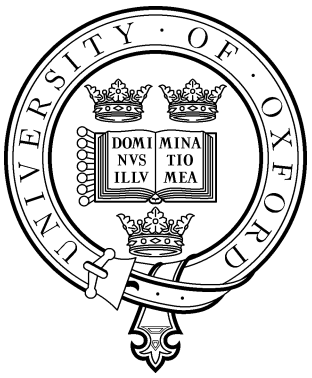


Molecular biophysics of strong DNA bending and the RecQ DNA helicase

Ryan M Harrison

Wolfson College
University of Oxford

Supervisor: Prof. Jonathan P. K. Doye

A thesis submitted for the degree of
Doctor of Philosophy in Physical & Theoretical Chemistry

Trinity 2014

Abstract

Molecular biophysics of strong DNA bending and the RecQ DNA helicase
Ryan M Harrison
Wolfson College, University of Oxford
Submitted for the degree of Doctor of Philosophy in Physical & Theoretical Chemistry
Trinity 2014

Molecular biophysics is a rapidly evolving field aimed at the physics-based investigation of the biomolecular processes that enable life. In this thesis, we explore two such processes: the thermodynamics of DNA bending, and the mechanism of the RecQ DNA helicase. A computational approach using a coarse-grained model of DNA is employed for the former; an experimental approach relying heavily on single-molecule fluorescence for the latter.

There is much interest in understanding the physics of DNA bending, due to both its biological role in genome regulation and its relevance to nanotechnology. Small DNA bending fluctuations are well described by existing models; however, there is less consensus on what happens at larger bending fluctuations. A coarse-grained simulation is used to fully characterize the thermodynamics and mechanics of duplex DNA bending. We then use this newfound insight to harmonize experimental results between four distinct experimental systems: a 'molecular vise', DNA cyclization, DNA minicircles and a 'strained duplex'. We find that a specific structural defect present at large bending fluctuations, a 'kink', is responsible for the deviation from existing theory at lengths below about 80 base pairs.

The RecQ DNA helicase is also of much biological and clinical interest, owing to its essential role in genome integrity via replication, recombination and repair. In humans, heritable defects in the RecQ helicases manifest clinically as premature aging and a greatly elevated cancer risk, in disorders such as Werner and Bloom syndromes. Unfortunately, the mechanism by which the RecQ helicase processes DNA remains poorly understood. Although several models have been proposed to describe the mechanics of helicases based on biochemical and structural data, ensemble experiments have been unable to address some of the more nuanced questions of helicase function. We prepare novel substrates to probe the mechanism of the RecQ helicase via single-molecule fluorescence, exploring DNA binding, translocation and unwinding. Using this insight, we propose a model for RecQ helicase activity.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Symbols

A Helmholtz free energy

aa amino acid residue(s)

A Absorbance, *e.g.* A_{280} denotes absorbance at 280 nm

a.u. arbitrary unit(s)

bp base pair(s)

B_s Bending modulus

δt Simulation time step

E Potential energy or Förster resonance energy transfer (FRET) efficiency

ε Extinction coefficient

G Gibbs free energy

H Enthalpy

I Fluorescence intensity

I_A Fluorescence intensity, acceptor

I_D Fluorescence intensity, donor

J Florescence spectral overlap

j_{dyn} Kinetic j-factor

j_{eq} Equilibrium j-factor, first introduced by Jacobson & Stockmayer[1]

k Kinetic rate constant

κ Harmonic spring constant (Part I). Orientation factor for FRET (Part II)

k_B Boltzman Constant

K_d Dissociation constant

K_{eq} Equilibrium constant

λ Wavelength

Lk Linking number

L_P Persistence length

l Segment length

l_s	Segment length, single-stranded DNA	R_0	Förster distance
n	Refractive index or Hill coefficient	$R_{contour}$	Contour length of duplex DNA
N_A	Avogadro's number	R_{ee}	End-to-end distance
N_{bp}	Total length in base pairs	S	Stoichiometry, fluorescence
N_d	Duplex length in base pairs	σ	Superhelical density
N_s	Single-stranded length in nucleotides	T	Temperature
nt	nucleotide	t	Time
φ	Quantum yield of flourphore	τ	Characteristic decorrelation time
Q_{bp}	Order parameter, number of base pairs formed	τ_c	Critical torque
		θ	Angle, usually bending angle
Q_{ee}	Order parameter, end-to-end distance	Tw	Twist number
r	Fluorescence anisotropy	Wr	Writhe number

List of Abbreviations

ADP adenosine diphosphate

AFM atomic force microscopy

ALEX alternating laser excitation

AMP-PNP adenylylimidodiphosphate

ATP adenosine triphosphate

BCA bicinchoninic acid assay

BER base excision repair

BLM Bloom syndrome protein

BME β-mercaptoethanol

BSA bovine serum albumin

cryo-EM cryo-electron microscopy

dsDNA double-stranded DNA

DTNB 5,5′-dithiobis-(2-nitrobenzoic acid, also known as Ellman's reagent,

DTT dithiothreitol

EDTA ethylenediaminetetraacetic acid

EED end-to-end distance

EM-CCD electron multiplying charge coupled device

ESI-MS electrospray ionization mass spectrometry

ESI-MS/MS electrospray tandem mass spectrometry

EtBr ethidium bromide

FENE finitely extensible nonlinear elastic

FPLC fast protein liquid chromatography

FRET Förster resonance energy transfer

HPLC high-performance liquid chromatography

HRDC helicase and RNaseD C-terminal

IMAC ion affinity chromatography

ITPG isopropyl β-D-1-thiogalacto pyranoside

KWLC kinkable worm-like chain

MC Monte Carlo

MD molecular dynamics

MMR methyl-directed mismatch repair

NADH nicotinamide adenine dinucleotide

NER nucleotide excision repair

NHEJ non-homologous end-joining

NTA nitrilotriacetic acid

NTP nucleoside triphosphate

O/N overnight

PAR 4-(2-pyridylazo)resorcinol

PBS phosphate buffered saline

PDB Protein Data Bank

PEG polyethylene glycol

PIFE protein induced fluorescent enhancement

POT1 telomere end binding protein

PSF point spread function

QY quantum yield

R/T room temperature

RECQ4 RecQ protein-like 4

RecQwt full-length wild-type RecQ

RPA replication initiator protein A

RQC RecQ C-terminal

SD standard deviation

SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEM standard error of the mean

SF superfamily

SNR signal-to-noise ratio

SSB single stranded binding protein

ssDNA single-stranded DNA

TCEP tris(2-chloroethyl) phosphate

TEV Tobacco Etch Virus nuclear inclusion a endopeptidase

TGase transglutaminase

TIR total internal reflection

TIRFM total internal reflection fluorescence microscopy

TopoIII topoisomerase III

VMMC virtual-move Monte Carlo

WH winged-helix

WHAM weighted histogram analysis method

WLC worm-like chain

WRN Werner syndrome RecQ helicase-like protein

Acknowledgements

This work would not have been possible without ample personal and professional support. Prof. Jonathan Doye, my advisor for the computational work (Part I), took a risk in accepting an experimentalist, with undergraduate training in an unrelated 'engineering' (biomedical) discipline, little formal knowledge of theory and uncertain funding. Fortunately, the collaboration has been a fruitful one, and I thoroughly enjoyed my time under his tutelage. I am also grateful for the mentorship of Dr Flavio Romano, without whose initial day-to-day supervision, and subsequent guidance, such rapid progress in transitioning from bench to theory would not have been possible.

Dr Keir Neuman, my advisor for the experimental work (Part II), which was performed primarily at the National Institutes of Health, will always have my respect for his sincerity and openness, and my admiration for his uncanny ability to intuit biological mechanisms. I'll fondly remember Dr Susanta Sarkar, a senior postdoc and mentor, who through his patience and dedication, taught me most everything I know about the art of single-molecule fluorescence.

In addition to this cadre of advisors, I've received bountiful support at both the National Institutes of Health and in Oxford. For the computational work:

- Dr Thomas Ouldridge, benevolent dictator (for life?) of the oxDNA model, who has taught me to question my own reflexive experimentalist hand waving.
- Prof. Ard Louis, a co-advisor and very close collaborator of Prof. Jon Doye, who has on many occasions helped me appreciate the 'big picture'.
- Drs John Schreck, Majid Mosayebi and Dennis Palagin. Flushing out experimentalist hand waving is a team effort, and what thirsty work.
- Dr Aleks Reinhardt for raising my awareness of digital typesetting and LaTeXsupport.

- Drs Flavio Romano, Petr Šulc, Lorenzo Rovigatti and Mr Benedict Snodin for development of the oxDNA simulation software, enabling the computational work.
- Miss Esther Miller for the kindest possible introduction to Britishness.

And for the experimental work:

- Dr Jim Sellers for guidance on the transition from experiment to theory.
- Dr Yasuharu (Harry) Takagi for personal and professional guidance.
- Dr Achillefs Kapanidis, who hosted me in his lab during Hilary 2012.
- Drs Timothy Craggs and Johannes Hohlbein for training in the ALEX technique.
- Dr Marie-Paule Strub, for training in protein expression and purification.

This work would not have been possible without the generous financial support of:

- National Science Foundation Graduate Research Fellowship
- National Heart, Lung, and Blood Institute; in particular Dr Robert Balaban, who provided the fourth year of funding required to complete the work on Part I.
- NIH-Oxford/Cambridge Scholar program
- NIH Office of Intramural Training & Education and Wolfson College, University of Oxford, for supporting conference travel.

On a more personal note, I would like to recognize Mrs Charlotte Saylor who singularly inspired and enlivened my nascent interest in science, those many moons ago. In gratitude I remember Prof. Jeffrey Gray, Dr Michael Daily (then Mr), Dr Kazunori Sugiyasu, Prof. Wendy Thomas, and Prof. Joel Bader, who each, in their own masterful way, helped forge a bushy-tailed pupil into something more worthy. Finally, I would like to applaud my family, in particular my parents, Robert & Sharon Harrison, and my grandmother, to whom this thesis is dedicated, for their unwaivering support. University, let alone a research doctorate, would have been untenable without their steady personal guidance, and the occasional (usually gentle) kick.

For my grandmother, Minnie Bryant

1 Preface

Molecular biophysics is an interdisciplinary approach to addressing biomolecular questions. In this thesis, we will explore two distinct classes of biomolecules, DNA in Part I and proteins, specifically DNA helicases, in Part II. The approaches used span the gamut of modern biophysical techniques, bridging both simulation and experiment.

In Part I, we explore DNA bending using a coarse-grained simulation of DNA. First, we characterize the thermodynamic and structural properties of DNA bending (Chapter 4). Next, we apply our newfound understanding to four distinct experimental model systems, harmonizing often apparently conflicting experimental results for a 'molecular vise' (Chapter 5), DNA cyclization (Chapter 6), DNA minicircles (Chapter 7) and a 'strained duplex' (Chapter 8). Taken together, this part highlights how simulation is a valuable tool for helping to elucidate experiment, directly observing states that may only be indirectly probed via experiment, and explaining a great diversity of observed phenomena in terms of a few physical principles.

In Part II, we probe the behavior of the RecQ DNA helicase via experiment. First, we develop novel substrates to investigate the helicase in both ensemble and single-molecule assays (Chapter 11). Next, we apply our substrates to resolve key questions about the structure and function of RecQ, ultimately proposing a mechanism for RecQ DNA helicase activity (Chapter 12).

The applications and techniques employed in the separate parts are quite distinct, highlighting the breadth of biophysics-based inquiry.

Part I Coarse-grained modelling of strong DNA bending

2 Introduction

Strong bending of double-stranded DNA (dsDNA) plays an important biological role in genome maintenance and regulation, most famously evident in DNA-protein complexes such as those formed in conjunction with the Lac repressor,[2] the TATA binding protein[3] and the nucleosome.[4,5] In worm-like chain (WLC) models, large bending fluctuations have concomitantly large free-energy penalties to formation; however, large bending fluctuations, that are not an artifact of DNA-protein co-crystallization, have been observed in a variety of sensitive solution-phase assays.[6,7]

Although WLC models[8–11] are sufficient to describe the thermodynamics of DNA bending found in many experiments,[12,13] deviations from WLC behavior have been reported via cyclization,[14–16] multimerization,[17] DNA minicircles,[18] a stressed ring system[19–21] and a molecular vise,[22] calling for further investigation of the strong bending regime.

The presence of structural defects may resolve the mystery of non-WLC behavior. For example. a structural defect that may appear under conditions of high stress is a 'kink',[23–25] whereby stacking and/or base pairs are disrupted in a localized region. Kinking would in principle allow excess twist and bend to localize in the disrupted region, thereby allowing the remainder of the molecule to explore highly bent and/or twisted states at a lower free-energy cost. Although there is strong experimental evidence for kinking (reviewed in Vologodskii *et al.*[26]), as of yet there is no consensus on the precise nature and characteristic onset of kinking: many claims remain controversial. For example, in DNA cyclization assays, cyclized molecules at $\sim$100 base pairs (bp)[14,15] were reported to be orders of magnitude more likely than predicted by a WLC model.[8] However, these results have been disputed, with others suggesting that non-WLC behavior only begins at shorter lengths.[13,16] Similarly, apparent evidence of high flexibility of DNA at short length scales obtained via atomic force microscopy (AFM)[27] has since been

attributed to artefacts of the curve-fitting algorithms.[28]

Given the uncertainty surrounding kinking, and the range of systems with which kinking has been associated, it is important to develop a sound theoretical underpinning for such a structural transition. Doing so would enable comparison between distinct systems, to establish whether results were consistent; for example, can a generally applicable free-energy cost of kinking be found? Further, if kinking is attributable to the disruption of base-pairing and stacking in the duplex, the observed behaviour should be explicable in terms of the basic thermodynamic and mechanical properties of DNA – establishing whether this is indeed true would be a substantial step in understanding the nature of kinking.

In this chapter, we outline the principles of DNA structural deformations, such as kinking, outline a few essential WLC models and highlight our specific aims for Part I.

2.1 DNA structural deformation

2.1.1 Continuous bending

In the continuously bent regime, bending energy is relatively uniformly distributed along the chain. WLC models have proven themselves successful in describing the free-energy landscape of continuously bent DNA. Modified WLC models that deviate from a harmonic potential at large bend angles have also been proposed.[29] Models of anharmonic bending are non-quadratic for large bending angles, softening the potential. The non-quadratic function used by Mazur *et al.*[29] was dependent on both sequence and applied torque; however, the structural basis for an anharmonic potential could be attributable to a variety of sources, and remains somewhat unclear.

2.1.2 Kinking

A kink is an area of extreme DNA curvature localized to a small segment instead of being continuously distributed along the chain. Kinks are thought to play an important role in DNA flexibility at length scales below the persistence length. The concept of DNA kinking is not a new idea, being first proposed by Crick and Klug in 1975.[23]

Non-WLC localized bending at a defect occurs when bending energy is concentrated at a structural defect, the 'kink'. Kinking allows the system to relieve bending energy in the duplex at the cost of disrupting hydrogen bonding and stacking at the defect. Since the free-energy penalty of kinking may be approximated as fixed by the cost of disrupting hydrogen bonding and stacking, while the free-energy cost of continuous bending grows rapidly as the curvature increases, a transition between continuous bending and kinking may be expected as the bending stress becomes more extreme. To address the potential for kinking at large bending stress, kinkable worm-like chain (KWLC) models[24,26,30] have been developed. These are briefly explored in Section 2.2.1.

Structural defects, such as kinking, do not have a standard quantitative definition. Some define the strong bending regime structurally, seeking explanations in terms of defects such as kinks and bubbles,[19] or in the subtleties of anisotropic DNA bending,[31,32] characteristic curvature,[26] or nonstandard stacking arrangements.[25] The effect of the introduction of a 'bulge' site, where one of the two strands is longer and has unpaired bases in an otherwise complementary duplex, has also been explored, with kinking observed for 120 bp duplexes with a 3–8 nucleotide (nt) bulge.[33]

2.2 Worm-like chain (WLC) models

WLC models, which represent DNA bending, and somtimes twisting, as continuously distributed along the DNA chain, accurately describe the free-energy landscape for small bending fluctuations;[9–11] however, deviations have been observed for large bending fluctuations in several model systems. The WLC model has been widely used to model DNA flexibility at small bending fluctuations since 1972.[34,35] It has proven itself valuable in accurately describing many DNA behaviors, including force-extension of dsDNA, linking number variation, and various mechanical parameters such as persistence length.[8–11]

The discrete WLC model takes DNA as a set of nodes, usually representing individual base pairs, connected by harmonic springs (Figure 2.1). The angle $\theta_{i,i+1}$ between two adjacent nodes determines the magnitude of the harmonic potential. In the absence of intrinsic curvature, deviations of $\theta_{i,i+1}$ from zero increase the energy quadratically. Both small and large bending

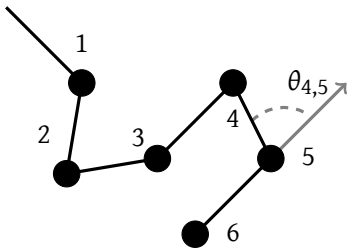


Figure 2.1 Diagram of a worm-like chain (WLC) model
Representation of a discrete WLC model. Each node experiences a quadratic bending potential in angle θ.

perturbations are described by

$$E_i = \frac{1}{2}\kappa\theta_{i,i+1}^2 \tag{2.1}$$

$$E_{\mathrm{WLC}} = \sum_{i=1}^{n} E_i, \tag{2.2}$$

where κ is the harmonic spring constant.

An exponential decay is observed in the decorrelation of $\theta_{i,i+\ell}$ along the chain, with a characteristic length scale. This persistence length is defined as

$$\langle \cos\theta_{i,i+\ell} \rangle = \exp\left(-\frac{\ell}{L_{\mathrm{P}}}\right), \tag{2.3}$$

where $\langle \cos\theta_{i,i+\ell} \rangle$ is the ensemble average value of $\cos\theta_{i,i+\ell}$, ℓ is the segment length of interest, and L_{P} is the persistence length (the length scale over which the polymer becomes flexible). Note that, assuming WLC behavior, the spring constant κ from Equation (2.2) can be determined empirically via fitting persistence length.

Alternatively, $\cos\theta$, and thereby the persistence length, may be described in terms of orientation vectors, where the total angle $\theta_{i,j}$ between the two orientation vectors would be

$$\cos\theta_{i,j} = \frac{\boldsymbol{a}_i \cdot \boldsymbol{a}_j}{\|\boldsymbol{a}_i\|\|\boldsymbol{a}_j\|} = \hat{\boldsymbol{a}}_i \cdot \hat{\boldsymbol{a}}_j, \tag{2.4}$$

where $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$ are the base pair orientation of base pair indices i and j respectively.

Note that polymers are often characterized in terms of the bending modulus B_S, related to L_P via

$$L_P = \frac{B_S}{k_B T}. \tag{2.5}$$

For dsDNA, the persistence length has been characterized experimentally by a variety of methods, generally yielding values of ∼40–50 nm (120–150 bp) at physiological salt conditions. While the WLC model is an excellent descriptor of long duplex DNA (>L_P) with small to moderate bending, DNA flexibility has been observed to diverge from that predicted by the WLC model at large bending.[13–16,18,22,36] Several models have been put forward to explain this discrepancy, all centering around the role of either localized DNA structural deformations or anharmonic bending. Later, we will show that a WLC model works well for a 30 bp duplex for fluctuations below $\approx 18\, k_B T$ (Chapter 4).

2.2.1 Kinkable worm-like chain (KWLC) model

The WLC model has been extended to account for the presence of localized bending in the DNA duplex,[24,26,30,37] a kink (Section 2.1.2). A simple functional form for this kinkable worm-like chain (KWLC) is given by[26]

$$E_i = \min\left(\frac{1}{2}\kappa\theta_{i,i+1}^2, k_B T\left(\theta_{i,i+1} - b\right)^6 + h\right), \tag{2.6}$$

where akin to Equation (2.2), κ is the harmonic spring constant and $\theta_{i,i+1}$ the bend angle. Additionally, h is the enthalpic energy barrier of kink formation, and b specifies the typical bending angle at the kink.

In conjunction with an unlooping-based FRET assay,[16] the KWLC expression has been used to predict the free-energy cost of kink formation, which is sensitive to the salt concentration, with a low-salt ($[Na^+] = 50\,\mathrm{mM}$) value of $18\, k_B T$ ($h = 22\, k_B T, b = 0.3\,\mathrm{rad}$) and a high-salt ($[Mg^+] = 10\,\mathrm{mM}$) value of $12\, k_B T$ ($h = 17\, k_B T, b = 0.7\,\mathrm{rad}$).

2.2.2 End-to-end distance (EED)

In practice, applying a WLC model relies on a number of system-specific assumptions. Importantly, the free energy of the WLC model is usually reported as a function of the end-to-end distance (EED), the separation between the two ends of a duplex.

Becker *et al.*[11] have developed expressions for the probability distribution of the end-to-end distance that harmonize between several physical regimes, the rigid rod (very stiff chain), semiflexible (moderately stiff chain, DNA is in this regime) and flexible (not stiff, approaches freely-jointed chain). Note that when we compare to such distributions in Chapter 4, there will be deviations because the WLC model of Becker *et al.*[11] takes into account neither the finite thickness of the duplex (relevant at short EED), nor the extensibility of dsDNA (relevant at extensions greater than the contour length).

2.2.3 DNA cyclization: Shimada & Yamakawa *j*-factor

One application of the WLC model that will be particularly important to this thesis will be the estimation of '*j*-factor' in Chapter 6, where we will consider DNA cyclization. In particular, we will make frequent comparisons to a form for the *j*-factor developed by Shimada & Yamakawa (S&Y).[8] Their *j*-factor is effectively the local concentration of one end of a duplex at the other end, when the two ends are 'on-register'.

The S&Y WLC model, in accounting for chain end orientation, approximates the lowest free-energy contour for an elastic twisted worm-like chain with no intrinsic curvature. There are two conceptual terms, one for bending potential energy, the other for torsional potential energy. The free energy, and thereby the *j*-factor, are found by taking the partition function over contour space. Bend potential energy is length dependent, and is governed by the persistence length. Torsional potential energy is imposed by the requirement that chain ends be in an appropriate orientation. This requirement superimposes a periodic oscillation on the bending potential energy, with a period of the pitch length: maxima occur (on-register) when the duplex length is an integer number of pitch lengths, and so the duplex does not require twisting to bring the ends into register; minima occur (off-register) at half-integer pitch lengths where the

requirement to twist is most pronounced. The magnitude of the oscillations is governed by the torsional stiffness.

Later, we will show that our results deviate from S&Y due to either a breakdown of the assumption of the chain ends being on-register, or the assumptions of the WLC model itself, namely continuous bending rather than kinking. (Chapter 6).

2.3 Coarse-grained models of DNA

In addition to WLC models (also known as polymer models, see Section 2.2), other approaches such as quantum mechanical, all-atom and coarse-grained models have been applied to examining DNA,[38] and other biomolecules.[39]

Quantum mechanical simulations do not address time- and length-scales relevant to bulk DNA flexibility. All-atom simulations are, albeit to a lesser degree, limited to short time scales and small system sizes.[i] Here, we discuss coarse-grained models that cover time- (μs – ms) and length-scales (nm – μm) appropriate for the investigation of DNA flexibility (reviewed with respect to DNA nanotechnology in reference 41).

The transition from all-atom simulations to coarse-grained models is a continuum; models that may be used to directly address questions of DNA flexibility are a subset with certain characteristics. Thermodynamic models, such as the Peyrard–Bishop family[42] and SantaLucia models,[43, 44] while reproducing dsDNA melting temperatures with impressive accuracy (generally a deviation of less than 3 K), take no account of structure, and are therefore oblivious to structural defects, which may play an important role in sub-persistence-length flexibility.

"Beads on a string" models that take a structural representation of multiple[45] or single[46, 47] base-pairs as opposed to nucleotides have a similar limitation. While they do capture structure at or above the persistence length, it is unclear whether their structural representation could account for base-specific defects that may occur at high curvature in sub-persistence length fragments. Additionally, the lack of a representation of single-stranded DNA (ssDNA) flexibility limits their applicability to some experimental systems (Chapter 6, Chapter 8).

[i] Advances in force-fields and much greater computational power do lead to some exceptions, such as the 100 ns simulation of a DNA origami (3 million atoms) demonstrated in reference 40.

Those that include explicit solvent, like their all-atom cousins, are very computationally demanding.[48] If biasing were implemented, they may hold promise for investigating relatively small systems (Chapter 4); however, their application to larger, experimentally relevant systems, would still be computationally limited.

Despite these limitations, coarse-grained models have been used to investigate DNA flexibility in specific systems; for example, supercoiling in DNA minicircles.[49] While not yet used to directly simulate an experimental system, the de Pablo model[49,50] reproduces the kinetics of DNA denaturation/rehybridization and the dsDNA persistence length. Further, with some reparameterization, mechanical denaturation (force-extension) is captured.[51] These accomplishments suggest that it may be applicable to probing experimental systems directly.

2.4 Specific aims

In the following chapters, using a coarse-grained model of DNA, we will fully characterize the structure and thermodynamics of strong DNA bending (Chapter 4). Specifically, there is thought to be an expectation of a crossover from WLC to non-WLC behavior for sufficiently strong bending. A recent review of the field by Vologodskii *et al.*[26] highlighted two open questions: (i) the nature of strong bending and (ii) the characteristic curvature required for strong bending. We address both of these question, in addition to revealing the microscopic properties of the non-WLC state.

As our model (Section 3.1) describes the general behavior of DNA, this insight can be applied to a variety of systems. To demonstrate the validity of our predictions, we probe four distinct experimental systems without reparameterization, considering a 'molecular vise'[22] (Chapter 5), DNA cyclization (Chapter 6), minicircles (Chapter 7) and a strained-duplex (Chapter 8). The general behavior of these disparate systems is elucidated in terms of a basic set of structural defects.

3 Simulation methods

Simulation methods have been developed to simulate the behavior of a diverse array of physical systems. We will restrict the discussion to a high-level overview to refresh the reader, highlighting only those technical nuances that will be of importance to the particular model systems under study. Complete methods are available in Appendix A.

The oxDNA simulation software,[i] a reduction to practice of the oxDNA model developed by Ouldridge *et al.*,[52,53] encapsulates sampling, scoring and biasing into a convenient open-source software package, which will be used to run our simulations.

3.1 oxDNA: A coarse-grained DNA model

OxDNA is a powerful coarse-grained model and simulation package for nucleic acids. The model was developed by Thomas Ouldridge while working as a graduate student in the group of Ard Louis at the University of Oxford. His seminal work has spawned a series of papers detailing the subtle interplay of the physical effects that drive DNA assembly, structure and thermodynamics.[41,52,54–57] Although the model's thermodynamic, structural and dynamic properties are fully characterized elsewhere,[52,53] we will nonetheless give a brief introduction.

OxDNA is designed to bridge the gaps between all-atom simulations, coarse-grained methods that do not represent individual nucleotides and analytic expressions. As a coarse-grained model, it is able to probe length and time scales inaccessible to traditional all-atom simulations. It has been employed successfully for a plethora of systems, reviewed in reference 41, beginning with the thermodynamic and structural characterization of DNA nanotweezers.[58]

[i] oxDNA available from: `https://dna.physics.ox.ac.uk/`

The model consists of a rigid body with three sites per nucleotide, with two representing the base (*i.e.* adenine, cytosine, guanine or thymine), and one representing the backbone (*i.e.* sugar and phosphate). The model takes the classical approximation, and does not account for quantum effects. The sites interact via a complex set of anisotropic potentials,[53] including terms for hydrogen bonding and base-base stacking. The backbone connectivity is modeled as a finitely extensible nonlinear elastic (FENE) spring, and an excluded volume interaction is applied to all three sites. Simulations are run in a box of defined size, but with periodic boundary conditions to minimize boundary effects.

The model is parameterized to reproduce the thermodynamics of duplex melting in *high salt* conditions ($[Na^+] = 500\,\mathrm{mM}$), where electrostatic repulsion is short-ranged due to counter-ion screening. Non-Watson–Crick base pairing arrangements, such as Hoogsteen base pairing, are not included. The model includes two parameterizations for representing bases, average-base and sequence-dependent (Section 3.1.1).

The success of the oxDNA model is underpinned by its robust description of both single- and double-stranded DNA, including hybridization thermodynamics, persistence length and torsional modulus. DNA geometry (*e.g.* finite thickness, excluded volume, helicity) is also taken into account. Thus, DNA behavior arises naturally from the oxDNA model, without additional parameterization. This allows oxDNA to harmonize results between experimental systems, elucidating a basic set of thermodynamics which inform the varied behaviors observed experimentally.

Importantly, oxDNA is not designed to reproduce a particular experiment, but rather to serve as a guide for sound thermodynamic and structural reasoning, elucidating the basic physics underlying observations from desperate sources.

3.1.1 Parameterization: Average-base and sequence-dependent

The oxDNA model includes both average-base and sequence-dependent parameterizations, both of which are tuned to reproduce the thermodynamics of double-stranded DNA (dsDNA) via the melting temperature. In the average-base parameterization, the base pairing and stacking interactions of each base are identical, whereas in the sequence-dependent parameterization,[55]

interactions vary by base identity (*e.g.* AT base pairs are weaker than GC base pairs).

Given that an inherently more detailed parameterization is available, the careful reader may wonder why the average-base parameterization is ever employed. Note however, that the two parameterizations are complementary: the average-base parameterization allows for the examination of *generic* effects, unmodulated by sequence. For example, in Chapter 6, we will show that the average-base parameterization is sufficient to explain the generic phenomena of interest, namely the characteristic length at which a deviation from worm-like chain (WLC) occurs. Sequence-dependence is a relatively minor modulation.

Indeed, as a coarse-grained model, many known physical effects are purposefully omitted, even from the sequence-dependent parameterization: base-pair rise, the width of the major/minor groove (the grooves are identical), twist (and thereby Lk_0; see Section 7.1.1), *etc.* , are all perturbed by sequence-dependence. That oxDNA reproduces experimental observations without accounting for these effects implies that they are either i) implicitly parameterized into the model, or ii) not critical to the specific effects investigated. For both the average-base and sequence-dependent parameterizations, coarse-graining demonstrates its utility by reducing numerous complex interactions to a minimal set that conforms to experiment.

3.1.2 Molecular dynamics (MD)

Molecular dynamics (MD) is a simulation technique, that starting with a model system (a potential function or Hamiltonian) and a specification of initial conditions (starting configuration and momenta), follows the time-evolution of a system by numerical integration of Newton's equations of motion. In the ergodic limit, this algorithm samples the equilibrated distribution in the *NVE* ensemble. A thermostat and/or barostat may be applied to approximate other ensembles.

In oxDNA, we are interested in the *NVT* ensemble.[ii] We do not treat the solvent explicitly, but rather diffusive dynamics are achieved by an Andersen-like thermostat[59] (Appendix A.4). Typical simulation parameters are summarized in Table 3.1. A brief description of application-specific simulation methods is given in each chapter. Complete details, again by chapter, are available in

[ii] Since the partial pressure of DNA is assumed to be negligible, $\Delta G = \Delta A$. For convenience and ease of comparison with experiment, we choose to report ΔG

Parameter	Value
Ensemble	*NVT*
Temperature T	298 K
Number of MD steps N_{steps}	1×10^8
Simulation time step δt	15 fs
Diffusion coefficient D	600 $\text{nm}^2\ \text{ns}^{-1}$

Table 3.1 Default simulation parameters, molecular dynamics (MD)
Default MD simulation parameters.

Parameter	Value
Ensemble	*NVT*
Temperature T	298 K
Number of VMMC steps per particle N_{steps}	1×10^7
Rotational seed move δx_{rot}	0.21 nm
Translational seed move δx_{trans}	0.21 nm

Table 3.2 Default simulation parameters, virtual-move Monte Carlo (VMMC)
Default VMMC simulation parameters.

Appendix A.

3.1.3 Virtual-move Monte Carlo (VMMC)

In oxDNA, MD is typically used to examine fluctuations, such as transient structural defects. We employ Monte Carlo (MC) methods because for certain systems, their sampling may be more efficient. With MC, the system is evolved in configuration-space instead of time, by proposing trial moves, and then accepting or rejecting these moves with respect to an acceptance criterion designed to reproduce a distribution, usually the Boltzmann distribution.

Virtual-move Monte Carlo (VMMC) is one such MC method, proposing cluster trial moves instead of single-particle trial moves. This approach is computationally efficient for sparse systems with strong local interactions, and systems that rely on cooperative motion as a primary relaxation mode, such as DNA. Typical parameters are summarized in Table 3.2. Application-specific simulation methods are available in Appendix A.

3.1.4 Umbrella sampling

MD and VMMC simulations excel at capturing very short-time scale phenomena with rapid interconversion; for example, structural deformations such as kinking and fraying. Where time scales are longer, or interconversion between states slow, rare event techniques may be employed to more efficiently sample the relevant states of the system of interest.

In Part I we will frequently use umbrella sampling with VMMC to efficiently sample rare states, and the equilibrium between two states separated by significant free-energy barriers. Umbrella sampling involves biased sampling to explore these rare or transition states; this bias is removed at the analysis stage. Without the bias, rare and transition states would be probed infrequently. The free-energy profile that we obtain with simulation can also provide information about dynamics, assuming that variations in rates are dominated by changes in the magnitude of the relevant free-energy barrier, as is elaborated in Chapter 6.

To further speed-up simulations, windowing may also be employed, usually with two windowed simulations, overlapping in a single well-defined transition region. This obviates the need for more advanced window stitching techniques, such as weighted histogram analysis method (WHAM).[60,61]

3.2 Definitions

Our study investigates the role of several structural defects, the detection and reporting of which rely on practical definitions. Namely, we require definitions of fraying, bubble formation and kinking.

3.2.1 Base pair, bubble and fray

The identification of bubbles and fraying first requires a definition of a formed base pair. Following previous work,[52] we define a base pair as formed if the hydrogen-bonding interaction between the two complementary bases is stronger than 4.142×10^{-21} J ($\sim 1\, k_{\mathrm{B}}T$ at $T = 298$ K). We define fraying as the loss of base pairs at the duplex ends, while bubbles are defined as the

opening of base pairs in otherwise continuous stretches away from the duplex ends.

3.2.2 Kink

A practical definition of a kink is less straightforward. Qualitatively, a kink is an area of pronounced curvature localized to a small region (Section 2.1.2). There is some diversity of opinion on the quantitative description and classification of kinks;[25] however, they have in common the breaking of base pairs and/or stacking.[iii] Other requirements may include a change in the relative orientation of nucleotides. Due to the ambiguity in the definition of a kink, we explore both structural and energetic criteria for kinking.

Throughout Part I, we rely primarily on the structural criterion. The similarities and differences between the two criteria are discussed in-depth in Section 4.4.1; though, in general, the structural detector can be considered a *lower* bound for kinking, while the energetic detector is an *upper* bound for kinking.

3.2.2.1 Structural kink criterion

Structurally, we define a kink using the relative orientation of consecutive nucleotides. In oxDNA, the orientation of each nucleotide is unequivocally determined by two orthogonal unit vectors, the base-backbone vector and the base-normal vector. The base-backbone vector connects the backbone and base interaction sites of each nucleotide. The model is designed such that, in a relaxed duplex, the base-normal vector at base i, $\hat{\boldsymbol{a}}_i$, and at the next base $i+1$, $\hat{\boldsymbol{a}}_{i+1}$, are approximately parallel; that is, $\hat{\boldsymbol{a}}_i \cdot \hat{\boldsymbol{a}}_{i+1} \approx 1$. A kink is defined to be present if $\hat{\boldsymbol{a}}_i \cdot \hat{\boldsymbol{a}}_{i+1} < 0$, a condition implying the antiparallel orientation of consecutive nucleotides along one strand.[iv]

Naturally, for duplex regions, we consider kinking along either strand; although, if a kink is present in the duplex, both strands are usually kinked. To limit false positives due to fraying,

[iii] The diligent reader may wonder how exactly bubble formation is distinguished from kinking. While kinks and bubbles are often contemporaneous, this is not always the case (Section 4.4.3).

[iv] We have considered slight modifications to the aforementioned scheme to address the issue of potential underreporting. While structures may indeed appear qualitatively kinked, they may not satisfy the rather stringent $\hat{\boldsymbol{a}}_i \cdot \hat{\boldsymbol{a}}_{i+1} < 0$ criteria; this is especially true for kinks at a nick. Considering partially parallel orientations (*e.g.* $\hat{\boldsymbol{a}}_i \cdot \hat{\boldsymbol{a}}_{i+1} < 0.5$) leads to the opposite problem, potential overreporting. In an alternative, albeit unphysical approach, we treat the nicked strand as contiguous for the purposes of kink detection. Unsurprisingly, this approach yields a higher estimate for kinking. The qualitative, and in most cases quantitative, interpretation of our data is robust to these changes in the kink criteria.

we do not include the first and last 3 pairs of nucleotides in the duplex region in our analysis. Further, frays are explicitly detected and ignored as kink locations.

Kinks in nicked regions are treated slightly differently, as our criterion is only well defined along an intact strand. A kink at a nick is detected on the opposite strand. As kinks at a nick may diffuse slightly, we define a region of 3 base pairs on either side of the nick on the intact strand. If the intact strand is kinked in this 6 base—pair region, then the molecule is considered kinked at that nick. In the case of a strand with two nicks (*e.g.* Chapter 6), kinking at a nick is monitored for both strands.

3.2.2.2 Energetic kink criterion

Energetically, a kink can be defined as a disruption in stacking between adjacent base pairs, where the stacking potential energy is less than one-tenth its typical value in a duplex. For the oxDNA model, the structural and energetic criteria for kinks are very similar when kinks are reasonably probable due to strong bending of the duplex. In this regime, definition of a kink is reasonably unambiguous and straightforward as favorable kinks are necessarily strongly bent, and both criteria do a reasonable job of identifying them.

However, at small bending fluctuations where kinked states are unlikely, the two criteria can give results that are qualitatively consistent but differ quantitatively. Caution should be used when interpreting the data; a fuller discussion is provided in Section 4.4.1.

4 Thermodynamics of DNA bending

There is much interest in understanding the physics of DNA bending, due to both its biological role in genome regulation and its relevance to nanotechnology. For small bending fluctuations, worm-like chain models give a satisfactory description of DNA bending, predicting continuously bent configurations as expected from a semi-flexible rod. However, there is less consensus on what happens at larger bending fluctuations, with experimental evidence pointing to a transition from continuously bent configurations to a state with strongly localized bending. Here, we use a coarse-grained model of DNA to explore both regimes and fully characterize the thermodynamics and mechanics of duplex DNA bending. We find that DNA can form a kinked state in which the stress is localized at the expense of disrupting a small number of base-pairing and stacking interactions.

4.1 Introduction

In this chapter, we use oxDNA (Section 3.1) to investigate the thermodynamics of DNA bending, systematically addressing questions such as the characteristic length and energy scale of kink formation (Section 2.4). OxDNA is particularly well-suited to this study because of its physical description of single- and double-stranded DNA, including the thermodynamics of hybridization and mechanical properties such as persistence length and torsional modulus.[52,53] As such, it is an ideal model for exploring kinking, and is capable of highlighting general properties as well as providing evidence as to whether the reported observations of kinking are consistent with the known thermodynamic and mechanical properties of DNA.

4.2 Simulation methods

In order to probe the free-energy cost of bending, simulations were performed with the virtual-move Monte Carlo (VMMC) algorithm[62] (Section 3.1.3) in an *NVT* ensemble at 298 K. As the entire DNA bending regime is of interest, including strongly bent states that are not thermally accessible, umbrella sampling[63] (Section 3.1.4) was employed to efficiently sample the free-energy landscape of DNA bending. The end-to-end distance (EED) R'_{ee}, defined as the distance between the base interaction sites of the first and last bases along a strand was, used as an order parameter for the umbrella sampling procedure. For clarity, we use the average-base parameterization to elucidate the fundamental thermodynamics of DNA bending. Additionally, sequence-dependent thermodynamics are used to elucidate sequence effects in bending free-energy profiles. Further details of simulation procedure are given in Appendix A.1.1.

To simplify interpretation of our results, base pairing is enforced through the umbrella potential at the ends of the duplex to explicitly disallow 'fraying' of duplexes. Fraying is a possible mechanism for relaxation of stress in bent systems;[22] however, we do not consider it here as we are primarily concerned with the response of the duplex to stress, rather than with end effects. Also, when we consider the 'molecular vise' in Chapter 5, the system is explicitly designed to minimize fraying, as discussed by Fields *et al.*[22]

We opted to constrain the end base pairs (one pair on either end of the duplex). Imposing such a constraint allows us to sensibly define and measure the EED between the centre of the helical axes, R_{ee}, as the first and last base pairs are guaranteed to be formed. Error bars represent the standard error of the mean from 5 independent simulations. Results are essentially identical for a condition where 1–5 nucleotides at either end of the duplex are constrained to be base paired (Appendix A.1.2, Figure A.1).

4.3 Free energy of bending

We start our discussion by characterizing the basic thermodynamics of DNA bending in oxDNA. We simulate a 30 bp DNA duplex (Figure 4.1 (a)), and report the free-energy profile $\Delta G(R_{ee})$ as a

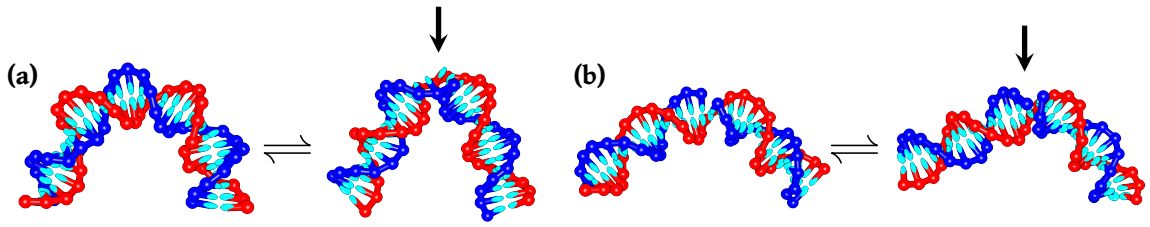


Figure 4.1 OxDNA representations of the kink/unkink equilibrium in intact and nicked duplexes
Representative oxDNA illustrations of kinking in a 30 bp DNA duplex. **(a)** The kinking transition of an intact 30 bp DNA duplex at $R_{ee} = 4$ nm. Kinking intact DNA induces the formation of an 1–3 bp bubble, as well as disrupting stacking (**arrow**). **(b)** The kinking transition of a nicked 30 bp DNA duplex at $R_{ee} = 8$ nm. Only disruption of stacking is required to kink nicked DNA (**arrow**).

function of the end-to-end distance R_{ee}. The free energy of duplex bending has two distinct regimes, as is indicated by a change in slope at $R_{ee}^{trans} \approx 5.5$ nm (Figure 4.2 (a)). The obvious interpretation of this change in slope is that it is attributable to the appearance of a specific structural defect, a kink (Section 2.1.2). To verify, we decompose the overall free-energy profile into contributions from continuously bent and kinked macrostates, revealing with clarity that the change in slope is indeed the crossover between the regimes.

The probability of kinking is similar for the energetic and structural criteria (see Section 3.2.2 for definitions), suggesting that both stacking and base pairing disruptions are usually required to form a kink in the duplex (Figure 4.1 (b)). Indeed, representative oxDNA structures for the continuously bent and kinked 30 bp duplex (Figure 4.1 (a)) reveal that duplex kinking typically induces a 1–3 bp bubble at the kink location (Section 4.4.3), and that this kink is located near the centre of the duplex (Section 4.4.2).

4.3.1 Comparison to worm-like chain (WLC) model

In Figure 4.2 (a) we compare the predictions of oxDNA to those of a WLC model (Section 2.2). We compare specifically to the approximations to the WLC EED distribution developed by Becker, Rosa & Everaers[11] (Section 2.2.2), hereafter referred to as the Becker model. As expected, when using the base pair rise and persistence length appropriate to oxDNA, agreement between the Becker model and oxDNA is excellent for intermediate R_{ee}.

Deviations arise between oxDNA and the Becker model at shorter R_{ee} due to kinking. Deviations

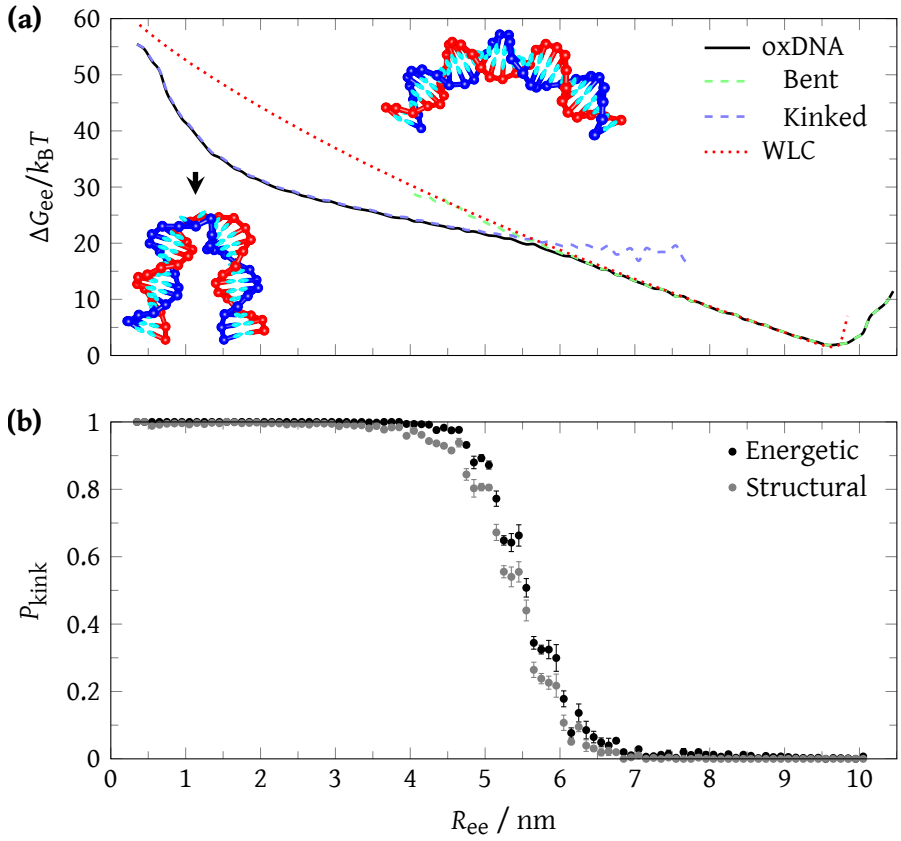


Figure 4.2 Free energy of bending an intact duplex
(a) Free-energy of bending $\Delta G(R_{ee})$ for oxDNA (**black**), with the decomposed free energies of the continuously bent (**light green** dashed) and kinked (**light blue** dashed) regimes overlaid. OxDNA is compared to the worm-like chain (WLC) model approximation of Becker *et al.*[11] (**red** dotted). The comparison is fit-free, using known oxDNA parameters for persistence length (41.82 nm) and rise per base pair along the helical axis (0.34 nm).[53] To perform the fairest comparison between the Becker model and oxDNA, rather than setting the minimum $\Delta G(R_{ee})$ to be zero as elsewhere in the text, we normalise $\Delta G(R_{ee})$ so that $\exp\left(-\Delta G/k_BT\right) = P(R_{ee})$ in both cases. **(b)** Probability of kinking P_{kink} using energetic (**black**) and structural (**gray**) criteria for a kink. Inset figures are oxDNA representations for a 30 bp duplex, depicting the kinked (**arrow**) and continuously bent configurations at $R_{ee} = 4$ nm and $R_{ee} = 8$ nm respectively.

also occur at longer R_{ee}, where R_{ee} is less than the duplex DNA contour length $R_{contour}$ ($R_{ee} \lesssim R_{contour}$), due to the Becker model neglecting the finite intrinsic extensibility of DNA. Additionally, the Becker model does not account for the finite thickness of DNA, which causes excluded volume effects at very short R_{ee}. In oxDNA, a sharp rise in ΔG is present at $R_{ee} \approx 2\,\text{nm}$.

4.3.2 Length dependence

According to the WLC model, the enthalpic cost of bending a duplex to some fraction of its contour length decreases with duplex length. By contrast, there is no obvious reason why the free-energy cost of forming a kink at the centre of the duplex should be strongly length-dependent. We might therefore expect that, as duplex length is increased, the location of the kink transition should shift to smaller values of $R_{ee}/R_{contour}$.

For sufficiently long DNA, we expect to see unkinked duplexes even for small $R_{ee}/R_{contour}$. These expectations are confirmed in Figure 4.3; long duplexes ($\geq$ 73 bp) do not preferentially kink at any R_{ee}, and for shorter duplexes the length-dependence of the kink transition is as expected. The re-entrant behavior at extremely small $R_{ee}/R_{contour}$ arises from a change in geometry that is enforced by excluded volume effects (Section 4.4.4).

It seems natural to define a free-energy cost $\Delta G_{trans} = \Delta G(R_{ee}^{trans}) = 18.1\,k_B T$ for the 30 bp duplex. This is essentially a measure of work that must be done on the duplex by pulling its ends together before kink formation occurs. This numerical value is comparable to the estimate of '$\Delta G_{collapse}$' made by Fields *et al.*,[22] the bending stress released upon kinking for a similar system. However, free energies must be carefully defined if quantitative comparisons are to be made. In particular, ΔG_{trans} is not a generic free-energy cost of kink formation that can be directly compared with other systems in which kink transitions may occur.

For clarity, we present the bending free energy and end-to-end separation at the mid-point of the kink transition for all simulations in Table 4.1.[i]

[i] It is important to note that ΔG_{trans} is a function of duplex length, even in the context of systems in which kink formation is driven by pulling two ends of a duplex together, because kinking occurs at smaller $R_{ee}/R_{contour}$ as duplex length is increased.

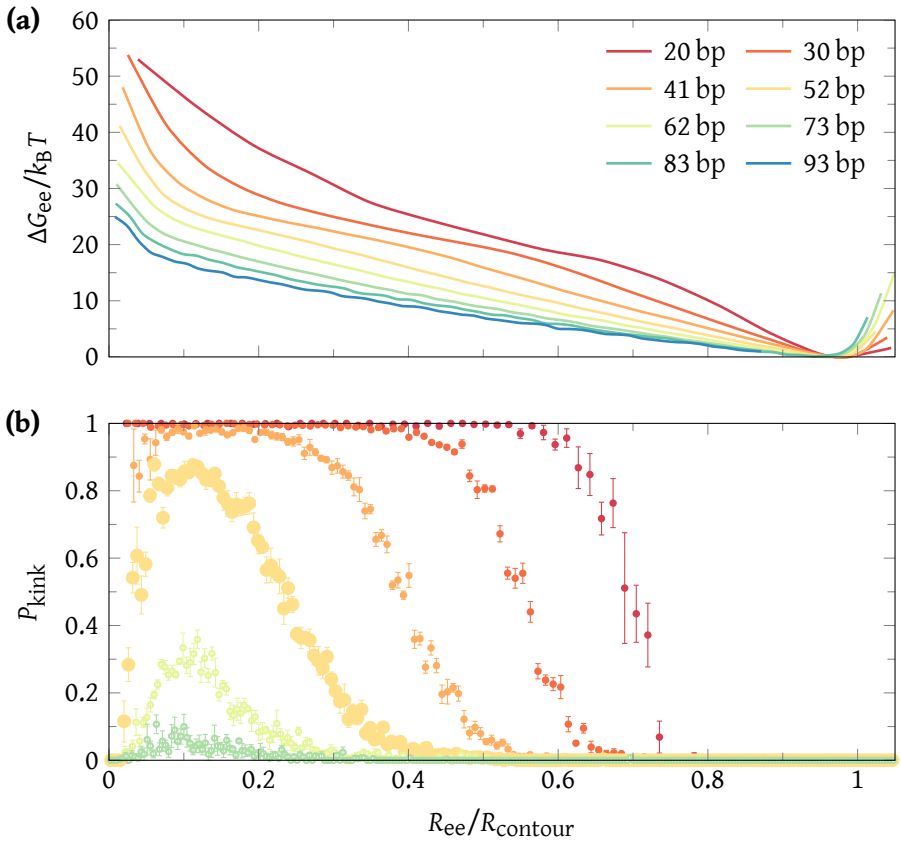


Figure 4.3 Free energy of bending, length dependence
(a) Length dependence of $\Delta G(R_{ee})$, the duplex DNA bending free energy, as a function of reduced end-to-end distance $R_{ee}/R_{contour}$. Error bars are comparable to the line thickness.
(b) Length dependence of duplex DNA kinking. The structural criterion for kinking is shown; results using the energetic criterion are very similar in the kink transition region (Section 4.4.1).

	ΔG_{trans} / $k_{\text{B}}T$	$R_{\text{ee}}^{\text{trans}}$ / nm
Length / bp[†]		
20	15.9	4.7
30	18.1	5.5
[n]30	6.8	7.8
41	19.7	5.3
52	21.4	4.3
Sequence (28 bp)		
poly(GC)	21.1	5.1
Avg-base	18.0	5.4
poly(AT)	14.8	5.8
Mismatch (28 bp)		
Intact	16.1	5.7
[m]GG	13.0	6.0
[m]GT	13.3	6.0
[m]AA	12.8	6.1
[m]CC	13.5	5.9
Nicked (28 bp)		
[n]poly(GC)	7.9	7.2
[n]Avg-base	6.7	7.3
[n]poly(AT)	4.8	7.6

Table 4.1 Free energy of bending and end-to-end distance at kink transition mid-point
The bending free energy at the mid-point of the kink transition ΔG_{trans}, indicated by the crossover between continuously bent and kinked regimes, occurs at the end-to-end separation, $R_{\text{ee}}^{\text{trans}}$. Error bars are ± 0.1 for both ΔG_{trans} and $R_{\text{ee}}^{\text{trans}}$.
[n] Nick at the mid-point of the duplex.
[m] Mismatch in the duplex from Fields *et al.*[22]
[†] Lengths of 62 bp and 73 bp not shown because their kink transition free-energies are not well defined. While kinked microstates do occur, their probability never reaches $1/2$, our definition of the kink transition mid-point.

4.3.3 Comparison to free energy of kink formation

One might define a quantity ΔG_{kink} through the probability of spontaneous kink formation at a specific location in the duplex, in the absence of stress. Although in principle ΔG_{kink} would be expected to be a generic (sequence-dependent) quantity, in practice its value strongly dependents on the criteria used to identify kinks (Section 4.4.1). This is in sharp contrast to ΔG_{trans}, which is almost identical for both criteria used to identify kinking in this work. In the presence of sharp bending, the overwhelming majority of disruptions to the basic duplex structure are sharply bent regions with broken stacking interactions, which are unambiguously 'kinks'. In an unstressed system, however, the most common deviations from the canonical duplex are not unambiguously identifiable as kinks, making the definition of a meaningful ΔG_{kink} problematic.

Even with well-defined criteria for identifying kinks, ΔG_{kink} is not directly relatable to ΔG_{trans}. Two reasons for this difference are particularly evident in the context of bending duplexes. Firstly, to be a generic quantity, ΔG_{kink} must be defined through the probability that a kink occurs that is centred on a specific base pair (or base-pair step) in a duplex. In the context of duplex bending, however, kink formation can occur over a range of locations (Section 4.4.2), and this fact is reflected in ΔG_{trans}. Secondly, in pulling the ends of the duplex together to $R_{\text{ee}}^{\text{trans}}$, we do work on both the unkinked *and* the kinked state. Although kinks are more flexible than unkinked duplexes, due to entropic effects and the need to strain the kink itself, kinks posses a finite resistance to bending. This is evidenced by the post-kinking slope in $\Delta G(R_{\text{ee}})$ in Figure 4.2 (a). To compensate, we must apply even more stress to the system to reach the point at which kinks become favourable relative to unkinked duplexes, tending to increase ΔG_{trans}.

That the bending stress at which kinking occurs in a specific system (ΔG_{trans}) is not immediately related to the intrinsic cost of kink formation (ΔG_{kink}) emphasizes the difficulty in comparing between distinct experimental setups, in which the factors contributing to ΔG_{trans} may well be different. For example, in some systems, it may be possible to relieve stress by kinking at one of a number of positions, as observed in simulations of short minicircles (Section 7.4.3, Section 7.4.4). In others, such as the 'molecular vise' system,[22] the kinking location may be constrained to be near the mid-point of a duplex. The free-energy cost of bending the kinked

state may also be distinct in different systems.

Unless care is taken, such differences will hamper quantitative comparison of free energies. Analytic models[22,36] used to extract free energies from data must therefore be carefully chosen if quantitative comparisons are to be drawn. Indeed, one of the advantages of simulations employing coarse-graining is that direct comparisons can be made between distinct systems at the level of physical observables, rather than indirectly via free energies derived from analytic models. In this work, we will only make direct comparisons of ΔG_{trans} for systems that are nearly identical, probing the impact of a specific change, such as sequence or the presence of a mismatch. Thus, deviations will largely reflect differences in ΔG_{kink} and not system-specific effects.

4.3.4 Specific effects

4.3.4.1 Sequence dependence

Given that the relative free energies of bending and kinking can also be perturbed with sequence variation,[64] we investigated how sequence variation affects the kink transition. We report sequence dependence in the free energy $\Delta G(R_{\text{ee}})$ of bending for a 28 bp duplex (Figure 4.4 (a,d)). It is clear from this data that the poly(AT) sequence undergoes kinking more easily than the poly(GC) sequence; ΔG_{trans} is lower, presumably due to the weaker base-pairing in the poly(AT) sequence. A secondary observation is that the poly(AT) sequence is slightly more flexible prior to the kink transition in oxDNA. These two effects should, in principle, have opposing consequences for $R_{\text{ee}}^{\text{trans}}$, tending to increase and decrease $R_{\text{ee}}^{\text{trans}}$ respectively, *i.e.* its persistence length is smaller. It is clear from Figure 4.4 (d), however, that the ease of kinking dominates and poly(AT) has a longer $R_{\text{ee}}^{\text{trans}}$ and smaller ΔG_{trans} compared to poly(GC).

We also observe that a 65 % GC sequence has a slightly lower ΔG_{trans} than the average-base model (Table 4.1). This behavior is expected since the duplex will kink and bend preferentially in AT-rich regions. Due to the lack of easily kinkable regions, the average-base model is harder to kink than would be naïvely expected from an ensemble of 50 % GC-content sequences.

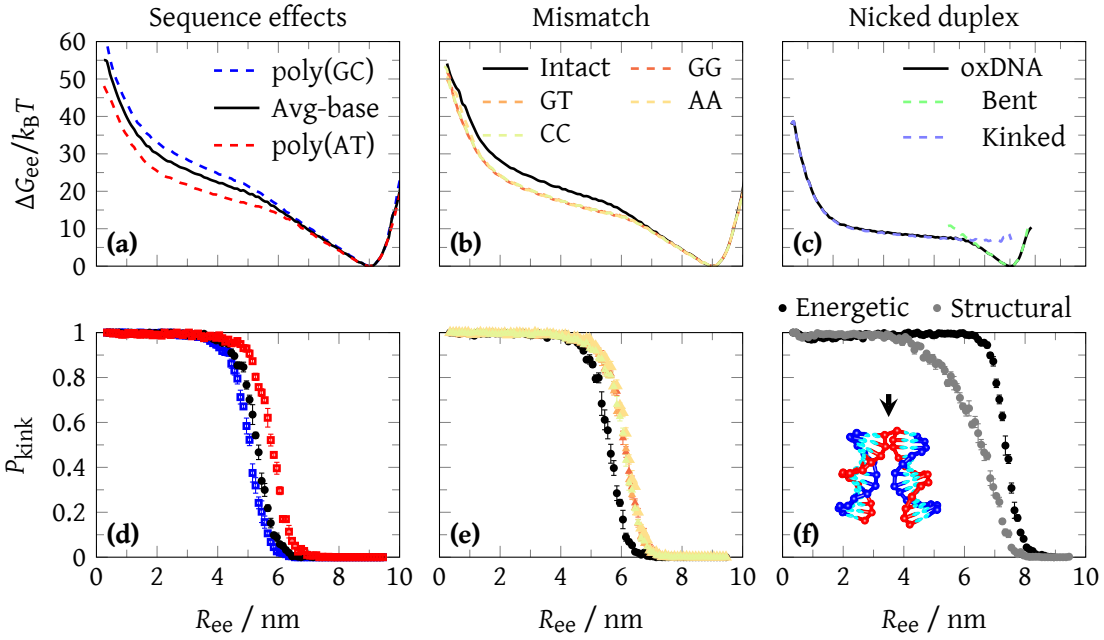


Figure 4.4 Effects of sequence-dependence, nicks and mismatches on the free energy of bending
(a) Sequence dependence of bending for a 28 bp intact DNA duplex, comparing the sequence-dependent and average-base oxDNA parameterizations. Error bars are comparable to the line thickness. **(b)** Impact of mismatches on the bending of a 28 bp DNA duplex, compared to intact DNA, using the oxDNA sequence-dependent parameterization. Mismatches, ordered by increasing free-energy cost, are as follows: GG, GT, AA and CC. Sequences from Fields *et al.*[22] **(c)** Impact of a nick on the bending of a 28 bp duplex for oxDNA average-base parameterization (**black**), with the decomposed free energies of the continuously bent (**light green** dashed) and kinked (**light blue** dashed) regimes overlaid. **(d)** Sequence-dependence of kinking using the structural criterion for a kink; results using the energetic criterion are very similar. **(e)** Impact of mismatches on kinking. **(f)** Impact of a nick on the kinking, via energetic (**black**) and structural (**gray**) criterions. OxDNA representation of a kink (**arrow**) in the nicked duplex at $R_{ee} = 4$ nm.

4.3.4.2 *Mismatches*

Having investigated sequence variation, we now turn our attention to a system in which we introduce a mismatch, *i.e.* a structural motif that substitutes a base pair with a non-Watson–Crick pair. It is reasonable to expect that kinking will localize to the mismatch, provided it is conveniently positioned, and ΔG_{trans} will be reduced. Following Fields *et al.*,[22] a mismatch was introduced at the mid-point of an otherwise intact duplex. We investigated GG, GT, AA and CC mismatches.

As expected, oxDNA reports that mismatched sequences have a lower ΔG_{trans} and undergo the kink transition at longer $R_{\text{ee}}^{\text{trans}}$ (Figure 4.4 (b,e)). The change in ΔG_{trans} is consistent with typical free-energy costs associated with mismatches:[44] oxDNA reports 16.1 $k_{\text{B}}T$ for the perfectly-matched duplex versus 12.8–13.5 $k_{\text{B}}T$ for the mismatches, a deviation of $\sim$3 $k_{\text{B}}T$ (Table 4.1). Fields *et al.* also found evidence that the ease of kinking was correlated with the sequence-dependent mismatch cost predicted by the SantaLucia model of DNA thermodynamics.[22] OxDNA's treatment of mismatches is extremely simple, neglecting any residual interactions or possible steric clashes between geometrically incompatible bases. Consequently, it does not strongly distinguish between different mismatches, and does not reproduce sequence-dependent mismatch variation.

4.3.4.3 *Nicks*

Like a mismatch, a nick, which is a break in an otherwise contiguous stretch of DNA whereby the backbone phosphodiester bond is severed, is expected to impact DNA bending.[ii] In the relaxed state, base pairing is preserved and coaxial stacking may occur between the adjacent, but now non-covalently bonded, nucleotides. It is relatively easy for a nicked system to sharply kink while only disrupting stacking and not hydrogen-bonding (Figure 4.1 (b)). We observe, unsurprisingly, that the presence of a nick dramatically reduces the bending free energy at the kink transition (Figure 4.4 (c)). In accordance with this lower bending free energy ($\Delta G_{\text{trans}} = 6.7\, k_{\text{B}}T$), the crossover from continuously bent to kinked regimes, and therefore the kink transition mid-point, is shifted from $R_{\text{ee}}^{\text{trans}} = 5.4$ nm to $R_{\text{ee}}^{\text{trans}} = 7.3$ nm for the 28 bp

[ii] As oxDNA is a coarse-grained model, there does not exist a 1:1 relationship between an oxDNA nick and a physical nick, i.e. the precise disruption to the phosphodiester bond is ambiguous, and beyond relieving torsional stress, structural details and rearrangements due to the presence of a nick should not be inferred.

systems (Table 4.1).

It is also worth noting that DNA with a nick is far easier to bend then normal duplex DNA after the kink transition: the post-kinking slope at $R_{ee} \approx 2$–5 nm for the nicked system (Figure 4.4 (c)) is much shallower than for the intact duplex (Figure 4.4 (b)). The absence of the constraint of the additional backbone bond in the nicked system makes extremely sharp bending very easy once coaxial stacking has been broken; this contrast highlights the work that must be done to strongly bend a kinked duplex without a nick. We note that for the nicked system, a more substantial difference between energetic and structural criteria for kinks is evident (Figure 4.4 (f)). This distinction arises because nicked systems undergo significant kinking at relatively large R_{ee}, when it is less advantageous for kinked sections to bend extremely sharply; hence, because the structural criterion implies an effective angle $> 90°$ between consecutive bases, it can underestimate the prevalence of kinks. We refer the reader to Section 4.4.1 for further details.[iii]

4.3.5 Angular dependence

While we usually choose to report our bending free energy results as a function of EED R_{ee} (Section 4.3), it may be instructive to instead examine the free-energy profiles as a function of the bending angle θ, formed in-between the two duplex sections that meet at the kink. We show bending free energy versus the bending angle θ using the same simulations as Figure 4.2 (1×10^8 VMMC steps per particle per simulation for each of 5 independent simulations, outputting configurations every 1×10^4 VMMC steps per particle).

Qualitatively, the bending angle θ is the total angle between the relatively straight duplex sections at the location of the kink. Small θ ($\theta \to 0°$) roughly corresponds to large R_{ee} ($R_{ee} \to R_{contour}$, the contour length of DNA), and large θ ($\theta \to 180°$) to small R_{ee} ($R_{ee} \to 0$ nm).

[iii] The careful reader may note that the bending free-energy appears dominated by hydrogen-bonding as opposed to stacking. For intact systems, this observation is consistent with the oxDNA sequence-dependent parameterization and the structural definition of a kink (Section 3.2.2.1). Base pairing varies by 50 % (2 for A-T versus 3 for G-C base pairs), whereas stacking interactions vary by only 6 %. Further, the structural definition of kinking implicitly requires the breaking of base pairs. While the detector is unimportant for kinking in intact duplexes, because kinking at a nick does not require the breaking of base pairs, the structural and energetic criteria of a kink differ for nicked systems Section 4.3.4.3. As we do not directly probe the sequence-dependence of kinking at a nick, and the impact of stacking interactions versus fraying at the nick is ambiguous, for nicked systems, we suggest against attributing sequence-dependent variation in the free-energy of bending to either hydrogen-bonding or stacking.

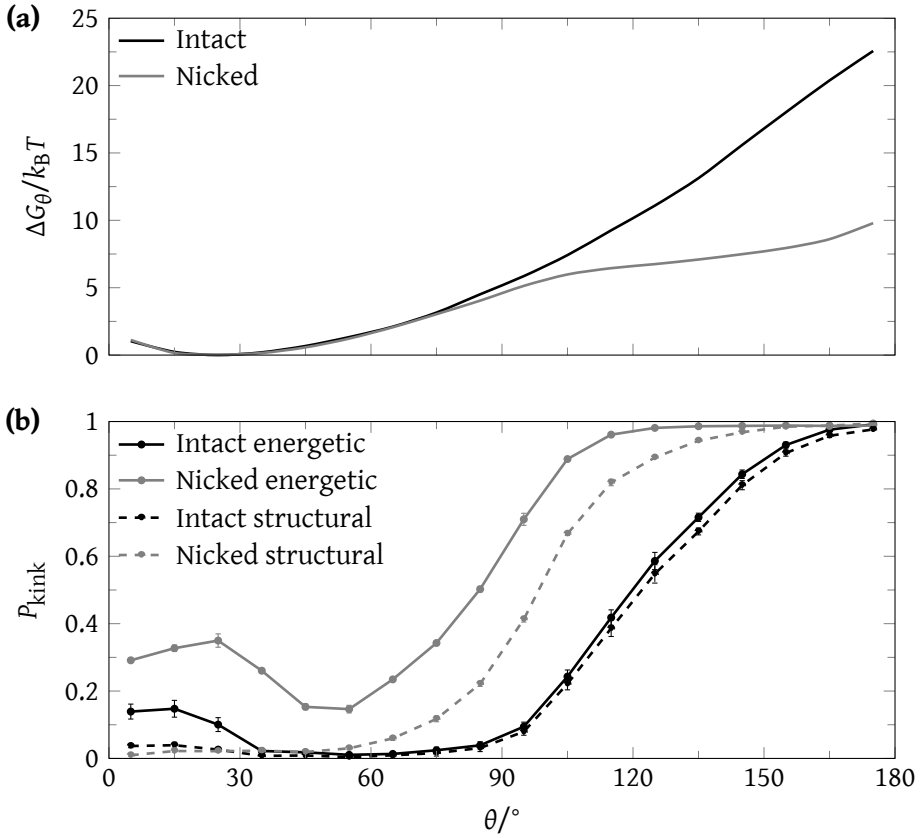


Figure 4.5 Free energy of bending, bending angle θ
(a) Free energy of bending and **(b)** probability of kinking for a 28 base pair intact and nicked duplex as a function of the bending angle θ. Data extracted from Figure 4.2.

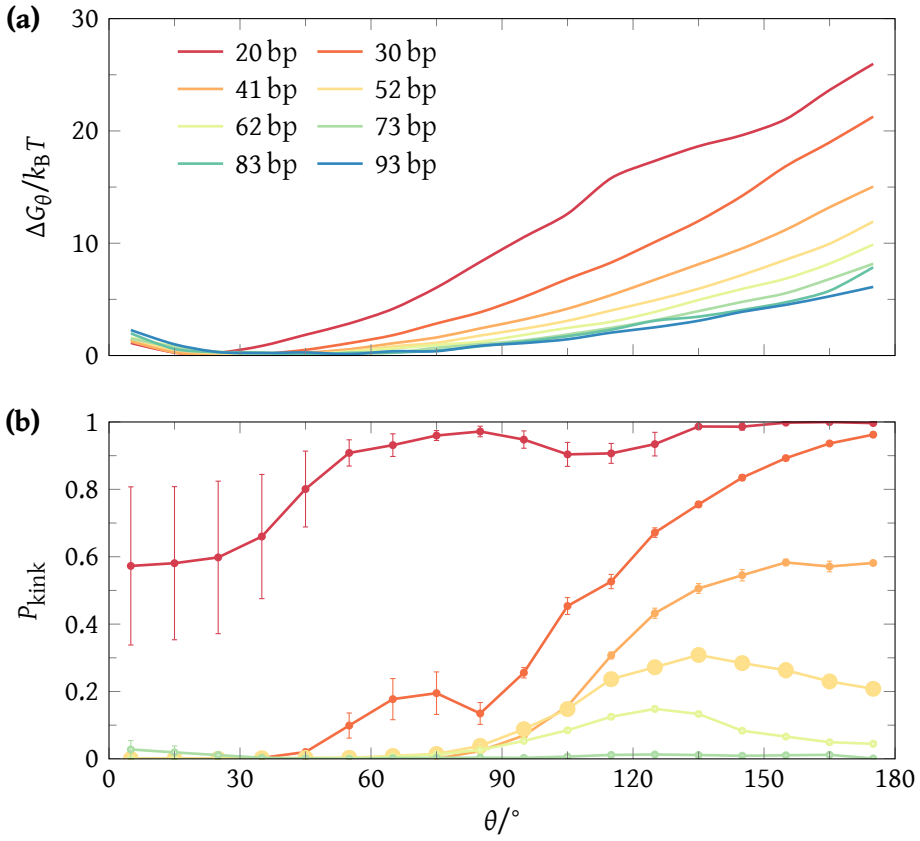


Figure 4.6 Free energy of bending, length dependence
(a) Length dependence of $\Delta G(\theta)$, the duplex DNA bending free energy as a function of bending angle θ. With the exception of $N_{bp} = 20$, error bars are comparable to line thickness.
(b) Length dependence of duplex DNA kinking. The structural criterion for kinking is shown; results using the energetic criterion are very similar in the kink transition region (Figure 4.5).

A qualitative definition of θ is fairly straight-forward, but a robust quantitative definition is somewhat more troublesome. For our quantitative definition, we exploit the facts that (i) for bent duplexes kinking generally occurs near the middle of the duplex and (ii) that we disallow non-native base pairing and fraying. First we determine the helical axis vectors along the first and last 5 base pairs of the duplex. At each base pair index, we take the midpoint between bases, then define a local helical axis as the normalized vector from midpoint i to midpoint $i + 1$. The resulting 4 unit vectors are then normalized to determine the average helical axis for each of the two sections. Finally, we take the inverse cosine of the dot product of the two segment's average helical axis vectors, yielding a proxy for the bending angle θ.

For the intact duplex, unlike the EED case, there is no clear partitioning between kinked and unkinked states. The bending free energy is also highly angle dependent, with non-negligible angular dependence for a kink in the duplex. For the nicked duplex, there is clear partitioning between the kinked and unkinked states, with an angle-dependent bending free energy for the unkinked states, and negligible angle-dependence for the kinked states.

For completeness, and comparison with EED (Figure 4.3), the length dependence of intact duplexes as a function of bending angle θ is shown in Figure 4.6.

4.4 Structure and thermodynamics of kinking

4.4.1 Energetic versus structural kink criteria

In Figure 4.2 we showed that the energetic and structural kink criteria were in good agreement for the kink transition region; however, far from R_{ee}^{trans}, this agreement can be much worse. The basic problem is that kinked states are likely to be bent; however, at EEDs R_{ee} close to the contour length $R_{contour}$, the duplex is straight. At $R_{ee} \approx R_{contour}$, our structural criterion, which relies upon the duplex being sufficiently bent (a local bend angle of $> 90°$), is probably inappropriate. In principle, the energetic detector could work when duplex is straight, but kinked states typically involve a variety of disruptions (Section 4.4.1.2), normally involving the breaking of stacking, but also often broken base pairs (Section 4.4.3). The latter is not included in the energetic criterion, so states at $R_{ee} \approx R_{contour}$ will only include the stacking disruption

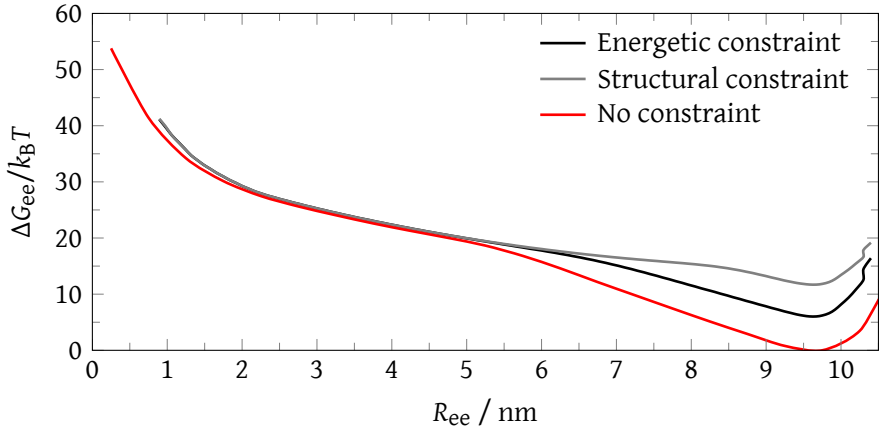


Figure 4.7 Free energy of bending, kinking near the contour length
Free energies as a function of R_{ee} for a 30 bp duplex constrained to kink by either the energetic or structural criteria. The free energy is normalized by manually matching to the data reported in Figure 4.2 (a) for kinked states at small R_{ee}. Error bars are comparable to line thickness.

associated with kinks. Additionally, unlike at $R_{ee} \approx R_{contour}$ where a structural disruption can occur anywhere along the duplex, at small R_{ee}, the kink is localized to the midpoint along the duplex.

4.4.1.1 Sampling kinked states near the contour length

We performed additional simulations for a 30 bp duplex in which an additional umbrella bias was used to forbid unkinked states (separate simulations were performed for each kink criterion. In Figure 4.7, we plot the free energy of bending as a function of the EED R_{ee}) obtained from these simulations. Curves were normalized by manually matching to the data obtained for small R_{ee} in the original system – the zero of free energy is then the same as in Figure 4.2 (a).

Although both energetic and structural criteria identify disruptions of the typical duplex structure, it is clear that the free energy of the kinked state at larger values of R_{ee} is strongly dependent on the criteria used to identify kinking. At small R_{ee}, disruptions that are clearly strongly bent kinks are prevalent, far more so than disruptions which do not involve strongly bent kinks. At larger values of R_{ee}, disruptions to the helix that are not unambiguously identifiable as qualitative kinks become more important, explaining the discrepancy.

The energetic criterion gives an upper bound on the amount of kinking. It is difficult to imagine a structure which could reasonably be described as 'kinked' without breaking stacking; however,

stacking may also be disrupted without substantial bending of the duplex through a relatively small structural rearrangement. Such configurations dominate the states which satisfy the energetic criterion at larger R_{ee}.

The structural criterion is more stringent, and might be expected to give a lower bound on the amount of kinking – satisfying the structural criterion certainly requires substantial disruption of the duplex state. It is also possible for the duplex to be substantially disrupted and for significant bending to occur without meeting the criterion. That the structural criterion can underestimate kinking is particularly evident from the data for nicked systems (Figure 4.4 (c,f)). Due to the ease of kinking, kinks become prevalent at large R_{ee} at which point the system, although obviously kinked, can be insufficiently bent to trigger the structural detector. Albeit a minor effect, note that in the nicked system, the kink criteria are only well defined along the continuous strand, and that in the intact system, reporting kinking along only one intact strand slightly under reports the probability of kinking, *i.e.* a kink on either strand.

4.4.1.2 Type-I versus type-II kinks in oxDNA

Despite our expectation that the structural criterion should be a lower bound on kinking, we find that at large R_{ee} there is a substantial contribution to the structurally-kinked macrostate from configurations with 'flipped-out' bases, or other perturbations to the duplex structure that are substantial but are not typical of the kinks that facilitate strong bending. It is not immediately clear whether such states should be classified as kinks. Overall, this ambiguity makes the definition of an intrinsic free energy of kinking extremely problematic. Any numerical value given would be strongly dependent on the criteria used to identify kinking, and a universally appropriate definition of kinked states is not obvious.

Nomenclature has been developed in all-atom molecular dynamics (MD) simulations[25,65,66] to distinguish between type-I (no disruption of hydrogen bonds, only stacking) and type-II kinks (disruption of both hydrogen bonds and stacking). Due to its coarse-grained nature, we have less confidence that oxDNA has sufficient molecular detail to contribute reliably to this discussion, e.g. oxDNA does not account for sugar puckering or other base-specific structural deviations from canonical double-stranded DNA (dsDNA) (Section 3.1.1).

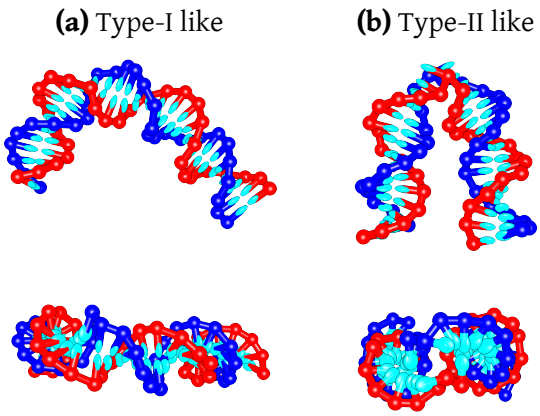


Figure 4.8 Type-I versus type-II kinks
OxDNA representations of type-I like and type-II like kinks in 30 bp duplex. Front view at top, top view at bottom. **(a)** Type-I like kinks have disrupted stacking with preserved base pairing, while **(b)** type-II like kinks disrupt both stacking and base pairing.

Furthermore, the distinction between type-I and type-II kinks are of less direct relevance to this study. We therefore do not use the type-I/type-II classification system of kinks, although we note that we see configurations which could be assigned to both types (Figure 4.8).

While the energetic and structural detectors yield very similar results for intact duplexes in the kink transition region, their behavior differs as $R_{ee}/R_{contour} \to 1$, with the structural detector saturating to $P_{kink} = 0$ and the energetic detector saturating to $P_{kink} \approx 0.01$ (Figure 4.7). This difference is attributable to spontaneous unstacking, and is not reflective of either type-I or type-II like kinks.

4.4.2 Localization

When the ends of a duplex are pulled together, it would be expected that kinking occurs near the midpoint of the duplex. Kinks far from the midpoint would result in incommensurate relaxed duplex sections, which are then less likely to have their ends in close proximity. Analysis of the kink probability distribution confirms that kinks are localized near the midpoint (Figure 4.9).

For the intact duplexes the distribution is quite wide, ∼10 base pairs, suggesting that entropic effects are quite prevalent and some degree of incommensurately can be tolerated. The distribution also appears to become slightly broader for longer duplexes. Importantly, the preference for localization about the midpoint of the duplex is due to bending: in relaxed duplex

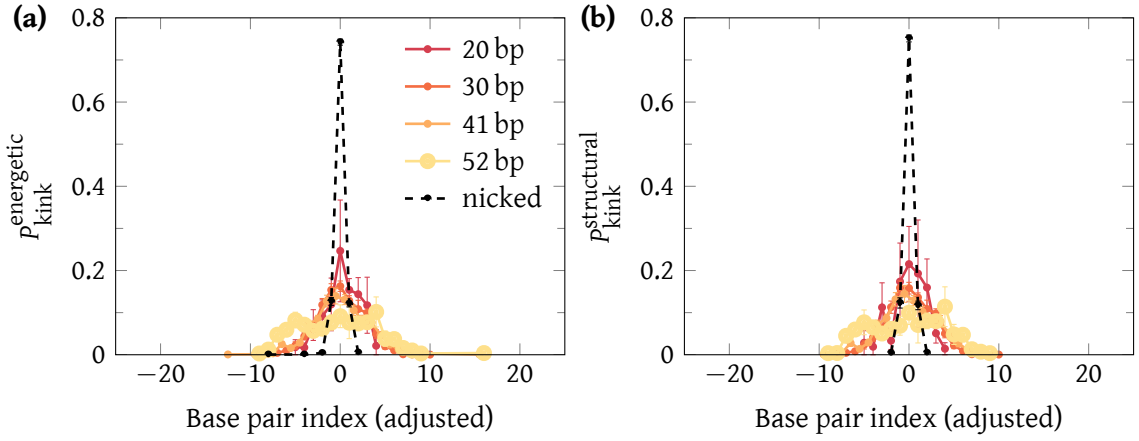


Figure 4.9 Kink localization
Normalized histogram for the probability distribution of kink localization for the **(a)** energetic and **(b)** structural criteria for intact ($N_{\text{bp}} = 20, 30, 41, 52$) and nicked ($N_{\text{bp}} = 30$) duplexes. We sample all duplexes at constant end-to-end distance from the umbrella potential bin $R'_{\text{ee}} = 4.26$–4.69 nm, which corresponds to $\langle R_{\text{ee}} \rangle \approx 4.5$ nm.

DNA, or in DNA minicircles, one would not necessarily expect a preferred site for kinking with the average-base parameterization of oxDNA.

For the nicked duplex at $N_{\text{bp}} = 30$, the distribution is quite narrow, ∼2 base pairs, reflecting the strong preference for localization of kinking at the nick (Figure 4.9). Contrary to the intact duplexes, the preference for kinking at a nick is not due to the bending stress imposed on the system. Because only the cost of disrupting coaxial stacking must be incurred, and not that of breaking base pairing, nicked sites have a lower free-energy cost of kinking.

Since structural and energetic kinks are a property defined between consecutive bases (Section 3.2.2), we report the kink probability distribution in base steps, where base step i refers to a kink between base index i and $i + 1$. To define the location of a kink, all base steps along one strand that satisfy either the structural or energetic kink criteria are included in the histogram. To normalize, we divide by the total number of detected kinks along both strands. Kinks are considered distinct if they are separated by more than 6 base steps. Thus, the normalized kink distribution will sum to the average number of base steps satisfying the kink criterion rather than to 1. To compare multiple lengths, we report the probability of kinking versus the *adjusted* base pair index, centered around the mid-point of the duplex, $^{(N_{\text{bp}} - 2)}/_{2}$. The -2 offset accounts for zero-based numbering and indexing the base step to base index i instead of $i + 1$.

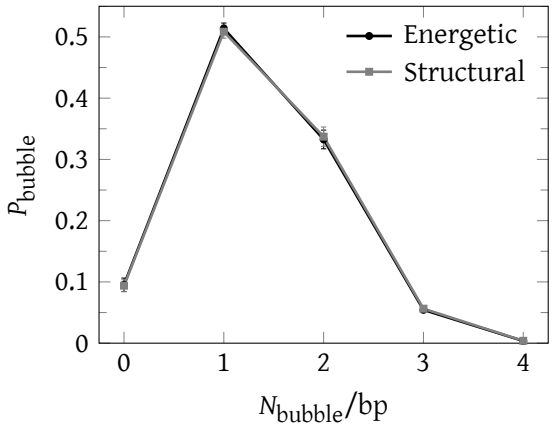


Figure 4.10 Bubble size distribution
Normalized histogram for the probability distribution P_{bubble}, the total bubble size in base pairs (N_{bubble}) given the presence of a kink via either the energetic (**black**) or structural (**gray**) criteria for $N_{\text{bp}} = 30$ at $\langle R_{\text{ee}} \rangle \approx 4.5$ nm from the umbrella potential bin $R'_{\text{ee}} = 4.26$–4.69 nm. Error bars are smaller than the points.

4.4.3 Relationship to bubble nucleation

The distribution of bubble sizes reveals the frequent presence of a 1–3 bp bubble in kinked duplexes (Figure 4.10). Briefly, bubbles are defined as disruption of base pairs away from the duplex ends, while disruption at duplex ends is classified as fraying (Section 3.2.2). Unsurprisingly, bubbles are almost always co-localized with kinks, so we take total bubble size along the duplex as a proxy for bubbles at a kink. Results for total bubble size given the energetic and structural criteria for a kink are very similar.

4.4.4 Relationship to re-entrant continuous bending

As $R_{\text{ee}}/R_{\text{contour}} \to 0$, kinking appears to be less effective at relieving bending stress; for all but the shortest duplexes, there is re-entrant behavior from kinking back to continuous bending (Figure 4.3 (b)). A typical kinked structure is shown in Figure 4.11 (a); it consists of two long (relatively relaxed) duplex sections and a highly bent kink region. However, such a configuration cannot bring the ends into extremely close proximity ($R_{\text{ee}} < 2$ nm) due to steric effects. Instead, the blunt ends of the duplex must bend to face each other. This bending of the duplex ends fundamentally changes the constraints placed on the rest of the system, and makes kinking opposite the blunt ends less effective at relieving the stress (the system cannot localize the majority of bending stress into a single kink).

