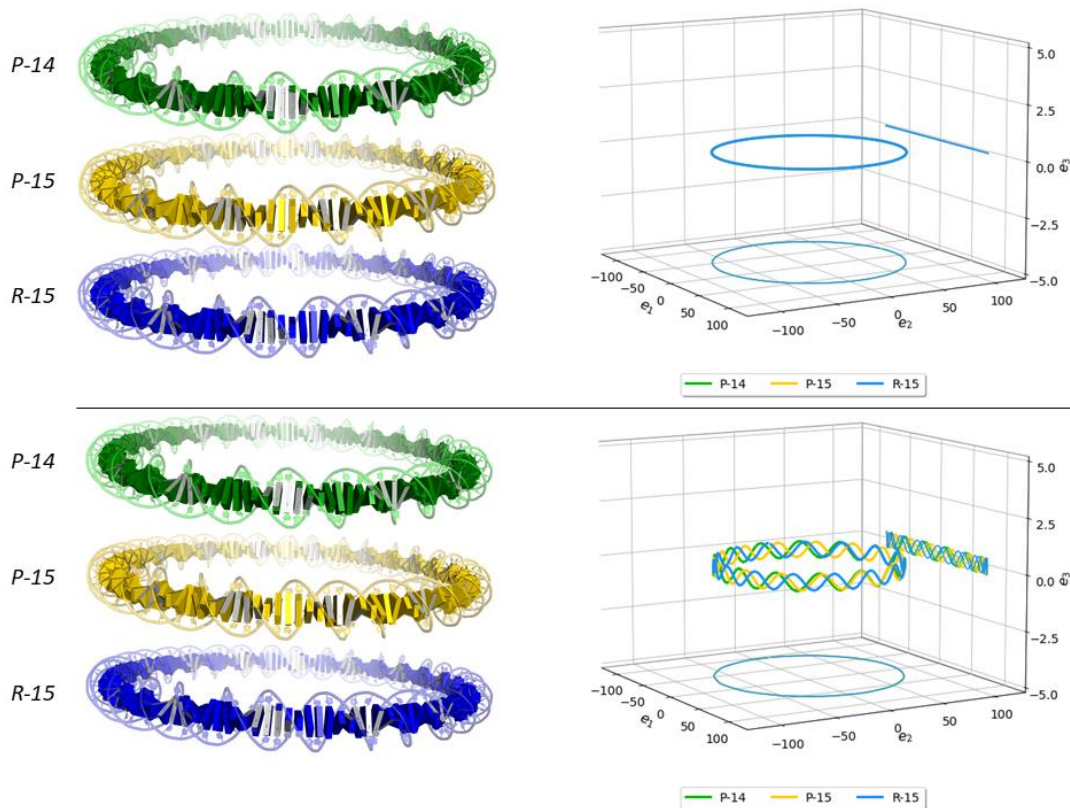




Figure A.IV. 2. The bp-step roll and tilt parameters with respect to the  $\frac{1}{2}$ - $k_B T$  energy barrier of the 150-bp minicircles. (Left) Optimized bp step roll and tilt from IdealDNA optimizations of the P-14 (green  $\circ$ ), P-15 (yellow  $\blacktriangle$ ), and R-15 (blue  $\blacktriangle$ ) initial circles resulted in a uniform circle where each point results in a bend angle of  $2.4^\circ$  (Equation 6). The center of both the energy barrier (dashed line) and the circle made by the optimized parameters lies at  $(0^\circ, 0^\circ)$ , the intrinsic values of Ideal DNA. (Right) Optimized bp step roll and tilt from BDNA98 optimizations of the P-14, P-15, and R-15 initial circles results in a similar dispersion of points where, again, each point corresponds to a bend angle of  $2.4^\circ$ , yet the center of the energy barrier shifts to a roll of  $1.4^\circ$ .

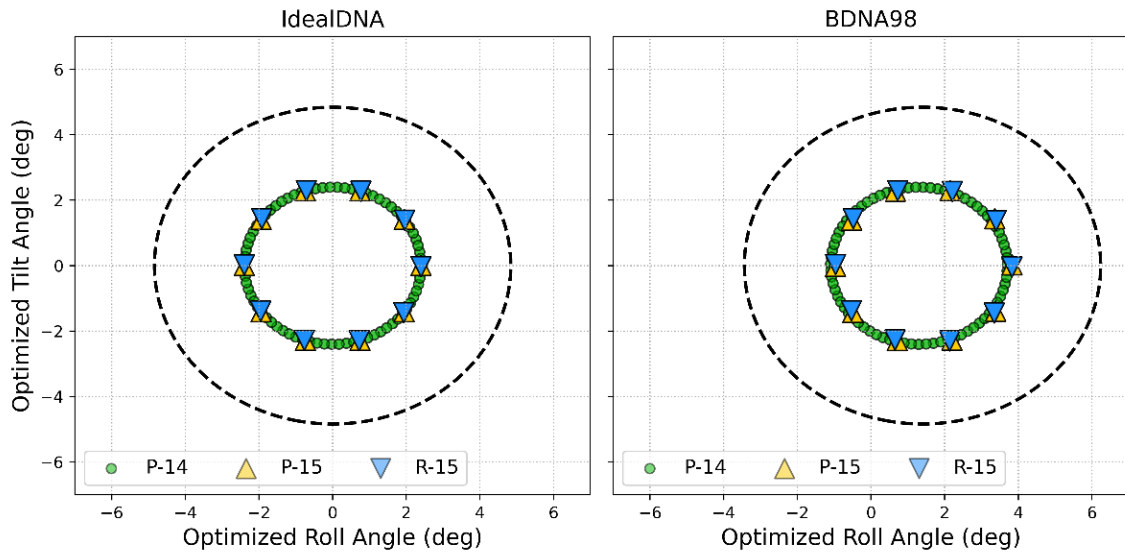


Figure A.IV. 3. The optimized elastic energy per base-pair step along the 150-bp minicircle sequence obtained using the Trifonov twist-only forcefield. The columns of this figure are organized based on initial circles (Lk 14 left, Lk 15 right) and the rows organized by total energy (top), dimer-twist energy (middle), and dimer-bend energy (bottom) in units of  $k_B T$ . In each plot, there is no variation in dimer energetic values across all sequences (colored lines:  $A_1G_1$  green,  $A_2G_2$  dark blue,  $A_3G_3$  red,  $A_4G_4$  blue, and  $A_5G_5$  gold). The differences in total energy among the different optimized states are directly tied to differences in twist energy as the bend energy does not deviate from  $0.123 k_B T/\text{step}$ , an energy value associated with the  $2.4^\circ$  dimer-level bend for a planar circle.