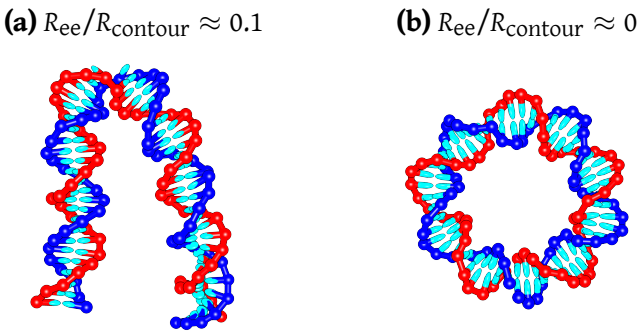


Figure 4.11 Re-entrant continuous bending at short end-to-end distance
OxDNA representations of a 52 bp duplex at small $R_{ee}/R_{contour}$. **(a)** Kinking helps alleviate bending stress by allowing for two more or less relaxed segment separated by an area of strong localized bending. **(b)** However, as $R_{ee}/R_{contour} \to 0$, due to excluded volume effects and the finite thickness of DNA relaxed segments may not overlap; therefore, the only way to satisfy the very small end-to-end separation is by bringing the duplex ends together, akin to a DNA minicircle. With this constraint, a single kink cannot relax the majority of the bending stress, unlike in case (a).

DNA adopts minicircle-like configurations for extremely small values of R_{ee}, as shown in Figure 4.11 (b). Minicircles can, of course, undergo kinking – however, the geometrical constraints mean that it is impossible to localize the majority of bending stress into a single kink,[24] unlike in systems with a slightly larger separation of the blunt ends. Thus we obtain the seemingly paradoxical result that kinking is more likely for systems that have some flexibility in their ends than for systems, such as minicircles, where the ends are constrained to be in extremely close proximity.

4.5 Conclusion

We have reported a full characterization of the thermodynamics of DNA bending in a coarse-grained model of DNA. A sequence-dependent transition from continuous bending to localized bending is observed.

Sequence-dependence is a generic property of kinking, which is manifest in this system as AT-rich sequences exhibiting localized bending at longer EED than GC-rich sequences (which are more difficult to kink). The transition from continuous to localized bending is due to a specific structural defect, a kink. Kinks are associated with the disruption of stacking and base-pairing interactions in the duplex.

Analysis of kinking reveals that one must be extremely careful in inferring free-energies of

kinking from specific systems. For example, unlike sequence-dependence, the reported length dependence of kinking is not a generic property of kinking, rather it is specific manifestation of the model system, which owing to the EED constraint preferentially kinks at the mid-point along the duplex. Thus, for the model system, short duplexes exhibit localized bending at longer EED than long duplexes, when normalized to the contour length. At sufficiently long duplex lengths ($\gtrsim$ 70 bp) there is no transition to localized bending.

A natural definition of ΔG_{kink} would be through the probability of spontaneous kink formation in relaxed DNA at a specific site. This should be a generic property depending on the local sequence, and would be the quantity measurable if relaxed duplexes were examined for kinking. If kinking is enforced through external bending of duplex DNA, however, the amount of work that must be done on the duplex state to enforce a transition, $\Delta G_{\mathrm{trans}}$, is potentially quite distinct from ΔG_{kink}. Reasons include the fact that work may also be done in bending the kinked state to reach the point of transition, and the fact that specific experimental systems (such as the molecular vise[22] in Chapter 5) may favour kink formation at certain locations whereas other systems, such as DNA minicircles (Section 7.4.4), may allow kinks anywhere along the strand. These considerations make direct comparisons of free-energies of kinking as inferred through simple analytic models difficult, hindering the comparison of kinking in distinct experimental systems.

Computational models such as oxDNA, in principle, allow comparison between distinct systems in which kinking is hypothesized, as they can be directly simulated with the same model and experimentally comparable observables can be recorded – factors such as the work done in bending a kink should be naturally incorporated, at least at a semi-quantitative level. Coarse-grained models parameterized to reproduce basic DNA mechanics and thermodynamics therefore offer a valuable tool for establishing whether proposed observations of kinking are consistent with each other and our current knowledge of DNA, as will be explored in subsequent chapters.

5 Model system: 'Molecular vise'

DNA bending is an important structural and thermodynamic property of DNA, and is non-trivial to experimentally probe. Using insight from a coarse grained model of DNA, we corroborate the results of recent 'molecular vise' experiments,[22] which provide one of the most direct probes of strong DNA bending currently available. Specifically, we reproduce the experimental high-salt buckling transition and approximate buckling length. We then directly probe the microscopic configurations responsible for this transition, validating the experimental interpretation of a kink-induced buckling transition.

5.1 Introduction

Fields *et al.* have developed a 'molecular vise' to probe the strong bending regime of DNA, providing evidence of a buckling transition[i] that cannot be explained by a worm-like chain (WLC) model.[22] Importantly, they infer kinking as the physical mechanism underlying this transition, a structural property that we can directly visualize. Having previously characterized DNA bending (Chapter 4), we are in a strong position to probe their system with simulation, and test whether their conclusions are reasonable given the known thermodynamic and mechanical properties of DNA, as encapsulated by our course-grained model.

The system they devise consists of a DNA hairpin and complement strand (Figure 5.1). The hairpin has two regions, a loop composed of N_{loop} nucleotides and a predominantly poly(AT) stem composed of 49 base pairs (39 AT base pairs adjacent to the loop and 10 terminal GC base pairs). Complement strands of different lengths are hybridized to the hairpin loop to form a

[i] Fields *et al.* have termed this a 'critical' buckling transition, but we find that this terminology may be ambiguous in the context of our work. They observe a thermodynamically smooth transition between (presumably) continuously bent and kinked states, which is *not* 'critical' in the classical thermodynamic sense, which refers to a second order transition of a bulk system.

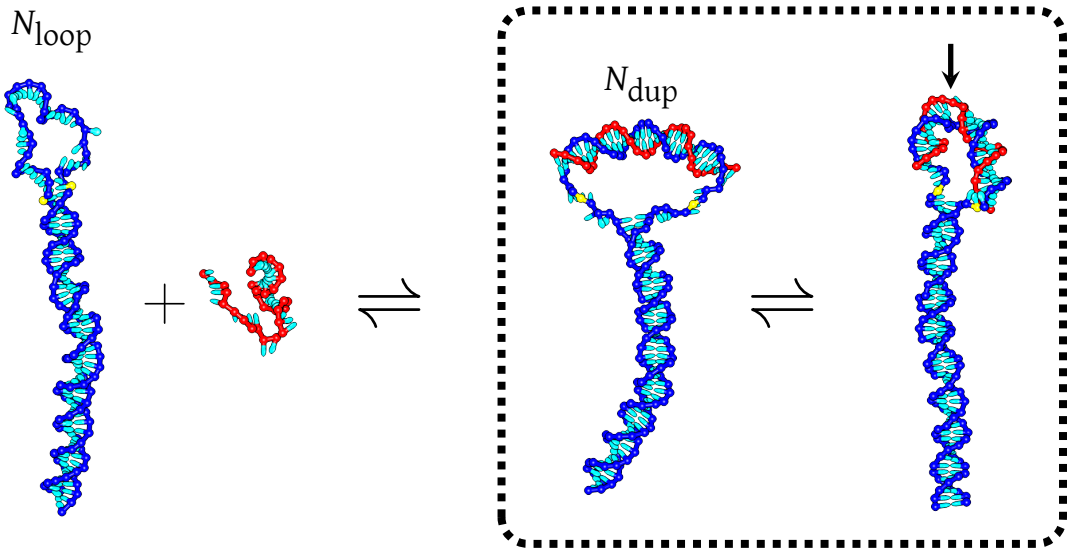


Figure 5.1 Schematic of molecular vise
OxDNA representation of the 'molecular vise' studied by Fields *et al.*[22] A duplex of N_{dup} base pairs is formed by hybridizing the N_{loop} nucleotide hairpin loop with a variable-length complementary strand. The duplex may be kinked (**arrow**) or unkinked, in which case there is an equilibrium between continuous bending in the duplex and unzipping in the stem. Stem-end thymines are used for fluorophore attachment and reporting hairpin separation (**yellow**). For the representative system shown, the hairpin stem is 49 base pairs, $N_{\text{loop}} = 36$ and $N_{\text{dup}} = 30$.

duplex of $N_{\text{dup}} < N_{\text{loop}}$ base pairs. The hairpin acts as a force clamp, imposing external bending stress on the duplex. Formation of a duplex within the loop may require opening part of the hairpin stem, which requires work via a salt-dependent unzipping force.[ii] Stem unzipping is monitored via Förster resonance energy transfer (FRET), reporting the separation between the fluorophores attached to the first and last bases of the hairpin loop (Figure 5.1).

This setup is powerful because it allows the observation of stem unzipping as a function of both the length of the complement strand and loop size. For example, at short N_{dup}, there is minimal stress on the system – it is necessary neither for the duplex to bend nor the stem of the hairpin to unzip. At long N_{dup} (comparable to N_{loop}), the system cannot maintain full base-pairing without bending the duplex. While at low-salt their data is in agreement with a WLC model, at high-salt their data suggests hairpin re-zipping at a shorter buckling length (of length N_{dup}) than implied by a WLC model.[11] Fields *et al.* infer that this discrepancy is due to the formation of a sharp kink, and not fraying, unzipping or continuous bending.

[ii] $\sim$9 pN at 23 °C and 200 mM monovalent salt for a 0 % GC-content sequence. For comparison, the value for a 100 % GC-content sequence is $\sim$21 pN.[67]

5.1.1 Specific aims

Having rigorously explored the thermodynamics of DNA bending in Chapter 4, we turn to the characterization of the 'molecular vise' of Fields *et al.*,[22] which provides an exquisite experimental probe of the behavior of duplex DNA under bending stress. We use oxDNA to explore the microscopic configurations that underlie their experimental observations. Our findings are compatible with the observations of Fields *et al.* at high salt concentrations.

5.2 Simulation methods

Simulations were conducted with the sequence-dependent parameterization of oxDNA using the virtual-move Monte Carlo (VMMC) algorithm[62] (Section 3.1.3) at $T = 25\,^{\circ}\mathrm{C}$. Error bars represent the standard error of the mean from 6 independent simulations. These simulations were performed without any biasing through the use of umbrella potentials. Further details of simulation procedure are given in Appendix A.2.

In our simulations, the thymines located at the first and last bases of the loop were chosen as the reporter for hairpin separation since they are the closest analogues to the experimental setup. The separation between these bases' centers of mass is reported as $R_{\mathrm{ee}}^{\mathrm{hairpin}}$.

5.3 Probing the molecular vise with oxDNA

We begin our discussion by considering the system with a loop length of $N_{\mathrm{loop}} = 36$ nucleotides and a variable length duplex of N_{dup} base pairs (Figure 5.2). OxDNA reveals distinct behavior for small, medium and large values of N_{dup}. These behaviors are related to the most prevalent deformation modes from the canonical relaxed duplex DNA structure at each given N_{dup} length. Possible deformations to consider involve unzipping of the hairpin stem, fraying, and bending or kinking of the duplex.

For short duplexes ($N_{\mathrm{dup}} < 20$) relative to the loop length $N_{\mathrm{loop}} = 36$, the hairpin separation $R_{\mathrm{ee}}^{\mathrm{hairpin}}$ is comparable to that of a hairpin without an annealed complement strand (Figure 5.2 (b)). Figure 5.2 (c) and Figure 5.2 (d) show that kinking and fraying of the duplex, and hairpin

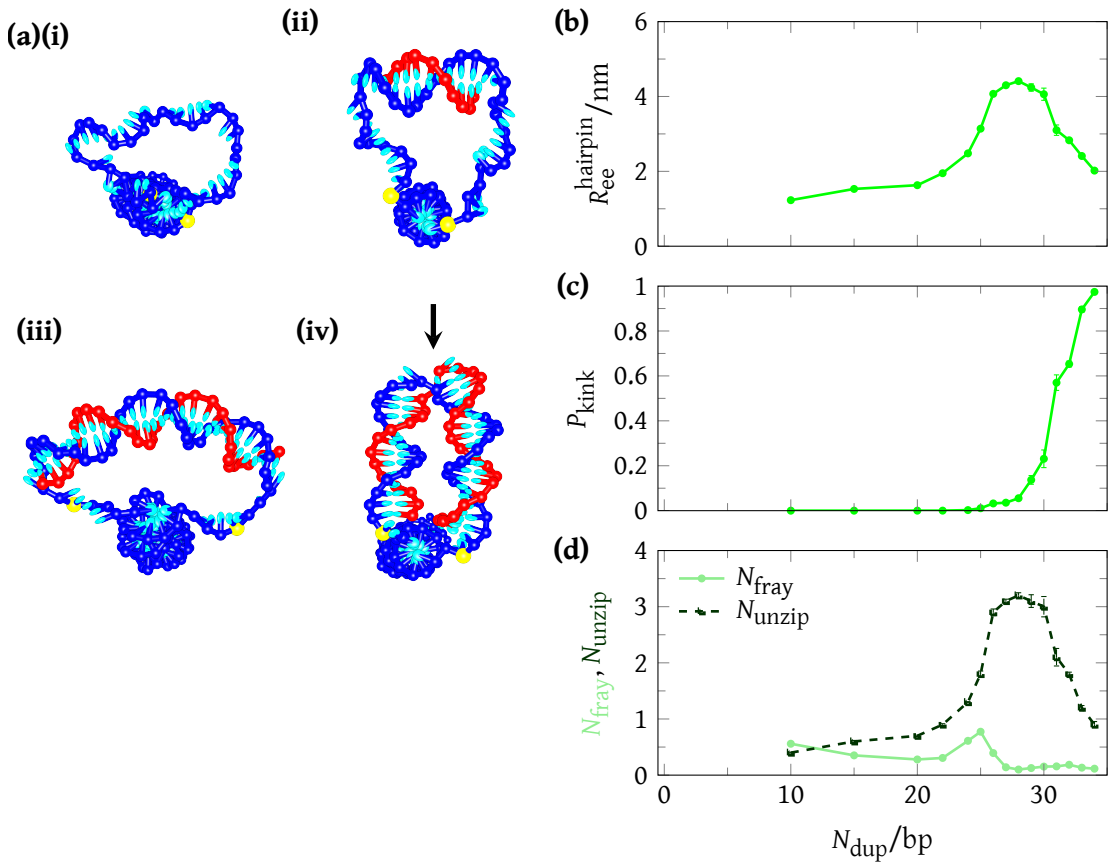


Figure 5.2 Properties of molecular vise, 36 bp
OxDNA structural properties of the duplex and hairpin stem for $N_{loop} = 36$ and variable N_{dup}. Points at $N_{dup} = 0$ are for the hairpin without an annealed complement strand. **(a)(i-iv)** oxDNA representations for $N_{dup} = 0, 10, 28, 34$, respectively. Stem-end thymines are used for fluorophore attachment and reporting hairpin separation (**yellow**). The kink in **(a)(iv)** is highlighted (**arrow**). **(b)** $R_{ee}^{hairpin}$ is the hairpin separation of end thymines, at the interface between the hairpin stem and loop. **(c)** P_{kink} is the probability of kinking in the duplex via the structural criterion. **(d)** N_{fray} and N_{unzip} are the number of disrupted base pairs via fraying in the duplex (**light green** circles) and unzipping the hairpin stem (**dark green** squares), respectively.

unzipping, are small, and simulation snapshots such as the ones shown in Figure 5.2 (a)(ii) do not show significantly bent duplexes. In this regime, the duplexes are too short to exert significant stress on the system.

For intermediate-length duplexes ($N_{\mathrm{dup}} = 20$–28), $R_{\mathrm{ee}}^{\mathrm{hairpin}}$ increases due to stem unzipping (Figure 5.2 (b)). The duplex is sufficiently long that the system cannot maintain full base-pairing without experiencing considerable mechanical stress. The system must either bend the duplex, or suffer a reduction in base pairing (either via unzipping the stem or fraying at the ends of the duplex). In fact, the system adopts configurations in which the deformation is shared between continuous bending of the duplex and unzipping, as can be seen from Figure 5.2 (a)(iii). The bending energy released for each base pair unzipped decreases as the duplex relaxes, whereas the cost of unzipping each base pair in the stem is approximately constant. The system can therefore minimize its free energy by adopting a state which combines unzipping with moderate bending. Unzipping the hairpin stem is favoured over fraying the duplex for two reasons. Firstly, stem unzipping contributes two nucleotides worth of single-stranded DNA to the hairpin loop, thereby reducing the tension, versus a single base for fraying the duplex. Secondly, the weaker poly(AT) stem melts easier than the more strongly bound GC-rich regions at both ends of the duplex. This was a deliberate design choice of Fields *et al.*.

For long duplexes ($28 < N_{\mathrm{dup}} \sim N_{\mathrm{loop}}$), $R_{\mathrm{ee}}^{\mathrm{hairpin}}$ decreases by re-zipping the stem and kinking, as reported in Figure 5.2 (b–d). A typical kinked structure is shown in Figure 5.2 (a)(iv). Increasing N_{dup} increases the bending stress in the system, which peaks at $N_{\mathrm{dup}} = 28$, leading to stronger unzipping and bending costs if the duplex does not kink. The cost of kinking, however, is more weakly dependent on N_{dup}, as was discussed in the general analysis of bending. Thus at sufficiently large N_{dup}, it is favourable to form a kink in the centre of the duplex, relieving the bending stress elsewhere and reducing the need to unzip.

5.3.1 Length dependence

As well as considering the behaviour at fixed N_{loop}, it is instructive to analyse the system as N_{loop} is varied. We focus our attention on the highly stressed regime of $N_{\mathrm{dup}} \sim N_{\mathrm{loop}}$ (Figure 5.3 (a) and Figure 5.3 (b)). We observe three regimes: For small $N_{\mathrm{loop}} < 30$, only fraying and continuous

bending are observed, with no signal for kinking in $R_{ee}^{hairpin}$. For $30 < N_{loop} < 60$, we see a sharp kink transition in which kinked states become dominant as $N_{dup} \rightarrow N_{loop}$. For larger N_{loop}, this transition is less pronounced, eventually fading away. Additional simulations at $N_{loop} = 66$–90 (Section 5.3.3), with $N_{dup} = N_{loop} - 2$, show the kinking probability decaying from ∼30 % to ∼2 %.

Why does a kink transition occur only for a finite range of N_{dup}? Unzipping of hairpin base pairs is extremely effective at relaxing stress for short duplexes: each unzipped base pair opened allows the duplex to straighten substantially. Thus at short lengths, stress is primarily relaxed by unzipping and duplex bending is small. Kinking does not play a role. As the duplex gets longer, each unzipped base pair relaxes the duplex less and less, as each disrupted base pair provides a smaller contribution relative to the overall length discrepancy between the intrinsic length of the duplex and the end-to-end distance enforced by the vise. Consequently, the bending of the duplex is enhanced, and eventually a kink transition is observed when this bending is sufficiently severe. For longer duplexes, the enthalpic cost of continuously bending to satisfy the separation imposed by the vise drops with duplex length (this can be shown directly using the Becker model[11]), whereas kink formation does not get easier. Thus the driving force for kinking is reduced, and kinking is absent at sufficiently long loop lengths.

Another notable point is that the maximum in $R_{ee}^{hairpin}$ prior to the kink transition (indicated by black dots in Figure 5.2 (b)) decreases with increasing N_{loop} at values of N_{loop} for which kinking is observable. Why should this be? As mentioned above, unzipping of stem base pairs is increasingly ineffective for larger systems, and bending of the duplex becomes increasingly easy. Therefore the degree to which the system unzips as $N_{dup} \rightarrow N_{loop}$ is reduced for larger systems, and $R_{ee}^{hairpin}$ does not grow as significantly prior to kinking.

5.3.2 Fraying in the molecular vise

Fraying involves melting base pairs at the ends of the duplex. Melted base pairs incur the free-energy cost of disrupting hydrogen bonding and stacking, but effectively increase the available loop length. To further the aim of experimentally probing DNA bending, the molecular

[iii] The molecular vise at $N_{bp} = 20$ is indeed shown in Figure 5.3 (a), but may not be visible because it does not kink for any N_{dup}.

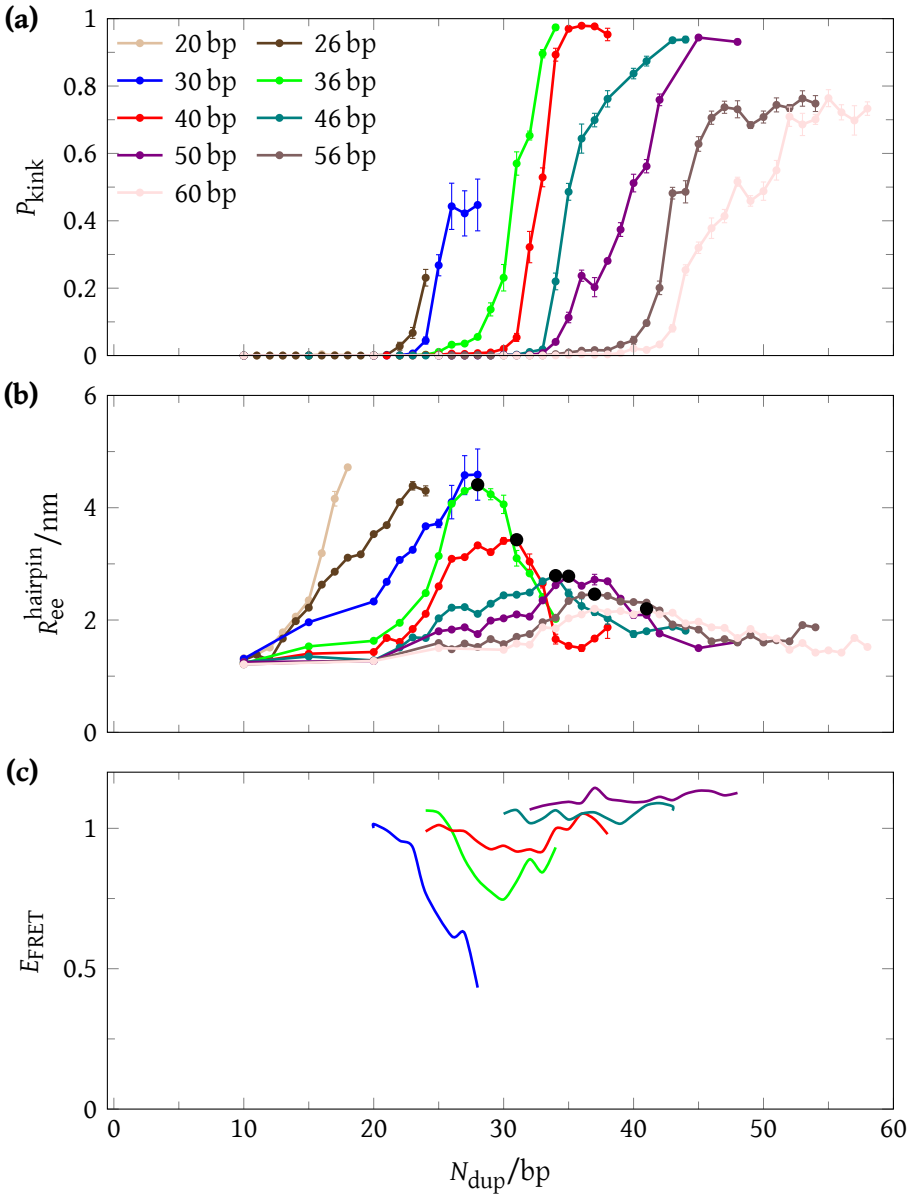


Figure 5.3 Properties of molecular vise, length-dependence
Structural properties for variable N_{loop}. **(a)** Kinking probability P_{kink} via structural criterion.[iii] **(b)** Hairpin separation $R_{\mathrm{ee}}^{\mathrm{hairpin}}$. Buckling lengths are highlighted (**black** marks). **(c)** The FRET efficiency E_{FRET} reproduced from molecular vise experiments. Available data for $N_{\mathrm{loop}} = 30, 36, 40, 46, 50$.[22]

vise is designed to minimize fraying. It is assumed that fraying will not be prevalent since the loop duplex ends are entirely GC base pairs, while the hairpin stem is poly(AT). Since the enthalpic cost of fraying depends on the number of base pairs melted, while the entropic gain depends on the length of the frayed segments, the free-energy cost of fraying is maximized by having fewer, longer stretches of single stranded DNA. For example, four frayed bases at one end of the duplex has a lower free energy than two bases frays at either end of the duplex. Both the relative ease of fraying the poly(AT) hairpin stem compared to the GC loop duplex, and the entropic benefit of maximizing fraying at one location, work in favour of selective unzipping in the stem.

5.3.3 Long length limit

In Figure 5.3 we showed the structural properties for variable loop length (N_{loop}) and duplex length (N_{dup}). The decay in the probability of kinking P_{kink} and the short-length plateau in hairpin separation $R_{\text{ee}}^{\text{hairpin}}$ both suggest the existence of a transition from kinking to continuous bending at long N_{loop}. Such a transition is indeed observed in our investigation of length-dependence in the free-energy of bending of intact duplexes (Figure 4.3), and would therefore be expected for the 'molecular vise'.

In Figure 5.4, by profiling the longest N_{dup} at a range of N_{loop} lengths, we examine the length-dependence of kinking, hairpin separation, fraying and stem unzipping. A transition from kinked to continuously bent states is observed at long N_{dup} (and therefore long N_{loop}), with the kink transition midpoint at $N_{\text{dup}} = 58$–64. By $N_{\text{dup}} = 88$, duplexes are almost exclusively continuously bent ($P_{\text{kink}} < 2\,\%$). Unsurprisingly, the length of the kink transition midpoint is consistent with our bending simulations (Figure 4.3).

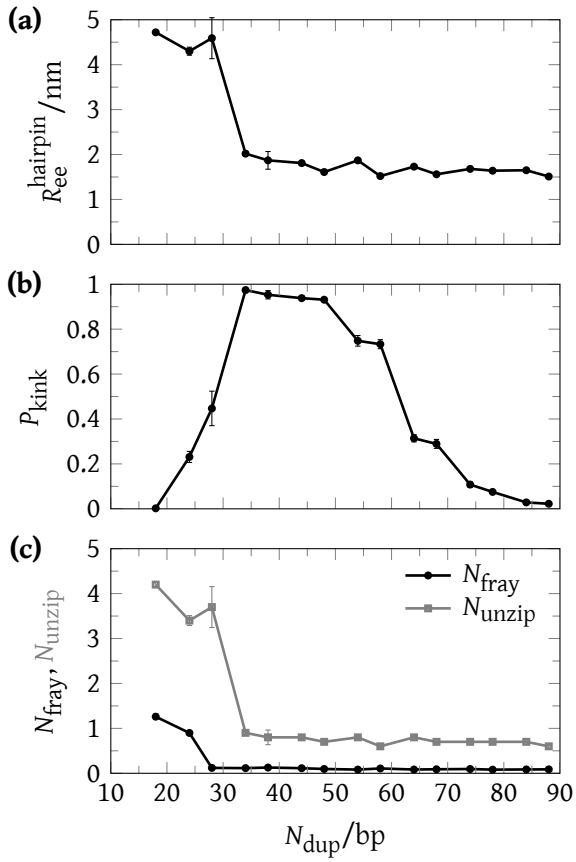


Figure 5.4 Properties of molecular vise in the long-duplex limit
Length-dependence of **(a)** hairpin separation $R_{ee}^{hairpin}$, **(b)** probability of kinking P_{kink}, as well as **(c)** fraying N_{fray} and stem unzipping N_{unzip}, at $N_{dup} = N_{loop} - 2 = 18$–$88$.

5.4 Comparison to experiment

5.4.1 FRET experiments

We now compare the oxDNA results to those of Fields *et al.*[22] Unlike oxDNA, Fields *et al.* could not directly observe kinked macrostates. Instead, they inferred kinking from non-monotonic behavior of the FRET signal between dyes attached to the first two bases of the loop (a proxy for fraying - Figure 5.2 (c)). They defined a buckling length at high-salt, which they interpreted as the onset of localized bending via kinking, as a local minimum in the FRET signal (Figure 5.3 (c). We can define an analogous quantity in oxDNA, the maximum in $R_{\mathrm{ee}}^{\mathrm{hairpin}}$ for a given N_{loop}, to compare with the FRET results.

In their higher salt experiments, Fields *et al.* saw evidence of fraying and no kinking for $N_{\mathrm{loop}} = 30$, identified non-monotonic behavior of FRET for $N_{\mathrm{loop}} = 36, 40$, and saw no substantial change in the FRET signal with N_{dup} for larger values of N_{loop}. Broadly speaking, these three regimes are consistent with our observations: kink transitions occur for intermediate values of N_{loop}, with only fraying at smaller values and no fraying or kinking at larger values. Quantitatively, we observe the beginning of a kink transition for $N_{\mathrm{loop}} = 30$, slightly earlier than Fields *et al.*, and see evidence of kinking at larger values of N_{loop}. The N_{dup} values at which kink transitions occur for $N_{\mathrm{loop}} = 36$ and $N_{\mathrm{loop}} = 40$ in oxDNA are within two base pairs of those inferred by Fields *et al.*, as can be seen in Figure 5.2 (c). Clearly, however, inferring the 'exact' onset of kinking is not trivial from the experimental data.

The quantitative consistency of results may be even stronger than immediately apparent. OxDNA is parameterized at $[\mathrm{Na}^{+}] = 500\,\mathrm{mM}$, meaning that the stem of the vise should be somewhat more resistant to fraying than in the experiments, which were performed at $[\mathrm{Na}^{+}] = 250\,\mathrm{mM}$, perhaps explaining the evidence of kinking for $N_{\mathrm{loop}} = 30$ that is absent in the work of Fields *et al.* Consistent with this argument, Fields *et al.* do not find evidence for kinking at all at much lower salt concentrations, when the stem of the vise is less stable. Slightly smaller values of N_{dup} at which buckling occurs for a given N_{loop} in oxDNA may also be explicable in this way.

Our results for larger values of N_{loop} also suggest that although kinking occurs, it does so following an increasingly small amount of unzipping. This is actually consistent with the

observations of Fields *et al.* – compare the drop in FRET prior to kinking for $N_{\text{loop}} = 36$ and $N_{\text{loop}} = 40$.[iv]

5.4.2 Limitations of FRET with respect to the molecular vise

We posit that unzipping prior to kinking may be undetectable through the methodology of Fields *et al.* for $N_{\text{loop}} \gtrsim 40$, who use gel-based ensemble FRET. One advantage of their method is that, based on differences in electrophoresis mobility with respect to a known DNA-ladder, they have access to a quality internal control for complement-loop hybridization in the formation of the loop duplex. Additionally, gel-based assays, as opposed to single-molecule assays, report on the ensemble, which is useful for making qualitative observations on thermodynamically smooth transitions.

Unfortunately, there is a highly non-linear relationship between FRET signal and fluorophore separation (Section 10.4.1.1), which is roughly equivalent to $R_{\text{ee}}^{\text{hairpin}}$. FRET efficiency scales as $1/r^6$ with fluorophore separation, and by definition, the FRET efficiency E is in the range $E = 0$–1 with a mid-point of $E = 0.5$ at the Förster radius, R_0. Due to this non-linearity, FRET is sensitive to distance changes at fluorophore separation close to the R_0; however, it also quickly saturates to its maximum and minimum at $E = 1$ and $E = 0$ respectively. Given that the Cy3B/Alexa647 fluorophore pair used to probe hairpin unzipping has an $R_0 = 5.7$–6.7 nm, FRET reporting is most sensitive at $r \approx 4$–9 nm ($E = 0.10$–0.90). Changes in FRET efficiency outside of this region will be difficult to distinguish experimentally with these fluorophores (Cy3B/Alexa647).

While Fields *et al.* observe constant high FRET at high-salt for $N_{\text{loop}} = 46$ and $N_{\text{loop}} = 50$, these lengths may indeed have undetected hairpin unzipping at a difficult to detect distance of $R_{\text{ee}}^{\text{hairpin}} < 4$ nm. As at $N_{\text{loop}} = 36$, there may exist a transition to the kinked state at these lengths. While Fields *et al.* make strong claims about the high-salt unbent to kink transition, they are silent on the kinked to continuously bent transition, possibly in respect to this experimental limitation.

OxDNA reports $R_{\text{ee}}^{\text{hairpin}} < 3$ nm for $N_{\text{loop}} \geq 46$, suggesting that it is possible that kinking occurs

[iv] Fields *et al.* suggest that kinking occurs for both $N_{\text{loop}} = 36$ and $N_{\text{loop}} = 40$, yet the FRET signal is stronger for $N_{\text{loop}} = 36$. It is plausible that this trend continues, with kinking present for the longer $N_{\text{loop}} = 46$ and $N_{\text{loop}} = 50$ lengths – kinked, but not detected by FRET with their flourophores, as we will argue in Section 5.4.2.

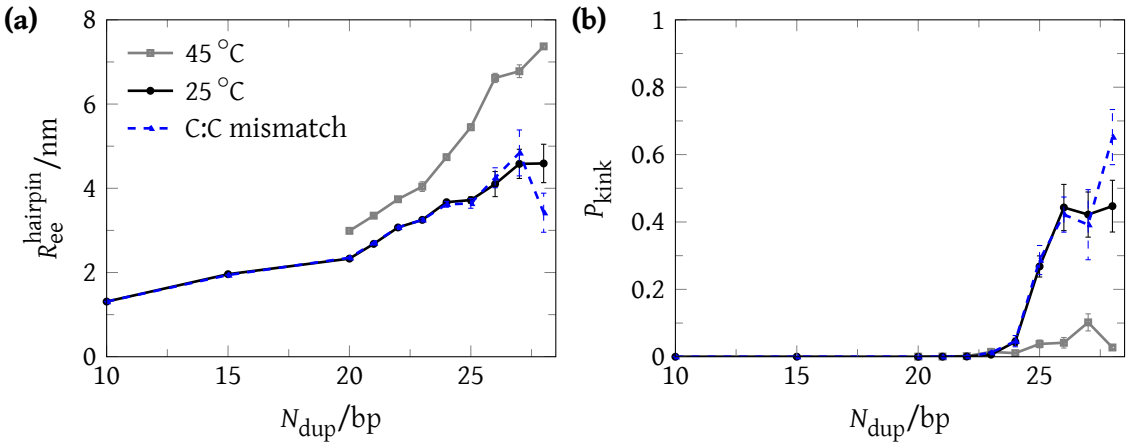


Figure 5.5 Properties of molecular vise, temperature-dependence
Temperature-dependent structural properties **(a)** hairpin separation and **(b)** kinking for $N_{loop} = 30$ at $T = 25\ ^{\circ}C$ (**black** circle) and $T = 45\ ^{\circ}C$ (**gray** square). The C:C mismatch at $T = 25\ ^{\circ}C$ is also shown (**blue** triangle).

in experiment at larger values of N_{loop} without sufficient change in FRET signal for it to be observable. OxDNA does suggest, however, that kinking is never observed for N_{loop} large enough that $N_{dup} \sim N_{loop}$ duplexes undergo continuous bending with sufficient ease.

5.4.3 Temperature-dependence and mismatches

We note that temperature is not well controlled in the Fields *et al.* gel-based experiments due to electrophoresis-induced heating of the gel running buffer. To explore the implications of electrophoretically-induced buffer heating, temperature-dependence was also explored with additional simulations at 35 °C (data not shown) and 45 °C (Figure 5.5). Increasing temperature, which increases the entropic cost of base pairing, reduces the unzipping force associated with the hairpin.[68] Unsurprisingly, we find negligible kinking at $T = 45\ ^{\circ}C$ for $N_{loop} = 30$, consistent with the reasoning that weakened base pairing in the stem leads to easier unzipping. The absence of a kink transition is reflected by a monotonically increasing $R_{ee}^{hairpin}$.

We also briefly explored the role of a C:C mismatch in the $N_{loop} = 30$ system (Figure 5.5). Since the kink transition is incomplete for $N_{loop} = 30$, the presence of a mismatch only slightly increases kinking. We would expect that a mismatch in a system with a complete kink transition ($N_{loop} = 36$–60) would lower the free-energy barrier to the kinked state, thereby decreasing the buckling length.

5.4.4 Summary

As with all experiments, the methods of Fields *et al.* are not without limitations. Temperature, an important determinant of DNA behavior, is not well controlled in gel based experiments due to electrophoresis-induced heating of the buffer. Due to this heating, evaporation increases the ion concentration over time. this effect can be mitigated by buffer addition to maintain a constant run buffer volume. Mapping from a qualitative to quantitative measure of FRET is non-trivial, requiring a number of assumptions and controls. While FRET efficiency is proportional to fluorophore separation, a quantitative mapping to distance is usually not possible.

The molecular vise of Fields *et al.* is a sensitive experimental measure of the duplex DNA buckling transition. While the evidence supporting their claim of a kinking transition at high-salt at the buckling length of 30 bp is strong, unbent (hairpin stem unzipping), kinked and continuously bent structures cannot be directly observed. Indeed, observed FRET efficiency reports on an ensemble of presumably unbent and kinked structures, reaching saturation at short and long complement lengths, respectively.

In addition to FRET experiments with the molecular vise, Fields *et al.* offer simulation and further biochemical probes to support their claims of a low-salt continuous bending transition and a high-salt kink transition, these include:

(i) Simulation at known DNA persistence lengths of 40–50 nm do not support a continuous bending transition at 30 bp: an implausibly low DNA persistence length of 25 nm would be required.
(ii) A single stranded DNA specific endonuclease cleaves the putative kinked molecules but not the continuously bent molecules.
(iii) Putative kinked molecules are more electrophoretically mobile than their continuously bent counterparts, reflecting the expected smaller effective size of the kinked molecules.

OxDNA is also not without limitation. For example, as oxDNA is parameterization to high-salt, we compare only to the Fields *et al.*[22] high-salt condition; their low-salt limit is not investigated. To redress this short-coming, a salt-dependent version of oxDNA using the Debye–Hückel approximation is under development. Additionally, specific sequence-dependent motifs are not

represented by oxDNA (Section 3.1.1). This is less of a concern since pathological sequences were explicitly avoided by Fields *et al.* to aid interpretability.

Taken together, the evidence for kinking is convincing, particularly when examined in light of other studies more indirectly examining DNA in the extreme bending regime, such as cyclization,[12–15] multimerization,[17] AFM,[27,28] and cryo-EM.[18] Even in model systems specifically designed to probe the presence of a buckling transition, such as the strained-duplex (Chapter 8),[19–21] the Field *et al.* molecular vise is methodologically superior.

5.5 Conclusion

We have simulated the 'molecular vise' of Fields *et al.*,[22] a direct probe of DNA bending. Our results are consistent with experiment, supporting their interpretation of non-WLC behavior via a specific structural defect, *i.e.* a kink, and providing a detailed mechanistic perspective on the data. Specifically, we reproduce their high-salt kink transition at an appropriate buckling length, and directly probe the microscopic states responsible for the transition, further validating the structural interpretation of experimental observations. We also propose that duplex kinking might be present for longer duplexes than can be directly inferred from the experimental data, due to the limits of FRET precision and minimal unzipping prior to kinking.

OxDNA is by no means a perfect model, and the level of agreement with Fields *et al.* should not be over-interpreted. Although reproduction of this data is evidence that the model is reasonable, a number of modelling simplifications, and the $2\times$ difference in salt conditions in experiment and oxDNA, might give rise to small quantitative deviations. The important point is that the general agreement between the experimental data and oxDNA helps to show that kinking, as inferred by Fields *et al.*, is consistent with the known thermodynamics and mechanics of DNA *vis-à-vis* the predictions of the oxDNA model. Furthermore, as oxDNA can also be used to probe kinking in DNA cyclization (Chapter 6) and minicircles (Chapter 7), we conclude that the evidence of kinking in the systems investigated are broadly consistent. This is particularly important given the controversy surrounding the kink transition.[16,28] We note that our observation that kinking may go undetected for intermediate loop sizes in the setup of Fields *et al.* is potentially important in this regard; otherwise the absence of kinking for systems of intermediate N_{loop}

($N_{\text{loop}} = 46, 50$) might be surprising when compared to digestion assays on minicircles,[18,36] particularly as simple mechanical arguments suggest that the geometry of torsionally relaxed minicircles is less condusive to kinking than the geometry of the molecular vise.

As of yet, direct experimental observation of kinked DNA macrostates has been elusive; however, simulation helps bridge the gap between indirect experimental observations and the fundamental thermodynamics underlying DNA bending.

6 Model system: DNA Cyclization

DNA cyclization is a powerful technique for examining DNA bending, reporting agreement with the worm-like chain (WLC) model at small to moderate bending fluctuations. Using a course-grained simulation we investigate the thermodynamics of DNA cyclization, finding agreement with worm-like chain (WLC) prediction to lengths of 80 base pairs. We find that a specific structural defect, a kink, is responsible for non-WLC behavior at lengths below 80 base pairs. Our findings are consistent with ligase and Förster resonance energy transfer (FRET)-based measurements of DNA cyclization.

6.1 Introduction

6.1.1 Historical context

DNA cyclization is a convenient model system used to probe DNA bending. Cyclization experiments were first reported in 1966, albeit on 48.5 kbp λ-DNA.[69,70] In 1981, Shore *et al.* developed a method to probe the bending of shorter 126–4361 bp fragments,[71] later noting periodicity in the cyclization efficiency of 237–254 bp fragments.[12]

More recently, there has been a particular interest in probing the cyclization of sub-persistence length DNA, to explore whether this regime is accurately described by the WLC model. For example, in 2004 Cloutier & Widom (C&W)[14] challenged the conventional wisdom of WLC flexibility established by Shore *et al.*,[12,71] claiming much greater than predicted cyclization efficiency in 93–95 bp DNA fragments. This controversial finding spurred debate on the characteristic length at which DNA cyclization efficiency deviates from the predictions of the WLC model. Despite much experimental[13–16,72] and theoretical effort,[24,26,30,73–76] a consensus has not yet been established.

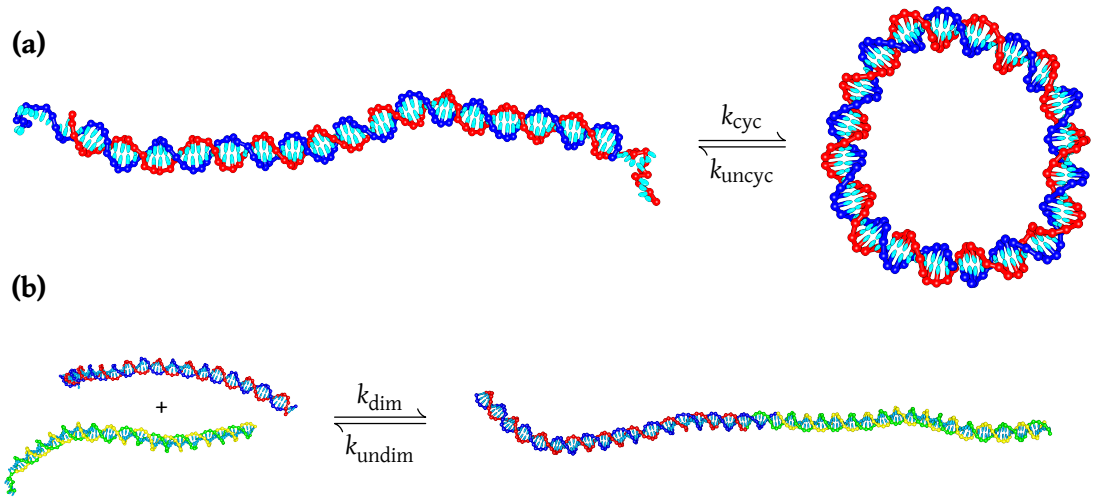


Figure 6.1 Schematic of cyclization and dimerization
Schematic representations of **(a)** cyclization where k_{cyc} and k_{uncyc} are the forward and reverse rate constants respectively, and **(b)** dimerization where the rate constants are k_{dim} and k_{undim}. Note that for the dimerization system there is only one complementary sticky end per monomer, the other end being blunt to allow for only one reaction product, a linear dimer. Figures are oxDNA representations for monomers of length $N_{bp} = 101$, including complementary sticky ends of length $N_s = 10$. (Dimer length is $2N_d + N_s$).

6.1.2 Cyclization experiments

A typical cyclization experiment, as depicted in Figure 6.1, uses a cyclization substrate with complementary sticky ends, N_s bases in length, on both ends of a N_d base pair duplex. Cyclization leads to the formation of a N_{bp}-base pair duplex, where $N_{bp} = N_s + N_d$. The resultant structure is not a closed minicircle – two backbone 'nicks' are present at either end of the sticky ends. Either the forward rate or equilibrium constant of the cyclization reaction is reported. Experiments differ in how exactly they measure cyclization: methods based on ligation,[12–14,71] Förster resonance energy transfer (FRET)[15,16] and multimerization[17] have been reported.

In ligase-based experiments, cyclized molecules are kinetically trapped by ligation of the two backbone nicks. The open and ligated cyclized molecules can then be resolved by gel electrophoresis and the concentration of different products measured. FRET-based experiments can be performed in equilibrium, with the molecules allowed to cyclize and uncyclize indefinitely. Fluorophores are attached to both ends of the molecule as FRET reporters: a high FRET signal will be reported when the duplex ends are in close proximity (cyclized), low FRET when apart (open). Although non-WLC behavior has been suggested by the ligase-based experiments of

Cloutier & Widom (C&W)[14] and the FRET-based experiments of Vafabakhsh & Ha (V&H),[15] these results have been contested.[13,26] There is not yet a consensus on their interpretation.

6.1.3 *j*-factor

6.1.3.1 Concept

Cyclization efficiency is traditionally reported in terms of a *j*-factor, first introduced by Jacobson & Stockmayer,[1] which is a measure of the effective local concentration of duplex ends at zero end-to-end distance separation. The *j*-factor enables the ring closure probability to be calculated, and importantly, may be related to a ratio of equilibrium constants:

$$j\text{-factor} = j_{\mathrm{eq}} \equiv \frac{K_{\mathrm{eq}}^{\mathrm{cyc}}}{K_{\mathrm{eq}}^{\mathrm{dim}}}, \tag{6.1}$$

where $K_{\mathrm{eq}}^{\mathrm{cyc}}$ and $K_{\mathrm{eq}}^{\mathrm{dim}}$ are the equilibrium constants for cyclization and dimerization, respectively (for further discussion, see Appendix A.3.3).

Multimerization of the cyclization substrate yields a mixture of predominantly linear and circular dimers.[13,14] To avoid this complication, a separate dimerization substrate may be prepared with the same sequence as the cyclization substrate, but with only one N_{s} base complementary stick end per molecule[15] (Figure 6.1 (b)). Following hybridization, the total length of the system is then $2N_{\mathrm{d}} + N_{\mathrm{s}}$ base pairs, with blunt as opposed to sticky ends. The consequences of this choice for $K_{\mathrm{eq}}^{\mathrm{dim}}$ and hence j_{eq} are discussed in Appendix A.3.3.

6.1.3.2 A worm-like chain (WLC) model for DNA cyclization

Assuming the contribution from base-pairing to $K_{\mathrm{eq}}^{\mathrm{cyc}}$ and $K_{\mathrm{eq}}^{\mathrm{dim}}$ is the same, the WLC model can be used to estimate j_{eq}. A common assumption is that the cyclized state is fully stacked, with coaxial stacking across the two nicks. For this situation, the analytic expression derived by Shimada & Yamakawa (S&Y)[8] (Section 2.2.3), which includes both the bending energy cost of bringing the two ends together and the twist energy cost of bringing the two helices ends into register, is the most appropriate.

To avoid the need to address the thermodynamics of base-pairing, S&Y assume a duplex without complementary sticky ends, which is a good assumption in the limit where base pairing is identical for the cyclized and dimerization substrates, and the molecule is much larger than the sticky ends. The Shimada-Yamakawa model only considers continuous duplex DNA, does not consider the effects of single-stranded hybridization and has no description for structural defects, such as kinking.

6.1.4 Thermodynamic versus kinetic *j*-factor

Some of the confusion surrounding claims of cyclization efficiency greater than that predicted by the WLC model revolves around the use and interpretation of *j*-factors. While the *j*-factor relation using the ratio of cyclization to dimerization *equilibrium* constants is well established, the ratio of forward *rate* constants is usually measured experimentally.

6.1.4.1 Ligase-based assays

In the case of ligase-based assays (reviewed in Peters *et al.*[77]), the *j*-factor is given as

$$j_{\mathrm{dyn}}^{\mathrm{ligase}} = \frac{k_{\mathrm{cyc}}^{\mathrm{ligase}}}{k_{\mathrm{dim}}^{\mathrm{ligase}}} \approx j_{\mathrm{eq}}, \tag{6.2}$$

where $k_{\mathrm{cyc}}^{\mathrm{ligase}}$ and $k_{\mathrm{dim}}^{\mathrm{ligase}}$ are the forward rate constants for the formation of the *ligated* circle and dimer respectively.

In the experimental limit where the ligation rate is very slow compared to the rate constants for uncyclization (k_{uncyc}) and undimerization (k_{undim}), the concentrations of unligated circles and dimers will reach an equilibrium with reactants. If this condition is met, $j_{\mathrm{dyn}}^{\mathrm{ligase}}$ should be equivalent to j_{eq}. In practice, this limit is valid for low ligase concentrations ([Ligase] $\ll$ [DNA]) and short sticky ends.[13,75] The importance of this condition is illustrated by Du *et al.*,[13] who suggested that the non-WLC behavior in the C&W experiments was due to an insufficiently low ligase concentration (reviewed in Vologodskii *et al.*[26]).

An additional complication with ligase experiments relates to the structure of the substrate, specifically the DNA surrounding the nick. It is unclear whether the ligase will act on all nicked

duplexes, or only a subset which happen to be coaxially stacked, at one or both nicks. As we show in this work, to alleviate stress, cyclized systems are more likely than dimers to break coaxial stacking at a nick. The ligation rate of hybridizing complementary sticky ends may therefore vary substantially depending on the system.

6.1.4.2 FRET-based assays

The FRET experiments of V&H have the advantage of directly monitoring the transition between cyclized and open states. While it is possible to report thermodynamics from FRET experiments, V&H report dynamics:

$$j_{\mathrm{dyn}}^{\mathrm{FRET}} = \frac{k_{\mathrm{cyc}}}{k_{\mathrm{dim}}} \approx j_{\mathrm{eq}}, \tag{6.3}$$

where k_{cyc} and k_{dim} are the forward rate constants for the formation of the *unligated* circle and dimer respectively.

j-factors extracted from FRET-based experiments must also be interpreted with care. Making a comparison between $j_{\mathrm{dyn}}^{\mathrm{FRET}}$ and the S&Y prediction for $j_{\mathrm{eq}}^{\mathrm{WLC}}$, V&H make a claim of greater than WLC flexibility at $N_{\mathrm{bp}} \approx 100$ bp. However, $j_{\mathrm{dyn}}^{\mathrm{FRET}} \approx j_{\mathrm{eq}}$ only in the limit where $k_{\mathrm{uncyc}} \approx k_{\mathrm{undim}}$, a condition that V&H, as well as another more recent FRET experiment,[16] have shown not to be met.

Given that $k_{\mathrm{uncyc}} \neq k_{\mathrm{undim}}$, one should not expect $j_{\mathrm{dyn}}^{\mathrm{FRET}} \approx j_{\mathrm{eq}}$ at short N_{bp}. Thus, the observed deviation of $j_{\mathrm{dyn}}^{\mathrm{FRET}}$ from $j_{\mathrm{eq}}^{\mathrm{WLC}}$ is not necessarily an indication of non-WLC behavior.

6.1.5 Specific aims

The interpretation of cyclization experiments is not straightforward; in particular, the microscopic states, responsible for the putative non-WLC flexibility, cannot be directly observed. Additionally, establishing whether these observations are mutually consistent with the other systems for which non-WLC behavior has been reported is a challenging task due to the distinct interplay of mechanics, geometry and topology inherent in each experimental system, namely DNA minicircles,[18,36] a 'molecular vise',[22] and a 'strained duplex',[20,78] all of which are explored in Part I.

Fortunately, simulations can help bridge this gap. In this chapter we will probe cyclization equilibrium as a function of duplex length. The aim of our simulations will be to establish: (i) whether experimental results are truly indicative of non-WLC behavior; (ii) whether the behavior can be attributed to certain types of structure, such as 'kinks';[25,65,66] (iii) whether the results are consistent with the other systems showing evidence of non-WLC behavior; and (iv) whether the inferred occurrence of disruptions to duplex structure is consistent with our current understanding of DNA thermodynamics. OxDNA is particularly well-suited for this study, as it provides a good description of both single- and double-stranded DNA, including hybridization thermodynamics, persistence length, torsional modulus and basic structure, all of which play a role in cyclization.

6.2 Simulation methods

6.2.1 Cyclization simulations

Simulations of the cyclization equilibrium were performed with the virtual-move Monte Carlo (VMMC) algorithm[62] (Section 3.1.3) at 298 K. As the free-energy barrier between typical open and cyclized states is large (Figure 6.7), and the transition between the two macrostates constitutes a rare-event, umbrella sampling[63] (Section 3.1.4) was employed to sample the barrier crossing in reasonable computational time.

We use a two-dimensional order parameter $Q = (Q_{ee}, Q_{bp})$ to characterize the transition. Q_{ee} is a discretized measure of the distance of closest approach between the complementary sticky ends. Q_{bp} is the number of base pairs formed between complementary sticky ends, where $0 \leq Q_{bp} \leq N_s$. Further details of the definitions are available in Appendix A.3.1.

To further improve computational efficiency, the umbrella sampling was windowed to separately sample the open and cyclized states of each molecule. For the window associated with the open state, the system was restricted to $Q_{bp} = 0$; for the window associated with the cyclized state, the system was restricted to $Q_{ee} = Q_{ee}^{min}$ (the value corresponding to the shortest distances between sticky ends). Simulations were run until convergence to within ± 5 % for each window.

The sampling windows overlap at $Q = (Q_{ee}^{min}, Q_{bp}^{0})$; results were combined by normalizing each

window so that free energies were equal for this value of the order parameter. As there is only one well-defined overlap between the values of the order parameters for both windows, more complex approaches, such as the weighted histogram analysis method,[60,61] were unnecessary. Further details are available in Appendix A.3.1. The simulation of dimerization equilibrium is roughly analogous to cyclization, and is elaborated in Appendix A.3.2. To compute the equilibrium constants, we deemed all states with $Q_{\mathrm{bp}} \geq 1$ to contribute to the cyclized and dimerized states (Appendix A.3.3).

To highlight the basic thermodynamics of DNA cyclization, we use oxDNA's average-base parameterization; where appropriate, we use the sequence-dependent parameterization to more directly compare with experiment. A complete list of sequences is available in Table A.2. To further simplify sampling, we forbid the formation of base pairs that are not intended in the design of the system (non-native base pairs). Error bars represent the standard error of the mean from 5 independent simulations.

6.2.2 Use of a structural kink criterion

In this chapter, we have opted to report kinking using the structural criterion. Such a choice will underestimate the prevalence of kinks, as it is possible for sharply bent configurations to fail to satisfy the criterion. However, for indicative purposes, such a choice is sufficient. More details, along with subtleties related to kinking at a nick, are discussed in Section 3.2.2.1. An alternative energetic definition of kinking, defined by a disruption in the stacking interaction between consecutive bases, is also valid; however, for the oxDNA model, structural and energetic definitions usually yield very similar results (Section 4.4.1).

6.3 Thermodynamic *j*-factor

6.3.1 Overview

We simulate a large range of system sizes; some illustrative configurations are shown in Figure 6.2. Quantitatively, we first consider the behavior of j_{eq} as a function of length. For the dimerization system we only computed the equilibrium constant $K_{\mathrm{eq}}^{\mathrm{dim}}$ for a few lengths

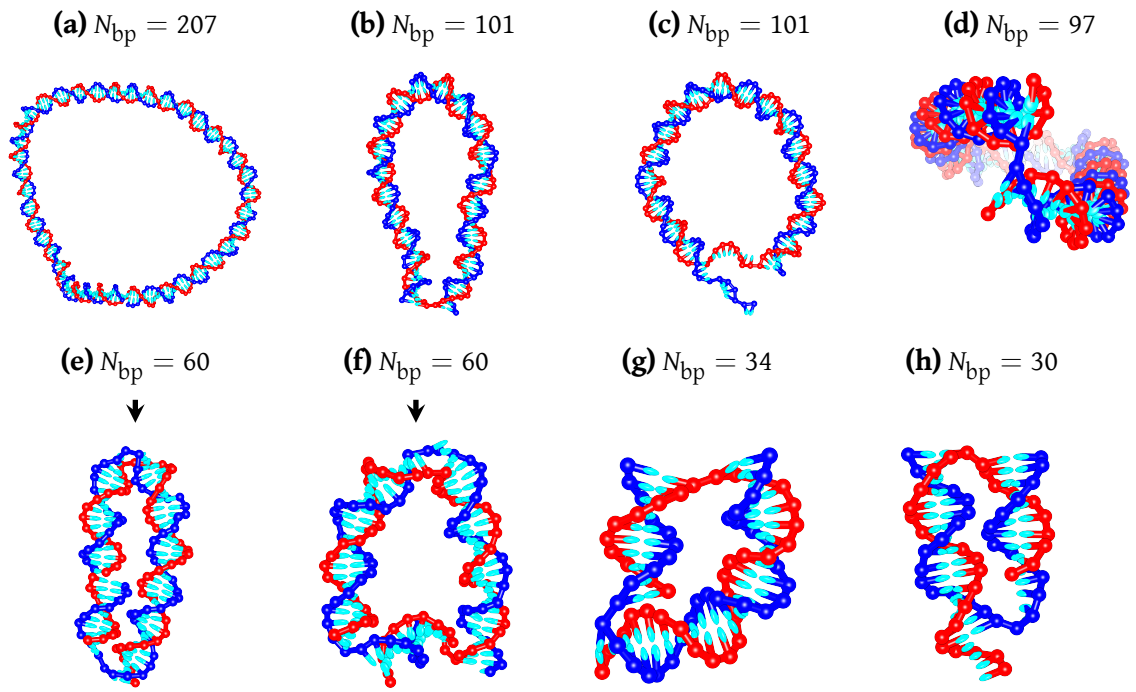


Figure 6.2 Menagerie of cyclized configurations
OxDNA representations of different cyclized configurations. Kinks in the duplex, which disrupt stacking and induce a 1–3 bp bubble, are indicated with an arrow. **(a)** A fully stacked 'circle'. **(b)** A 'teardrop' configuration with a kink at one of the nicks. **(c)** A $Q_{bp} = 1$ 'transition state' configuration. **(d)** Teardrop configurations with $N_{bp} = (n + 1/2) \times$ pitch length can reduce the stress associated with chain continuity at the nick by out-of-plane bending. Configuration with a kink in the duplex and either **(e)** a kink at one of the nicks or **(f)** kinks at both nicks. For short duplexes where N_d is not much larger than N_s, the sticky ends can associate either by **(g)** relatively minor bending of the duplex or **(h)** fraying a few base pairs.

(monomers $N_{bp} = 30, 67, 73, 101$). As expected, we found K_{eq}^{dim} to be largely length-independent (Section 6.3.4, Table 6.1); therefore, we use an average value of $K_{eq}^{dim} = 0.92 \pm 0.20 \times 10^{12}$ M^{-1} in our j-factor calculation. In contrast, we found K_{eq}^{cyc}, and thereby j_{eq} (Equation (6.1)), to vary substantially with length.

In Figure 6.3, we show j_{eq}^{oxDNA} values for 81 different lengths in the range $N_{bp} = 30$–207 for fixed $N_s = 10$ using the average-base parameterization of oxDNA. These results are compared to j_{eq}^{WLC} predictions based on the Shimada & Yamakawa (S&Y) expression[8] using previously calculated values for structural and mechanical properties of oxDNA.[52] The S&Y expression is appropriate for the formation of a fully stacked 'circle' configuration with coaxial stacking at both nicks (Figure 6.2 (a)). Note that the comparison in Figure 6.3 is fit-free. In addition to investigating cyclization as a function of intact duplex DNA length, we also explore the thermodynamics of changing the length of the single-stranded sticky ends (Section 6.5.2), as well as introducing

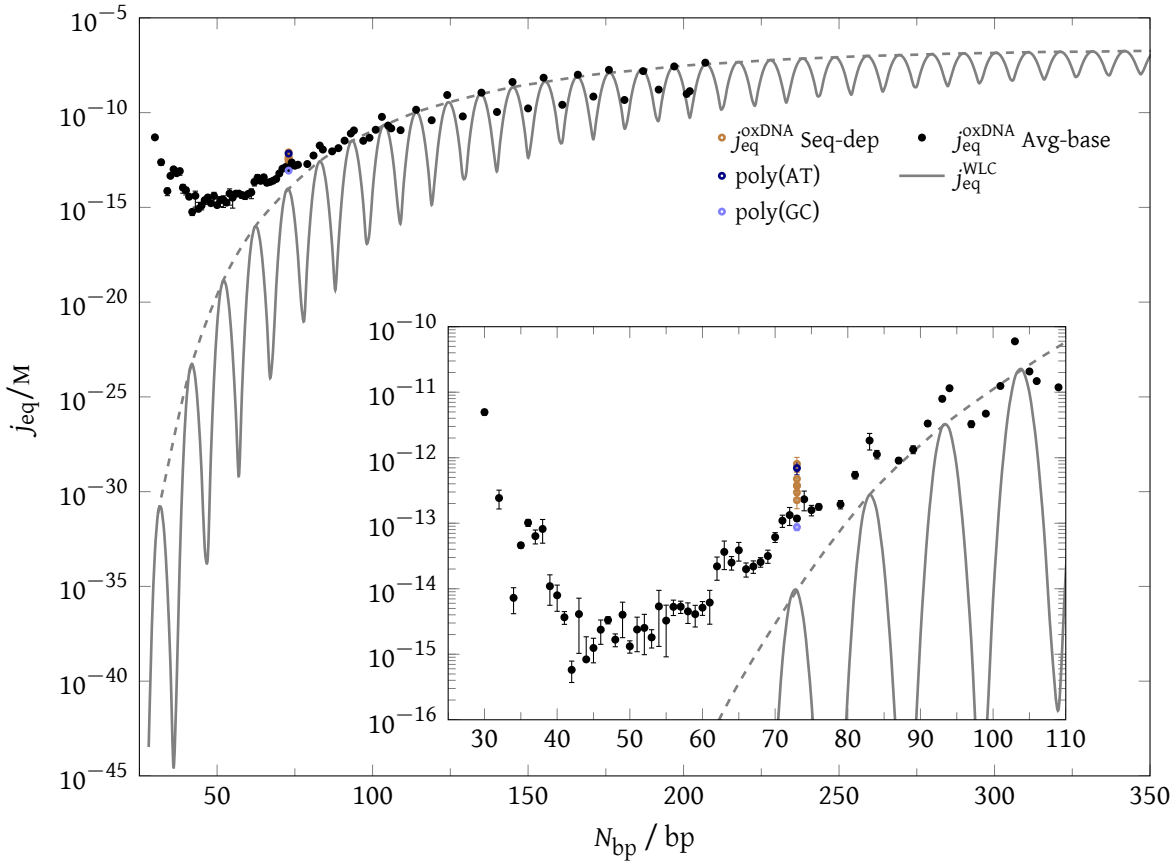


Figure 6.3 Thermodynamic *j*-factor
j_{eq}^{oxDNA} for the oxDNA average-base parameterization (**black** circles) as a function of N_{bp} for $N_s = 10$. For comparison, the Shimada & Yamakawa (S&Y) WLC prediction $j_{eq}^{WLC\,8}$ (**gray** solid line) is plotted using values of torsional stiffness (4.75×10^{-28} J m^{-1}), persistence length (41.82 nm) and pitch length (10.36 bp/turn) appropriate to oxDNA.[52,57] The dashed line gives the maxima envelope for the S&Y prediction (**gray** dashed line). Sequence-variation in j_{eq}^{oxDNA} is shown using the sequence-dependent parameterization with six $N_{bp} = 73$ bp sequences from V&H (**brown** open marks).

point-defects such as nicks and mismatches (Section 6.5.3).

The behavior of the S&Y expression is well understood. In the regime of interest, shortening N_{bp} tends to make cyclization less favourable as a result of increased bending stress within the duplex. This problem becomes particularly acute for DNA lengths below the persistence length (41.82 nm/126 bp for the curve plotted in Figure 6.3). On top of this systematic behavior, a periodic oscillation is associated with the need to over- or under-twist the duplex when the natural twist is not commensurate with that required to form a closed circle. The magnitude of this oscillation increases at shorter N_{bp} because a greater twist per base pair is required.

When considering the behavior of j_{eq}^{oxDNA} in light of the S&Y expression, three length-scale

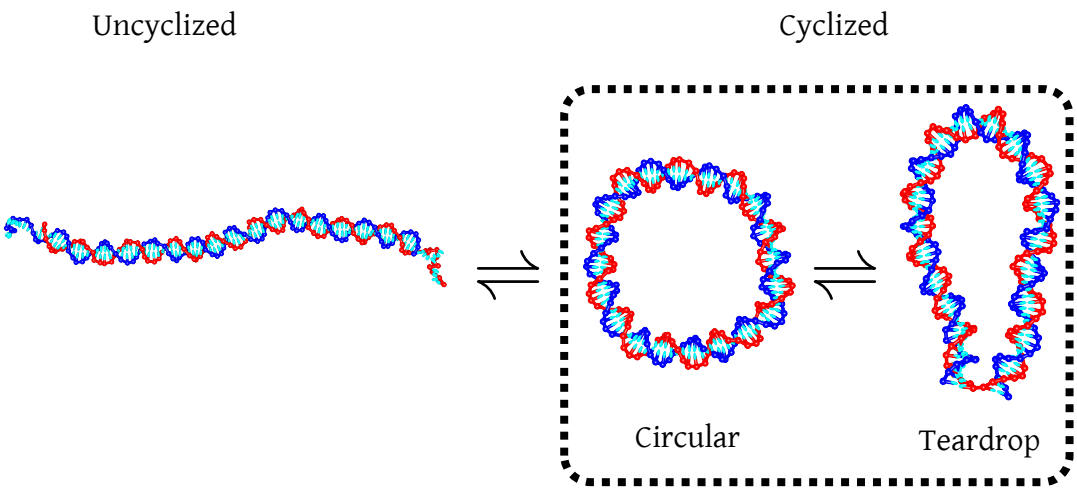


Figure 6.4 OxDNA representation of cyclization equilibrium between open and cyclized states
Representative snapshots of the cyclization equilibrium between open (left) and cyclized (right) states for an $N_{bp} = 101$ ($N_s = 10$, $N_d = 91$). The cyclized state is a mixture of 'circular' molecules, where both nicks are coaxially stacked, and 'teardrop' molecules, where a kink is present at one of the nicks.

dependent regimes become apparent: long ($N_{bp} \gtrsim 80$), intermediate ($N_{bp} \approx 45$–80 bp) and short ($N_{bp} \lesssim 45$).

6.3.1.1 *Long length limit (> 80 bp)*

In the long length regime ($N_{bp} \gtrsim 80$), oxDNA reproduces the periodic oscillations predicted by the S&Y expression, and values of j_{eq} coincide at the maxima of these oscillations. However, even at the longest N_{bp} lengths we consider, the magnitude of the oscillation is smaller for oxDNA than predicted by the S&Y expression. At shorter N_{bp}, the magnitude of the oscillation decreases for oxDNA, also at odds with the S&Y expression. The difference stems from the possibility of adopting alternative 'teardrop' configurations (Figure 6.2 (b)), in which most of the twisting stress, and some of the bending stress, can be relieved by kinking at one of the nicks, breaking coaxial stacking. Thus, the cyclized state is not monolithic, rather it is composed of circular and teardrop configurations (Figure 6.4).

The constraint of DNA continuity at the nick can be more easily satisfied by out-of-plane bending (Figure 6.2 (d)). The possibility of adopting this alternative configuration reduces the free-energy penalty for incommensurate values of N_{bp}, thus suppressing the oscillations. Further, at shorter N_{bp}, the bending stress increases more slowly for these kinked teardrop configurations than for the coaxially stacked circles. Thus, the difference in j_{eq} between the 'on-register' (coaxially

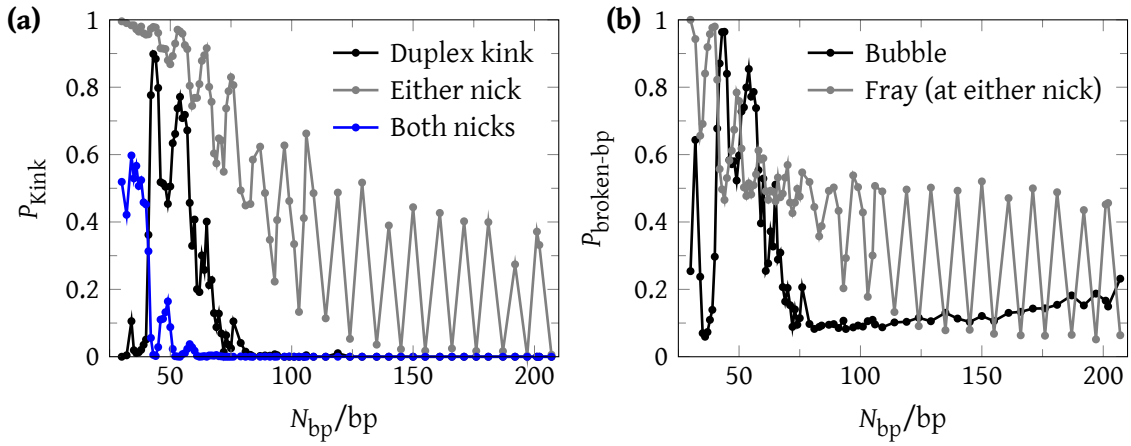


Figure 6.5 Kinking in DNA cyclization
(a) Probability of kinking P_{kink} as a function of length N_{bp}, in the duplex (**black**), at either nick (**gray**) and at both nicks (**blue**). Details of our structural kink criteria are available in Section 3.2.2.1. **(b)** Probability of broken base pairs $P_{broken\text{-}bp}$ as a function of length N_{bp} for fraying (base pairing disruption at either nick) (**gray**) and bubble formation (base pairing disruption in the duplex region) (**black**).

stacked) and 'off-register' (kinked at a nick) molecules decreases at shorter N_{bp}, rather than increasing as predicted by the S&Y model. Although not taken into account for most analyses of cyclization, the possibility of the cyclized molecule exhibiting a 'teardrop' configuration has previously been suggested by Vologodskii *et al.*[24,26]

Further evidence in support of this argument is given in Figure 6.5 (a), which shows the probability of kinking[i] as a function of N_{bp}. At long lengths ($N_{bp} \gtrsim 80$), kinking does not occur in the duplex regions, but can occur at the nick sites. There is clear periodicity in kinking at a nick. For example, at $N_{bp} = 145 \approx 14 \times$ pitch length, the percentage probability of kinking at either of the nicks is ~0 % and the system virtually always adopts a coaxially stacked circle configuration; however, at the longer $N_{bp} = 201 \approx 19.5 \times$ pitch length, the percentage probability is ~40 %.

As N_{bp} is shortened, the percentage probability of kinking at a nick gradually increases for these 'off-register' lengths; it is not until $N_{bp} = 114 \approx 11 \times$ pitch length that the bending stress along the duplex is sufficient to cause ~10 % of molecules to kink at a nick for an 'on-register' length.

[i] A slight additional complication stems from the use of umbrella sampling, a procedure needed to greatly enhance the sampling efficiency. To report kinking we must first remove the umbrella weight applied for each order parameter to yield the unweighted kinking distribution. Naturally, the unbiased kinking distribution is presented in Figure 6.5 (a)

6.3.1.2 Intermediate lengths

In the intermediate-length regime ($N_{bp} \approx 45$–80 bp), we observe enhanced cyclization efficiency compared to the S&Y prediction. Although j_{eq}^{oxDNA} continues to decrease with N_{bp}, it does so more gradually than the S&Y expression would predict; consequently, j_{eq}^{oxDNA} is in excess of the peak envelope of the S&Y expression. In this regime, bending stress in the circular coaxially stacked configuration is sufficiently large that kinking at one of the nicks occurs for even the on-register systems (Figure 6.5 (a)). As a result, cyclization is more favourable for oxDNA than predicted by the S&Y model, with all points lying clearly above the envelope of the S&Y maxima.

On-register effects also seem to contribute to j_{eq}^{oxDNA} in the upper end of the intermediate-length regime, with shallow maxima occurring at $N_{bp} = 63$ and $N_{bp} = 74$, approximately 6 and 7 times the pitch length. However, these sizes no longer correspond to minima in the probability of kinking at a nick (Figure 6.5 (a)) so the structural underpinnings of these variations is less clear. For shorter lengths, although there are size-dependent variations in j_{eq}^{oxDNA} (*e.g.* maxima at $N_{bp} = 43, 49, 54$), there is no longer a simple relationship to the pitch length, instead reflecting more complex geometric compatibilities that allow cyclized states at these lengths to be particularly stable compared to nearby lengths.

In the intermediate regime, most cyclized molecules are kinked at one of the two nicks. At shorter N_{bp}, the bending stress in the duplex region of the teardrop configurations increases, with the highest curvature being localized opposite the nick that is kinked. Consequently, it becomes increasingly favorable to localize bending stress into a kink in the duplex (unnicked) region, with the probability of this duplex kinking (Figure 6.2 (e)) increasing from near zero at $N_{bp} = 81$ to near one at $N_{bp} = 43$ (Figure 6.5 (a)). A kink in the duplex will generally be located opposite a kink at a nick because this arrangement minimizes the residual bending stress in the unkinked portions of the duplex by equalizing the lengths of the double-helical segments between the two kinks. In addition to a loss of stacking, kinking in the duplex typically involves breaking 1–3 base pairs (Section 4.4.3). A typical configuration for such a kinked duplex state is shown in Figure 6.2 (e). As with kinking at the nick, the existence of this alternative configuration lowers the free-energy cost of cyclization, raising j_{eq}^{oxDNA} further above the S&Y prediction.

At the lower end of the intermediate-length regime, the cost of bending without kinking in the duplex region is so high that molecules with duplex kinking dominate the cyclized state. States with two kinks localize the majority of the bending stress at the kinking sites, as shown in Figure 6.2 (e), with relatively relaxed duplex sections . Consequently, the free-energy cost of looping is largely independent of N_{bp} in this regime, causing $j_{\mathrm{eq}}^{\mathrm{oxDNA}}$ to level-off and reach an approximately constant value. In contrast, the S&Y $j_{\mathrm{eq}}^{\mathrm{WLC}}$ prediction decreases very rapidly. For example, at $N_{\mathrm{bp}} = 43$, $j_{\mathrm{eq}}^{\mathrm{oxDNA}}$ is 10^9 times greater than the S&Y prediction.

There exists a rich landscape of structures informed by slight differences in local geometries as a function of N_{bp}. In addition to the canonical two-kink structures, containing one kink in the duplex and one kink at a nick (Figure 6.2 (e)), we observe several nontrivial arrangements, albeit with relatively low probability. For example, the configuration in Figure 6.2 (f) contains a kink at both nicks as well as in the duplex.

6.3.1.3 Short lengths (< 45 bp)

In the short-length regime ($N_{\mathrm{bp}} \lesssim 45$), we observe an increase in $j_{\mathrm{eq}}^{\mathrm{oxDNA}}$, in stark contrast to the rapidly decreasing $j_{\mathrm{eq}}^{\mathrm{WLC}}$ predicted by the S&Y expression. Given the many assumptions violated at this length scale, a deviation is unsurprising; however, we did not anticipate an increase in $j_{\mathrm{eq}}^{\mathrm{oxDNA}}$.

As N_{d} is now not much larger than N_{s}, kinking at both nicks allows the single-stranded sticky ends to hybridize without duplex kinking. The system now adopts a conformation of two parallel duplexes, with the stress now borne by a mixture of continuous bending and fraying of a few base pairs at the ends of the duplexes. (Figure 6.2 (g,h)). Figure 6.5 (a) shows a very clear crossover to cyclized states with kinks at both nicks and no kinks present in the duplex, occurring abruptly at $N_{\mathrm{bp}} = 40$–42. At the same time, fraying at the duplex ends also increases (Figure 6.5 (b)).[ii] In this regime, as N_{bp} is shortened, the difference in length of the duplexes decreases and the stress in the system tends to drop, leading to a higher value for j_{eq} (Figure 6.3). Overlaid on this overall trend are non-trivial geometric effects associated with whether the periodicity of the two duplexes is convenient for connecting them, which leads to the non-monotonic behavior of

[ii] Since distinguishing fraying from kinking at low Q_{bp} is problematic, we report unweighted kinking only for the most probable values, *e.g.* $Q_{\mathrm{bp}} = 8$–10. The lower Q_{bp} configurations appear in our simulations only because of biased sampling and have a negligible contribution to the unweighted kinking probability.

j_{eq} and kinking.

The N_{bp} length of the crossover between the short- and intermediate-length regimes will be very dependent on N_s, occurring at shorter N_{bp} lengths for shorter N_s. The thickness of the double helix is also now significant compared to the lengths of the double helical segments, influencing the amount of bending required for cyclization.

6.3.2 Sequence dependence

So far we have only reported results using the oxDNA average-base parameterization, in which the strength of base-pairing and stacking interactions are independent of base identity. As the free-energy cost of disrupting a duplex to form a kink is sequence-dependent, it is important to consider how sequence might perturb the general trends we have elaborated thus far. We therefore studied six $N_{bp} = 73$ sequences used in the V&H[15] experiments using the parameterization of oxDNA that includes sequence-dependent thermodynamics (Figure 6.3).

Duplex kinking for oxDNA typically involves the breaking of base pairs. We observe that kinking preferentially occurs at AT base pairs, which are weaker than GC base pairs. Base pairs that are weaker than average introduce preferred sites for kinking; thus, when duplex kinking is relevant, j_{eq}^{oxDNA} is expected to be larger for the sequence-dependent than for the average-base parameterization. This is indeed the case for $N_{bp} = 73$. In oxDNA, the sequence-induced variation for the V&H $N_{bp} = 73$ sequences is a factor of $\sim$4 (GC-content 11–52 %). This compares to a factor of $\sim$8 between the extrema in GC-content, poly(AT) and poly(GC).

That we find sequence heterogeneity generally makes duplex kinking easier is consistent with our explicit investigation of duplex kinking (Section 4.3.4.1), and a seperate study which found an enhanced likelihood for bubbles to form at the tips of plectonemes.[57] Our results imply that the crossover to cyclized configurations with duplex kinking occurs at slightly longer N_{bp} when sequence-dependence is included than for the average-base parameterization.

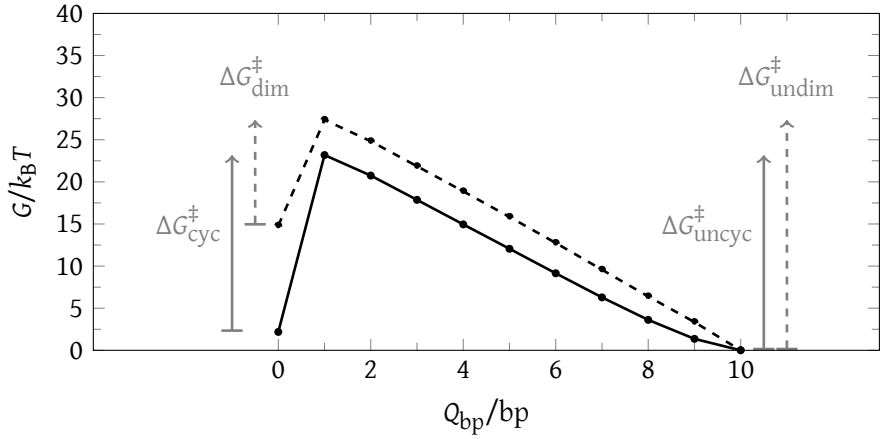


Figure 6.6 Free-energy profiles of cyclization/dimerization
Free-energy profiles of cyclization (**solid**) and dimerization (**dashed**) for a $N_{bp} = 101$ system ($N_s = 10$, $N_d = 91$). The activation free-energy barriers, $\Delta G^{\ddagger}$ for the forward (cyclization, dimerization) and reverse (uncyclization, undimerization) reactions are labelled.[iii] $\Delta G^{\ddagger}_{cyc}$ reflects the free-energy cost of bending to form the first base pair in a cyclization system, whereas $\Delta G^{\ddagger}_{dim}$ reflects the entropic cost of bringing two monomers together within the simulation volume. The dimerization simulations are for a simulation cube of side 85.18 nm, corresponding to a duplex concentration of 2.69 µM.

6.3.3 Free-energy profiles

Further insight can be gained by going beyond the equilibrium constants and studying the free-energy profiles of hybridization as a function of Q_{bp}, as illustrated for cyclization and dimerization at $N_{bp} = 101$ (Figure 6.6). One interesting feature of the profiles is that the free-energy gain from hybridizing the complementary sticky ends once an initial base pair has formed is less for cyclization than for dimerization: $\Delta G^{\ddagger}_{uncyc} < \Delta G^{\ddagger}_{undim}$. Physically, this indicates that substantial additional bending stress develops as subsequent base pairs form for the cyclization system, reducing the free-energy gain upon zippering of the sticky ends, relative to dimerization. For example, the $Q_{bp} = 1$ configuration in Figure 6.2 (c) is clearly less bent than the $Q_{bp} = 1$ configuration in Figure 6.2 (b) (see Section 6.7.2 for additional discussion).

6.3.4 Dimerization equilibrium

The equilibrium constant for dimerization K^{dim}_{eq} is length-independent to within two times the standard error of the mean for the oxDNA average-base parameterization (Table 6.1). This

[iii] For clarity, we have depicted $\Delta G^{\ddagger}_{uncyc}$ and $\Delta G^{\ddagger}_{undim}$ as the free energy difference between the fully base-paired closed state $Q_{bp} = 10$ and the transition state $Q_{bp} = 1$; however, in practice, frayed states $Q_{bp} = [2, 9]$ do contribute to the closed state. While the distinction does not significantly impact our results ($< 1/4\, k_B T$), we do include the contribution of frayed states in both $\Delta G^{\ddagger}_{uncyc}$ and $\Delta G^{\ddagger}_{undim}$.

suggests that the role of excluded volume effects is small between 336 nM and 2.69 µM.

For the sequence-dependent parameterization, there is some variation in $K_{\mathrm{eq}}^{\mathrm{dim}}$, but this does not represent length-dependence per-se; rather, it is most likely due to sequence variation in the bases at the interface between the duplexes and the sticky ends.

For the computation of j_{eq} and j_{dyn}, we use average values ($K_{\mathrm{eq}}^{\mathrm{dim}}$, $\Delta G_{\mathrm{dim}}^{\ddagger}$ and $\Delta G_{\mathrm{undim}}^{\ddagger}$) for 336 nM. For sequence-dependent results at $N_{\mathrm{bp}} = 73$, a length where the dimerization equilibrium was determined, we use the appropriate length-specific value instead of the average.

[iv] For [Conc]=336 nM, we refer to the N_{d} of the two monomers, e.g. for $N_{\mathrm{d}} = 20; 20$, given that all sequences are $N_{\mathrm{s}} = 10$, this corresponds to a *monomer* $N_{\mathrm{bp}} = 30$ in Table A.2. For [Conc]=2690 nM, we again refer to the N_{d} of the two monomers with $N_{\mathrm{s}} = 10$; however, we use the sequence that corresponds to the *dimer*, e.g. for $N_{\mathrm{d}} = 45; 46$, the *dimer* sequence is $N_{\mathrm{bp}} = 101$ in Table A.2.

N_d/bp*	[Conc]/nM	Average-base			Sequence-dependent		
		K_{eq}^{dim} / 10^{12} M^{-1}	$\Delta G_{dim}^{\ddagger}$ / $k_B T$	$\Delta G_{undim}^{\ddagger}$ / $k_B T$	K_{eq}^{dim} / 10^{12} M^{-1}	$\Delta G_{dim}^{\ddagger}$ / $k_B T$	$\Delta G_{undim}^{\ddagger}$ / $k_B T$
20;20	336	1.11 ± 0.11	14.42 ± 0.11	27.09 ± 0.10	0.42 ± 0.05	14.13 ± 0.07	25.75 ± 0.11
57;57	336	0.87 ± 0.05	14.62 ± 0.04	27.05 ± 0.05	0.54 ± 0.03	14.01 ± 0.01	25.79 ± 0.06
63;63	336	0.78 ± 0.04	14.69 ± 0.15	27.00 ± 0.05	0.33 ± 0.05	14.19 ± 0.05	25.60 ± 0.13
91;91	336	0.91 ± 0.15	14.66 ± 0.07	27.13 ± 0.16	0.44 ± 0.03	14.20 ± 0.04	25.86 ± 0.08
Avg	336	0.92 ± 0.20	14.60 ± 0.16	27.07 ± 0.20	0.43 ± 0.08	14.12 ± 0.10	25.75 ± 0.20
10;10	2690	1.22 ± 0.19	12.54 ± 0.05	27.38 ± 0.16	0.46 ± 0.02	12.11 ± 0.11	25.91 ± 0.04
28;29	2690	1.11 ± 0.09	12.66 ± 0.10	27.41 ± 0.08	0.82 ± 0.12	12.60 ± 0.04	27.03 ± 0.15
30;33	2690	1.11 ± 0.17	12.64 ± 0.12	27.40 ± 0.15	1.15 ± 0.12	12.53 ± 0.05	27.32 ± 0.11
45;46	2690	1.04 ± 0.09	12.55 ± 0.08	27.24 ± 0.09	0.48 ± 0.05	12.08 ± 0.04	25.90 ± 0.11
Avg	2690	1.12 ± 0.28	12.60 ± 0.18	27.36 ± 0.25	0.73 ± 0.18	12.32 ± 0.13	26.54 ± 0.22

Table 6.1 Dimerization equilibrium constants

Equilibrium constant for dimerization (K_{eq}^{dim}) and activation free-energy barriers to dimerization ($\Delta G_{dim}^{\ddagger}$) and undimerization ($\Delta G_{undim}^{\ddagger}$) as defined in Figure 6.6, for both the oxDNA average-base and sequence-dependent parameterizations. Simulations were performed at $T = 298$ K. Sequences from Table A.2.[iv]

* Monomer duplex length N_d. Complementary sticky ends are of identical sequence and length $N_s = 10$. Total length $N_{bp} = 2N_d + N_s$.

6.4 Kinetic *j*-factor

6.4.1 Context

As noted in the introduction, ligase experiments performed in the low ligase concentration limit should, in principle, be comparable to equilibrium *j*-factors. Indeed, for long DNA molecules (N_{bp} much longer than the persistence length) and low ligase concentrations, there has been consistent agreement between experiment and the WLC model.[12–16,71] However, Cloutier & Widom (C&W)[14] reported results for N_{bp} shorter than the persistence length ($N_{\mathrm{bp}} = 93, 94, 95, 105, 116$), showing an apparent deviation from WLC behavior, with $j_{\mathrm{dyn}}^{\mathrm{ligase}}$ (Equation (6.2)) enhanced over the S&Y WLC prediction $j_{\mathrm{eq}}^{\mathrm{WLC}}$[8] by a factor of $10^2 - 10^4$. We see no such enhancement at these lengths for oxDNA (Figure 6.3).

In contrast, Du *et al.*[13] found no deviation from WLC behavior for $N_{\mathrm{bp}} = 105$–130. Furthermore, they presented evidence suggesting that the C&W experiments used too high a ligase concentration to enable $j_{\mathrm{dyn}}^{\mathrm{ligase}}$ to be compared with $j_{\mathrm{eq}}^{\mathrm{WLC}}$. The results of Du *et al.* are in good agreement with the S&Y WLC expression, albeit with somewhat different parameters (torsional stiffness, persistence length and pitch length) than appropriate for oxDNA, owing to different buffer conditions. A more detailed comparison will be given in Section 6.6.

FRET measurements on DNA cyclization, as pioneered by Vafabakhsh & Ha (V&H),[15] provide a more direct measure of cyclization because k_{cyc}, k_{uncyc} and $K_{\mathrm{eq}}^{\mathrm{cyc}}$ are obtainable, although V&H mostly report k_{cyc}. V&H claim enhanced cyclization at $N_{\mathrm{bp}} \lesssim 100$, based on a comparison between their $j_{\mathrm{dyn}}^{\mathrm{FRET}}$ (Equation (6.3)) and the S&Y WLC expression $j_{\mathrm{eq}}^{\mathrm{WLC}}$. However, this is only a fair comparison if $k_{\mathrm{uncyc}} = k_{\mathrm{undim}}$, which, as k_{undim} is expected to be length independent, also implies that k_{uncyc} should be independent of length (N_{bp}). However, since both V&H[15] and more recent FRET measurements[16] suggest that k_{uncyc} increases with N_{bp}, this condition is not met. Specifically, V&H showing a 20-fold enhancement in the uncyclization rate compared to the undimerization rate for a $N_{\mathrm{d}} = 91$, $N_{\mathrm{s}} = 8$ molecule.[15] A reasonable physical justification for enhanced uncyclization, particularly as N_{bp} decreases, is that the greater extensibility of single-stranded DNA will allow some of the bending stress to be relieved as the sticky ends unbind.

6.4.2 Free energy of cyclization/uncyclization

Although we do not directly simulate the dynamics of cyclization, we can estimate the relative rates of processes using free-energy profiles such as Figure 6.6 and Kramer's theory for the rates of activated processes.[79] In agreement with experimental investigations, previous work on oxDNA has shown that duplex formation has an effective 'transition state' involving a very small number of base pairs (see discussion of reference 80). We therefore make the assumption that uncyclization and undimerization rates are given by:

$$k_{\text{uncyc}} = A \exp\left(\frac{-\Delta G^{\ddagger}_{\text{uncyc}}}{k_{\text{B}}T}\right), \tag{6.4}$$

$$k_{\text{undim}} = A \exp\left(\frac{-\Delta G^{\ddagger}_{\text{undim}}}{k_{\text{B}}T}\right). \tag{6.5}$$

where A is a universal constant for DNA melting, and $\Delta G^{\ddagger}_{\text{uncyc}}$ and $\Delta G^{\ddagger}_{\text{undim}}$ are defined in Figure 6.6. The physical content of this assumption is that increased favorability of base pair formation is manifested in slower unbinding (uncyclization) rates. This is important because bending stress in cyclized systems reduces $\Delta G^{\ddagger}_{\text{uncyc}}$ (Figure 6.6). Further, only $\Delta G^{\ddagger}_{\text{cyc}}$, and not $\Delta G^{\ddagger}_{\text{uncyc}}$ influences cyclization. The rate constants are equal ($k_{\text{uncyc}} = k_{\text{undim}}$) in the very long length limit (N_{bp} much longer than the persistence length), when their free-energy profiles are similar.

$\Delta G^{\ddagger}_{\text{cyc}}$ and $\Delta G^{\ddagger}_{\text{uncyc}}$ are plotted as a function of N_{bp} in Figure 6.7. In particular there is a general decrease in $\Delta G^{\ddagger}_{\text{uncyc}}$ at shorter lengths, suggesting k_{uncyc} increases with shorter N_{bp}, in agreement with experimental results.[15,16] Additionally, any torsional stress in the cyclized state is relieved as base pairs are disrupted, leading to the oscillations in $\Delta G^{\ddagger}_{\text{uncyc}}$ at long lengths, with the minima occurring at the more torsionally stressed off-register lengths ($N_{\text{bp}} \approx (n + 1/2) \times$ pitch length).

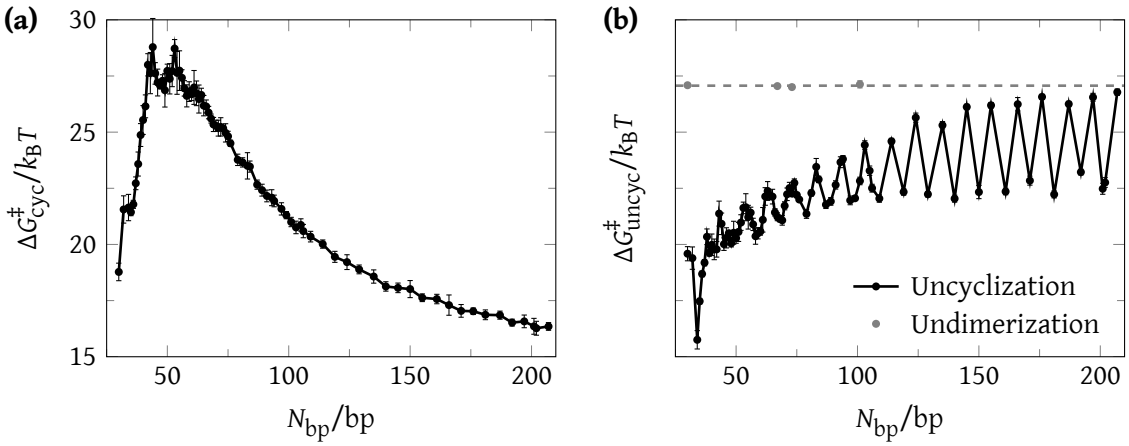


Figure 6.7 Free energy of cyclization/uncyclization, length-dependence
OxDNA average-base parameterization transition state free energies for **(a)** cyclization $\Delta G^{\ddagger}_{\mathrm{cyc}}$, $Q_{\mathrm{bp}=1} - Q_{\mathrm{bp}=0}$, **(b)** uncyclization $\Delta G^{\ddagger}_{\mathrm{uncyc}}$, $Q_{\mathrm{bp}=1} - Q_{\mathrm{bp}=[2,10]}$ (**black**) and undimerization (**gray**, average highlighted with dashed line). A graphical depiction of the transition states is available in Figure 6.6.

6.4.3 Derivation

To compare with V&H's results for $j^{\mathrm{FRET}}_{\mathrm{dyn}}$, we note that

$$j_{\mathrm{dyn}} = \frac{k_{\mathrm{cyc}}}{k_{\mathrm{dim}}} = \frac{K^{\mathrm{cyc}}_{\mathrm{eq}} k_{\mathrm{uncyc}}}{K^{\mathrm{dim}}_{\mathrm{eq}} k_{\mathrm{undim}}}. \tag{6.6}$$

Therefore, using Equation (6.5), our approximation for $j^{\mathrm{oxDNA}}_{\mathrm{dyn}}$ is

$$j^{\mathrm{oxDNA}}_{\mathrm{dyn}} = \frac{K^{\mathrm{cyc}}_{\mathrm{eq}} \exp\left(\frac{-\Delta G^{\ddagger}_{\mathrm{uncyc}}}{k_{\mathrm{B}}T}\right)}{K^{\mathrm{dim}}_{\mathrm{eq}} \exp\left(\frac{-\Delta G^{\ddagger}_{\mathrm{undim}}}{k_{\mathrm{B}}T}\right)}, \tag{6.7}$$

at each N_{bp}.

We are now in a position to estimate $j^{\mathrm{oxDNA}}_{\mathrm{dyn}}$ from oxDNA's equilibrium constants and activation free-energy barriers. We expect $K^{\mathrm{dim}}_{\mathrm{eq}}$ and $\Delta G^{\ddagger}_{\mathrm{undim}}$ to be length-independent as the excluded volume of the duplex far from the complementary single-stranded sticky ends is likely to have little effect on the dimerization process. Indeed, this appears to be the case to within $0.5\,k_{\mathrm{B}}T$ in ΔG (Table 6.1). Our resulting $j^{\mathrm{oxDNA}}_{\mathrm{dyn}}$ values are plotted in Figure 6.8, where they are compared to our $j^{\mathrm{oxDNA}}_{\mathrm{eq}}$ values, and V&H's $j^{\mathrm{FRET}}_{\mathrm{eq}}$.

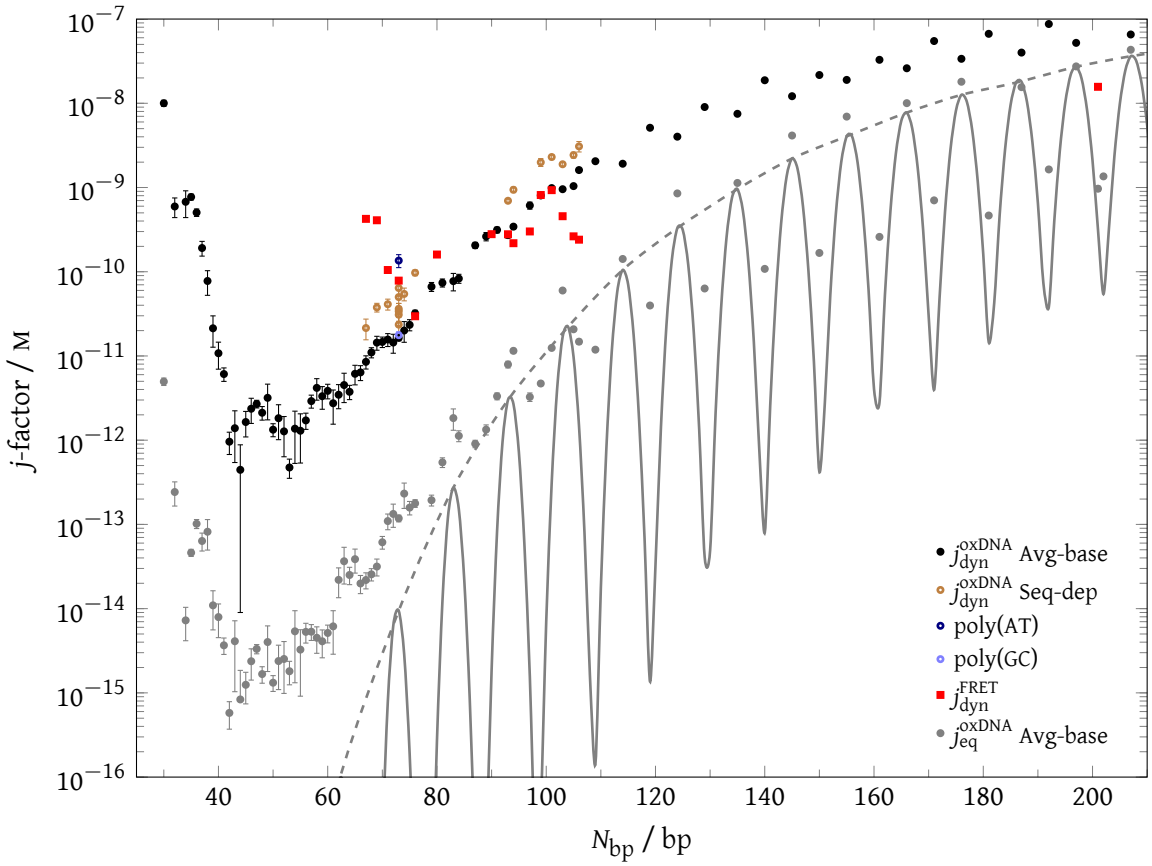


Figure 6.8 Kinetic *j*-factor
OxDNA dynamic *j*-factor j^{oxDNA}_{dyn} (**black** circle) compared to the FRET experiments of V&H j^{FRET}_{dyn} (**red** squares).[15] To highlight the role of sequence-variation, we also show results for the oxDNA sequence-dependent parameterization using 14 sequences of variable N_{bp}, and 6 poly(A) sequences at $N_{bp} = 73$. All sequences were also used by V&H,[15] and are available in Table A.2. For reference, j^{oxDNA}_{eq} and j^{WLC}_{eq} are also shown (**gray**).

6.5 Comparison to FRET experiments

We observe that oxDNA's $j_{\mathrm{dyn}}^{\mathrm{oxDNA}}$ values lie substantially above $j_{\mathrm{eq}}^{\mathrm{oxDNA}}$, except in the limit of long N_{bp} for on-register lengths. This is simply because the full bending and twisting stress present in the cyclized state is not yet present in the transition state where the two complementary sticky ends have formed their first base pair, *e.g.* Figure 6.2 (c). Instead of slowing down cyclization, stress which is not present in the transition state accelerates uncyclization. The observed behavior is similar to that seen by V&H; indeed, $j_{\mathrm{dyn}}^{\mathrm{oxDNA}}$ and $j_{\mathrm{dyn}}^{\mathrm{FRET}}$ agree remarkably well in the range $N_{\mathrm{bp}} = 90$–105, with reasonable agreement extending to $N_{\mathrm{bp}} = 70$. Our results suggest that variation of uncyclization rates with N_{bp} may be an important contribution to apparent non-WLC behavior in $j_{\mathrm{dyn}}^{\mathrm{FRET}}$.

We note that $j_{\mathrm{dyn}}^{\mathrm{oxDNA}}$, in contrast to $j_{\mathrm{eq}}^{\mathrm{oxDNA}}$, varies relatively smoothly with N_{bp}, with no strong periodicity on the length scale of the pitch length. Consistent with this, $\Delta G_{\mathrm{cyc}}^{\ddagger}$ also varies smoothly with N_{bp}. The periodicity in $j_{\mathrm{eq}}^{\mathrm{oxDNA}}$ comes from the free-energy gain when zippering-up the complementary sticky ends (*i.e.* $-\Delta G_{\mathrm{uncyc}}^{\ddagger}$), which is greater when N_{bp} is an integer number of the pitch length, allowing the formation of a relatively relaxed coaxially stacked circle. V&H suggest that their $j_{\mathrm{dyn}}^{\mathrm{FRET}}$ data at $N_{\mathrm{bp}} = 93$–106 displays a strong oscillation with a period of about one pitch length (Figure 6.8). In agreement with Vologodskii *et al.*,[26] we find no physical mechanism for such a strong oscillation and would suggest that the experimental evidence for this oscillation is not compelling.

In an attempt to account for the finite-size of the DNA sticky ends, V&H also attempted to compare their results to a WLC expression using a capture radius, a range in which the complementary sticky ends could associate, typically $\sim$5 nm; however, this comparison is somewhat arbitrary, and will be further explored in Section 6.7.1.

6.5.1 Sequence dependence

At the shortest lengths investigated, $j_{\mathrm{dyn}}^{\mathrm{FRET}}$ still lies above $j_{\mathrm{dyn}}^{\mathrm{oxDNA}}$. Some of this difference might be due to our use of the oxDNA average-base parameterization. Sequence does play a role in DNA flexibility (Section 4.3.4.1), but the impact of sequence variation is nontrivial. In particular,

because kinks tend to localize to AT base pairs, kinking in the duplex is easier for a sequence with 50 % GC-content than reported for our average-base parameterization.

Using the oxDNA parameterization with sequence-dependent thermodynamics for the six V&H $N_{bp} = 73$ sequences (GC-content 11–52 %),[15] we found j_{dyn} to increase compared to the value for the average-base parameterization. Consistent with our results for j_{eq}^{oxDNA}, we find a sequence-induced variation in j_{dyn}^{oxDNA} of a factor of $\sim$4 for these sequences, compared to a factor of $\sim$8 between the extrema in GC-content, poly(AT) and poly(GC).

This compares to a factor of $\sim$60 for experimental looping rates, a discrepancy which may be explained by V&H's use of poly(A) tracts, a sequence-motif well-known to introduce intrinsic curvature in duplex DNA.[81] As oxDNA's sequence-dependent parameterization is based on the nearest-neighbor thermodynamics of SantaLucia *et al.*,[43,44] alternative structural motifs such as poly(A) tracts are outside the scope of the model. Additionally, oxDNA doesn't reproduce sequence-dependent structural (*e.g.* the difference in size between purine and pyrimidine) or mechanical (*e.g.* flexibility) properties.

6.5.2 Length of complementary sticky ends

We also compare oxDNA to available V&H equilibrium data[15] for $N_d = 91, N_s = 8$–10 (Figure 6.9). Specifically, V&H report the fractional occupancy of the cyclized state f_{cyc}. Akin to the nomenclature in Appendix A.3.3, $f_{cyc} + f_{open} = 1$, where f_{open} is the fractional occupancy of the open (uncyclized) state.

Fractional occupancy can be a somewhat misleading measure by which to compare systems, because the entire transition region $f_{cyc} \approx 0.25$–0.75 is covered by a free-energy difference of $\sim 2\, k_B T$, while a similar free-energy difference at the extrema $\left(f_{cyc} \to 0 \text{ or } 1\right)$ would be indistinguishable. We therefore make our equilibrium comparison in terms of ΔG_{cyc}, where

$$\Delta G_{cyc} = -k_B T \ln\left(K_{eq}^{cyc}\right), \tag{6.8}$$

and

$$K_{eq}^{cyc} = \frac{f_{cyc}}{f_{open}} = \frac{f_{cyc}}{1 - f_{cyc}}. \tag{6.9}$$

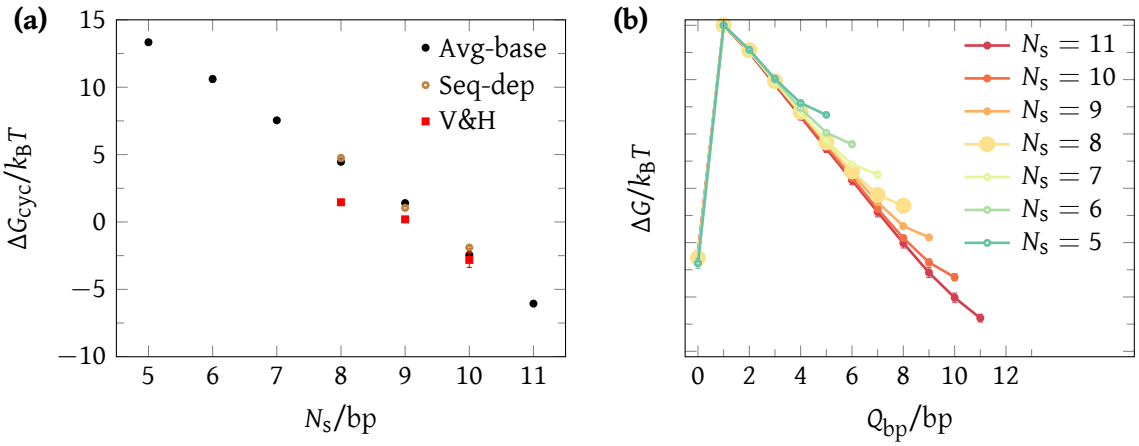


Figure 6.9 OxDNA versus FRET experiment, sticky ends
(a) Comparison between oxDNA and V&H[15] (Supplemental Material, Figure S1B) for ΔG_{cyc}, the free energy of the cyclized compared to the open state. **(b)** Free energy ΔG versus the order parameter reporting the number of base pairs formed between complementary sticky ends Q_{bp} for the oxDNA average-base parameterization. The reference state corresponding to $\Delta G = 0\, k_B T$ is $Q_{bp} = 1$.

OxDNA appears to be in good agreement with experiment; although, this may be partly coincidental as V&H do not report either the temperature, presumed to be 298 K, or a salt concentration: V&H give their imaging buffer cation concentration as $[Na^+] = 500$–$1000\,\text{mM}$ or $[Mg^{2+}] = 10$–$30\,\text{mM}$, while oxDNA is parameterized to $[Na^+] = 500\,\text{mM}$. At $N_s = 9$ and $N_s = 10$, oxDNA and experiment differ by $\sim 1\, k_B T$, while at $N_s = 8$, oxDNA under reports experiment by $\sim 3\, k_B T$ (cyclized state is less likely in oxDNA than experiment).

As expected from the SantaLucia model[43,44] of nearest-neighbor thermodynamics, the linear relationship between oxDNA results in Figure 6.9 (a) is reflective of the $\sim 3\, k_B T$ stabilization of each additional base pair. A possible explanation for the discrepancy between oxDNA and V&H[15] is that hairpin formation in the N_s region could decrease the fraction of cyclized molecules; but given the sequence, hairpin formation should be negligible. In particular, we cannot explain the discrepancy between oxDNA and the V&H data for $N_s = 8$, although we note that identifying a very low yield in experiment can be challenging, for instance due to the presence of impurities inducing large changes in ΔG_{cyc}.

6.5.3 Nicks and mismatches

We examine the role of structural defects, namely nicks and mismatches near the mid-point along the duplex, on cyclization using the average-base oxDNA parameterization. Unsurprisingly, we find that nicked and mismatches molecules cyclized more readily than their intact counterparts (Figure 6.10), consistent with our findings regarding the lower free-energy cost for kinking in these systems (Section 4.3.4.2, Section 4.3.4.3). We also find that kinks preferentially localize on the duplex region opposite the kink at a nick; for an example, see Figure 6.2 (e). The phenomena of correlated kink localization will explored with DNA minicircles in Chapter 7.[v]

V&H briefly examine the impact of nicks and mismatches on the rate of cyclization, recording f_{cyc} versus time (Figure S3 of V&H15). It does not appear that all molecules have reached an equilibrium plateau in f_{cyc}, therefore a concrete equilibrium comparison to oxDNA is difficult. We do, however, observe good agreement between the oxDNA and the experimental observations of V&H regarding the stabilizing effect of various motifs (Table 6.2). However, we do not reproduce the unexpected similarity between the experimental values for intact sequences at $N_{\mathrm{bp}} = 69$ and $N_{\mathrm{bp}} = 97$. This is consistent with our inability to reproduce V&H's j-factors for their shortest sequences.

Reassuringly, the results are at least internally self-consistent with the reported V&H j-factors, 4.08×10^{-10} M^{-1} and 3.00×10^{-10} M^{-1} for the nicked and mismatched systems, respectively.[15] The free-energy differences between intact molecules and their nicked/mismatched counterparts are also in good agreement with experiment (Table 6.2). Nonetheless, we caution the reader against over-interpretation of the exact numbers, and recommend in favor of qualitative comparisons.

6.5.4 Summary of Comparison

The origin of the remaining discrepancy between $j_{\mathrm{dyn}}^{\mathrm{FRET}}$ and $j_{\mathrm{dyn}}^{\mathrm{oxDNA}}$ at the shortest N_{bp} is not yet clear, but may indicate enhanced flexibility for V&H versus oxDNA due to WLC (*i.e.* lower persistence length) or non-WLC behavior (*i.e.* kinking at slightly longer N_{bp}). The

[v] The presence of two nicks in the DNA cyclization system slightly complicates the analysis of kink localization: for clarity, we present kink localization in the context of intact and singly nicked DNA minicircles in Section 7.4.4.

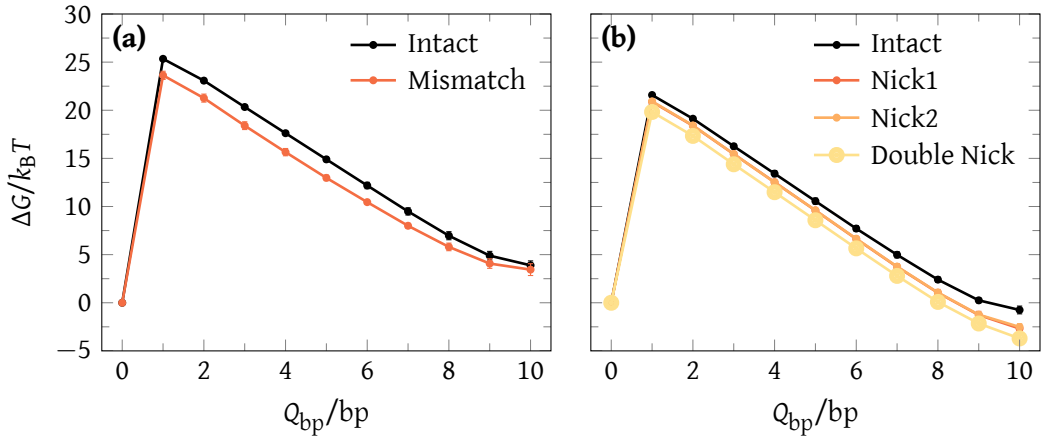


Figure 6.10 OxDNA versus FRET experiment, nicks and mismatches
Free-energy profile for **(a)** intact versus mismatch at $N_{bp} = 69$, and **(b)** intact versus nick(s) at $N_{bp} = 97$, using the oxDNA average-base parameterization and $N_s = 10$ for all sequences (Table A.2).

	OxDNA		Experiment*	
	K_{eq}^{cyc}	$\Delta\Delta G/k_BT$†	K_{eq}^{cyc}	$\Delta\Delta G/k_BT$†
$N_{bp} = 69$				
Intact	0.029 ± 0.01		1.10	
Mismatch	0.052 ± 0.02	−0.6	3.14	−1.1
$N_{bp} = 97$				
Intact	2.99 ± 0.35		0.93	
Nick 1	18.50 ± 0.93	−1.8	1.73	−0.6
Nick 2	16.18 ± 1.03	−1.7		
Double Nick	50.31 ± 2.35	−2.8	8.78	−2.3

Table 6.2 OxDNA versus FRET experiment, nicks and mismatches
K_{eq}^{cyc} for oxDNA and experiment (V&H,[15] Supplemental Material, Figures S3A and S3B). The experimental assays were not intended to probe the equilibrium behavior of the systems; as such, the last point of the kinetic cyclization versus time assay may not represent equilibrium values. To emphasize the qualitative difference between intact molecules and those with structural defects (nicks, mismatches), oxDNA results are for the average-base parameterization. Sequence-dependence introduces additional variability, and may complicate the comparison.
* Converted from yield f_{cyc} to K_{eq}^{cyc} using Equation (6.9).
† $\Delta\Delta G$ is computed as the difference in the ΔG of cyclization between intact states and those with structural defects (nick or mismatch).

oxDNA persistence length of 41.82 nm is within the range of experimental observations at $[Na^+] = 500\,mM$, but smaller values at $[Na^+] = 750\,mM$ are not implausible.[82] It is also possible that oxDNA slightly underestimates the prevalence of kinking within duplex regions (Section 7.4.1); an onset of kinking at slightly lower stress (longer N_{bp}) would make cyclization at shorter lengths more favourable. Finally, it is worth noting that the presence of fluorophores may cause perturbations in the V&H experiments.

V&H do report K_{eq}^{cyc} for some systems; oxDNA results are in good agreement for those lengths (Figure 6.9).

6.6 Comparison to ligase experiments

A comparison to the ligase experiments of Du *et al.*[13] and Cloutier & Widom (C&W)[14] highlights some interesting features of the oxDNA model (Figure 6.11). For example, while both the Du *et al.* experimental and oxDNA model results are comparable to the S&Y WLC expression for $N_{bp} \gtrsim 100$, there is a significant absolute deviation in j-factor due to differences in the relevant mechanical properties for oxDNA and those seemingly appropriate for the conditions of the experiment of Du *et al.* Specifically, they use NEB T4-ligase buffer, which has a salt concentration of $[Mg^{2+}] = 10\,mM$, while oxDNA is parameterized to $[Na^+] = 500\,mM$. Du *et al.*'s fit to their data gives a weaker torsional stiffness (2.4×10^{-28} versus $4.75 \times 10^{-28}\,J\,m^{-1}$), longer persistence length (47 versus 41.82 nm) and longer pitch length (10.54 versus 10.36 bp/turn) than for oxDNA. The comparison also highlights how relatively small deviations in DNA mechanical properties ($\sim$10 %) may shift the apparent j_{eq} by an order of magnitude.

The differences in S&Y WLC expression parameters are reasonable given the differential impact of divalent versus monovalent cations on DNA at high-salt concentrations. In the salt-concentration regime of Du *et al.* and oxDNA, melting temperatures are systematically lower in Mg^{2+} than Na^+.[83] For persistence length, the differential impact of divalent and monovalent salts is rather more complicated to disentangle. Briefly, the persistence length decreases from $\sim$100 nm at very low salt to a high-salt saturation value of $\sim$40 nm, with similar saturation persistence lengths for Mg^{2+} and Na^+. Given the measurement and mark-to-model uncertainty, persistence lengths of $\sim$35–50 nm are a reasonable high-salt limit.[82,84] By mark-to-model, we

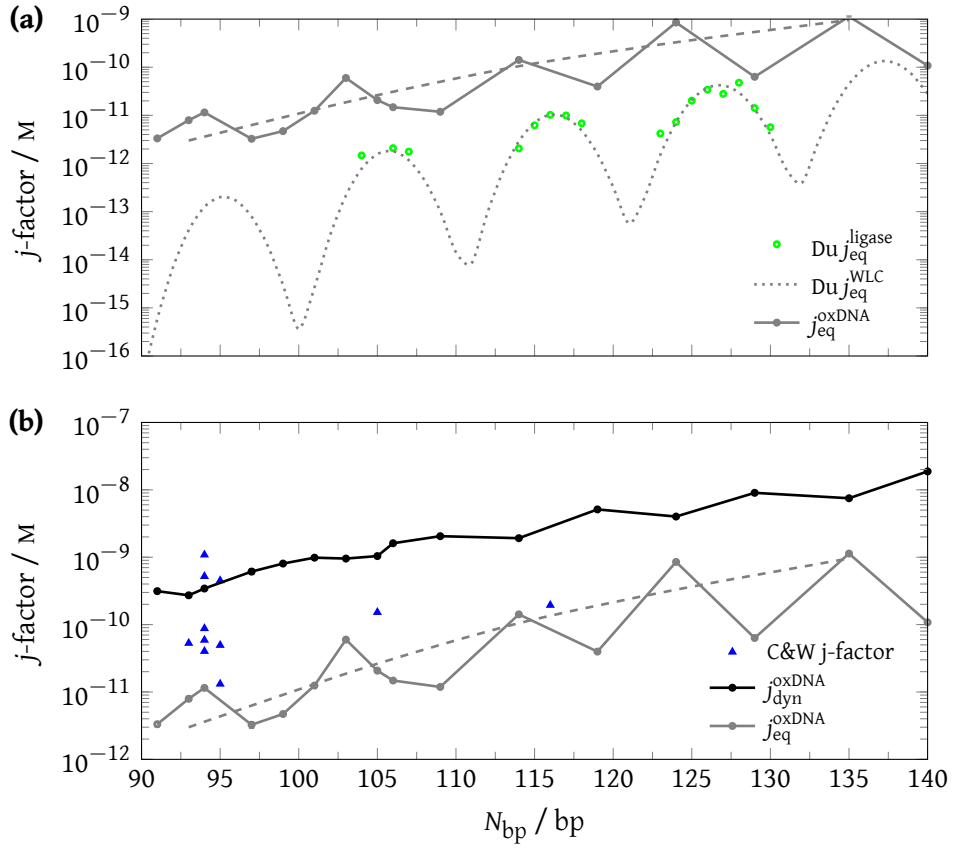


Figure 6.11 OxDNA versus ligase experiments
OxDNA comparison to the experiments of Du *et al.*[13] and Cloutier & Widom.[14] **(a)** Du *et al.* equilibrium *j*-factor via a ligase-based assay j_{eq}^{ligase} (**green** circles) compared to the S&Y WLC expression j_{eq}^{WLC} with Du *et al.* parameters (**gray** dotted) and the oxDNA model j_{eq}^{oxDNA} (**gray** marks). The top envelop of j_{eq}^{WLC} with oxDNA parameterized is dashed (**gray** dashed) **(b)** C&W ligase-based *j*-factor compared to j_{dyn}^{oxDNA} (**black** marks) and j_{eq}^{oxDNA} (**gray** marks). The maxima envelope of j_{eq}^{oxDNA} is given with a gray dashed line.

mean that experiments rarely report material properties directly, rather a model is used to translate experimental observables into estimates of material properties.

In this respect, atomic force microscopy (AFM) measurements of the persistence length are limited because Mg^{2+} is required to bind DNA to the surface.[85] Single-molecule trap-based measurements require a mark-to-model approximation from which it is difficult to precisely decouple bending (*i.e.* persistence length) as opposed to twisting (*i.e.* torsional stiffness).[86] Bulk optical measurements (*e.g.* magnetic birefringence, light scattering) have similar mark-to-model limitations in extrapolating their observables to persistence length.[84,87] There is also a slight temperature dependence, with j_{eq} enhanced by a factor of $\sim$10 between 5 °C and 42 °C.[88]

Du *et al.* only investigated lengths near the maxima in $j_{\mathrm{eq}}^{\mathrm{WLC}}$. As a consequence, their experiments are unable to confirm our prediction of a decrease in the magnitude of the oscillations in $j_{\mathrm{eq}}^{\mathrm{oxDNA}}$, which we attribute to the affect of teardrop configurations. Whether the experiments could potentially see the effects of teardrop configurations also depends on how the ligase activity depends on the nature of the DNA confirmation. Indeed, Vologodskii *et al.* mention the possibility of poor ligation efficiency for teardrop configurations.[26] For example, if the ligase were only active on fully-stacked circular configurations, then the affect of teardrop configurations would not be noticeable in the *j*-factor, even if they were present.

Since it is likely that C&W conducted their experiment at too high a ligase concentration, it is difficult to characterize their *j*-factor in terms of either j_{dyn} or j_{eq}. To isolate high ligase concentration as the culprit behind an apparent deviation from $j_{\mathrm{eq}}^{\mathrm{WLC}}$, Du *et al.* used conditions identical to C&W, namely buffer ($[Mg^{2+}] = 10\,\mathrm{mM}$) and temperature ($T = 21\,^{\circ}\mathrm{C}$ for Du *et al.*, $T = 20\,^{\circ}\mathrm{C}$ for C&W and $T = 25\,^{\circ}\mathrm{C}$ for oxDNA). Intriguingly, C&W's results mainly lie in-between $j_{\mathrm{dyn}}^{\mathrm{oxDNA}}$ and $j_{\mathrm{eq}}^{\mathrm{oxDNA}}$.

6.7 Discussion

6.7.1 Capture radius

To account for the role of the single-stranded tails in the cyclization rate, various authors have incorporated a 'capture radius' into the WLC *j*-factor calculation, as an approximation for how

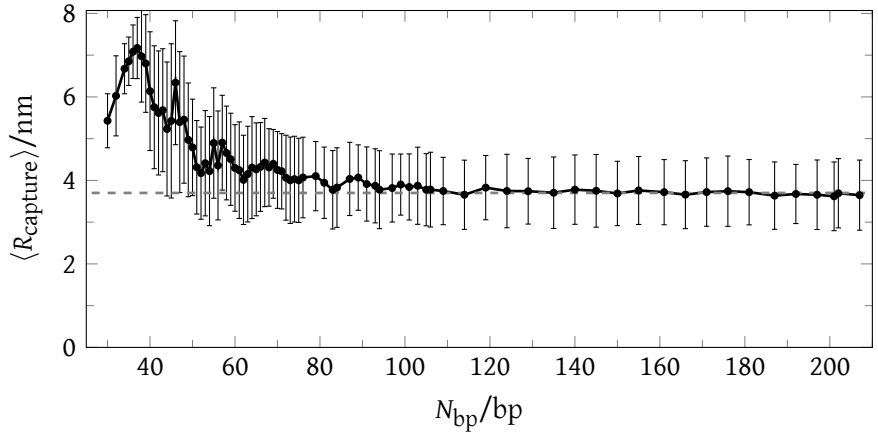


Figure 6.12 Estimate of capture radius
The ensemble average of the oxDNA capture radius $\langle R_{\text{capture}} \rangle$ as a function of length N_{bp} at fixed $N_{\text{s}} = 10$. The plateau at $\langle R_{\text{capture}} \rangle \approx 4$ nm at $N_{\text{bp}} \approx 50$–200 (gray dashed line), confirms that the typically assumed capture radius of 5 nm is reasonable. Error bars represent the standard deviation.

close the duplex ends must be in order for the sticky ends to hybridize.[15,16,26] Indeed, this approach leads to an enhanced j-factor, and while reasonable, the choice of capture radius is somewhat arbitrary. Again, we wish to emphasize that $j_{\text{dyn}} > j_{\text{eq}}$ is not in itself an indicator of non-WLC behavior. Also, as oxDNA naturally accounts for the finite-size of the sticky ends, we have the ability to test the capture radius assumption (Figure 6.12).

In oxDNA, we define the capture radius R_{capture} as the distance between the first and last bases of the duplex region (N_{d}) at $Q_{\text{bp}} = 1$ (Figure 6.13 (a,c)). Nucleotide positions are used to report distance, instead of the mid-point along the helical axis, because fraying may occur at the ends of the duplex. The choice of strand does not impact our results.

Our ensemble average capture radius, $\langle R_{\text{capture}} \rangle \approx 4$ nm, compares favorably with the typically assumed capture radius of 5 nm.[14,15] $\langle R_{\text{capture}} \rangle$ is roughly constant at $N_{\text{bp}} \approx 100$–$200$, increases slowly between $N_{\text{bp}} \approx 50$–100, and then more rapidly for $N_{\text{bp}} \approx 40$–50, before decreasing for $N_{\text{bp}} \gtrsim 37$.

The free-energy cost associated with forming the initial contact is partitioned between that for bending the duplex and that for stretching the single-stranded sticky ends. At shorter N_{bp}, the amount of bending required for initial binding increases, as does the stretching of the single-stranded sticky ends. At short lengths, this trend is reversed when the length of the unperturbed duplex approaches the capture radius at $Q_{\text{bp}} = 1$. Interestingly, the distribution of

R_{capture} is roughly Gaussian, with similar standard deviation for all lengths.

While we explicitly disallow misbonding, allowing misbonding may slightly increase the capture radius, an effect that should be more prominent for longer sticky ends. Overall, the commonly-used 5 nm value for the capture radius appears reasonable.

6.7.2 Initial stress in cyclized system

There is a general trend towards a larger activation free-energy barrier to cyclization ($\Delta G^{\ddagger}_{\text{cyc}}$) at shorter N_{bp}. This is due to the bending stress imposed upon the system by the formation of the initial base pair ($Q_{\text{bp}} = 1$). At the very shortest lengths $N_{\text{bp}} \approx 30$–45, this trend is reversed because the complementary single-stranded sticky ends (of length $N_{\text{s}} = 10$) are sufficiently long relative to N_{d} that an initial base pair can form with less bending stress in the duplex region. Cyclized versus dimerized states at $N_{\text{bp}} = 30, 101$ are depicted in Figure 6.13, clearly showing bending in the cyclized molecules, but not the dimerized molecules. Importantly, at $N_{\text{bp}} = 30$, the relatively long single-stranded region ($N_{\text{s}} = \frac{1}{2} \times N_{\text{d}} = 10$) reduces the requirement to bend the duplex (Figure 6.13 (a)), compared to the strong bending necessary at $N_{\text{bp}} = 101$ (Figure 6.13 (c)).

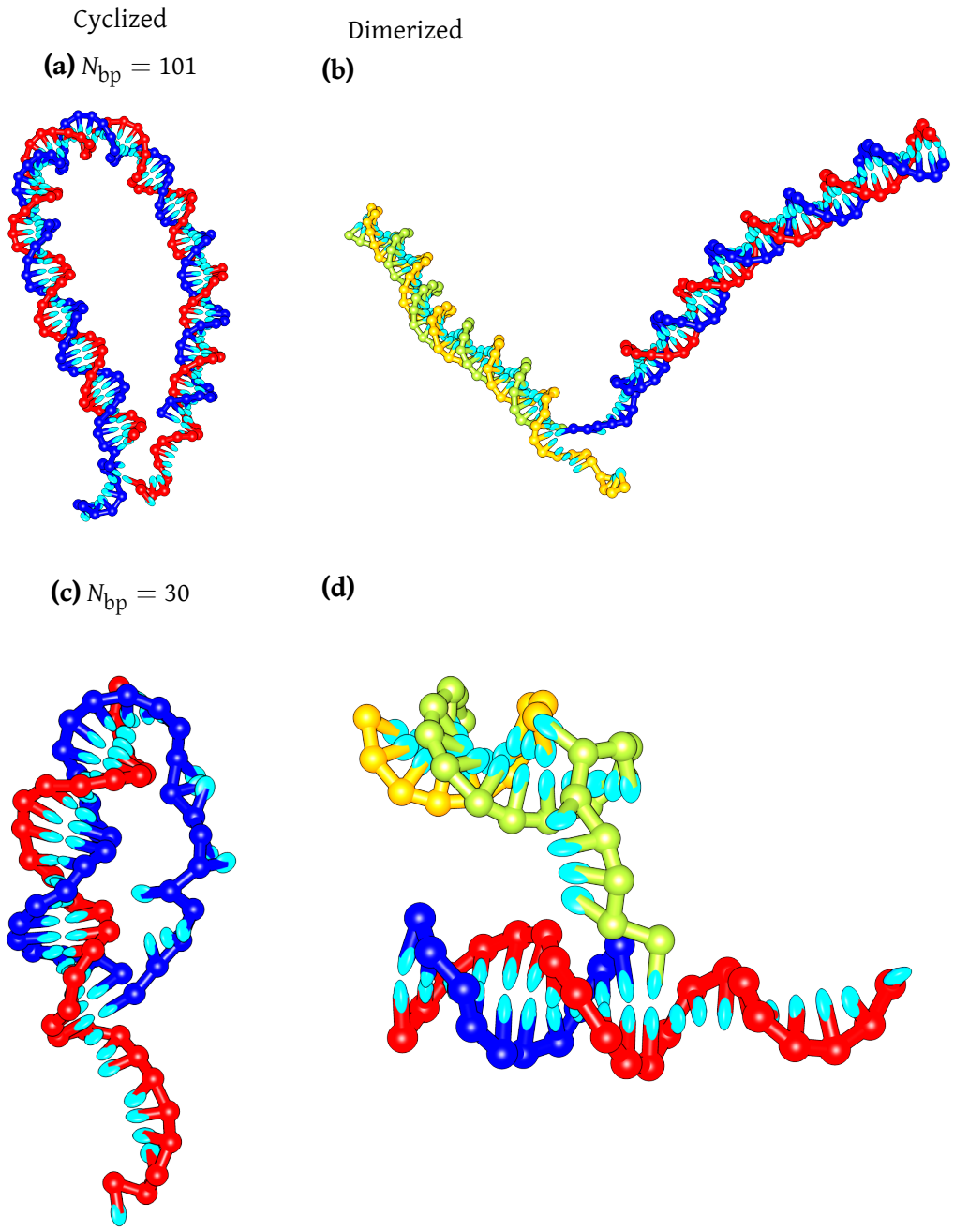


Figure 6.13 Examples of stress in cyclized system
OxDNA representations of $Q_{bp} = 1$ states for cyclized and dimerized configurations, highlighting the bending required to form the first base pair. Dimerized configurations in (b) and (d) are shown for $N_{d1} + N_{d2} + N_s = N_{bp}$, instead of length N_{bp} monomers. For (b), $N_{d1} = 45$, $N_{d2} = 46$ and $N_s = 10$. For (d), $N_{d1} = N_{d2} = 10$ and $N_s = 10$.

6.8 Conclusion

Cyclization is a system dependent manifestation of the general thermodynamics of strong DNA bending, elaborated in Chapter 4. The remarkable range of behavior in cyclized systems is explicable by the interplay between three specific deformation modes of stressed duplexes: continuous bending, kinking and fraying.

OxDNA reveals that each of these modes is present at a characteristic length-scale with respect to cyclization: continuous bending at long lengths ($N_{\mathrm{bp}} \gtrsim 80$), duplex kinking at intermediate lengths ($N_{\mathrm{bp}} \approx 45$–$80\,\mathrm{bp}$) and fraying at short lengths ($N_{\mathrm{bp}} \lesssim 45\,\mathrm{bp}$). In addition, as N_{bp} is shortened, there is an increase in kinking at the two nicks that remain after the hybridization of sticky ends. At longer lengths, kinking at a nick is only observed for 'off-register' molecules that cannot form torsionally relaxed coaxially stacked circles. The ability of said nicks to relax bending, as well as torsional stress, means that they become increasingly prevalent for shorter N_{bp}. At the shortest lengths, kinking at both nicks is dominant.

We use oxDNA to probe the reported observation of non-WLC behavior in FRET-based cyclization experiments.[15] In agreement with experiment, we observe that for shorter values of N_{bp}, the apparent j-factor lies substantially above the predictions of the Shimada & Yamakawa (S&Y)[8] WLC model. We also observe that the periodic oscillations predicted by the S&Y model are suppressed. This behavior arises from two conceptually distinct phenomena.

Firstly, highly stressed cyclized systems can adopt configurations that relax stress more effectively than through continuous bending, thereby reducing the overall free-energy cost of cyclization relative to a direct estimate based on a simple WLC-based model. At various values of N_{bp}, oxDNA identifies kinking at nicks, kinking within the duplex region and fraying of base pairs as key relaxation modes.

Secondly, oxDNA suggests that not all of the reduction in $K_{\mathrm{eq}}^{\mathrm{cyc}}$ relative to $K_{\mathrm{eq}}^{\mathrm{dim}}$ is due to stress manifest in the cyclization rate; uncyclization rates are also substantially increased relative to undimerization rates.[vi] The result is that dynamic j-factors based on the ratio of cyclization and dimerization rates lie even further above the S&Y prediction than their equilibrium j-factor

[vi] WLC and related statistical models do not predict absolute rates directly.

counterpart.

Only kinking within the duplex, not at a nick, can reasonably be described as truly non-WLC behavior. In oxDNA, kinking in the duplex only has a substantial effect for $N_{\mathrm{bp}} \lesssim 70$ bp. Classic WLC theory is applied to *duplex* DNA. Kinking at nicks and fraying can only occur when the DNA backbone is discontinuous, and any resultant effects are unrelated to whether WLC models accurately describes the DNA duplex.

Our results suggest that much of the apparent 'extreme bendability' reported by Vafabakhsh and Ha[15] can be attributed to factors that are not strictly speaking evidence of non-WLC behavior. Though we cannot account for deviations at their very shortest lengths ($N_{\mathrm{bp}} \lesssim 70$), we find no evidence for their claim of extreme bendability at $\sim$100 bp. In oxDNA, kinking within the duplex region is present, but not completely dominant. It is possible that oxDNA slightly overestimates the difficulty of kinking within a duplex – if this is the case, the data for the very smallest values of N_{bp} studied by V&H may be indicative of duplex flexibility over and above that predicted by the WLC. Although we note that to give a substantial effect on j_{eq}, kinking must not only be present, but must dominate the ensemble. Kinking dominating the ensemble at $N_{\mathrm{bp}} \approx 70$ is inconsistent with oxDNA predictions, and with available experimental evidence for a 'molecular vise' (Section 5.4) and DNA minicircles (Section 7.4.1).

Various authors have incorporated a 'capture radius' into the WLC j-factor calculation to capture phenomenologically some of the effects listed above.[15,16,26] Indeed, this approach leads to an enhanced j-factor, but the choice of capture radius is somewhat arbitrary and imprecise, making it difficult to assess whether non-WLC behavior is present. Further, it does not clearly differentiate between effects which reduce the free-energy cost of cyclization, and those which are partially manifested in uncyclization rates. Nonetheless, we find that the commonly used 5 nm capture radius is reasonable (Section 6.7.1).

OxDNA is only a model, and good correspondence with experimental results should not be over-interpreted. Nonetheless, the relaxation mechanisms identified are clearly physically plausible. For example, enhanced uncyclization rates have previously been noted in the literature.[15,16] It is clear that much of the apparent discrepancy between the data of V&H and the predictions of WLC-based models are due to effects that are not true violations of the WLC

model of duplex DNA flexibility. Our study also helps to reconcile the results of V&H with previous ligase-based assays which saw no evidence of enhanced flexibility at $N_{bp} \approx 100$,[13] and previous studies of minicircles which detected no evidence of duplex disruption at these length scales.[36]

6.8.1 Future directions

To explore whether the shortest lengths studied by V&H do show evidence of kinking and enhanced flexibility, we would propose experiments of shorter sequences and systematic collection of both dynamic and equilibrium data. The latter is extremely important; statistical WLC models make equilibrium predictions, and so the breakdown of a WLC description can only be confirmed with equilibrium data. Indeed, elucidating the subtleties of dynamic and quasi-dynamic (C&W) *j*-factors is one of the key issues addressed in this work.

7 Model system: DNA minicircles

Double-stranded DNA minicircles are a model system used to probe strong DNA bending due to the ability to easily control DNA topology, and thus their structure. By changing their length and linking number, relaxed as well as positively and negatively supercoiled minicircles may be probed. For sufficiently small minicircles, there is thought to be a transition from canonical duplex DNA to an alternative strongly bent configuration. The exact length and superhelical density at which this transition occurs is subject to debate, as is the nature of the strongly-bent state. We address these issues with oxDNA, supporting the view that relaxed DNA minicircles undergo a structural transition at a length of about six helical turns. Further, we find that in our model the structural transition is attributable to a specific defect, a kink, whereby bending is localized to a 1–3 base pair bubble. We explore kink-related emergent phenomena such as cooperative kink nucleation and the correlated diffusion of kinks. The behavior of supercoiled and nicked DNA minicircles is also explored.

7.1 Introduction

DNA minicircles are double-stranded DNA rings where the ends of the two strands are topologically sealed. While experimentally challenging to prepare, owing to their large free energies of formation relative to canonical duplex DNA, they are nonetheless a useful experimental and theoretical model system to investigate DNA bending. DNA minicircles possess a well-defined superhelical density σ, a measure of DNA supercoiling which may be precisely tuned by altering the linking number Lk and minicircle length N_{bp}. For a brief introduction to DNA topology, including supercoiling, see Mirkin 2001.[89]

7.1.1 Superhelical density

The DNA linking number Lk is an integer corresponding to the number of times the minicircle's two complementary single-strands wrap around each other.

Minicircles are relaxed when

$$Lk = Lk_0 \equiv \frac{N_{\mathrm{bp}}}{\text{base pairs per helical turn}} = \frac{N_{\mathrm{bp}}}{10.36}, \tag{7.1}$$

where 10.36 is the number of base pairs per helical turn in oxDNA. Typical experimental values are 10–11 bp/turn.[13,90–94]

The superhelical density σ is the normalized difference between the linking number Lk and the relaxed linking number Lk_0 given by

$$\sigma = \frac{\Delta Lk}{Lk_0} = \frac{Lk - Lk_0}{Lk_0}. \tag{7.2}$$

Relaxed DNA has $\sigma = 0$, positively supercoiled DNA $\sigma > 0$ (overtwisted) and negatively supercoiled DNA $\sigma < 0$ (undertwisted).

In practice, since the helical pitch length of DNA is non-integer, minicircles can only ever be approximately torsionally relaxed, being slightly undertwisted ($\sigma = 0^-$, $Lk < Lk_0$) or slightly overtwisted ($\sigma = 0^+$, $Lk > Lk_0$) with respect to the true torsionally relaxed state ($\sigma = 0$, $Lk = Lk_0$). The distinction is more pronounced for the shortest minicircles because the deviation between $\sigma = 0^-$, $\sigma = 0^+$ and $\sigma = 0$ is smaller for longer minicircles.

For a specific example, a relaxed minicircle at $Lk = 8$ and pitch length 10.36 bp/turn would correspond to $N_{\mathrm{bp}} = 82.88$, giving $N_{\mathrm{bp}} = 82$ as $\sigma = 0^+$, and $N_{\mathrm{bp}} = 83$ as $\sigma = 0^-$.

For reference, bear in mind that most naturally occurring minicircles will have $\sigma \approx 0$, and the widely studied *E. coli* genome is negatively supercoiled (undertwisted) to $\sigma \approx -0.06$.[95]

Note that linking number (Lk) is partitioned such that

$$Lk = Tw + Wr, \tag{7.3}$$

where *Tw* and *Wr* are the DNA twist and writhe, respectively. Physically, twist variation is a change in the number of base pairs per helical turn. Writhe is somewhat more difficult to conceptualize, but can be thought of as the number of topologically invariant strand crossing. For example, a typical 'figure-eight' formed from a loop will posses one point at which the loop crosses itself. This would be reported as $Wr = \pm 1$, depending on the orientation of the crossover.

7.1.2 Minicircle experiments

There is much interest in supercoiled DNA minicircles, specifically in the structural consequences of overtwist/undertwist ($\sigma > 0, \sigma < 0$), twist/writhe partitioning and how these relate to DNA bending. For example, single-strand specific nucleases, which cleave single-stranded but not double-stranded DNA, have been used to probe structural defects in negatively supercoiled minicircles.[96–98] Additionally, atomic force microscopy (AFM)[99] and cryo-electron microscopy (cryo-EM)[100] have revealed that writhing is the dominant effect in longer negatively supercoiled minicircles (178 bp and 339 bp with $\sigma \approx -0.12$ and $\sigma \approx -0.2$, respectively). Comparing the experiments suggests that greater $|\sigma|$ is required to writhe shorter minicircles, which is unsurprising given the greater ease of bending of a longer DNA duplex.

Alternatively, the effects of bending may be isolated by investigating relaxed ($\sigma \approx 0$) minicircles. A key question, highlighted by Vologodskii *et al.*,[26] is the length at which relaxed minicircles kink, whereby bending is localized to a structural defect at which, depending on the precise definition, base-base stacking and/or base pairing are disrupted (Section 3.2.2). For example, by probing relaxed minicircles with single-strand specific nucleases, Du *et al.* found no evidence of cleavage at $N_{bp} \approx 84$–85. However, cleavage was reported at $N_{bp} = 64$–65, suggesting the appearance of a structural defect (presumed to be a kink) at the shorter length.[36]

In agreement with Du *et al.*, a recent cryo-EM experiment reveals no evidence of kinking in a relaxed $N_{bp} = 94$ intact minicircle.[18] As a positive control for the study, double nicks or double single-stranded gaps were introduced 180° apart on the minicircle. Kinking was detected in the double-gapped but not in the double-nicked minicircles. Interestingly, 3D reconstructions of double-gapped minicircles suggest the presence of structures with zero, one or two kinks,[18]

raising interesting questions about variation in the number of kinks, the cooperativity of kink formation and spatial correlation in the localization of kinks.

7.1.3 Theory and simulation

Theory and simulation have been used to help elucidate the structural, mechanical and energetic phenomena underlying experimental findings. For example, Yan & Marko have demonstrated that strong bending localized to a structural defect could be responsible for enhanced DNA flexibility, compared to that expected for a worm-like chain.[24] Indeed, fully atomistic molecular dynamics (MD) simulations have reported single kinks in relaxed minicircles supporting the Yan & Marko prediction, while two kinks separated by $\sim$180° have been observed for positively supercoiled configurations.[25,65] Kinking was also reported in biased atomistic simulations on very short 12 bp DNA duplexes.[66]

Simulations of a simple model parameterized to the digestion studies of Du *et al.*[36] also suggest that the probability of structural defects increases for shorter minicircles.[101] Further supporting the hypothesis of kinking in short minicircles, another simulation study[102] reported that base-pair disruptions were much more likely to occur in short relaxed minicircles ($N_{\mathrm{bp}} = 66$) than in longer ones ($N_{\mathrm{bp}} = 86, 106$).

While the aforementioned studies focus primarily on structural defects, specifically kinking, twist/writhe coupling is also of interest. For example, investigating twist/writhe coupling in $N_{\mathrm{bp}} > 100$ minicircles with a coarse-grained model revealed that the partitioning is length-dependent;[103] specially, a planar twist dominated regime below a length-dependent superhelical density, with writhing becoming more prevalent at longer lengths.

7.1.4 Specific aims

Here we use oxDNA to examine the behavior of DNA minicircles; specifically, we address the length dependence of the structural defects and the nature of these states. Our model is particularly well-suited to the study of kinking, and the strong bending regime in general, as it applies to minicircles. For example, we have already seen that oxDNA provides a good description of the kinking transition in a 'molecular vise' (Chapter 5) and the role of kinking in

DNA cyclization (Chapter 6). Further, unlike all-atom simulations, oxDNA allows us to probe the relatively large time- and length-scales required for the characterization of DNA minicircles. Although kinked configurations have been observed via atomistic simulation, it is unclear whether these states are representative of equilibrium configurations. Further, atomistic simulations are too computationally expensive to observe multiple transitions, hampering the investigation of kink number variation, cooperativity and diffusion, all of which are explored in this study.

7.2 Simulation methods

Minicircle simulations are performed with MD (Section 3.1.2) at $T = 298$ K. Importantly, for the present study base pairs are only allowed to form between structurally 'correct' complements.

The DNA pitch length, the number of base pairs per helical turn, is an important model-dependent structural parameter for DNA supercoiling, defining Lk_0, the relaxed linking number, and therefore the superhelical density $\sigma = (Lk - Lk_0)/Lk_0$. In oxDNA, the pitch length is 10.36 bp/turn, within the 10–11 bp/turn range reported in a variety of assays.[13,90–94] Additional details are available in the Appendix A.4.

To highlight the basic thermodynamics of DNA minicircles, we use oxDNA's average-base parameterization; where appropriate, we use the sequence-dependent parameterization to highlight the role of sequence variation. Error bars represent the standard error of the mean from 5 independent simulations; where lines are shown, error bars are comparable to line thickness.

7.3 Results

The primary aim of this study is to investigate the length-dependent behavior of DNA minicircles, with a particular interest in the interconversion between continuously bent and kinked macrostates. We address the characteristic length scale at which kinking occurs for a variety of systems, as well as the number, localization and diffusion of kinks. First, we examine the case of torsionally relaxed minicircles ($\sigma \approx 0$), before turning our attention to the impact of nicks,

supercoiling ($\sigma \neq 0$) and cooperativity.

7.3.1 Torsionally relaxed minicircles

We begin by showing kinking in the duplex for torsionally relaxed intact minicircles as a function of length N_{bp} (Figure 7.1 (a)). We report results for the closest N_{bp} to $\sigma = 0$, $\sigma = 0^{-}$ and $\sigma = 0^{+}$, corresponding to slightly undertwisted and slightly overtwisted circles respectively.

The average number of kinks rises from 0 to 2, with a transition between $N_{\mathrm{bp}} = 63$, where less than 5 % of minicircles are kinked, and $N_{\mathrm{bp}} = 42$, where over 90 % are kinked (Section 7.4.3). We define the midpoint of the kink transition to be the point where the average number of kinks is 1, which is half the saturation value. In our model, the kink transition midpoint for torsionally relaxed intact minicircles (Figure 7.2 (a–c)) occurs at $N_{\mathrm{bp}} = 41$–42, with the onset of kinking at $N_{\mathrm{bp}} = 62$–63 (Figure 7.1 (a)).

By $N_{\mathrm{bp}} = 31$–32, molecules invariably exhibit 2 kinks, localized opposite each other (Figure 7.2 (c)). This configuration localizes bending stress to the two kinks, allowing for two relatively straight duplexes to form inbetween. On the contrary, in a 'teardrop' configuration with only one kink, bending stress is still substantial, as evidenced by the large curvature of the continuously bent duplex opposite the kink (Figure 7.2 (b)). Finally, in the absence of a kink, the entire structure is continuously bent (Figure 7.2 (a)).

7.3.2 Nicked minicircles

Comparison of intact minicircles to their counterparts with a nick on one of the strands (nicked minicircles) highlights interesting physics in both systems. Using oxDNA's average-base parameterization, we find that nicked minicircles (Figure 7.2 (d–f)) kink at longer lengths than intact minicircles (Figure 7.1 (e)). Easier kinking in nicked minicircles is due to the much lower free-energy cost of a kink at a nick, as opposed to a kink in the DNA duplex (Section 4.3.4.3). Thus, for virtually all states with one kink, the kink occurs at the nick, saturating to a value of $\langle \mathrm{Kink} \rangle_{\mathrm{nick}} = 1$ at $N_{\mathrm{bp}} \approx 75$.

Overlaid on the general trend towards more kinking at shorter lengths, there is an oscillation in

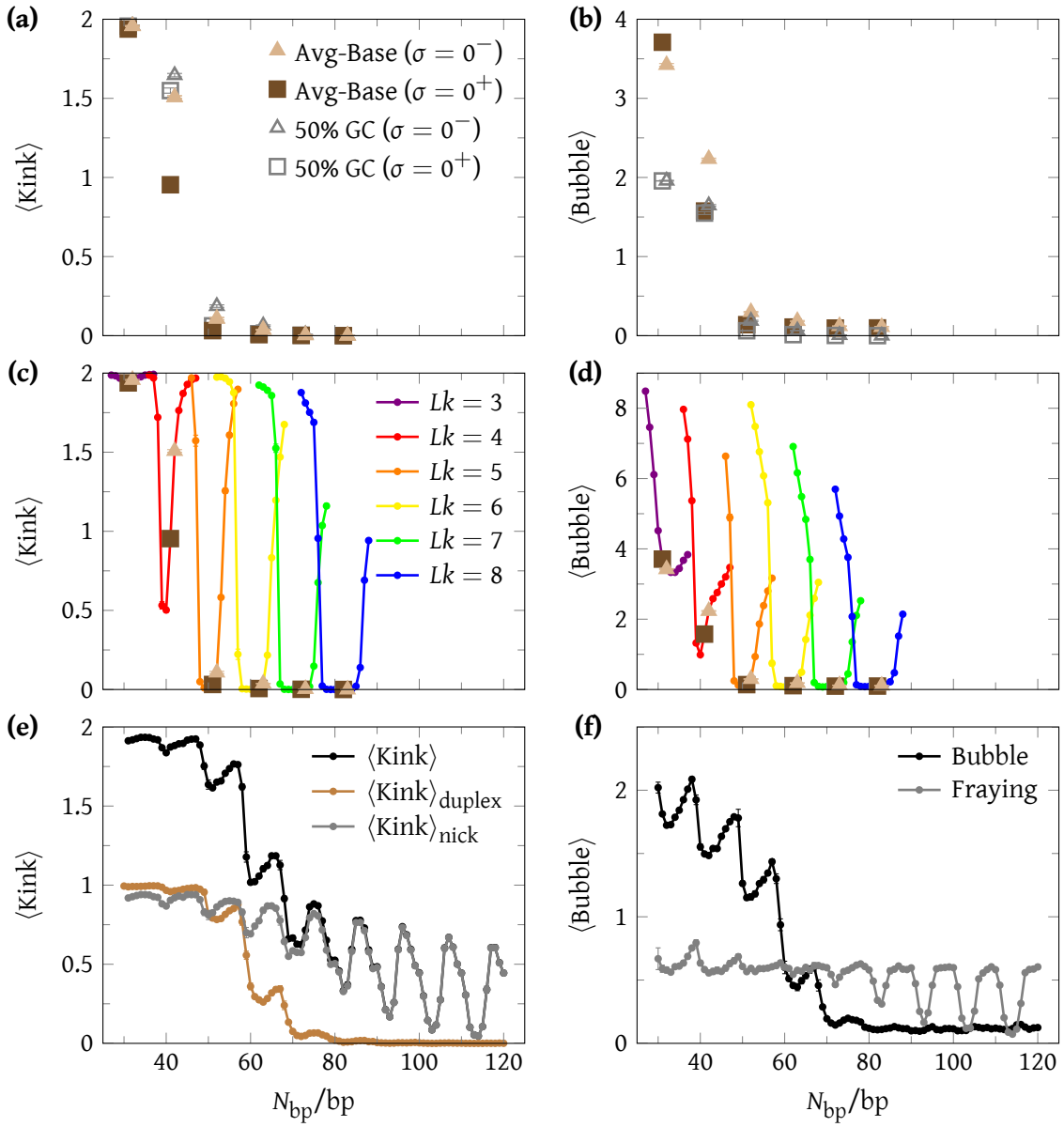


Figure 7.1 Length dependence of the average number of kinks and total bubble size
Average number of kinks ⟨Kink⟩ and average total bubble size ⟨Bubble⟩ versus minicircle length N_{bp} using the oxDNA average-base parameterization. **(a)** Kinking and **(b)** bubble size for intact relaxed minicircles ($\sigma \approx 0$), including sequence-dependent results for sequences with 50 % GC-content (open marks). **(c)** Kinking and **(d)** bubble size for intact minicircles ($-0.2 \lesssim \sigma \lesssim 0.2$) with variable linking number Lk to highlight the impact of supercoiling. **(e)** Kinking and **(f)** bubble size for nicked minicircles, partitioning average kinking between a kink in the duplex (⟨Kink⟩$_{\text{duplex}}$) and a kink at a nick (⟨Kink⟩$_{\text{nick}}$). By definition, bubbles may only occur in duplex regions; disrupted base pairing at strand ends (including nicks) is classified as fraying. Kinking is reported via the structural criterion (Section 3.2.2.1).

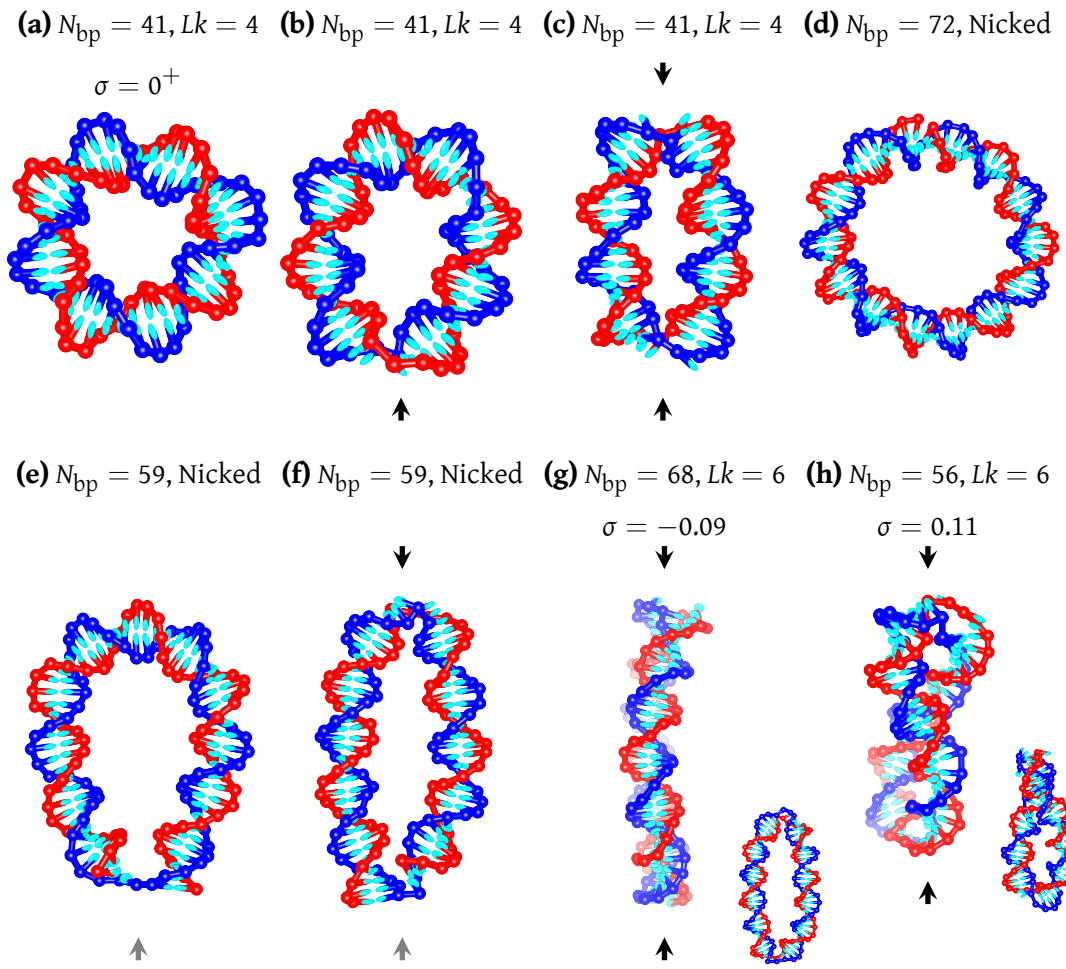


Figure 7.2 Representative minicircle structures
Intact relaxed minicircle ($\sigma = 0^+$) **(a)** in a 'circle' configuration, **(b)** with 1 kink, and **(c)** 2 kinks. Nicked minicircle in a **(d)** circle configuration, **(e)** in a 'teardrop' configuration with a kink at the nick, and **(f)** with both a kink at the nick and a kink in the duplex. Supercoiled minicircles with **(g)** undertwisting relaxed via bubble formation and twisting ($\sigma = -0.09$), and **(h)** overtwisting relaxed via writhing ($\sigma = 0.11$). Insets are rotated 90°. Arrows indicate kinks in the duplex (**black**) and kinks at a nick (**gray**).

$\langle \text{Kink} \rangle_{\text{nick}}$ with a period of one pitch length, with local maxima occurring for 'off-register' lengths ($N_{\text{bp}} = (n+1/2) \times$ pitch length) and local minima for 'on-register' lengths ($N_{\text{bp}} = n \times$ pitch length). The maxima for off-register lengths are due to the additional torsional strain present in such a system if stacking across the nick is to be maintained. Most of this stress can be simply relieved by kinking at the nick, with the kink acting as a relatively free hinge. However, some stress remains even after kinking, due to the constraint of backbone connectivity across the nick, which usually leads to out-of-plane bending. Unsurprisingly, oxDNA has revealed similar behavior for DNA cyclization (Section 6.3.1.1), as cyclization leads to minicircles with two nearby nicks.

Intriguingly, kinking in the duplex of nicked molecules also occurs at longer lengths than for intact minicircles, with a saturation value of $\langle \text{Kink} \rangle_{\text{duplex}} = 1$, a kink transition midpoint at $N_{\text{bp}} \approx 59$ ($\langle \text{Kink} \rangle_{\text{duplex}} \approx 50\,\%$), and detectable kinking at $N_{\text{bp}} \gtrsim 80$ ($\langle \text{Kink} \rangle_{\text{duplex}} \lesssim 1\,\%$).

Like the kink at the nick, oscillations with a period of one pitch length are also observed for $\langle \text{Kink} \rangle_{\text{duplex}}$, with maxima and minima for 'off-register' and 'on-register' lengths respectively (Figure 7.1 (e)). Although, the structural basis of this enhanced kinking at off-register lengths is not yet entirely clear, the additional bending stress due to the backbone connectivity across the nick, and the preference for duplex kinking into one of the grooves, probably plays a role (Section 7.4.4).

As previously suggested from cryo-EM observations,[18] we find that kinks for both intact and nicked minicircles tend to localize opposite each other (Figure 7.2 (c,f)). For the nicked minicircles, duplex kinks nucleate and diffuse only in a $\sim$10 base pair region opposite the kink at the nick (Figure 7.3 (a)). On the other hand, for the intact minicircles, kinks are free to nucleate and diffuse anywhere along the strand (Figure 7.3 (b)). The physical reason is quite clear: unlike the nicked minicircles (Figure 7.8 (b,d)), there are no strongly preferred sites at which to kink (Section 7.4.4).

Bear in mind that oxDNA's simplified representation of DNA neglects some features that may induce preferential kinking sites in the absence of nicks: with the average-base parameterization there are no AT base pairs at which to preferentially localize, and oxDNA's use of symmetric grooves eliminates the possibility of preferential kinking into either the major or minor groove.

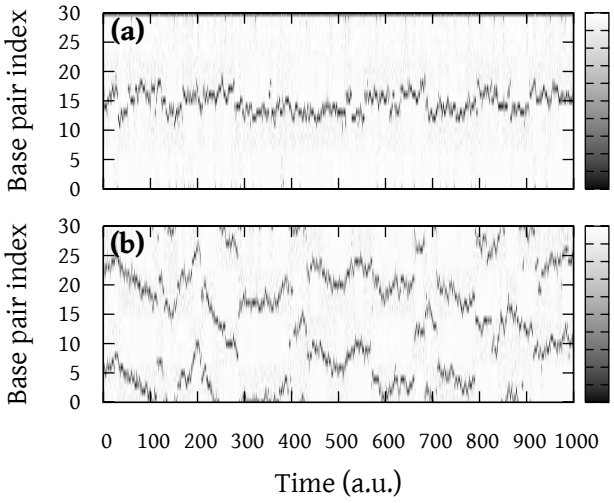


Figure 7.3 Correlated kink diffusion
Orientation between consecutive bases versus simulation time for **(a)** a nicked $N_{bp} = 31$ minicircle and **(b)** an intact relaxed $N_{bp} = 31$ minicircle. The base pair index i indicates the orientation between bases i and $i + 1$. For each base pair index and configuration, points are labeled with a gradient from unkinked parallel orientations (**white**) to kinked anti-parallel orientations (black). We choose an intermediate cutoff for kinked states (**gray**), but most kinked states are very anti-parallel. In **(a)**, note that the kink at the nick in the nicked minicircle lies at base pair index 30, along the top of the plot.

However, whenever two kinks are present, in either intact or nicked minicircles, their relative positions are strongly correlated. The resulting 'correlated diffusion' is clear in Figure 7.3 (b). Correlated diffusion appears to be quite robust; for example, even when kinks dissipate, they re-nucleate opposite an existing kink, if present.

7.3.3 Supercoiling

7.3.3.1 Background

Investigating supercoiled minicircles probes physics, such as twist/writhe partitioning, unavailable in torsionally relaxed systems ($\sigma \approx 0$). In our study of intact relaxed minicircles, there is a hint of the impact of supercoiling on kinking in minicircles in the observation that slightly undertwisted ($\sigma = 0^-$) minicircles kink more readily than their slightly overtwisted ($\sigma = 0^+$) counterpart. In the following, we will systemically investigate the effects of larger superhelical densities.

Supercoiled minicircles and supercoiled linear DNA under tension have been extensively studied

via experiment,[18,26,36,96–100] theory[16,24,36,101] and simulation.[25,65,66,102–104] The basic physics are reasonably well understood. For linear positively supercoiled DNA, excess twist beyond a certain characteristic twist density is partitioned to plectonemic writhe. For negatively supercoiled DNA, plectonemes may also form, but excess twist may also be absorbed into bubbles, which possess a naturally lower twist than double-helical DNA. The applied force controls the response to negative supercoiling, with plectonemes favored at low force, where the cost of the resulting reduced extension is not prohibitive. One more interesting feature found in recent oxDNA simulations of plectonemes[57] is that when the plectonemes are sufficiently tightly wound (as becomes the case as the force is increased) it may be favorable for kinks/bubbles to form at the plectoneme tip. This relieves bending stress, and in the case of very negative supercoiling, absorbs some of the excess twist.

7.3.3.2 Small superhelical density

It is well established that undertwisting (negative supercoiling, $\sigma < 0$) promotes, while overtwisting (positive supercoiling, $\sigma > 0$) suppresses bubble formation.[25,105–107] By probing kinking and bubble formation as a function of linking number Lk, we find that these general trends hold for intact minicircles (Figure 7.1 (c)): overtwisting suppresses bubbles, while undertwisting makes bubbles more likely (Figure 7.1 (d)).

As highly bent kinks generally necessitate the opening of a 1–3 base pair bubble (Section 4.4.3), suppressed bubble formation, in turn, reduces the propensity to kink. Therefore, undertwisted molecules kink more readily than their overtwisted counterparts. Indeed, the minima in $\langle \text{Kink} \rangle$ occur some 2–3 base pairs before the relaxed lengths.

That the minima in kinking occur for overtwisted as opposed to more relaxed minicircles may be more clearly seen on a plot of $\langle \text{Kink} \rangle$ versus superhelical density σ; for, example at $Lk = 5$ in Figure 7.4 (a,c). Given constant linking number, increasing the length N_{bp} effectively sweeps from overtwisted to relaxed to undertwisted minicircles. The minimum in $\langle \text{Kink} \rangle$ does not correspond to a relaxed minicircle, rather it occurs for an overtwisted minicircle of $\sigma \approx 0.05$.

Accordingly, for undertwisted minicircles we observe enhanced bubble formation *vis-à-vis* larger total bubble size (Figure 7.4 (b,d)). In contrast, mild overtwisting suppresses bubble formation,

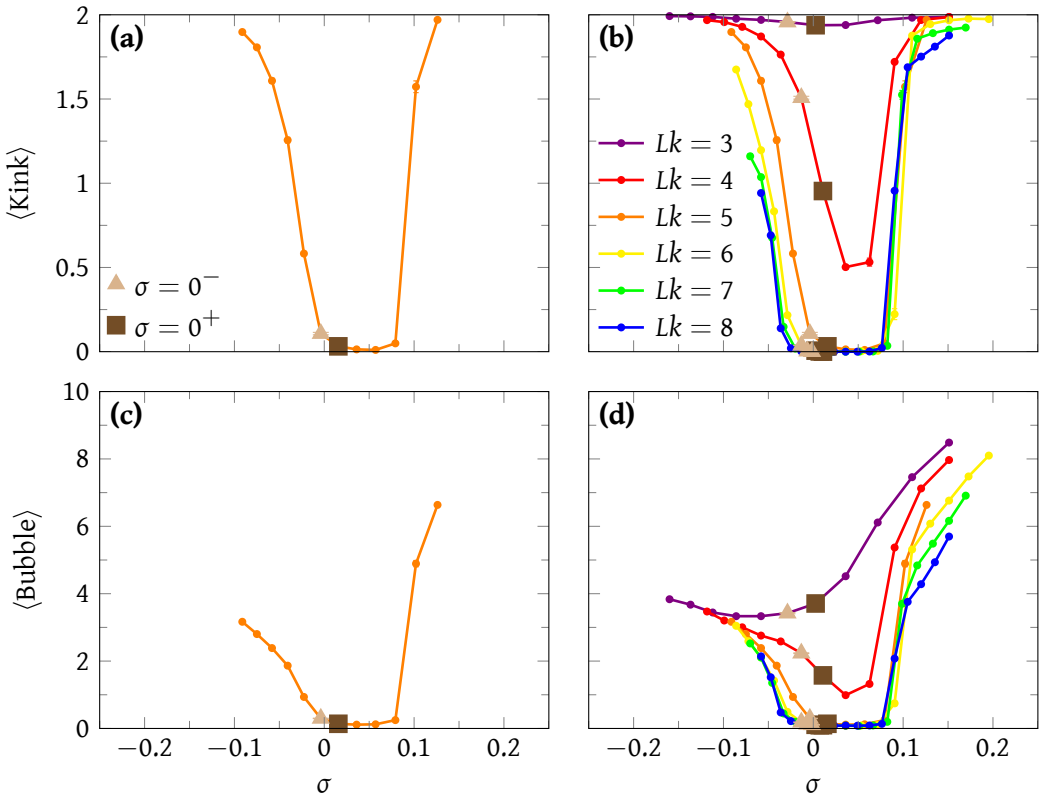


Figure 7.4 Average number of kinks and total bubble size, superhelical density
Average number of kinks ⟨Kink⟩ versus σ for intact minicircles with **(a)** $Lk = 5$, and **(b)** $Lk = 3$–8. Average total bubble size ⟨Bubble⟩ versus σ for **(c)** $Lk = 5$, and **(d)** $Lk = 3$–8.

with a small constant bubble size across the entire minimum in $\langle \text{Kink} \rangle$ (Figure 7.1 (d)).

7.3.3.3 Large superhelical density

Curiously, at $\sigma \gtrsim 0.1$, kinking in overtwisted minicircles is accompanied by larger bubble sizes than achieved by undertwisted minicircles of either comparable $\langle \text{Kink} \rangle_{\text{duplex}}$ or $|\sigma|$. What then, causes enhanced kinking and bubble formation for large overtwist (Figure 7.4 (b,d))?

First, we note that short minicircles ($N_{\text{bp}} \lesssim 35$, $Lk = 3$) are so highly stressed that two kinks are necessary to relax all minicircles (undertwisted, relaxed and overtwisted), whereas the relaxation mechanisms are distinct at longer lengths.

Secondly, we observe that beyond a certain superhelical density σ and length N_{bp}, oxDNA reveals that overtwisted minicircles relax the superhelical stress via writhing, rather than by increased twist (*i.e.* shorter pitch length). For sufficiently large minicircles, we would expect a 'figure-eight' plectonemic structure involving continuous bending; however, for the minicircle lengths considered in this study we do not observe this behavior because such a structure would lead to extreme curvatures at the two plectoneme 'tips' of the figure-eight.

Thirdly, that kinking occurs for different physical reasons in undertwisted and overtwisted minicircles resolves the dilemma. For overtwisted minicircles, instead of a figure-eight, kinking occurs at the plectoneme tips, yielding 'X-like' structures with two relatively straight duplex sections separated by two kinked regions of high curvature (Figure 7.2 (h)). The enhanced bubble formation for large overtwisting occurs because such extreme curvature at the plectoneme tips requires a significant number of base pairs to open, in addition to kinking. At longer N_{bp}, we would expect the duplex sections to become more curved, with less extreme bending at the plectoneme tips, eventually approaching a continuously bent figure-eight at sufficiently long N_{bp}. We do not probe this regime.

On the other hand, undertwisted minicircles do not writhe, but instead form two bubbles (Figure 7.2 (g)). Here, plectonemic writhe is disfavored because, as noted above, extreme curvature would be induced at the plectoneme tips. As bubbles readily absorb undertwist, but not overtwist, it is more favorable to absorb the undertwisting as bubbles than as writhe.

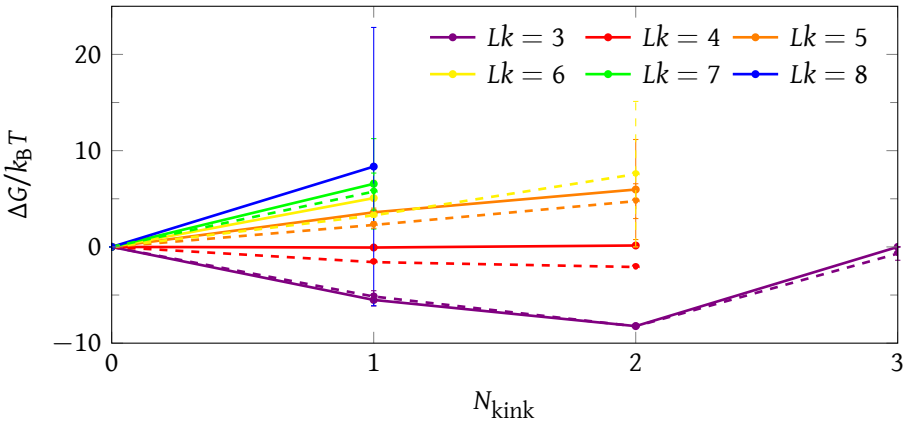


Figure 7.5 Lack of cooperativity in the number of kinks, intact relaxed minicircles
Free energy versus number of kinks N_{kink} for $\sigma = 0^+$ (**solid**) and $\sigma = 0^-$ (**dashed**) intact minicircles. The reference state is $N_{kink} = 0$ for each system.

7.3.4 Thermodynamics of kinking

The oxDNA model allows for the study of minicircle thermodynamics in fine detail; in particular, the underlying cause of the emergent phenomena of kinking, which we rationalize in terms of the properties of duplex DNA bending.

One interesting effect is the possibility of cooperativity in kink formation for intact relaxed minicircles. Yan & Marko[24] have determined that if a kink acts as a free joint, the expectation from worm-like chain (WLC) theory is that the first kink forms a 'teardrop' configuration, with some bending stress localized to the sole kink and the remainder to the continuously bent region opposite the kink. This configuration relieves ∼30 % of the bending stress, while the formation of a second kink in the previously continuously bent region relieves the remaining ∼70 % of the original bending stress.

Earlier we noted that kinks in the duplex form at much longer lengths in nicked minicircles, compared to their intact relaxed counterparts. This is contrary to our *a priori* expectation. It is intuitive that for nicked minicircles, there is a smaller local cost of forming a kink at the nick than in the duplex. Therefore, kinking at the nick occurs at longer N_{bp} than for kinking in the duplex (Figure 7.1). This is not the dilemma, rather it is why kinking in the *duplex* occurs at longer N_{bp} in nicked than in intact minicircles.

We find that cooperativity in the nicked minicircles explains why kinks in the duplex are easier

to nucleate in nicked, as opposed to intact relaxed minicircles (Section 7.4.2, Table 7.1). Further, oxDNA reveals a structural and thermodynamic basis for the observed cooperativity: kinks in the duplex do not behave as a free joint, while kinks at a nick do.

To better understand these cooperative effects, we evaluate the free-energy as a function of the number of kinks for intact relaxed minicircles (Figure 7.5). For the intact relaxed minicircles, the free-energy gain/loss for the first and second kinks are comparable, suggesting a lack of cooperativity. For the nicked system, the free energy of forming of the second kink is clearly more favorable, as evidenced by the longer length at which duplex kinks begin to occur.

The reason for the different behavior of the nicked and intact minicircles can be rationalized from the free-energy profile of a DNA duplex as a function of end-to-end distance.[i] For a nicked strand, once the kink at the nick is formed, there is very little free-energy cost associated with bringing the two duplex ends closer together (*i.e.* increasing the bend angle at the nick). On the other hand, there is a significant free-energy penalty in bringing the two ends of an intact duplex closer together, even after a kink is formed.

The different costs of changing the bending angle at a kink at a nick, versus a kink in a duplex region, underlies the divergent behavior of the nicked and intact minicircles. In particular, the free-energy cost of forming a second kink involves both the cost of forming a new tightly-bent kink, and the cost of increasing the bending angle at the first kink. This latter cost is significantly smaller when the first kink is at a nick, making a subsequent kink in the duplex region opposite more favorable.