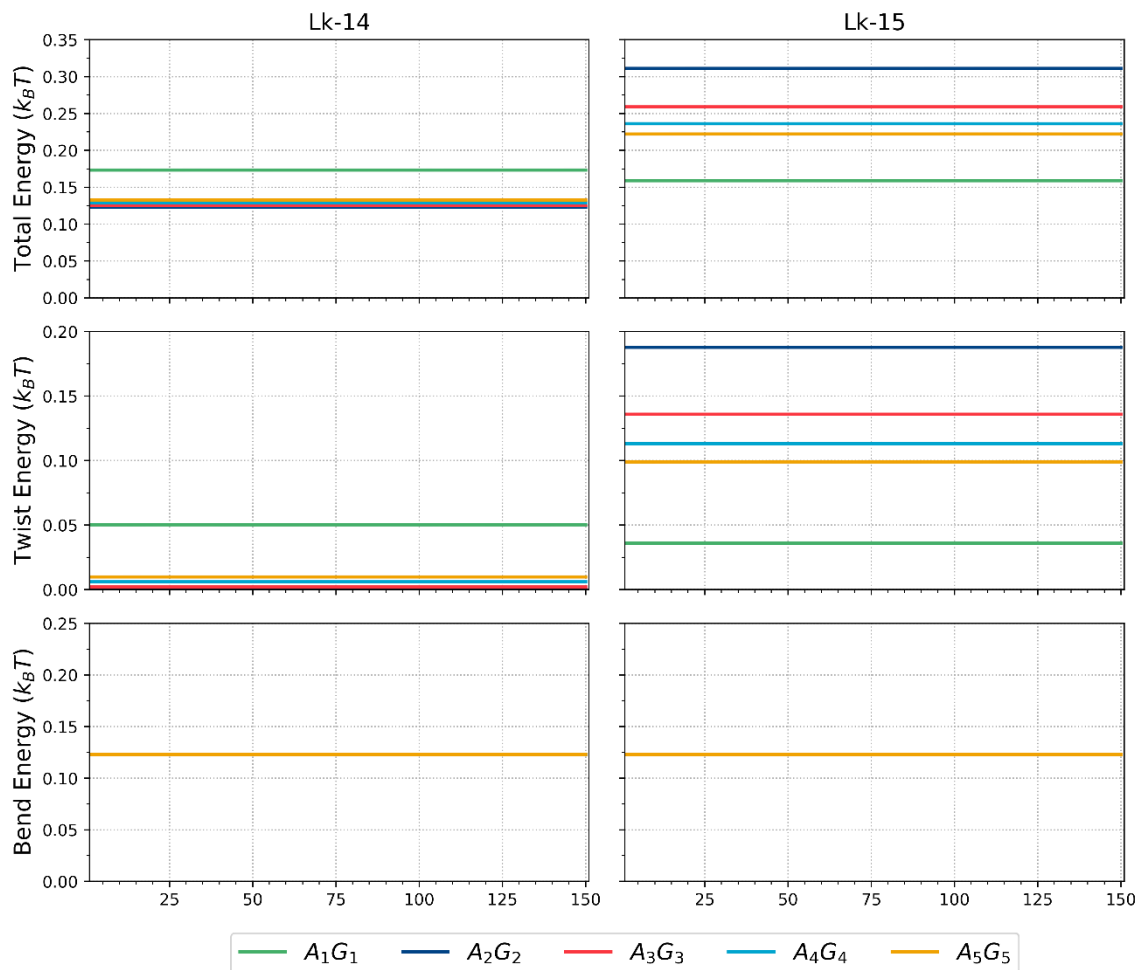


Figure A.IV. 4. The optimized elastic energy per base-pair step along the 150-bp minicircle sequence obtained using the Zhurkin roll-specific forcefield. The columns of this figure are organized based on initial circles (Lk 14 left, Lk 15 right) and the rows are organized by total energy (top), dimer-twist energy (middle), and dimer-bend energy (bottom) in units of  $k_B T$ . Use of a roll-specific forcefield resulted in a variance in dimeric energy which appears to increase as the number of homodimeric tracts increase. While periodic behavior can be seen for the  $A_1G_1$  (green),  $A_2G_2$  (dark blue) and  $A_3G_3$  (red) sequences in both topological states, the total energy of the  $A_4G_4$  (blue) sequences has a high region in the middle of the sequence and lower energy regions at the start. Additionally, the difference in  $A_5G_5$  (gold) energy values is striking, with the Lk 14 circle containing regions of high and low energy while the Lk 15 circle has the lowest total energy/step. The breakdown of the dimer twist and bend energies shows out-of-phase behavior of the energy profiles: local twist energy extrema coincide with bend energy extrema. The regions of low and high energy are tied to bend energy as these profiles show similar overall increases and decreases along the sequence. See Supplemental Table A.IV.1 for mean values and associated deviations.