7.4 Discussion

7.4.1 Comparison with experiment

The characteristic length at the onset of oxDNA's kink transition ($N_{\mathrm{bp}} = 62$–63) is in agreement with the predictions of a kinkable worm-like chain model.[16] There is also good agreement with experiment; specifically the minicircle digestion experiments of Du *et al.*,[36] who report detectable levels of relaxed DNA minicircle digestion at $N_{\mathrm{bp}} = 63$–64, lengths that correspond

[i] See Section 4.3 for a discussion of end-to-end distance or Section 4.3.5 for a discussion of the bending angle.

to the torsionally relaxed lengths $\sigma = 0^-$ and $\sigma = 0^+$ for six helical turns in their experimental conditions. The agreement between oxDNA and Du *et al.* should not be overinterpreted, as the comparison is muddled by the roles of salt concentration, sequence dependence and any small potential bias introduced by using enzymatic digestion as a reporter.

Firstly, oxDNA is parameterized to a different salt concentration ($[Na^+]$=500 mM) than the experimental conditions of Du *et al.* ($[Na^+] = 120$ mM, $[Mg^{2+}] = 2$ mM, $[Ca^{2+}] = 2$ mM). These solution conditions are not directly comparable because monovalent and divalent cations have differential impacts on DNA. For example, Mg^{2+} may aid kinking through structural effects,[18] and the role of even small concentrations of Mg^{2+} on DNA stability is non-trivial.[83]

Secondly, 50 % GC-content sequences, such as those used by Du *et al.*, are expected to kink more readily than one would naïvely infer from results obtained with the oxDNA average-base parameterization. Indeed, we have previously reported that kinks tend to localize to AT base pairs owing to the lower free-energy cost of disrupting AT as opposed to GC base pairs (Section 4.3.4.1, Section 6.5.1). As expected, the minicircles exhibit the same behavior, with an increased likelihood of kinking for 50 % GC-content sequences versus their average-base counterparts (Figure 7.1 (a), most pronounced at $N_{bp} = 41$–42). This shifts the midpoint of the kink transition to a somewhat longer N_{bp}. We also show enhanced kinking in AT-rich regions (Section 7.4.4).

Thirdly, while digestion assays can report a qualitative signal of whether a particular substrate is liable to digestion or not, it is non-trivial to extract a quantitative report of the fraction of kinked molecules. A quantitative interpretation would require knowledge of the characteristic time required for the digestion enzyme to bind and digest a duplex bubble/kink. Additionally, enzymatic activity may displace the equilibrium between kinked and unkinked states, requiring careful experimental conditions (*e.g.* very low enzyme concentrations and an enzyme bind/digest time-scale much longer than that of the kinked/unkinked interconversion). Supporting this reasoning, Du *et al.* only report cleavage of $N_{bp} = 63$–64 minicircles using the BAL-31 nuclease, not the S1 nuclease, which they find to only cleave negatively supercoiled minicircles. Moreover, the lack of cleavage by S1 nuclease raises questions about nuclease sensitivity, kink stability in the buffer conditions required for enzyme activity and the presence of structural defects not

	Torsionally relaxed intact		Nicked	
N_{bp}	$\Delta G^1_{duplex}/k_BT$	$\Delta G^2_{duplex}/k_BT$	$\Delta G^1_{nick}/k_BT$	$\Delta G^1_{duplex}/k_BT$
31	-5.51 ± 1.74	-2.72 ± 0.11		-4.45 ± 0.16
32	-5.15 ± 1.74	-3.10 ± 0.12		-4.65 ± 0.31
41	-0.06 ± 0.06	0.21 ± 0.06	-4.49 ± 1.01	-3.06 ± 0.12
42	-1.58 ± 0.21	-0.51 ± 0.06	-5.03 ± 1.73	-3.22 ± 0.08
51	3.59 ± 0.12	2.37 ± 0.88	-4.11 ± 0.34	-1.30 ± 0.13
52	2.28 ± 0.15	2.47 ± 0.41	-3.72 ± 0.22	-1.28 ± 0.09
62	5.07 ± 0.36		-2.30 ± 0.09	0.90 ± 0.10
63	3.28 ± 0.06	4.28 ± 1.00	-2.54 ± 0.15	1.00 ± 0.08
72	6.55 ± 0.72		-0.93 ± 0.06	2.82 ± 0.05
73	5.75 ± 0.34		-1.20 ± 0.12	2.76 ± 0.21
82	8.35 ± 1.73		0.22 ± 0.06	4.26 ± 0.43
83			0.25 ± 0.03	4.02 ± 0.40

Table 7.1 Free energy of kink formation in intact and nicked minicircles
Free energy of kinking in the duplex (ΔG_{duplex}) and at a nick (ΔG_{nick}) for the intact relaxed and nicked minicircles. Error bars represent standard error of the mean from 5 independent simulations. Blank spaces in the table indicate a lack of sampling, and therefore an undefined free energy. For a fair comparison of cooperativity in the nicked minicircles, compare ΔG^1_{duplex} from the intact with ΔG^1_{duplex} from the nicked minicircles, the second and fifth columns (see text for details).

detected by the enzyme.

7.4.2 Cooperativity in kink formation

The free energy versus the number of kinks is a useful metric for examining potential cooperativity in the number of kinks; however, it can be somewhat misleading to compare intact and nicked minicircles through diagrams such as Figure 7.5. In Table 7.1 we present an alternate representation to fairly compare kinks in both intact and nicked minicircles.

For the intact relaxed minicircle, the free energies for forming the first and second duplex kinks are given as:

$$\Delta G^1_{duplex} = -k_BT \ln\left(\frac{P_1}{P_0}\right), \tag{7.4}$$

$$\Delta G^2_{duplex} = -k_BT \ln\left(\frac{P_2}{P_1}\right), \tag{7.5}$$

where $\Delta G^1_{\text{duplex}}$ and $\Delta G^2_{\text{duplex}}$ are free energies of the first and second duplex kinks respectively, and P_0, P_1 and P_2 ($P_0 + P_1 + P_2 = 1$) are the probabilities of having zero, one and two kinks, respectively (Section 7.4.3, Figure 7.6).

For the nicked minicircle, the first kink is formed at the nick, while the second is formed in the duplex:

$$\Delta G^1_{\text{nick}} = -k_{\text{B}}T \ln\left(\frac{P_1}{P_0}\right), \tag{7.6}$$

$$\Delta G^1_{\text{duplex}} = -k_{\text{B}}T \ln\left(\frac{P_2}{P_1}\right), \tag{7.7}$$

where ΔG^1_{nick} is the free energy of a kink at the nick.

With the data available in Table 7.1, we now infer cooperativity in the intact relaxed minicircles. Cooperativity of a kink in the duplex of the nicked minicircles is available by comparison of $\Delta G^1_{\text{duplex}}$ between the intact and nicked systems.

For intact minicircles, if there is cooperativity in the number of kinks, given that lower free energies are more favorable, then we would expect the free-energy of the first kink to be greater than the second ($\Delta G^1_{\text{duplex}} > \Delta G^2_{\text{duplex}}$). For anti-cooperative behavior, the relationship would be reversed ($\Delta G^1_{\text{duplex}} < \Delta G^2_{\text{duplex}}$), while the values would be about the same in the absence of cooperativity ($\Delta G^1_{\text{duplex}} \approx \Delta G^2_{\text{duplex}}$).

We observe a lack of cooperative, or possibly slight anti-cooperativity, for the intact minicircles, with $\Delta G^1_{\text{duplex}} \gtrsim \Delta G^2_{\text{duplex}}$ for all lengths where there is sufficient sampling.

For the nicked minicircles, we of course always find that it is more favorable to form a kink at the nick than a kink in the duplex. To detect cooperativity, we should instead compare the free-energy of forming a duplex kink in the nicked system to that of forming the first kink in an intact minicircle of equivalent size. The observation of a lower free-energy of formation for the duplex kink in the nicked system indicates cooperativity.

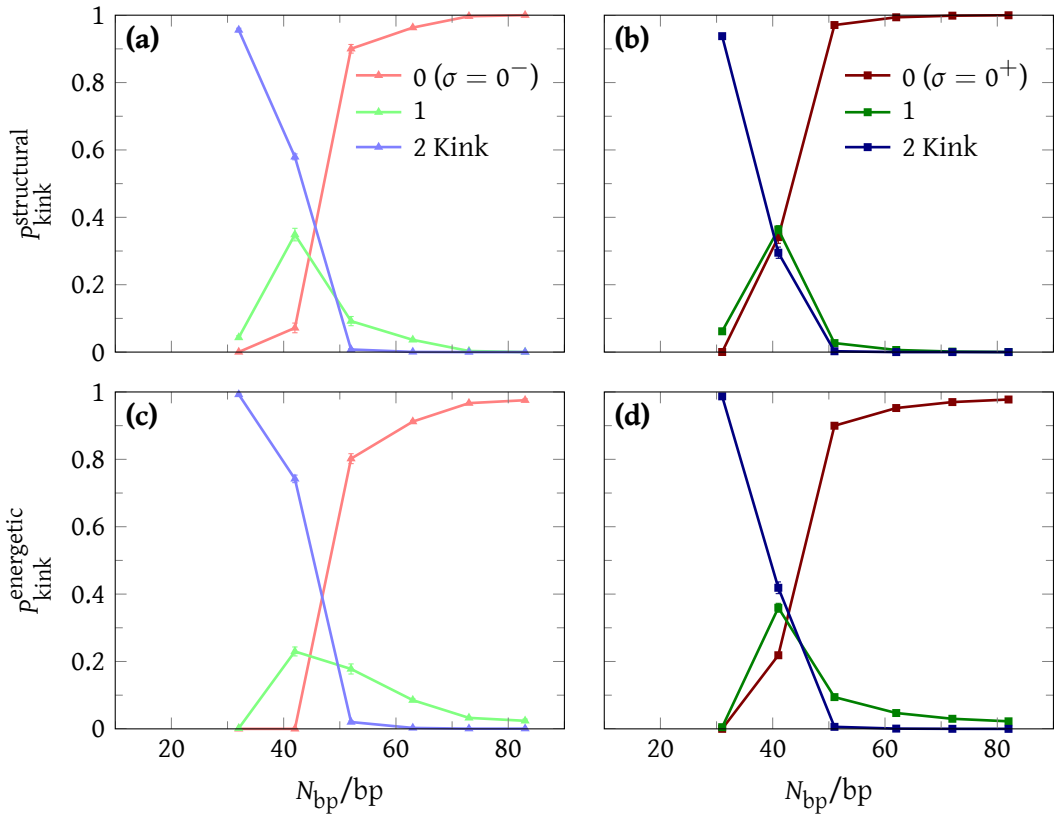


Figure 7.6 Probability of kinking, intact relaxed minicircles
Probability of kinking P_{kink} for 0, 1 and 2 kinks ($P_{\text{kink}} = P_0 + P_1 + P_2$) for intact relaxed minicircles with the **(a,b)** structural criterion and **(c,d)** energetic criterion for a kink with oxDNA's average-base parameterization.

7.4.3 Probability of kinking

In Figure 7.1 (a,c,e) we showed the average number of kinks versus minicircle length N_{bp} using a structural criterion for a kink. The choice of the average number of kinks may be somewhat misleading when comparing over N_{bp} because the total number of kinks changes as a function of length. It may also be difficult to distinguish configurations with true intermediate kinked states versus those that are an ensemble average of kinked and unkinked populations. Distinguishing low occupancy configurations may also be problematic.

In Figure 7.6 and Figure 7.7 we present the probability of zero (P_0), one (P_1) and two (P_2) kinks versus minicircle length N_{bp} for intact relaxed and nicked minicircles, respectively. As $P_0 + P_1 + P_2 = 1$, this representation resolves the aforementioned issues, but sacrifices at a glance knowledge of the number of kinks.

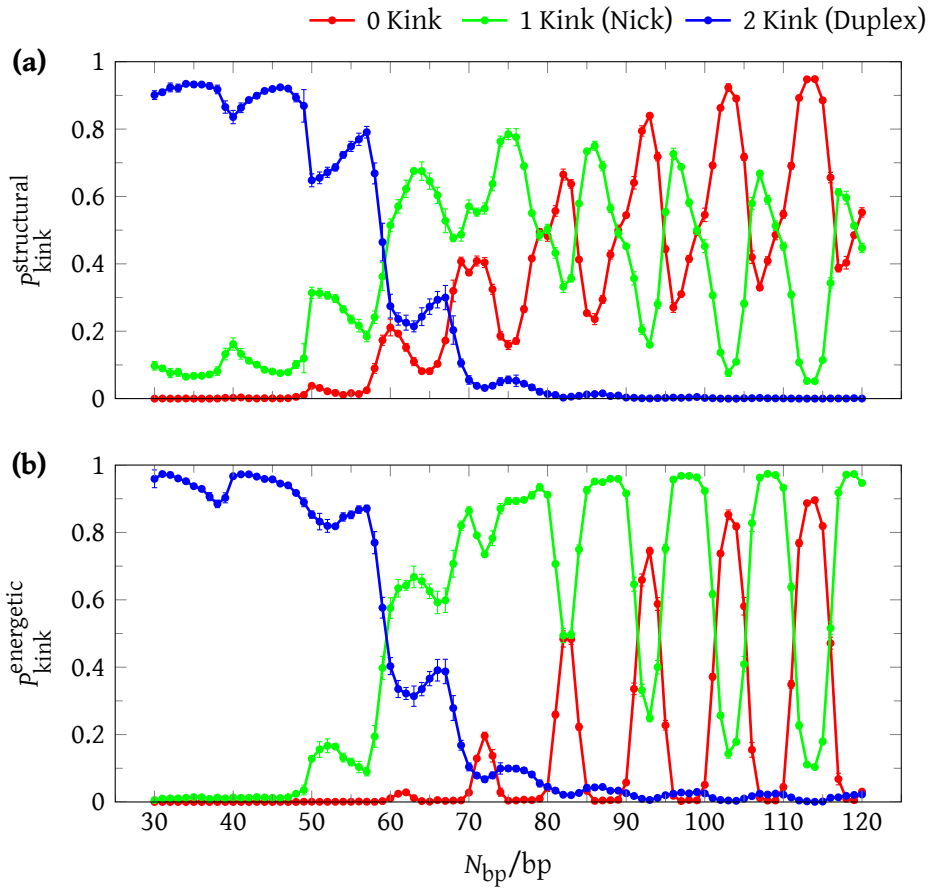


Figure 7.7 Probability of kinking, nicked minicircles
Probability of kinking P_{kink} for 0, 1 and 2 kinks ($P_{\text{kink}} = P_0 + P_1 + P_2$) for nicked minicircles with **(a)** structural kink criterion and **(b)** energetic kink criterion.

For example, in Figure 7.6 (a,b) an intermediate state with one kink, a 'teardrop' configuration, becomes apparent at $N_{bp} = 41–42$. From looking only at the number of kinks (Figure 7.1 (a)), one might have guessed that the transition is from $0 \rightarrow 2$ kinks, with no intermediate state with one kink.

As another example, in Figure 7.7 (a), the anti-correlated periodicity in P_0 and P_1 is quite clear at $N_{bp} = 80–120$. This behavior can be assumed from Figure 7.1 (e), but is strikingly obvious when examining kink probability instead of number of kinks. We have also investigated minicircles with two opposing nicks; unsurprisingly, the onset of kinking occurs at even longer N_{bp} than the single-nicked minicircles (data not shown).

The choice of the structural detector may also underestimate the prevalence of kinking (Section 4.4.1). To demonstrate that the structural definition indeed captures the basic physics of interest, we also show the probability of kinking with an alternative energetic kink criterion. Recall that energetically, a kink is defined as a disruption in stacking between adjacent base pairs (Section 3.2.2.2).

For minicircles, the results for the structural and energetic criteria are quite similar (Figure 7.6, Figure 7.7). Further explanation of the subtle differences between the two detectors are available in Section 4.4.1.

7.4.4 Localization of kinks

In Figure 7.8 we report the localization of kinks, defined with respect to the base pair index, where base pair index i is the position between consecutive bases i and $i + 1$ along an intact strand. For comparison of different lengths we shift the base pair index by $(N_{bp} - 2)/2$.

The localization of kinks for intact relaxed minicircles is relatively flat (Figure 7.8 (a,c)), as one would expect from the average-base parameterization. Unsurprisingly, for the sequence-dependent parameterization, the distribution is not flat at the shortest lengths, and contains peaks corresponding to AT-rich regions.

For nicked minicircles (Figure 7.8 (b,d)), we might naïvely expect kinks in the duplex to localize in a unimodal distribution opposite the kink at the nick. Curiously, unlike a kink at a nick,

whose position along the strand is unimodally distributed around the kink, the position of a kink in the duplex has a bimodal distribution, with maxima $\pm 1/4$ pitch length from the position opposite the nick. Bimodal distributions are observed for both on-register and off-register lengths. The structural basis for the observation of bimodal distributions and enhanced kinking at 'off-register' lengths is not yet entirely clear, though it may be due to a preference for in-plane kinking into the groove.

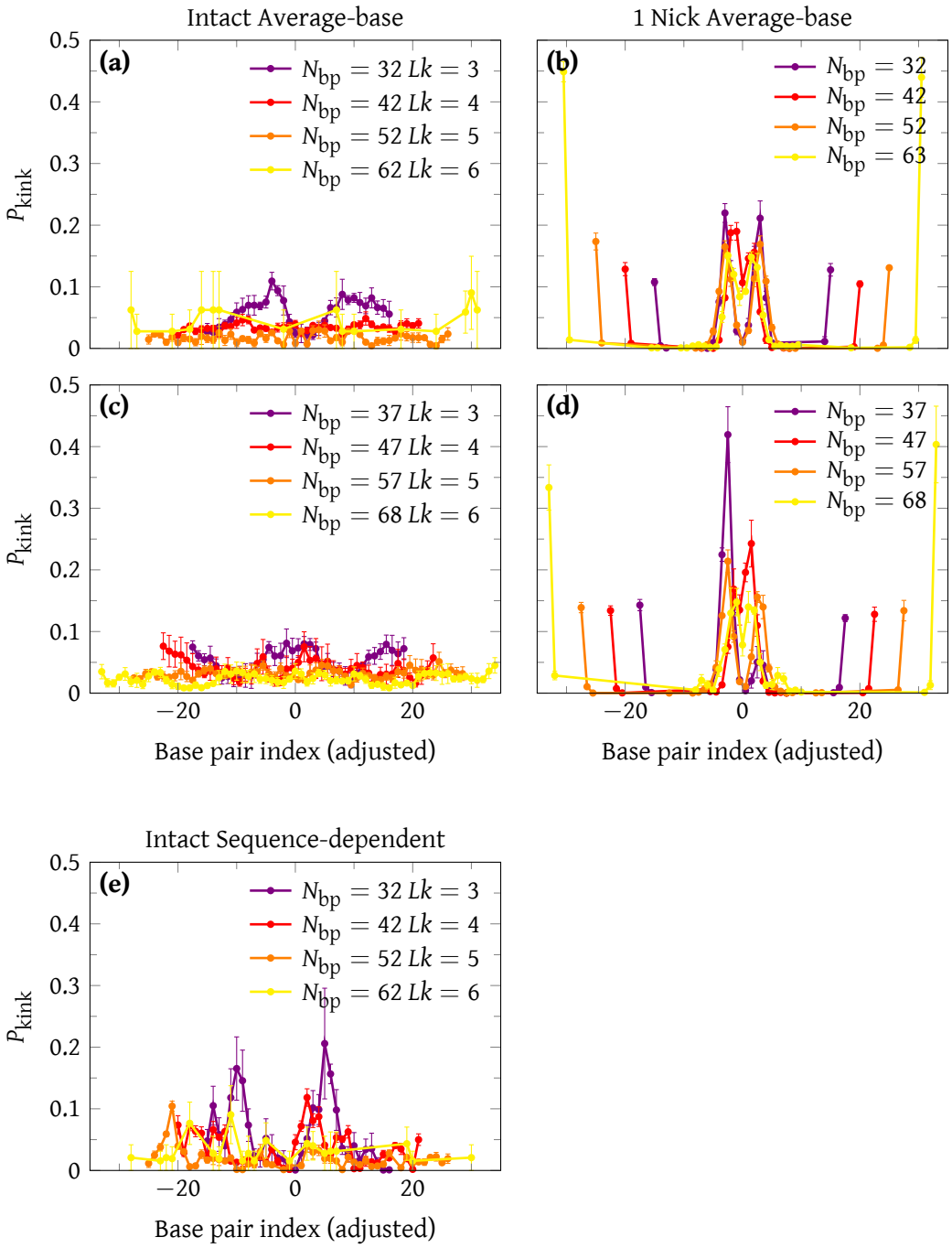


Figure 7.8 Kink localization, nicked minicircles
Localization of kinks along the intact strand with the average-base parameterization for 'on-register' **(a)** intact relaxed and **(b)** nicked minicircles, as well as 'off-register' **(c)** intact relaxed and **(d)** nicked minicircles. The sequence-dependent parameterization for **(e)** 'on-register' intact relaxed minicircles is also shown. P_{kink} is the probability of kinking via the structural criterion at a given base pair index, normalized by the number of kinked molecules. In **(b,d)**, the maxima at base pair index $\pm(N_{bp})/2$ corresponds to a kink at the nick, while the maxima at $N_{bp} = 0$ corresponds to kinks in the duplex region.

7.5 Conclusion

The oxDNA coarse-grained model elucidates the basic physics and underlying thermodynamics of strong DNA bending in DNA minicircles. The model is in good agreement with existing theory,[16,24,101] all-atom MD simulations[25,65,66] and experiment.[18,26,36] For example, oxDNA finds kinking in relaxed DNA minicircles at 62–63 base pairs, but not 82–83 base pairs: this apparent agreement with minicircle digestion experiments[36] suggests that kinking is indeed the structural defect reported in experiment.

While it is unlikely that the agreement between digestion assays and oxDNA is coincidental, we caution the reader against over-interpretation. The experimentally observed digestion of relaxed $N_{\mathrm{bp}} = 63$–64 minicircles does not necessarily imply quantitative bounds on kinking, and while reasonable, the comparison of the Du *et al.* data with the *onset* of kinking in oxDNA is arbitrary. Taken together with strong agreement to a 'molecular vise' (Chapter 5) and DNA cyclization (Chapter 6), we interpret comparable minicircle results as a promising sign that the oxDNA predictions are physically reasonable and useful in helping to elucidate the basic physics of highly bent DNA.

Beyond direct experimental comparison, oxDNA supports kinking as the structural basis behind experimental observations. For example, in supercoiled minicircles, we find that overtwisted minicircles kink to accommodate writhing, while their negatively supercoiled counterparts' undertwist is absorbed by bubbles. These also act as kinks to relieve bending stress.

OxDNA also makes several, physically reasonable, predictions; for example, cooperative localization of kinks and correlated kink diffusion, finding that duplex kinks localize opposite each other on intact minicircles and opposite the nick for nicked minicircles. Further, oxDNA directly addresses the conundrum of duplex kinking at longer lengths in nicked, as opposed to intact minicircles, revealing cooperativity of kink formation in the former, but not in the latter. The observation of cooperativity is substantiated by both thermodynamic and structural arguments based on the angular dependence of the free-energy of kink formation present in duplex kinks, but not in kinks at a nick.

8 Model system: 'Strained duplex'

8.1 Introduction

Experimentally, it is difficult to examine strong DNA bending in linear duplex DNA without the aid of a model system. The laboratory of Giovanni Zocchi at UCLA (University of California, Los Angeles), amongst others, has characterized the use of single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) segments as molecular springs and tethers. First applied to DNA-protein chimeras,[108,109] Qu *et al.* extended the method to examine the thermodynamics of sharply bent DNA duplexes.[19,20,78]

8.1.1 Concept of the 'strained duplex'

The model system, which we have dubbed the 'strained duplex' for simplicity, is produced by the hybridization of a linear to a circular strand (Figure 8.1 (a)), creating a molecule with a distinct double-stranded region, which is useful for probing the bending of intact DNA. Alternatively, hybridization of a linear strand to two ends of another longer linear strand yields a molecule suitable for probing the bending of a nicked duplex (Figure 8.1 (b)). For clarity, schematic representations of a linear duplex, strained intact and strained nicked duplex are depicted in Figure 8.2.

For convenience, the molecules are labeled by the lengths of their double-stranded (N_d) and single-stranded (N_s) regions, where the circular ssDNA length is given by $C = N_d + N_s$. For nicked molecules, C is the length of the nicked molecule, which is invariably longer than the complement linear strand.[i]

[i] DNA Minicircles, discussed in Chapter 7, occur in the limit where $C = N_d$.

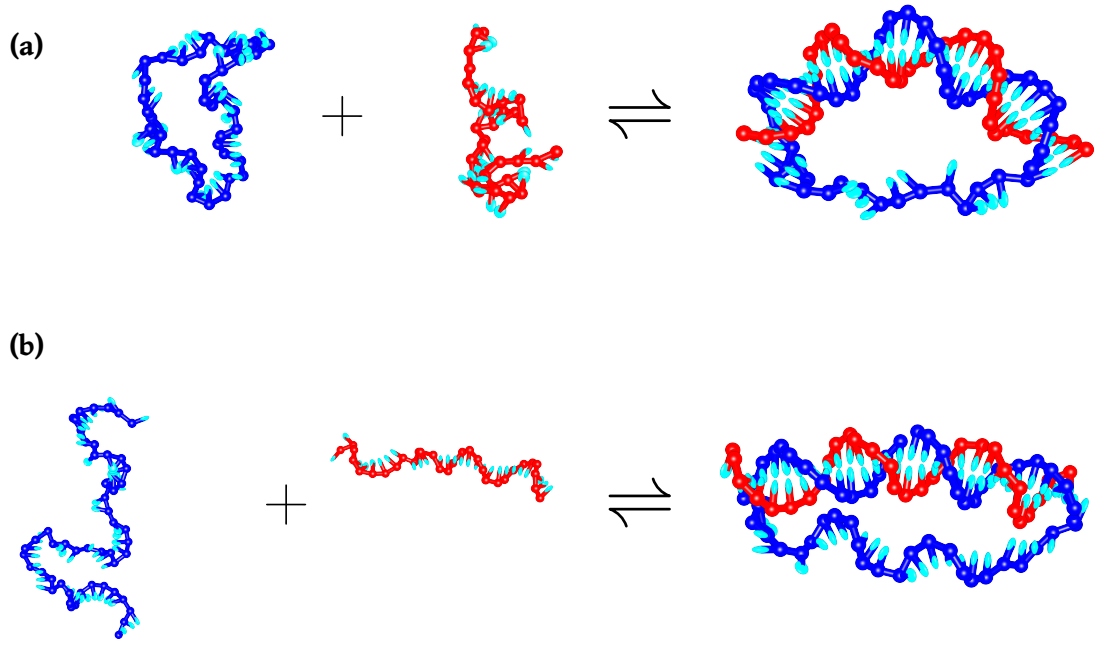


Figure 8.1 Assembly of intact and nicked 'strained duplexes'
OxDNA representation of the assembly of **(a)** intact and **(b)** nicked 'strained duplexes'. The intact strained duplex is assembled from a 45 bp circular strand and a 30 bp duplex yielding a strained duplex with a duplex of length $N_d = 30$ and a single-stranded tether of length $N_s = 15$. The nicked strand is assembled from a 30 bp and a 54 bp duplex, yielding a nicked duplex of length $N_d = 30$ (nick is at the mid-point of the duplex), and a single-stranded tether of length $N_s = 24$.

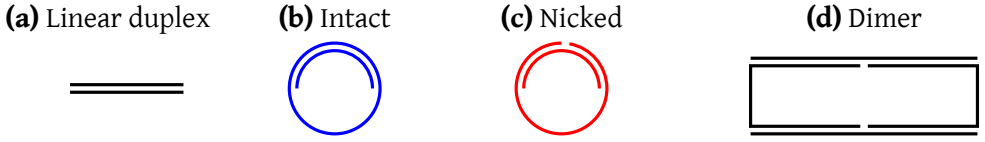


Figure 8.2 Schematic of strained duplexes
Schematic of strained duplex model system developed by Qu *et al.*[78]

Importantly, varying the length of the single-stranded region (of length N_s) modulates the stress imposed on the double-stranded region (of length N_d). As N_s becomes shorter, an increasing amount of stress is imposed on the duplex. Following our thermodynamic characterization of strong bending in Chapter 4, for sufficiently large bending stress, *i.e.* sufficiently short N_s, structural defects would be expected.

Indeed, Qu *et al.* propose kinking (Section 2.1.2) as the structural defect responsible for relieving bending stress in the duplex.[19] While all-atom MD simulations of DNA minicircles do support this view,[25,65] it is difficult to experimentally probe specific structural defects; instead, Qu *et al.* develop models to infer the presence of kinks from their experiments.

8.1.2 Specific aims

We use oxDNA to probe DNA bending in the 'strained duplex' model system. We directly observe kinking of the strained duplex, one of several structural defects that facilitate the relaxation of bending stress. Continuous bending, fraying and bubble formation are also observed. Where possible, we compare directly to experiment.

8.2 Experimental approaches

Two complementary experimental approaches have been developed to probe strong DNA bending with the strained duplex, 'monomer-dimer' equilibrium[78] and melting curve analysis.[19] The monomer-dimer equilibrium approach is only applicable to nicked systems, while melting curve analysis is applicable to both nicked and intact molecules. These approaches are not without precedent, as melting curve analysis in combination with a statistical model has previously been used to probe DNA flexibility.[110]

Unsurprisingly, Qu *et al.* report similar results for both the monomer-dimer equilibrium and melting curve results. Importantly, they assert that a discontinuous buckling instability between continuously bent and kinked dsDNA exists beyond a 'critical' torque (τ_c).[ii] They also conclude

[ii] The transition is not 'critical' in a classical thermodynamic sense, a second-order phase-transition of a bulk system. Unfortunately, given the abundance of references to 'critical' in the work of the Zocchi laboratory, it would be even less clear to omit the phrase. Therefore, we interpret 'critical' to mean a discontinuity between the two states, as we suspect was the author's intention: indeed, they use 'critical' and 'buckling instability' interchangeably.

that the presence of a nick in the middle of the double-stranded region has negligible impact on this 'critical' torque ($\tau_c^{\text{intact}} = 30\,\text{pN nm}$[19] versus $\tau_c^{\text{nick}} = 27\,\text{pN nm}$[20]).

In Section 8.2.1 and Section 8.2.2 we overview the approach of Qu *et al.*, laying a foundation for our critique in Section 8.6.

8.2.1 Monomer-dimer equilibrium

Here we introduce the theory Qu *et al.*[78] have developed to interpret their monomer-dimer equilibrium data. The Qu *et al.* analysis has two conceptual parts. First, we overview their approach to extracting the free energy of dimer formation from their gel-based assays (Section 8.2.1.1), then we turn our attention to the analytic expressions used to extract the elastic energy (free energy of bending) of the stressed monomer (Section 8.2.1.2).

8.2.1.1 Experimental concept

Nicked systems may anneal in either a monomeric (Figure 8.1 (b)) or a dimeric (Figure 8.1 (d)) configuration. In the dimer, two duplex regions (with nicks) are joined by two single-stranded regions at either end. If the monomer and dimer chemical potentials are written as

$$\mu_M = \mu_M^{\text{ref}} + k_B T \ln(X_M), \tag{8.1}$$

$$\mu_D = \mu_D^{\text{ref}} + k_B T \ln(X_D), \tag{8.2}$$

where μ^{ref} are reference chemical potentials, then using the condition for monomer-dimer equilibrium

$$2\mu_M - \mu_D = 0, \tag{8.3}$$

it follows that

$$\Delta G_{M\to D}^{\text{ref}} = \mu_M^{\text{ref}} - \frac{\mu_D^{\text{ref}}}{2}, \tag{8.4}$$

$$= \frac{1}{2} k_B T \ln\left(\frac{X_D}{X_M^2}\right), \tag{8.5}$$

where X_D and X_M are the total mole fractions of the dimer and monomer respectively. Qu *et al.* determined X_D and X_M by interpolating monomer and dimer concentration from sample loading concentrations and the relative intensities of gel electrophoresis bands.[78] They ignore the impact of higher order oligomers.

8.2.1.2 Analytic expression for elastic energy $\left(E_{el}\right)$

Importantly, Qu *et al.*[78] assume that $\Delta G^{ref}_{M\to D}$ is simply related to the change in elastic energy for dimerization, *i.e.*

$$\Delta E_{el} = \Delta G_{M\to D}. \tag{8.6}$$

They further assume that ΔE_{el} is dominated by the elastic energy of the monomer, applying a small correction to account for screened electrostatic repulsion between the double-helical sections of the dimer, and some resulting strain in the connecting single-stranded sections. Thus, they obtain E_{el}, the elastic energy stored in the monomer, which they partition such that the elastic energy in the duplex is due to bending, while the elastic energy in the single-stranded section is due to non-zero extension:

$$E_{el} = E_d(x_{eq}) + E_s(x_{eq}), \tag{8.7}$$

where E_d is the duplex energy, E_s is the single-stranded energy, and x_{eq} is the end-to-end distance (EED) of both the single-strand and the duplex. Amongst other omissions, this expression assumes that there will be no fraying of the duplexes, a potential response to applied stress (Section 5.3.2, Appendix A.1.2). At equilibrium:

$$\left.\frac{\partial E_d}{\partial x}\right|_{x_{eq}} + \left.\frac{\partial E_s}{\partial x}\right|_{x_{eq}} = 0. \tag{8.8}$$

For E_d they choose the analytic expression

$$E_d = \begin{cases} \tau_c \arccos\left(\dfrac{x}{2R}\right) & \text{for } 0 < x < x_c \\ \dfrac{5B_s}{L}\dfrac{x_0 - x}{2L} - k_B T \ln\left(\dfrac{2L - x}{2L - x_0}\right) & \text{for } x_c < x < x_0, \end{cases} \tag{8.9}$$

where $2L = 0.33\,\text{nm} \times N_\text{d}$ is the contour length (0.33 nm is for the rise per base pair along the helical axis), B_S is the bending modulus of double-stranded DNA ($50\,k_\text{B}T\,\text{nm}$),[iii] $R = L\left(\frac{1-2\gamma^2}{45}\right)$ and $\gamma = \frac{L\tau_\text{C}}{2B_\text{S}}$. x is the end-to-end separation between duplex ends, while x_0 is the mean of the distribution, which due to entropic effects will be shorter than the contour length, and is given as $x_0 = 2L\left(1 - \frac{Lk_\text{B}T}{5B_\text{S}}\right)$.

Values in parentheses are taken from Qu *et al.*[19] The values would be slightly different for oxDNA:[52,57] $B_\text{S} = 41.82\,k_\text{B}T\,\text{nm}$ at 298 K and the rise per base pair along the helical axis is 0.34 nm.

Importantly, x_C is a 'critical' length at which there is a discontinuous jump from a sharply bent to a continuously bent regime. The equations for the latter is an approximate worm-like chain (WLC) expression which does not account for duplex force-extension, and is therefore only valid in the range $x_\text{C} < x < x_0$. The former equation is their expression for non-WLC behavior ($0 < x < x_\text{C}$), which they presume to be a kink, and which is governed by a single free parameter, the 'critical' torque τ_C. The presumption of a 'critical' torque and the resulting discontinuous transition between states (as opposed to a thermodynamically smooth transition), as well as the limitation of having only one free parameter, will be important for our subsequent discussion (Section 8.6).

E_S is the truncated polynomial expansion of the Marko-Siggia expression:[111]

$$E_\text{S} = \frac{9k_\text{B}T}{4N_\text{S}l_\text{S}^2}\left[x^2 + \frac{x^3}{N_\text{S}l_\text{S}} + \frac{3x^4}{(N_\text{S}l_\text{S})^2}\right], \tag{8.10}$$

where l_S is the single-stranded segment length ($l_\text{S} = 0.764\,\text{nm}$). Finally, to determine τ_C, Equation (8.9) is fit to the monomer-dimer equilibrium values of E_el. Again, note that τ_C is explicitly given as the sole free parameter; that is, use of the preceding equations presumes the existence of a 'critical' torque.

[iii] Recall that the persistence length L_P may be related to the bending modulus B_S by Equation (2.5).

8.2.2 Melting curve analysis

In a complementary approach Qu *et al.*[19] determine the melting curves for linear, intact (Figure 8.1 (a)) and nicked (Figure 8.1 (b)) model systems. Akin to the monomer-dimer equilibrium experiments (Section 8.2.1), the Qu *et al.* analysis has two conceptual parts. The first concerns the acquisition of experimental melting profiles (Section 8.2.2.1); the second concerns the development of their analytical model (Section 8.2.2.2).

8.2.2.1 Experimental concept

Melting is the result of an equilibrium between DNA dissociation and annealing, given by the reaction

$$\mathrm{A} + \mathrm{A}' \underset{k_\mathrm{d}}{\overset{k_\mathrm{a}}{\rightleftharpoons}} \mathrm{AA}' \tag{8.11}$$

where k_a is the rate constant for association, k_d is the rate constant for dissociation, A is ssDNA, A′ its complement stand and AA′ the annealed dsDNA.

DNA dissociation and annealing are impacted by a host of 'exogenous' environmental factors including temperature, pH and salt concentration, as well as 'endogenous' factors such as DNA concentration and sequence. All of these factors influence the melting curve, a measure of the dissociation characteristics of duplex DNA, which can be conveniently expressed in terms of the fractional yield of dissociated molecules (f_d) by either concentration (assuming equilibrium) or rate (assuming steady-state). Expressed in terms of concentration, f_d is given as

$$f_\mathrm{d} = \frac{[\mathrm{A}]}{[\mathrm{A}] + [\mathrm{AA}']}, \tag{8.12}$$

It is plausible then, that if other factors, such as molecular strain, alter the relative rates of dissociation and association, that they impact f_d and therefore the melting curve. Qu *et al.*[19] report melting curves, determined by optical absorbance at 260 nm (A_{260}), for intact, nicked and linear duplex molecules.

8.2.2.2 Analytic expression for elastic energy (E_{el})

To extract the elastic free energy of the nicked and intact systems from the melting curves, Qu *et al.* fit a zipper model[112] to the data for the three systems.[iv] Fitting to the linear duplex result allows the enthalpy cost (ε) and entropy gain (σ) of opening a linear duplex base pair to be obtained, a negative control. For the nicked system they assume that the elastic free energy for the kinked state is *purely enthalpic* and takes the form

$$A_{\mathrm{el}}^{\mathrm{nicked}} = E_{\mathrm{el}}^{\mathrm{nicked}} = \alpha_{\mathrm{nicked}} \sqrt{N_{0,\mathrm{nicked}}}, \tag{8.13}$$

when fully base paired, where α_{nicked} and $N_{0,\mathrm{nicked}}$ are parameters. The absolute values of ε and σ are not fully constrained by the fit to the linear duplex, and so values are chosen that allow $E_{\mathrm{el}}^{\mathrm{nicked}}$ to agree with the room temperature dimerization experiments.

For the intact system, the elastic free energy for the kinked state[v] is assumed to have the form

$$A_{\mathrm{el}}^{\mathrm{intact}} = E_{\mathrm{el}}^{\mathrm{intact}} = \alpha_{\mathrm{intact}} \frac{\sqrt{N_{0,\mathrm{intact}}}}{T} + 3\left(\varepsilon - \sigma T\right), \tag{8.14}$$

when fully base-paired where α_{intact} and $N_{0,\mathrm{intact}}$ are parameters, and there is assumed to be an entropic component to the elastic free energy that is equivalent to the breaking of 3 terminal base pairs.[vi]

From the parameters, both assumed, and extracted from fit to the melting profiles, the room-temperature elastic free energies are estimated using Equation (8.13) and Equation (8.14). 'Critical' torques are then obtained from these A_{el} using the same procedure as for the experiment for dimerization of linear duplexes. Typical values obtained are $\tau_{\mathrm{c}}^{\mathrm{nicked}} = 27$ pN nm and $\tau_{\mathrm{c}}^{\mathrm{intact}} = 31$ pN nm.[20]

[iv] Although the use of a zipper model is somewhat antiquated, it is well-established.
[v] Recall that they presume that the non-WLC expression governed by τ_{c} in Equation (8.9) represents kinked states.
[vi] Note that the entropic gain for breaking three internal base pairs in a bubble would be lower.

8.3 Simulation methods

We conduct structural and thermodynamic simulations of the stranded duplex. Structural simulations are conducted with virtual-move Monte Carlo (VMMC) (Section 3.1.3) at $T = 23\,^{\circ}\mathrm{C}$. Thermodynamic simulations for the melting profiles and base pairing were performed at either $T = 300\,\mathrm{K}$ or $T = 340\,\mathrm{K}$. Melting profiles were determined from histogram reweighing over the temperature range $T = 315$–$360\,\mathrm{K}$ (Appendix A.5.4). Bending free energy simulations (umbrella sampling over EED) are excerpted from Chapter 4. The average-base parameterization is used to elucidate the basic physics of the system, while the sequence-dependent parameterization is used for a closer comparison with experiment. Complete details are given in Appendix A.5

8.4 Structural characterization of the 'strained duplex'

8.4.1 Intact 'strained duplex'

As discussed in the introduction (Section 8.1.1), the length of the single-stranded region (N_s) modulates the bending stress imposed on the duplex region (N_d), with shorter N_s inducing more bending stress than longer N_s. In Figure 8.3, we show the N_d and N_s dependent behavior of the intact strained duplex with respect to kinking and fraying for the oxDNA average-base parameterization.

It should come as no surprise that the relative contributions of various structural deformations, namely continuous bending, kinking and fraying, differ markedly by length. We find that for the longest duplexes ($N_d = 60$–65) continuous bending is the predominant deformation at large bending stress (small N_s). For intermediate-length duplexes ($N_d = 25$–55), kinking is the predominant structural defect. Finally, for short duplexes ($N_d < 25$), fraying becomes increasingly prevalent.

As the underlying physical mechanisms are identical, the ordering of structural defect by duplex length, from shortest to longest, fraying $\rightarrow$ kinking $\rightarrow$ continuous bending, is in line with oxDNA results for other model systems, *e.g.* the 'molecular vise' (Chapter 5) and DNA cyclization (Chapter 6).

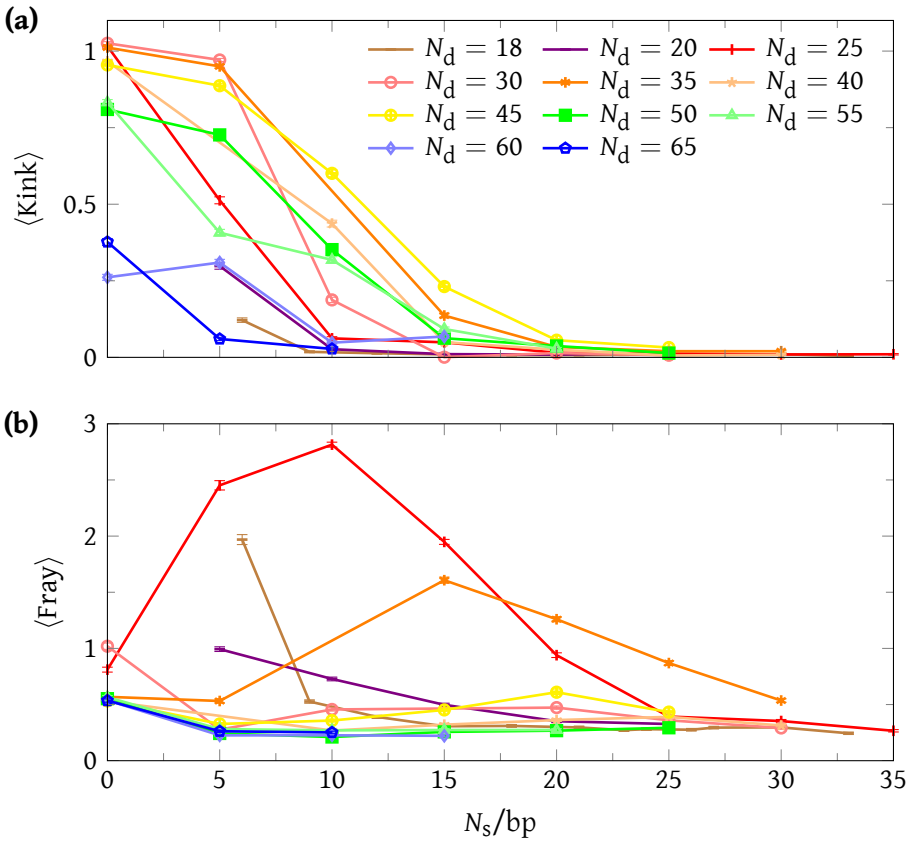


Figure 8.3 Length-dependent behavior of strained duplexes
(a) Kinking for intact strained duplexes as a function of length using the average-base parameterization at $T = 23\,^{\circ}\mathrm{C}$, where $\langle \mathrm{Kink} \rangle$ is the average number of kinks in the duplex. **(b)** Fraying, where $\langle \mathrm{Fray} \rangle$ is the average number of frayed base pairs. The different length-dependent behavior regimes are summarized as: Short ($N_d < 25$): fraying predominates. Intermediate ($N_d = 25$–55): kinking predominates, and saturates to $\langle \mathrm{Kink} \rangle \to 1$ as $N_s \to 0$. Long ($N_d = 60$–65): continuous bending dominates, with kinking decreasing to the very long length ($N_d \gg L_P$) limit. Sequences from Qu *et al.*[19,20,78]

In very long duplexes ($N_d \gg L_P$, where L_P is the DNA persistence length), negligible structural deformations would be expected. Mind, however, that negligible is not zero. Thermally driven structural deformation will still occur, albeit exceedingly rarely for relatively high-energy states such as kinks. For example, at the biologically-relevant twist density of $\sigma = -0.06$, although often safely ignored, spontaneous denaturation does occur.[113] We expect that kinking will eventually approach its very long length scale limit, which is often negligible, but rarely strictly zero. Although we do not probe the very-long-length limit with the strained duplex, the approach towards negligible kinking and deformation by continuous bending is evident (Figure 8.3 (a)).

The impact of changing the duplex length N_d below the persistence length ($N_d \times 0.332\,\text{nm} < L_P$)[vii] is somewhat less obvious. At the longest lengths examined ($N_d = 60$–65), bending stress is partitioned between continuously-bent and kinked states, which are in equilibrium (Figure 8.4 (a)). As we have previously reported (Section 4.3.2, Section 5.3.3, Section 6.3.1.1), the approach of kinking to the very-long-length limit ($\langle\text{Kink}\rangle \to 0$) appears faster than might be naïvely expected from a very-long-length limit of $N_d \times 0.332\,\text{nm} \gg L_P$. At $N_d = 60$, less than ½ of the DNA persistence length, saturation kinking is already only ¼ of its value at $N_d = 30$. Therefore, kinking approaches the very long length scale limit as $N_d \times 0.332\,\text{nm} \to L_P$ is stricter and more accurate statement of the systems observed behavior. This result was not predicted *a priori*, but is consistent with our understanding of DNA bending thermodynamics, which are fully explored in Chapter 4.

For intermediate duplex lengths ($N_d = 25$–55), N_d saturates to $\langle\text{Kink}\rangle \to 1$ at the shortest N_s. As expected in the presence of less bending stress, kinking decrease as N_s increases. This behavior is analogous to that of the longer N_d molecules.

The trend towards kinking at large bending stress does not hold for the shortest duplexes ($N_d \lesssim 25$); instead, unlike the long and intermediate duplex length molecules, the absence of kinking is due to fraying, not continuous bending. For example, at $N_d = 25$, as N_s decreases fraying goes through a maximum at $N_s = 10$, before again decreasing and giving way to kinking at shorter N_s.

[vii] Note that 0.332 nm is the base pair rise in oxDNA;[52] necessary because L_P is given in units of nm instead of bp.

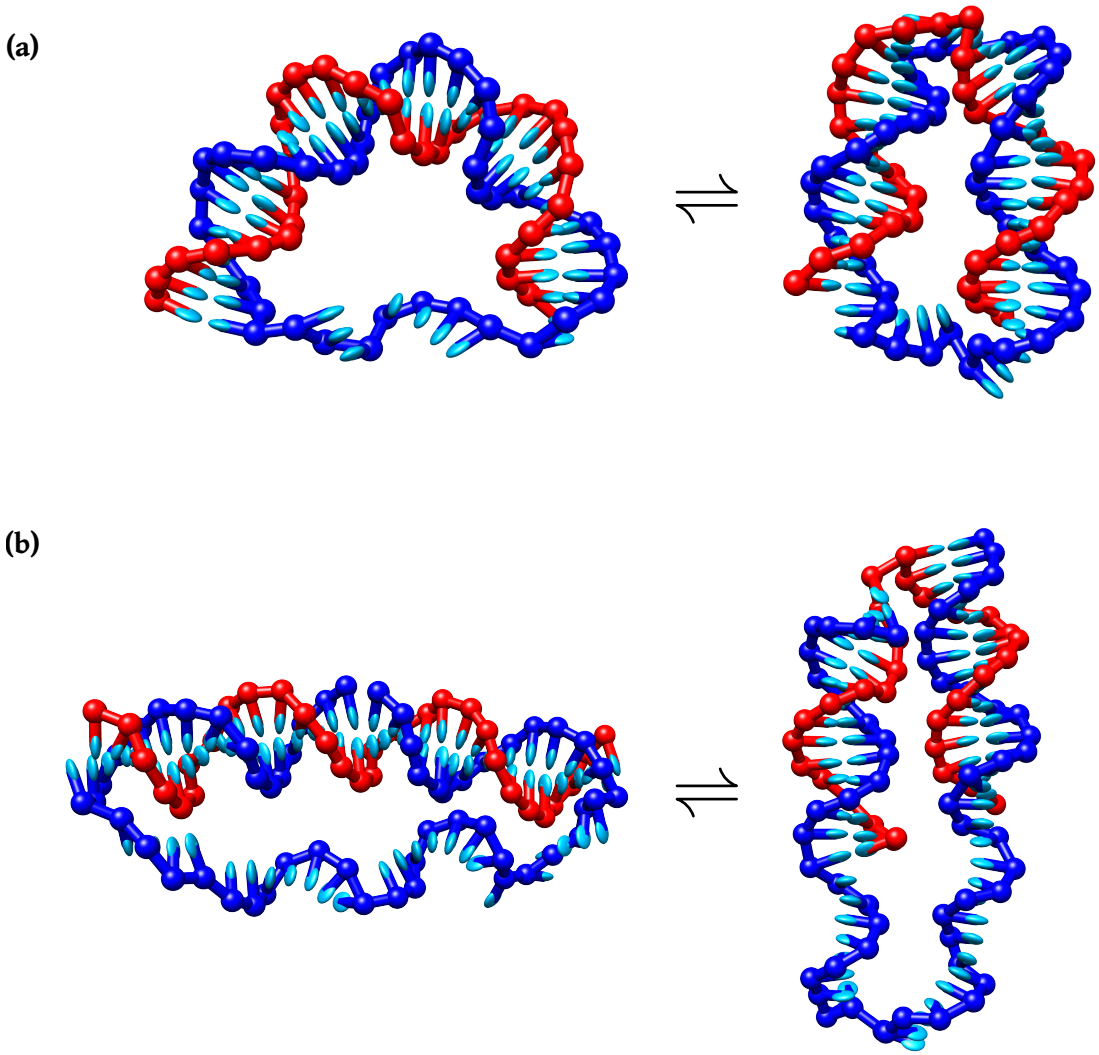


Figure 8.4 Equilibrium between continuously bent and kinked strained duplexes
Equilibrium between continuously bent and kinked states for **(a)** an intact molecule ($N_d = 30, N_s = 8$) and **(b)** a nicked molecule ($N_d = 30, N_s = 24$), both in the kink transition region.

8.4.2 Nicked 'strained duplex'

Comparison of strained intact and nicked duplexes highlights interesting physics in both systems. We find that for a given duplex length, kinking occurs more readily in the nicked (Figure 8.4 (b)), as opposed to intact molecules(Figure 8.4 (a)). Curiously, we also find that the transition to fraying at the shortest duplex lengths is conspicuously absent for the nicked molecules, with kinking occurring in the shortest lengths examined, $N_\mathrm{d} = 18$.

The relative ease of kinking a duplex can be given in terms of the kink transition mid-point, the N_s at which the fractional occupancy of the kinked state is 50 %. For example, in an intermediate-length duplex ($N_\mathrm{d} = 30$), we find that nicked molecules have a kink transition mid-point at $N_\mathrm{s} \approx 25$, as opposed to $N_\mathrm{s} \approx 10$ for their intact counterpart (Figure 8.5 (a)). On the contrary, fraying is only slightly enhanced for intact molecules (Figure 8.5 (b)).

Different kink transition points for the intact and nicked molecules should be unsurprising given the oxDNA results and supporting experimental comparisons of previous chapters (Chapter 6, Chapter 7), as well as the thermodynamic investigation of Chapter 4, where we demonstrated the lower free-energy cost of kinking at a nick and a clear structural basis for the phenomenon.

In practice, kinking of intact duplexes requires bubble formation; however, bubble formation is not necessary for kinking at a nick. In addition to not requiring the disruption of base pairs, much more rotational freedom is allowed around a kink at a nick as opposed to a kink in the duplex (Compare the nicked molecule in Figure 8.4 (b) with the intact molecule in Figure 8.4 (a)).

Figure 8.5 (a,c) shows that by $N_\mathrm{d} = 30$, which is representative of intermediate-length duplexes, fraying is mainly thermal and is no longer responsible for DNA deformation under strain. However, the same does not hold for short duplexes ($N_\mathrm{d} \lesssim 25$). Although fraying is the dominant relaxation mode for short intact duplex, kinking continues as the predominant structural deformation for the nicked molecules. For a representative short duplex length ($N_\mathrm{d} = 18$), depicted in Figure 8.5 (c), the intact molecules do not have a kink transition, while the kink transition mid-point occurs at $N_\mathrm{s} \approx 14$ for the nicked molecule.

As the transition mid-points for oxDNA are significantly different from the conclusions of Qu *et al.* (we will give a detailed structural comparison in Section 8.4.5), we explore some of the

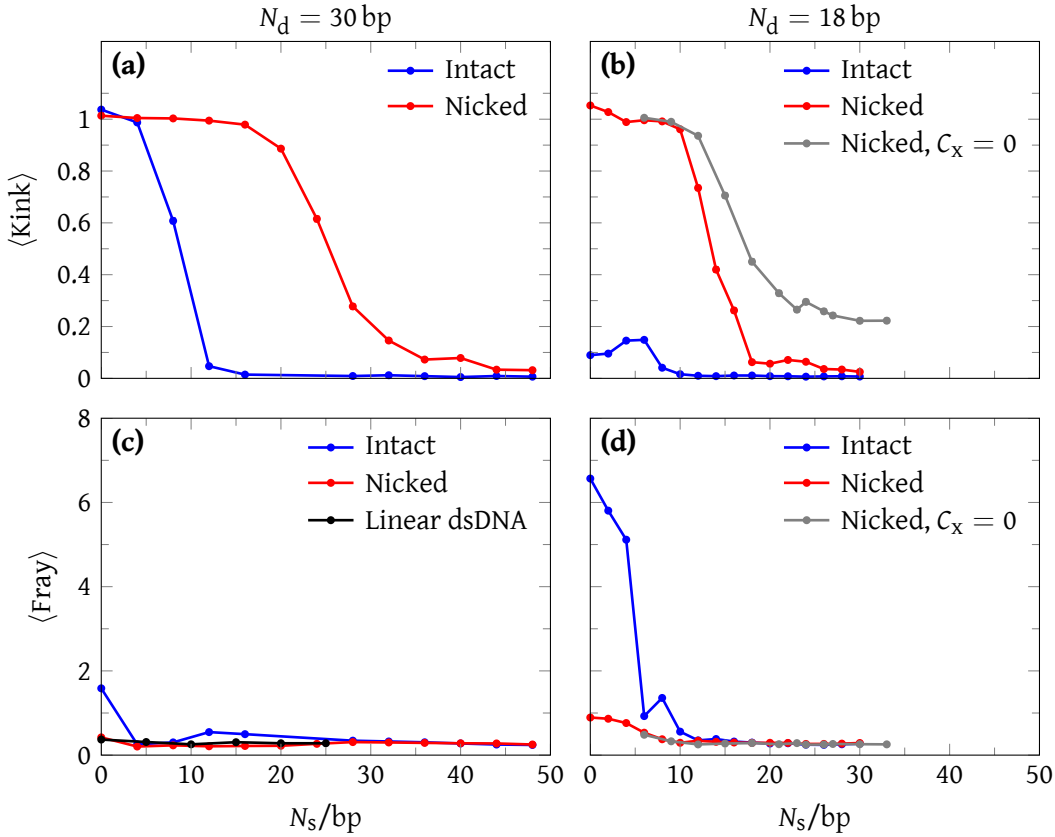


Figure 8.5 Comparison of intact and nicked strained duplexes
(a) Kinking for intact and nicked $N_d = 30$ molecules. The kink transition ($\langle \text{Kink} \rangle = {}^1\!/_2 \langle \text{Kink} \rangle|_{\max}$) occurs at a shorter N_s for the intact versus the nicked system. Given that lower N_s corresponds to a more strained system, this suggest that it requires more bending stress to kink the intact system.
(b) Kinking for $N_d = 18$ molecules. Kinking is abolished for the intact molecules, but not for their nicked counterpart.
(c) Fraying of intact and nicked $N_d = 30$ molecules.
(d) Fraying for $N_d = 18$ molecules. While fraying is the primary structural deformation for the intact system, the nicked systems exhibits minimal fraying. Also note that setting $E_{C_x} = 0$ (coaxial stacking) has a negligible affect upon fraying.
All systems use the average-base parameterization at 23 °C with a poly(A) N_s region (no secondary structure).

factors that might affect the position of the kink transition.

8.4.3 Structural defects: kink, bubble and fray

Frequency histograms, such as those at $N_d = 30$ (Figure 8.6) and $N_d = 18$ (Figure 8.7), neatly summarize the interplay and relative contribution of various deformations, particularly when comparing molecules possessing more (*e.g.* $N_s = 4$) and less (*e.g.* $N_s = 24$) strain.

For example, frequency histograms make evident the reduced total bubble size of nicked duplexes,[viii] and the fact that 5′ and 3′ fraying is symmetric, and therefore thermally governed, for the intermediate length intact system at $N_d = 30$, but not the short system at $N_d = 18$.

Specifically, for the intact molecules, fraying rather than kinking is the predominant structural deformation. This is evidenced by a very minor change in kinking with strain, negligible bubble formation, symmetric thermally-driven fraying for the $N_s = 24$ molecule, and asymmetric strain-induced fraying for the $N_s = 4$ molecule. For the nicked molecule, kinking predominates; however, unlike at intermediate duplex length, bubble and symmetric fraying formation in excess of the nicked duplex are observed. This implies higher relative strain free energies for short versus intermediate duplex lengths.

8.4.4 Specific structural effects

8.4.4.1 Secondary structure

The diligent reader may have already questioned the potential role of secondary structure in the single-stranded region on the kink transition mid-point. Sequences with more secondary structure would be expected to have a longer N_s at the kink transition mid-point. Energetically, maintaining a straight duplex region in the presence of single-stranded secondary structure would require overcoming the free-energy cost of breaking base pairing in the single-stranded region. Structurally, the presence of secondary structure effectively shortens the single-stranded region (*e.g.* DNA hairpins have a much smaller EED than their equivalent non-bonded single-strand).

[viii] A ‘bubble’ in a nicked duplex is actually fraying on either side of the nick. The term bubble is borrowed to distinguish from terminal fraying and for comparison with the intact duplex.

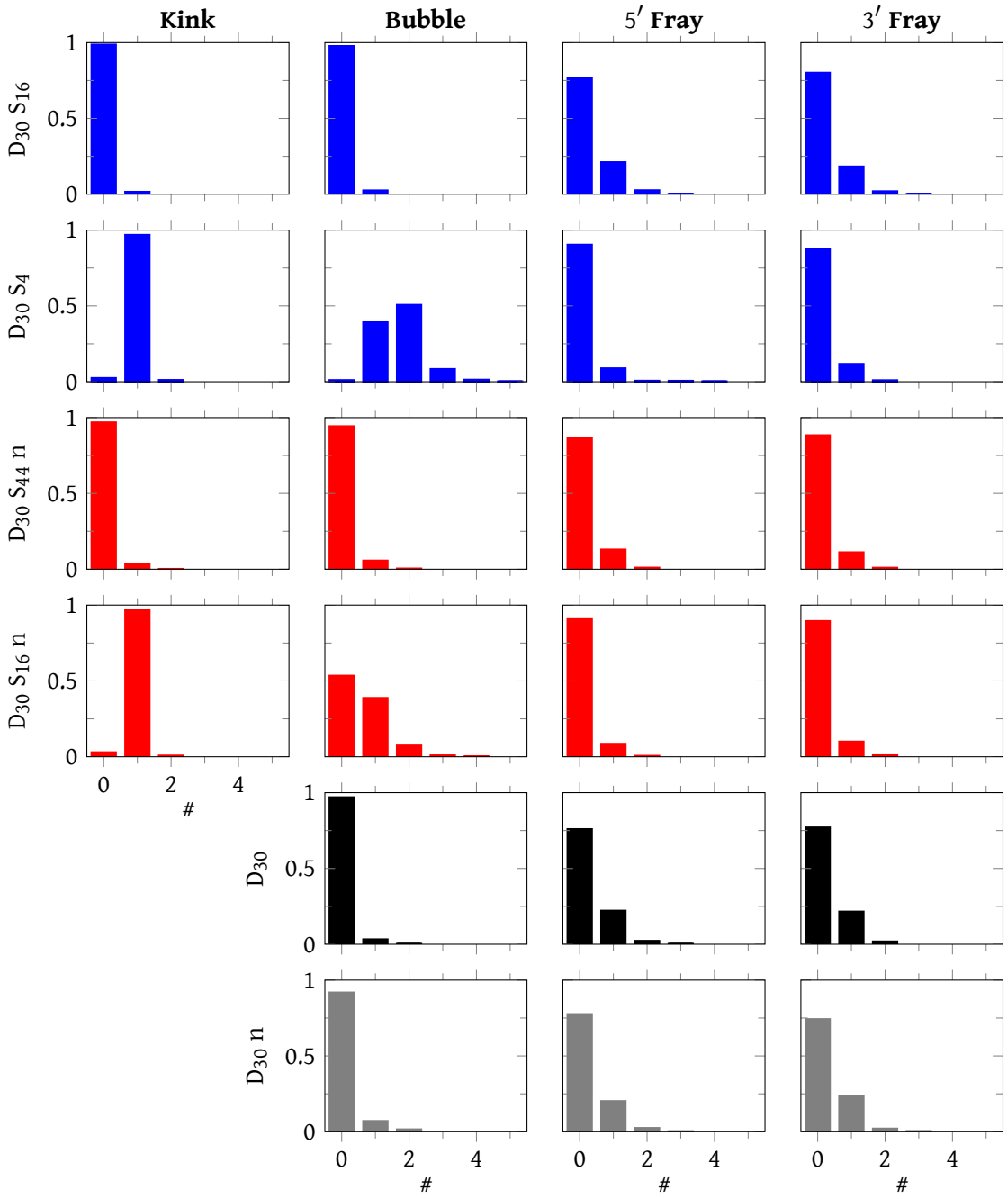


Figure 8.6 Kink, bubble and fray histograms for intermediate duplex length $(N_d = 30)$
Frequency histograms of Kink, Bubble, 5′ and 3′ Fray for intact (**blue**) ($N_d = 30$, $N_s = 16, 4$, poly(A)), nicked (**red**) ($N_d = 30$, $N_s = 44, 16$, poly(A)), 30 bp dsDNA (**black**) and 30 bp nicked dsDNA (**gray**). To eliminate ambiguity, and better illustrate the profiles for the various kinked states, histograms for Bubble, 5′ and 3′ Fray only include statistics for the dominant kinking behavior, either $\langle \text{Kink} \rangle = 0, 1$.
Nomenclature: D (N_d), S (N_s), n (nicked)

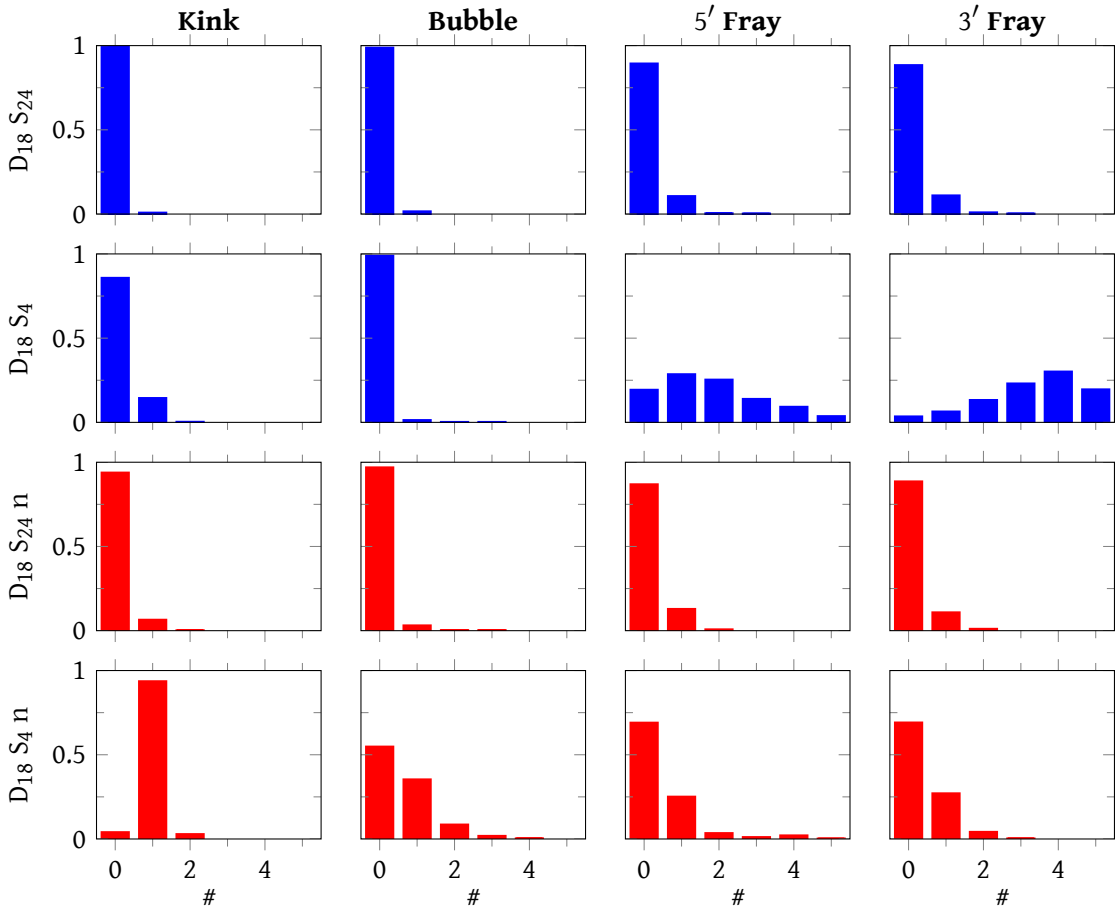


Figure 8.7 Kink, bubble and fray histograms for short duplex length $(N_d = 18)$
Frequency histograms of Kink, Bubble, 5′ and 3′ Fray for intact (**blue**) ($N_d = 18$, $N_s = 24, 4$, poly(A)), nicked (**red**) ($N_d = 18$, $N_s = 24, 4$, poly(A)), 30 bp dsDNA (**black**) and 30 bp nicked dsDNA (**gray**). To eliminate ambiguity, and better illustrate the profiles for the various kinked states, histograms for Bubble, 5′ and 3′ Fray only include statistics for the dominant kinking behavior, either $\langle \text{Kink} \rangle = 0$ or $\langle \text{Kink} \rangle = 1$. Nomenclature: D ($N_d$), S ($N_s$), n (nicked)

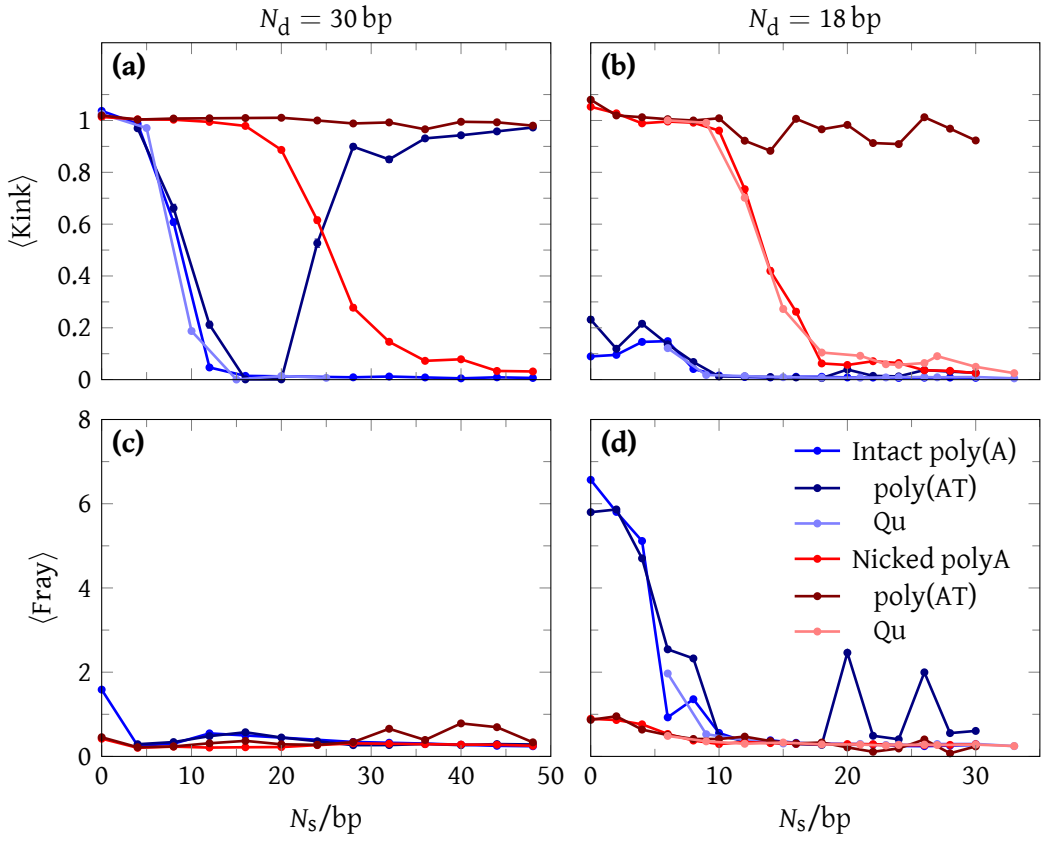


Figure 8.8 Effect of secondary-structure on the single-stranded region
Kinking and fraying for intact and nicked molecules of lengths $N_d = 30$ and $N_d = 18$. Kinking at **(a)** $N_d = 30$ and **(b)** $N_d = 18$. Fraying at **(c)** $N_d = 30$ and **(d)** $N_d = 18$.

To probe the impact of secondary structure, we vary the N_s sequence from that of Qu *et al.* For clarity, we use the average-base parameterization; sequence-dependent effects are explored in a later section (Section 8.4.4.2).

We choose the single-stranded region to be poly(A), which does not form secondary-structure.[ix] The effects of secondary structure are negligible for the Qu *et al.* sequence, the results being nearly identical to poly(A) (Figure 8.8).

As a positive control, we used the pathological AT-repeat (Figure 8.8). Hairpin formation effectively reduces N_s, by acting as an additional component of the ssDNA molecular spring. At $N_d = 30$, for the nicked molecule, extensive hairpin formation occurs, which forces the N_d region into the kinked state. For the intact case, hairpin formation does not occur until longer N_s because it is at first resisted by the greater cost of kinking in duplex DNA. Once a

[ix] By definition, secondary structure involves the formation of Watson–Crick base pairs, which is impossible for poly(A); however, poly(A) may form structures through interactions neglected by oxDNA.

sufficiently long hairpin is formed, the system adopts the kinked state. At $N_d = 18$, secondary structure induced by the pathologic sequence forces kinking in the nicked system by effectively shortening the single-stranded region. As kinking does not predominate at this length for the intact molecule, secondary structure has no effect.

8.4.4.2 Sequence dependence

Sequence plays a role in both duplex bending, and the strain imposed on the duplex by the single-stranded region. As the sequence-dependence of bending intact and nicked duplexes has already been examined in some detail (Chapter 4), we turn our attention to the role of the single-stranded region. We have previously noted that secondary structure does indeed perturb system behavior, with hairpin formation inducing kinking in both nicked and intact molecules. Of interest is whether the base pair composition of the single-stranded tether influences the kink transition mid-point. We choose to compare a poly(A) sequence with a poly(AG) sequence as they represent the strongest and weakest stacking sequences that cannot form secondary structure.

With the oxDNA sequence-dependent parameterization, we find that there is negligible difference in the behavior of these two sequences (Figure 8.9). Specifically, the kink transition mid-point and characteristic fraying is nearly identical for all N_s sequence variations of both the intact and nicked molecules. This is not surprising, given that the difference in stacking strength between the two sequences leads to quite minor differences in oxDNA's force-extension curves (see Figure 7 of 55).[x]

8.4.4.3 Temperature dependence

As both fraying and bubble formation are expected to become easier with temperature, we have briefly examined the effect of temperature dependence on the nicked $N_d = 18$ molecule (Figure 8.10). Below the melting transition, increased temperature has a minor effect on kinking; for example, the profiles for 23 °C and 40 °C are quite similar. Fraying, like kinking, is also largely temperature insensitive below the melting temperature. Unsurprisingly, fraying is slightly enhanced at higher temperatures.

[x] The oxDNA model is parameterized such that, enthalpically, A-A stacking is ∼6 % stronger than A-G stacking.

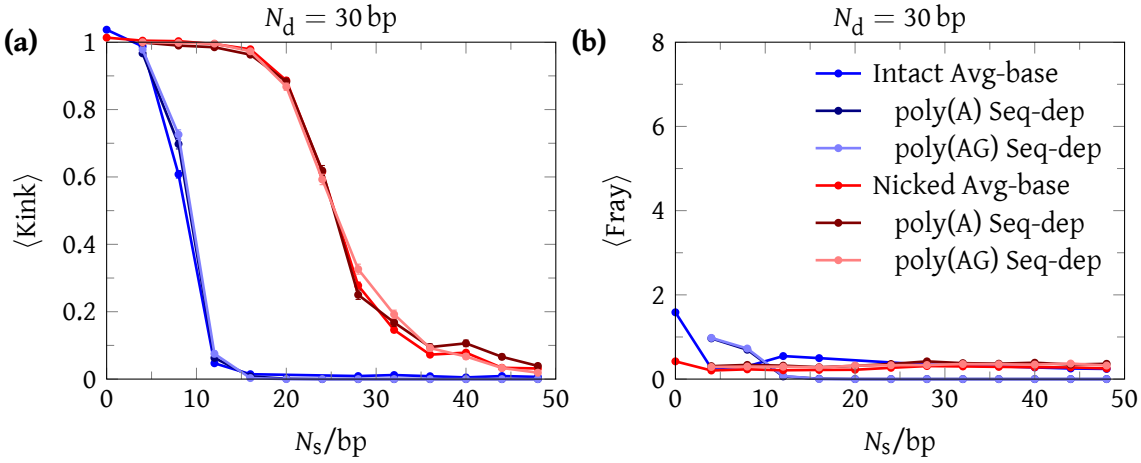


Figure 8.9 Role of sequence in the single-stranded region
(a) Kinking and **(b)** fraying for intact and nicked molecules of length $N_d = 30$ with the sequence-dependent parameterization. The sequence of the N_s region is varied, comparing the average-base parameterization (poly(A)) with the sequence-dependent parameterization for poly(A) and poly(AG).

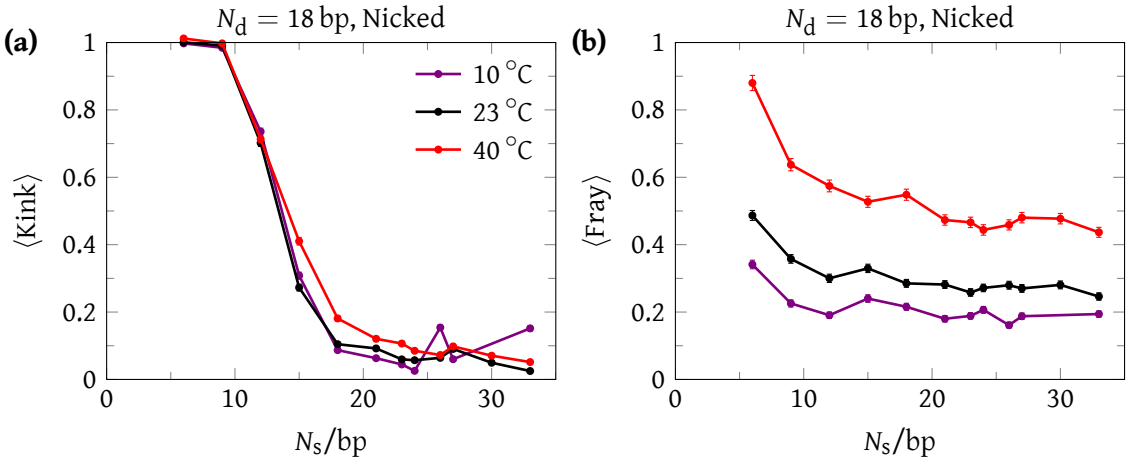


Figure 8.10 Temperature-dependence of nicked strained duplex
Kinking **(a)** and fraying **(b)** of a nicked $N_d = 18$ molecules as a function of temperature, using the sequence-dependent parameterization with sequences from Qu *et al.* .

8.4.5 Comparison with Qu *et al.*, structure

It is important to note that the structural properties of the strained duplex, in particular the kink transition, differ between oxDNA and Qu *et al.* Thermodynamic and free-energy comparisons are discussed in Section 8.5. Importantly, the discrepancy is a consequence of the analysis developed by Qu *et al.* (Section 8.2.1, Section 8.2.2), not the underlying experimental observations.

While oxDNA reports a kink transition midpoint of a thermodynamically smooth transition between continuously bent and kinked states, as a consequence of their analysis, Qu *et al.* presume a 'critical' buckling instability with a discontinuity between their WLC expression for continuous bending and their non-WLC expression for kinking (Equation (8.9)). For the monomer-dimer equilibrium analysis,[20] which is limited to nicked systems, Qu *et al.* report a 'critical' buckling instability at $N_s \approx 24$ for the $N_d = 18$ system and $N_s \approx 40$ for the $N_d = 24$ system.

These values of N_s are much longer than for oxDNA, which has kink transition mid-points of $N_s \approx 14$ for $N_d = 18$, and $N_s \approx 24$ for $N_d = 30$ (Section 8.4.2, specifically Figure 8.5). The longer N_s implies a combination of either a lower free-energy cost of kinking at the nick for Qu *et al.*, or a much stiffer single-stranded region as compared to oxDNA.

To test the former scenario we probed the role of coaxial stacking,[114] which dominates the interaction between the bases on either side of the nick and therefore the rigidity at the nick, for the nicked $N_d = 18$ molecule (Figure 8.5 (b,d)). The coaxial stacking interaction must be disrupted to kink at a nick (Section 2.1.2), so disabling the interaction should decrease the free-energy penalty to kink formation. Indeed, simulations run with the coaxial stacking term of the potential disabled reveal that coaxial stacking shifts the nicked kink profile to the right, but the effect is relatively modest. There is no impact on fraying (Figure 8.5 (d)).

Therefore, even when coaxial stacking is disabled, abolishing most of the attractive interaction between the two duplexes on either side of the nick, we cannot recapitulate the result of the Qu *et al.* analysis, finding that the kink transition midpoint is shifted just 4 nt, from $N_s = 14$ to $N_s = 18$ for $N_d = 18$.

As an observation about the latter scenario, in the oxDNA model, the Qu *et al.* sequences have negligible secondary structure (Section 8.4.4.1), suggesting that any potential difference in

effective stiffness is not due to the effects of secondary structure in the single-stranded region.

For melting curve analysis,[19] which can be used to investigate the intact and nicked strained duplexes, Qu *et al.* report nearly identical 'critical' torques for the nicked and intact systems: namely $N_s \approx 27$ for the $N_d = 18$ system and $N_s \approx 36$ for the $N_d = 24$ system.[xi] In our opinion, this is a consequence of their analysis, and does not necessarily reflect the underlying melting profiles (Section 8.5.2).

With oxDNA, we report a kink transition midpoint of $N_s \approx 8$ for $N_d = 30$ (Figure 8.5). In agreement with the underlying thermodynamics of DNA bending (Chapter 4, specifically Section 4.3.4.3, as well as Section 8.5.1), as well as results for DNA cyclization (Chapter 6, Section 6.5.3) and DNA minicircles (Chapter 7, Section 7.3.2), this shorter length suggests that the free-energy of kinking is lower for the nicked than intact duplexes. The results of Qu *et al.*'s analysis and oxDNA are even more dissimilar for $N_d = 18$, where we do not report kinking at any N_s for the intact system. Instead, the system relaxes by fraying, also in accordance with our thermodynamic understanding (Appendix A.1.2), and prior observation in other systems (Section 5.3.2,, Section 6.3.1.3).

The assumption of a 'critical' buckling instability aside, we have exhaustively investigated other plausible sources for the much longer N_s, and therefore lower kinking free energy, implied by the Qu *et al.* analysis. The effects of temperature (Section 8.4.4.3), sequence dependence (Section 8.4.4.2) and secondary structure (Section 8.4.4.1) do not explain the deviation. Rather, it is the inappropriate analysis of Qu *et al.* , as discussed at-length in Section 8.6.

8.5 Thermodynamics

Now that we have a clear structural overview of the strained duplex system, let us now turn our attention to a discussion of thermodynamics.

[xi] These values differ slightly from the monomer-dimer equilibrium assays, but τ_c is unaffected.

8.5.1 Free energy of bending, end-to-end distance (EED)

In Section 8.4, strain is imposed on intact and nicked duplex regions, N_d, by varying the length of a single-stranded molecular spring, N_s. The caveat of secondary structure aside (Section 8.4.4.1), by bringing the ends of the duplex region closer together, shorter N_s imposes a greater bending stress than longer N_s.

Although the use of single-stranded DNA as a molecular spring is a convenient experimental method to impose bending stress on a duplex, we are not restricted by physical mechanisms in simulation. With simulation, the underlying EED distribution can be sampled directly (Section 4.3), without the proxy of N_s. This approach allows direct access to the free-energy profile with respect to EED, providing a useful complement to the structural examination presented in Section 8.4.

Most duplexes will posses an EED near the contour length of $\sim$0.34 nm/bp. In accordance with the WLC model, we expect higher energies at EED far from the contour length: the free-energy cost of bending at shorter EED or the free-energy cost of over-stretching at longer EED. Duplex DNA overstretching with oxDNA is examined elsewhere,[56] so we concern ourselves only with DNA bending.

As explored in detail in Section 4.3, ΔG(EED) versus EED profiles reveal a characteristic shoulder (Figure 4.2 (a)). This shoulder is shifted to the right for the nicked molecule (Figure 4.4 (c)), implying that the transition occurs at longer EED, and has a lower free-energy cost of kinking.

Decomposing the free-energy profiles into continuously bent and kinked sub-populations, reveals that this shoulder is in fact due to a structural transition from continuously bent to kinked states (Figure 4.2 (a), Figure 4.4 (c)). At longer EED, continuously bent states are dominant, while kinked states are dominant at shorter EED (Figure 4.2 (b), Figure 4.4 (f)). The shoulder, indicating a change in slope of the free-energy profiles, coincides with the kink transition mid-point.

Taken together, these data suggest a thermodynamically smooth transition between continuously bent and kinked states. Moreover, they demonstrate the different EED transition regions of the intact and nicked molecules. In agreement with structural studies (Section 8.4), kinking at a nick

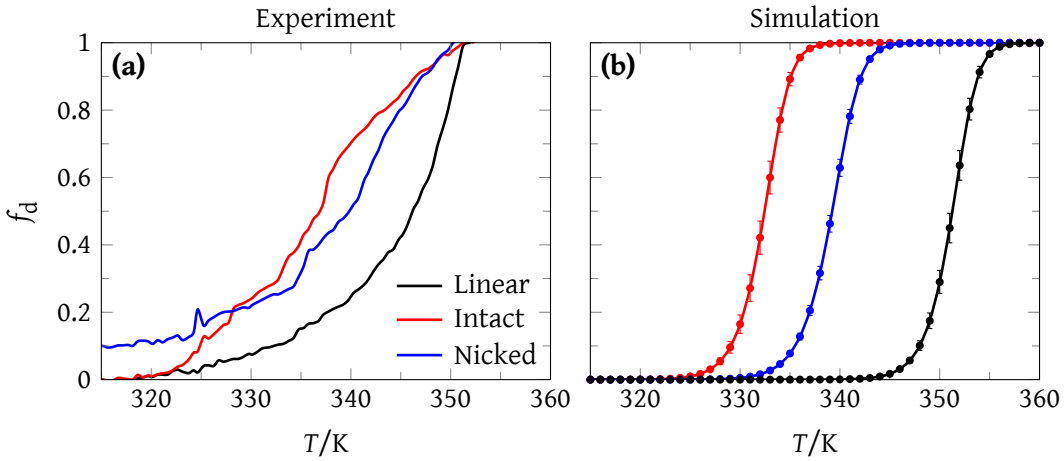


Figure 8.11 Melting profiles
Experiment versus simulation for the linear duplex (**black**), intact (**blue**) and nicked (**red**) model systems. The resulting melting temperatures (T_m) are given in Table 8.1.
(a) Experimental melting curves ($N_d = 31, N_s = 14$) by A_{260} at 100 nM [DNA] in 10 mM Tris buffer with 1 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl and 5 mM $MgCl_2$, and pH = 7.9.
(b) Simulation melting curves ($N_d = 30, N_s = 15$) with the sequence-dependent model. Statistics from windowed umbrella sampling runs at 340 K were histogram-reweighed to cover the entire melting transition region. Concentration correction (21.5 µM to the experimental concentration of 100 nM), and heteroduplex DNA finite size correction[115] were then applied. Errors bars are standard error of the mean (SEM) from 10 independent simulations.

is less costly than kinking duplex DNA. The underlying free-energy difference between kinking duplex versus nicked DNA accounts for the difference in behavior between the two systems.

8.5.2 Melting behavior

As discussed in Section 8.2.2, Qu *et al.* determined melting curves of linear, intact and nicked model systems.

As expected, due to the strain on the molecule, the intact system has a lower melting temperature (T_m) than the linear duplex (Figure 8.11 (a)). Despite a reduction in the free-energy cost of kinking, the nicked molecule has a lower T_m than the intact. At first glance this is counter-intuitive; however, the puzzle is resolved by considering the higher *entropic* cost of forming the nicked system, compared to the intact system, may more than offset the mostly *enthalpic* benefit of easier kinking. As we will show in Section 8.5.3, examining free energy as a function of base pairing reveals that this is indeed the case.

Simulation results are in semi-quantitative agreement with experiment, with a decrease in

System	Experiment		Simulation	
	T_m/K	ΔT_m/K	T_m/K	ΔT_m/K
Linear duplex	346		351 ± 2	
Intact	340	-6	339 ± 2	-12 ± 3
Nicked	337	-9	333 ± 2	-18 ± 3

Table 8.1 Melting temperature (T_m), experiment versus simulation
Melting temperature (T_m) for experimental and simulation at a concentration of 100 nM. ΔT_m is taken as the difference from the linear duplex. Full melting curves are depicted in Figure 8.11. Error bars represent SEM from 10 independent simulations.

T_m when comparing the linear duplex (highest T_m), intact and nicked (lowest T_m) systems (Table 8.1). Quantitatively, the linear duplex is somewhat over-estimated, while the nicked system is somewhat underestimated.

It is unclear whether the linear duplex molecule of Qu *et al.*[19] includes single-stranded overhangs, which the simulated system does not (*i.e.* blunt-ended, Figure 8.2 (a)). Further, their melting curve normalization involved manually identifying the transition between 'real' DNA dissociation and spurious signal from the hyperchromic absorbance of single-stranded DNA (a change in absorbance due to base unstacking). Melting curves were then offset to align these regions under the assumption that ssDNA hyperchromic absorbance is the same for the duplex and strained molecules, and that the manually selected transition points were accurate. Less importantly, note that experimental systems differ slightly, $N_s = 14, N_d = 31$, versus $N_s = 15, N_d = 30$. Given these experimental caveats, the lack of error bars in their experimental results and potential errors due to extrapolation in the simulation (Appendix A.5.4), we caution against overinterpretation.

8.5.3 Free-energy profiles, base paring

Previously (Section 8.5.2), we have discussed shifts in the T_m of melting profiles. The origins of these differences in T_m become apparent when we examine free-energy profiles for the formation of these structures using the number of base pairs Q_{bp} as our order parameter (Figure 8.12). The initial entropy-dominated free-energy barrier between the bound and unbound molecules is, as expected, similar for all molecules.

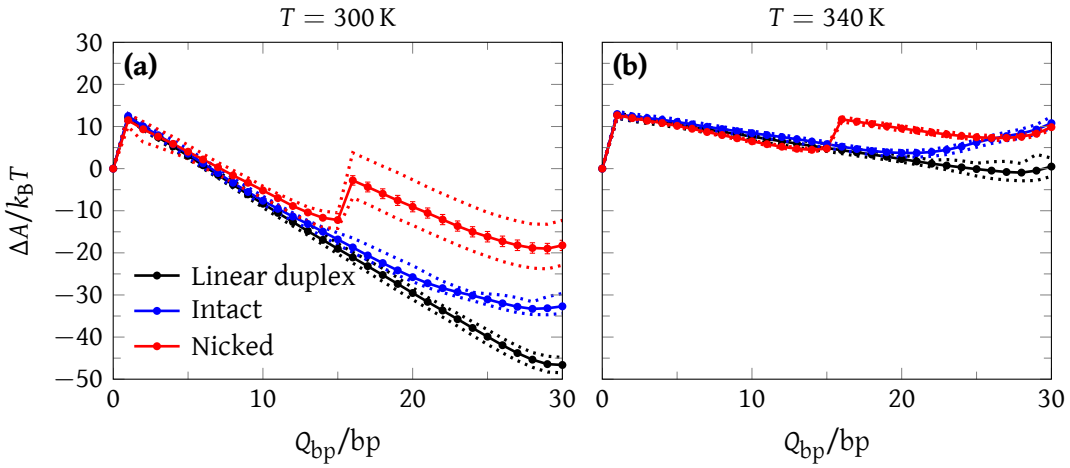


Figure 8.12 Free-energy profile, base pairing
Free-energy profile versus number of formed base-pairs for linear duplex (**black**), intact (**blue**)and nicked (**red**) DNA ($N_d = 30, N_s = 15$) with the average-base parameterization at **(a)** 300 K and **(b)** 340 K. The simulation contains one molecule in a cubic box of dimension 42.6 nm, an effective concentration of 21.5 µM. Finite-size corrections[115] have been applied. Errors bars represent SEM from 10 independent simulations. Dotted lines are minimum and maximum values over all simulations.

Firstly, we compare the duplex and intact systems as both involve the formation of single double-helical section (Figure 8.1). Initially, the downward slope of the free-energy as a function of the number of base pairs Q_{bp} is approximately the same for the linear duplex and intact strained duplex; however, as Q_{bp} increases, these is an increasing deviation of the slope. This deviation is due to the increasing build-up of bending stress in the intact system, which for $N_d = 30$ and $N_s = 15$ is stored as continuous bending (Section 8.4.1).

In contrast, the nicked system involves the formation of two double-helical segments (Figure 8.1 (b)), apparent from the free-energy profile as the local maximum in free energy at $Q_{bp} = 16$. In particular, it is favorable to first form one double-helical section, before forming the second, as is evident by the increase in free energy between $Q_{bp} = 15$ and $Q_{bp} = 16$ (Figure 8.12 (a)). This barrier is associated with the loss of conformational entropy upon the binding and subsequent hybridization of the second duplex section of the nicked molecule.

It is this extra entropic cost that causes the T_m of the nicked system to be lower than that of the intact system (Section 8.5.2), even though the free-energy cost of bending is lower for nicked duplexes than intact duplexes (Section 4.3.4.3). While sequence does impact the free-energy profiles, the entropy-dominated local maximum for the nicked system does not appear to be sequence dependent, with both the sequence-dependent and average-base models reporting

quantitatively comparable results (data not shown).

Supporting this conclusion, we observe from the free-energy profile of the intact system at $T = 340$ K (Figure 8.12 (b)), which corresponds to the melting point of the intact system, that the degree of fraying is significantly enhanced. This enhancement is indicated by the free-energy minimum of the intact system with 20 of 30 base pairs formed ($Q_{\mathrm{bp}} \approx 20$). Clearly, the magnitude of the strain-induced deviations in T_{m} will vary with N_{s} and N_{d}.

8.6 Critique of Qu *et al.*

Thus far, we have made comparisons only to the experimental observations of Qu *et al.*,[19,78] purposefully avoiding comparisons to their analytical model. Although we do not directly compare to the monomer-dimer equilibrium experiments (Section 8.2.1),[xii] the 'critical' value for kinking, in terms of the length of the single-stranded tether N_{s}, inferred by Qu *et al.* from their experiments are significantly longer than for oxDNA (Section 8.4.5). For example, Qu *et al.* report $N_{\mathrm{s}} \approx 36$ for the $N_{\mathrm{d}} = 24$ intact system,[19] compared to $N_{\mathrm{s}} \approx 6$ for oxDNA (Figure 8.3).

Similarly, although we have semi-quantitative agreement with the experimental T_{m} between the different systems, Qu *et al.* infer from their experiments that both the nicked and intact systems are kinked at these values of N_{s} and N_{d}, whereas oxDNA predicts only the nicked system to be kinked. We find that these differences stem not from an inability of oxDNA to capture the extreme bending of DNA – the comparisons with experiment in the previous chapters (Chapter 5, Chapter 7, Chapter 6) give us confidence in the oxDNA model in this regard – but rather certain assumptions in the Qu *et al.* analytical model.

8.6.1 Critique of monomer-dimer equilibrium analytical expression

8.6.1.1 Experimental context

We wish to highlight a few of intriguing features of the Qu *et al.* dimerization experiments. Firstly, in the regions of the 'critical' buckling instability, the monomer band from their DNA electrophoresis gels was split between two peaks. These peaks were assigned to the kinked and

[xii] Such calculations, though potentially feasible, would be very computationally expensive and are beyond the scope of this thesis.

unkinked states; however, the time scales for interconverting between these two states would be expected to be very rapid compared to the time scales for gel migration; therefore, we would expect a single peak for the monomer band. The substates of the monomer peaks are unclear.[xiii]

Secondly, the gels showed the presence of not only monomer and dimer bands, but also trimers and other higher-order oligomers. Although the removal of the bending stress provides a clear driving force for dimerization, it is less clear what could drive the formation of trimers given that the process would involve a reduction in translational entropy.[xiv]

To explore the Qu *et al.* analytical model, we will examine the duplex (Equation (8.9)) and single-stranded (Equation (8.10)) components in-turn.

8.6.1.2 Comparison to E_d

Firstly, we consider the Qu *et al.* approach to extracting the change in elastic energy from the equilibrium constants. Although the expressions in Section 8.2.1 capture the concentration dependence of the free-energy difference, there are other contributions that are not considered, namely changes in translational and rotational entropy upon dimerization. In particular, Qu *et al.* assume that only changes in the elastic energy contribute to $\Delta G^{\mathrm{ref}}_{\mathrm{M}\to\mathrm{D}}$ (Equation (8.6)), the free-energy difference between monomer and dimer. Further, correctly capturing the internal entropy of the dimer, which comes from the relative position and orientation of the two double-helices and the connecting single strands, is likely to be more complicated than the correction they apply, which assumes a particular parallel and aligned geometry for the dimer (Figure 8.2 (d)).

Secondly, we consider the expressions for elastic free energy (Equation (8.7), Equation (8.9), Equation (8.10)). Qu *et al.* neglect fraying in their analysis, which our oxDNA results suggest is an unreasonable approximation for the nicked system. Neglecting fraying is even more problematic for the intact system, particularly at shorter N_d. For a specific example, note the dominance of fraying over kinking for $N_\mathrm{d} = 18$ at small N_s (Figure 8.5 (d))).

[xiii] One interesting possibility is that the fluorescent intercalating agent used to image the gel kinetically traps the states, effectively decreasing the interconversion rate, and allowing the two substates to be resolved. Wang *et al.*[116] have partially addressed the role of ethidium bromide (EtBr), but have not explicitly examined interconversion.

[xiv] Taking a concentration profile as an experimental control is a straight-forward assay that would help clarify this issue.

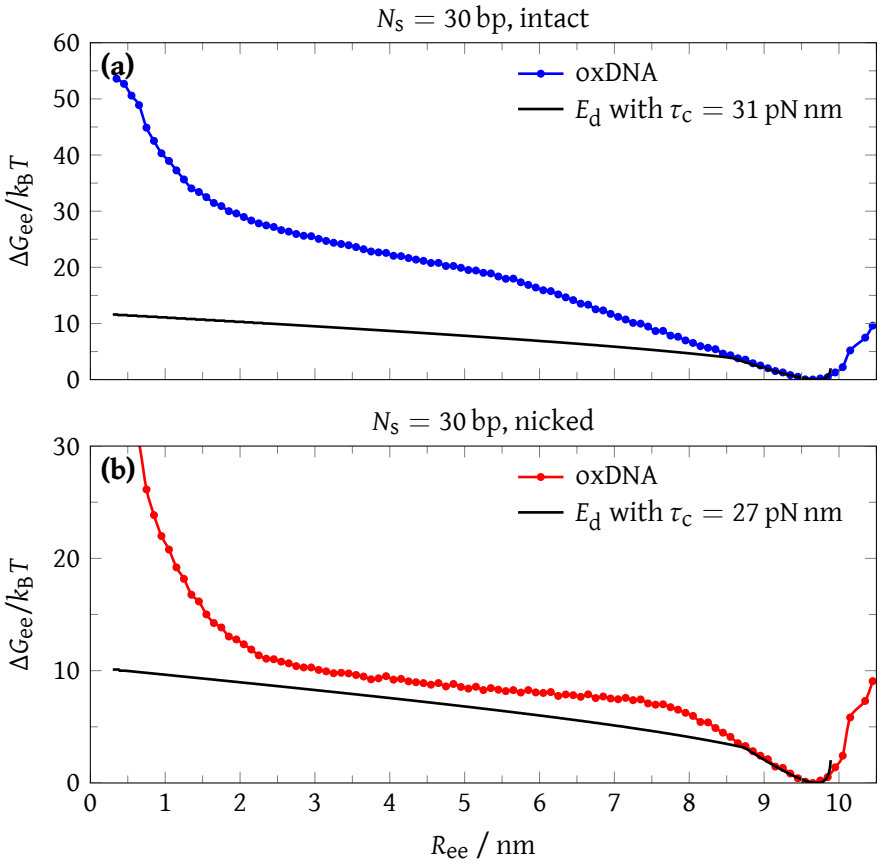


Figure 8.13 Double-stranded DNA bending, oxDNA versus E_d
Comparison of the free energy of bending **(a)** intact **(b)** nicked double-stranded DNA for oxDNA versus the E_d expression of Qu *et al.* (Equation (8.9)). We compare with parameter values: $N_d = 30$, $B_s = 50\,k_B T$ nm and $T = 298$ K.[xvii] The 'critical' torques are from Qu *et al.*,[19] and are $\tau_c^{\text{intact}} = 31$ pN nm and $\tau_c^{\text{nicked}} = 27$ pN nm for the intact and nicked systems, respectively.

Probing further, we directly compare the Qu *et al.* form for duplex bending (Equation (8.9)) with the oxDNA free energy-profiles computed in Chapter 4. The comparisons are shown in Figure 8.13. The approximate WLC-form for the unkinked state (continuously bent) behaves similarly to oxDNA; however, there are considerable differences in the free energies of the kinked state. The most significant of these is the position of the kink transition,[xv] the slopes of the free energy once kinked, and the lack of a rise at small $R_{ee}/R_{contour}$. The third is less important, as it will only play a role at very small N_s, and is due to the neglect of excluded volume effects which arise due to the finite thickness of DNA[xvi], a reasonable simplification that has been successfully used elsewhere.[8,11]

[xv] Or, if you prefer the Qu *et al.* approach, the equivalent torques at the transition point. Though, this would be a torque corresponding to the mid-point of a thermodynamically smooth kink transition, not a 'critical' torque τ_c.

[xvi] At these short separations, it also makes a difference as to whether the EED is measures in terms of the center of the end base pair, or in terms of the backbone position.

[xvii] Qu *et al.* use $T = 295$ K instead of $T = 298$ K. The difference is slight, and we opt to compare at the same

The first is the most important, and is a reflection of the previously noted lower values of N_S required to kink the duplex in oxDNA, than inferred by Qu *et al.* (Section 8.4.5). Alternatively, the bending free energy at the kink transition ΔG_{trans}, is 3.9 $k_B T$ and 3.2 $k_B T$ for the intact and nicked systems of Qu *et al.* resectively, compared to 18.1 $k_B T$ and 6.8 $k_B T$ for oxDNA (Chapter 4, specifically Table 4.1).

In the language of 'critical' torques, the oxDNA values would be significantly larger than those reported by Qu *et al.* The differences in slope are partially related to the one-parameter nature of the expression for the free energy of the kinked duplex (Equation (8.9)). It would probably be much more natural to have a two-parameter expression, where the parameters would govern the 'raw' free-energy cost associated with forming a structural defect (kink or otherwise), and a force constant describing the cost of changing the bending angle at the defect. In contrast, τ_C in Equation (8.9) controls both.

8.6.1.3 *Comparison to single-stranded energy* (E_S)

We can also compare oxDNA to the Qu *et al.* analytical expression for ssDNA. By differentiating Equation (8.10), one can obtain their prediction of the force-extension curve. This is compared to results previously obtained for oxDNA (Figure 8.14).[56] It is clear that Equation (8.10) predicts ssDNA to be too stiff, but why? Although it can be argued whether the WLC approximation on which Equation (8.10) is based is the best model for ssDNA, more importantly, some of the parameters used are inappropriate. In particular, Equation (8.10) assumes that the chain consists of N_S persistence lengths where the ssDNA persistence length is $l_S = 0.764$ nm. Although the appropriate value of the contour length is somewhat subtle because of stacking interactions, the persistence length is clearly too long. With the Qu *et al.* persistence length and a $N_S = 100$ nt segment, a contour length of 25.5 nm would be predicted. For comparison, the contour length would be ∼65 nm for a fully unstacked chain, while a fully stacked helical configuration would have a contour length of ∼34 nm.

temperature for clarity. The Qu *et al.* bending modulus ($B_S = 50\ k_B T$ nm) is stiff than that of oxDNA ($B_S = 41.82\ k_B T$ nm). The difference small, but slightly more pronounced than temperature. As it does not impact result of the comparison, we omit the comparison for clarity.

[xviii] We find F_S by taking the taking the derivative of Equation (8.10) with respect to EED x: $F_S = -dE_S/dx$. As the sign simply indicates the directionality of the force, we report the force as positive for convenience.

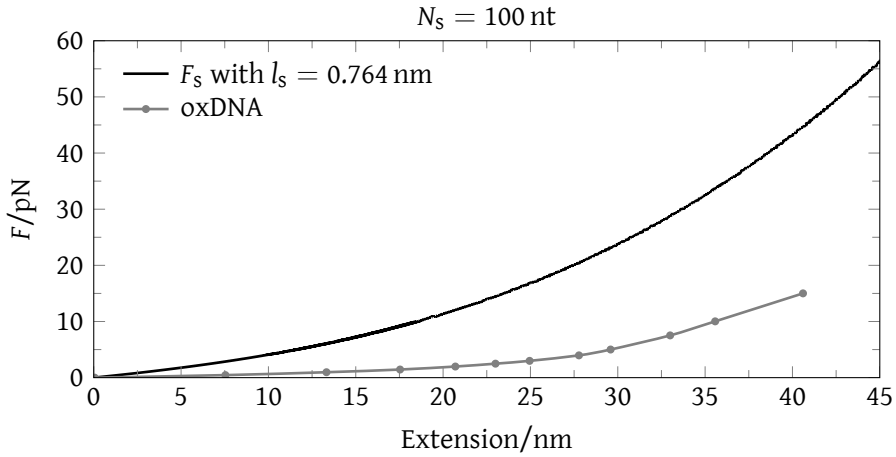


Figure 8.14 Single-stranded force-extension, oxDNA versus E_S
Comparison of the single-stranded for extension for oxDNA (excerpted from Figure 7 of 56) versus the F_S,[xviii] the force expression derived from E_S (Equation (8.10)) for $N_S = 100$ nt, $T = 23\,^{\circ}$C and $l_S = 0.764$ nm.

8.6.1.4 *Single-stranded length N_S and end-to-end distance $\langle$EED$\rangle$*

The combined effects of ssDNA that is too stiff, and nicked duplexes that kink more easily leads to analytical expressions that report kinking at much larger N_S than oxDNA. The expressions of Qu *et al.* also assume that there is an abrupt transition between the kinked and continuously bent states. This leads to a jump in the predicted end-to-end separation (Figure 8.15). In oxDNA, unsurprisingly, the transition occurs over a range of N_S lengths.[xix]

For the nicked system, the oxDNA end-to-end separation has a sigmoid form because the relative flexibility of the kink at the nick leads to a significant decrease in the end-to-end separation upon kinking. By contrast, there is a much higher free-energy cost for changing the bending angle for the intact system, and so the kink transition in the intact system is barely evident from Figure 8.15.

[xix] An observant reader will note that sequence variation or secondary structure could contribute to ssDNA EED, possibly explaining some of the variation between Qu *et al.* and oxDNA. We doubt that this is the case. The Qu *et al.* single-stranded tether has negligible secondary structure in the single-stranded tether (Section 8.4.4.1). Further, given that the average-base and sequence-dependent parameterization yield nearly identical results in the absence of secondary structure (Section 8.4.4.2), strongly suggests that sequence-variation should have minimal impact on $\langle$EED$\rangle$. These results suggest that neither secondary structure, nor sequence-variation account for the discrepancy between the Qu *et al.* and oxDNA.

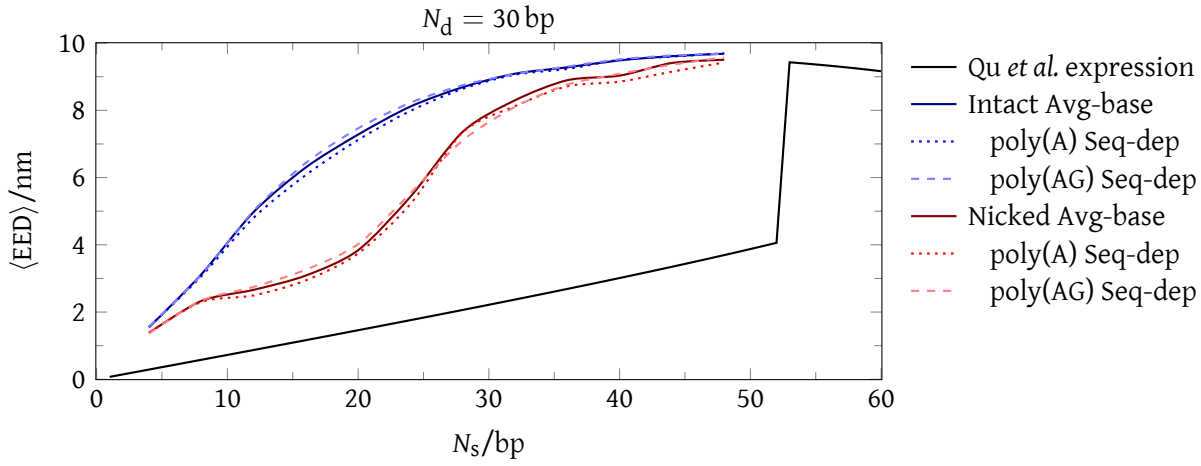


Figure 8.15 Single-stranded length N_s versus end-to-end distance ⟨EED⟩
Relationship between ⟨EED⟩ and N_s for oxDNA and Qu *et al.* Simulation results for intact (**blue**) and nicked (**red**) $N_d = 30$ using the duplex sequence from Qu *et al.* .[19] Shading and texture represent N_s sequence and model: average-base poly(A) (**dark, solid**), sequence dependent poly(A) (**dotted**), sequence dependent poly(AG) (**light, dashed**). Numerical solution of $x_{eq}(N_s)$ for the analytical model of Qu *et al.*[20,78] from Equation (8.8) (**black**), solved using Wolfram Mathematica 7. This yields a relationship between x_{eq} and N_s. Parameters appropriate for the nicked system were used: $N_d = 30$ ($L = 5$ nm), $\tau_c^{nicked} = 27$ pN nm, $B_s = 50$ nm $\times k_B T$ and $l_s = 0.764$ nm. The intact system has a slightly higher 'critical' torque of $\tau_c^{intact} = 31$ pN nm; this does not significantly alter the result of the expression, so we omit the intact system for clarity.

8.6.2 Critique of melting curve analysis

For the melting experiments (Section 8.2.2), our simulations (Section 8.5.2) have revealed a discrepancy between oxDNA and the Qu *et al.* interpretation. For the formation of the nicked system, there is an additional entropy cost compared to the intact strained duplex, because the resulting single-stranded region has a loop that was not present in the starting strands (Figure 8.1). This is the reason for the lower melting temperature of the nicked system. However, because Qu *et al.* do not consider this entropic contribution, they predict a significantly higher elastic free energy for the fully base-paired nicked system ($8.62\, k_B T$) than for the intact system ($4.82\, k_B T$) for $N_d = 30$ and $N_s = 15$ at their melting points. Such relative values are in clear disagreement with both oxDNA, and other experiments that indicate that it is easier to kink at a nick. The relative free energies of the intact and nicked systems are reversed only when they extrapolate to room temperature. To cause the reversal, Qu *et al.* assume that A_{el} is purely enthalpic for the nicked system,[xx] but has a large entropic component for the intact system. Further, the entropic component of the intact system's free energy is arbitrarily chosen, being

[xx] Later dimerization experiments suggest that there is a mild temperature dependence to A_{el}.[117]

equivalent to the breaking of 3 terminal base pairs. Breaking base pairs within a duplex (*i.e.* a bubble) should have a significantly lower entropic gain than breaking terminal base pairs (*i.e.* fraying). This implies a bubble size of > 3 bases pairs which, unsurprisingly, is inconsistent with oxDNA prediction (Section 4.4.3).

Curiously, in a later publication,[117] they argue for a similarly mild temperature dependence of the intact system to that of the nicked system, for which they 'measure' the temperature dependence of A_{el}. They also argue that bubble formation at duplex kinks is unlikely. One of Qu *et al.*'s supporting rationales was that this leads to the ''paradoxical prediction'' that the elastic energy of the intact system would then be lower than that for the nicked system near the melting temperature. Ironically, this is exactly what their interpretation of melting curve data implies. The realization that their interpretation implies that kinking at the nick is *more* difficult than kinking in the duplex, which is inconsistent with the corpus of experimental and theoretical study we have enumerated (Chapter 4, Chapter 5, Chapter 6, Chapter 7), is a subtle but important caveat.

8.6.3 Summary

We have purposefully disentangled the underlying experiments from the analytical model of Qu *et al.* While there are a few curious inconsistencies in their underlying experimental approaches (Section 8.2.1.1, Section 8.2.2.1), the crux of the discrepancy between the oxDNA and Qu *et al.* lie with their analytical model, specifically extracting of the elastic energy from their experimental data (Section 8.2.1.2, Section 8.2.2.2).

The disagreement between oxDNA and the results of the Qu *et al.* analytical model serve to highlight several important issues, including:

(i) The presumption of a 'critical' (discontinuous) buckling instability, as opposed to a thermodynamically smooth transition, between kinked and unkinked states.
(ii) Similar difficulty in kinking intact and nicked systems, as opposed to nicked systems being significantly easier to kink.
(iii) The neglect of entropic effects that cause the lower melting temperature of nicked systems.

8.7 Conclusion

We use oxDNA to probe the structure and thermodynamics of the 'strained duplex' model system of Qu *et al.* . We find that specific structural defects are responsible for relaxing bending stress in the duplex: continuous bending for the long duplexes ($N_d = 60$–65), kinking for intermediate length duplexes ($N_d = 25$–55), and fraying for the shortest duplexes ($N_d < 25$). Based on the work in previous chapters, in particular the investigation of the 'molecular vise' (Chapter 5), the Qu *et al.* experiments are consistent with our expectations.

We find that the N_s at which kinking occurs is a sensitive measure of bending in the duplex, liable to changes due to sequence, temperature and structure. In the case of structure, we again see clear evidence that the free-energy cost of kinking at a nick is lower than kinking in the duplex. This is indicated by kinking at longer N_s (less bending stress) for the nicked, as opposed to the intact molecules.

While we do not agree with results of the Qu *et al.* analytical model (Section 8.6), we are in semi-quantitative agreement with their underlying experimental data;[xxi] for example, finding agreement between the oxDNA and Qu *et al.* melting profiles. Contrary to the analytical model of Qu *et al.*, we find thermodynamically smooth transitions between continuous bending and kinking. Further, we find that kinking at a nick is much easier than kinking in the duplex.

[xxi] Unlike the other chapters, some of the data contained herein are development runs instead of publication-quality production runs (Appendix A.5). This means that the reported errors, albeit small, are slightly larger than in the other chapters. The observant reader may note that sampling may be insufficient in development-quality data; however, we have explicitly checked sampling, and find that this is not the case (Appendix A.5.2). However, the trends are clear, and in practice, given the enormous qualitative discrepancies between oxDNA and Qu *et al.*, the use of some development data is a moot point.

9 Conclusion

In Part I, we use oxDNA, a coarse-grained simulation of DNA, to characterize the basic thermodynamics of DNA bending. We apply these insights to 4 distinct experimental systems. Our inquiry directly addresses the difficulties inherent in interpreting experimental data, and the art of applying simulation to untangle generic versus system-specific phenomena.

Specifically, we find that a small set of structural defects, namely continuous bending, kinking and fraying, explain a very wide range of observations in the 4 experimental systems investigated. While the interplay between these defects is governed by the peculiarities of each system, simulation allows us to resolve the essential physical principles underlying apparently contradictory experimental observations.

Continuous bending dominates at lengths above about 80 base pairs and is in strong agreement with prevailing models of DNA flexibility, namely the worm-like chain (WLC) model. Where structural defects are absent, DNA behavior is well described by the worm-like chain (WLC) model, highlighting the enduring utility of WLC models in describing the structure and thermodynamics of DNA bending. We find no evidence for extreme bending (*i.e.* non-WLC behavior) at lengths longer than 80 base pairs.

At intermediate lengths, between 40 and 80 base pairs, it becomes favorable to localize bending to a kink, whereby base pairing and/or stacking are disrupted in a 1–3 base pair region, thereby allowing the remaining duplex DNA to adopt a relaxed canonical configuration. Exogenous factors such as salt concentration and temperature aside, sequence and local DNA defects are the most important determinants of the transition from continuously bent to kinked states. For example, we find that AT-rich regions kink more readily than their GC-rich counterparts, and that kinks selectively localize to pre-existing DNA defects, such as nicks and mismatches.

At the very shortest lengths, usually less than 40 base pairs, fraying of a few base pairs allows for sufficient relaxation to obviate the need for both kinking and continuous bending. Owing to the importance of local geometric and steric considerations, relaxation via fraying is more system-dependent than other structural aberrations.

The promise of the oxDNA model is its ability to simulate experimental systems directly, detangling generalizable and system-specific phenomena. Our unique contribution is in harmonizing the interpretation of disparate experimental observations, drawing meaningful comparisons between theory and experiment with direct simulation of experimental systems.

Part II Single-molecule studies of the RecQ DNA helicase

Abstract

DNA helicases are ATP-dependent translocates essential to genome integrity through their role in replication, recombination and repair. As such, they are highly conserved in prokaryotes and eukaryotes, from *E. coli* to man. In humans, heritable defects in helicases manifest clinically as premature aging and a greatly elevated cancer risk, in disorders such as the Werner and Bloom syndromes. These clinical manifestations follow from aberrant recombination and deficient DNA damage repair. Despite their important biological role, the mechanisms by which helicases translocate on and unwind their substrates remains poorly understood. Although several models have been proposed to describe the mechanics of helicases based on biochemical and structural data, ensemble experiments have been unable to address some of the more nuanced questions of helicase function.

We have elucidated a potential mechanism of RecQ, a prototypical superfamily II DNA helicase, by developing a model that integrates the catalytic cycle with specific conformational and binding states. Specifically, via novel fluorescently labeled mutants, we have observed ATP-binding cleft movements upon binding and hydrolysis. Additionally, we have used single-molecule fluorescence (Förster resonance energy transfer (FRET)) to recapitulate ensemble binding dynamics and observe putative RecQ unwinding at the single-molecule level.

10 Introduction

10.1 Impetus

Genome maintenance through DNA replication, recombination and repair is indispensable to genomic stability. Mutagenic lesions such as thymine-dimers, AP-sites (apurinic or apyrimidinic), spontaneous deamination of adenine/cytosine, aberrant methylations/alkylations, ring openings, cross-linking, single strand breaks, double strand breaks and mismatches disrupt genomic stability. In the absence of adequate proof-reading and repair mechanisms, these lesions impair DNA replication, distort protein function and ultimately lead to cell death or aberrant cell behavior (reviewed in reference 118).

Both prokaryotic and eukaryotic cells employ a variety of well conserved pathways to repair or circumvent mutagenic DNA lesions. High-fidelity repair mechanisms include nucleotide excision repair (NER), base excision repair (BER), methyl-directed mismatch repair (MMR) and recombination repair. There are other repair mechanisms, such as non-homologous end-joining (NHEJ) for double-strand breaks and SOS repair, but they are themselves mutagenic.

Helicases are essential to the proper functioning of the cell through their integral roles in replication, transcription, repair and recombination. Most helicases can be thought of as modular nucleoside triphosphate (NTP)-dependent translocases, which hydrolyse NTP to drive movement (translocation) and activity (primarily unwinding the duplex, but also other functions) along the DNA. Additionally, most helicases possess additional domains supporting the core helicase activity, with each module/domain conferring properties such as directionality, processivity, substrate specificity and unwinding activity.[119]

While traditional ensemble assays have garnered much biochemical information about the

relationship between helicase structure and function, many basic biophysical questions remain unanswered. For example, do superfamily (SF) 1 and SF2 helicases translocate single-stranded DNA (ssDNA) by one of the proposed mechanisms[120–122] (Section 10.3.1), or by some as of yet unknown process? How is translocation coupled to unwinding? How exactly do accessory protein interactions impact translocation and unwinding?

As opposed to ensemble assays, single-molecule techniques allow for the direct detection of intermediate and low-occupancy states within a population. While the interpretation of ensemble assays often relies upon kinetic models, the underlying signal from single-molecule data, intensity versus time for most fluorescence-based techniques, is model-free. By sampling individual events instead of ensembles, single-molecule methods can provide insights inaccessible to their ensemble counterparts.[123]

We will use the *E. coli* RecQ helicase, the prototypical helicase of the RecQ helicase family, as a model system with which to probe SF2 helicase function. The RecQ helicases are of particular biological and clinical interest because of their significance in genome maintenance pathways, particularly DNA damage repair. First, as a foundation from which to better elucidate its mechanical and enzymatic properties, we have designed, produced and extensively tested previously unreported RecQ substrates. Secondly, we apply ensemble and single-molecule fluorescence techniques to examine the binding, translocation and unwinding of the *E. coli* RecQ helicase.

The investigation of helicases with single-molecule fluorescence is not without precedent.[124,125] For example, in two notable single-molecule fluorescence studies of *E. coli* Rep helicase (SF1A), the Ha group was able to uncover behavior and domain movements invisible to ensemble methods. In one,[124] they used doubly labeled 3′-duplex DNA substrates to monitor burst, stalls and rewinding with monomeric Rep. They also demonstrated that monomeric Rep could translocate and position itself at the ss/double-stranded DNA (dsDNA) junction, but not unwind the duplex substrate without oligomerization. In the other,[125] they used singly labeled Rep to probe the nature of the 'open' (apo) and 'closed' (DNA bound) states. Both of these discoveries would be difficult, if not impossible, to observe in the ensemble.

10.2 Clinical relevance of human RecQ helicases

The RecQ helicases are genome maintenance enzymes implicated in DNA recombination and repair.[126,127] They are conserved in prokaryotes and eukaryotes, from *E. coli* and yeast to man. In humans, it is clinically relevant in several rare hereditary disorders. Three of the five human RecQ homologues are implicated in disease states: Bloom syndrome protein (BLM) in Bloom syndrome, RecQ protein-like 4 (RECQ4) in Rhothmund–Thomson syndrome and Werner syndrome RecQ helicase-like protein (WRN) in Werner syndrome. All three of these diseases are characterized by high frequencies of chromosomal damage, including double stranded breaks and rearrangements. Cellular phenotypes of these disease states include genomic instability, telomere dysfunction and premature senescence. Patient phenotypes vary, but invariably include a predisposition to developing cancer at an early age.[128]

While the clinical symptoms are well characterized, the molecular basis of genome maintenance by the RecQ helicases is not entirely understood.[129] It is known that human RecQ helicases interact with a wide variety of proteins implicated in DNA replication, repair and telomere regulation, discussed at length in Section 11.2.3.

10.3 DNA helicases

There are five core families of helicases: the SF1, SF2 and SF3 helicases, the DnaB-like family and the Rho-like family. SF1 and SF2 helicases are the most numerous and generally function as either monomers or dimers. SF3 helicases are generally found in DNA/RNA viruses. DnaB-like and Rho-like helicases are generally hexameric. All but the SF3 helicases are conserved across prokaryotes and eukaryotes, from *E. coli* to man.

SF1 and SF2 helicases contain conserved helicase motifs (I, II, III, IV, V, and VI), which are implicated in both ATP binding and hydrolysis, as well as translocation.[119] Structurally, the two superfamilies are somewhat similar, containing core RecA-like domains – where the conserved helicase motifs are located – along with flanking function-dependent domains. It has been suggested[119] that a key functional differentiator between SF1 and SF2 helicases is the ability of SF2 helicases to translocate both ssDNA and dsDNA, while SF1 helicases can only translocate on

ssDNA.

10.3.1 Helicase mechanisms

The translocation mechanism of monomeric and dimeric (usually SF1/2) helicases fall into three main categories: active rolling,[120] inchworm[121] and Brownian motor.[122] In the active rolling and inchworm models, two (or more) distinct DNA binding sites alternate their binding to the substrate. In the Brownian motor model, a single DNA binding site is responsible for translocation.

Under the active rolling model, the binding sites should be identical and able to bind either ssDNA or dsDNA substrates (Figure 10.1 (a)). The inchworm model differs in that the DNA binding sites are unique – one specializes in ssDNA, the other in dsDNA (Figure 10.1 (b)). An additional distinction is that the active rolling model disallows binding of both sites simultaneously, while the inchworm model usually necessitates binding of both sites during at least one step of the catalytic cycle. Also, one would expect large step sizes for an active rolling model – approximately the number of occluded nucleotides. This is not a requirement in the inchworm model. Although the DNA binding domains can be on the same or multiple proteins, the active rolling model is generally associated with dimers or higher order oligomers. The actual conformational change responsible for the inchworm or rolling movement is usually thought to correspond with either ATP-binding, phosphate release, or both.[130] The Brownian motor is similar to inchworm model, but relies on a single DNA binding site alternating between weakly and tightly bound states (Figure 10.1 (c)).

Note that the unwinding mechanism is not necessarily the same as the translocation mechanism. This is exemplified in a study by Soultanas *et al.*[131] in which lysine/threonine to alanine point mutations on the dsDNA binding site of *B. stearothermophilus* PcrA were able to impair helicase activity, while leaving ATPase activity and single-stranded translocation mostly unchanged. They conclude that the unwinding activity is impaired by a loss of destabilizing interactions with the dsDNA and the translocation likely functions via an inchworm, as opposed to active rolling, mechanism.

Strand-displacement, separation pin and separation loop unwinding models which are

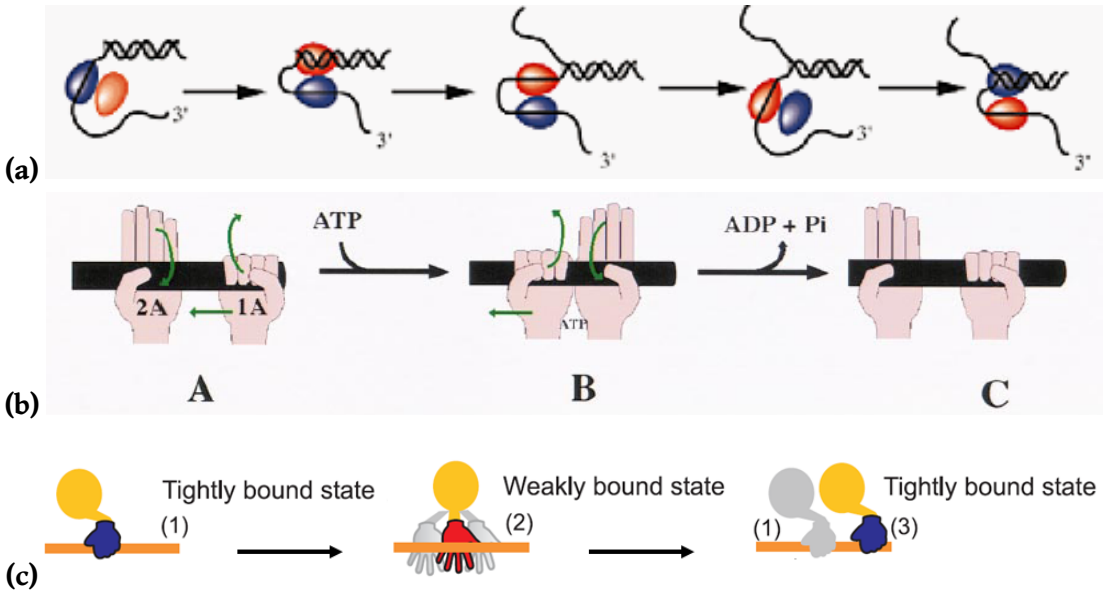


Figure 10.1 Models of helicase translocation
Three models of helicase translocation: the **(a)** active rolling model showing alternation between two identical DNA binding sites;[120] **(b)** the inchworm model showing alternation between two non-identical DNA binding sites;[121] and **(c)** the Brownian motor model showing translocation by alternating weakly and tightly bound states.[122] Figures from references 120–122.

independent of translocation mechanism have been proposed. In a strand-displacement mechanism, the mechanical force from translocation is enough to physically shear duplex DNA (or other obstructions) from the bound ssDNA strand. Separation pin and loop mechanisms involve a physical element actively destabilizing interactions at the ss/dsDNA junction. For example, Lee and Yang[132] have used a variety of crystal structures to suggest a translocation and unwinding mechanism for UvrD, supporting the view that Y621 acts as a separation pin while unwinding the ss/dsDNA junction (for further elaboration, see Section 11.2.2). Unfortunately, crystal structures are ensemble averaged static 'snapshots'; as such, while dynamics can be inferred, they cannot be observed directly.

While a plethora of kinetic parameters have been determined for RecQ through ensemble (stopped flow fluorescence anisotropy) and single-molecule experiments (processivity, step-size), the structural and mechanistic basis for RecQ translocation and unwinding have yet to be determined. Zhang *et al.* have suggested that RecQ translocates via an inchworm model.[133] We explore the RecQ helicase mechanism in Chapter 12.

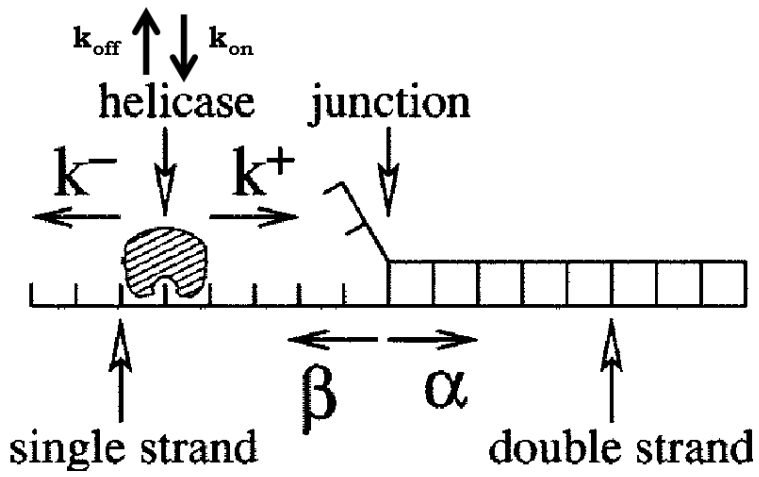


Figure 10.2 Kinetic parameters of active/passive helicases
Rate of strand-opening at the ss/dsDNA junction (α), rate of strand-closing at the ss/dsDNA junction (β), rate of forward progression (k^+), rate of backwards progression (k^-), ssDNA binding rate (k_{on}) and ssDNA dissociation rate (k_{off}).
Figure modified from modified from Betterton & Jülicher 2005.[134]

10.3.2 Active versus passive helicases

Betterton & Jülicher (B&J)[134] have developed a theoretical model for helicase translocation and unwinding, which supports the classification of helicases as either active or passive. Passive helicases unwind DNA by waiting for thermal fluctuations to destabilize nucleic acid interactions, and then translocating forward to prevent re-annealing. Active helicases destabilize nucleic acid interactions directly. A shortfall in the B&J framework is that both step-size and rate of slippage ($k_- \neq 0$) must be known in order to accurately determine the helicase interaction potential (Figure 10.2). Since these are difficult to measure experimentally, several helicases have conflicting classifications.

A complementary helicase activity model has been developed by Manosas *et al.*[135] This model uses the empirical ratio of unwinding velocity to ssDNA translocation velocity (V_{un}/V_{trans}) to classify active versus passive helicases. Since passive helicases require thermal denaturation at the ss/dsDNA junction to unwind, it would be expected that $V_{trans} > V_{un}$; whereas for an active helicase that destabilizes the ss/dsDNA junction directly, $V_{trans} \approx V_{un}$.

The unwinding velocity of GC versus AT duplex DNA ($V_{un}^{\%GC=100}/V_{un}^{\%GC=0}$) and the unwinding velocity with and without applied force ($V_{un}^{F=0}/V_{un}^{F=F_r}$) are two other key ratios thought to correspond to passive and active helicase activity. GC pairs are more thermally stable than AT

pairs; therefore, thermal fraying would occur less frequently on GC pairs than AT pairs. Thus, for a passive helicase, it would be expected that $V_{\mathrm{un}}^{\%\mathrm{GC}=0} > V_{\mathrm{un}}^{\%\mathrm{GC}=100}$. Likewise, applied force increases the dissociation of both GC and AT pairs; thus for passive helicases we would expect $V_{\mathrm{un}}^{F=F_{\mathrm{r}}} > V_{\mathrm{un}}^{F=0}$.

Fitting the B&J model with a step-size of one and zero slippage yields a cutoff between active and passive helicases of 0.25. That is when the three key ratios ($V_{\mathrm{un}}/V_{\mathrm{trans}}$, $V_{\mathrm{un}}^{\%\mathrm{GC}=100}/V_{\mathrm{un}}^{\%\mathrm{GC}=0}$, and $V_{\mathrm{un}}^{F=0}/V_{\mathrm{un}}^{F=F_{\mathrm{r}}}$) are greater than 0.25, the enzyme follows an active model. Both RecQ and UvrD are active helicases by the Manosas and Croquette criteria,[135] with $V_{\mathrm{un}}/V_{\mathrm{trans}} \approx 1$. While not assayed by Manosas *et al.*, there is strong mutational evidence suggesting PcrA functions via an active mechanism.[131] Hexameric helicases, such as *E. coli* DnaB, tend to be passive.

10.4 Overview of single-molecule fluorescence

In Chapter 12, we employ single-molecule assays to the RecQ DNA helicase; therefore, we provide a brief recap of fluorescence-based techniques, particularly the phenomena of Förster resonance energy transfer (FRET). We also review a conceptual understanding of single-molecule microscopy, specifically total internal reflection fluorescence microscopy (TIRFM). In single-molecule fluorescence studies the signal-to-noise ratio (SNR) – achieved through high quantum yield (QY), relatively photostable fluorophores and efficient detectors (usually highly sensitive back-thinned electron multiplying charge coupled device (EM-CCD) cameras) – is sufficient to resolve the intensity profile of a single fluorophore. This supreme sensitivity is what makes our single-molecule investigation of RecQ possible.

10.4.1 Fluorescence techniques

10.4.1.1 Förster resonance energy transfer (FRET)

Förster resonance energy transfer (FRET) is a method of reporting sub-diffraction limit interactions on the order of 1–20 nm. It is based on the non-radiative energy transfer from an excited donor fluorophore to an acceptor fluorophore in close proximity (∼3–10 nm). The FRET efficiency (E) is determined by the relative brightness of the donor and acceptor fluorophores,

in other words, what fraction of energy is transferred from donor to acceptor. The efficiency of the energy transfer is related to the sixth-power of the distance between the two fluorophores. R_0, which generally ranges from 2 nm to 8 nm depending on the fluorophores involved and their local environment, is the distance at which the FRET efficiency $E = 50\,\%$, given by,

$$E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \approx \frac{I_A}{I_A + I_D} = \frac{I_A - \beta I_D}{(I_A - \beta I_D) + \gamma I_D}, \tag{10.1}$$

where R is the fluorophore separation, R_0 is the Förster distance, I_D and I_A are the donor and acceptor fluorophore intensities, and β and γ are correction factors.

In single-molecule studies we detect the donor (I_D) and acceptor (I_A) intensity directly. To accurately report E, these intensities must be corrected for leakage of donor light into the acceptor channel (β) and differences in the quantum yield and detection efficiency (γ),[136]

$$\gamma = \left(\frac{\eta_A}{\eta_D}\right) \times \left(\frac{\varphi_A}{\varphi_D}\right) \approx \frac{\left(I_A^{Pre} - I_A^{Post}\right)}{\left(I_D^{Post} - I_D^{Pre}\right)} = \gamma^{photobleach}, \tag{10.2}$$

where η is the detector efficiency, φ is the QY, Pre refers to the average intensity before photobleaching and Post refers to the average intensity after acceptor photobleaching, and

$$R_0^6 = \frac{9\varphi_D \ln 10 \kappa^2 J}{128\pi^5 n^4 N_A}, \tag{10.3}$$

where φ_D is donor QY, κ^2 is a measure of dipole orientation,[i] n is the refractive index ($\sim$1.33 in H_2O), N_A is Avogadro's number and J is a measure of the fluorescence spectral overlap of the donor and acceptor fluorophores. J is given as

$$J = \int f_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda, \tag{10.4}$$

where $f_D(\lambda)$ is the normalized donor emission spectrum and $\varepsilon_A(\lambda)$ is the molar extinction coefficient of the acceptor in medium.[ii] Note that this is the generally the same $\varepsilon_A(\lambda_{max})$

[i] κ^2 has a range of 0–4, with $\kappa^2 = 2/3$ under isotropic conditions with random orientations of the donor and acceptor. A review and derivation of this 'κ^2 assumption' is available in reference 137. For comparison, note that $\kappa^2 = 4$ corresponds to anti-parallel, and $\kappa^2 = 0$ to perpendicular, orientations of the donor and acceptor dipole moments.[119,137]

[ii] For reference, we have provided a compendium of relevant fluorophore structures (Figure B.5) and properties (Table B.4), including molar extinction coefficient.

commonly reported for determining approximate dye concentrations.

10.4.1.2 Alternating Laser EXcitation (ALEX)

Alternating laser excitation (ALEX) expands upon traditional continuous donor excitation FRET by allowing the direct detection of both the donor and acceptor fluorophores by alternating between two excitation sources. Let us consider a surface immobilized system where FRET occurs between donor and acceptor fluorophores, that is with donor excitation (D_{ex}) we observe fluorescence in both the donor emission (D_{em}) and acceptor emission (A_{em}) channels. A reaction occurs that displaces one or both fluorophores. After the reaction, there is no fluorescence in the D_{em} or A_{em} channels with D_{ex}. Which fluorophores were displaced? In a traditional single-molecule FRET experiment, it is not possible to distinguish between the following cases:

- Neither fluorophore is displaced. The donor fluorophores photobleaches.
- The acceptor fluorophore is displaced. The donor fluorophores photobleaches.
- Both fluorophores are displaced or photobleach.

By alternating A_{ex} and D_{ex}, ALEX solves the problem, albeit at the expense of lower time-resolution. In addition to $D_{ex}D_{em}$ and $D_{ex}A_{em}$, we now have two additional channels of information, $A_{ex}D_{em}$ and $A_{ex}A_{em}$. Ignoring odd anti-Stokes effects or faulty filter sets, no fluorescence should be observed in the $A_{ex}D_{em}$ channel. $A_{ex}A_{em}$ allows us to distinguish the cases:

- Neither fluorophore is displaced versus donor fluorophore photobleaching (signal in $A_{ex}A_{em}$).
- The acceptor fluorophore is displaced. The donor fluorophores photobleaches. (Observe more molecules. Assuming that donor photobleaching and acceptor displacement are independent, it is unlikely that these events will always occur simultaneously.)
- Both fluorophores are displaced versus photobleaching (no signal in $A_{ex}A_{em}$).

This is the power of ALEX-FRET. Now, in addition to the continuous wave FRET equation for E (Equation (10.1)), we have an additional stoichiometry term S,

$$S \approx \frac{I_{DA} + I_{DD}}{I_{DA} + I_{DD} + I_{AA}}, \tag{10.5}$$

using the convention I_{xy}, where x is excitation and y is emission. For example, I_{DA} means donor excitation (D) and acceptor emission (A). Stoichiometry is useful in separating populations of molecules. Excitation intensity is adjusted such that $S = 1$ for donor-only species, $S = 0.3$–0.8 for donor-acceptor species and $S = 0$–0.2 for acceptor-only species.[138] ALEX has been implemented on both TIRFM and confocal setups.

10.4.1.3 Caveats in the interpretation of FRET data

The studies outlined in this thesis use synthetic dye pairs Cy3 / Cy5 ($R_0 \approx 5$ nm), Cy3B / Atto647N ($R_0 \approx 5.5$ nm) and Cy3 / Alexa647 ($R_0 \approx 6$ nm). FRET using synthetic dies and transition metal quenchers, for example a Ni^{2+} coordinated by histidines at the n and $n + 3$ positions of an α-helix, is also possible. This approach is useful for reporting inter-domain motions as dye-metal pairs generally have a short R_0 ($R_0 \approx 1$–2 nm[139,140]). Single-molecule FRET is not without its nuances and subtle caveats,[141] which we carefully address where appropriate. Key among these is that inter-molecular distance is non-trivial to extract from FRET data: R_0, on which distance calculations rely, is sensitive to a slew of environmental variables; synthetic dyes have a host of photophysical issues, most prominently photobleaching; surface effects may also pose problems in a total internal reflection (TIR) setup. Mechanical movements suggested by *relative* changes in FRET are often more biologically useful than an absolute distance. Given the difficulty in calculating R_0 from experiment, Förster distances for common dye pairs are widely reported in manuals and handbooks.[142,143] Given that φ_D, J, and most troubling the κ^2 assumption, may not hold for fluorophores conjugated to biomolecules,[119,137] absolute distances calculated with a handbook R_0 should be avoided. In our studies, we believe that relative changes in FRET are sufficient to address the biologically relevant questions at issue, and do not attempt to compute absolute distances from our data.

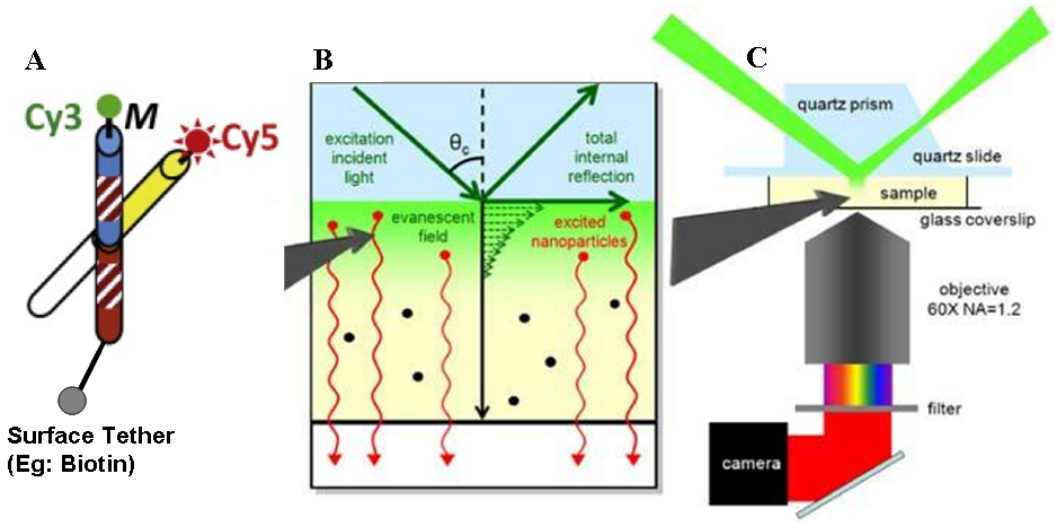


Figure 10.3 smTIR-FRET schematic
(a) Surface attached model system with two protein-linked fluorophores interacting via FRET. Bright, relatively photostable synthetic dyes are used for best SNR.
(b) Schematic of evanescent wave. Only fluorophores very close (50–300 nm) to the surface are excited. Bulk solution is not exposed to excitation illumination.
(c) Schematic of a prism-type single-molecule TIRFM setup.
Figure modified from Sarkar *et al.* and Bumb *et al.*[147,148]

10.4.2 Microscopy

10.4.2.1 Total internal reflection fluorescence (TIRFM)

The principles of TIRFM, FRET, and their applications to single-molecule studies are well established;[144,145] however, a briefly introduction is warranted to cover a few essential concepts.

TIR is a method to eliminate background light and image a very thin slice (50–300 nm) immediately adjacent to the TIR surface interface (Figure 10.3 (b)). This is accomplished by generating an exponentially decaying evanescent wave at an interface between high and low refractive index materials. The two most common approaches for generating this evanescent wave are quartz prisms (Figure 10.3 (c)) and high numerical aperture (NA) objectives. Though the setups are dated, see Axelrod *et al.*[146] for a concise theoretical understanding of TIR and evanescent waves. Complete details of our single-molecule setups are available in Appendix B.4.

10.4.2.2 Confocal

In a confocal setup, the excitation beam is focused on a small (fL) spot in the solution.[123] A small pinhole ($^1/_3$–$^2/_3$ point spread function (PSF) width) is used to eliminate emission signal from all but the focal spot. Smaller pinholes increase SNR at the expense of total signal. At very low concentrations (pM), fluorophores can be observed diffusing through the confocal spot.[149] FRET, if any, can be determined. Diffusion time through the confocal spot is a crude measure of particle size.

10.5 Specific aims

We wish to better elucidate the chemo-mechanical coupling, translocation and unwinding mechanisms of *E. coli* RecQ helicase *in vitro*. This is a multi-faceted and challenging undertaking, involving several interrelated sub-aims:

(i) Design, production and characterization of site-specific fluorophore labeled *E. coli* RecQ mutants for single-molecule studies.
   (a) Single-site specifically labeled mutants for use in bimolecular single-molecule experiments with site-specifically labeled DNA.
   (b) Double-labeled mutants for examining domain motions. The H1/H2 domains, on either side of the ATP-binding cleft, are of particular interest (Section 11.3.1.1).

(ii) Examine dynamics of binding, translocation and unwinding.
   (a) Unwinding with unlabeled protein and double labeled DNA.
   (b) Binding dynamics to short tailed DNA substrates using acceptor labeled RecQ and donor labeled DNA.
   (c) Inter-domain motions during translocation and unwinding with double labeled RecQ mutants.

The first aim, which constitutes Chapter 11, lays the foundation for future work using site-specifically labeled RecQ in single-molecule fluorescence assays. The second aim, which constitutes Chapter 12, is intended to clarify the translocation and unwinding mechanisms of RecQ.

11 RecQ substrate preparation

In this chapter we present the high-level methodological employed in the design, production and testing of novel RecQ substrates for single-molecule fluorescence studies. We have generated two previously unreported fluorescently labeled RecQ mutants, a family of four single-site specifically labeled mutants (two on the H1, and one each on the H2 and Zn domains), and a double-labeled mutant (H1/H2 domains).

To highlight unresolved questions concerning the function of the RecQ helicases, we begin by introducing what is known of the structure and function of the RecQ helicases (Section 11.2). Next, we examine existing RecQ mutants (Section 11.3), which reveals residues of particular functional importance; a double-edged sword when designing constructs. On the one hand, mutagenesis highlights key sites for protein stability, DNA and accessory protein interactions; one the other hand, they restrict manipulation of the protein, which may hamper construct design. We present our approach in Section 11.4, building on the insight garnered from previous experimental studies. Detailed experimental methods are omitted for clarity, and are presented in Appendix B.

11.1 Nomenclature

Throughout Part II, we discuss a variety of proteins and protein mutants. For clarity, we apply a consistent nomenclature.

11.1.1 Substitutions

<initial amino acid><residue index><substituted amino acid>

Where both <initial amino acid> and <substituted amino acid> are given as one letter codes, e.g. A for Alanine (Ala). For example, an asparagine (N) to cysteine (C) substitution at residue 65 would be given as N65C.

11.1.2 Deletions

<initial amino acid><residue index>del

Where both <initial amino acid> is given as one letter codes, e.g. A for Alanine (Ala). For example, a phenylalanine (F) deletion at residue 561 would be given as F561del. The initial amino acid may be omitted, e.g. 561del. Range deletions are also permissible, e.g. 490–492del corresponds to 490del-491del-492del.

11.1.3 Truncations

<protein id>$_{\text{<start>–<end>}}$

Instead of specifying all deletions, truncations are given with the protein name, and subscripted *included* residues. For example $\text{RecQ}_{1–524}$ contains only the first 524 residues of the RecQ protein. As the biological role of the truncated residues is not obvious for the residue numbers, biologically relevant truncations are often abbreviated, e.g. $\text{RecQ}_{1–524}$ as RecQΔHRDC, where HRDC is the biologically relevant helicase and RNaseD C-terminal domain.

11.1.4 N-terminal modifications

<N-terminal modification><protein id>

For ease of purification and expression, N-terminal modifications are commonly applied to existing proteins. For example, TGase-RecQ represents a transglutaminase (TGase) tag appended to the N-terminus of the RecQ protein. A complete list of N-terminal modifications are provided in Table 11.1.

System	Description
His	N-terminal polyhistidine tag (HHHHHH)
Throm	N-terminal thrombin tag (LVPRGS)
TEV	N-terminal Tobacco Etch Virus nuclear inclusion a endopeptidase tag (ENLYFQ)
TGase	N-terminal transglutaminase Q1-tag (PNP**Q**LPF)

Table 11.1 List of N-terminal modifications

11.1.5 Fluorescent labels

<protein id>-<fluorescent label>

Labels, such as fluorophores dyes, may be conjugated to proteins. We employ a variety of fluorescent dyes (see Figure B.5 and Table B.4 for structural and chemical properties respectively), and two complementary labeling schemes, maleimide and cadaverine.

Maleimide dyes selectively label surface exposed thiol groups, i.e. cysteine labeling. For exampple, RecQ-Alexa647maleimide corresponds to the Alexa647-maleimide dye conjugated to RecQ.

Fluorophores may also be conjugated to certain N-terminal tags, e.g. TGaseRecQ-Alexa647cadaverine implies conjugation of TGaseRecQ to Alexa647-cadaverine at the central glutamine (Q) residue (Table 11.1).

If the conjugation chemistry is not specified, maleimide may be assumed, i.e. Alexa647 implies Alexa647-maleimide.

11.2 Overview of structure & function

11.2.1 RecQ

E. coli RecQ is prototypical, and as such, is frequently used as a model system. It is closely related to other superfamily (SF)2 helicases, such as the hepatitis C virus NS3 RNA helicase and *Saccharomyces cerevisiae* Sgs1. It is astonishing how far our understanding of the RecQ-family

of helicases has come in the quarter century since *E. coli* RecQ was first identified[150] and determined to posses helicase activity[151] (reviewed in reference 128).

All RecQ helicases share two central helicase domains (H1 and H2) containing the conserved helicase motifs for adenosine triphosphate (ATP) binding and hydrolysis. Though not required for helicase activity, many RecQ helicases also contain RecQ C-terminal (RQC) and helicase and RNaseD C-terminal (HRDC) domains. These domains are thought to aid substrate specificity and mediate accessory protein interactions (Section 11.2.3). The RQC domain can be further subdivided into the Zn-binding domain, the site of a zinc-binding cysteine tetrad, and the winged-helix (WH) domain (Table B.5). For clarity, the domains locations of several RecQ-family helicases are outlined in Figure 11.1 (schematic) and Table B.5 (detailed, including residue numbers).

Structural[152] and kinetic[133,153] studies strongly suggest that RecQ functions as a monomer *in vitro*;[133,153] however, its *in vivo* oligomerization state is still a matter of some debate. As with other 3′ to 5′ translocating helicases, RecQ preferentially binds 3′-duplex DNA with a 5–50 nt 3′ overhang. It also binds to 5′ overhangs and single-stranded DNA (ssDNA), but with much lower affinity than ss/double-stranded DNA (dsDNA) junctions.[133] ATPase activity is only stimulated by ssDNA DNA does not stimulate the enzyme.[151] Unlike other helicases implicated in DNA replication and repair, RecQ is able to catalyze the unwinding of unusual substrates such as G-quartets, telomeric D-loops and Holliday junctions.

Zinc binding is implicated in the proper folding of RecQ; however, the cation is not required for RecQ activity.[154] While no *E. coli* RecQ/DNA crystal structures exist, crystal structure of RecQ homologues suggest that the WH domain is implicated in dsDNA binding,[155] while the HRDC[156] and H1/H2[157] domains are implicated in ssDNA binding (Figure 11.2). Seemingly in contrast to other motor proteins that exhibit a pronounced domain rearrangement upon ATP-binding (reviewed by in reference 130), the apo and ATP-γ-S crystal structures of the N-terminal full-length wild-type RecQ (RecQwt) domains reveals no large scale change in structure, with a minute H1-centroid to H2-centroid distance change 0.3 Å (Figure 11.3, Table 11.2). As the structure is not a DNA co-crystal, it is possible that a domain rearrangement only occurs in the presence of DNA.

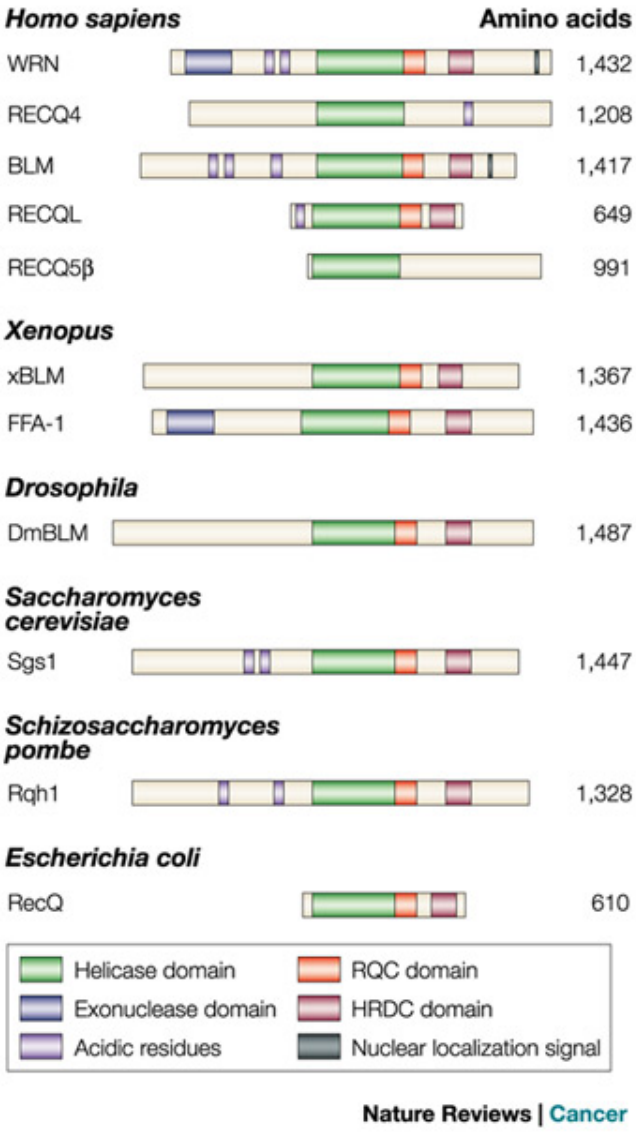


Figure 11.1 Conserved domains of the RecQ-like helicases
The catalytic helicase domain is conserved across all RecQ-like helicases. The RecQ C-terminal (RQC) domain (which is in-turn composed of zinc-binding and winged-helix (WH) sub-domains), and the helicase and RNaseD C-terminal (HRDC) domains are also highly conserved. Figure from Hickson 2003.[128]

	H1/H2 (Å)	N65C/S245C (Å)
Apo	52.3	52.0
ATP-γ-S	30.1	30.4

Table 11.2 H1/H2 distance
H1/H2 centroid distance and selected amino acid C_A-C_A distances.

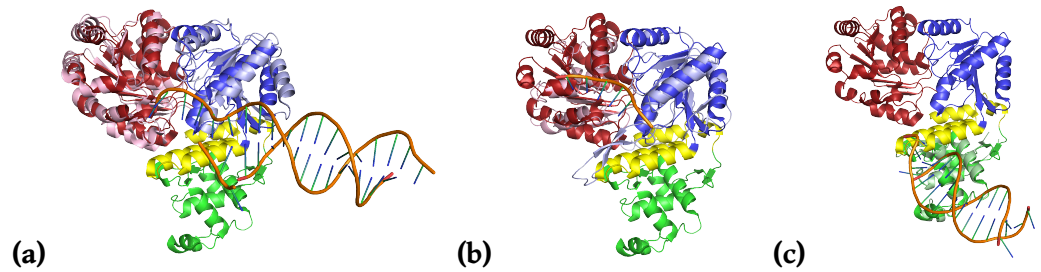


Figure 11.2 Structural alignments of *E. coli* RecQ (PDB: 1OYY)
E. coli RecQ domains: H1 (**red**), H2 (**blue**), Zn-binding (**yellow**) and winged-helix (**green**). Corresponding domains for RecQL1, NS3 and Werner syndrome RecQ helicase-like protein (WRN) are lighter shades of same color. Note the striking structural homology of the helicase domains and suggestive ssDNA binding orientation of the *E. coli* RecQ / human RecQL1 alignment.
(a) N-terminal domains of human RecQL1 in complex with 3′-duplex DNA (Protein Data Bank (PDB): 2WWY, alignment of H1/H2)
(b) Hepatitis C virus NS3 in complex with ssDNA (PDB: 1A1V, alignment of H1/H2)
(c) RQC-domain of human WRN in complex with blunt-duplex DNA (PDB: 3AAF, alignment of RQC domain).

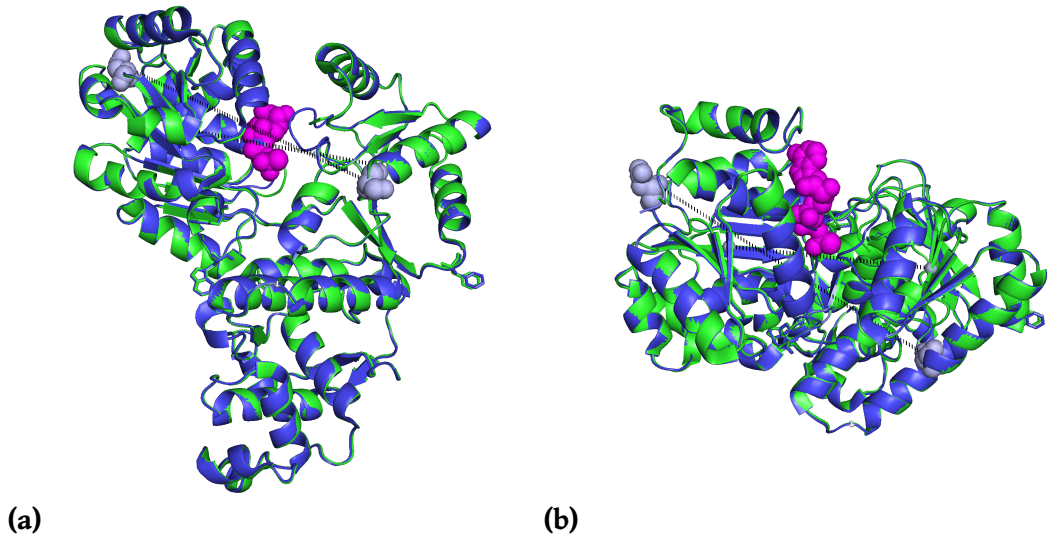


Figure 11.3 Crystal structure reveals no cleft closure upon ATP-γ-S binding
N-terminal domains of RecQwt in apo (**blue**, PDB: 1OYW) and ATP-γ-S bound (**green**, PDB: 1OYY) crystals. ATP-γ-S (**magenta** sphere), apo H1/H2 C_A-atom centroid (**light blue** sphere) and N65C C_G atom (H1) / S245C O_G atom (H2) (**light blue** sphere) distances are highlighted with black dashed lines.
(a) Front-view
(b) 90° *x*-axis rotation

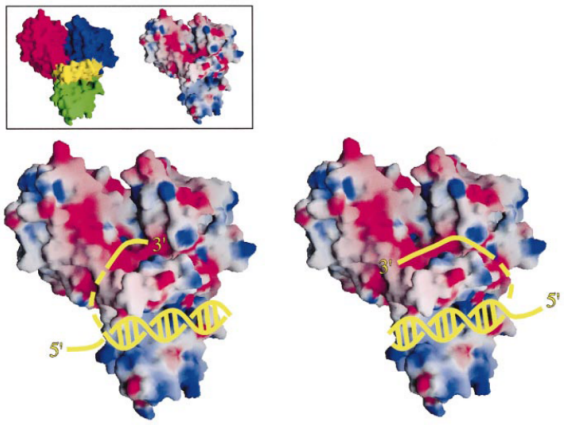


Figure 11.4 Visualization of *E. coli* RecQ
Main: Electrostatic surface rendering of protein, with positive (**blue**), negative (**red**) and neutral (**white**) charges. The HRDC domain, attached via a flexible loop to the winged-helix domain, is not shown.
Inset of helicase domains: H1 at residues 1-208 (**red**), H2 at residues 209-340 (**blue**), Zn-binding domain at residues 341-406 (**yellow**) and winged-helix domain at residues 407-516 (**green**).
Main: Bernstein, Zittel and Keck putative ss/dsDNA binding sites.[152] 3′-duplex DNA in yellow. Note the two ssDNA binding orientations. Figure from Bernstein Zittel and Keck.[152]

Biochemical studies of RecQ ss/dsDNA binding are decidedly mixed, some studies suggest competitive binding of ss/dsDNA to a single site,[158] whereas truncation mutant studies suggest localization of ss/dsDNA binding to distinct domains.[159] A putative model of the DNA binding orientations were proposed by Bernstein Zittel and Keck (Figure 11.4).[152] For convenience, a listing of RecQ-family helicases and their 4-character PDB identifiers are available in Table B.6.

11.2.2 UvrD

UvrD (Helicase II) is a SF1 prokaryotic helicase implicated in DNA repair through the nucleotide excision repair (NER) and methyl-directed mismatch repair (MMR) mechanisms. UvrD shares about 40 % sequence identity with SF1 helicases PcrA and Rep. It is also closely related to the eukaryotic *S. cerevisiae* Srs2 helicase, which like UvrD dissociates Rad51 (RecA) nucleofilaments.[160] There is some debate about the nature of the UvrD active unit. Competing models for monomeric[161] and dimeric[162,163] activity have been put forth. It has however been established that monomeric UvrD can translocate on ssDNA *in vitro*.[162]

Like RecQ, UvrD translocates unidirectionally from $3' \rightarrow 5'$.[164] As expected, it has a preference for binding/unwinding 3′-tailed duplex DNA, but can also process forked,[165] nicked and blunt ended substrates.[166] Interestingly, Runyon, Bear and Lohman suggest – from electron

microscopy images – that UvrD can unwind bidirectionally from nicks.[166]

UvrD is known to interact with a variety of other proteins implicated in DNA repair, including UvrABC[167,168] and MutS/MutL/MutH.[144] *In vivo*, UvrD is required to both unwind long stretches of dsDNA during MMR and dissociating RecA filaments from ssDNA.[145]

Several crystal structures have been produced for UvrD/DNA in complex with a variety of transition state ligands; for example, in complex with 3′-duplex DNA and a variety of transition state analogues have been determined (PDBs: 2IS1, 2IS2, 2IS4, 2IS6).[132] These structures have lead to the development of a detailed model of UvrD unwinding via a modified inchworm (rotating ratchet / two-part power stroke) mechanism. This model features a monomeric UvrD as the active unit. Given that each ss/dsDNA junction can only accommodate one UvrD monomer in the crystal structures, the mechanism of dimeric or oligomeric UvrD unwinding DNA is unclear. Structural similarities to *B. stearothermophilus* PcrA (SF1), were also used to assemble the model, specifically the DNA-binding induced domain motions.[121] The proposed UvrD 3′-duplex DNA unwinding cycle is presented in Figure 11.5.

11.2.3 Accessory protein interactions

11.2.3.1 E. coli RecQ

The active-type *E. coli* RecQ helicase functionally interacts with a variety of other DNA replication and repair proteins, including topoisomerase III (TopoIII) and single stranded binding protein (SSB). For example, *E. coli* RecQ specifically stimulates TopoIII strand passage activity, a function that neither enzyme can accomplish alone. However, a specific physical interaction has not yet been identified.[169–171]

E. coli RecQ and SSB physically interact as shown by tandem affinity purification isothermal titration calorimetry, and RecQ helicase activity is improved by pre-binding SSB.[172] Further work has localized this interaction to the RecQ winged-helix domain and the C-terminus of SSB. Indeed SSBΔC8 (SSB without the C-terminus) actually inhibits helicase activity. SSB also stimulated the concatenation of circular dsDNA in the presence of TopoIII and RecQ.[170]

The mechanisms of these accessory protein interactions are not well understood; however,

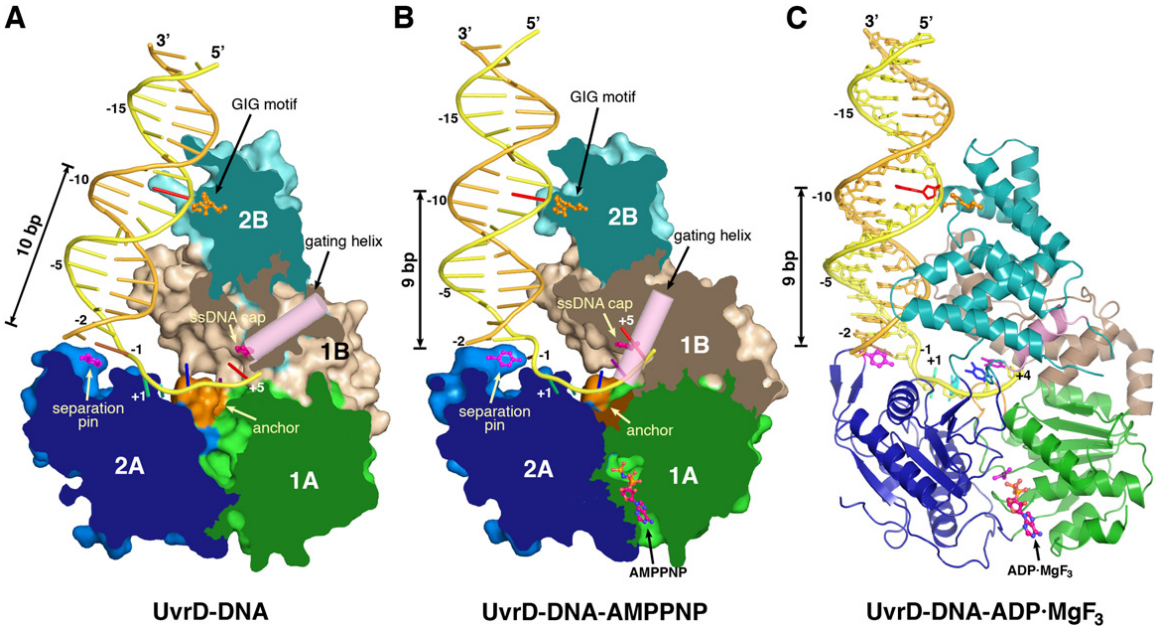


Figure 11.5 The 4 domains of UvrD in complex with the ss/dsDNA junction
(a) The separation pin Y621 forms a base stacking interaction at the ss/dsDNA fork. The gating helix is in the open position.
(b) 1A, 1B and 2B move together with respect to 2A during ATP binding. The separation pin changes conformation and physically interrupts the base pairing at the ss/dsDNA junction. The gating helix closes to prevent exit of the 3′ ssDNA.
(c) As the hydrolyzed ADP unbinds, the 1A/1B/2B and 2A domains return to their original position, the Y621 reforms base stacking with the next ss/dsDNA junction base pair, the gating helix opens and the ssDNA is pulled through.
Note that based on the crystal structure, only one UvrD monomer can physically occupy the ss/dsDNA junction.
Figure from Lee & Yang 2006.[132]

tellingly, the structural and functional equivalents of *E. coli* SSB from other species, such as the bacteriophage T4 gp32, the *S. cerevisiae* replication initiator protein A (RPA) and the human RPA, do not stimulate *E. coli* RecQ helicase activity, again suggesting a highly-specific interaction. Furthermore, SSB does not alter the ATP hydrolysis rate of RecQ. The stimulatory affect of SSB on RecQ helicases strongly suggests that the interaction is both conserved and specific. That is, the stimulatory affect is due to a specific physical interaction with RecQ and not solely by preventing the re-annealing of recently displaced strands.

11.2.3.2 Comparison with human RecQ helicases

The helicase stimulatory affect of accessory proteins is not unique to *E. coli* RecQ. Human RecQL1,[173] WRN[174] and Bloom syndrome protein (BLM)[175] all physically interact with, and are stimulated by human RPA, but not by *E. coli* SSB or T4 gp32. In fact, *in vitro*, RPA is required for the human helicases to fully unwind longer DNA substrates.

Analogous to SSB stimulation of *E. coli* RecQ, human RPA specifically stimulates the human RecQ helicases RecQL1,[173] WRN[174] and BLM.[175] Analogous to the role of *E. coli* RecQ in clearing RecA filament assemblies, the human RecQ helicases clear filaments of Rad51 nucleoproteins.

The role of BLM and WRN in telomere regulation is even less clear than their role in replication and repair. Telomere end binding protein (POT1) physically interacts with BLM and WRN, specifically stimulating the unwinding of telomeric D-loop and G-quadruplex DNA *in vitro*.[176,177] *In vivo*, these proteins are colocalized with, and known to associate with, telomeric DNA, further suggesting a role in telomere maintenance by resolving telomeric DNA structures.[178] BLM[179] and WRN[178] are also known to associate with TRF1 and TRF2, proteins that bind double stranded telomeric repeats (TTAGGG). These proteins are known to protect telomeric DNA from WRN exonuclease cleavage, and inhibit (TRF1) or stimulate (TRF2) BLM unwinding of telomeric substrates.[178,179]

11.2.3.3 Comparison with passive helicases

Accessory proteins interactions are not limited to active-type (Section 10.3.2) helicases, such as the RecQ-family. The unwinding velocity of passive helicases such as T7 gp4[180] and *E. coli*

DnaB[181] is dramatically increased by accessory proteins. For example, the presence of the replication machinery, specifically DNA polymerase, increase the unwinding rate of the DnaB replication helicase. Although the biophysical underpinnings of this stimulation have yet to be elucidated, there is a plausible physical basis – DNA polymerase and the replisome machinery acting as an additional motor which 'push' the helicase along the strand.

11.3 Overview of reported RecQ mutants

The central RecA-like domains are conserved across all SF2 helicases; however, mutagenesis studies strongly suggest that the unwinding mechanism is not identical for all RecQ-like proteins. Mutagenesis studies can be broadly classified into three focus areas:

(i) Putative unwinding pins (located on the H1, H2 and WH domains)
(ii) ATP-binding residues (located in the cleft between the H1 and H2 domains)
(iii) Truncation mutants (both N- and C-term).

11.3.1 Unwinding-pin mutants

There have been three putative unwinding pins identified in the RecQ family of helicases. These consist of a collection of aromatic residues on the H1 domain, H2 domain, and a β-hairpin on the WH domain. In *E. coli* RecQ, core residues of these unwinding pins are as follows (Figure 11.6 (a)):

- W154, F158, R159 (H1-pin, H1 domain)
- F221 (H2-pin, RecA-pin, resi 219-223 on H2 domain)
- W347 (Zn-pin, located on helix-turn-helix resi 342-347 in Zn-binding domain)
- H487 (WH-pin, located at tip of β-hairpin resi 484-496 in WH domain)

Naming conventions differ between laboratories, and residue numbers vary by protein, so for the sake of clarity, we follow the aforementioned naming convention throughout. For clarity, sequence alignments of several RecQ-family helicases are listed in Figure 11.7. These alignments highlight a high degree of unwinding pin conservation amongst the SF2 RecQ-like helicases, with expected decreases in conservation among the SF2 helicases, typified by UvrD.

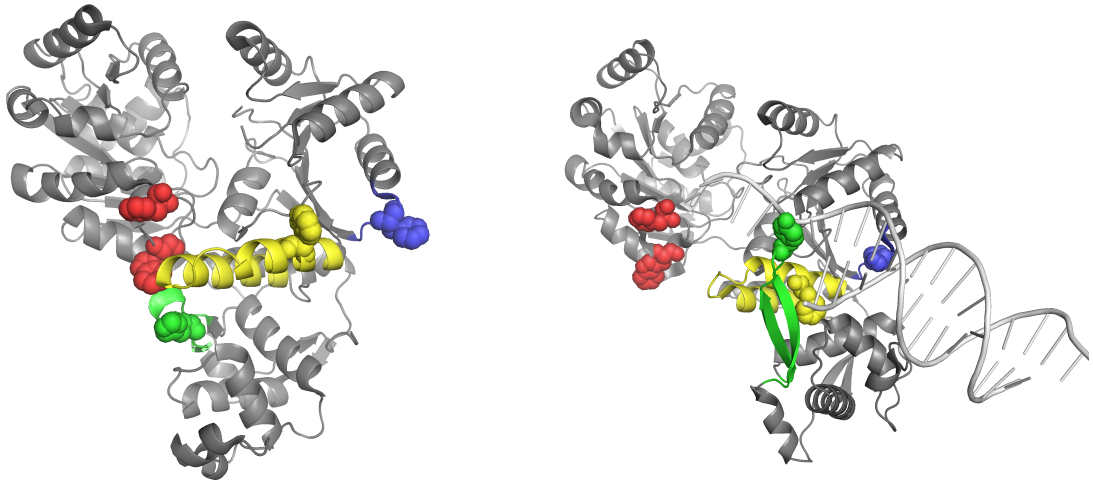


Figure 11.6 Unwinding pin locations
(a) *E. coli* RecQ (PDB: 1OYY) with the N-terminal domains ($RecQ_{1-516}$) in complex with ATP-γ-S, showing several motifs: H1-pin at W154 & F158 (**red**), H2-pin at F221 & residues 219-223 (**blue**), Zn-Helix-turn-Helix at residues 342-347 & Zn-pin at W347 (**yellow**), and the winged-helix (WH) β-hairpin at residues 484-496 and WH-pin at H489 (**green**).
(b) Human RecQL1 (PDB: 2WWY) with most of the N-terminal domains ($RecQL1_{49-616}$) in complex with 3′-duplex DNA, showing analagous motifs to **(a)**: H1-pin at W227 & F231 (**red**), H2-pin at P294 and residues 292-297 (**blue**), Zn-Helix-turn-Helix at residues 419-449 and Zn-pin at F424 (**yellow**), and the WH β-hairpin at residues 554-573 & WH-pin at Y564 (**green**).