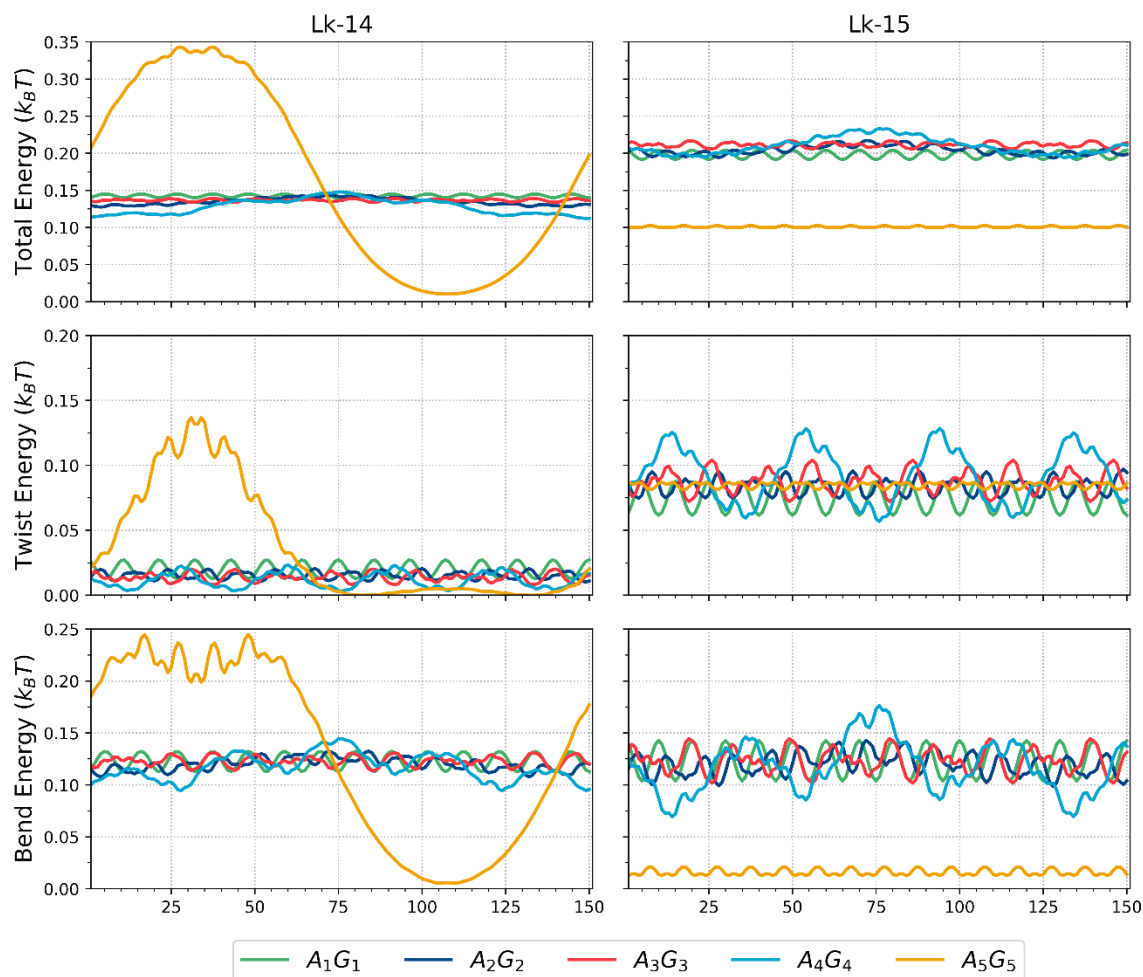


Table A.IV. 1 The mean elastic energy/step of the 150-bp minicircles optimized with the Zhurkin roll-specific forcefield. The table includes the total, twist, and bend energy/step with each associated standard deviation.

	Sequence	Total Energy/step ( $k_B T, \pm \sigma_E$ )	Twist Energy/step ( $k_B T, \pm \sigma_E$ )	Bend Energy/step ( $k_B T, \pm \sigma_E$ )
<b>Lk 14</b>	<b>A<sub>1</sub>G<sub>1</sub></b>	$0.143 \pm 0.002$	$0.020 \pm 0.005$	$0.123 \pm 0.007$
	<b>A<sub>2</sub>G<sub>2</sub></b>	$0.135 \pm 0.005$	$0.015 \pm 0.003$	$0.120 \pm 0.006$
	<b>A<sub>3</sub>G<sub>3</sub></b>	$0.137 \pm 0.001$	$0.014 \pm 0.003$	$0.123 \pm 0.005$
	<b>A<sub>4</sub>G<sub>4</sub></b>	$0.128 \pm 0.010$	$0.011 \pm 0.006$	$0.117 \pm 0.013$
	<b>A<sub>5</sub>G<sub>5</sub></b>	$0.168 \pm 0.124$	$0.036 \pm 0.045$	$0.131 \pm 0.087$
<b>Lk 15</b>	<b>A<sub>1</sub>G<sub>1</sub></b>	$0.198 \pm 0.005$	$0.074 \pm 0.009$	$0.124 \pm 0.014$
	<b>A<sub>2</sub>G<sub>2</sub></b>	$0.204 \pm 0.006$	$0.084 \pm 0.006$	$0.120 \pm 0.011$
	<b>A<sub>3</sub>G<sub>3</sub></b>	$0.211 \pm 0.003$	$0.088 \pm 0.009$	$0.123 \pm 0.012$
	<b>A<sub>4</sub>G<sub>4</sub></b>	$0.210 \pm 0.011$	$0.093 \pm 0.021$	$0.117 \pm 0.025$
	<b>A<sub>5</sub>G<sub>5</sub></b>	$0.101 \pm 0.001$	$0.085 \pm 0.002$	$0.016 \pm 0.003$

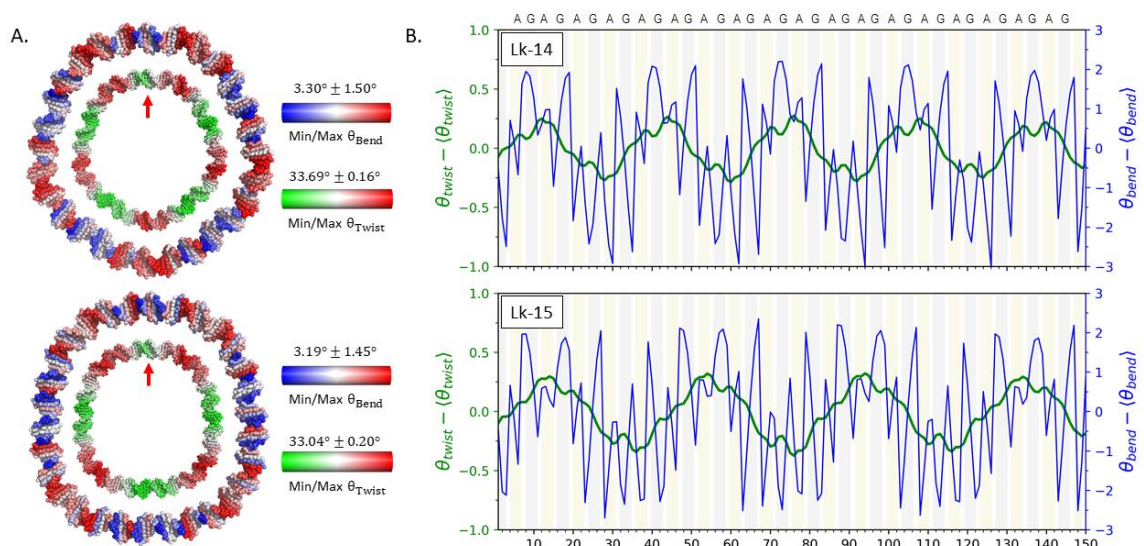


Figure A.IV. 6. Linear models of A<sub>4</sub>G<sub>4</sub> and A<sub>5</sub>G<sub>5</sub> sequences of length 150-bp optimized with the Zhurkin roll-specific forcefield. DNA base pairs are represented in block form with silver edges representative of minor groove faces. The A<sub>4</sub>G<sub>4</sub> linear model (blue) appears to have a regular spiral about the helical axis while the A<sub>5</sub>G<sub>5</sub> linear model (yellow) is a highly curved structure such that a 2-D projection would show a half-circular configuration. Both model fragments are arranged such that the central base pair has the same orientation in space while the views are either along the profile (top left), rotated 90° about the global x-axis (bottom left), and rotated 100° about vertical axis (right).