11.3.1.1 *H1-pin*

In *E. coli* RecQ, the H1-pin (Figure 11.6 (a), Figure 11.7 (a)) is indispensable to efficient unwinding of 3′-duplex substrates.[182] Single mutants (W154L, F158L and R159L) reduce DNA stimulated ATPase activity, but maintain basal activities somewhat comparable to wild type, slightly stimulated by W154L and R159 and increased for F158L. Unwinding activity is abolished for some mutants (W154L, W154L_F158L, R159L) and significantly impaired for others (F158L). None of the mutants impact ssDNA binding, suggesting their non-trivial role in unwinding, possibly through chemo-mechanical coupling of ATP hydrolysis and strand separation. Likewise the analogous RecQL1 H1-pin residues W227A and F231A (Figure 11.6 (b), Figure 11.7 (a)) are required for unwinding activity, but do not impact DNA-binding or oligmerization.[183] As an interesting aside, the H1 and H2 pins are swapped on the $5' \rightarrow 3'$ RecD2 helicase (SF1B).[184]

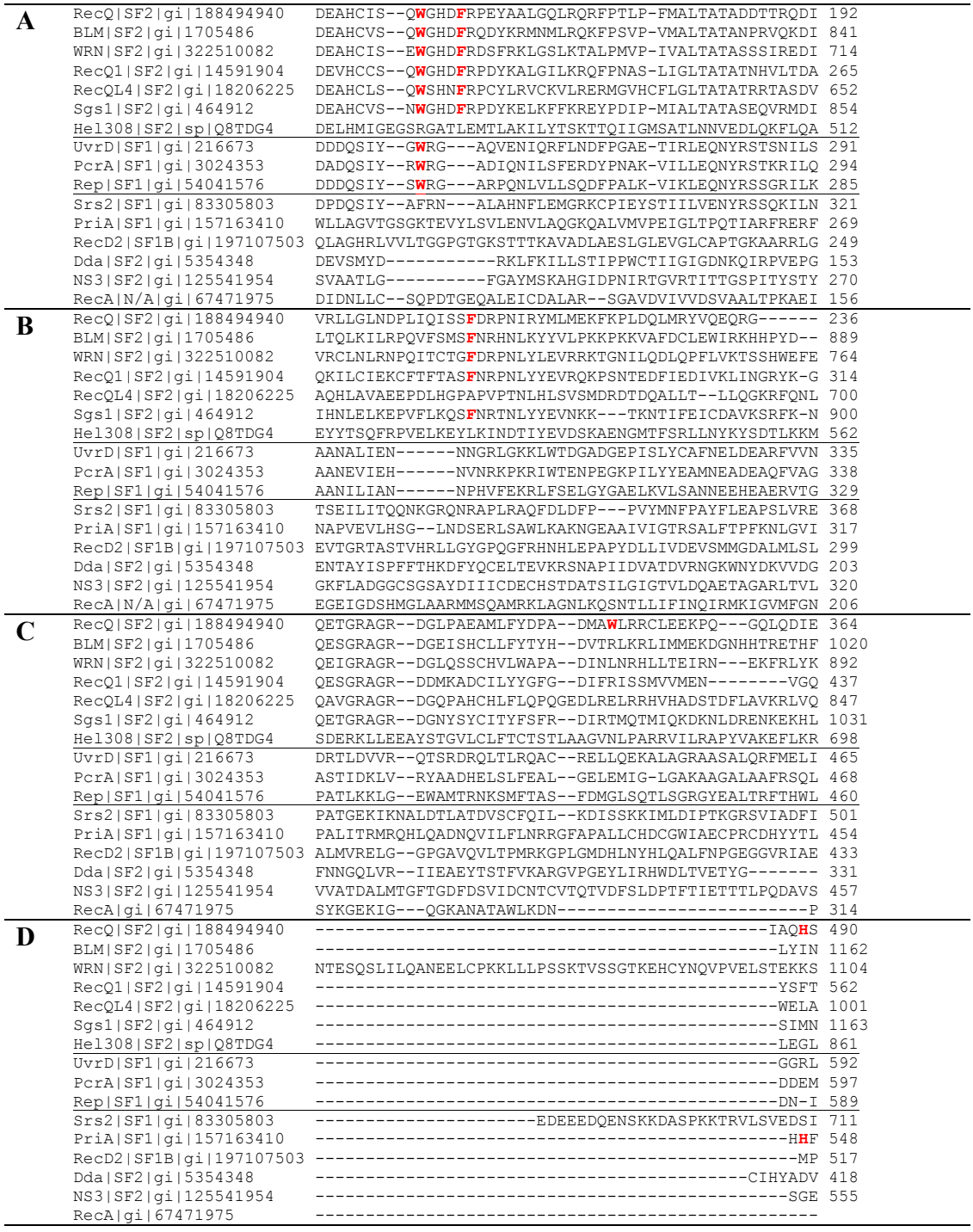


Figure 11.7 Sequence alignment of RecQ-family helicase unwinding pins
(a) H1-pin alignment (W154, F158), **(b)** H2-pin alignment (F221), **(c)** Zn-pin alignment (W347) and **(d)** WH-pin alignment (H489). Relevent residue numbers for the *E. coli* RecQ pins are highlighted (**red**).

11.3.1.2 WH-pin

The WH-pin is not essential for unwinding in all RecQ-family helicases. Wild type human RecQL1 exists as a mixture of oligomers.[185] The WH β-hairpin located at residues 554–573 (Figure 11.6 (b), Figure 11.7 (c)), specifically mutants RecQL1$_{49-616}$-F561del-T562del-T566del and RecQL1$_{49-616}$-Y559A-S560A-F561A, are responsible for dimerization.[183] Cross linking experiments reveal that these mutants are monomeric, even in the presence of ssDNA. In DNA fractional unwinding assays, helicase activity of forked dsDNA is diminished, but not abolished, indicating that a completely intact WH β-hairpin is not required for unwinding. The WH-pin residue (Y564A) is more important for forked DNA unwinding, abolishing unwinding activity in the RecQL1$_{49-616}$ dimer and reducing unwinding activity in RecQL1$_{FL}$ oligimer (Figure 11.6 (b)).[182,183]

The WH β-hairpin is also important for WRN, another human RecQ-family protein. Like human RecQL1 it also has a pronounced β-hairpin. This β-hairpin is tipped by an aromatic residue (F1037) which is found unpairing dsDNA in a co-crystal of dsDNA and WRN WH domain (PDB: 3AAF).[155] Unsurprisingly, a F1037A mutation is shown to severely impair WRN RQC binding to 5′-duplex DNA.

Analogous β-hairpin deletion (490–492del) and WH-pin mutants (H491A) on *E. coli* RecQ do not perturb DNA unwinding or DNA-stimulated ATPase activity (Figure 11.6 (a)).[186] Note however that the *E. coli* RecQ β-hairpin (484–496) is significantly shorter than the RecQL1 equivalent (554-573), and is not oriented towards the putative DNA binding channel. Hoadley and Keck[187] have posited that this is due to either a fundamentally different unwinding mechanism or that the winged-helix domain might reorient upon DNA binding. In either case, mutations in the WH-pin and β-hairpin do not perturb function.

Likewise the BLM WH-pin, which unlike *E. coli* RecQ and human RecQL1 is not an aromatic residue, is not required for strand separation.[188] This suggests that the role of the WH-pin in unwinding subtly differs between the human RecQ helicases, and that like the N-term and C-term accessory domains, it may be tuned to the particular role of each helicase.

11.3.2 ATP-binding mutants

Several residues on the H1 domain comprise the ATP binding site: Y23, R27, Q30, K53 (sometimes reported as K55), S54, L55 and N146. K55A has been demonstrated in several studies to abolish helicase activity, but to have no impact on DNA binding.[189] ATP analogues, useful for approximating intermediate hydrolysis steps, are reviewed elsewhere.[190] For example, adenylylimidodiphosphate (AMP-PNP), a non-hydrolyzable analogue thought to mimic bound ATP, is used later in this study (Section 12.1).

11.3.3 Truncation mutants

Both N- and C-term truncation mutants have been reported for RecQ-family helicases.

As a general trend in higher organisms, the N-terminal domains before the Walker A/B ATPase motifs appear to govern oligomerization state; the C-terminal domains, substrate specificity. Given this general observation, it is expected that N- and C-term domains should not be essential for core unwinding activity on uncomplicated forked dsDNA substrates, but that they may be required to resolve more complex DNA substrates such as Holliday junctions or other recombination intermediates. To understand the precise roles of these domains, there have been efforts to identify the minimal unwinding unit of the human RecQ helicases, and to examine the role of the C-terminal HRDC and WH domains of *E. coli* RecQ.[188]

It has been shown that the double truncation mutant $BLM_{642–1077}$ of the 1417 residue Bloom syndrome protein is sufficient for duplex DNA helicase activity.[188] The N-terminal domain is suspected to be involved with oligomerization. Curiously, in sequencing studies on Bloom syndrome patients, several polymorphisms (V1198M, S1209T, V1321I, H1324Y) occur at positions presumably not necessary for BLM activity.[191] Analogous double truncation mutants on the 1447 residue yeast BLM homologue, $Sgs1_{400–1268}$, showed DNA binding, basal/DNA-stimulated ATPase activity and forked DNA unwinding comparable to wild type Sgs1.[192]

Likewise *E. coli* RecQ C-terminal truncation mutants ($RecQ_{1–524}$; abbreviated as RecQΔHRDC for clarify) are just as effective as wild type in unwinding forked dsDNA.[186] A similar truncation mutant, $RecQ_{1–516}$, showed reduced ssDNA binding.[156,159] Presumably, residues 516–524 occur

in an unstructured loop connecting the N-terminal and HRDC domains. Interestingly, the WRN HRDC domain does not bind DNA,[155] whereas the *E. coli* RecQ HRDC has been shown to preferentially bind ssDNA over dsDNA; even a putative binding channel has been localized.[156] Substrate preferences, revealed by binding affinities to a variety of isolated DNA substrates, are broadly similar between the WRN and *E. coli* RecQ RQC domain.

11.4 Novel RecQ substrates

Non-specifically labeled RecQwt, site-specifically labeled RecQ mutants and a double-labeled RecQ mutant, have been purified and quality controlled for DNA-binding, basal/DNA-stimulated ATPase activity and ensemble Förster resonance energy transfer (FRET). Neither the site-specifically labeled nor double-labeled RecQ constructs have been reported in the literature. For convenience, a listing of all purified RecQ substrates is shown in Table B.1. For the reader not familiar with commonly used fluorophores, their structures and properties are listed in Figure B.5 and Table B.4, respectively.

11.4.1 Non-specifically labeled

As a proof of concept, wild-type RecQ was non-specifically labeled with Alexa647-maleimide. Given that RecQ contains 11 cysteines (Figure 11.8 Native CYS highlighted in gray), and that at least 6 (Figure 11.9 (a)) of these cysteines appear solvent exposed from the available crystal structures (PDB: 1OYW, 1OYY), the precise Alexa647-maleimide labeling sites could not be ascertained. Overall labeling stoichiometry of RecQ:Alexa647 labeling was approximately 1:1.2 as indicated by A_{280} ($\varepsilon_{280}^{\text{RecQ}} = 48\,820\ \text{M}^{-1}\,\text{cm}^{-1}$) and A_{650} ($\varepsilon_{650}^{\text{Alexa647}} = 238\,000\ \text{M}^{-1}\,\text{cm}^{-1}$; see Table B.4 for comparison to other fluorophores) (Figure 11.10 (a)). The A_{280} contribution of Alexa647 is negligible. Unsurprisingly given 6 surfaces exposed cysteines, these results do suggest that RecQwt posses multiple liable sites.

Interpreting data from single-molecule assays using this non-specifically labeled substrate proved difficult, due in part to the possibility of multiple acceptor fluorophore positions. Knowledge of the labeling locations or site-specifically labeled RecQ should facilitate data interpretation. Electrospray tandem mass spectrometry (ESI-MS/MS) was attempted to identify

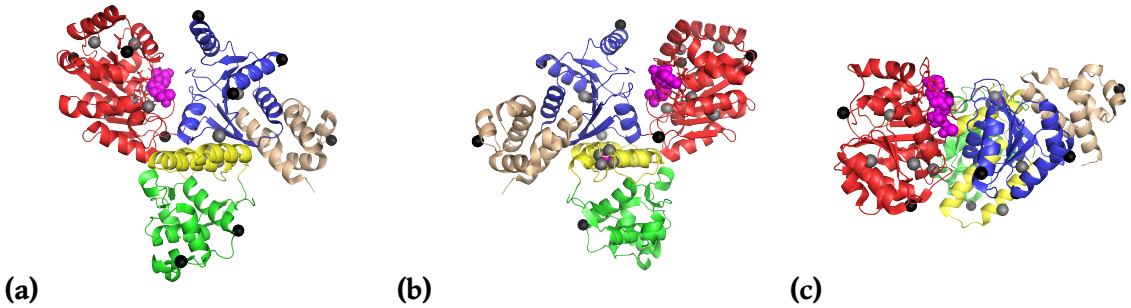


Figure 11.8 RecQ cysteine mutants (CYS → SER and NNN → CYS)
Cartoon representation of Helicase 1 (**red**), Helicase 2 (**blue**), Zn-binding (**yellow**), Winged-Helix (**green**) and HRDC (**wheat**) domains. ATP analogue (**magenta**). Structure based on RosettaDock[193] prediction of N-terminal domains (PDB: 1OYY) and HRDC domain (PDB: 1WUD).
Sphere representation of Native CYS → SER (**gray**) and NNN → CYS (**black**). S_G or equivalent atom highlighted. H1 (1-208), H2 (209-340), Zn-binding (341-406), Winged-Helix (407-516), HRDC (517-609)
Native CYS → SER (C43S, C56S, C94S, C111S, C150S, C243S, C351S, C380S, C397S, C400S, C403S)
Mutant NNN → CYS (N65C, S97C, S207C, S245C, S258C, N273C, Q445C, S465C, S579C)
(a) Front-view
(b) Back view, 180° y-axis rotation from front-view
(c) Top view, 90° *x*-axis rotation from front-view

the labeling locations of Alexa647-maliemide. While coverage was comparable between labeled and unlabeled substrates, the labeling positions could not be localized (data not shown). From quadrupole MS experiments on labeled RecQ, we suspected that the ionization of the labeled substrates was inferior to the native protein. An additional Mono Q purification step was attempted to improve the effective labeling stoichiometry. Unfortunately, there was insufficient separation between the labeled and unlabeled RecQ peaks for enrichment (data not shown).

11.4.2 Site-specifically labeled, single-label

Following experiments with non-specifically labeled RecQ, we moved to generate site-specifically labeled RecQ, in the hope that these substrates would yield more interpretable data in single-molecule assays. Two distinct labeling schemes were attempted, maleimide and transglutaminase (TGase)-mediated cadaverine transfer to an N-terminal tag. TGase, and other enzymatic labeling schemes, are reviewed in reference 194. Ideally, maleimide site-specific labeling is done in the presence of only one liable surface exposed cysteine. Therefore, generating a 'CysLite' mutant, not liable to maleimide dyes, is a necessary first step in creating a site-specifically liable mutant (Figure 11.11). With a CysLite construct, we may substitute a cysteine any desired position (Figure 11.12). For reference, native cysteines are depicted in gray in Figure 11.8.

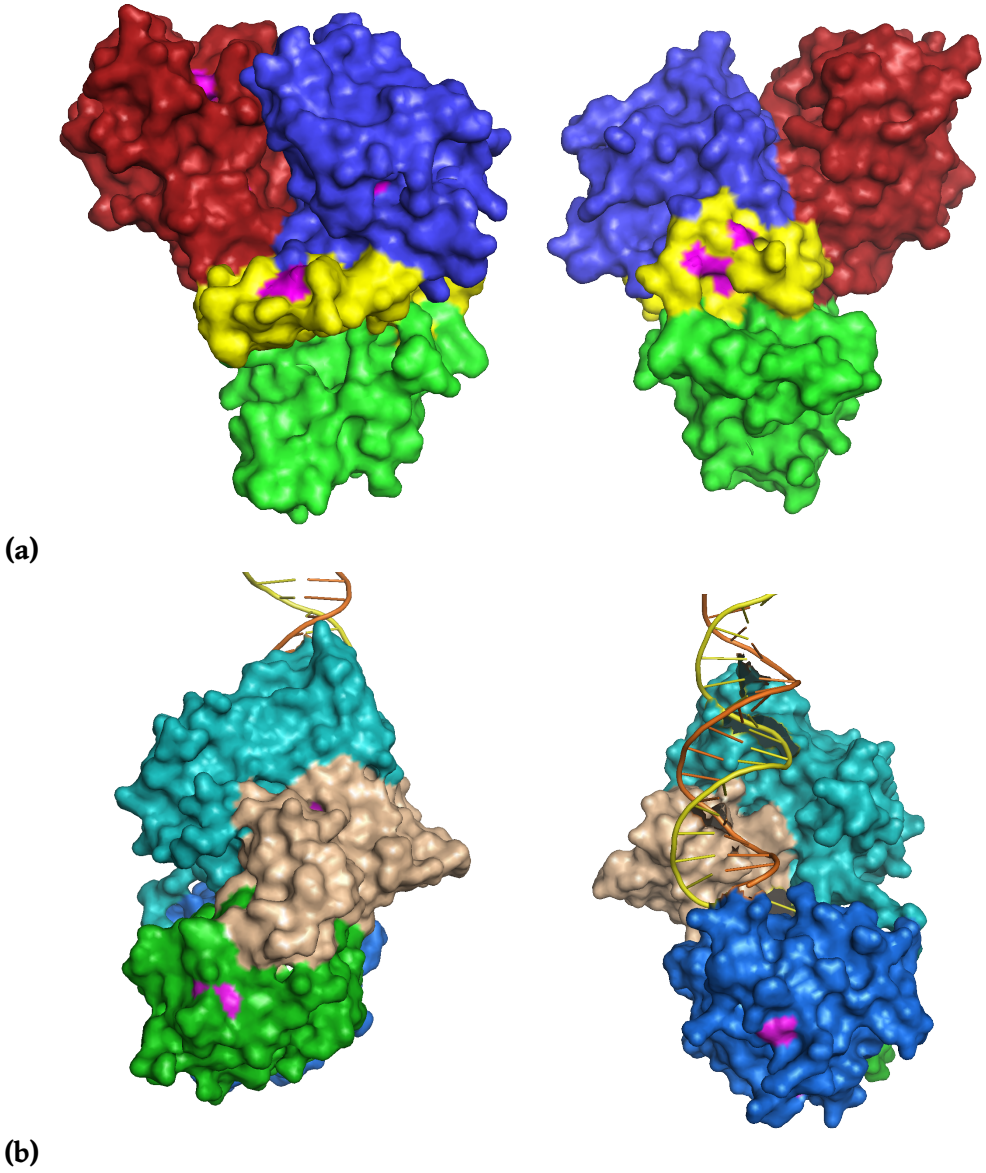
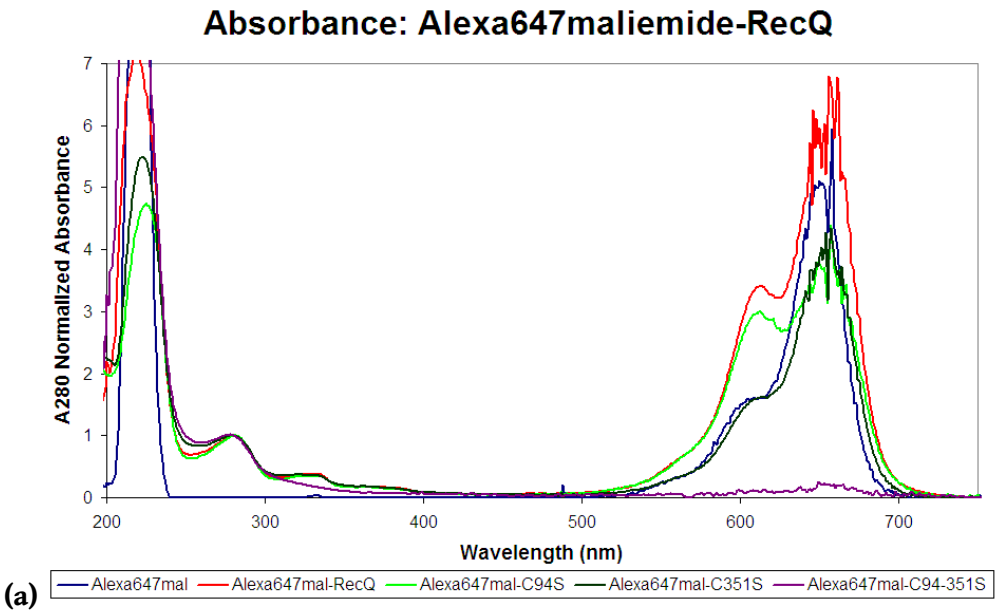
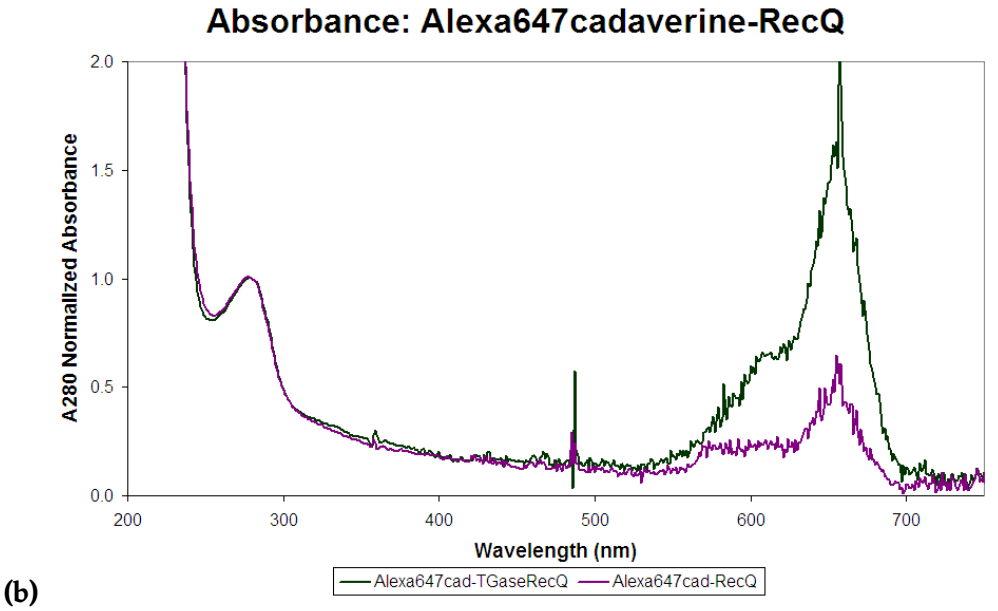


Figure 11.9 Surfaced exposed cystines of RecQ and UvrD
Surface exposed cystines highlighted (**magenta**).
(a) RecQ domains: Helicase 1 (**red**), Helicase 2 (**blue**), Zn-binding (**yellow**), Winged-Helix (**green**). The H1 (1-208), H2 (209-340), Zn-binding (341-406) and domains contain 5, 1 and 5 cysteine respectively. The Winged-Helix (407-516) and HRDC (517-609) domains contain no cysteines. C94 (H1) and C351 (WH) appear to be partially exposed. The cysteine tetrad (C380, C397, C400, C403) coordinates Zn^{2+}.
(b) UvrD domains: 1A (**green**), 1B (**wheat**), 2A (**blue**) and 2B (**teal**). The RecA-like domains are 1A and 2A (**green** and **blue**, respectively). The 1A (1-89, 215-280), 1B (90-214), 2A (281-377, 551-647) and 2B (378-550) domains contain 1, 1, 3 and 1 cysteines respectively. C52 (**1A**) and C322 (**2A**) appear to be partially solvent exposed.



(a)



(b)

Figure 11.10 Alexa647 labeling of RecQ
(a) Alexa647-maleimide/RecQ labeling. Alexa647-maliemide (**blue**), RecQ-Alexa647 (**red**), RecQC94S-Alexa647 (**green**), RecQC351S-Alexa647 (**dark green**) and RecQC94-351S-Alexa647 (**purple**). Note that HisThromRecQ C94-351S (**purple**) has negligible A_{650} absorbance. This indicates that the vast majority of labeling occurs at C94 and C351. Labeling stoichiometry based on A_{650}/A_{280} and the extinction coefficients of RecQ and Alexa647 ($\varepsilon_{280}^{\text{RecQ}} = 48\,820\ \text{M}^{-1}\,\text{cm}^{-1}$, $\varepsilon_{650}^{\text{Alexa647}} = 238\,000\ \text{M}^{-1}\,\text{cm}^{-1}$) were as follows: RecQ (1.22), C94S (0.77), C351S (0.79) and C94-351S (0.05).
(b) Alexa647-cadeverine/RecQ labeling. TGase-RecQ-Alexa647cadaverine (**dark green**) and RecQ-Alexa647cadaverine (**purple**). Labeling stoichiometry was as follows: TGase-RecQ (0.29) and RecQ (0.09). All traces normalized to A_{280}, with the exception of the Alexa647-maliemide dye, which is scaled proportional to the A_{220}, since its A_{280} is negligible.

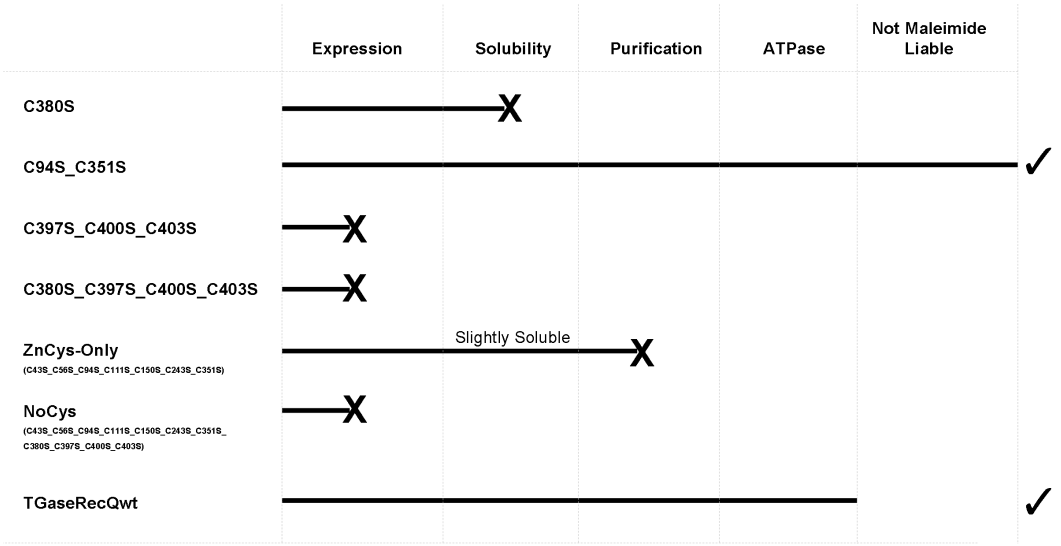


Figure 11.11 Overview of selected CysLite mutants
Note that the TGase-RecQwt mutant is maleimide liable. This is by design, as labeling is intended to be labeled via transglutaminase mediated cadaverine transfer to the N-terminal TGase tag.

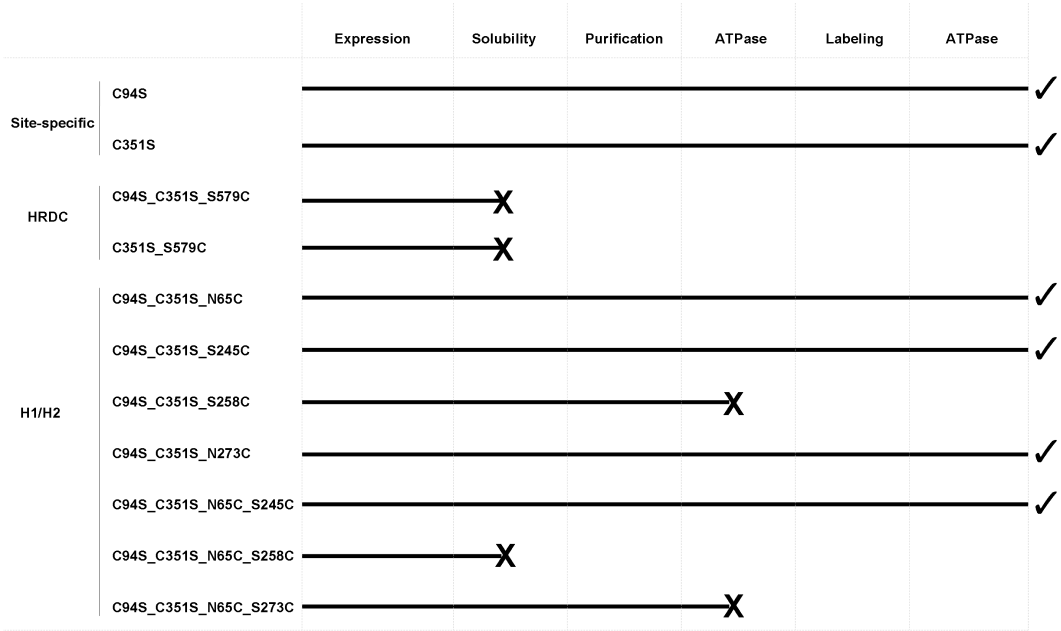


Figure 11.12 Overview of selected single and double mutants
Note that the C94S and C351S site-specific mutants are liable on the opposite residue (i.e. C94S labels C351 and C351S labels C94). Note that two active single mutants do not necessarily imply an active double mutant. For example, both C94S_C351S_N65C and C94S_C351S_N273C are active, but C94S_C351S_N65C_N273C is not.

11.4.2.1 Zinc protects surface exposed cysteine tetrad

Previous studies have shown that mutations to the Zn-binding cysteine tetrad have a deleterious impact on RecQ purification (probably inclusion bodies) and function.[154] In demonstrating that tetrad mutants (C380S and C380S_C397S_C400S_C403S) do not express (Figure 11.11), we have confirmed this observation. Liable cysteines that cannot be removed by mutagenesis would ordinarily be problematic; however, a prior study using 4-(2-pyridylazo)resorcinol (PAR), an absorbance based reporter of free Zn^{2+},[195,196] and 5,5′-dithiobis-(2-nitrobenzoic acid, also known as Ellman's reagent, (DTNB),[197] an absorbance based reporter of free thiols, suggests that the tetrad cysteines are protected in the presence of zinc.[152] In this study, the ratio of free thiol to protein increases from 3.7:1 to 7.2:1 after the removal of Zn^{2+} by dialysis in dithiothreitol (DTT) and ethylenediaminetetraacetic acid (EDTA). Given that the Zn-finger is highly conserved, sensitive to mutations, and may be protected from labeling, we proceeded in attempts to generate a CysLite mutant without mutating the Zn-finger region, and use $ZnCl_2$ in our labeling conditions to protect the otherwise surface exposed Zn-finger cysteines.

Using serial mutagenesis, we generated a HisTEVRecQ ZnCys-Only mutant with a 6-His tag, Tobacco Etch Virus nuclear inclusion a endopeptidase (TEV)-protease site, with all 7 non-tetrad cysteines mutated to serine. This mutant was only very slightly soluble and could not be purified (for a discussion of the approaches attempted to improve the solubility of this, and other, RecQ mutants see Section 11.5.2).

11.4.2.2 CysLite mutant

After careful inspection of the available RecQ crystal structures (Figure 11.9), several intermediate Cys → Ser mutations were expressed, purified and tested for Alexa647-maliemide labeling. The HisThromRecQ C94S_C351S double mutant, dubbed CysLite, does not label at a 6:1 dye:protein ratio (Figure 11.10 (a)). This CysLite mutant was used as a base from which to generate a small family of mutants with a single liable cysteine.[i]

These mutants were then site-specifically labeled with maleimide dyes. For example, HisThromRecQ C94S and HisThromRecQ C351S have a stoichiometry of approximately 1:0.8

[i] A similar two-step, CysLite to site-specifically labeled, methodology could prove tractable for UvrD (Figure 11.9 (b)).

(RecQ:Alexa647). Characterization included DNA-stimulated ATPase (Figure 12.4) and binding to $3'$-duplex DNA (Figure 12.5). See Section 12.2 for further discussion.

11.4.2.3 *Transglutaminase (TGase)-mediated construct*

TGase mediated site-specific labeling for single-molecule studies has been established previously.[198,199] The TGase site is specifically labeled with stoichiometry Alexa647:RecQ $\approx$ 0.3 (Figure 11.10 (b)). The biggest downsides of this approach, compared to maleimide labeling, is that TGase-site positions are limited to termini and long flexible loop regions, such as the linker between the RecQ WH and HRDC domains. Single-molecule experiments were attempted with both maleimide and cadaverine labeled substrates.

11.4.3 Site-specifically labeled, double-label

Following the identification of a CysLite base and subsequent single site-specifically labeled constructs, additional mutants were attempted to generate a double-labeled construct. Rationally designed sites were selected on the H1, H2, Winged-helix and HRDC domains by the following procedure:

(i) Maximum distance sensitivity is achieved around the R_0 for a given fluorophore pair. Given that the R_0 for common donor (Cy3/Cy3B) and acceptor (Cy5/Alexa647/Atto647N) dyes range from 5–6 nm, pairs with C_A-C_A distances of 4–7 nm are desirable (Figure 11.13 (a)).

(ii) Maleimide reactivity requires that the S_G cysteine atom is solvent accessible. Buried or disulfide-bonded cysteines are not liable. Therefore, it is required that sites have solvent accessible surface areas > 20 Å^2 (Figure 11.13 (b)).

(iii) RecQ mutants must retain DNA binding, basal ATPase and DNA-stimulated ATPase activities comparable to wildtype. To minimize disruption, only small, non-aromatic, uncharged residues (SER, THR, GLN, ASN) were considered for mutagenesis (Figure 11.13 (c)).

(iv) Observed FRET is correlated to the distance between fluorophore positions, not residue positions. The size of the dye, length of the linker, and flexibility of the protein local environment all contribute to fluorophore mobility. It is generally desirable that the fluorophore remain rotationally unconstrained such that the kappa-squared assumption

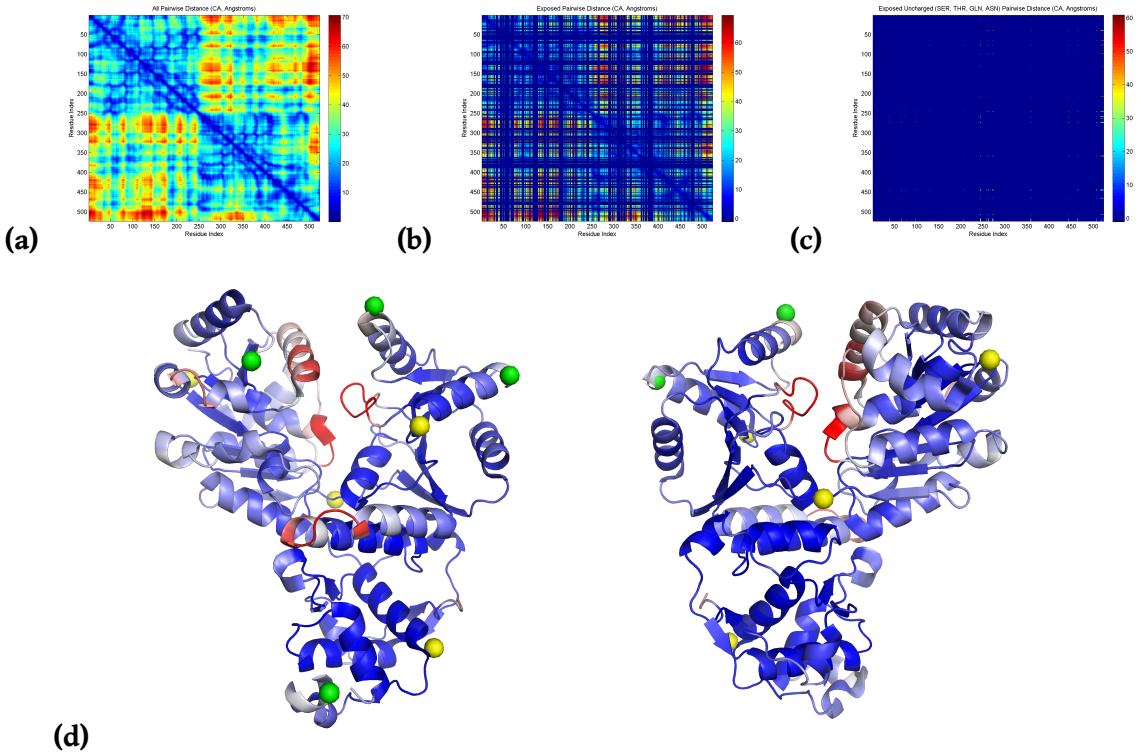


Figure 11.13 Rational design procedure for NNN → CYS mutants
(a) C_α-C_α pair-wise distance distribution (ideally 4–7 nm).
(b) Residues with solvent assessable surface area > 20 Å.[119]
(c) Small, non-aromatic, uncharged residues to minimize disruption (SER, THR, GLN, ASN).
(d) *B*-factor map of RecQ: $B < 10$ (**blue**) to $B > 50$ (**blue**). Putative labeling site residues with $B < 20$ (**yellow**) and $20 < B < 30$ (**green**)

holds ($\kappa^2 = 2/3$).[119,137] This is why most maleimide dyes, including the ones used in these studies, contain a C_6-C_{12} linker between the reactive maleimide and dye. Additional fluorophore mobility attributable to the protein local environment is undesirable because it further decouples the desired measurement (distance between residue positions and the domains upon which they reside) and the observed measurement (distance between fluorophore positions). B-factor, a measure of the dispersion from theoretical electron density, is a proxy for local environment flexibility. Lower B-factors generally correspond to less flexible regions, while higher B-factors correspond to more flexible regions. Since we wish to detect relative distance changes between relatively rigid domains, residue positions with low B-factor are preferred (Figure 11.13 (d)).

Selected sites are highlighted in black in Figure 11.8. Given the possibility of a H1/H2 cleft opening/closing during translocation and unwinding, emphasis was afforded to these mutants (Figure 11.12). The double labeled HisTEVRecQ C94S_C351S_N65C_S245C, designed to monitor

H1/H2 cleft movement, labels with high efficiency, demonstrates basal and DNA stimulated ATPase turnover comparable to wildtype and shows FRET in ensemble fluorometry assays (Section 12.1).

11.5 Discussion

11.5.1 Controversy surrounding Thrombin cleavage of RecQ

Many research groups employ a protocol using Thrombin[159] or biotinylated-Thrombin[153] for full-length RecQ preparation. Briefly, Thrombin cleavage was conducted at room temperature overnight at 10 µU Thombin (Sigma) per mg of protein in standard buffer (10 % Glycerol, 300 mM NaCl and 20 mM Tris HCl pH7.5) supplemented with 10 mM CaCl.

However in our study, Thrombin cleaved both the tag and RecQ, generating ∼60 kDa and ∼10 kDa fragments (by gel electrophoresis) that do not bind a Ni-nitrilotriacetic acid (NTA) column (evidence that the 6-His tag is removed). Examination of the sequence by the Expasy peptide cutter reveals a thrombin site at position 514, immediately before the HRDC domain. The theoretical mass of this 96 residue fragment is 10.8 kDa, which corresponds to that of our observed fragment in electrospray ionization mass spectometry (ESI-MS). N-terminal sequencing revealed that the fragment began 6 residues upstream of the theoretical cut site, yielding a theoretical mass of 9.9 kDa. This mass was confirmed via ESI-MS (data not shown). We therefore conclude that these two fragments correspond to RecQΔHRDC and the HRDC domain.

An alternative His-tag free purification protocol for RecQ purification has been previously reported;[200] however, we have developed a simpler protocol using a TEV protease site instead of Thrombin, which indeed produces full-length wild-type RecQ (RecQwt) (see Methods Appendix B.1 for detailed proceedure). Ensemble kinetics studies on RecQ, RecQΔC (equivalent to RecQΔHRDC), RecQΔN and RecQΔNΔC have been performed.[159] While the study reported that the ATPase turnover rates of what was reported to be RecQwt and RecQΔC are indistinguishable, we initially questioned their result as both substrates utilized a Thrombin cleavage in their preparation. As such, we have generated wild-type and site-specifically labeled Thrombin-cleaved RecQΔHRDC substrates (Table B.1). RecQΔHRDC was separated from the HRDC domain

via gel filtration. Full length RecQ and RecQΔHRDC were determined to have comparable basal ATPase, DNA-stimulated ATPase and binding properties (Section 12.2). Both the RecQwt from the TEV protocol and RecQΔHRDC from the Thrombin protocol were used in single-molecule experiments to explore the role of the HRDC domain.

We cannot explain how other experimental groups[153,159] obtained a full-length wild-type RecQ from Thrombin cleavage. From our data, we must conclude that they inadvertently used RecQΔHRDC owing to its very similar ensemble properties (Section 12.2). The use of RecQΔHRDC instead of RecQwt may further complicate the interpretation of an already confused literature.

11.5.2 Insolubility of RecQ mutants

It was observed that RecQ mutants have reduced solubility compared to RecQwt. A variety of methods were used to increase the overall yield of RecQ mutants by increasing their solubility:

- Temperature (15 °C, 25 °C and 37 °C)
- Supplemental 10 mM $ZnCl_2$ in growth media
- Expression system (Rosetta2(DE3)pLys, Lemo21 and BL21(DE3)pLys)
- Media (Invitrogen MagicMedia™ *E. coli* Expression Medium SoluPouch™ and LB-Broth[201])

Mutant RecQ solubility was not improved under any *E. coli* growth condition assayed, and was insoluble for all post-induction time points (data not shown).

11.5.3 Use of 'inline' fluorophores for labeled DNA preparation

While a variety of DNA substrates were used in the preliminary single-molecule total internal reflection fluorescence microscopy (TIRFM) assays, all labeled DNA substrates used an inline labeling scheme (Figure 11.14). A major drawback of this labeling scheme is that internal labeling sites necessitate the replacement of an otherwise natural base pairing. Other schemes, such as an amino modified C_6dT branch from the base and do not perturb natural base pairing. This is a consideration often overlooked in single-molecule studies, and one potential reason why the interpretation of the preliminary results has proven problematic. Unfortunately, branched labeling may interfere with RecQ translocation and unwinding.

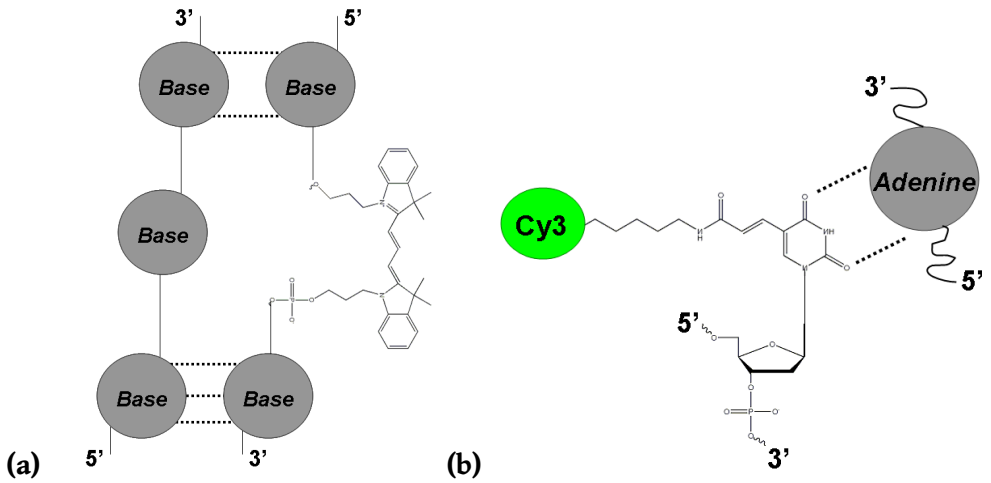


Figure 11.14 DNA labeling schemes, inline versus branched
(a) Inline: Int Cy3, and **(b)** branched: Int amino modifier C_6dT. Note that the inline labeling scheme **(a)** would interfere with natural base pairing since it replaces a base. A branched labeling scheme **(b)** does not interfere with base pairing but may act as an obstruction along the backbone.

Another consideration is the close proximity of the Cy3 and Cy5 dyes in the labeled DNA/unlabeled protein experiments. Cyanine dyes are notorious for blinking; this is exacerbated by very close physical proximity to other dyes.

In selecting DNA substrates, one should consider minimizing base pairing disruption, increasing the distance between the acceptor/donor and dye photostablity (*e.g.* Alexa series, Cy3B and Cy5.5). The advantages of Cy3B over Cy3, namely increased brightness and photostability, are discussed elsewhere.[202] Schematics of all DNA substrates are shown in Figure 12.6 (a).

11.6 Conclusion

We have reported the design, production and quality-control of fluorescently labeled RecQ substrates, sufficient for use in single-molecule assays. Most importantly, we have revealed a potentially problem with a commonly used RecQ purification protocol, suggesting that the RecQ species used in many experiments is a truncated RecQΔHRDC mutant, and not RecQwt (Section 11.5.1).

We have also overcome a number of technical hurdles, applicable to not only RecQ, but also other proteins which are challenging to site-specifically label owing to a large number of surface exposed cysteines. Namely, we generated a CysLite substrate with which we can freely introduce

site-specific fluorophore labeling sites, despite the large number of maleimide-liable surfaces exposed cysteines. The most important technical feat was the use of Zn^{2+} in the protection of otherwise liable cysteines (Section 11.4.2.1), the protection/deprotection of reactive sites being common in synthetic organic chemistry, but not molecular biology. This approach made the production of the CysLite substrate tractable. We also explored the use of alternative labeling schemes in producing a double site specifically labeled substrate, which are challenging to produce, and therefore uncommon in the literature. The substrates reported therein, particularly the CysLite precursor, should be broadly applicable to research where site-specific labeling of RecQ is demanded.

12 Putative mechanism for the RecQ helicase

As a starting point, we hypothesize that RecQ operates via an inchworm or modified inchworm (rotating ratchet) mechanism, posses DNA binding sites akin to those proposed by the Keck[152] and Yang[132] labs and translocate/unwind DNA via iterative engagement/disengagement of the H1/H2 unwinding pins.

12.1 Direct determination of H1/H2 cleft movement

Using the double labeled H1/H2 double mutant, HisTEVRecQ CysLite N65C S245C (Section 11.4.3), we examine the cleft opening/closing of RecQ during its turnover cycle.

Both labeling reactions (**1:** Cy3B, 2:1, 1hr, R/T $\rightarrow$ Atto647N, 8:1, O/N, R/T; **2:** Atto647N, 2:1, 1hr, R/T $\rightarrow$ Cy3B, 8:1, O/N, R/T) yield the comparable labeling efficiency and stoichiometry, 1:0.74:0.61 and 1:0.64:0.61 RecQ:Cy3B:Atto647N respectively (Figure 12.1 (a)). Both labeling reactions yield comparable basal ATPase and DNA-stimulated ATPase activity (Figure 12.1 (b)). Curiously, labeling Rxn1 (Cy3B $\rightarrow$ Atto647N) yields an apo high-Förster resonance energy transfer (FRET) state, while labeling Rxn2 yields an apo low-FRET (Figure 12.1 (c)). This could be due to the relative position of the dyes artificially biasing the Rxn1 labeled mutant towards a closed conformation, and/or the Rxn2 labeled mutant towards an open conformation. As the fluorescent dyes themselves could bias the underlying open/closed distribution via a variety of mechanisms, it is impossible to determine the true native distribution. We will therefore focus on changes in the open/closed distribution. At the ensemble, changes in the underlying distribution will be obfuscated, resulting in changes in overall apparent FRET. At the single-molecule level, it should be possible to tease apart the change in distribution more precisely; that is, answer the question of whether intermediate FRET values at the ensemble are

true intermediates, or mixtures of binary open/closed states.

12.1.1 Ensemble fluorescence

Given the apo FRET states for Rxn1 (apo high FRET) and Rxn2 (apo low FRET), we set out to determine how the addition of analytes would change the apparent FRET signal. Although it is impossible to determine whether any intermediate apparent FRET values are mixtures of binary open/closed states or true intermediates, we take the former as a plausible simplifying assumption until more advanced single-molecule techniques can be brought to bear. With the double labeled RecQ mutant (HisTEVRecQ CysLite N65C S245C, Cy3B_Atto647N), we address a number of previously inaccessible questions, including:

(i) Does the binding of dT_{54} single-stranded DNA (ssDNA) induce a cleft opening/closure?
(ii) Does adenosine triphosphate (ATP) induce a cleft opening/closure?
(iii) Given that ATP is hydrolysable by RecQ, and that the enzyme presumably changes conformation to couple the energy release of ATP hydrolysis to translocation/unwinding, what can analogues (e.g. AMP-PNP / ATP-bound state, ADP-AlF_4 / ADP•P_i) reveal about the conformational state of RecQ during the catalytic cycle?

Ensemble fluorometer results for Rxn1/apo- high FRET/apo- closed state and Rxn2/apo- low FRET/apo- open state are depicted in Figure 12.2 and Figure 12.3 respectively. Panel A of both figures shows RecQ, RecQ + dT_{54}, RecQ + ATP and RecQ + AMP-PNP, all under saturating conditions. AMP-PNP, as a non-hydrolysable ATP analogue, is generally thought to mimic the ATP-bound state pre-hydrolysis.

For Rxn1 labeled RecQ mutant, none of the aforementioned analytes causes a change in apparent FRET from the apo- high FRET state (Figure 12.2 (a)). For Rxn2 labeled RecQ mutant, the addition of ATP causes a dramatic increase in FRET efficiency, while dT_{54} and AMP-PNP yield no change from the apo- low FRET state (Figure 12.3 (a)). These results suggest that ssDNA and ATP-binding (AMP-PNP, pre-hydrolysis ATP binding) dot not induce a change in cleft conformation. They also suggest that ATP induces a cleft closing at some point in the catalytic cycle.

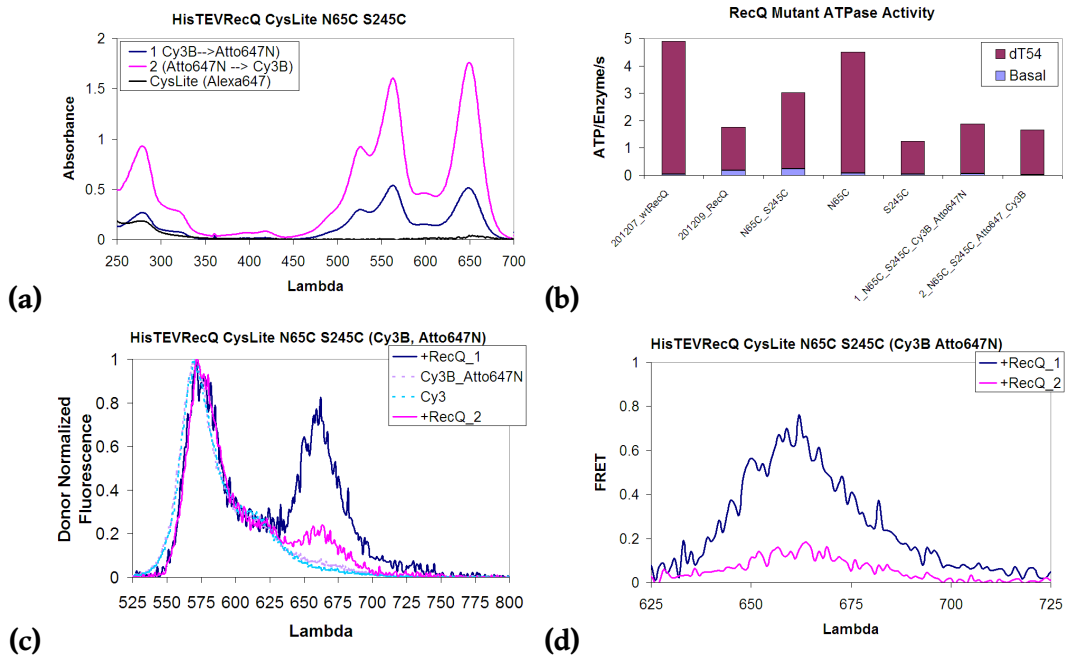


Figure 12.1 Double labeling of HisTEVRecQ CysLite N65C-S245C
(a) Absorbance Rxn1 HisTEVRecQ CysLite N65C S245C, Cy3B → Atto647N (**blue**), Rxn2 HisTEVRecQ CysLite N65C S245C,Atto647N → Cy3B (**magenta**) and HisThromRecQ CysLite, Alexa647 (**black**).
Labeling stoichiometry based on A_{280}:A_{560}:A_{655} and the extinction coefficients of RecQ ($\varepsilon_{280}^{\text{RecQ}} = 48\,820\ \text{M}^{-1}\,\text{cm}^{-1}$), Cy3B ($\varepsilon_{560}^{\text{Cy3B}} = 130\,000\ \text{M}^{-1}\,\text{cm}^{-1}$), Atto647N ($\varepsilon_{655}^{\text{Atto647N}} = 150\,000\ \text{M}^{-1}\,\text{cm}^{-1}$) and Alexa647 ($\varepsilon_{649}^{\text{Alexa647}} = 239\,000\ \text{M}^{-1}\,\text{cm}^{-1}$) were:

	RecQ	Cy3B	Atto647N
Rxn 1, Cy3B → Atto647N	1	0.74	0.61
Rxn2, Atto647N → Cy3B	1	0.64	0.61

Note that the labeling efficiency and stoichiometry of both labeling reactions are comparable. Alexa647 labeling of the CysLite mutant was less than 2 %.
(b) ATPase Mutant basal (400 nM) and dT_{54} DNA-stimulated (1 µM) ATPase activity. Note that both double mutants are active and show ∼20 fold DNA stimulation.
(c) Ensemble Fluorescence Ex500 fluorescence spectra for both Cy3B → Atto647N (Rxn1, **blue**) and Atto647N → Cy3B (Rxn2, **magenta**) with appropriate controls. RecQ mutant (500 nM), dT_{54} (10 µM, data not shown) and adenylylimidodiphosphate (AMP-PNP) (1 mM, data not shown).
(d) Ensemble FRET for Cy3B → Atto647N (Rxn1, **blue**) and Atto647N → Cy3B (Rxn2, **magenta**) depicted by subtracting the contribution of Cy3B and Atto647N free dye. Note that the reference curves were shifted before subtracting to account for the slightly different Stokes shift of free dye versus protein conjugated dye. Note that despite comparable labeling stoichiometry, (**Rxn1**) yields a high FRET state, while (**Rxn2**) yields a low FRET state. Reaction conditions were identical except for the order of dye addition.

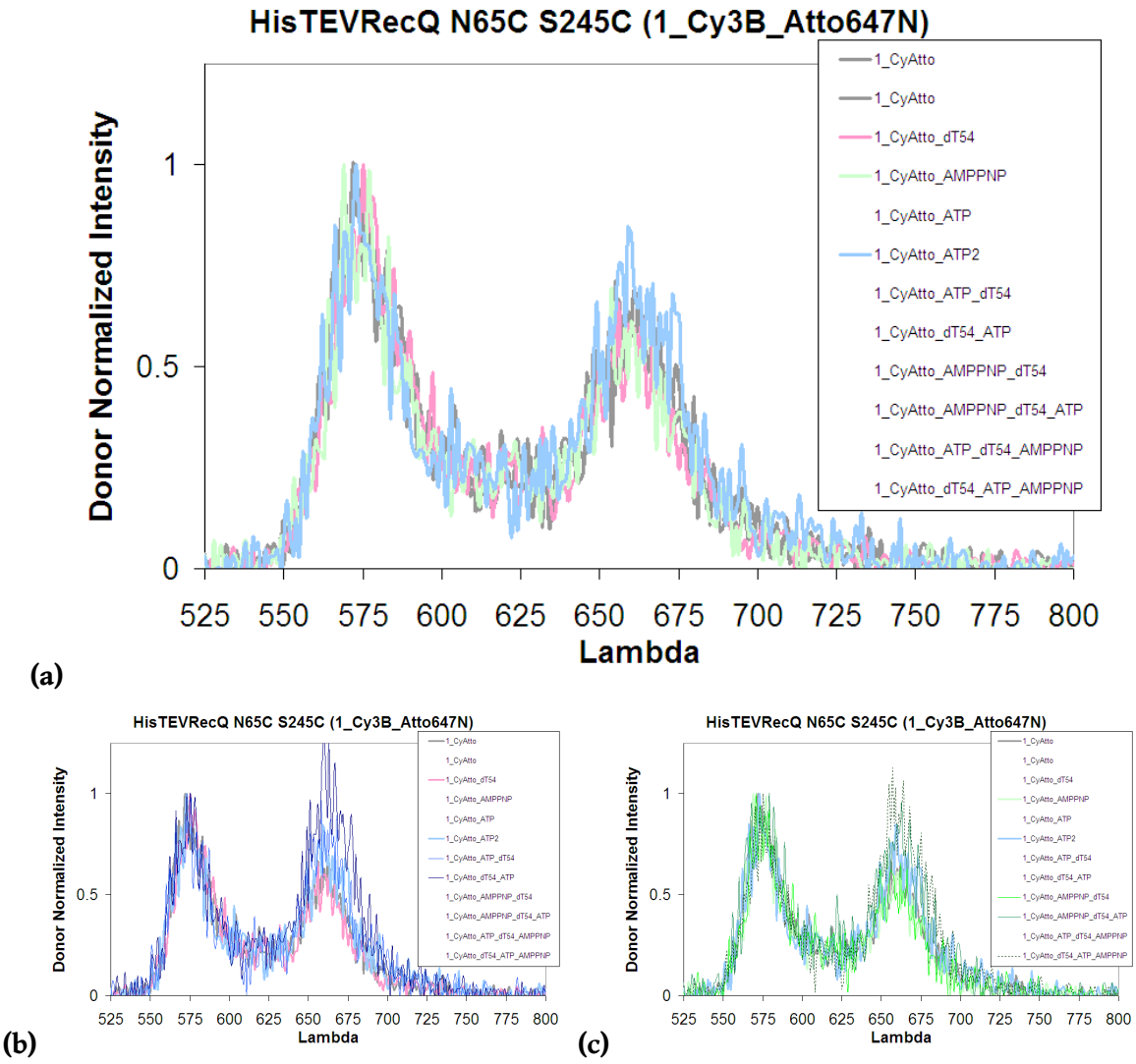


Figure 12.2 Minimal change in FRET for Rxn1 labeled RecQ mutant
(a) The addition of 5 mM dT_{54} ssDNA (**magenta**), 5 mM ATP (**blue**) or 400 µM AMP-PNP (**green**) to 200 nM HisTEVRecQ CysLite N65C S245C Rxn1 labeling, Cy3B → Atto647N (**gray**) does not significantly alter FRET.
(b) The addition of 5 mM dT_{54} ssDNA and 5 mM ATP may slightly increase FRET.
(c) The addition of 5 mM dT_{54} ssDNA and 400 µM AMP-PNP may slightly increase FRET.
All concentrations are final value in imaging buffer (Appendix B.7.2.1).

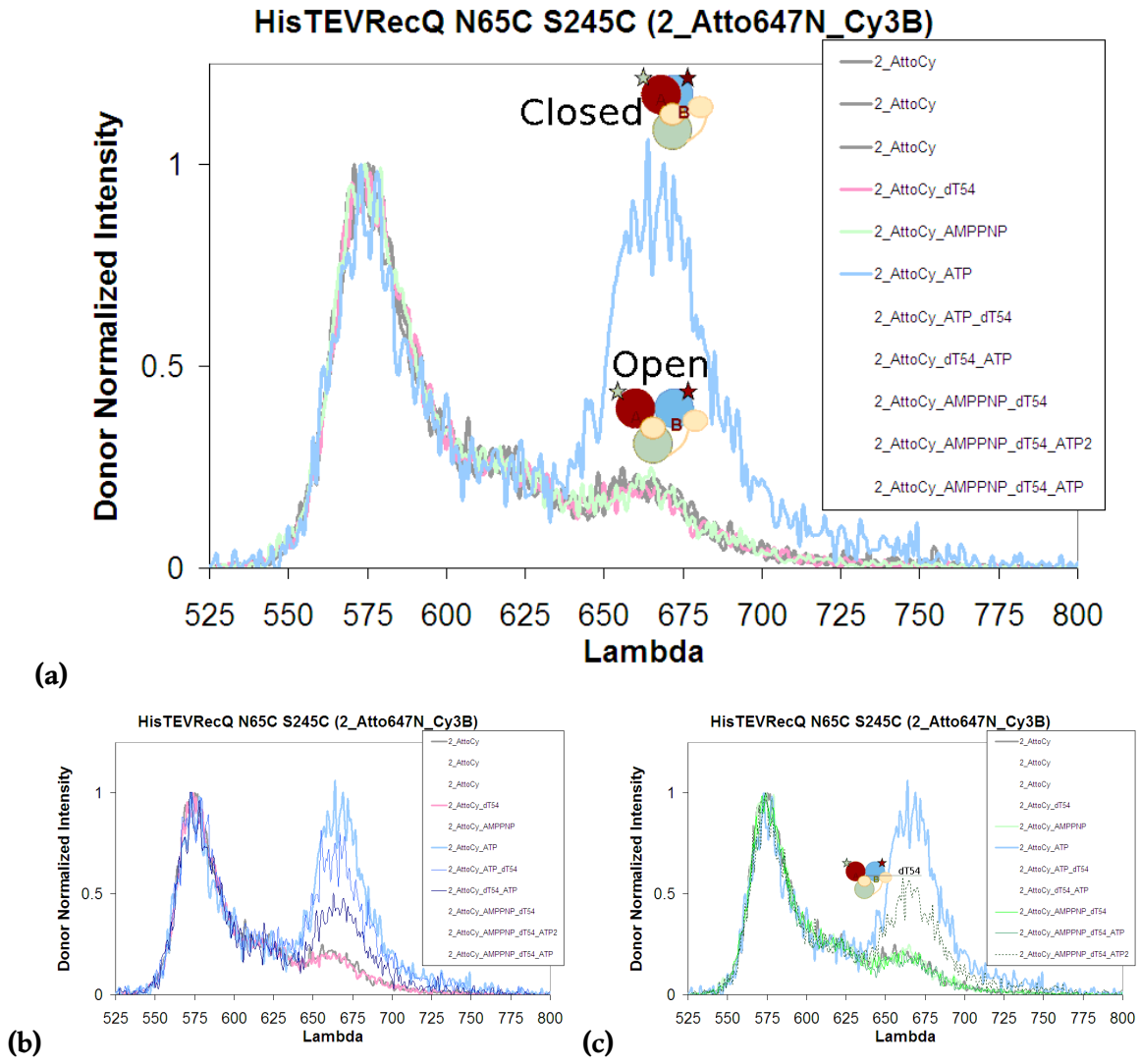


Figure 12.3 ATP causes change in FRET of Rxn2 labeled RecQ mutant
(a) The addition of 5 mM ATP (**blue**), but not 5 mM dT_{54} ssDNA (**magenta**) or 400 µM AMP-PNP (**green**), to 500 nM HisTEVRecQ CysLite N65C S245C Rxn2 labeling, Atto647N → Cy3B (**gray**) dramatically increases the FRET signal.
(b) The addition of 5 mM dT_{54} ssDNA and 5 mM ATP cause intermediate FRET values.
(c) The addition of 5 mM dT_{54} ssDNA and 400 µM AMP-PNP yield no change in FRET. An intermediate FRET value can be recovered after 5 minute incubation with a high 10 mM ATP concentration (**dashed green**).
All concentrations are final value in imaging buffer (Appendix B.7.2.1).

Next, we ask questions about H1/H2 cleft movement in the ssDNA bound state; specificially, whether ATP still induces cleft closing, and whether AMP-PNP has any affect on the cleft in the presence of ssDNA. Figure 12.2 (b) and Figure 12.3 (b) show RecQ and RecQ + dT_{54} + ATP for Rxn1 and Rxn2 double labeled RecQ mutants, respectively. For Rxn1, there is no dramatic change in FRET; for Rxn2, intermediate FRET values are observed.

12.2 Dynamics of DNA binding

12.2.1 Ensemble ATPase turnover

Following prior work,[133,153,159,200] we observe ssDNA stimulated enhancement of RecQ ATPase activity (Figure 12.4). Fortuitously, we find that there is no significant change in ATPase activity for the site-specific C→S substitutions. Further, changes in ATPase turnover upon labeling with Alexa647 are also present in the wildtype (wt_20120404), suggesting that this decrease is not due to the presence of the fluorophore. We also find that there is no significant difference in ATPase activity between full-length RecQ and truncated RecQΔHRDC (Section 11.5.1).

12.2.2 Ensemble binding assays

We report qualitatively similar dissociation constants (K_d) in the 100 nM range for both full-length wild-type RecQ (RecQwt) (Figure 12.5 (a,c,e)) and RecQΔHRDC (Figure 12.5 (b,d,f)), suggesting similar binding behavior. Curiously, while the reported dissociation constants are comparable, the Hill coefficient[i] differs between the protein induced fluorescent enhancement (PIFE) ($n = 2$) and anisotropy ($n = 1$) assays, under identical conditions. Hill coefficients of 1,[153] 2[203] and 3[204,205] have been reported in the literature.

We cannot explain these variations. They may simply be a consequence of overfitting; however, one intriguing possibility lies with the PIFE phenomena itself. It is conceivable that multiply binding sites encourage a robust PIFE response, while the anisotropy response is largely independent of the number of binding sites (unless each individual bound configuration is

[i] The Hill coefficient is a measure of binding cooperativity, and is given as n in the expression $\theta = \frac{[L]^n}{K_d + [L]^n}$, where θ is the fractional occupancy of the bound state, [L] is the ligand (DNA for our assays) and K_d is the dissociation constant.

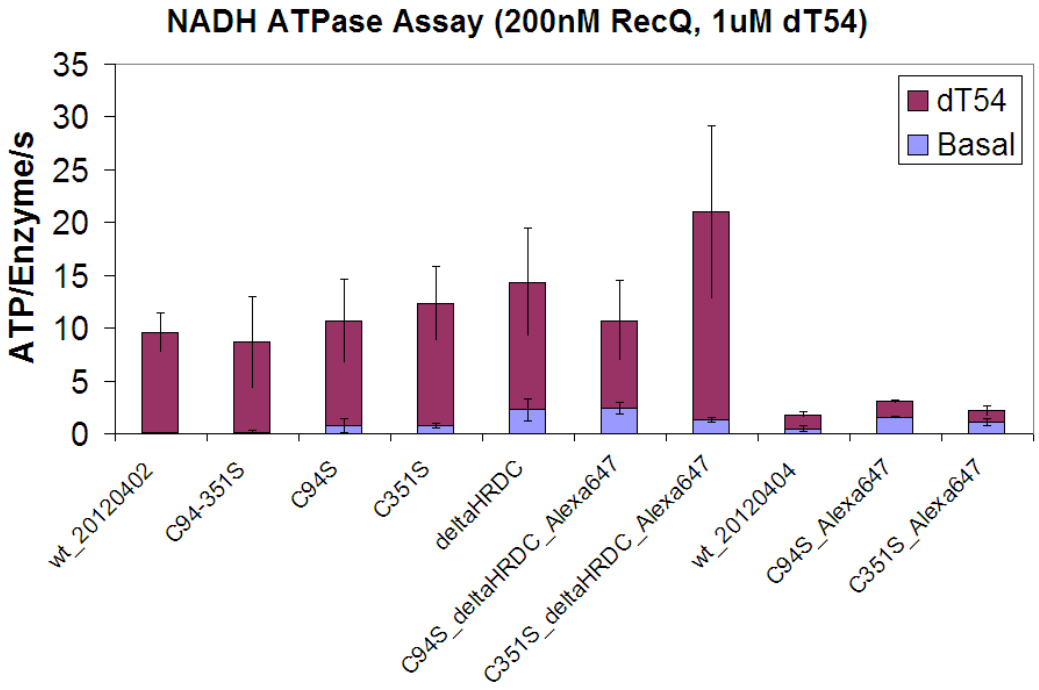


Figure 12.4 ATPase assay on CysLite and selected site-specific mutants
Basal (200 nM RecQ) and DNA-stimulated (1 µM dT_{54}) ATPase activity with NADH coupled assay. wt_20120402 and wt_20120404 are RecQwt. Mutants are HisThromRecQ unless otherwise stated. ΔHRDC mutants are generated with Thrombin, and therefore have their Thrombin-tags cleaved as well. wt_20120404, C94S_Alexa647 and C351S_Alexa647 received the same treatment (other than the dye addition), so wt_20120404 (as opposed to wt_20120402) is the appropriate internal control.

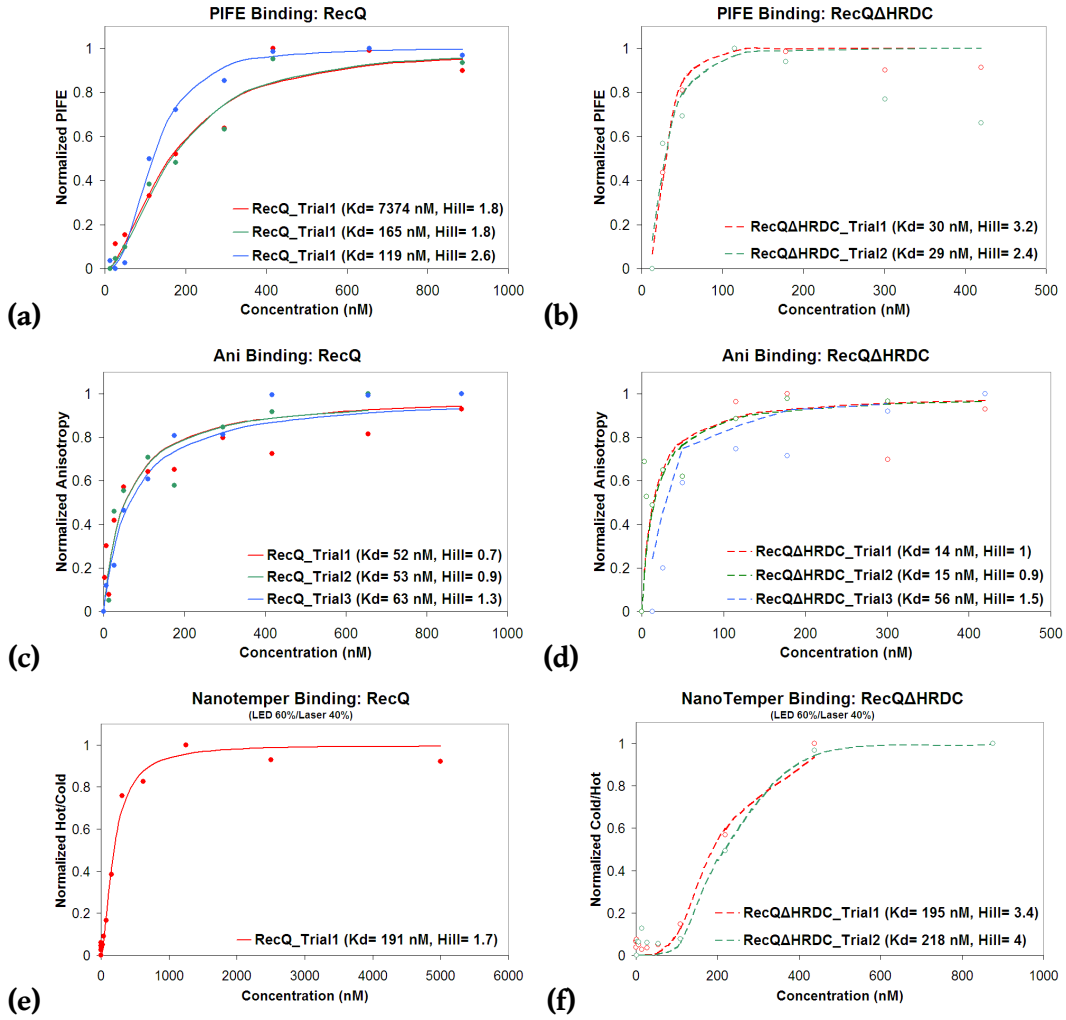


Figure 12.5 RecQ and RecQΔHRDC binding to 3′-duplex DNA
K_d determined by hill equation for PIFE, Anisotropy and NanoTemper based binding measurement with **(a)** 10–15 nM 25 nt / 25 bp 3′-duplex DNA (iCy3 at ss/dsDNA junction) on RecQwt and **(b)** RecQΔHRDC yielded a Hill coefficient of ∼2, while an anisotropy-based determination on **(c)** RecQwt and **(d)** RecQΔ HRDC under identical conditions yielded a Hill coefficient of ∼1. NanoTemper results for **(e)** RecQwt and **(f)** RecQΔHRDC for comparison.

quite short-lived). Given the Hill coefficient of $n = 1$ in the anisotropy assays, it is unlikely that cooperativity, if any, is due to RecQ multimers. Multimerization has been suggested in other experimental work;[203–205] however, anisotropy should be quite sensitive to low-order multimers (*e.g.* dimers, trimers and tetramers). It seems unlikely that such a short DNA substrate could accommodate more than one enzyme.[ii]

Taken together, we conclude that RecQ binding to DNA may be slightly cooperative. Importantly, any putative cooperativity is associated with the DNA-RecQ interaction, and not RecQ multimerization, supporting the view that RecQ is able to bind appropriate DNA substrates as a monomer.[153]

12.2.3 Single-molecule binding assays

Single-molecule binding dynamics experiments were conducted with non-specifically labeled Alexa647-maleimide RecQwt (10 nM and 50 nM) and Cy3 labeled 3′-duplex DNA (Figure 12.6 (d,e)) in the absence of ATP. We observe pulse-train-like single-molecule trajectories (Figure 12.7 (a,b)), expected to correspond to the association (bound state) and dissociation (unbound state) of labeled RecQ (acceptor, Alexa647-maliemide) to/from the 3′-duplex DNA junction (donor, IntCy3).

Analysis of these step-like time-series yields dwell-time distributions (Appendix B.5), corresponding to bound and unbound states at 10 nM and 50 nM RecQ (Figure 12.8). Double Gaussian fits on the dwell-time distribution were then used to segment two discrete FRET states. Bound and unbound time constants (τ^{bound}, τ^{unbound}) were determined by exponential fits to the E segmented dwell-time distributions with a 1 s imposition of resolution (Figure 12.9). Imposition of resolution was used to exclude short, likely spurious, events. Time constants are reported in Table 12.1.

For example, for the simplest system with a single bound and unbound state, rate constants are

[ii] Although, the substrate does not rule-out RecQ-RecQ interactions, gel filtration suggest that RecQ is indeed a monomer (Appendix B.1.2).

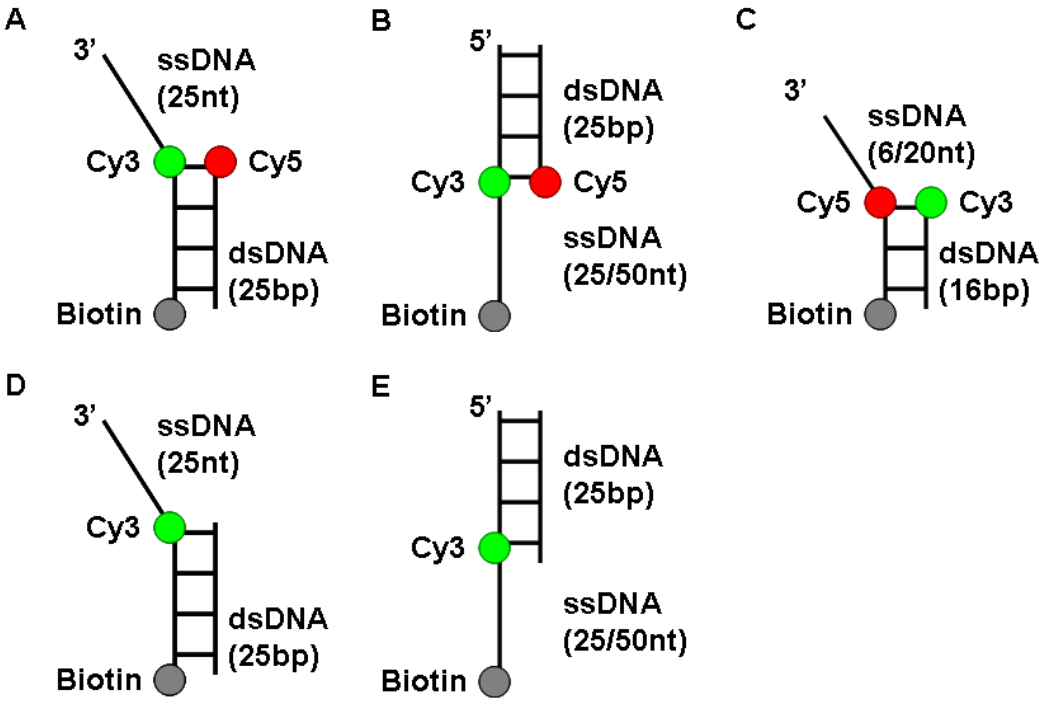


Figure 12.6 Labeled DNA schematic
(a) Cy3 / Cy5 5′-biotin 25 bp/25 nt 3′-duplex DNA (Cy3 Bottom 5′-biotin 50 nt / Cy5 Top)
(b) Cy3 / Cy5 3′-biotin 25 bp/25 nt 3′-duplex DNA (Cy3 Bottom 3′-biotin 50 nt / Cy5 Top) and 3′-biotin 25 bp/50nt 3′-duplex DNA (Cy3 Bottom 3′-biotin 75 nt / Cy5 Top)
(c) Substrate from Pan *et al.*[206]
(d) Cy3 5′-biotin 25 bp / 25 nt 3′-duplex DNA (Cy3 Bottom 5′-biotin 50 nt / Top Unlabeled)
(e) Cy3 3′-biotin 25 bp / 25 nt 3′-duplex DNA (Cy3 Bottom 3′-biotin 50 nt / Top Unlabeled) and 3′-biotin 25 bp / 50 nt 3′-duplex DNA (Cy3 Bottom 3′-biotin 75 / Top Unlabeled)
Note that all labels use an inline Int Cy3 or Cy5. Sequences available in Table B.3.

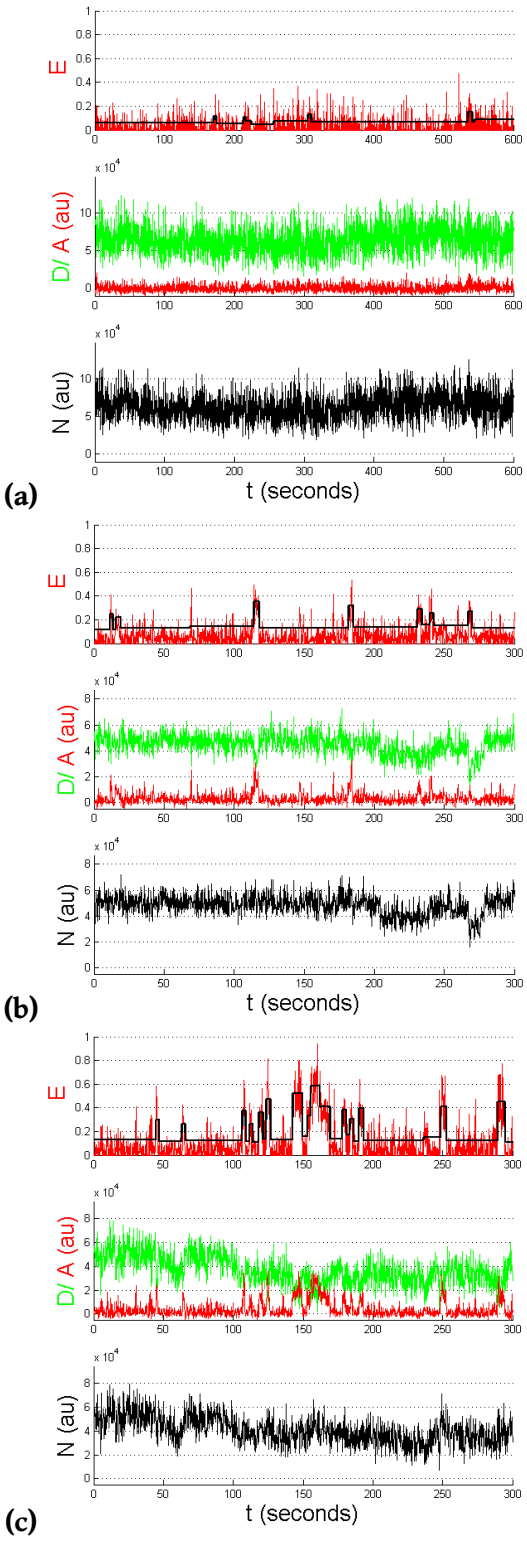


Figure 12.7 Single-molecule trajectories with donor labeled DNA
Results for non-site-specifically labeled RecQ.
(a) DNA Only
(b) 50 nM RecQ. Note the pulse train, particularly the anti-correlation in the donor and acceptor channels. This corresponds to acceptor labeled RecQ association/dissociation.
(c) 10 nM RecQ, 1 mM ATP. Note the unusually behavior, including *E*-values higher than those found without ATP.
Conditions: 5′-biotin 25 bp/25 nt 3′-duplex DNA (Figure 12.6 (a)) and standard imaging buffer (Appendix B.7.2.1); down-sampled from 100 ms to 500 ms for clarity; *E* panel fit with 1D Gaussian Kernel edge detector (Scale/0.6, Threshold/0.3). All *y*-axis scales in a.u..

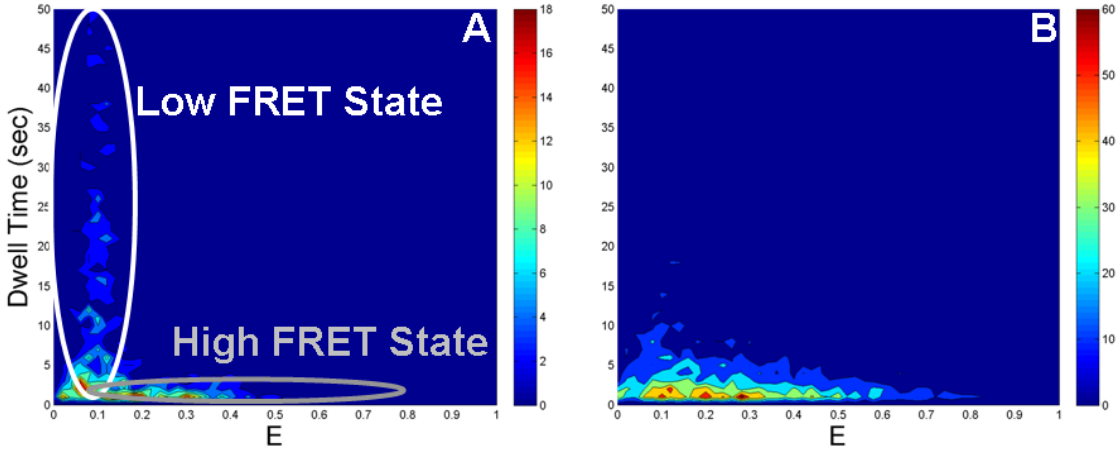


Figure 12.8 Dwell-time distribution for acceptor labeled RecQ and donor DNA
Results for non-site-specifically labeled RecQ.
(a) 10 nM RecQ
(b) 50 nM RecQ
Note how the total number of events increases at higher concentration. Also note that the low FRET state occupancy is decreased. Both experiments carried out at 3 mW 532 nm (donor) excitation with 100 ms integration time in standard imaging buffer (Appendix B.7.2.1).

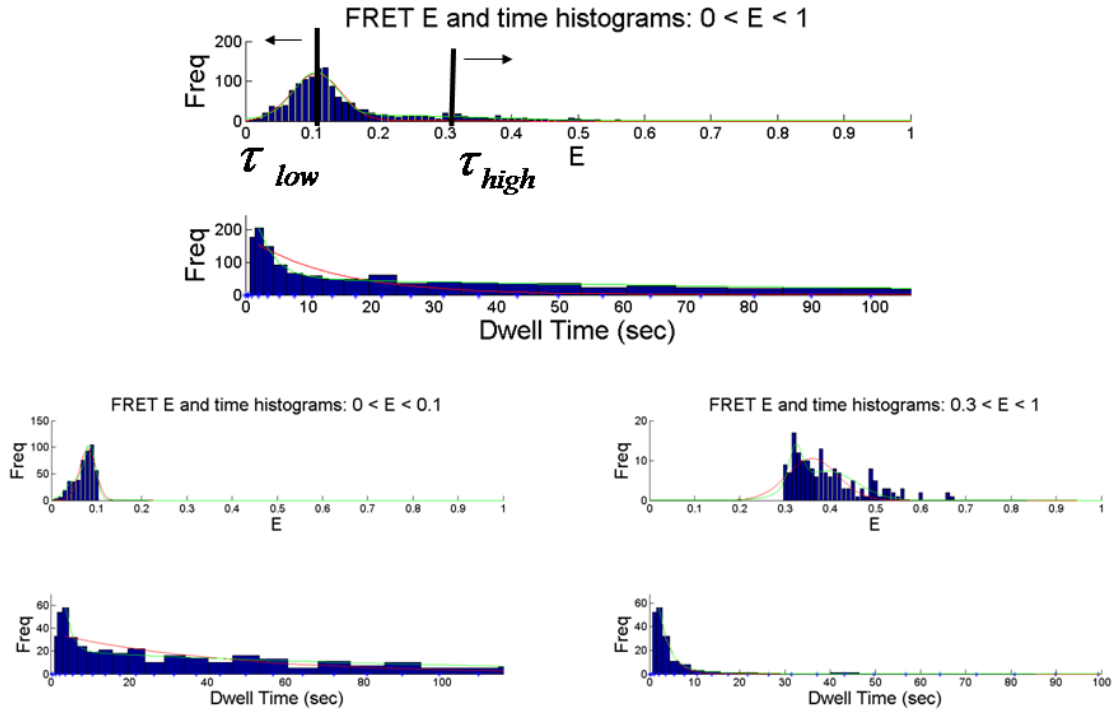


Figure 12.9 Exponential fits of dwell-time distribution
Single exponential/Gaussian (**red**) and Double exponential/Gaussian (**green**) fits for non-site-specifically labeled RecQ.
(a) The FRET distribution is fit to one or more Gaussian distributions. The centroid of these Gaussians defines the cut offs for τ_{low}, which corresponds to the unbound state, and τ_{high}, which corresponds to the bound state.
(b) The dwell-time distribution for the unbound state is fit to exponentials.
(c) The dwell-time distribution for the bound state is fit to exponentials.
The results of these exponential fits are depicted in Table 12.1

10 nM RecQ (Non-specifically labeled with Alexa647-maliemide)			
Method	τ^{unbound} / s	τ^{bound} / s	K_{eq}
Mean	46.5	3.3	0.07
Exp1	44.4	2.4	0.05
Exp2 Slow	105.7	11.2	0.11
Exp2 Fast	1.8	2.1	1.17

50 nM RecQ (Non-specifically labeled with Alexa647-maliemide)			
Method	τ^{unbound} / s	τ^{bound} / s	K_{eq}
Mean	22.8	7.3	0.32
Exp1	14.4	4.9	0.34
Exp2 Slow	62.4	11.1	0.18
Exp2 Fast	6.2	1.8	0.29

Table 12.1 RecQ dissociation constants from single-molecule prism-type TIRFM
τ^{unbound} corresponds to the unbound FRET state ($E < 0.1$); τ^{bound} corresponds to the bound FRET state ($E > 0.3$). $K_{\text{eq}} = \left(\tau^{\text{bound}}/\tau^{\text{unbound}}\right)$ is reported because K_{d} is model dependent.

computed from time constants as,

$$k_{\text{on}} = \frac{1}{\tau^{\text{bound}} \times [\text{RecQ}]} \left(\text{M}^{-1}\,\text{s}^{-1}\right), \tag{12.1}$$

$$k_{\text{off}} = \frac{1}{\tau^{\text{unbound}}} \left(\text{s}^{-1}\right), \tag{12.2}$$

where k_{on} and k_{off} are the forward (association) and reverse (dissociation) rate constants respectively, and τ^{bound} and τ^{unbound} are the characteristic lifetimes of the bound and unbound states from exponential fits.