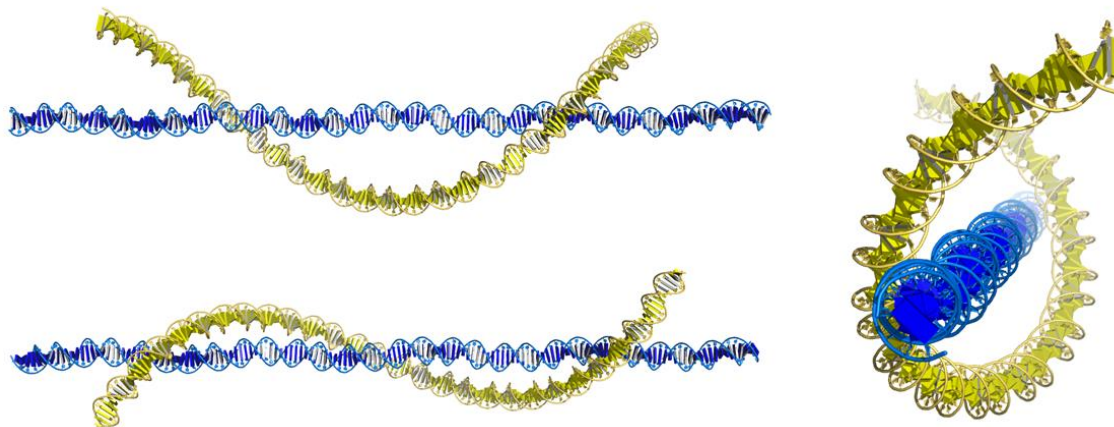


Figure A.IV. 7. Linear models of the  $A_4G_4$  (top) and  $A_5G_5$  sequences (bottom) of length 150-bp optimized with all three sequence-specific forcefields. For each sequence, the linear models are superimposed onto the first base pair. The linear models of the  $A_4G_4$  sequence (top) show relatively linear tracts for each with the Zhurkin-optimized (green) and combined Zhurkin/Trifonov-optimized (red) structures deviating from Trifonov-optimized (orange) trajectory. This is again observed with the  $A_5G_5$  sequence models (bottom) where the roll-influenced models show the circular path seen in Supplemental Figure A.IV.6 although the ends are in the same space as the  $A_4G_4$  models. Supplemental Table A.IV.2 shows the sequence alignment RMSD scores that relates each of these structural models.

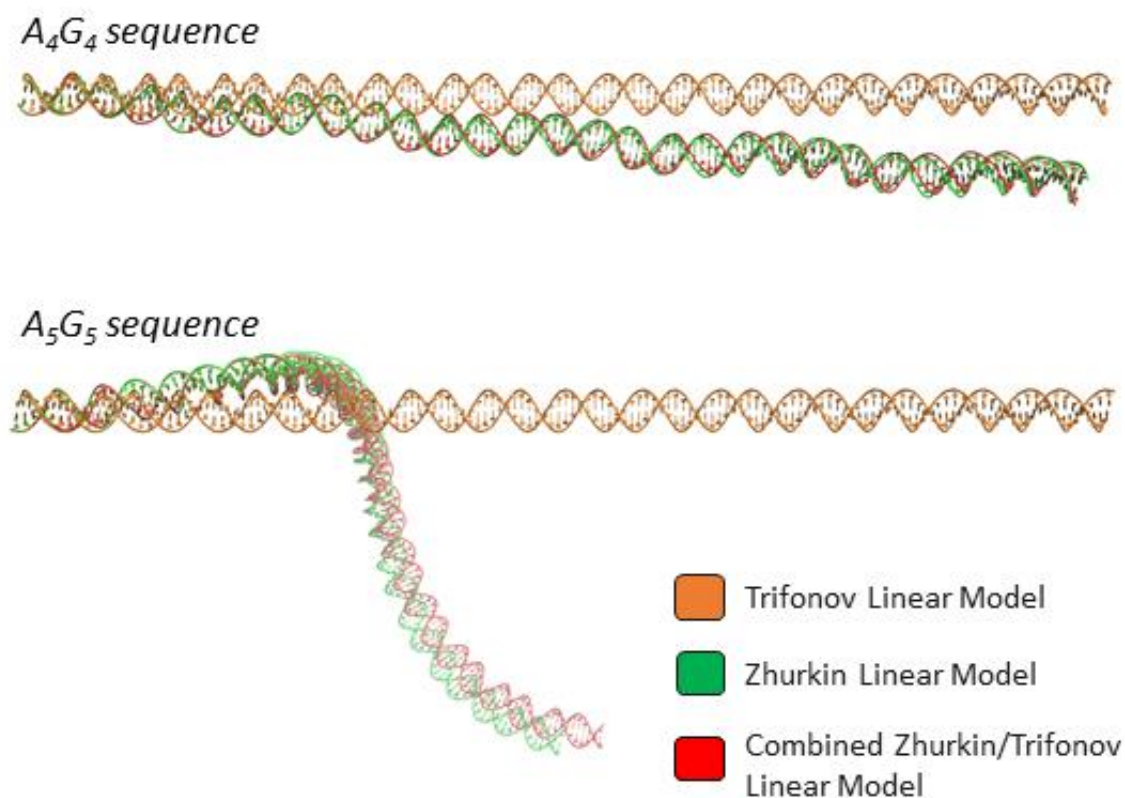


Table A.IV. 2. Sequence alignment scores\* of the A<sub>4</sub>G<sub>4</sub> and A<sub>5</sub>G<sub>5</sub> linear models seen in Supplemental Figure A.IV.7.

	Trifonov model	Zhurkin model	Zhurkin/Trifonov model
Trifonov model		3.11 Å	2.95 Å
Zhurkin model	56.56 Å		1.24 Å
Zhurkin/Trifonov model	54.61 Å	2.71 Å	

\* The RMSD values (in Å) are calculated from a pair of rest-state models used.

Cells in blue are associated with the A<sub>4</sub>G<sub>4</sub> linear models and cells in gold are for A<sub>5</sub>G<sub>5</sub> linear models



Figure A.IV. 8. The optimized elastic energy per base-pair step along the 150-bp minicircle sequence obtained using the combined Zhurkin/Trifonov forcefield. The columns of this figure are organized based on initial circles (Lk 14 left, Lk 15 right) and the rows are organized by total energy (top), dimer-twist energy (middle), and dimer-bend energy (bottom) in units of  $k_B T$ . As seen previously with the Zhurkin roll-specific optimization findings in Supplemental Figure A.IV.4, use of non-zero roll values results in variance of dimeric energy which, again, appears to increase as the number of homodimeric tracts increases. However, as this forcefield also contains sequence-specific twist values, a separation of total and twist energy values is observed and is somewhat reflective of the energy profiles seen previously with the Trifonov twist-specific optimization findings in Supplemental Figure A.IV.3. This combined model does bring into sharper relief the periodic behavior of the dimer energy, as seen with the Zhurkin optimization. However, while the pattern of high and low total energy/step regions is again observed with the  $A_4G_4$  sequence for both Lk 14 and Lk 15 circles, it is also seen with the  $A_2G_2$  sequence. This is further observed in the bend energy/step plots. See Supplemental Table A.IV.3 for mean values and associated deviations.

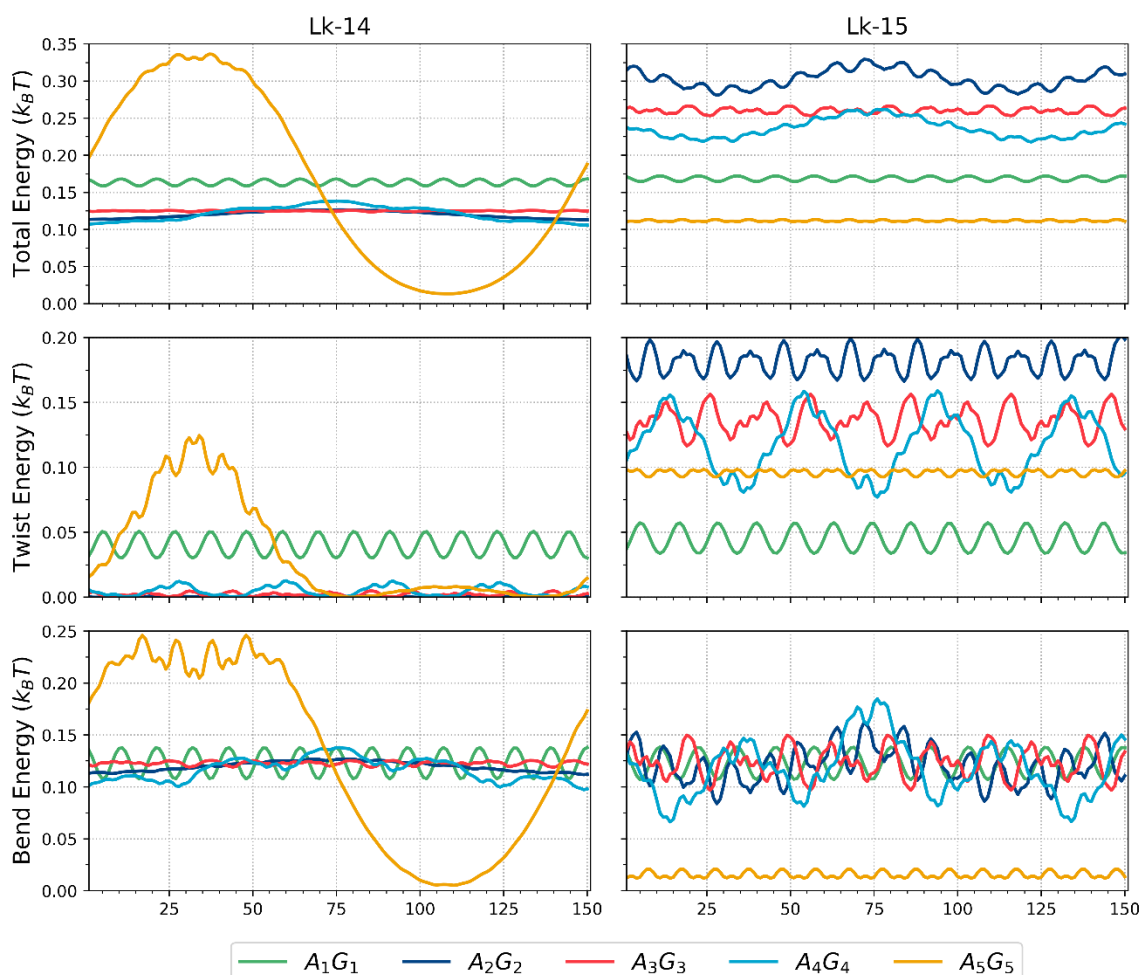


Table A.IV. 3. The mean elastic energy/step of the 150-bp minicircles optimized with the combined Zhurkin/Trifonov forcefield. The table includes the total, the twist, and the bend energy/step with each associated standard deviation.

		<b>Total Energy/step</b> <b>(<math>k_B T, \pm \sigma_E</math>)</b>	<b>Twist Energy/step</b> <b>(<math>k_B T, \pm \sigma_E</math>)</b>	<b>Bend Energy/step</b> <b>(<math>k_B T, \pm \sigma_E</math>)</b>
<b>Lk 14</b>	<b>A<sub>1</sub>G<sub>1</sub></b>	0.163 $\pm$ 0.003	0.040 $\pm$ 0.007	0.123 $\pm$ 0.011
	<b>A<sub>2</sub>G<sub>2</sub></b>	0.120 $\pm$ 0.005	0.0003 $\pm$ 0.0003	0.120 $\pm$ 0.005
	<b>A<sub>3</sub>G<sub>3</sub></b>	0.125 $\pm$ 0.001	0.002 $\pm$ 0.001	0.123 $\pm$ 0.002
	<b>A<sub>4</sub>G<sub>4</sub></b>	0.121 $\pm$ 0.010	0.005 $\pm$ 0.004	0.116 $\pm$ 0.011
	<b>A<sub>5</sub>G<sub>5</sub></b>	0.164 $\pm$ 0.120	0.032 $\pm$ 0.039	0.131 $\pm$ 0.088
<b>Lk 15</b>	<b>A<sub>1</sub>G<sub>1</sub></b>	0.169 $\pm$ 0.003	0.045 $\pm$ 0.008	0.123 $\pm$ 0.011
	<b>A<sub>2</sub>G<sub>2</sub></b>	0.303 $\pm$ 0.012	0.182 $\pm$ 0.010	0.121 $\pm$ 0.018
	<b>A<sub>3</sub>G<sub>3</sub></b>	0.260 $\pm$ 0.004	0.136 $\pm$ 0.011	0.124 $\pm$ 0.015
	<b>A<sub>4</sub>G<sub>4</sub></b>	0.236 $\pm$ 0.013	0.118 $\pm$ 0.024	0.117 $\pm$ 0.028
	<b>A<sub>5</sub>G<sub>5</sub></b>	0.112 $\pm$ 0.001	0.096 $\pm$ 0.002	0.016 $\pm$ 0.003