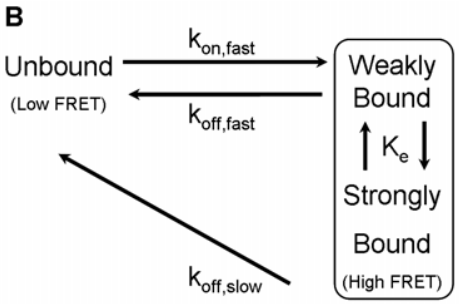
The dissociation equilibrium constant (K_{d}), in turn, can be computed from the rate constants,

$$K_{\text{d}} = \frac{k_{\text{off}}}{k_{\text{on}}} \left(\text{M}\right). \tag{12.3}$$

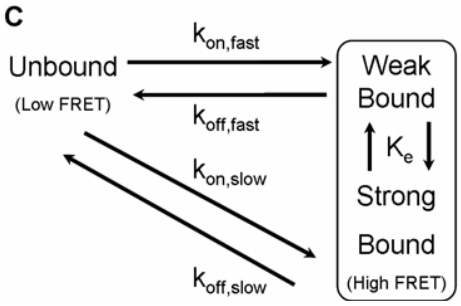
Analogous computations were carried out for the higher order binding models. Since both single and double exponential time constants were consistent with the data, a variety of binding models are plausible (Figure 12.10).

Our reported 'slow' and 'fast' time constants are separated by about one order of magnitude,

	k_{on}	k_{off}	K_d (nM)
10 nM RecQ			
Mean	2.2×10^6	0.30	140
Exp1	2.3×10^6	0.41	180
50 nM RecQ			
Mean	8.8×10^5	0.14	160
Exp1	1.4×10^6	0.2	150



	k_{on}	k_{off}	K_d (nM)
10 nM RecQ			
Exp1	2.3×10^6		
Slow		0.09	40
Fast		0.48	220
50 nM RecQ			
Exp1	1.4×10^6		
Slow		0.09	65
Fast		0.55	400



	k_{on}	k_{off}	K_d (nM)
10 nM RecQ			
Slow	9.5×10^5	0.09	95
Fast	5.5×10^7	0.48	9
50 nM RecQ			
Slow	3.2×10^5	0.09	280
Fast	3.2×10^6	0.55	170

Figure 12.10 Plausible binding models for RecQ/DNA interaction
Results for non-site-specifically labeled RecQ.
(a) Two-state association/dissociation.
(b) Single association rate, biphasic dissociation.
(c) Biphasic association and dissociation.

on the order of the qualitative detection limit. However, that the double exponential fit is significant over the single exponential fit to $p = 0.01$, suggests two-state binding behavior, as indicated in Figure 12.10 (c). As these results are a consequence of non-site-specifically labeled RecQ, and both C94 and C351 are known to be liable (Figure 11.10 (a)), we caution the reader against over-interpretation. We cannot rule-out the possibility that the two binding modes are a result of differentially labeled sub-populations.

Single-molecule dissociation constants were slightly higher than ensemble literature reports: $K_d = 18$ nM for a 16 bp/40 nt 3′-duplex DNA substrate,[133] and $K_d = 26$ nM for a 32 nt ssDNA.[205] We observe higher dissociation constants at higher RecQ concentration in double exponential models (Figure 12.10 (b,c). If true, this would suggest that there is binding inhibition at higher RecQ concentrations. Inhibition of unwinding initiation has been reported in the literature, adding to the plausibility of this conclusion.[206] Using single-molecule total internal reflection fluorescence microscopy (TIRFM) with double labeled DNA and unlabeled RecQ, they observe that: (i) longer tailed 3′-duplex substrates (16 bp/20 nt) take longer to initiate unwinding than shorter tailed substrates (16 bp/6 nt) and (ii) once unwinding begins, longer tailed substrates unwind faster than shorter tailed substrates. Furthermore, using multiple turnover ensemble kinetics on the same substrate (16 bp/20 nt), they observe two phase unwinding, with a fast and a slow phase. At low RecQ concentration (10 nM), fast phase unwinding prevails, while at higher RecQ concentration (50 nM) slow phase unwinding is more prevalent. Both of these observations could be explained by binding inhibition at higher RecQ concentrations. Note, however, that they use biotinylated Thrombin[153] in their RecQ preparation. As we have shown in Section 11.5.1, the use of Thrombin in the preparation of RecQ may actually yield a truncated RecQΔHRDC substrate, instead of RecQwt.

Overall, the results are surprisingly consistent considering the limitations of the single-molecule approach: (i) the input concentration likely differs from the experimental concentration at the surface of the flow cell due to nonspecific absorption of RecQ, (ii) no correction was made for the fact that there is a non-specifically labeled population of RecQ (some of which are unlabeled), (iii) the close proximity of the surface could impede RecQ binding and (iv) that the acceptor fluorophore could interfere with DNA binding.

DNA binding experiments were also conducted with site-specific Alexa647 labeled RecQ. Qualitatively, there was no difference between non-specifically and site-specifically labeled observations. There were insufficient traces for site-specifically labeled RecQ to draw quantitative conclusions.

12.3 Dynamics of translocation and unwinding

Two complementary approaches were used as a preliminarily assay of RecQ unwinding using single-molecule TIRFM. In one approach, we used acceptor/donor labeled 3′-duplex DNA and unlabeled RecQ; in the other, acceptor labeled RecQ and donor labeled 3′-duplex DNA. A double-labeled RecQ mutant has also been prepared for use in a third set of experiments aimed at directly probing inter-domain motions during translocation and unwinding.

12.3.1 Labeled DNA and unlabeled RecQ

It is difficult to distinguish between unwinding events and acceptor photobleaching in assays with acceptor/donor labeled DNA and unlabeled RecQ (Figure 12.11). Given our substrate design (Figure 12.6 (a,b)), one would expect a high FRET initial value ($E^{\mathrm{Pre}} \approx 0.9$), transitioning to a low FRET ($E^{\mathrm{Post}} \approx 0.1$) value as the acceptor containing strand is displaced by a RecQ helicase. This is clearly observed in a multiple turnover ensemble assay (Figure 12.12).

While donor photobleaching can be distinguished by the lack of recovery in donor signal, acceptor photobleaching cannot be so easily accounted for. One way around this problem would be the observation of steps, discrete FRET transitions as RecQ possessively unwound the DNA substrate. A variety of ATP concentrations were assayed, which based on available ensemble kinetics data, should have allowed for the detection of base-pair unwinding with the Neuman lab instrumentation (100 ms).

Rarely (13 of 314 trajectories), at 50 nM RecQ and 100 µM ATP, what appear to be steps are visible. The data are too sparse to accurately identify discrete E states via multi-Gaussian fitting of the FRET efficiency distribution (Appendix B.5.2), the technique used to extract binding kinetics. Another approach is therefore required.

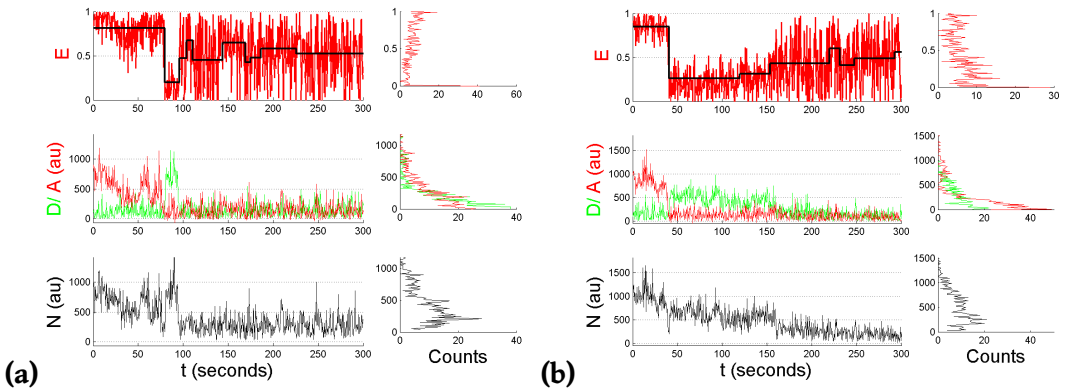
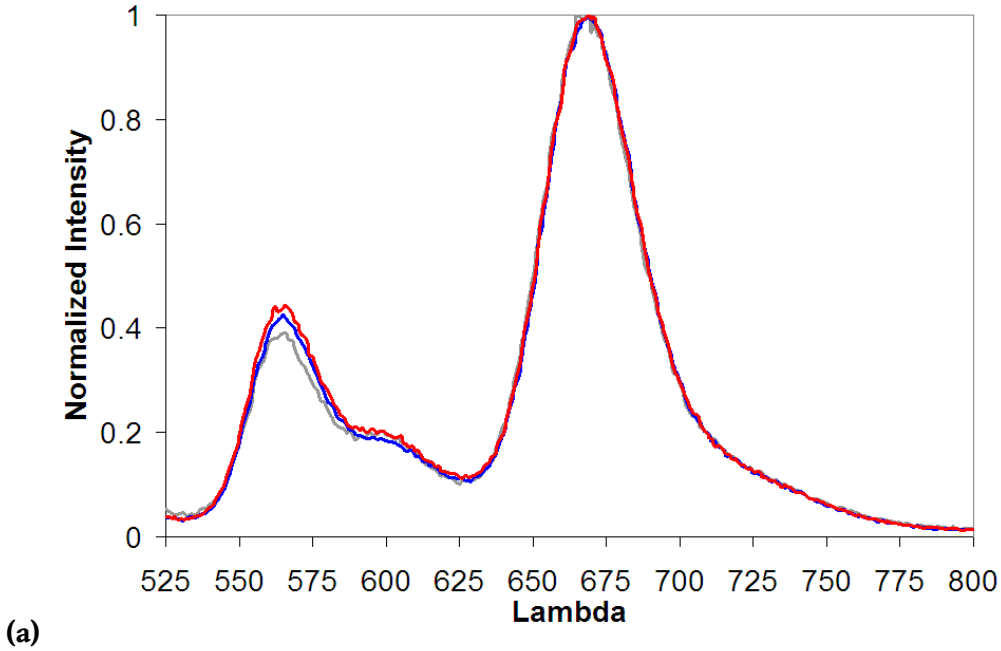


Figure 12.11 Single-molecule trajectories with double labeled DNA
(a) DNA Only
(b) 50 nM RecQ, 10 µM ATP with regeneration system.
Note the overall similarity of the traces. Conditions: 5′-biotin 25 bp/25 nt 3′-duplex DNA (Figure 12.6 (a)) and standard imaging buffer (Appendix B.7.2.1); down-sampled from 100 ms to 500 ms for clarity; *E* panel fit with 1D Gaussian Kernel edge detector (Scale/0.6, Threshold/0.3). All *y*-axis scales in a.u..

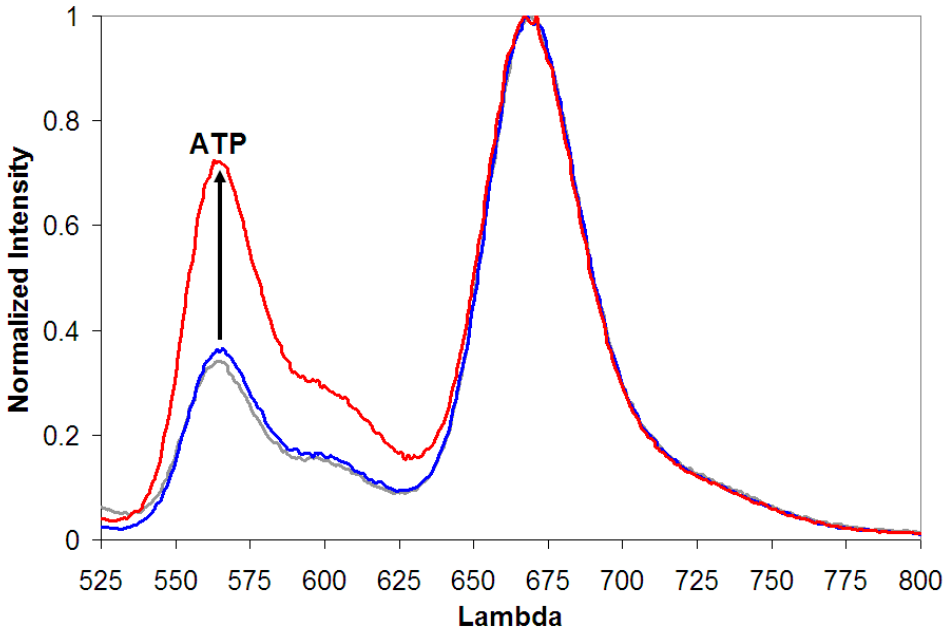
A variational Bayesian inference model, vbFRET,[iii] was used to fit the 13 stitched traces to a 4 state model (Figure 12.13 (a)).[207] Reproducible fits required down-sampling from 50–100 ms to 500 ms (ignoring fast transitions). Due to stitching, the lowest FRET state is an inaccurate representation of unwound DNA; occupancy and transitions out of this state are therefore ignored. Note however, that transitions from higher states into the unwound state are still valid. The transition matrix reveals that unwinding transitions occur in both directions; that is, while high to low *E* transitions are preferred, at all states there remains a non- negligible probability of transitioning to a higher FRET state (Figure 12.13 (b)). These data suggest that, while biased in the $3' \rightarrow 5'$ direction, RecQ does in fact translocate in reverse ($5' \rightarrow 3'$). However, given the rare occurrence of step-like traces and the now partially resolved problem of discriminating acceptor photobleaching, interpretation should be met with skepticism. While not yet attempted, slowly hydrolysable ATP analogues may reveal more stepped traces by slowing down RecQ at saturating ATP.

Another approach to side-step unwinding event detection versus acceptor photobleaching, useful in cases where RecQ unwinds DNA too quickly for steps to be observed, is to measure the slope of the transition from high to low FRET. Unfortunately, given the noise in the FRET trajectories ($\sigma \approx 0.1E$, $0 < E < 1$), there was no appreciable difference between the slopes of RecQ+DNA+ATP and a control of RecQ+DNA. Therefore, we have been unable to distinguish

[iii] vbFRET: `http://vbfret.sourceforge.net/`



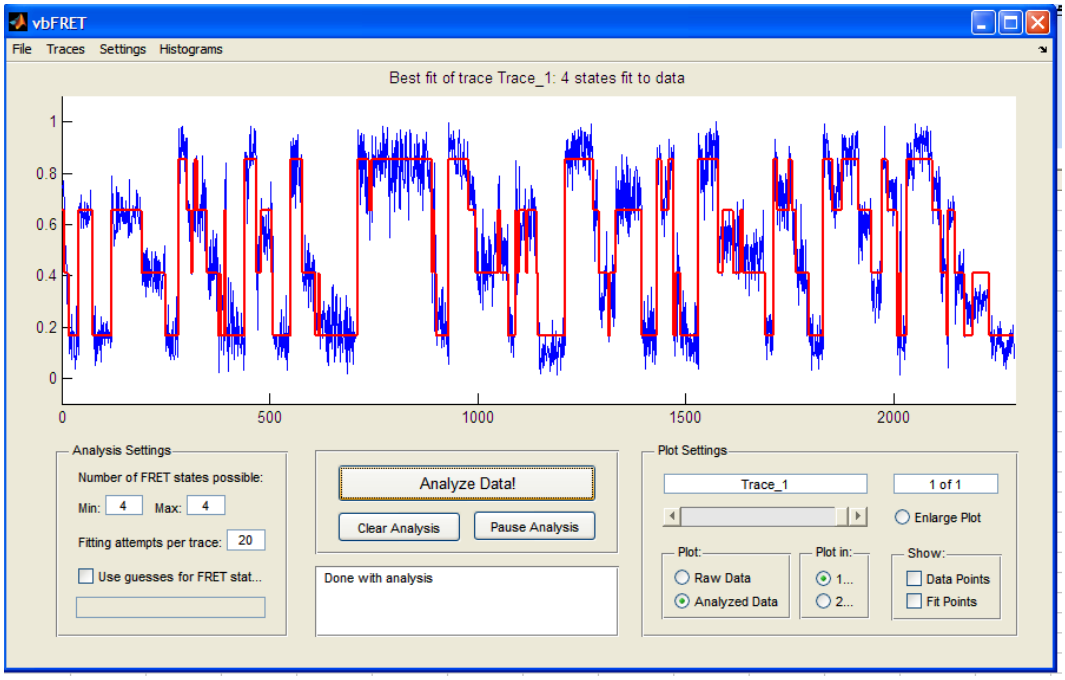
(a)



(b)

Figure 12.12 Ensemble FRET of steady-state, RecQ mediated DNA unwinding
DNA (**gray**), DNA+RecQ (**blue**), and DNA+RecQ+ATP (**red**). RecQ unwinds 3′-duplex DNA only in the presence of ATP. Steady-state unwinding in **(a)** imaging buffer and **(b)** imaging buffer supplemented with 3 mM $MgCl_2$.
Conditions are as follows: 15 nM DNA (15 nM 25 bp/25 nt 3′-duplex DNA: Cy3 Bottom, Cy5 Top), 30 nM RecQ, 75 µM ATP in imaging buffer (Appendix B.7.2.1).

(a)



(b)

Initial	**Occupancy**$^{\chi}$	**Forward** $(3' \rightarrow 5')$ **§**	**Reverse** $(5' \rightarrow 3')$ **§**	**Final** $^{\psi}$			
				0.16	0.41	0.65	0.86
0.16$^{\zeta}$							
0.41	0.30	0.61	0.39	0.61		0.33	0.06
0.65	0.33	0.77	0.23	0.13	0.65		0.22
0.86	0.38	1.00	0.00		0.19	0.81	

Figure 12.13 Analysis derived from vbFRET raw transition matrix
Results for non-site-specifically labeled RecQ.
(a) Screenshot of vbFRET depicting 13 stitched acceptor/donor labeled DNA, Unlabeled RecQ step traces.
(b) Data derived from vbFRET transition matrix. vbFRET outputs the raw transition matrix including self-to-self transitions, which are omitted.$^{\zeta}$
ψ For each initial FRET state, if a non-self transition occurs, what is the relative probability of transitioning into each state. Transition probabilities sum to unity row-wise.
χ Occupancy does not include the unwound (lowest initial FRET) state.
§ RecQ is assumed to move in the forward, $3' \rightarrow 5'$, direction if the FRET state transitions from higher to lower (unzipping of the DNA substrate); likewise, it is assumed to move in the reverse, $5' \rightarrow 3'$, if the FRET state transitions from lower to higher (rezipping of the DNA substrate). For visualization, refer to the substrate schematic depicted in Figure 11.14.
ζ Due to trajectory stitching (concatenating), the lowest initial FRET state ($E \approx 0.16$) is an inaccurate measure of the unwound condition. Occupancy and transitions out of this state are physically meaningless. Also note that the transition from $0.16 \rightarrow 0.86$ is a result of stitching.
See Bronson *et al.*[207] for details about vbFRET.

between acceptor photobleaching and RecQ mediated DNA unwinding at the single-molecule level.

Despite our difficulties, there is a single-molecule report of RecQ unwinding using a nearly identical substrate and setup (Figure 12.6 (a,c)).[206] They claim that photobleaching is negligible in their setup. While an oxygen scavenger system extends the life of fluorophores ~100-fold, in our hands, bleaching is still observed (Appendix B.4.1.1). Furthermore, they claim that the initiation time (the time until the beginning of the FRET transition), partial unwinding time (transition from high to low FRET) and unwinding time (disappearance of Cy3 signal, which is on the leaving strand) all decrease at increased ATP concentrations. We have been unable to reproduce these results, and suggest caution when drawing conclusions from their data.

12.3.2 Unwinding using smALEX-TIRF

By directly exciting and observing the acceptor fluorophore (A/A), alternating laser excitation (ALEX) circumvents the problems caused by indistinguishable acceptor photobleaching.[208,209] The Neuman lab prism-type TIRFM setup does not have this capability, but the Kapanidis lab objective-type TIRFM setup does. Experiments with double-labeled DNA and unlabeled protein were repeated with a more advanced ALEX smTIRFM setup. Signal loss in the acceptor fluorophore was attributable to acceptor photobleaching, not strand displacement by RecQ, again, leading us to question the results in reference 206, who without ALEX were unable to distinguish strand-displacement from acceptor photobleaching.

RecQ-mediated total unwinding of substrates on the surface was not demonstrated. This could either be due to surface effects or RecQ's inability to fully unwind substrates on the surface. The aforementioned results could be explained by partial unwinding, RecQ stalling, eventual dissociation and subsequent rezipping. Curiously, repetitive unwinding and rezipping have been observed at the single-molecule level for Bloom syndrome protein (BLM), a human RecQ homologue.[210]

12.3.3 Labeled DNA / non-specifically labeled RecQ

Another approach for observing unwinding dynamics is the use of donor labeled DNA (Figure 12.6 (d,e)) and acceptor labeled RecQ, in the presence of ATP. Unlike the labeled DNA / unlabeled RecQ approach, this technique is able to distinguish acceptor photobleaching from unwinding initiation. There are, however, multiple drawbacks: (i) RecQ dissociation and acceptor photobleaching cannot be distinguished, (ii) since the leaving strand is unlabeled, ssDNA cannot be distinguished from 3′-duplex DNA, and (iii) the maximal RecQ concentration is limited by background fluorescence. In the prism-type TIRFM setup, RecQ concentrations up to 50 nM were accessible, while 10 nM was the limit for the objective-type TIRFM setup. This is not surprising given the more precise tuning of the laser incident angle ($> \theta_c$ obviously) in prism-type setups, and the subsequent reduction possible in signal-to-noise ratio (SNR).

A small percentage of traces exhibited stair-case like behavior, presumed to be RecQ mediated unwinding/translocation of the substrate (Figure 12.7 (c)). As a first pass, discrete FRET states were determined by using a 1D Gaussian Kernel edge detector on multiple stitched traces exhibiting stair-case behavior. This approach proved untenable, as the peaks in E could not be deconvolved from the FRET efficiency distribution. Furthermore, even after parameter tuning (Appendix B.5.2), dwell-time fits varied considerably more than in the two-state static association/dissociation case.

12.3.4 Labeled DNA / site-specifically labeled RecQ

We suspected that interpretability would be improved with a site-specifically labeled RecQ. This approach has not been reported in the literature for RecQ, and is only possibly due to our substrate preparation efforts (Section 11.4.2). While qualitatively similar binding events were observed (Figure 12.14), there were insufficient traces for quantitative analysis. Labeled RecQ concentrations in excess of 10 nM were inaccessible due to increases background fluorescence on the objective-type TIRFM. Given a K_d for 3′-duplex DNA of 200 nM, we would expect roughly 1 % bound at any given time. Hampered by the small viewing area on the objective-type TIRFM setup (20–50 molecules per field of view) and inability to distinguish ssDNA and 3′-duplex DNA after ATP and RecQ are introduced (can only image one field of view per slide/well), it was

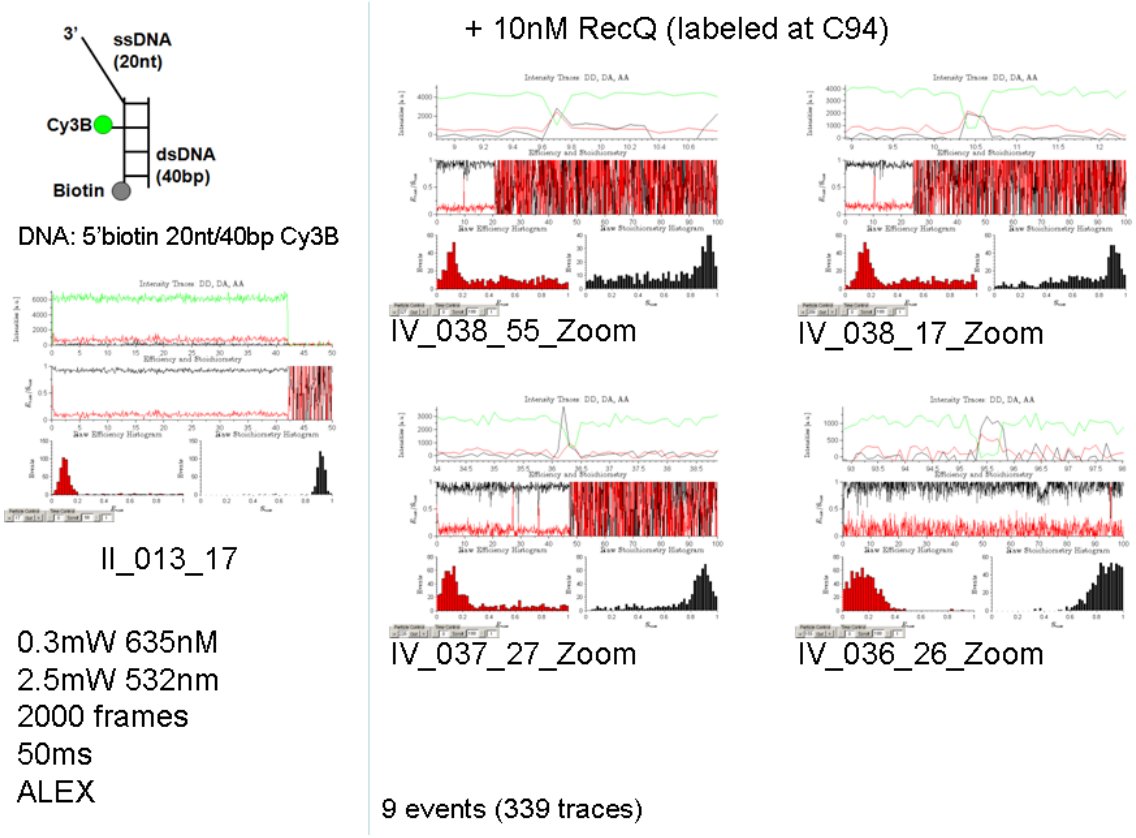


Figure 12.14 Site-specifically labeled RecQ with ALEX
Single-molecule trajectories with donor labeled DNA (Amino-dT near junction labeled with Cy3b and site-specifically labeled RecQ are qualitatively similar to non-specifically labeled. Insufficient traces for quantitative analysis.

considered infeasible to collect the number of events (∼1000) required for quantitative analysis.

12.4 RecQ functions via an inchworm mechanism

12.4.1 Comparison to ensemble kinetics data

Sarlós, Gyimesi, and Kovács[211] report the kinetics of the RecQ ATP hydrolysis cycle using AMP-PNP (ATP-bound), ADP•AlF_4 (ADP•P_i) and ADP, both apo and in the presence of ssDNA. Kinetic parameters are comparable between the apo and ssDNA bound for all analogue states except ADP•AlF_4, mimicking P_i release. Using quenched flow and assuming the irreversibility of P_i release, they found that P_i release preceded ∼150 times slower for apo than ssDNA bound

RecQ ($k_3 \approx 0.2\,s^{-1}$ versus $30\,s^{-1}$). Assuming that the intermediate states represent a mixture of open and closed conformation and not a true intermediate, and that the pre-hydrolysis ATP bound state (AMP-PNP) does not alone induce cleft opening, we propose that the ADP•P_i state corresponds to cleft closure. An intermediate FRET state, that is more open on average, is observed because given the increased rate of P_i release of the ssDNA bound RecQ, the enzyme spends less time in the closed state. If this is the case, the intermediate FRET state should be dependent on the dT_{54} concentration; furthermore, saturating ADP•AlF_4 (ADP•P_i mimic) should yield high-FRET in Rxn2, corresponding to a closed state. Figure 12.2 (c) and Figure 12.3 (c) show RecQ and RecQ + dT_{54} + AMP-PNP for Rxn1 and Rxn2 double labeled RecQ mutants, respectively. Like in panel A, dT_{54} + AMP-PNP do not cause a change in FRET for either substrate. This suggests that AMP-PNP (pre-hydrolysis ATP-binding) does not induce cleft closure.

Taken together, these data suggest the following about the nature of H1/H2 conformational changes:

(i) ssDNA binding does not mediate H1/H2 cleft movement (**1:** no FRET change, **2:** no FRET change).

(ii) ATP binding (presumably mimicked by AMP-PNP) does not induce cleft movement (**1:** no FRET change, **2:** no FRET change). This is corroborated by the fact that crystal structures of the apo and ATP-γ-S bound (a slowly hydrolysable ATP analogue similar to AMP-PNP) N-terminal RecQwt domains are nearly indistinguishable.[152]

(iii) ATP hydrolysis induces cleft closing (**1:** no FRET change, **2:** low-FRET to high FRET). In light of the kinetics of the ATP hydrolysis cycle revealing a ∼150 times slower P_i release in the absence of ssDNA, and the observation of apparent intermediate FRET values in the presence of ssDNA (faster P_i release resulting in less relative time spent in the ADP•P_i state), that hydrolysis and the subsequent ADP•P_i state induce cleft closure.

None of these observations have been previously reported in the literature.

12.4.2 Proposed model: Swap, Shuffle and Pull

Given these observations and the results of separation pin mutagenesis in the RecQ family helicases (discussed in Section 11.3, specifically Section 11.3.1), we propose a complete model

for RecQ translocation and unwinding with several specific predictions for RecQ function (Figure 12.15). The crux of the model is H1/H2 mediated inchworming in the $3' \rightarrow 5'$ direction. H1 domain binding to ssDNA is mediated by the H1-pin (W154, F158), H2 domain binding by the H2-pin (F221), and non-specific dsDNA binding by the winged-helix (WH) domain. The H1-pin and H2-pin alter their engagement to ssDNA, and their relative conformations (open versus closed) throughout the catalytic cycle. The engagement of either the H1-pin or H2-pin to ssDNA corresponds to a weakly bound state; the engagement of both the H1-pin and H2-pin to a strongly bound state. The presence of two distinct binding states is supported by smtotal internal reflection (TIR)-FRET binding studies and unpublished work by Sarlós, Gyimesi, and Kovács from their 2012 manuscript.[211]

$3'$-duplex DNA binds with the junction under the H1 domain, with $3'$ ssDNA threading around the helicase and RNaseD C-terminal (HRDC) domain in its ssDNA binding channel. The inchworm mechanism proceeds in three steps: Swap, Shuffle and Pull. First, RecQ binds ssDNA by engaging the H2-pin. Upon ATP binding, the H1-pin engages and the H2-pin disengages (Swap). The model does not predict a conformational change due to ATP binding; this is contrary to the commonly accepted inchworm[130] (or inchworm and ratchet) model which predicts a conformational change leading to base unwinding from ATP binding. Crystal structures of UvrD[130,132] (SF1) and NS3[212] (SF2) both support the view of an ATP induced conformational change.

Upon ATP hydrolysis, the H1/H2 cleft closes, resulting in the H2 domain being pulled towards the H1 domain; the H2-pin then reengages such that both the H1- and H2-pins are engaged (Shuffle). The enzyme remains locked in the strongly bound state until P_i release. Upon P_i release the H1-pin disengages and the H1/H2 cleft opens, resulting in the H2 domain pulling on the $3'$ end of the strand, in essence using the H1 domain as a wire-stripper for strand-displacement (Pull). No conformational change upon ADP release. The 'Swap' step was placed at ATP-binding, but the model still holds in its entirety if the Swap occurs during ADP release. The only difference would be that initial DNA binding would occur with the H1-pin (Swap during ADP release) instead of the H2-pin (Swap during ATP-binding). The enzyme is now reset, and ready for another Swap, Shuffle, Pull cycle.

Unwinding of simple substrates (forked and $3'$-duplex DNA) follows the same process as

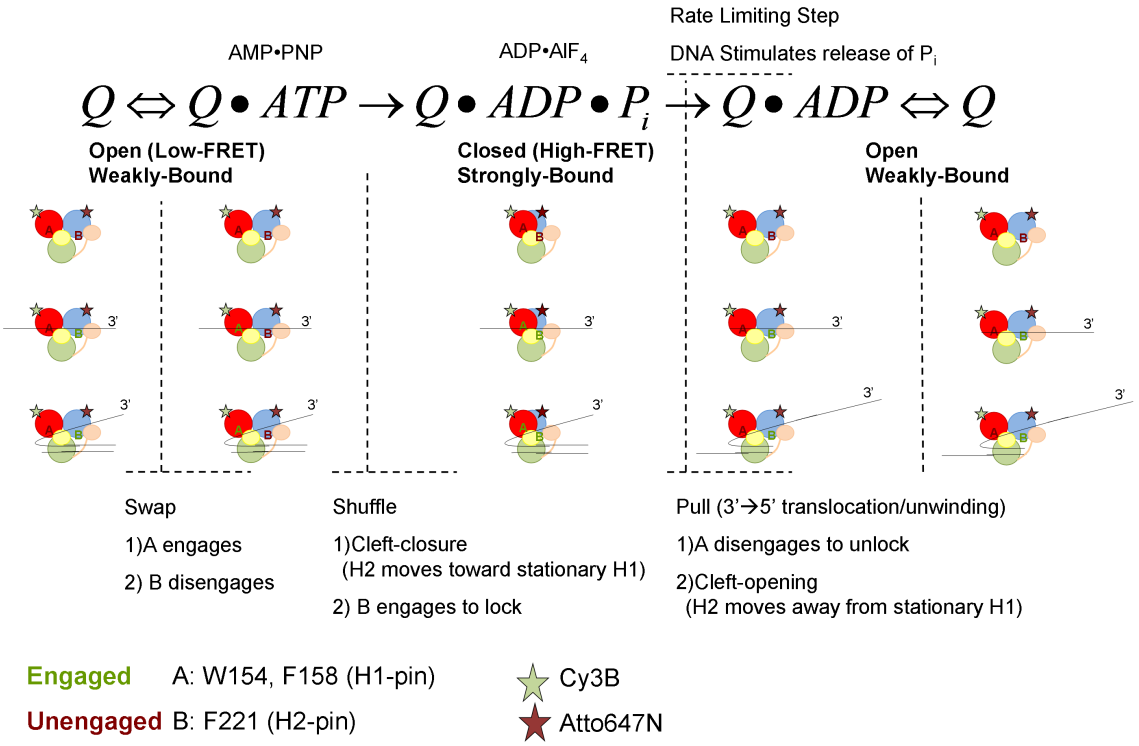


Figure 12.15 Model of RecQ translocation and unwinding
Swap, Shuffle and Pull model of RecQ translocation and unwinding.

RecQ binding DNA: H2-pin engages.

ATP-binding: Swap. H1-pin engages. H2-pin disengages.

ATP-hydrolysis: Cleft closure pulling the H2 domain towards the H1 domain. H1- and H2-pin engages until phosphate release.

Phosphate (P_i) release: H1-pin disengages. Cleft opens with H2 moving away from H1 domain. Since the H2-pin is still engaged, this pulls the 3′ ssDNA, effectively using the H1 domain as a wire stripper for strand-displacement.

Adenosine diphosphate (ADP) release: No conformational change.

Note that the Swap could occur at either ATP-binding or ADP release; ATP-binding was chosen for clarity.

translocation, and operates via forcible strand-displacement instead of a fixed unwinding pin. The WH-pin and Zn-pin may have ancillary usage in resolving more complex substrates, such as Holliday junctions, but are non-essential for simple substrates.

One of the distinguishing characteristics of the Swap, Shuffle, Pull model are its many specific and directly testable hypotheses. The ensemble characterization of the double-labeled RecQ mutant (HisTEVRecQ CysLite N65C S245C labeled with Cy3B-maliemide / Atto647N-maliemide) warrants its use in single-molecule assays to directly monitor the putative H1/H2 cleft movement, a direct test of the model. Explicitly probing the unwinding pin residues (W154L, F158L, F221L, H489L) through mutagenesis is also a possibility.

For example, if the H1-pin were eliminated by mutation to bulky non-aromatic residues (W154L, F158L), the H1-pin would be forced to remain in a disengaged state. This would eliminate the strongly-bound state, eliminate the Swap step, and result in the H2-domain repetitively shuffling and pulling without moving. P_i release would also be facilitated by the lack of a strongly bound state, meaning higher basal turnover. Initial DNA binding would be unaffected. Cleft opening/closure would be preserved, but would no longer be coupled to translocation/unwinding. Many of these predictions have already been borne out.[182]

13 Conclusion

Using novel engineered substrates in conjunction with advanced single-molecule fluorescence-based techniques, we probe the behavior of the RecQ DNA helicase.

Combining our experiments with previously reported ensemble work, we developed a model coupling the ATP catalytic cycle, specific H1/H2 conformational change and DNA binding. Specifically, we propose a mechanism for RecQ helicase activity: Swap, Shuffle and Pull (Section 12.4.2), whereby the enzyme sequentially engages and disengages two distinct DNA binding sites (also responsible for the biphasic binding behavior) to actively pull along the bound single-stranded DNA strand, akin to a wire stripper. This model supports the view that the translocation and unwinding modes are mechanistically identical.

To the best of the author's knowledge, a similarly detailed model linking the ATP catalytic cycle to specific conformational states associated with translocation/unwinding has not been proposed for a SF2 helicase. The model, and the double labeled RecQ mutant designed to directly report H1/H2 conformational changes, are novel and exciting additions to the understanding of SF2 helicases, and DNA motor proteins in general.

To undertake the our experiments, we have developed a family of previously unreported RecQ mutants, including single- and double- site-specifically labeled RecQ (Chapter 11). Using these substrates, we directly report that the relative movements of the two helicase domains (H1/H2 cleft movements with/without ATP and ATP-analogues) are indeed correlated with translocation and unwinding. We find that RecQ likely posses biphasic binding, where there are two distinct binding/unbinding modes (a strong and a weak), separated by an order of magnitude in time scale. We have also observed putative RecQ unwinding with labeled RecQ and labeled DNA at the single-molecule level, finding that while RecQ predominantly translocates and unwinds in the

$3' \rightarrow 5'$ direction, the enzyme occasionally reverses, a process dubbed slippage.

13.1 Future Directions

A better understanding of site-specific and double RecQ mutants, and a detailed model, should lay the foundation for more detailed single-molecule TIRF and confocal experiments. Noting that these aims are independent, one could proceed as follows:

(i) smFRET with site-specifically labeled RecQ and DNA. Given the kinetics ($K_d \approx 200\,\mathrm{nM}$ for $3'$-duplex substrates) of RecQ-DNA complexes, monitoring bimolecular events with labeled RecQ and labeled DNA free in solution, is untenable. If it is possible to increase the local concentration without increasing background, vesicle encapsulation for example, it should be possible to monitor trajectories of labeled RecQ translocation/unwinding labeled DNA.

(ii) smFRET with double site-specifically labeled RecQ to examine suspected inter-domain motions. Good places to begin would be relative domain motions of H1/H2 (mutant available), H1/winged-helix (WH) (not available) and H1/helicase and RNaseD C-terminal (HRDC) (not available). Confocal measurements would be ideal for determining the cleft position of each state of the catalytic cycle. smTIR measurements would be ideal for directly observing the transitions between states. If the cleft motion is sub-nanometer (it doesn't appear to be from the ensemble experiments of the double-labeled mutant) instead of using traditional pairs of synthesis dyes, transition-metal[139,140] Förster resonance energy transfer (FRET) or dye-quencher[213] FRET could be attempted.

None of the aforementioned aims have been reported in the literature because the substrates to conduct them did not exist. We believe that individually, they should be attainable within the near-term (2 years).

A future direction, which may be further afield, is:

(i) Using the site-specifically labeled RecQ mutants, particularly the double labeled mutant, in a combined magnetic tweezers / TIRF. This would allow direct corroboration of stepping (H1/H2) and unwinding.

This sort of combined experiment is difficult, and would require an intermediate-time scale (2–5 years).

Future directions, which in the author's opinion are of lesser importance, include:

(i) Assign the location and orientation of the putative ss/dsDNA binding sites. Using the family of site-specific RecQ mutants and appropriate DNA substrates, it should be possible to generate a list of constraints for modeling the DNA binding channel(s) of RecQ. A similar approach has been used to model the binding channel, and infer changes in the protein conformation during ATP hydrolysis using several ATP-analogues, with the *E. coli* Rep helicase using a site-specifically labeled enzyme and junction labeled DNA.[214]

(ii) smFRET with unlabeled RecQ. Having considered the possibility that unwinding is too fast to observe at 50–100 ms resolution, it may be possible to slow down the system using non-hydrolysable analogues such as ATP-γ-S. The hope is that either unwinding slopes will become better resolved to allow for the detection of individual steps. Given that single-molecule FRET with double labeled DNA and unlabeled RecQ has already been reported, this approach is less interesting than working with the newly reported site-specifically labeled RecQ.

(iii) Experiments with site-specifically labeled UvrD. Given the time and effort required to generate suitable mutants for RecQ, it cannot be expected that creating a CysLite UvrD and subsequence site-specific mutants will be trivial. A labeled transglutaminase (TGase)-UvrD would be significantly easier to achieve, but also limited by the intrinsic flexibility of the TGase tag. Due to this intrinsic flexibility, TGase-labeled substrates are likely unsuitable for studies of relatively small inter-domain motions.

These aims have either been partially reported in the literature[125] or are an application of the lessons learned in this thesis to a related system.[150] They are, in the author's opinion, about as difficult as the near-term aims but with significantly less impact.

Part III Technical Appendices

Introduction

These technical appendices include supplementary materials, mainly methods, designed to (i) aid reproducibility by serving as a technical reference document and (ii) address technical caveats that are not essential the central line of inquiry, but might nonetheless be of interest to a particularly careful reader. Unlike the main corpus of the thesis, where we do not necessarily assume familiarity with the simulation and experimental methods employed, the technical appendices, by their very nature, demand a fairly strong background. For ease of use, these technical appendices are organized by part, with appendix Appendix A referring only to Part I, and appendix Appendix B referring only to Part II.

A Part I: Detailed simulation methods

A.1 Thermodynamics of DNA bending

A.1.1 Bending simulations

Bending simulations were performed in three phases: exploratory, equilibration and production. In the exploratory phase, we iteratively adjusted the umbrella sampling weights to yield a flat probability distribution $P(R'_{ee})$, where R'_{ee} is the end-to-end distance (EED) between the base interaction sites of the first and last bases on a single strand using a discrete potential with increments of 0.4259 nm. The duplexes were constrained not to fray (Appendix A.1.2), allowing us to sample in terms of R_{ee}, the EED between the centres of the helical axis at the first and last base pairs, choosing an R_{ee} bin-width of 0.085 nm.[i] Non-native base-pairing was also forbidden. As the iteration was performed manually, simulation times varied in the range 10^6–10^7 virtual-move Monte Carlo (VMMC) steps per particle.

In the equilibration phase, we equilibrated the system for 10^7 VMMC steps per particle using the aforementioned umbrella weights. Adequate sampling (number of transitions between EED bins) and decorrelation (via a block average decorrelation method) of potential energy, bubble size, fraying (where applicable) and structural kinking were checked. For the production phase, each production simulation (five independent simulations per measurement) was initialized with a statistically independent starting configuration and a unique random seed. This is accomplished by randomly drawing starting configurations from the equilibration phase, one for each production run, ignoring the first 2×10^6 VMMC steps per particle of equilibration.

[i] While the difference between the terminal base-base (umbrella sampled distribution) and terminal helical axis (reported distribution) definitions of R_{ee} are negligible, defining R_{ee} in terms of the latter explicitly disallows geometric effects that might have been present as a consequence of umbrella sampling the EED of only one strand. It also allows for re-binning R_{ee} without re-running the simulation.

The production run time for each simulation was 10^8 VMMC steps per particle. For reference, the characteristic decorrelation time for the potential energy is $\sim 10^4$ VMMC steps per particle, while the decorrelation time for kinking is $\sim 10^5$ VMMC steps per particle. The 'seed moves' used to build clusters in the VMMC algorithm were:

- Rotation of a nucleotide about its backbone site, with an axis chosen uniformly on the unit sphere, and with an angle drawn from a normal distribution with a mean of zero and a standard deviation of 0.10 radians.
- Translation of a nucleotide, where the displacement along each Cartesian axis is drawn from a normal distribution with a mean of zero and a standard deviation of 0.085 18 nm.

Additional simulations were performed to measure the free energy of kinked states for a 30 bp duplex at larger values of R_{ee}. In this case, as well as biasing R_{ee}, the umbrella potential was used to forbid states without kinks (separate simulations were performed for each kink criterion). Umbrella potentials for R_{ee} were obtained via manual iteration at a preliminary stage. Simulations were started from kinked configurations and thermalized for 3.3×10^6 VMMC steps per particle. Subsequent data collection lasted for 3.3×10^8 VMMC steps per particle. Five independent simulations were performed. The VMMC 'seed moves' differ slightly from those mentioned previously:

- Rotation of a nucleotide about its backbone site, with the axis chosen from a uniform random distribution and the angle from a normal distribution with mean of zero and a standard deviation of 0.15 radians;
- Translation of a nucleotide with the direction chosen from a uniform random distribution and the distance from a normal distribution with mean of zero and a standard deviation of 1.28 Å.

A.1.2 Constraints on fraying

To isolate the effects of DNA bending, as opposed to alternative deformation modes such as fraying of the terminal base pairs, we impose a condition preventing fraying in our DNA bending simulations. Using an additional order parameter, the 30 bp system was constrained such that the terminal base pairs must remained hydrogen bonded, thereby eliminating the possibility

of fraying at duplex ends. We compared constraints where 1, 3 and 5 base pairs at the duplex ends were prevented from opening, finding that the various conditions yield comparable results (Figure A.1).

To further probe the behavior of the 30 bp system, in separate simulations we allowed the system to fray freely or preserved hydrogen bonding along the entire duplex. Unsurprisingly, allowing fraying as an additional deformation mode reduces the bending free-energy penalty in the strong bending regime, while preserving hydrogen bonding increases the bending free-energy penalty by eliminating kinked states that require disrupted base pairing, thereby reducing the number of possible deformation modes.

Since the system where fraying is allowed may not have terminal base pairs, the usual definition of R_{ee} reporting the separation between the helical axis centres of the terminal base pairs, is not well defined. Therefore, for this system we instead use R'_{ee}, defined as the separation between the base interaction sites of the first and last bases along a single strand.

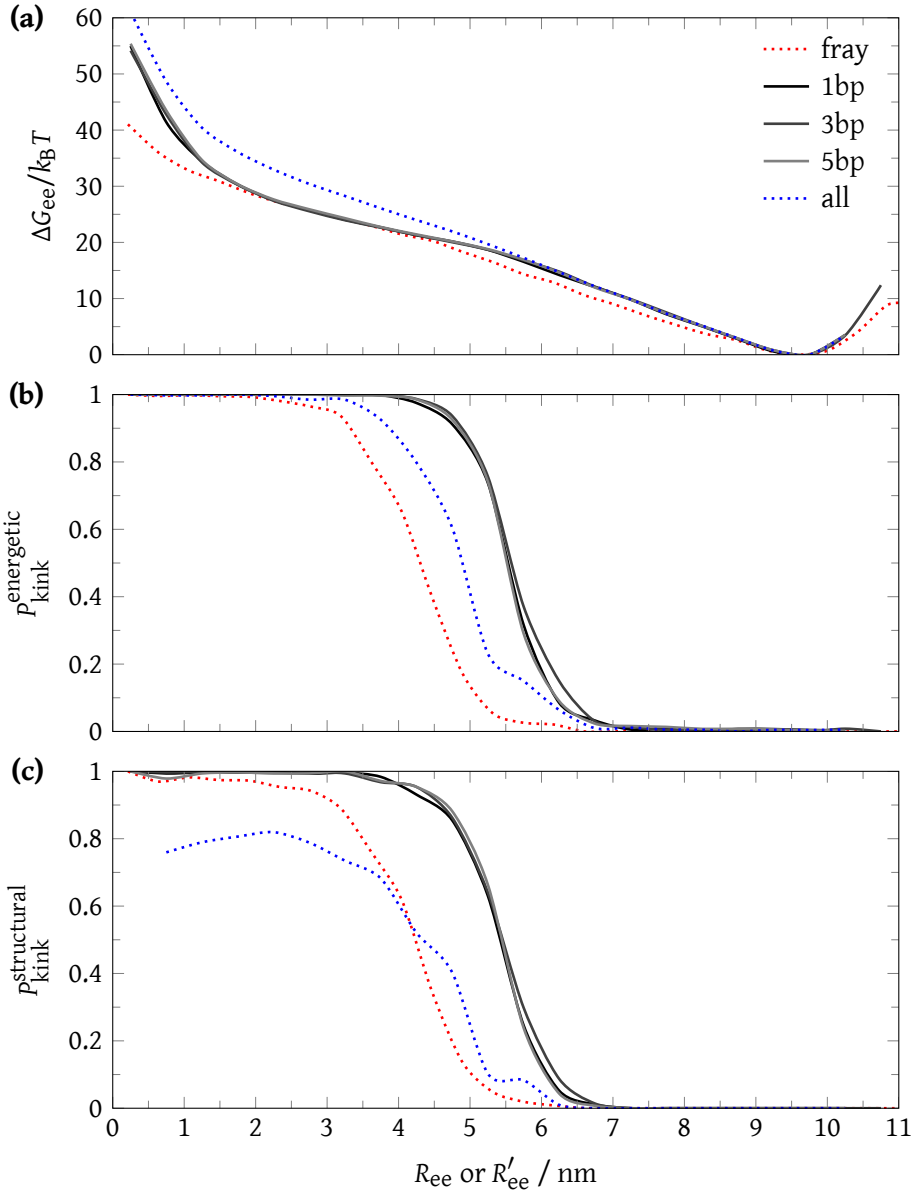


Figure A.1 Free energy of bending, impact of fraying
The impact of constraining the terminal 1, 3 and 5 base pairs from fraying on the free energy of bending $\Delta G(R_{ee})$ and kinking in an $N_{bp} = 30$ duplex. Limiting cases allowing fraying and enforcing hydrogen bonding along the entire duplex are also shown. **(a)** Free energy of bending. Probability of kinking based on the **(b)** energetic or **(c)** structural criteria for kinking. For the frayed system R_{ee} (the EED between the centres of the helical axis at the first and last base pairs) is ill-defined, so the very similar R'_{ee} (the EED between the base interaction sites of the first and last bases along the umbrella sampled strand) is used instead.

A.2 'Molecular vise'

Akin to the bending simulations of duplex DNA (Appendix A.1.1), the molecular vise simulations were performed in three phases: assembly, equilibration and production. An exploratory phase is not required since the simulations were unbiased, *i.e.* not using an umbrella sampling potential. As in the bending simulations, non-native base pairing was disallowed for simplicity. In the assembly phase, the system was initialized as two single strands. Harmonic traps were used to bring the stem ends, as well as the loop and complement strand, into proximity. Assembly with traps requires $\sim 5 \times 10^5$ VMMC steps per particle, significantly faster than diffusion limited assembly.

The assembled systems were equilibrated for 10^7 VMMC steps per particle. For the production phase, each production simulation (six independent simulations per point) was initialized with a statistically independent starting configuration and a unique random seed. This was accomplished by randomly drawing starting configurations from the equilibration phase, one for each production run, ignoring the first 2×10^6 VMMC steps per particle of the equilibration phase. The length of each production run was $\sim 10^8$ VMMC steps per particle.

All simulations were performed with the sequence-dependent oxDNA parameterization. Where available ($N_{\text{loop}} = 30, 36, 40, 46, 50$), the sequences are identical to those used in the experiments of Fields *et al.*;[22] where not available , similar sequences are used. All sequences are written $5'$ to $3'$.

The Fields *et al.* stem sequence is used for all systems: `GCC CGG CGG CTT ATA AAA TTT ATT AAT TAT ATA TTT TAT TTA ATA TAA T-Loop`

A complete list of loop sequences is available in Table A.1.

N_{loop}/bp	Loop sequence
20	TAC CGA TAA GCT TGG TCA TT
26	TCC CAC CGA TAA GCT TGG TCA TGC CT
30‡*	TCG CCC ACC GAT AAG CTT GGT CAT GCC CGT
36‡	TGC CCG CCC ACC GAT AAG CTT GGT CAT GCC CGC CGT
40‡	TCC GCC CGC CCA CCG ATA AGC TTG GTC ATG CCC GCC GCCT
46‡	TCC GCC GCC CGC CCA CCG ATA AGC TTG GTC ATG CCC GCC GCC GCC T
50‡	TGC CCG CCG CCC GCC CAC CGA TAA GCT TGG TCA TGC CCG CCG CCG CCC GT
56	TGC CGC CCG CCG CCC GCC CAC CGA TAA GCT TGG TCA TGC CCG CCG CCG CCC GCC GT
60	TCC GCC GCC CGC CGC CCG CCC ACC GAT AAG CTT GGT CAT GCC CGC CGC CGC CCG CCG CCT
66	TCG CCC GCC GCC CGC CGC CCG CCC ACC GAT AAG CTT GGT CAT GCC CGC CGC CGC CCG CCG CCC GCT
70	TGC CGC CCG CCG CCC GCC GCC CGC CCA CCG ATA AGC TTG GTC ATG CCC GCC GCC GCC CGC CGC CCG CCG T
76	TCC GGC CGC CCG CCG CCC GCC GCC CGC CCA CCG ATA AGC TTG GTC ATG CCC GCC GCC GCC CGC CGC CCG CCG GCC T
80	TGC CCG GCC GCC CGC CGC CCG CCG CCC GCC CAC CGA TAA GCT TGG TCA TGC CCG CCG CCG CCC GCC GCC CGC CGG CCC GT
86	TCG CGC CCG GCC GCC CGC CGC CCG CCG CCC GCC CAC CGA TAA GCT TGG TCA TGC CCG CCG CCG CCC GCC GCC CGC CGG CCC GCG CT
90	TGC CGC GCC CGG CCG CCC GCC GCC CGC CGC CCG CCC ACC GAT AAG CTT GGT CAT GCC CGC CGC CGC CCG CCG CCC GCC GGC CCG CGC CGT

Table A.1 Molecular vise sequences

Loop sequences for the molecular vise system. Terminal loop thymines, used to compute $R_{ee}^{hairpin}$ are highlighted (**red**).

* Mismatch locations from Fields *et al.*[22] are highlighted (**blue**). Following the experimental design, the mismatch is introduced on the complement strand.

‡ Sequences from Fields *et al.*[22]

A.3 Cyclization

A.3.1 Cyclization simulations

Cyclization simulations were performed in three phases: exploratory, equilibration and production. In all cases, we use a virtual-move Monte Carlo (VMMC) algorithm[62] in combination with umbrella sampling.[63]

In the exploratory phase, we iteratively adjusted the umbrella sampling bias for the windows associated with the open and cyclized states, yielding a flat population of states for both windows. We use discrete potentials for both dimensions of the order parameter (Section 6.2.1). For Q_{ee}, we use variable width distance increments: 0–1.7036 nm, 3.4072, 5.1108, 8.518, 12.777, 17.036, 21.295, 25.554, 34.072, 42.59, 51.108 and $>$ 51.108 nm. For Q_{bp}, the increment is fixed, and is simply the number of base pairs formed ($0 \leq Q_{bp} \leq N_s$). As the weighting iterations were semi-automated, simulation times varied in the range 10^6–10^7 VMMC steps per particle. Simulations were performed with one molecule in a cube of side 170.36 nm, which corresponds to a unimolecular concentration of 336 nM. To further simplify sampling, we forbid the formation of base pairs that are not intended in the design of the system (non-native base pairs).

In the equilibration phase, we equilibrated the system for 10^7 VMMC steps per particle using the aforementioned umbrella weights. Adequate sampling (number of transitions in Q_{ee} and Q_{bp}) and decorrelation (via a block average decorrelation method) of potential energy, bubble size, fraying and structural kinking were checked. For the production phase, each simulation (five independent simulations per measurement) was initialized with a statistically independent starting configuration and a unique random seed. This is accomplished by randomly drawing starting configurations from the equilibration phase, one for each production trial, ignoring the first 2×10^6 VMMC steps per particle of equilibration. The production trial run time is 10^7 VMMC steps per particle. For reference, the characteristic decorrelation time for the potential energy is $\sim 10^4$ VMMC steps per particle, while the decorrelation time for kinking is $\sim 10^5$ VMMC steps per particle.

The‘seed moves’ used to build clusters in the VMMC algorithm were:

- Rotation of a nucleotide about its backbone site, with an axis chosen uniformly on the unit sphere, and with an angle drawn from a normal distribution with a mean of zero and a standard deviation of 0.10 radians.
- Translation of a nucleotide, where the displacement along each Cartesian axis is drawn from a normal distribution with a mean of zero and a standard deviation of 0.085 18 nm.

A.3.2 Dimerization simulations

Dimerization simulations follow a very similar procedure as cyclization simulations (Appendix A.3.1), but adapted for a bimolecular system. With the exception of the starting configuration (bimolecular), exploratory, equilibration and production procedures are identical.

As introduced in Section 6.1.2, each molecule has one complementary sticky end and one blunt end. This precludes the formation of multimers and circular dimers, allowing only the formation of linear dimers. Simulations were performed with these two complementary, non-palindromic, sticky-ended duplexes, in a cube of side 170.36 nm, which corresponds to a dimer concentration of 336 nM. To estimate the impact of excluded volume effects, dimerization systems were also performed in a cube of $^1/_2$ the size and therefore $8\times$ the concentration (85.18 nm per side, which corresponds to a dimer concentration of 2.69 µM). Like the cyclization simulations, non-native base pairing was disallowed.

Analogously to the cyclization simulations, the dimerization simulations are windowed using the order parameters Q_{ee} and Q_{bp}, respectively the distance of closest approach and the number of base pairs formed between complementary sticky ends. The window associated with the dimerized state is given by $Q_{bp} \geq 1\,\text{bp}$ ($Q_{ee} = Q_{ee}^{min}$, the small separation between base sites of the base paired nucleotides), while the window associated with the undimerized state is given by $Q_{bp} = 0\,\text{bp}$.

A.3.3 Computation of equilibrium constants

The cyclization reaction is unimolecular (Figure 6.1 (a)):

$$\mathrm{A} \underset{k_{\mathrm{uncyc}}}{\overset{k_{\mathrm{cyc}}}{\rightleftharpoons}} \mathrm{B}, \tag{A.1}$$

where A and B are the open and cyclized states respectively, k_{cyc} is the forward rate constant (cyclization) and k_{uncyc} is the reverse rate constant (uncyclization). The equilibrium constant for cyclization, $K_{\mathrm{eq}}^{\mathrm{cyc}}$, can be estimated directly from simulations of a unimolecular, isolated system as

$$K_{\mathrm{eq}}^{\mathrm{cyc}} = \frac{P_{\mathrm{B}}}{P_{\mathrm{A}}}. \tag{A.2}$$

Here P_{A} and P_{B} are the probabilities with which uncyclized and cyclized systems are observed in simulation ($P_{\mathrm{A}} + P_{\mathrm{B}} = 1$).

The dimerization reaction is a bimolecular association of distinct molecules:

$$\mathrm{A} + \mathrm{B} \underset{k_{\mathrm{undim}}}{\overset{k_{\mathrm{dim}}}{\rightleftharpoons}} \mathrm{AB}. \tag{A.3}$$

In this case, the bimolecular equilibrium constant can be inferred from a simulation of a single pair of dimerizing monomers via

$$K_{\mathrm{eq}}^{\mathrm{dim}} = \frac{P_{\mathrm{AB}}}{P_{\mathrm{A/B}}[\mathrm{A}_0]}, \tag{A.4}$$

in which $[A_0]$ is the total concentration of strand type A in simulation, and $P_{\mathrm{A/B}}$ and P_{AB} are the probabilities with which monomers and dimers are observed in simulation. This expression follows from Equation (7) of Ouldridge *et al.*[115] The resultant $K_{\mathrm{eq}}^{\mathrm{dim}}$ obeys the standard law of mass action for bulk systems in equilibrium,

$$K_{\mathrm{eq}}^{\mathrm{dim}} = \frac{[\mathrm{AB}]}{[\mathrm{A}][\mathrm{B}]}. \tag{A.5}$$

Recall from Section 6.1.3 that we use a definition for the equilibrium *j*-factor j_{eq} that is consistent

N_{bp}	Note	Cyclization sequence
$N_s = 10$		
30‡		CAG AAT CCG TGC TAC ACC TCC ACC GTT TCA
67		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTT TGA CCT GAC TAT CCT CAC CTC CAC CGT TTC A
		CGA TCA TGG AGT TAT ATC TGA GGG AAA CTG GAC TGA TAG GAG TGG AGG TGG CAA AGT GTC TTA GGC A
69		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTT TGA CCC ATG ACT ATC CTC ACC TCC ACC GTT TCA
71		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTT CTA ATT GAC TGA CTA TCC TCA CCT CCA CCG TTT CA
73		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTT CTA ATT GAC CAT GAC TAT CCT CAC CTC CAC CGT TTC A
74		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTT TAA GTT GAC CCA TGA CTA TCC TCA CCT CCA CCG TTT CA
76		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTT CTA AGT TGA CCT CAT GAC TAT CCT CAC CTC CAC CGT TTC A
93		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTT AAT ACT TCT CCT ATG ACT TCT AAT TGA CCC ATG ACT ATC CTC ACC TCC ACC GTT TCA
94		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTT TAA TAC TTC TCC TAT GAC TTC TAA TTG ACC CAT GAC TAT CCT CAC CTC CAC CGT TTC A
97		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTA TGT TAA TAC TTC TCC TAT GAC TTC TAA TTG ACC CAT GAC TAT CCT CAC CTC CAC CGT TTC A
99		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTA GAT GTT AAT ACT TCT CCT ATG ACT TCT AAT TGA CCC ATG ACT ATC CTC ACC TCC ACC GTT TCA
101		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTG TAG ATG TTA ATA CTT CTC CTA TGA CTT CTA ATT GAC CCA TGA CTA TCC TCA CCT CCA CCG TTT CA
103		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTA CGT AGA TGT TAA TAC TTC TCC TAT GAC TTC TAA TTG ACC CAT GAC TAT CCT CAC CTC CAC CGT TTC A
105		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTG AAC GTA GAT GTT AAT ACT TCT CCT ATG ACT TCT AAT TGA CCC ATG ACT ATC CTC ACC TCC ACC GTT TCA
106		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTA GAA CGT AGA TGT TAA TAC TTC TCC TAT GAC TTC TAA TTG ACC CAT GAC TAT CCT CAC CTC CAC CGT TTCA
207		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTA TCA GTA CGA AGC TGG GCT ATA CCG TTC TTA TTG TCC TTA ATA CCA CCG ACG AGT TGT ACG CCC TCT CAT CCG AAG ACG ACA CGT ACC TGG GAA AAG AAC GTA GAT GTT AAT ACT TCT CCT ATG ACT TCT AAT TGA CCC ATG ACT ATC CTC ACC TCC ACC GTT TCA
$N_d = 91$		
101		CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTG TAG ATG TTA ATA CTT CTC CTA TGA CTT CTA ATT GAC CCA TGA CTA TCC TCA CCT CCA CCG TTT CA
100‡		AGA ATC CGT GCT AGT ACC TCA ATA TAG ACT CCC TGT AGA TGT TAA TAC TTC TCC TAT GAC TTC TAA TTG ACC CAT GAC TAT CCT CAC CTC CAC CGT TTC A
99‡		GAA TCC GTG CTA GTA CCT CAA TAT AGA CTC CCT GTA GAT GTT AAT ACT TCT CCT ATG ACT TCT AAT TGA CCC ATG ACT ATC CTC ACC TCC ACC GTT TCA
98‡		AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTG TAG ATG TTA ATA CTT CTC CTA TGA CTT CTA ATT GAC CCA TGA CTA TCC TCA CCT CCA CCG TTT CA
97‡		ATC CGT GCT AGT ACC TCA ATA TAG ACT CCC TGT AGA TGT TAA TAC TTC TCC TAT GAC TTC TAA TTG ACC CAT GAC TAT CCT CAC CTC CAC CGT TTC A
96‡		TCC GTG CTA GTA CCT CAA TAT AGA CTC CCT GTA GAT GTT AAT ACT TCT CCT ATG ACT TCT AAT TGA CCC ATG ACT ATC CTC ACC TCC ACC GTT TCA
95‡		CCG TGC TAG TAC CTC AAT ATA GAC TCC CTG TAG ATG TTA ATA CTT CTC CTA TGA CTT CTA ATT GAC CCA TGA CTA TCC TCA CCT CCA CCG TTT CA
Variable-sequence N_d		
73	TA	CAG AAT CCG TAG CTC TAG CAC CGC TTA AAC GCA CGT ACG CGC TGT CTA CCG CGT TTT AAC CGC CAA TAG GAT T
	E8A10	CAG AAT CCG TTT TTA TTT ATC GCC TCC ACG GTG CTG TTT TTT TTT TCT GTT GGC CGT GTT ATC TCG AGT TAG T
	E8A17	CAG AAT CCG TTT TTA TTT ATC GCC TCC ACG GTG CTT TTT TTT TTT TTT TTT GGC CGT GTT ATC TCG AGT TAG T
	E8A26	CAG AAT CCG TTT TTA TTT ATC GCC TCC TTT TTT TTT TTT TTT TTT TTT TTT TTC CGT GTT ATC TCG AGT TAG T
	E8A38	CAG AAT CCG TTT TTA TTT ATC GTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT ATC TCG AGT TAG T
Structural defects		
69[m]	C:C	CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTT TGA CCC ATG ACT ATC CTC ACC TCC ACC GTT TCA
97[n]	Nick 1	CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTA | TGT TAA TAC TTC TCC TAT GAC TTC TAA TTG ACC CAT GAC TAT CCT CAC CTC CAC CGT TTC A
		CGA TCA TGG AGT TAT ATC TGA GGG ATA CAA TTA TGA AGA GGA TAC TGA AGA TTA ACT GGG TAC TGA TAG GAG TGG AGG TGG CAA AGT GTC TTA GGC A
97[n]	Nick 2	CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTA TGT TAA TAC TTC TCC TAT GAC TTC TAA TTG ACC CAT GAC TAT CCT CAC CTC CAC CGT TTC A
		CGA TCA TGG AGT TAT ATC TGA GGG ATA CAA TTA TGA AGA GGA TAC TGA AGA TTA ACT GGG TAC | TGA TAG GAG TGG AGG TGG CAA AGT GTC TTA GGC A
97[n]	Double	CAG AAT CCG TGC TAG TAC CTC AAT ATA GAC TCC CTA | TGT TAA TAC TTC TCC TAT GAC TTC TAA TTG ACC CAT GAC TAT CCT CAC CTC CAC CGT TTC A
		CGA TCA TGG AGT TAT ATC TGA GGG ATA CAA TTA TGA AGA GGA TAC TGA AGA TTA ACT GGG TAC | TGA TAG GAG TGG AGG TGG CAA AGT GTC TTA GGC A

Table A.2 Cyclization sequences

Unless otherwise stated,‡, cyclization sequences are from V&H,[15] with sticky end (N_s) highlighted in red, and the remainder of the sequence (N_d) in black. Sequences are written 5′ to 3′. For conciseness, other than the initial example of $N_{bp} = 67$, a second strand is omitted where it is the reverse complement of the first strand.

‡ Non-V&H sequence. [m] C:C Mismatch highlighted (**blue**). [n] Nicks highlighted with bar | (**blue**).

with V&H:[15]

$$j_{\mathrm{eq}} \equiv \frac{K_{\mathrm{eq}}^{\mathrm{cyc}}}{K_{\mathrm{eq}}^{\mathrm{dim}}}. \tag{A.6}$$

This is a measure of the effective concentration of one sticky end in the vicinity of the other when the dimerization reaction involves distinct monomers, forming a heterodimer, with only one sticky end per monomer. Researchers often estimate $K_{\mathrm{eq}}^{\mathrm{dim}}$ using the dimerization of identical monomers, forming a homodimer, with palindromic sticky ends. In this case, given the same underlying interaction strength between sticky ends (and hence the same $K_{\mathrm{eq}}^{\mathrm{cyc}}$), $K_{\mathrm{eq}}^{\mathrm{dim}}$ would be increased by a factor of two due to combinatorial (4x the number of possible dimers) and symmetry effects (1/2 due to homodimer versus heterodimer).[115,215] As a result, with this approach the equilibrium j-factor is estimated as

$$j_{\mathrm{eq}} = 2\frac{K_{\mathrm{eq}}^{\mathrm{cyc}}}{K_{\mathrm{eq}}^{\mathrm{dim}}}. \tag{A.7}$$

In the case of identical monomers with two non-palindromic sticky ends, there is no pre-factor in the j-factor estimate, as the combinatorial (2x number of possible dimers) and symmetric effects (1/2 due to homodimer versus heterodimer) cancel. Overall, under the approximation of ideal behavior of separate complexes, these three approaches are equivalent.

A.4 Minicircles

Minicircle simulations were performed with a molecular dynamics (MD) algorithm using a simple Andersen-like thermostat.[59] Briefly, Newton's equations of motion are integrated (NVE ensemble) with a time-step of $\delta t = 15$ fs. To approximate Brownian motion,[ii] every N_{newt} steps, the velocity and momenta of each particle are refreshed with a refresh probability (pt) from a Maxwell-Boltzmann distribution of the temperature of interest ($T = 298$ K), which emulates the *NVT* ensemble. The refresh probability (pt) was set by the newtonian step ($N_{\mathrm{newt}} = 103$ MD

[ii] Ballistic motion occurs in the limit where $pt = 0$

steps) and the diffusion coefficient (D) by

$$pt = \frac{2\frac{k_B T}{m} \times N_{\text{newt}} \times \delta t}{\frac{k_B T}{m} \times N_{\text{newt}} \times \delta t + 2D}, \tag{A.8}$$

where m is the oxDNA unit mass, 5.24×10^{-25} kg (for the reader's convenience, we have written the equation in SI units, instead of oxDNA reduced units). The diffusion coefficient was chosen to have small temperature deviation from the set point, and to decorrelate potential energy and kinking relatively rapidly. We choose $D = 120\,\text{nm}^2\,\text{ns}^{-1}$, yielding a refresh probability of $pt = 9.7\,\%$. This choice leads to a sufficiently small discrepancy between the intended value of the temperature and that measured from the kinetic energy, a difference of about 0.3 K at $T = 298$ K. This diffusion coefficient is $100 - 1000\times$ greater than those reported experimentally;[216] however, artificial increases in D have been previously reported for hybridization kinetics with oxDNA,[80] improving sampling without otherwise perturbing the result. Configurations are output every 10^5 steps, which corresponds to the characteristic potential energy decorrelation time.

We forbid the formation of base pairs that are not intended in the design of the system (non-native base pairs). For example, with two strands (both written $5' \rightarrow 3'$), the first base of the complement strand would only be allowed to form base pairs with the last base of the reverse complement strand, and vise-versa.

Minicircle simulations were performed in two phases: equilibration and production. In the equilibration phase, appropriate starting configurations (intact with given Lk or nicked) were run for 10^7 MD steps. For the production phase, five independent simulations were initialized with random seed and starting configurations drawn from the equilibration phase (configurations at 6×10^7, 7×10^7, 8×10^7, 9×10^7 and 1×10^8 MD steps).

Results are identical to those obtained with a VMMC algorithm[62] at 298 K. Sequences are available in Table A.3 for the intact relaxed minicircles with 50 % GC-content.

N_{bp}	Lk	σ	GC/%	Sequence
31	3	0.003	48.39	CGC TCA TCA TAT ACG GGC AGA AGT CAA GCT A
32	3	−0.029	50.00	GGT CCA CTC AAG TTT GCT TTG ATG TGA CTG CG
41	4	0.011	48.78	AGA ACC CTG CGC TAC GAT GAA GGC CGA AAC TAT ATC TAT CG
42	4	−0.013	50.00	AGC AAG GAA CTC AGT TGA CGC TAT AGA GAA GCC ACC AGT TGC
51	5	0.016	49.02	AGC AGT ATT TCC TTG CTG TAT AGA CCG GGA GAC ACA GAC GCT TGA CAT CAG
52	5	−0.004	50.00	TTC TAT TCC AAA CAA CGC CAG TCG CCG CAA ACT GCT GTC TGA GGG AAA CTA C
62	6	0.003	50.00	CCC ACA GTA ACG ATC TTC TGT TGT AAC GTG ACT CAG CTG TCG TAG CAG GTA GTC TGG CTC TA
63	6	−0.013	49.21	GCC CAA CGT CCA GAC ACA CTG CCT TCA GTA AGT GAT GAA TGA AAA GGC TGA GCT GGG GAT AAA
72	7	0.007	50.00	CAC TGG GCC ATT CAA GAT GAA GCA GGC TGA GAC AGC TCA CAT TTC GTT TTC CGT ATT GGA GTG CAT GTT GGG
73	7	−0.007	49.32	AAA CTC AGC TCA GTC ATG AGG CGC TAA GTA ACG GAA GCC GCT CTC ACT ATC ATT TTA CGC TGT CCC AGA GAC T
82	8	0.011	50.00	CAC GCG GCG GCA AAG ATA GAA AGC AAC TAC CAT TGT GCA TCT TTG AAT CCT ACG GAC GTG CAG CGT TGC ATT AAG CGC CTT A
83	8	−0.001	49.40	ACA CTT CAT GAT TAT ATG CGC ACC GTT TGC ATG CCT CTG CTC AAA CGC GCT ACA CAC TAC GGG GAC AGG TAG ATG TTC GGA AA

Table A.3 Minicircle sequences
50 % GC-content sequences used for investigating intact relaxed minicircles with sequence-dependence. Sequences were randomly generated, choosing the first sequence to satisfy the ∼50 % GC-content criterion. Misbonding is explicitly disallowed, obviating the need to screen-out pathologic sequences, if present. Superhelical density σ computed from Equation (7.2).

A.5 'Strained duplex'

A.5.1 Structural simulations

Structural simulations were performed with both MD and VMMC moves. Unless otherwise stated, simulations were performed at $T = 23\,^{\circ}\mathrm{C}$. The results for both methods are nearly identical (Appendix A.5.2). Structural simulations are development studies, *i.e.* there is only one trial for each molecule (N_{d}, N_{s}). Instead of pre-equilibration, simulations are initialized with unique random seeds and an unequilibrated starting configuration; the first 10 % of the trajectory is ignored as potentially unequilibrated. Error bars represent the standard deviation (SD) (usually we report standard error of the mean (SEM)), and where not shown, are comparable to line thickness.

For MD, simulations are run for 2×10^8 MD steps, with trajectories output every 2×10^5 MD steps. Unless otherwise stated in the text, all other parameters are the defaults discussed in Section 3.1.2. For VMMC, simulations are run for 2×10^7 VMMC steps per particle, with trajectories output every 1×10^4. VMMC 'seed moves' were zero mean with standard deviation 0.2 radians for rotation, and zero mean with standard deviation 0.05 simulation units (0.0425 nm) for translation (Section 3.1.3).

A.5.2 Structural simulations, sampling

Adequate sampling and statistics are imperative for the robustness of course grain simulations; that the simulation is properly equilibrated cannot be assumed *a priori*, and must be demonstrated. It is non-trivial to prove that a system is at equilibrium since the various constituent parts of a system often equilibrate on different time and length scales.

For example, let us take a simple chemical reaction $\mathrm{A} \longrightarrow \mathrm{B} \longrightarrow \mathrm{C}$. The equilibration time scale of molecular rearrangement could be on the order of picoseconds, $\mathrm{A} \longrightarrow \mathrm{B}$ on the order of seconds, and $\mathrm{B} \longrightarrow \mathrm{C}$ on the order of days. A millisecond time scale simulation could be useful for studying the presumably equilibrated process of molecular rearrangement, but would be useless for the out of equilibrium reactions $\mathrm{A} \longrightarrow \mathrm{B}$ and $\mathrm{B} \longrightarrow \mathrm{C}$. Each process under study must be independently checked for robustness, and the simulation run long enough to equilibrate the

longest time-scale phenomena.

At long time scales, Markovian processes result in exponential distributions

$$\chi(t) \sim \exp\left(\frac{t}{\tau_c}\right), \tag{A.9}$$

where χ is the autocorrelation function, t is time and τ_c the characteristic time. By definition, the data are completely correlated at $\chi = 1$, and completely decorrelated at $\chi = 0$.

The discrete time autocorrelation function is given by

$$\chi(t) = \frac{1}{N}\sum_{n=1}^{N} x_t \bar{x}_{t+n} - \langle x \rangle_t^2, \tag{A.10}$$

where N is the number of data points and $\langle x \rangle_t$ is the time average of x. Assuming ergodicity,

$$\langle x \rangle_{\text{ensemble}} = \langle x \rangle_{\text{time}}, \tag{A.11}$$

the ensemble average may be substituted for the time average; for example, when computing the autocorrelation of Monte Carlo (MC) trajectories.

In practice, the autocorrelation function is noisy at short time scales in MD and MC simulations. To reduce noise, we use the block averaged autocorrelation function

$$\chi(t) = \frac{\langle x^2 \rangle_t - \langle x \rangle_t^2}{N} \sum_{t=1}^{N-M_b} \sum_{j=1}^{M_b} x_t \bar{x}_{t+j} - \langle x \rangle_t^2; \tag{A.12}$$

where M_b is the number of blocks. The prefactor $1/N(\langle x^2 \rangle_t - \langle x \rangle_t^2)$ produces an estimate of the accuracy of the standard deviation. Deviations from $\chi(t = 0) = 1$ indicate an inaccurate estimation of the standard deviation. The block averaged autocorrelation reduces to the simple autocorrelation function in the limit $N \sim M_b$.

Simulations were taken as equilibrated where $\chi \to 0$, usually after $\sim 5\tau_c$. Likewise, data points were considered independent after this time interval; therefore the SEM is given by

$$SEM_{\bar{x}} = \sqrt{5\tau_c}\frac{\sigma}{\sqrt{N}} \tag{A.13}$$

where $SEM_{\bar{x}}$ is the SEM of the distribution, σ is the uncorrected (biased) standard deviation and $\sqrt{5\tau_c}$ is my approximation for uncorrelated states.

The characteristic time, a proxy for equilibration time-scale, differ with respect to the metric under study. In the case of structural deformations, it is unclear, *a priori*, the relative equilibration time scales of potential energy, EED, fraying, kinking and bubble formation. To ensure that simulation results are indeed sampling equilibrium states, and to correct the unbiased standard deviation to account for correlation between states, block averaged autocorrelation is taken for these properties. Representative results for intact molecule ($N_d = 30, N_s = 8$), a system in the kink transition region, are shown in Figure A.2.

Ergodicity cannot be assumed *a priori*, so time (MD) and ensemble averaged (VMMC) trajectories are explicitly shown to be equivalent. Block averaged autocorrelation (χ) shows a clear separation of time scales between EED/Fray (shorter) and Bubble/Kink (longer). This separation is more pronounced for VMMC than MD.

A.5.3 Thermodynamic simulations, melting and base pairing

Umbrella sampling was used to determine melting and base pairing. We use a one-dimensional order parameter for the number of base pairs formed (Q_{bp}), with windows for the open ($Q_{bp} = 0$) and closed ($Q_{bp} > 0$) states. In the exploratory phase, weights are iteratively adjusted to yield approximately even sampling. In the development phase, one simulation is run to yield many equilibrated states. In the production phase, starting configurations are drawn from the development phase, initializing 5 independent trials with unique random seeds. Error bars represent the SEM, and where not shown are comparable to line thickness.

For base pairing, simulations were run at 300 K. For melting, we extrapolate the partition function via histogram reweighting for $T = 310$–360 K from the ensemble at 340 K. Melted states are taken as $Q_{bp} = 0$, with a molecule considered associated if there is any base pairing ($Q_{bp} > 0$). Unsurprisingly, there is an extrapolation error between 300 K and 340 K, but this should not impact the qualitative result (Appendix A.5.4).

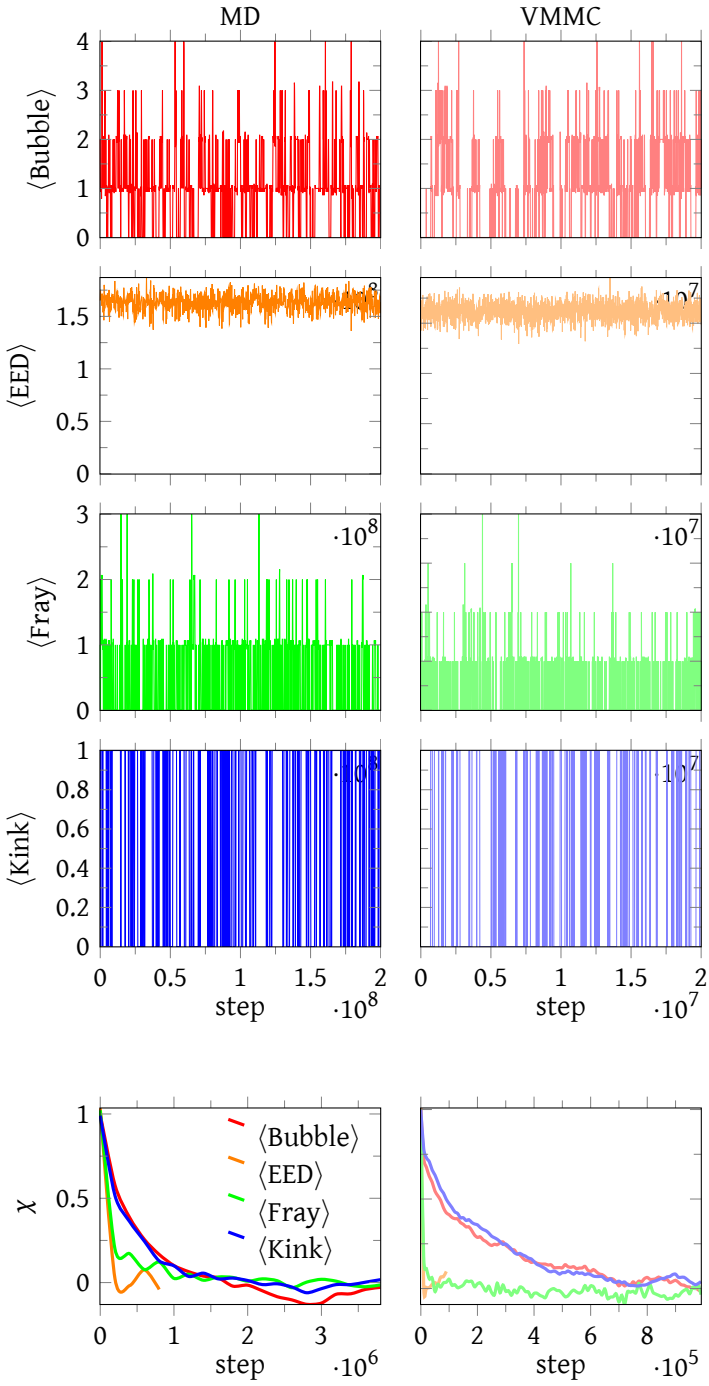


Figure A.2 Representative sampling for 'strained duplex'
Representative MD and VMMC trajectories for Bubble (**red**), EED (**orange**), Fray (**green**) and Kink (**blue**) for a intact molecule ($N_d = 30, N_s = 8$). Block averaged autocorrelation function (χ).

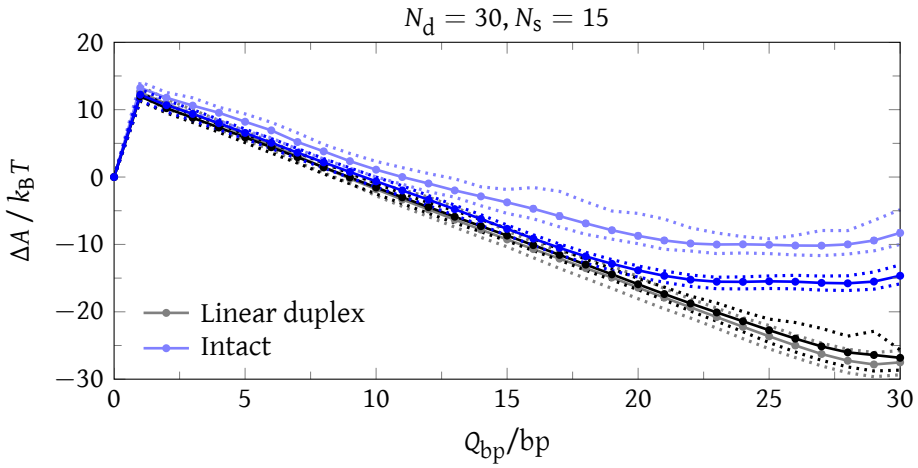


Figure A.3 Free energy profile, temperature extrapolation
Free energy profile versus of number of formed base-pairs for linear duplex (**black**) and intact (**blue**) systems ($N_d = 30, N_s = 15$) with the average-base parameterization. Simulation results at 300 K (**light coloured**) and 340 K (**dark coloured**) extrapolated to 320 K using histogram reweighting. Errors bars represent SEM from 10 independent simulations. Dotted lines are minimum and maximum values over all simulations.

A.5.4 Thermodynamic simulations, temperature extrapolation

Since the histogram reweighting employed in Section 8.5.2 uses an identical configuration space over all temperatures, it becomes inaccurate if the accessible state-space changes as a function of temperature. This is of course true; however, state-space accessibility can be assumed as constant within narrow temperature ranges, where the bounds of 'narrow' are system dependent. To test the applicability of this approximation on the strained duplex model system, the 300 K and 340 K simulations are extrapolated to 320 K (Figure A.3). Given the qualitative agreement in energy profile shape, the difference between the two extrapolations ($\sim 5\, k_B T$) sets an approximate maximum value for the extrapolation error around 320 K. Owing to the large deviations of the strained molecules, precise quantitative conclusions should not be drawn from the histogram extrapolation of the strained molecules. Extrapolation over a 45 K window ($T = 315$–360 K from an configuration space at $T = 340$ K), as is the case for melting profile determination, is too large for quantitative comparison. Indeed, extrapolation errors likely explain some of the deviation between experimental and simulation values of T_m (Table 8.1).

Quantitative prediction of T_m is not, however, the aim of this work; rather, an examination of the short to intermediate length-scale behavior of DNA flexibility, and the resulting stringent critique of the Qu *et al.* analytical model. Experimental comparisons, such as T_m, are useful

in so far as they highlight the strengths and deficiencies of the model. If anything, the close agreement between experimental and simulation melting curves highlights the robustness of the oxDNA model, and its appropriate treatment of real-world DNA thermodynamics.