## BIBLIOGRAPHY

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Møller, H. D. *et al.* CRISPR-C: Circularization of genes and chromosome by CRISPR in human cells. *Nucleic Acids Research* **46**, (2018).
3. Franklin, R. E. & Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature* **171**, 740–741 (1953).
4. Braun, G., Tierney, D. & Schmitzer, H. How Rosalind Franklin discovered the helical structure of DNA: Experiments in diffraction. *The Physics Teacher* **49**, 140–143 (2011).
5. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
6. Klug, A. The Discovery of the DNA double helix. *Journal of Molecular Biology* **335**, 3–26 (2004).
7. Garcia, H. G. *et al.* Biological consequences of tightly bent DNA: The other life of a macromolecular celebrity. *Biopolymers* **85**, 115–130 (2007).
8. Lavelle, C. Pack, unpack, bend, twist, pull, push: The physical side of gene expression. *Current Opinion in Genetics and Development* **25**, 74–84 (2014).
9. Norouzi, D., Katebi, A., Cui, F. & B. Zhurkin, V. Topological diversity of chromatin fibers: Interplay between nucleosome repeat length, DNA linking number and the level of transcription. *AIMS Biophysics* **2**, 613–629 (2015).
10. Bowater, R. P. Supercoiled DNA: Structure. in *Encyclopedia of Life Sciences* (John Wiley & Sons, Ltd, 2005). doi:10.1038/npg.els.0003899.
11. Camerini-Otero, R. D. Homologous recombination proteins in prokaryotes and eukaryotes. *Annual Review of Genetics* **29**, 509–552 (1995).
12. Karymov, M. A. *et al.* Structure, dynamics, and branch migration of a DNA Holliday junction: A single-molecule fluorescence and modeling study. *Biophysical Journal* **95**, 4372–4383 (2008).
13. Franklin, R. E. & Gosling, R. G. The structure of sodium thymonucleate fibres. I. The influence of water content. *Acta Crystallographica* **6**, 673–677 (1953).
14. Marko, J. F. & Siggia, E. D. Statistical mechanics of supercoiled DNA. *Physical Review E* **52**, 2912–2938 (1995).
15. Ussery, D. W. DNA structure : A-, B- and Z-DNA helix families. *Life Sciences* 1–11 (2002) doi:10.1038/npg.els.0003122.
16. Lu, X.-J. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Research* **31**, 5108–5121 (2003).
17. Li, S., Olson, W. K. & Lu, X.-J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Research* **47**, W26–W34 (2019).
18. Zimmerman, S. B. The three-dimensional structure of DNA. *Annual Review of Biochemistry* **51**, 395–427 (1982).

19. Vargason, J. M., Henderson, K. & Ho, P. S. A crystallographic map of the transition from B-DNA to A-DNA. *Proceedings of the National Academy of Sciences* **98**, 7265–7270 (2001).
20. Srinivasan, A. R. *et al.* Properties of the nucleic-acid bases in free and Watson-Crick hydrogen-bonded states: Computational insights into the sequence-dependent features of double-helical DNA. *Biophysical Reviews* **1**, 13–20 (2009).
21. Portugal, J. & Subirana, J. A. Counterions which favour the C form of DNA. *The EMBO Journal* **4**, 2403–2408 (1985).
22. Hud, N. V. & Plavec, J. A unified model for the origin of DNA sequence-directed curvature. *Biopolymers* **69**, 144–158 (2003).
23. Pastor, N. The B- to A-DNA transition and the reorganization of solvent at the DNA surface. *Biophysical Journal* **88**, 3262–3275 (2005).
24. Roca, J. Transcriptional inhibition by DNA torsional stress. *Transcription* **2**, 82–85 (2011).
25. Reymer, A., Zakrzewska, K. & Lavery, R. Sequence-dependent response of DNA to torsional stress: A potential biological regulation mechanism. *Nucleic Acids Research* **46**, 1684–1694 (2018).
26. Vafabakhsh, R. & Ha, T. Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science* **337**, 1097–1101 (2012).
27. Zhu, J. *et al.* Molecular characterization of cell-free eccDNAs in human plasma. *Scientific Reports* **7**, 1–11 (2017).
28. Shoura, M. J. *et al.* Intricate and cell type-specific populations of endogenous circular DNA (eccDNA) in *Caenorhabditis elegans* and *Homo sapiens*. *G3: Genes/Genomes/Genetics* **7**, 3295–3303 (2017).
29. Kim, S.H. *et al.* Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* **185**, 435–440 (1974).
30. Møller, H. D. *et al.* Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nature Communications* **9**, 1–12 (2018).
31. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465 (1981).
32. Metzenberg, S. & Agabian, N. Mitochondrial minicircle DNA supports plasmid replication and maintenance in nuclei of *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences* **91**, 5962–5966 (1994).
33. Taanman, J. W. The mitochondrial genome: Structure, transcription, translation and replication. *Biochimica et Biophysica Acta - Bioenergetics* **1410**, 103–123 (1999).
34. Catanese, D. J., Fogg, J. M., Schrock, D. E., Gilbert, B. E. & Zechiedrich, L. Supercoiled minivector DNA resists shear forces associated with gene therapy delivery. *Gene Therapy* **19**, 94–100 (2012).
35. Irobalieva, R. N. *et al.* Structural diversity of supercoiled DNA. *Nature Communications* **6**, 8440 (2015).
36. Hardee, C. L., Arévalo-Soliz, L. M., Hornstein, B. D. & Zechiedrich, L. Advances in non-viral DNA vectors for gene therapy. *Genes* **8**, (2017).
37. Vologodskii, A. Exploiting circular DNA. *Proceedings of the National Academy of Sciences* **95**, 4092–4093 (1998).

38. Mirkin, S. M. DNA Topology: Fundamentals. in *Encyclopedia of Life Sciences* (John Wiley & Sons, Ltd, 2001). doi:10.1038/npg.els.0001038.
39. Vologodskii, A. *Biophysics of DNA*. (Cambridge University Press, 2015). doi:10.1017/CBO9781139542371.
40. Corless, S. & Gilbert, N. Effects of DNA supercoiling on chromatin architecture. *Biophysical Reviews* **8**, 51–64 (2016).
41. White, J. Self-linking and the Gauss integral in higher dimensions. *American Journal of Mathematics* **91**, 693–728 (1969).
42. Fuller, F. B. The writhing number of a space curve. *Proceedings of the National Academy of Sciences* **68**, 815–819 (1971).
43. Călugăreanu, G. Sur les classes d'isotopie des noeuds tridimensionnels et leurs invariants. *Czechoslovak Mathematicl Journal* **11**, 588–625 (1961).
44. Dennis, M. R. & Hannay, J. H. Geometry of Calugareanu's theorem. *Article submitted to Royal Society* 1–10 (2005) doi:10.1098/rspa.2005.1527.
45. Swigon, D. The mathematics of DNA structure, mechanics, and dynamics. in *The IMA Volumes in Mathematics and its Applications* vol. 150 293–320 (Springer, 2009). doi: 10.1007/978-1-4419-0670-0\_14.
46. Clauvelin, N. & Olson, W. K. The synergy between protein positioning and DNA elasticity: energy minimization of protein-decorated DNA minicircles. *The Journal of Physical Chemistry B* **125**, 2277–2287 (2021).
47. Bauer, W. R. Structure and Reactions of Closed Duplex DNA. *Annual Review of Biophysics and Bioengineering* **7**, 287–313 (1978).
48. Bauer, W. R., Lund, R. a & White, J. H. Twist and writhe of a DNA loop containing intrinsic bends. *Proceedings of the National Academy of Sciences* **90**, 833–837 (1993).
49. Shimada, J. & Yamakawa, H. Ring-closure probabilities for twisted wormlike chains. Application to DNA. *Macromolecules* **17**, 689–698 (1984).
50. Dill, K. & Bromberg, S. *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience*. (CRC Press, 2010). doi: 10.4324/9780203809075.
51. Wang, M. D., Yin, H., Landick, R., Gelles, J. & Block, S. M. Stretching DNA with optical tweezers. *Biophysical journal* **72**, 1335–46 (1997).
52. Moroz, J. D. & Nelson, P. Torsional directed walks, entropic elasticity, and DNA twist stiffness. *Proceedings of the National Academy of Sciences* **94**, 14418–14422 (1997).
53. Bustamante, C., Bryant, Z. & Smith, S. B. Ten years of tension: single-molecule DNA mechanics. *Nature* **421**, 423–427 (2003).
54. Nomidis, S. K. *et al.* Twist-bend coupling, twist waves, and the shape of DNA loops. *Physical Review E* **100**, (2019).
55. Marko, J. F. Biophysics of protein-DNA interactions and chromosome organization. *Physica A: Statistical Mechanics and its Applications* **418**, 126–153 (2015).
56. Marko, J. F. M. & Siggia, E. D. S. Bending and twisting elasticity of DNA. *Macromolecules* **27**, 981–988 (1994).
57. Olson, W. K. Simulating DNA at low resolution. *Current Opinion in Structural Biology* **6**, 242–256 (1996).

58. Balaeff, A., Mahadevan, L. & Schulten, K. Modeling DNA loops using the theory of elasticity. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **73**, 1–23 (2006).
59. Nomidis, S. K., Kriegel, F., Vanderlinden, W., Lipfert, J. & Carlon, E. Twist-bend coupling and the torsional response of double-stranded DNA. *Physical Review Letters* **118**, 1–6 (2017).
60. Olson, W. K., Swigon, D. & Coleman, B. D. Implications of the dependence of the elastic properties of DNA on nucleotide sequence. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **362**, 1403–1422 (2004).
61. Perez, P. J. & Olson, W. K. Insights into genome architecture deduced from the properties of short Lac repressor-mediated DNA loops. *Biophysical Reviews* **8**, 135–144 (2016).
62. Skoruppa, E., Laleman, M., Nomidis, S. K. & Carlon, E. DNA elasticity from coarse-grained simulations: The effect of groove asymmetry. *Journal of Chemical Physics* **146**, (2017).
63. Shore, D., Langowski, J. & Baldwin, R. L. DNA flexibility studied by covalent closure of short fragments into circles. *Proceedings of the National Academy of Sciences* **78**, 4833–4837 (1981).
64. Cloutier, T. E. & Widom, J. Spontaneous sharp bending of double-stranded DNA. *Molecular Cell* **14**, 355–362 (2004).
65. Czapla, L., Swigon, D. & Olson, W. K. Sequence-dependent effects in the cyclization of short DNA. *Journal of Chemical Theory and Computation* **2**, 685–695 (2006).
66. Skoruppa, E., Nomidis, S. K., Marko, J. F. & Carlon, E. Bend-induced twist waves and the structure of nucleosomal DNA. *Physical Review Letters* **121**, 088101 (2018).
67. Olson, W. K., Colasanti, A. V., Czapla, L. & Zheng, G. Insights into the sequence-dependent macromolecular properties of DNA from base-pair level. in *Coarse-Graining of Condensed Phase and Biomolecular Systems* (ed. Voth, G. A.) 205–223 (CRC Press, 2008). doi:10.1201/9781420059564.
68. Dickerson, R. E. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Research* **17**, 1797–1803 (1989).
69. El Hassan, M. A. & Calladine, C. R. The assessment of the geometry of dinucleotide steps in double-helical DNA; a new local calculation scheme. *Journal of Molecular Biology* **251**, 648–664 (1995).
70. Beveridge, D. L. *et al.* Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophysical Journal* **87**, 3799–3813 (2004).
71. Dixit, S. B. *et al.* Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: Sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophysical Journal* **89**, 3721–3740 (2005).
72. Drew, H. R. *et al.* Structure of a B-DNA dodecamer: conformation and dynamics. *Proceedings of the National Academy of Sciences* **78**, 2179–2183 (1981).

73. Peck, L. J. & Wang, J. C. Sequence dependence of the helical repeat of DNA in solution. *Nature* **292**, 375–378 (1981).
74. Koo, H.-S., Wu, H.-M. & Crothers, D. M. DNA bending at adenine · thymine tracts. *Nature* **320**, 501–506 (1986).
75. Trifonov, E. N. & Sussman, J. L. The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proceedings of the National Academy of Sciences* **77**, 3816–3820 (1980).
76. Ulanovsky, L. E. & Trifonov, E. N. Estimation of wedge components in curved DNA. *Nature* **326**, 720–722 (1987).
77. 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**, 4227–4234 (1990).
78. Goodsell, D. S. & Dickerson, R. E. Bending and curvature calculations in B-DNA. *Nucleic Acids Research* **22**, 5497–5503 (1994).
79. Barbic, A., Zimmer, D. P. & Crothers, D. M. Structural origins of adenine-tract bending. *Proceedings of the National Academy of Sciences* **100**, 2369–2373 (2003).
80. 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. *Proceedings of the National Academy of Sciences* **95**, 11163–11168 (1998).
81. Gorin, A. A., Zhurkin, V. B. & Olson, W. K. B-DNA twisting correlates with base-pair morphology. *Journal of Molecular Biology* **247**, 34–48 (1995).
82. 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 Conformation and Transcription* 18–34 (2005).
83. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28**, 235–242 (2000).
84. Burley, S. K. *et al.* RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research* **49**, D437–D451 (2021).
85. Berman, H. M. *et al.* The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophysical Journal* **63**, 751–759 (1992).
86. Coleman, B. D., Olson, W. K. & Swigon, D. Theory of sequence-dependent DNA elasticity. *The Journal of Chemical Physics* **118**, 7127–7140 (2003).
87. Colasanti, A. V. *et al.* Weak operator binding enhances simulated lac repressor-mediated DNA looping. *Biopolymers* **99**, 1070–1081 (2013).
88. Schlick, T., Olson, W. K., Westcott, T. & Greenberg, J. P. On higher buckling transitions in supercoiled DNA. *Biopolymers* **34**, 565–597 (1994).
89. Vologodskii, A. Disentangling DNA molecules. *Physics of Life Reviews* **18**, 118–134 (2016).
90. Lu, X.-J. & Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nature protocols* **3**, 1213–27 (2008).

91. Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. The structure of an oligo(dA)oligo(dT) tract and its biological implications. *Nature* **330**, 221–226 (1987).
92. Barbič, A., Zimmer, D. P. & Crothers, D. M. Structural origins of adenine-tract bending. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 2369–2373 (2003).
93. Zhurkin, V. B., Lysov, Y. P. & Ivanov, V. I. Anisotropic flexibility of DNA and the nucleosomal structure. *Nucleic Acids Research* **6**, 1081–1096 (1979).
94. Tolstorukov, M. Y., Virnik, K. M., Adhya, S. & Zhurkin, V. B. A-tract clusters may facilitate DNA packaging in bacterial nucleoid. *Nucleic Acids Research* **33**, 3907–3918 (2005).
95. Kabsch, W., Sander, C. & Trifonov, E. N. The ten helical twist angles of B-DNA. *Nucleic Acids Research* **10**, 1097–1104 (1982).
96. *Mathematics of DNA Structure, Function and Interactions*. (Springer, 2009).
97. Olson, W. K., Marky, N. L., Jernigan, R. L. & Zhurkin, V. B. Influence of fluctuations on DNA curvature. *Journal of Molecular Biology* **232**, 530–554 (1993).
98. Horowitz, D. S. & Wang, J. C. Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling. *Journal of Molecular Biology* **173**, 75–91 (1984).
99. Heath, P. J., Clendenning, J. B., Fujimoto, B. S. & Schurr, M. J. Effect of Bending Strain on the Torsion Elastic Constant of DNA. *Journal of Molecular Biology* **260**, 718–730 (1996).
100. Miyazawa, T. Molecular vibrations and structure of high polymers. II. Helical parameters of infinite polymer chains as functions of bond lengths, bond angles, and internal rotation angles. *Journal of Polymer Science* **55**, 215–231 (1961).
101. Ho, P. S. & Carter, M. DNA structure: Alphabet soup for the cellular soul. in *DNA Replication- Current Advances* (ed. Seligmann, H.) (InTech, 2011). doi:10.5772/52807.
102. The PyMOL Molecular Graphics System, Version 2.4.
103. Rose, I. A., Hanson, K. R., Wilkinson, K. D. & Wimmer, M. J. A suggestion for naming faces of ring compounds. *Proceedings of the National Academy of Sciences* **77**, 2439–2441 (1980).
104. Liljas, A. *et al. Textbook of Structural Biology. Textbook of Structural Biology* (World Scientific, 2009). doi:10.1142/6620.
105. Leontis, N. B. & Westof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512 (2001).
106. Lavery, R., Zakrzewska, K., Sun, J. sheng & Harvey, S. C. A comprehensive classification of nucleic acid structural families based on strand direction and base pairing. *Nucleic Acids Research* **20**, 5011–5016 (1992).
107. Hoogsteen, K. The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Crystallographica* **12**, 822–823 (1959).
108. Felsenfeld, G., Davies, D. R. & Rich, A. Formation of a three-stranded polynucleotide molecule. *Journal of the American Chemical Society* **79**, 2023–2024 (1957).



109. Vasquez, K. M. & Glazer, P. M. Triplex-forming oligonucleotides: principles and applications. *Quarterly Reviews of Biophysics* **35**, 89–107 (2002).
110. Nikolova, E. N. *et al.* A historical account of Hoogsteen base-pairs in duplex DNA. *Biopolymers* **99**, 955–968 (2013).
111. Xu, Y., McSally, J., Andricioaei, I. & Al-Hashimi, H. M. Modulation of Hoogsteen dynamics on DNA recognition. *Nature Communications* **9**, 1473 (2018).
112. Yakovchuk, P. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research* **34**, 564–574 (2006).
113. Feng, B. *et al.* Hydrophobic catalysis and a potential biological role of DNA unstacking induced by environment effects. *Proceedings of the National Academy of Sciences* **116**, 17169–17174 (2019).
114. Olson, W. K. *et al.* A standard reference frame for the description of nucleic acid base-pair geometry. *Journal of Molecular Biology* **313**, 229–237 (2001).
115. Lu, Y., Weers, B. & Stellwagen, N. C. DNA persistence length revisited. *Biopolymers* **61**, 261–275 (2001).
116. Hagerman, P. J. Flexibility of DNA. *Annual Review of Biophysics and Biophysical Chemistry* **17**, 265–286 (1988).
117. Clauvelin, N. & Olson, W. K. Synergy between protein positioning and DNA elasticity: Energy minimization of protein-decorated DNA minicircles. *The Journal of Physical Chemistry B* **125**, 2277–2287 (2021).

ProQuest Number: 28772682

INFORMATION TO ALL USERS

The quality and completeness of this reproduction is dependent on the quality and completeness of the copy made available to ProQuest.



Distributed by ProQuest LLC (2022).

Copyright of the Dissertation is held by the Author unless otherwise noted.

This work may be used in accordance with the terms of the Creative Commons license or other rights statement, as indicated in the copyright statement or in the metadata associated with this work. Unless otherwise specified in the copyright statement or the metadata, all rights are reserved by the copyright holder.

This work is protected against unauthorized copying under Title 17,  
United States Code and other applicable copyright laws.

Microform Edition where available © ProQuest LLC. No reproduction or digitization of the Microform Edition is authorized without permission of ProQuest LLC.

ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 - 1346 USA