B Part II: Detailed Experimental Methods

B.1 Expression, purification and mutagenesis

B.1.1 RecQwt DNA template preparation

E. coli full-length wild-type RecQ (RecQwt) in a pET15b vector was generously provided by Professor James L. Keck, Biomolecular Chemistry, University of Wisconsin. In our hands, the previously reported thrombin treatment[159] to remove the N-terminal 6-His tag led to protein cleavage between the Helicase/RecQ-Ct and helicase and RNaseD C-terminal (HRDC) domains, not a full-length product as reported (Section 11.5.1). To remedy this problem, the full-length wild-type RecQ (RecQwt) insert was cut (BamHI/NdeI) and inserted into a pET28b vector (Figure B.1). The thrombin site (GLVPRGS) was then replaced with a Tobacco Etch Virus nuclear inclusion a endopeptidase (TEV) protease site (ENLYFQ\G) through two-rounds of mutagenesis using the Stratagene QuikChange II XL site-directed mutagenesis kit. Subsequent 6-His removal was performed with TEV protease, resulting in full-length RecQwt by both Tris-Glycine sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and reverse-phase high-performance liquid chromatography (HPLC)/electrospray ionization mass spectometry (ESI-MS) (data not shown).

B.1.2 RecQ expression and purification

Competent BL21(DE3)gold pLysS cells were transformed with the relevant wild-type or mutant RecQ. After an initial 50 mL isopropyl β-D-1-thiogalactopyranoside (ITPG) induction test, RecQ was expressed overnight (O/N) in 1 L complete MagicMedia supplemented with 25 µg mL^{-1} Kanamycin (pET28b) and 34 µL mL^{-1} Chloramphenicol (pLysS). Cells were pelleted (Sorvall RC 5C Plus with Swing Bucket SH3000, 4000 rpm, 4 °C, 45 min) and stored at −80 °C.

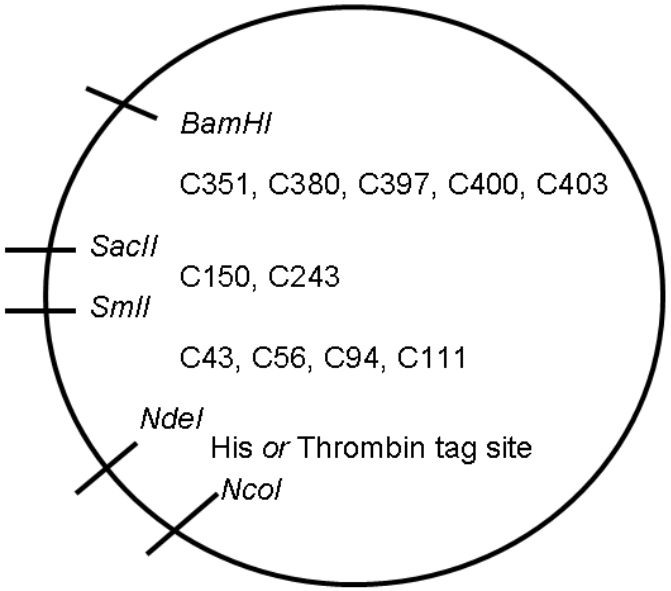


Figure B.1 Cartoon of RecQ inserted into pET28b vector
The approximate positions of relevant restriction sites, 11 wild-type cysteines, and tag-location are highlighted.

Following resuspension in buffer A (namely 20 mM Imidazole, Appendix B.7.1.1) and the addition of a Roche cOmplete, ethylenediaminetetraacetic acid (EDTA)-free Protease Inhibitor Cocktail tablet, the cell pellet was broken by emulsification (IKA-Werke Ultra-Turrax T25), dry ice freeze/thawing and sonication (Branson 450). Following two centrifugation steps to remove bulk cell contents (Eppendorf 5810R with FA-46-6-30 Fixed Rotor, 9000 rpm, 4 °C, 30 min; Beckman-Coulter Optima L-100 XP Ultracentrifuge with Type 45 Ti Rotor, 40 000 rpm, 4 °C, 1 h), 6-His tagged RecQ was purified at room temperature (R/T) from the supernatant via fast protein liquid chromatography (FPLC) and ion affinity chromatography (IMAC) (GE Healthcare HisTrap HP 5 mL column or self-packed 5 mL Qiagen Ni-nitrilotriacetic acid (NTA) Superflow Cat. No. 30430) with a linear gradient from buffer A to buffer B (buffer A with 300 mM Imidazole, Appendix B.7.1.2). In addition to absorbance, fractions were analyzed via SDS-PAGE before pooling. HisTEVRecQ elutes at about 70/30 buffer A/bufferB. To simplify purification, a step-gradient from 100 % buffer A to 100 % buffer B (no linear transition) was also used. In our quality control assays, there is no difference between the step and linear gradient purification product (data not shown).

To remove the His-tag (note that not all substrates have their His-tags removed) pooled fractions were concentrated (Millipore Amicon Ultra-15 Centrifugal Filter) and cleaved with TEV

protease (plasmid DNA for TEV protease was generously provided by Dr David Waugh, Protein Engineering Section, Macromolecular Crystallography Laboratory, National Cancer Institute). Briefly, 50:1 RecQ:TEV protease by A_{280} absorbance was incubated under reducing conditions (1 mM dithiothreitol (DTT)) at R/T for 2.5 h. To separate cleaved RecQ from TEV/HisRecQ, the reaction mixture was IMAC purified (GE Healthcare HisTrap HP 5mL column) with a linear gradient from buffer A to buffer B. Again, fractions were analyzed via absorbance and SDS-PAGE before pooling. 6-His cleaved RecQ eluted at 100 % buffer A.

Pooled fractions were further purified via gel filtration on a GE Healthcare HiPrep 16/60 Sephacryl S-300 HR gel filtration column with 100 % buffer C (Appendix B.7.1.3).[i] In our hands, it is imperative that the GF be performed at R/T; RecQ precipitation is accelerated, and subsequent yields lowered, at 4 °C. Each fraction was quality controlled via SDS-PAGE. Pooled fractions are quality controlled via reverse-phase HPLC/ESI-MS and nicotinamide adenine dinucleotide (NADH)-coupled adenosine triphosphate (ATP) activity assay; sometimes, fluorescence-based ensemble DNA-unwinding was also employed. RecQ is stored at −80 °C or liquid nitrogen.

B.1.2.1 Useful observations and words of warning

IMAC: Addition of 1 mM β-mercaptoethanol (BME) to IMAC buffer A/B was observed to slow, but not eliminating precipitation, and was therefore useful in overnight steps.

Storage: In our hands, storage for $>$ 3 month at −80 °C leads to protein degradation into ∼60 kDa and ∼8 kDa fragments, as shown by SDS-PAGE (data not shown). For long-term storage, liquid nitrogen is advised. Overall activity is reduced, but not eliminated, by multiple freeze/thaw cycles; store in small one-time use aliquots.

Stability: RecQ (presumably RecQΔHRDC due to degradation) retains its DNA-stimulated ATPase activity, albeit with absolute and relative turnover reduced, after multiple freeze/thaw cycles and storage at −20 °C for ∼18 months and counting.

[i] RecQ ellutes as one peak, suggesting a monomeric species; however, very weakly associating, or very short-lived multimers cannot be ruled out with gel filtration.

B.1.3 RecQ mutagenesis

Three distinct rounds of RecQ mutagenesis were been performed:

(i) Cys-lite
(ii) Site-specific labeling
(iii) Double labeling

The Stratagene QuikChange II XL site-directed mutagenesis kit and the primers listed in Table B.2 were used for all rounds. Plasmid DNA was verified by 1 % agarose gels (data not shown) and sequencing (sequencing primers also available in Table B.2). RecQ protein mutants are listed in Table B.1. All RecQ mutants were expressed and purified following similar procedures.

B.1.3.1 CysLite

CysLite is an *E. coli* RecQ protein without maleimide liable surface exposed cysteines. Toward this goal, cysteines were systematically removed from a HisThrombRecQ base via consecutive rounds of mutagenesis. Single cysteine → serine mutants were generated at all positions except C397S_C400S_C403S, which was generated from a single primer pair. Serine codon usage was optimized for that of highly expressed enteric bacterial proteins. Consecutive rounds of mutagenesis were used to generate multiple mutations in a construct. HisTEVRecQ mutants were generated by digestion (NEB restriction enzymes and buffers), gel-purification (Qiagen MinElute Gel Extraction Kit after running on 1 % Agarose gels) and re-ligation (NEB T4 DNA Ligase) of the relevant constituent fragments into a HisTEV backbone vector. For example, HisTEVRecQ ZnCysOnly was generated by NdeI/BamHI HisTEV pET28b (no RecQ insert) and NdeI/SacII HisThromRecQ C43S_C56S_C94S_C111S_C150S_C243S_C351S.

B.1.3.2 Single-site, double-site labeling and separation-pin mutants

Using a CysLite mutant (HisThromRecQ C94S_C351S or HisTEVRecQ C94S_C351S) that shows no reactivity with Alexa647-maliemde, as starting material, site-specific NNN → CYS mutants were generated. An additional NNN → CYS site was generated for double mutants. Separation pin mutants were introduced into HisTEVRecQ.

HisThromb	HisTEV
RecQwt (§, ζ)	RecQwt (§, χ, χ*)
C43S_1	HisTEV RecQ Zn-only (§) (C43-56-94-111-150-243-351S)
C56S_4	HisTEV RecQ No-Cys (§) (C43S-56-94-111-150-243-351-380-397-400-403S)
C94S_6 (§ ζ ζ*)	
C111S_3	C94S_C351S (§χ)
C150S_1	C94S_C351S_N65C (§*)
C243S_1	C94S_C351S_S97C
C351S_1 (§ ζ ζ*)	C94S_C351S_S207C
C380S_2 (§)	C94S_C351S_S245C (§*)
C397_400_403S_2 (§)	C94S_C351S_S258C (§)
C380_397_400_403S_3 (§)	C94S_C351S_N273C (§*)
C351_380_397_400_403S_3	C94S_C351S_Q445C
C43_56S_2	C94S_C351S_S465C
C43_56_94S_3	C94S_C351S_S579C (§)
C43_56_94_111S_3	C94S_S579C (§)
C43_56_94_111_150S_2	C94S_C351S_N65C_S245C (§**)
C43_56_94_111_150_243S_5	C94S_C351S_N65C_S258C (§)
C94_351S_3 (§)	C94S_C351S_N65C_N273C (§**)
C150_243S_4	

Other	TGase
pET28b (No RecQ insert)	HisThromTEVTGase RecQwt pET15b (§,χχ*)

Table B.1 List of RecQ DNA constructs
All backbone vectors are pET28b unless otherwise noted.
§ Purified protein
§ Protein was insoluble or did not express
χ TEV cleavage of His-tag
ζ Thrombin cleavage (ΔHRDC)
* Alexa647-maliemide or Alexa647-cadaverine labeling
**Double labeling with Cy3B-maliemide/Atto647N-maliemide

Primer Name	Primer Sequence (5′ to 3′)
Sequencing Primers	
Primer_R1 (Forward) (Novagen T7 promoter primer #69348-3)	TAATACGACTCACTATAGGGG
Primer_R2 (Reverse)	CGTGGCCCCATTGGGAG
Primer_R3(Forward)	GGAATTAGCGCGGCGGC
Primer_R4(Reverse)	CATCGCTTCCGCGGGC
Primer_R5(Forward)	CGTGGTGCTAATAACCAGC
Primer_R6 (Reverse) (Novagen T7 terminator primer #69337-3)	CCGCTGAGCAATAACTAGC
CysLite Primers	
Throm → TEV Round 1	5′-CATCATCACAGCAGCGGCCTGTACTTCCAGGGCAGCCATATGGCGCAGGC-3′ 3′-GTAGTAGTGTCGTCGCCGGACATGAAGGTCCCGTCGGTATACCGCGTCCG-5′
Throm → TEV Round 2	5′-CAGCCATCATCATCATCATCACAGCGAAAACCTGTACTTCCAGGGCAGCCATATG-3′ 3′-GTCGGTAGTAGTAGTAGTAGTGTCGCTTTTGGACATGAAGGTCCCGTCGGTATAC-5′
C43S_G188C_S	TTTCCGGCCGCGATTCCCTCGTCGTC
C43S_G188C_A	GACGACGAGGGAATCGCGGCCGGAAA
C56S_G227C_S	GGCGGAAAATCCCTTTCCTATCAAATCCCTGCC
C56S_G227C_A	GGCAGGGATTTGATAGGAAAGGGATTTTCCGCC
C94S_G341C_S	CGTGGCGGCGGCGTCCCTTAACTC
C94S_G341C_A	GAGTTAAGGGACGCCGCCGCCACG
C111S_G392C_S	TGAAGTGATGACAGGCTCCCGCACCGG
C111S_G392C_A	CCGGTGCGGGAGCCTGTCATCACTTCA
C150S_G509C_S	GTTGATGAAGCGCACTCTATCTCCCAATGGGGC
C150S_G509C_A	GCCCCATTGGGAGATAGAGTGCGCTTCATCAAC
C243S_G788C_S	CAGGCATTATCTACTCCAACAGCCGCGCG
C243S_G788C_A	CGCGCGGCTGTTGGAGTAGATAATGCCTG
C351S_G1112C_S	CGTGGCTGCGCCGTTCTCTGGAAGAGAAG
C351S_G1112C_A	CTTCTCTTCCAGAGAACGGCGCAGCCACG
C380S_G1199C_S	TTGCCGAAGCGCAAACTTCCCGTCGTCTG
C380S_G1199C_A	CAGACGACGGGAAGTTTGCGCTTCGGCAA

C397/400/403S_S	CAGGAGCCGTCCGGGAACTCCGATATCTCCCTCGATC
C397/400/403S_A	GATCGAGGGAGATATCGGAGTTCCCGGACGGCTCCTG

Site-specific labeling Primers

N65C_S	CTATCAAATCCCTGCCTTATTGCTA
N65C_A	CACGGTAAGGCCGCATAGCAATAA
S97C_S	GGCGGCGTGCCTTAACTGCACGCA
S97C_A	CGGGTTTGCGTGCAGTTAAGGCAC
S207C_S	CGCTGATTCAAATCAGCTGTTTTGA
S207C_A	ATTCGGACGGTCAAAACAGCTGATT
S245C_S	GCATTATCTACTGCAACTGCCGCG
S245C_A	TTCTACTTTCGCGCGGCAGTTGCA
S258C_S	CGCTGCGCGCCTGCAATGCAAGGG
S258C_A	CTAATTCCCTTGCATTGCAGGCGC
N273C_S	CCTATCATGCCGGGCTGGAAAATT
N273C_A	TCGGCGCGAACACAATTTTCCAGC
Q445C_S	GAAGTGATTCGTGGTGCTAATAACT
Q445C_A	CCATAGTCGCGGATACGGCAGTTA
S465C_S	TCTATGGCATGGGCCGTGATAAAT
S465C_A	AATGTICATGGCATTIATCACGGCC
S579C_S	GCCAGCGAAATGCTCTGCGTIAAC
S579C_A	CAACGCCGTTAACGCAGAGCATTT

Separation Pin Primers

W154L_S	GATGAAGCGCACTGTATCTCCCAACTGGGCCACGATTTC
W154L_A	GAAATCGTGGCCCAGTTGGGAGATACAGTGCGCTTCATC
F158L_S	CCCAATGGGGCCACGATCTGCGCCCGGAA
F158L_A	TTCCGGGCGCAGATCGTGGCCCCATTGGG
F221L_S	CTACATGCTGATGGAGAAGCTGAAACCGCTCGATCAGTTGA
F221L_A	TCAACTGATCGAGCGGTTTCAGCTTCTCCATCAGCATGTAG
H489L_S	ACGCAAAATATTGCCCAGCTGTCTGCCCTACAACTGACAG
H489L_A	CTGTCAGTTGTAGGGCAGACAGCTGGGCAATATTTTGCGT

Other	
dT_{54} ssDNA	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

Table B.2 Primers for sequencing and mutagenesis
Sequencing primers: Six sequencing primers were used to verify each construct. Primers were ordered free of salt at 50 nmol scale from IDT unless otherwise specified. Sequencing direction is noted in parenthesis.
CysLite Primers: Primers used in conjunction with the Stratagene QuikChange II XL site-directed mutagenesis kit to create HisTEV (Round1 and Round2 on HisTEV pET28b) and Cys → Ser mutations (all others on HisThromRecQ). CXXS (e.g. C43S) refers to the protein mutations; numbering with respect to wild-type *E. coli* RecQ. GXXC (e.g. G188C) refers to the DNA mutation; number with respect to HisThrom RecQ. S and A refer to the sense (forward) and anti-sense (reverse) strands respectively. Mutation sites highlighted in red. There primers were ordered free of salt at 50 nmol scale from Operon.
Site-specific labeling primers: Same nomenclature as CysLite Primers. Mutagenesis applied to CysLite vector that demonstrates negligible maleimide labeling.
Separation pin primers: Same nomenclature as CysLite Primers.
'_S' and '_A' refer to the sense and anti-sense primers respectively.

B.2 Fluorophore labeling

B.2.1 DNA labeling

Labeled and unlabeled oligos were purchased from either Operon or IDT (Table B.3). All labeled oligos were purchased with an Int Cy3 or Int Cy5 modification. Briefly, complementary oligos were annealed (1 h at 80 °C, followed by an 8 h linear cooling to 4 °C), and then gel purification in an unstained 1 % agarose gel in 1x TBE buffe. Bands, visualized by the natural absorbance of the Cyanine dyes, were excised and soaked O/N in 1x TE buffer. After removal of the agarose fragment, absorbance was checked and the DNA stored at −20 °C.

B.2.2 Protein labeling

B.2.2.1 Maleimide-conjugated fluorophore labeling of RecQ

RecQ labeling was performed with a variety of dyes, with both maleimide and TGase mediated cadaverine labeling schemes. For convenience, dye properties are outlined in Table B.4; chemical structures Cy3, Cy3B, Cy5, Alexa647 and Atto647N in Figure B.5.

Oligo Name	Oligo Sequence (5′ to 3′)
	Neuman Lab
Bottom 3′-biotin 50 Unlabeled	CTG GCA GGC CCA GTC CAG GTC GCT CTG TGT GCG TCA CGC CGG TGA GCG TG/3BioTEG/
Cy3 Bottom 3′-biotin 50 nt	CTG GCA GGC CCA GTC CAG GTC GCT C/iCy3/TG TGT GCG TCA CGC CGG TGA GCG TG/3BioTEG/
Cy3 Bottom 5′-biotin 50 nt	/5BioTEG/CTG GCA GGC CCA GTC CAG GTC GCT C/iCy3/TG TGT GCG TCA CGC CGG TGA GCG TG
Cy3 Bottom 3′-biotin 75 nt	CTG GCA GGC CCA GTC CAG GTC GCT C/iCy3/TG TGT GCG TCA CGC CGG TGA GCG TGT TTT TTT TTT TTT TTT TTT TTT /3BioTEG/
Cy5 Top	/5Cy5/GAG CGA CCT GGA CTG GGC CTG CCA G
Top Unlabeled	GAG CGA CCT GGA CTG GGC CTG CCA G
	Kapanidis Lab
DNA G326 (Amino dT)	/5BioTEG/TGA ACC AGG AGC CCG GAT GAT CCT GAC GAC GGA GAC CGC CGX CGT CGA CAA GCC GGC CGA CT
JH_DNA2	ACG CGG CGG TCT CCG TCG TCA GGA TCA TCC GGG CTC CTG GTT CA
FRET Standards	
STD45T-NH2B	TAA ATC TAA AGT AAC ATA AGG TAA CAT AAC GTA AGC TCA XTC GCG
STD45B-NH2dT13	CGC GAA TGA GCT XAC GTT ATG TTA CCT TAT GTT ACT TTA GAT TTA
STD45B-NH2dT18	CGC GAA TGA GCT TAC GTX ATG TTA CCT TAT GTT ACT TTA GAT TTA
STD45B-NH2dT23	CGC GAA TGA GCT TAC GTT ATG TXA CCT TAT GTT ACT TTA GAT TTA
STD45B-NH2dT28	CGC GAA TGA GCT TAC GTT ATG TTA CCT XAT GTT ACT TTA GAT TTA
STD45B-NH2dT33	CGC GAA TGA GCT TAC GTT ATG TTA CCT TAT GTX ACT TTA GAT TTA
STD45B-NH2dT38	CGC GAA TGA GCT TAC GTT ATG TTA CCT TAT GTT ACT TXA GAT TTA

Table B.3 DNA oligos for single-molecule TIRF experiments
X is the amino-dT position for labeling.
In the case of DNA_G326, this position was labeled with Cy3B-NHS-ester. In the case of the FRET standards, X was labeled with Cy3B-NHS-ester (STD45T-NH2B) or Alexa647-NHS-ester (complement strands) to produce standards with FRET ranging from low (STD45T-NH2B/ STD45B-NH2dT13) to intermediate (STD45T-NH2B/ STD45B-NH2dT28) to high (STD45T-NH2B/ STD45B-NH2dT38).

B.2.2.2 TGase-mediated cadaverine-conjugated labeling of RecQ

Transglutaminase (TGase) labeling is an alternative to maleimide-based cysteine labeling. A N-terminal TGase tag (PKPQQFM) was cloned into the HisThromTEVRecQwt pET15b vector immediately following the TEV site (HisThromTEVTGaseRecQwt). Following purification, expression and TEV-mediated 6-His-tag cleavage, TGase was used to selective transfer Alexa647-cadaverine onto the Gln residue in the TGase-tag.

B.2.3 Single labeling of RecQ

All protein labeling was performed with Alexa647-maliemide (Molecular Probes). Briefly, protein is dialyzed (20 mM Tris HCl pH 8, 300 mM NaCl, 10 % glycerol) to remove sulfur compounds, such as DTT, that interfere with labeling. The dialyzed protein is then reduced with tris(2-chloroethyl) phosphate (TCEP) buffer (100 mM PO_4 pH 8, 10 mM TCEP). Since TCEP is known to interfere with maleimide labeling (albeit much less so that sulfer containing compounds such as DTT and BME),[217] it is removed via desalting (GE PD-10 desalting column with 100 mM PO4 pH8). In a septum sealed glass vial under nitrogen, DMSO dissolved Alexa647-maleimide and protein (2–30:1) are incubated at R/T for 5 hr, followed by 4 °C O/N incubation. 100 mM BME at R/T for 2 h is added to quench the reaction. Dye, labeled and unlabeled proteins are separated via FPLC (GE Superdex Peptide HR 10/30 or GE Superdex 200 16/60 column). Fractions are checked by A_{280}/A_{650} absorbance and SDS-PAGE before pooling. Alexa647/RecQ stoichiometry is determined by A_{280}/A_{650} absorbance, Bradford and bicinchoninic acid assay (BCA) assays. The same procedure was used to label non-specifically and site-specifically. A RecQ mutant is considered site-specifically labeled if its control construct (native residue instead of cysteine) has a negligible A_{650} at 6:1 Alexa647-maliemide:RecQ after gel filtration.

B.2.4 Double labeling of RecQ

Double labeling was performed with Cy3B-maliemide (GE) and either Alexa647-maliemide (Molecular Probes) or Atto647N-maliemide (Sigma). Dialysis and reduction are identical to that of the single labeling procedure. Dye1 and protein (2:1) are incubated for 1 h at R/T. Dye2 and protein (8:1) are incubated O/N at R/T. The reaction is quenched with 100 mM BME. Following

separation of free dye via gel filtration, fractions are checked at $A_{280}/A_{550}/A_{650}$ absorbance and SDS-PAGE before pooling. Before visualization with coomassie, gels are checked for in-gel Förster resonance energy transfer (FRET) on a FluorChem Q gel imager (Filter set Ex:Cy3/Em:Cy5, with Ex:Cy3/Em:Cy3 and Ex:Cy5/Em:Cy5 as controls). A RecQ mutant was considered doubly labeled if in-gel FRET was observed, FRET was observed via ensemble fluorometry and labeling stoichiometry was better than 1:0.5:0.5 (RecQ : Cy3B : Atto647N or Alexa647).

B.3 Ensemble assays

B.3.1 ATPase assay

ATPase activity is determined using an absorbance based NADH coupled ATP assay consisting of imaging buffer (Appendix B.7.2.1) supplemented with 200 µM NADH, 2 % PK/LDH (pyruvate kinase / lactate dehydrogenase), 1 mM PEP (phosphoenolpyruvate), 1 mM ATP (Figure B.2).

The procedure is as follows:

(i) Establish a baseline for NADH turnover by monitoring A_{340} in the absence of enzyme.

(ii) Measure basal enzyme activity by monitoring A_{340} upon the addition of 200–1000 nM of RecQ.

(iii) Measure DNA-stimulated enzyme activity by monitoring A_{340} upon the addition of 100 µM dT_{54}.

Crude basal (2) and DNA-stimulated (3) ΔA_{340} are corrected by subtracting baseline NADH turnover (1). Using this corrected ΔA_{340}, ATP turnover (k_{cat}) is determined as follows:

$$k_{cat} = \frac{\Delta A_{340}^{\text{corrected}}}{[\text{Enzyme}] \times \varepsilon_{\text{NADH}}^{340\,\text{nm}} \times \ell} \left(\frac{1}{s}\right), \tag{B.1}$$

where ℓ is the cross-section of the flowcell (usually 1 cm), ε is the extinction coefficient of NADH ($\varepsilon_{\text{NADH}}^{340\,\text{nm}} = 6220\ \text{M}^{-1}\,\text{cm}^{-1}$) and the enzyme concentration is determined by either BCA assay or A_{280} using $\varepsilon_{\text{RecQ}}^{280\,\text{nm}} = 48\,800\ \text{M}^{-1}\,\text{cm}^{-1}$ for RecQ, as calculated by Expasy ProtParam (tags and point-mutations are negligible). Assays were performed on a variety of comparable

[ii] A useful resource for ATP-turnover assays: `http://www.proweb.org/kinesin/Methods/ATPase_assay.html`.

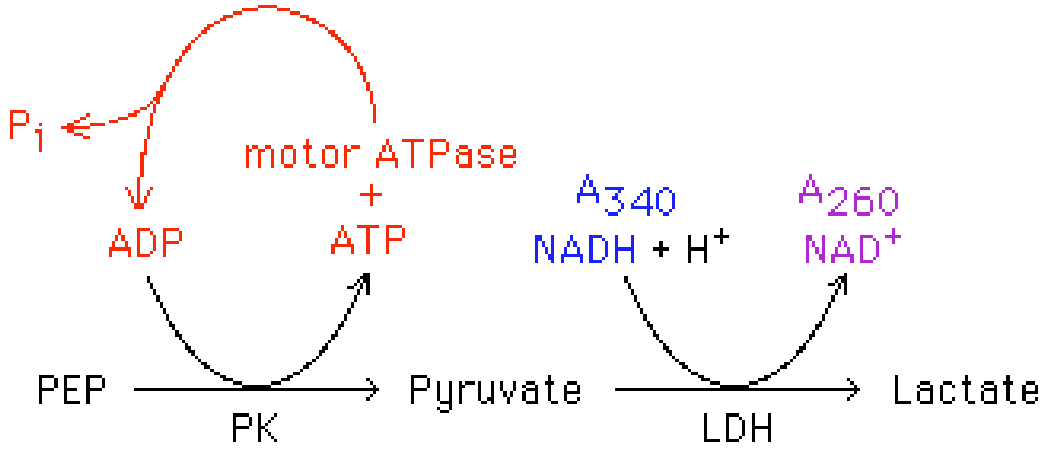


Figure B.2 Schematic for NADH coupled ATP-turnover assay
Enzyme-mediated ATP hydrolysis is linked to NADH turnover. NADH absorbance at 340 nm is monitored as a reporter of ATP turnover.[ii]

spectrophotometers: Agilent 8453, Beckman DU640 and HP8453.

B.3.2 Fluorometer assays

Fluorescence based assays are used to measure enzyme-DNA binding and ensemble FRET. All experiments were carried out on a PTI Fluorometer with two independently adjustable detection paths (separate polarizer and PMT). In anisotropy measurements, excitation light was vertically polarized (0°), with detection channels A ($F^{\perp}$/Horizontal/90°) and B ($F^{\parallel}$/Vertical/0°) polarized orthogonal to each other. In all other measurements, including protein induced fluorescent enhancement (PIFE) assays, excitation and detection channels were polarized identically (54.7°). Slit widths were set to 2 mm, and all measurements were collected with 1 nm step-size and 0.1 s integration time per step.

Anisotropy was determined by,

$$r = \frac{F^{\parallel} - F^{\perp}}{F^{\parallel} + 2F^{\perp}}, \tag{B.2}$$

where $F^{\parallel}$ detection is polarized parallel to the excitation light and $F^{\perp}$ is polarized perpendicular. In non-anisotropy assays, the counts for both detection channels were combined.

Unless otherwise stated, experiments are conducted in imaging buffer (Appendix B.7.2.1). Typically samples were excited at 500 nm and 600 nm, with emission spectra collected at 525–800 nm (Cy3, Cy3B) and 625–800 nm (Cy5, Alexa647, Atto647N) respectively.

B.4 Single-molecule experiments

B.4.1 Single-molecule TIRFM (Prism-type)

Full details of the single-molecule prism-type total internal reflection fluorescence microscopy (TIRFM) setup and flow cell preparation are described in Sarkar *et al.*[147]

Briefly, it is a Olympus IX71inverted microscope (Olympus, Melville, NY), with 60x numerical aperture 1.2 water-immersion objective (UPLSAPO60XW, Olympus) with additional 1.6x tube lens magnification, Back Illuminated and iXon EM 897 camera and a bracket mounted prism. Either a 532 nm (GCL-025-L-0.5 %, CrystaLaser, Reno, NV) or 650 nm continuous-wave circularly polarized laser of $\sim$3 mW are directed through the prism, generating an evanescent wave slightly larger than the viewing area of $\sim 150 \times 75$ mm[119] (penetration depth is 100–150 nm).

Our flowcell is amine functionalized quartz (Vectabond, Vector Laboratories, Burlingame, CA) with sandwich of biotinylated polyethylene glycol (PEG), strepavadin and biotinylated DNA (2 pM, which corresponds to 500–1000 bound molecules per field of view). Additional passivation is provided by 0.1 mg/ml bovine serum albumin (BSA). Before addition of biotinylated DNA, we take the additional step of photobleaching the slide with 30 mW 532 nm to reduce background signal from dust. Unless otherwise stated, experiments were carried out in imaging buffer (Appendix B.7.2.1), supplemented by an oxygen scavenger system (Appendix B.7.2.2). All reagents are flushed well in excess of our flow cell volume of $\sim$100 nL (4 cm $\times$ 2 cm $\times$ 0.12 mm). Experimental procedures vary from this point, but involve the introduction of various concentrations of RecQ and ATP.

RecQ concentrations up to 1 mM were used in unlabeled RecQ experiments. Excess background noise from acceptor labeled RecQ in solution limited me to concentrations of 10–50 nM for experiments using labeled protein. ATP concentrations assayed ranged from 1 µM to 10 mM. For non-saturating ATP concentrations (below 1 mM), a Creatine Kinase ATP regeneration system was used to clamp the concentration (Appendix B.7.2.3). This was used in addition to the aforementioned Oxygen scavenger system.

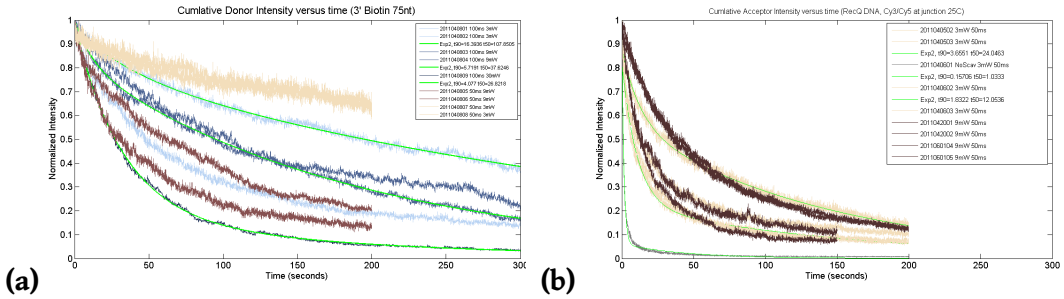


Figure B.3 D/D and D/A lifetimes by laser power (excitation/emission)
(a) Donor lifetime (ex:D / em:D) and **(b)** acceptor lifetime (ex:D / em:A)
Note that t_90 (90 % of fluorophores are still intact) is used instead of fluorescence lifetime. This metric was used since we cannot easily distinguish between unwinding, bleaching and dark-stating on our non-ALEX TIRFM setup.

B.4.1.1 Instrumentation controls

To ensure the best signal-to-noise ratio (SNR) possible, a detailed determination of the noise characteristics of the Neuman lab single-mocule TIRFM setup, described in Sarkar *et al*,[147] was required. Single-point mean variance (Back Illuminated iXon EM 897, bit-level 2^14) yielded a post multiplication gain of approximately 0.02 electrons per grey—level. All single-molecule results will be reported in electrons using this conversion factor. Dark and read noise were effectively zero. Total camera noise, presumably clock induced charge noise, was 1.4 e- (no light entering the system). Donor/donor and donor/acceptor noise from a Cy3 Bottom $3'$-biotin 50 nt/Cy5 Top preparation over 1 s (10 frames) were 3.5 (mean 2.2) and 4.4 (mean 2.8) e- respectively. When factoring in camera noise and assuming an EM multiplication gain factor $F^2 = 2$, we are operating at 1.5 2 times the shot noise limit.

In our hands, Cy3 and Cy5 t90 (90 % of fluorophores are still active; we use this metric because it is very difficult to differentiate between unwinding and bleaching in a non-alternating laser excitation (ALEX) setup) are 11 and 4 seconds respectively at 3 mW donor excitation with the aforementioned substrate (Figure B.3). As expected, SNR increases (data not shown) and lifetime decreases at higher excitation power.

B.4.2 Single-molecule ALEX-TIRFM (Objective-type)

Full details of the three-color (473, 532 and 635 nm) objective-type ALEX-TIRFM setup are described in Lee *et al.*[218] Only two channels (532 and 635 nm) were used.[iii].

#1.5 glass cover slips (22 × 40 mm) with mounted silicon gasket (4×, ∼25 µL wells) were used instead of flow-cells. The use of wells instead of flow-cells allows for more direct access and easier buffer exchange. Gaskets are reused and cleaned via iterative washing with MeOH, filtered H_2O and acetone. Glass cover slips are baked in high temperature furnace (heat 90 °C h^{-1}, 500 °C for 1 h, cool 150 °C h^{-1}) before use and not reused. Neutravidin (0.5 mg/ml in phosphate buffered saline (PBS), wash with PBS + 0.1 mg/ml BSA to remove) is used instead of Strepavadin. Objective-type and prism-type experimental procedures are otherwise identical.

B.5 Data processing and analysis

B.5.1 Extracting single-molecule time-series

Movies were recorded to raw tiff format with Andor software for the iXon camera. TwoTone[219] was used to extract single-molecule trajectories of acceptor and donor intensities.[219] While the freeGaussElipse point spread function (PSF) fitting feature is by far the most accurate, fixedGauss was used because it is significantly less computationally intensive (∼2 h for a 3000 frame 512 × 512 tiff stack on a 32-bit 2.8 GHz Intel Core2 Duo with 2 GB ram versus 12+ hours). FRET efficiency was calculated from TwoTone donor/acceptor intensities using Equation (10.1), with $\beta \approx 0.1$ and $\gamma^{\text{photobleach}} \approx 1$ empirically determined from control traces with neither RecQ nor ATP. A custom MATLAB suite on top of TwoTone was written for plotting and analysis of continuous donor excitation trajectories. Seneca, a Kapanidis lab software package, was used to visualize ALEX trajectories.

[iii] A more practical summary of the instrument is available from the Kapanidis lab webpage: `http://www.physics.ox.ac.uk/Users/kapanidis/Group/Methods.SMFMethods.html`

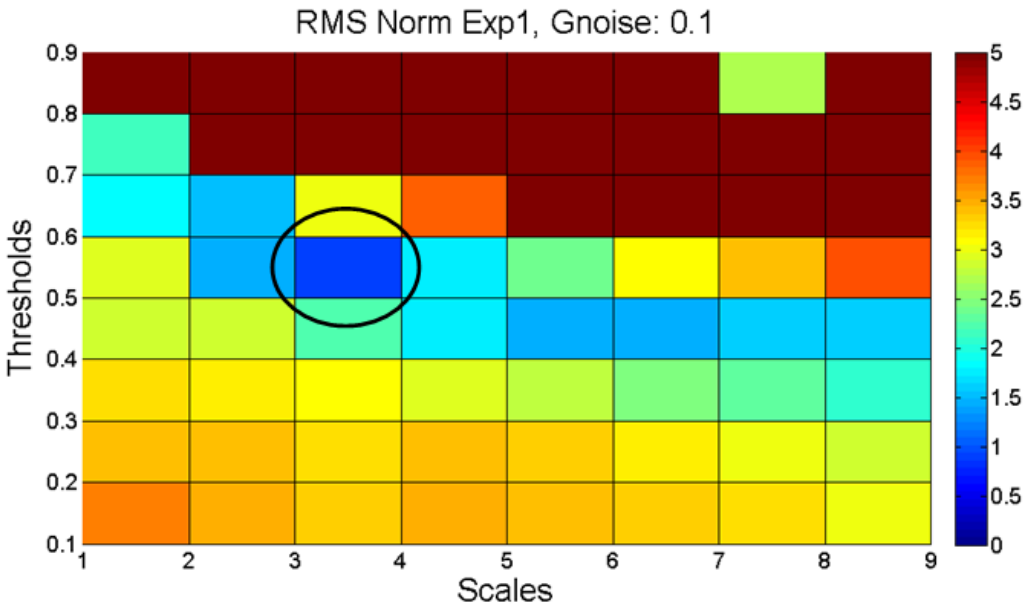


Figure B.4 Parameter tuning of edge detection using simulated data
Simulation parameters were as follows: two state exponentially distributed system with $\tau_1 = 10s$ and $\tau_2 = 10s$. States were defined as $E_1 = 0.1$ and $E_2 = 0.3$ with 0.1 Gaussian white noise overlaid. These were the approximate characteristics of the experimental signal. The lowest scoring parameters were used for all edge detection, where the scoring function was normalized RMS given by:

$$\mathrm{RMS}_{\mathrm{norm}} = \sqrt{\left(\frac{\tau^{\mathrm{low}}_{\mathrm{pred}} - \tau^{\mathrm{low}}_{\mathrm{actual}}}{\tau^{\mathrm{low}}_{\mathrm{actual}}}\right)^2 + \left(\frac{\tau^{\mathrm{high}}_{\mathrm{pred}} - \tau^{\mathrm{high}}_{\mathrm{actual}}}{\tau^{\mathrm{high}}_{\mathrm{actual}}}\right)^2}.$$

B.5.2 Kinetic data extraction

For extraction of kinetic information from single-molecule studies of static binding/unbinding of acceptor labeled RecQ to donor labeled DNA, a 1D Gaussian Kernel edge detector[iv] was used. The accuracy of the edge detection was sensitive to the window size and threshold parameters; therefore, these were calibrated by simulation, a simple two state model with exponentially distributed dwell times plus variable Gaussian white noise (Figure B.4). This implementation was similar to the two state model examined by Shi, Gafni and Steel.[220]

[iv] 1D edge-detector available from: `http://www.cs.unc.edu/Enanowork/cismm/download/edgedetector/index.html`

B.6 Compendia

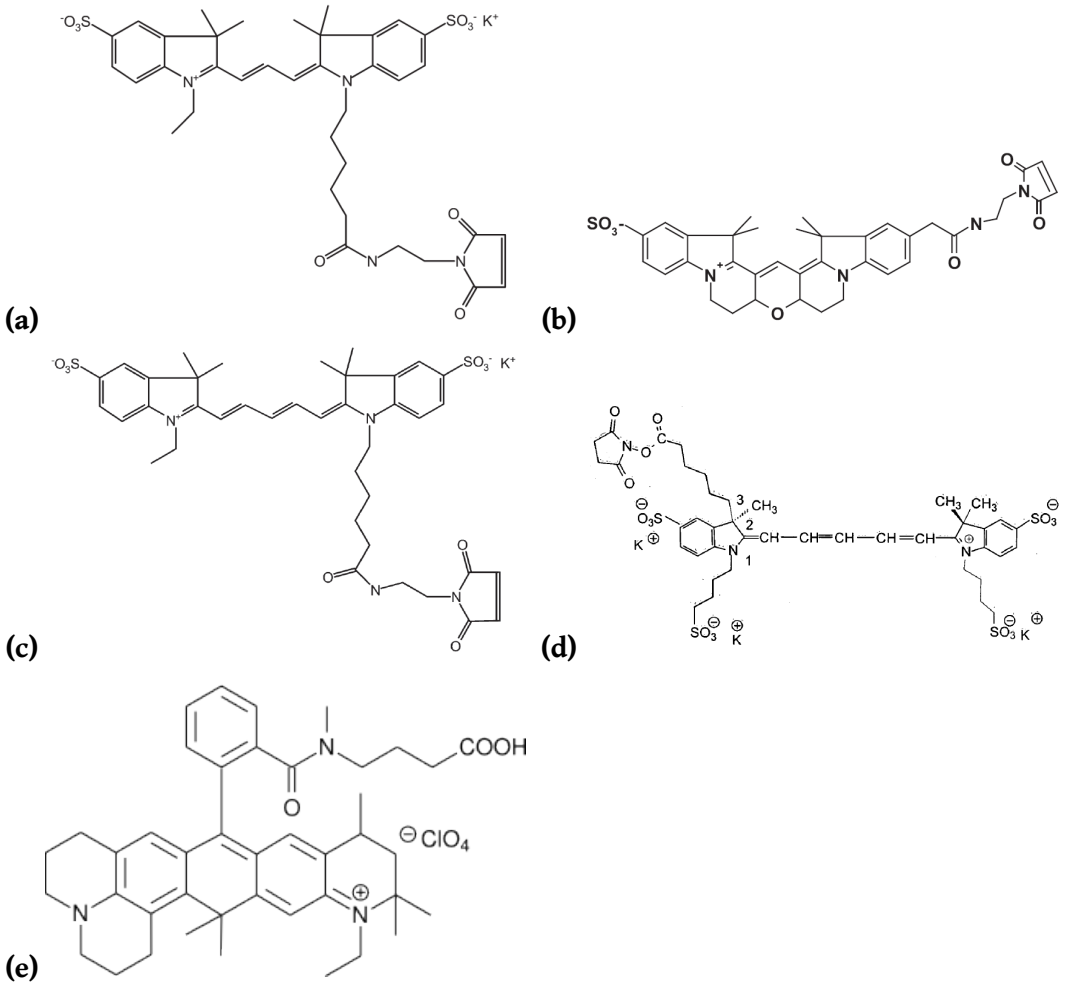


Figure B.5 Chemical structures of Cy3/Cy3B/Cy5, Alexa647 and Atto647N
(a) Cy3[221]
(b) Cy3B[222]
(c) Cy5[223]
(d) Alexa647[226]
(e) Atto647N (Carboxyl-Functionalized)[227]
Dyes are maleimide functionalized unless otherwise stated.

Dye (maleimide)	MW (Da)	Abs (nm)	Em (nm)	Extinction Coefficient ($M^{-1}\,cm^{-1}$)	QY (a.u.)
Cy3[221]	791	550	570	150,000	0.15
Cy3B[222]	796	558	572, 620	130,000	$>$ 0.70
Cy5[223]	817	649	670	150,000	0.28
Alexa647[142]	$\sim$1300	652	668	239,000	0.33
Atto647N[224]	868	644	669	150,000	0.65

Table B.4 Compilation of fluorescent dyes
Absorbance (Abs), excitation (Ex) and emission (Em) are for maxima. Quantum yield (QY) relative to fluorescein. Experimental conditions differ slightly, but are roughly comparable. Succinimidyl ester or anti-body conjugated, in buffer (PBS) or methanol (extinction coefficients).

Protein	Domain	Sub-domain (if available)	Location (amino acid index)
RecQ (610 aa)	RecA	H1	1–208
		H2	209–340
	RQC	Zn	341–406
		WH	407–516
		HRDC	517–610
BLM (1417 aa)	RecA		642–993
	RQC	Zn	642–993
		WH	1069–1189
		HRDC	1190–1290
WRN (1432 aa)	RQC	WH	956–1064 (Protein Data Bank (PDB): 3AAF)
		HRDC	1142–1242
RecQL1 (649 aa)			
RecQL4 (1208 aa)			
SGS1 (1447 aa)	RecA		400–1268 includes entire RecA domain and portions of others.
NS3 (631 aa)	RecQ		189–430

Table B.5 Compilation of RecQ-family domain locations
H1/H2: Helicase domains 1 and 2. These are sometimes grouped and labeled the RecA domain. The ATP-binding domain containing the Walker A/B motif is located in the cleft between these two domains.
Zn: Zn-binding domain. A zinc-metal coordinating cysteine tetrad is located on this domain. This is also the location of the helix-turn-helix on which the Zn-pin residues.
WH: Winged Helix domain. The WHβ-hairpin is located on this domain, with the WH-pin residue located on the tip of the hairpin
HRDC domain: Helicase and RNaseD C-terminal domain. This domain is linked to the N-terminal domains by a 10–30 aa flexible linker on some RecQ-family helicases. The Zn, winged-helix (WH) and HRDC domains are together referred to as the RecQ C-terminal (RQC) domain.

	PDB	Description
RecQ (*E. coli*)	1OWY	N-terminal (1-516) without ATP-analogue bound
	1OYY	. . . with ATP-analogues bound
	1WUD	Core of HRDC domain (530–606)
BLM (Human)	2RRD	Core of HRDC domain
WRN (Human)	2DGZ	Winged-helix domain
	2E1E	Core of HRDC domain
	2E1F	Core of HRDC domain
	3AAF[155]	Winged-helix domain in complex with dsDNA
RecQL1 (Human)	2V1X	N-term domains
	2WWY	N-term domains in complex with 3′-duplex DNA
SGS1 (Yeast)	1D8B	Core of HRDC domain
NS3 (Hepatitis C)	1A1V[157]	N-term domains in complex with single-stranded DNA (ssDNA)
	3KQL[212]	ADP·AlF_4, dT_6 ssDNA. Starts at RecA domains. Enzyme does not posses HRDC.
	3KQN	ADP·BeF_3, dT_6 ssDNA
	3KQU	ADP·BeF_3, dT_{12} ssDNA
	3KQH	dA6 ssDNA
	3KQK	dT_6 ssDNA
Hel308	2P6R[225]	N-term in complex with 3′-duplex DNA
(*A. fulgidus*)	2P6U[225]	Apo
PcrA	3PJR	N-term in complex with 3′-duplex DNA

Table B.6 Compilation of RecQ-family PDB identifiers

B.7 List of buffers

All buffers are sterile filtered to 0.22 µM. This is particularly important for single-molecule reagents.

B.7.1 Protein purification

B.7.1.1 IMAC buffer A

- 20 mM Imidazole, pH 7.5 (w/ HCl)
- 10 % Glycerol
- 20 mM Tris HCl, pH 7.5
- 300 mM NaCl
- 1 mM BME

B.7.1.2 IMAC buffer B

- 300 mM Imidazole, pH 7.5 (w/ HCl)
- 10 % Glycerol
- 20 mM Tris HCl, pH 7.5
- 300 mM NaCl
- 1 mM BME

B.7.1.3 FPLC buffer C

- 10 % Glycerol
- 20 mM Tris HCl, pH 7.5
- 300 mM NaCl
- 1 mM DTT
- 1 mM EDTA

B.7.2 Single-molecule experiments

B.7.2.1 1x Imaging buffer

- 25 mM Tris HCl, pH 7.5 (50 mM was occasionally used)
- 50 mM NaCl
- 1 mM $MgCl_2$
- 0.1 mM DTT

B.7.2.2 1x O2 scavenger buffer

- 1 mM Trolox (10 µL from 100 mM Trolox stock)
- 0.1 mg/ml (10 µL of 10 U/mL Glucose oxidase)
- 0.025 mg/ml (250 U/mL) Catalase 5 µL, 200x Catalase
- 4 % w/w Glucose

B.7.2.3 1x ATP regeneration buffer

- 2 mM Creatine phosphate
- 0.1 mg/ml Creatine kinase

B.7.2.4 Vectabond solution

- 9 µL Vectabond (Vector Laboratories, Cat No: SP-1800)
- 450 µL Acetone

B.7.2.5 Neuman 100x PEG stock

- 100 mg, mPEG-SS, MW 5000 (SunBio USA Methoxy PEG succinimidyl succinate-5k, Cat No: P1SS-5)
- 1 mg, Biotin PEG NHS ester, MW 5000 (JenKem USA, Item Code**:** A5027-5)
- 100 mM $NaHCO_3$, pH 8.2 epp, 1 mL

B.7.2.6 Kapanidis 100x PEG stock

- 20 µL Sodium Carbonate, 100 mM pH 8.3
- 4 mg NHS-PEG (Lysan, mPEG SVA MW5000, mPEG-Succinimidyl valerate MW 5000)
- 50 µg biotin-PEG-NHS MW5000 (Lysan Biotin-PEG-SC)
- Eep 450 µL H_2O

B.7.2.7 25x Streptavidin stock

- 5 mg/mL Streptavidin (5 mg Streptavidin lyophilized powder, Invitrogen S-888)
- 50 % Glycerol 500 µL, 100 % Stock

References

[1] H. Jacobson and W. H. Stockmayer, "Intramolecular Reaction in Polycondensations. I. The Theory of Linear Systems," *The Journal of Chemical Physics*, vol. 18, pp. 1600–1606, Dec. 1950.

[2] M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu, "Crystal Structure of the Lactose Operon Repressor and Its Complexes with DNA and Inducer," *Science*, vol. 271, pp. 1247–1254, Mar. 1996.

[3] D. B. Nikolov, H. Chen, E. D. Halay, A. Hoffman, R. G. Roeder, and S. K. Burley, "Crystal structure of a human TATA box-binding protein/TATA element complex," *Proceedings of the National Academy of Sciences*, vol. 93, pp. 4862–4867, May 1996.

[4] T. J. Richmond and C. A. Davey, "The structure of DNA in the nucleosome core," *Nature*, vol. 423, pp. 145–150, May 2003.

[5] J. Widom, "Role of DNA sequence in nucleosome stability and dynamics," *Quarterly Reviews of Biophysics*, vol. 34, no. 03, pp. 269–324, 2001.

[6] A. R. Haeusler, K. A. Goodson, T. D. Lillian, X. Wang, S. Goyal, N. C. Perkins, and J. D. Kahn, "FRET studies of a landscape of Lac repressor-mediated DNA loops," *Nucleic Acids Research*, vol. 40, pp. 4432–4445, May 2012.

[7] N. A. Davis, S. S. Majee, and J. D. Kahn, "TATA Box DNA Deformation with and without the TATA Box-binding Protein," *Journal of Molecular Biology*, vol. 291, pp. 249–265, Aug. 1999.

[8] J. Shimada and H. Yamakawa, "Ring-closure probabilities for twisted wormlike chains. Application to DNA," *Macromolecules*, vol. 17, pp. 689–698, Apr. 1984.

[9] D. Thirumalai and B.-Y. Ha, "Statistical Mechanics of Semiflexible Chains: A Meanfield Variational Approach," *arXiv:cond-mat/9705200*, May 1997.

[10] A. K. Mazur, "Wormlike Chain Theory and Bending of Short DNA," *Physical Review Letters*, vol. 98, p. 218102, May 2007.

[11] N. B. Becker, A. Rosa, and R. Everaers, "The radial distribution function of worm-like chains," *The European Physical Journal E*, vol. 32, pp. 53–69, May 2010.

[12] D. Shore and R. L. Baldwin, "Energetics of DNA twisting: I. Relation between twist and cyclization probability," *Journal of Molecular Biology*, vol. 170, pp. 957–981, Nov. 1983.

[13] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, and A. Vologodskii, "Cyclization of short DNA fragments and bending fluctuations of the double helix," *Proceedings of the National Academy of Sciences*, vol. 102, pp. 5397–5402, Apr. 2005.

[14] T. E. Cloutier and J. Widom, "Spontaneous Sharp Bending of Double-Stranded DNA," *Molecular Cell*, vol. 14, pp. 355–362, May 2004.

[15] R. Vafabakhsh and T. Ha, "Extreme Bendability of DNA Less than 100 Base Pairs Long Revealed by Single-Molecule Cyclization," *Science*, vol. 337, pp. 1097–1101, Aug. 2012.

[16] T. T. Le and H. D. Kim, "Probing the elastic limit of DNA bending," *Nucleic Acids Research*, vol. 42, pp. 10786–10794, Dec. 2014.

[17] A. A. Podtelezhnikov, C. Mao, N. C. Seeman, and A. Vologodskii, "Multimerization-Cyclization of DNA Fragments as a Method of Conformational Analysis," *Biophysical Journal*, vol. 79, pp. 2692–2704, Nov. 2000.

[18] D. Demurtas, A. Amzallag, E. J. Rawdon, J. H. Maddocks, J. Dubochet, and A. Stasiak, "Bending modes of DNA directly addressed by cryo-electron microscopy of DNA minicircles," *Nucleic Acids Research*, vol. 37, pp. 1–12, May 2009.

[19] H. Qu, Y. Wang, C.-Y. Tseng, and G. Zocchi, "Critical Torque for Kink Formation in Double-Stranded DNA," *Physical Review X*, vol. 1, p. 021008, Nov. 2011.

[20] H. Qu and G. Zocchi, "The complete bending energy function for nicked DNA," *Europhysics Letters*, vol. 94, p. 18003, Apr. 2011.

[21] H. Qu, Y. Wang, C.-Y. Tseng, and G. Zocchi, "Nonlinear Elasticity of DNA Bending," *Biophysical Journal*, vol. 102, pp. 275a–276a, Jan. 2012.

[22] A. P. Fields, E. A. Meyer, and A. E. Cohen, "Euler buckling and nonlinear kinking of double-stranded DNA," *Nucleic Acids Research*, vol. 41, pp. 9881–90, Nov. 2013.

[23] F. H. C. Crick and A. Klug, "Kinky helix," *Nature*, vol. 255, no. 5509, pp. 530–533, 1975.

[24] J. Yan and J. F. Marko, "Localized Single-Stranded Bubble Mechanism for Cyclization of Short Double Helix DNA," *Physical Review Letters*, vol. 93, p. 108108, Sept. 2004.

[25] J. S. Mitchell, C. A. Laughton, and S. A. Harris, "Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA," *Nucleic Acids Research*, vol. 39, pp. 3928–3938, May 2011.

[26] A. Vologodskii and M. D. Frank-Kamenetskii, "Strong bending of the DNA double helix," *Nucleic Acids Research*, vol. 41, Aug. 2013.

[27] P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, J. Widom, C. Dekker, and P. C. Nelson, "High flexibility of DNA on short length scales probed by atomic force microscopy," *Nature Nanotechnology*, vol. 1, pp. 137–141, Nov. 2006.

[28] A. K. Mazur and M. Maaloum, "DNA Flexibility on Short Length Scales Probed by Atomic Force Microscopy," *Physical Review Letters*, vol. 112, p. 068104, Feb. 2014.

[29] A. Mazur, "Anharmonic Torsional Stiffness of DNA Revealed under Small External Torques," *Physical Review Letters*, vol. 105, no. 1, 2010.

[30] P. Wiggins, R. Phillips, and P. Nelson, "Exact theory of kinkable elastic polymers," *Physical Review E*, vol. 71, Feb. 2005.

[31] V. B. Zhurkin, Y. P. Lysov, and V. I. Ivanov, "Anisotropic flexibility of DNA and the nucleosomal structure," *Nucleic Acids Research*, vol. 6, pp. 1081–1096, Mar. 1979.

[32] A. A. Travers, "The structural basis of DNA flexibility," *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 362, pp. 1423–1438, July 2004.

[33] D. M. Lilley, "Kinking of DNA and RNA by base bulges.," *Proceedings of the National Academy of Sciences*, vol. 92, pp. 7140–7142, Aug. 1995.

[34] W. Gobush, "Statistical Mechanics of Wormlike Chains. I. Asymptotic Behavior," *The Journal of Chemical Physics*, vol. 57, no. 7, p. 2839, 1972.

[35] H. Yamakawa, "Statistical Mechanics of Wormlike Chains. II. Excluded Volume Effects," *The Journal of Chemical Physics*, vol. 57, no. 7, p. 2843, 1972.

[36] Q. Du, A. Kotlyar, and A. Vologodskii, "Kinking the double helix by bending deformation," *Nucleic Acids Research*, vol. 36, pp. 1120–1128, Mar. 2008.

[37] D. A. Sivak and P. L. Geissler, "Consequences of local inter-strand dehybridization for large-amplitude bending fluctuations of double-stranded DNA," *The Journal of Chemical Physics*, vol. 136, p. 045102, Jan. 2012.

[38] C. Maffeo, J. Yoo, J. Comer, D. B. Wells, B. Luan, and A. Aksimentiev, "Close encounters with DNA," *Journal of Physics: Condensed Matter*, vol. 26, p. 413101, Oct. 2014.

[39] W. G. Noid, "Perspective: Coarse-grained models for biomolecular systems," *The Journal of Chemical Physics*, vol. 139, p. 090901, Sept. 2013.

[40] J. Yoo and A. Aksimentiev, "In situ structure and dynamics of DNA origami determined through molecular dynamics simulations," *Proceedings of the National Academy of Sciences*, vol. 110, pp. 20099–20104, Dec. 2013.

[41] J. P. K. Doye, T. E. Ouldridge, A. A. Louis, F. Romano, P. Šulc, C. Matek, B. E. K. Snodin, L. Rovigatti, J. S. Schreck, R. M. Harrison, and W. P. J. Smith, "Coarse-graining DNA for simulations of DNA nanotechnology," *Physical Chemistry Chemical Physics*, vol. 15, pp. 20395–20414, Nov. 2013.

[42] M. Peyrard, "Nonlinear dynamics and statistical physics of DNA," *Nonlinearity*, vol. 17, p. R1, Mar. 2004.

[43] J. SantaLucia, "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics," *Proceedings of the National Academy of Sciences*, vol. 95, pp. 1460–1465, Feb. 1998.

[44] J. SantaLucia and D. Hicks, "The thermodynamics of DNA structural motifs," *Annual Review of Biophysics and Biomolecular Structure*, vol. 33, pp. 415–440, June 2004.

[45] C. A. Brackley, A. N. Morozov, and D. Marenduzzo, "Models for twistable elastic polymers in Brownian dynamics, and their implementation for LAMMPS," *The Journal of Chemical Physics*, vol. 140, p. 135103, Apr. 2014.

[46] A. Savelyev and G. A. Papoian, "Chemically accurate coarse graining of double-stranded DNA," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 20340–20345, Nov. 2010.

[47] O. Gonzalez, D. Petkevičiūtė, and J. H. Maddocks, "A sequence-dependent rigid-base model of DNA," *The Journal of Chemical Physics*, vol. 138, p. 055102, Feb. 2013.

[48] R. C. DeMille, T. E. Cheatham, and V. Molinero, "A Coarse-Grained Model of DNA with Explicit Solvation by Water and Ions," *The Journal of Physical Chemistry B*, vol. 115, pp. 132–142, Jan. 2011.

[49] F. Trovato and V. Tozzini, "Supercoiling and Local Denaturation of Plasmids with a Minimalist DNA Model," *The Journal of Physical Chemistry B*, vol. 112, pp. 13197–13200, Oct. 2008.

[50] T. A. K. Iv, N. Rathore, D. C. Schwartz, and J. J. d. Pablo, "A coarse grain model for DNA," *The Journal of Chemical Physics*, vol. 126, p. 084901, Feb. 2007.

[51] A.-M. Florescu and M. Joyeux, "Thermal and mechanical denaturation properties of a DNA model with three sites per nucleotide," *The Journal of Chemical Physics*, vol. 135, p. 085105, Aug. 2011.

[52] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, "Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model," *The Journal of Chemical Physics*, vol. 134, pp. 085101–085101–22, Feb. 2011.

[53] T. E. Ouldridge, *Coarse-Grained Modelling of DNA and DNA Self-Assembly*. Springer, July 2012.

[54] C. Matek, T. E. Ouldridge, A. Levy, J. P. K. Doye, and A. A. Louis, "DNA Cruciform Arms Nucleate through a Correlated but Asynchronous Cooperative Mechanism," *The Journal of Physical Chemistry B*, vol. 116, pp. 11616–11625, Sept. 2012.

[55] P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. K. Doye, and A. A. Louis, "Sequence-dependent thermodynamics of a coarse-grained DNA model," *The Journal of Chemical Physics*, vol. 137, pp. 135101–135101–14, Oct. 2012.

[56] F. Romano, D. Chakraborty, J. P. K. Doye, T. E. Ouldridge, and A. A. Louis, "Coarse-grained simulations of DNA overstretching," *The Journal of Chemical Physics*, vol. 138, pp. 085101–085101–10, Feb. 2013.

[57] C. Matek, T. E. Ouldridge, J. P. K. Doye, and A. A. Louis, "Plectoneme tip bubbles: Coupled denaturation and

writhing in supercoiled DNA," *Scientific Reports*, vol. 5, Jan. 2015.

[58] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, "DNA Nanotweezers Studied with a Coarse-Grained Model of DNA," *Physical Review Letters*, vol. 104, p. 178101, Apr. 2010.

[59] J. Russo, P. Tartaglia, and F. Sciortino, "Reversible gels of patchy particles: Role of the valence," *The Journal of Chemical Physics*, vol. 131, p. 014504, July 2009.

[60] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, "THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method," *Journal of Computational Chemistry*, vol. 13, pp. 1011–1021, Oct. 1992.

[61] J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, "Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations," *Journal of Chemical Theory and Computation*, vol. 3, pp. 26–41, Jan. 2007.

[62] S. Whitelam and P. L. Geissler, "Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles," *The Journal of Chemical Physics*, vol. 127, p. 154101, Oct. 2007.

[63] G. Torrie and J. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *Journal of Computational Physics*, vol. 23, no. 2, pp. 187–199, 1977.

[64] S. Geggier and A. Vologodskii, "Sequence dependence of DNA bending rigidity," *Proceedings of the National Academy of Sciences*, vol. 107, no. 35, pp. 15421–15426, 2010.

[65] F. Lankaš, R. Lavery, and J. H. Maddocks, "Kinking Occurs during Molecular Dynamics Simulations of Small DNA Minicircles," *Structure*, vol. 14, pp. 1527–1534, Oct. 2006.

[66] J. Spiriti, H. Kamberaj, A. M. R. de Graff, M. F. Thorpe, and A. van der Vaart, "DNA Bending through Large Angles Is Aided by Ionic Screening," *Journal of Chemical Theory and Computation*, vol. 8, pp. 2145–2156, June 2012.

[67] M. T. Woodside, W. M. Behnke-Parks, K. Larizadeh, K. Travers, D. Herschlag, and S. M. Block, "Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins," *Proceedings of the National Academy of Sciences*, vol. 103, pp. 6190–6195, Apr. 2006.

[68] J. M. Huguet, C. V. Bizarro, N. Forns, S. B. Smith, C. Bustamante, and F. Ritort, "Single-molecule derivation of salt dependent base-pair free energies in DNA," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 15431–15436, Aug. 2010.

[69] J. C. Wang and N. Davidson, "Thermodynamic and kinetic studies on the interconversion between the linear and circular forms of phage lambda DNA," *Journal of Molecular Biology*, vol. 15, pp. 111–123, Jan. 1966.

[70] J. C. Wang and N. Davidson, "On the probability of ring closure of lambda DNA," *Journal of Molecular Biology*, vol. 19, pp. 469–482, Aug. 1966.

[71] D. Shore, J. Langowski, and R. L. Baldwin, "DNA flexibility studied by covalent closure of short fragments into circles," *Proceedings of the National Academy of Sciences*, vol. 78, pp. 4833–4837, Aug. 1981.

[72] Y. Zhang and D. M. Crothers, "High-throughput approach for detection of DNA bending and flexibility based on cyclization," *Proceedings of the National Academy of Sciences*, vol. 100, pp. 3161–3166, Mar. 2003.

[73] Y. Zhang and D. M. Crothers, "Statistical Mechanics of Sequence-Dependent Circular DNA and Its Application For DNA Cyclization," *Biophysical Journal*, vol. 84, pp. 136–153, Jan. 2003.

[74] A. Travers, "DNA Dynamics: Bubble 'n' Flip for DNA Cyclisation?," *Current Biology*, vol. 15, pp. R377–R379, May 2005.

[75] R. A. Forties, R. Bundschuh, and M. G. Poirier, "The flexibility of locally melted DNA," *Nucleic Acids Research*, vol. 37, pp. 4580–4586, Aug. 2009.

[76] D. P. Wilson, A. V. Tkachenko, and J.-C. Meiners, "A generalized theory of DNA looping and cyclization," *Europhysics Letters*, vol. 89, p. 58005, Mar. 2010.

[77] J. P. Peters and L. J. Maher, "DNA curvature and flexibility in vitro and in vivo," *Quarterly Reviews of Biophysics*, vol. 43, no. 01, pp. 23–63, 2010.

[78] H. Qu, C.-Y. Tseng, Y. Wang, A. J. Levine, and G. Zocchi, "The elastic energy of sharply bent nicked DNA," *Europhysics Letters*, vol. 90, p. 18003, Apr. 2010.

[79] P. Hänggi, P. Talkner, and M. Borkovec, "Reaction-rate theory: fifty years after Kramers," *Reviews of Modern Physics*, vol. 62, pp. 251–341, Apr. 1990.

[80] T. E. Ouldridge, P. Sulc, F. Romano, J. P. K. Doye, and A. A. Louis, "DNA hybridization kinetics: zippering, internal displacement and sequence dependence," *Nucleic Acids Research*, vol. 41, pp. 8886–8895, Oct. 2013.

[81] C. Rivetti, C. Walker, and C. Bustamante, "Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility," *Journal of Molecular Biology*, vol. 280, pp. 41–59, July 1998.

[82] A. Savelyev, "Do monovalent mobile ions affect DNA's flexibility at high salt content?," *Physical Chemistry Chemical Physics*, vol. 14, no. 7, p. 2250, 2012.

[83] R. Owczarzy, B. G. Moreira, Y. You, M. A. Behlke, and J. A. Walder, "Predicting Stability of DNA Duplexes in Solutions Containing Magnesium and Monovalent Cations," *Biochemistry*, vol. 47, pp. 5336–5353, May 2008.

[84] P. J. Hagerman, "Flexibility of DNA," *Annual Review of Biophysics and Biophysical Chemistry*, vol. 17, pp. 265–286, 1988.

[85] S. Mantelli, P. Muller, S. Harlepp, and M. Maaloum, "Conformational analysis and estimation of the persistence length of DNA using atomic force microscopy in solution," *Soft Matter*, vol. 7, no. 7, p. 3412, 2011.

[86] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, "Ionic effects on the elasticity of single DNA molecules," *Proceedings of the National Academy of Sciences*, vol. 94, pp. 6185–6190, June 1997.

[87] G. Maret and G. Weill, "Magnetic birefringence study of the electrostatic and intrinsic persistence length of DNA," *Biopolymers*, vol. 22, pp. 2727–2744, Dec. 1983.

[88] S. Geggier, A. Kotlyar, and A. Vologodskii, "Temperature dependence of DNA persistence length," *Nucleic Acids Research*, vol. 39, pp. 1419–1426, Mar. 2011.

[89] S. M. Mirkin, "DNA Topology: Fundamentals," in *eLS*, John Wiley & Sons, Ltd, 2001.

[90] J. D. Watson and F. H. C. Crick, "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid," *Nature*, vol. 171, pp. 737–738, Apr. 1953.

[91] J. C. Wang, "Helical repeat of DNA in solution," *Proceedings of the National Academy of Sciences*, vol. 76, pp. 200–203, Jan. 1979.

[92] L. J. Peck and J. C. Wang, "Sequence dependence of the helical repeat of DNA in solution," *Nature*, vol. 292, pp. 375–378, July 1981.

[93] D. Rhodes and A. Klug, "Helical periodicity of DNA determined by enzyme digestion," *Nature*, vol. 286, pp. 573–578, Aug. 1980.

[94] S. Ido, K. Kimura, N. Oyabu, K. Kobayashi, M. Tsukada, K. Matsushige, and H. Yamada, "Beyond the Helix Pitch: Direct Visualization of Native DNA in Aqueous Solution," *ACS Nano*, vol. 7, pp. 1817–1822, Feb. 2013.

[95] E. L. Zechiedrich, A. B. Khodursky, S. Bachellier, R. Schneider, D. Chen, D. M. J. Lilley, and N. R. Cozzarelli, "Roles of Topoisomerases in Maintaining Steady-state DNA Supercoiling in Escherichia coli," *Journal of Biological Chemistry*, vol. 275, pp. 8103–8113, Mar. 2000.

[96] P. Beard, J. F. Morrow, and P. Berg, "Cleavage of Circular, Superhelical Simian Virus 40 DNA to a Linear Duplex by S1 Nuclease," *Journal of Virology*, vol. 12, pp. 1303–1313, Dec. 1973.

[97] J. C. Wang, "Interactions between twisted DNAs and enzymes: The effects of superhelical turns," *Journal of Molecular Biology*, vol. 87, pp. 797–816, Aug. 1974.

[98] H. B. Gray, D. A. Ostrander, J. L. Hodnett, R. J. Legerski, and D. L. Robberson, "Extracellular nucleases of Pseudomonas BAL 31. I. Characterization of single strand-specific deoxyriboendonuclease and double-strand deoxyriboexonuclease activities," *Nucleic Acids Research*, vol. 2, pp. 1459–1492, Sept. 1975.

[99] J. M. Fogg, N. Kolmakova, I. Rees, S. Magonov, H. Hansma, J. J. Perona, and E. L. Zechiedrich, "Exploring writhe in supercoiled minicircle DNA," *Journal of Physics: Condensed Matter*, vol. 18, pp. S145–S159, Apr. 2006.

[100] J. Bednar, P. Furrer, A. Stasiak, J. Dubochet, E. H. Egelman, and A. D. Bates, "The Twist, Writhe and Overall Shape of Supercoiled DNA Change During Counterion-induced Transition from a Loosely to a Tightly Interwound Superhelix: Possible Implications for DNA Structure in Vivo," *Journal of Molecular Biology*, vol. 235, pp. 825–847, Jan. 1994.

[101] X. Zheng and A. Vologodskii, "Theoretical Analysis of Disruptions in DNA Minicircles," *Biophysical Journal*, vol. 96, pp. 1341–1349, Feb. 2009.

[102] M. Zoli, "Twisting and bending stress in DNA minicircles," *Soft Matter*, vol. 10, pp. 4304–4311, May 2014.

[103] M. Sayar, B. Avşaroğlu, and A. Kabakçıoğlu, "Twist-writhe partitioning in a coarse-grained DNA minicircle model," *Physical Review E*, vol. 81, p. 041916, Apr. 2010.

[104] J. S. Mitchell and S. A. Harris, "Thermodynamics of Writhe in DNA Minicircles from Molecular Dynamics Simulations," *Physical Review Letters*, vol. 110, p. 148105, Apr. 2013.

[105] T. Strick, J.-F. Allemand, D. Bensimon, and V. Croquette, "Behavior of Supercoiled DNA," *Biophysical Journal*, vol. 74, pp. 2016–2028, Apr. 1998.

[106] J.-H. Jeon, J. Adamcik, G. Dietler, and R. Metzler, "Supercoiling Induces Denaturation Bubbles in Circular DNA," *Physical Review Letters*, vol. 105, p. 208101, Nov. 2010.

[107] J. Adamcik, J.-H. Jeon, K. J. Karczewski, R. Metzler, and G. Dietler, "Quantifying supercoiling-induced denaturation bubbles in DNA," *Soft Matter*, vol. 8, no. 33, p. 8651, 2012.

[108] C.-Y. Tseng, A. Wang, G. Zocchi, B. Rolih, and A. J. Levine, "Elastic energy of protein-DNA chimeras," *Physical Review E*, vol. 80, p. 061912, Dec. 2009.

[109] Y. Wang, A. Wang, H. Qu, and G. Zocchi, "Protein–DNA chimeras: synthesis of two-arm chimeras and non-mechanical effects of the DNA spring," *Journal of Physics: Condensed Matter*, vol. 21, p. 335103, Aug. 2009.

[110] G. Weber, J. W. Essex, and C. Neylon, "Probing the microscopic flexibility of DNA from melting temperatures," *Nature Physics*, vol. 5, pp. 769–773, Oct. 2009.

[111] J. F. Marko and E. D. Siggia, "Stretching DNA," *Macromolecules*, vol. 28, pp. 8759–8770, Dec. 1995.

[112] C. Kittel, "Phase Transition of a Molecular Zipper," *American Journal of Physics*, vol. 37, no. 9, p. 917, 1969.

[113] C. J. Benham, "Torsional stress and local denaturation in supercoiled DNA," *Proceedings of the National Academy of Sciences*, vol. 76, pp. 3870–3874, Aug. 1979.

[114] Ouldridge, Thomas, *Coarse-grained modelling of DNA and DNA self-assembly*. D.Phil Thesis, University of Oxford, Oxford, UK, 2011.

[115] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, "Extracting bulk properties of self-assembling systems from small simulations," *Journal of Physics: Condensed Matter*, vol. 22, p. 104102, Mar. 2010.

[116] J. Wang, H. Qu, and G. Zocchi, "Critical bending torque of DNA is a materials parameter independent of local base sequence," *Physical Review E*, vol. 88, p. 032712, Sept. 2013.

[117] D. S. Sanchez, H. Qu, D. Bulla, and G. Zocchi, "DNA kinks and bubbles: Temperature dependence of the elastic energy of sharply bent 10-nm-size DNA molecules," *Physical Review E*, vol. 87, p. 022710, Feb. 2013.

[118] M. Abdelhaleem, "Helicases: an overview," *Methods in Molecular Biology (Clifton, N.J.)*, vol. 587, pp. 1–12, 2010.

[119] M. R. Singleton and D. B. Wigley, "Modularity and Specialization in Superfamily 1 and 2 Helicases," *Journal of Bacteriology*, vol. 184, pp. 1819–1826, Apr. 2002.

[120] G. Waksman, E. Lanka, and J.-M. Carazo, "Helicases as nucleic acid unwinding machines," *Nature Structural & Molecular Biology*, vol. 7, pp. 20–22, Jan. 2000.

[121] S. S. Velankar, P. Soultanas, M. S. Dillingham, H. S. Subramanya, and D. B. Wigley, "Crystal Structures of Complexes of PcrA DNA Helicase with a DNA Substrate Indicate an Inchworm Mechanism," *Cell*, vol. 97, pp. 75–84, Apr. 1999.

[122] S. S. Patel and I. Donmez, "Mechanisms of Helicases," *Journal of Biological Chemistry*, vol. 281, pp. 18265 –18268, July 2006.

[123] J. G. Yodh, M. Schlierf, and T. Ha, "Insight into helicase mechanism and function revealed through single-molecule approaches," *Quarterly Reviews of Biophysics*, vol. 43, pp. 185–217, Aug. 2010.

[124] T. Ha, I. Rasnik, W. Cheng, H. P. Babcock, G. H. Gauss, T. M. Lohman, and S. Chu, "Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase," *Nature*, vol. 419, pp. 638–641, Oct. 2002.

[125] I. Rasnik, S. Myong, W. Cheng, T. M. Lohman, and T. Ha, "DNA-binding Orientation and Domain Conformation of the E.coli Rep Helicase Monomer Bound to a Partial Duplex Junction: Single-molecule Studies of Fluorescently Labeled Enzymes," *Journal of Molecular Biology*, vol. 336, pp. 395–408, Feb. 2004.

[126] W. K. Chu and I. D. Hickson, "RecQ helicases: multi-functional genome caretakers," *Nature Reviews Cancer*, vol. 9, no. 9, pp. 644–654, 2009.

[127] V. A. Bohr, "Rising from the RecQ-age: the role of human RecQ helicases in genome maintenance," *Trends in Biochemical Sciences*, vol. 33, pp. 609–620, Dec. 2008.

[128] I. D. Hickson, "RecQ helicases: caretakers of the genome," *Nature Reviews Cancer*, vol. 3, pp. 169–178, Mar. 2003.

[129] K. Hanada and I. D. Hickson, "Molecular genetics of RecQ helicase disorders," *Cellular and Molecular Life Sciences*, vol. 64, pp. 2306–2322, June 2007.

[130] W. Yang, "Lessons Learned From UvrD Helicase : Mechanism For Directional Movement," *Annual Review of Biophysics*, vol. 39, pp. 367–385, 2010.

[131] P. Soultanas, M. S. Dillingham, P. Wiley, M. R. Webb, and D. B. Wigley, "Uncoupling DNA translocation and helicase activity in PcrA: direct evidence for an active mechanism," *The EMBO Journal*, vol. 19, pp. 3799–3810, July 2000.

[132] J. Y. Lee and W. Yang, "UvrD Helicase Unwinds DNA One Base Pair At A Time By A Two-Part Power Stroke," *Cell*, vol. 127, pp. 1349–1360, Dec. 2006.

[133] X.-D. Zhang, S.-X. Dou, P. Xie, J.-S. Hu, P.-Y. Wang, and X. G. Xi, "Escherichia coli RecQ Is a Rapid, Efficient, and Monomeric Helicase," *Journal of Biological Chemistry*, vol. 281, pp. 12655 –12663, May 2006.

[134] M. D. Betterton and F. Jülicher, "Opening of nucleic-acid double strands by helicases: Active versus passive opening," *Physical Review E*, vol. 71, p. 011904, Jan. 2005.

[135] M. Manosas, X. G. Xi, D. Bensimon, and V. Croquette, "Active and passive mechanisms of helicases," *Nucleic Acids Research*, vol. 38, pp. 5518–5526, Sept. 2010.

[136] J. J. McCann, U. B. Choi, L. Zheng, K. Weninger, and M. E. Bowen, "Optimizing Methods to Recover Absolute FRET Efficiency from Immobilized Single Molecules," *Biophysical Journal*, vol. 99, pp. 961–970, Aug. 2010.

[137] B. W. van der Meer, "Kappa-squared: from nuisance to new sense," *Reviews in Molecular Biotechnology*, vol. 82, pp. 181–196, Jan. 2002.

[138] Y. Santoso, L. Hwang, L. Le Reste, and A. Kapanidis, "Red light, green light: probing single molecules using alternating-laser excitation," *Biochemical Society Transactions*, vol. 36, pp. 738–744, 2008.

[139] J. W. Taraska, M. C. Puljung, N. B. Olivier, G. E. Flynn, and W. N. Zagotta, "Mapping the structure and conformational movements of proteins with transition metal ion FRET," *Nature Methods*, vol. 6, pp. 532–537, July 2009.

[140] J. W. Taraska, M. C. Puljung, and W. N. Zagotta, "Short-distance probes for protein backbone structure based on energy transfer between bimane and transition metal ions," *Proceedings of the National Academy of Sciences*, vol. 106, no. 38, pp. 16227 –16232, 2009.

[141] S. S. Vogel, C. Thaler, and S. V. Koushik, "Fanciful FRET," *Science Signaling*, vol. 2006, p. re2, Apr. 2006. Förster resonance energy transfer (FRET) is used to measure protein-protein interactions in living cells. When FRET occurs, a "donor" fluorophore is excited by the absorption of a photon, and that energy is transferred by a nonradiative mechanism to a nearby acceptor molecule. The fraction of energy-transfer events relative to donor excitation events is called the FRET efficiency. It has a steep dependence on the separation distance between the donor and acceptor. Proteins of interest tagged with spectral variants of green fluorescent protein (GFP) can be readily used for FRET studies in living cells. The proper interpretation of FRET measurements, however, requires an understanding of the limitations of the numerous methods used for measuring FRET, as well as consideration of the subtleties of FRET theory, particularly its dependence on the abundance of donors and acceptors. This review highlights aspects of the acquisition of FRET measurements, and of FRET theory, that are vital for proper interpretation. The adoption of "standards" with known FRET efficiencies is a first step toward

eliminating erroneous interpretations of FRET experiments due to errors in the accuracy and precision of FRET measurements.

[142] I. Johnson, *The Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies, 11th Edition*. Life Technologies Corporation, 2010.

[143] P. Hinterdorfer and A. V. Oijen, *Handbook of Single-Molecule Biophysics*. Springer, Nov. 2009.

[144] M. C. Hall, J. R. Jordan, and S. W. Matson, "Evidence for a physical interaction between the Escherichia coli methyl-directed mismatch repair proteins MutL and UvrD.," *The EMBO Journal*, vol. 17, pp. 1535–1541, Mar. 1998.

[145] X. Veaute, S. Delmas, M. Selva, J. Jeusset, E. Le Cam, I. Matic, F. Fabre, and M.-A. Petit, "UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in Escherichia coli," *The EMBO Journal*, vol. 24, pp. 180–189, Jan. 2005.

[146] D. Axelrod, T. Burghardt, and N. Thompson, "Total internal reflection fluorescence," *Annual Review of Biophysics and Bioengineering*, vol. 13, no. 1, pp. 247–268, 1984.

[147] S. K. Sarkar, N. M. Andoy, J. J. Benítez, P. R. Chen, J. S. Kong, C. He, and P. Chen, "Engineered Holliday Junctions as Single-Molecule Reporters for Protein-DNA Interactions with Application to a MerR-family Regulator," *Journal of the American Chemical Society*, vol. 129, pp. 12461–12467, Oct. 2007.

[148] A. Bumb, S. K. Sarkar, X. S. Wu, M. W. Brechbiel, and K. C. Neuman, "Quantitative characterization of fluorophores in multi-component nanoprobes by single-molecule fluorescence," *Biomedical Optics Express*, vol. 2, pp. 2761–2769, Sept. 2011.

[149] S. Nie, D. T. Chiu, and R. N. Zare, "Probing individual molecules with confocal fluorescence microscopy," *Science*, vol. 266, pp. 1018–1021, Nov. 1994.

[150] H. Nakayama, K. Nakayama, R. Nakayama, N. Irino, Y. Nakayama, and P. C. Hanawalt, "Isolation and genetic characterization of a thymineless death-resistant mutant of Escherichia coli K12: identification of a new mutation (recQ1) that blocks the RecF recombination pathway," *Molecular & General Genetics*, vol. 195, no. 3, pp. 474–480, 1984.

[151] K. Umezu, K. Nakayama, and H. Nakayama, "Escherichia coli RecQ protein is a DNA helicase," *Proceedings of the National Academy of Sciences*, vol. 87, pp. 5363 –5367, July 1990.

[152] D. A. Bernstein, M. C. Zittel, and J. L. Keck, "High-resolution structure of the E.coli RecQ helicase catalytic core," *The EMBO Journal*, vol. 22, pp. 4910–4921, Oct. 2003.

[153] H. Q. Xu, E. Deprez, A. H. Zhang, P. Tauc, M. M. Ladjimi, J.-C. Brochon, C. Auclair, and X. G. Xi, "The Escherichia coli RecQ Helicase Functions as a Monomer," *Journal of Biological Chemistry*, vol. 278, no. 37, pp. 34925 –34933, 2003.

[154] J. L. Liu, P. Rigolet, S.-X. Dou, P.-Y. Wang, and X. G. Xi, "The Zinc Finger Motif of Escherichia coli RecQ Is Implicated in Both DNA Binding and Protein Folding," *Journal of Biological Chemistry*, vol. 279, pp. 42794 –42802, Oct. 2004.

[155] K. Kitano, S.-Y. Kim, and T. Hakoshima, "Structural Basis for DNA Strand Separation by the Unconventional Winged-Helix Domain of RecQ Helicase WRN," *Structure*, vol. 18, pp. 177–187, Feb. 2010.

[156] D. A. Bernstein and J. L. Keck, "Conferring Substrate Specificity to DNA Helicases: Role of the RecQ HRDC Domain," *Structure*, vol. 13, pp. 1173–1182, Aug. 2005.

[157] J. L. Kim, K. A. Morgenstern, J. P. Griffith, M. D. Dwyer, J. A. Thomson, M. A. Murcko, C. Lin, and P. R. Caron, "Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding," *Structure*, vol. 6, pp. 89–100, Jan. 1998.

[158] S.-X. Dou, P.-Y. Wang, H. Q. Xu, and X. G. Xi, "The DNA Binding Properties of the Escherichia coli RecQ Helicase," *Journal of Biological Chemistry*, vol. 279, pp. 6354 –6363, Feb. 2004.

[159] D. A. Bernstein and J. L. Keck, "Domain mapping of Escherichia coli RecQ defines the roles of conserved N- and C-terminal regions in the RecQ family," *Nucleic Acids Research*, vol. 31, pp. 2778–2785, June 2003.

[160] E. Antony, E. J. Tomko, Q. Xiao, L. Krejci, T. M. Lohman, and T. Ellenberger, "Srs2 Disassembles Rad51 Filaments by a Protein-Protein Interaction Triggering ATP Turnover and Dissociation of Rad51 from DNA," *Molecular Cell*, vol. 35, pp. 105–115, July 2009.

[161] L. E. Mechanic, M. C. Hall, and S. W. Matson, "Escherichia coli DNA Helicase II Is Active as a Monomer," *Journal of Biological Chemistry*, vol. 274, pp. 12488 –12498, Apr. 1999.

[162] N. K. Maluf, C. J. Fischer, and T. M. Lohman, "A Dimer of Escherichia coli UvrD is the Active Form of the Helicase In Vitro," *Journal of Molecular Biology*, vol. 325, pp. 913–935, Jan. 2003.

[163] J. A. Ali, N. K. Maluf, and T. M. Lohman, "An oligomeric form of E. coli UvrD is required for optimal helicase activity1," *Journal of Molecular Biology*, vol. 293, pp. 815–834, Nov. 1999.

[164] S. W. Matson, "Escherichia coli helicase II (urvD gene product) translocates unidirectionally in a 3'to 5'direction.," *Journal of Biological Chemistry*, vol. 261, no. 22, p. 10169, 1986.

[165] C. J. Cadman, S. W. Matson, and P. McGlynn, "Unwinding of Forked DNA Structures by UvrD," *Journal of Molecular Biology*, vol. 362, pp. 18–25, Sept. 2006.

[166] G. T. Runyon, D. G. Bear, and T. M. Lohman, "Escherichia coli helicase II (UvrD) protein initiates DNA unwinding at nicks and blunt ends," *Proceedings of the National Academy of Sciences*, vol. 87, no. 16, pp. 6383 –6387, 1990.

[167] G. F. Moolenaar, C. Moorman, and N. Goosen, "Role of the Escherichia coli Nucleotide Excision Repair

Proteins in DNA Replication," *Journal of Bacteriology*, vol. 182, pp. 5706–5714, Oct. 2000.

[168] J. Atkinson, C. P. Guy, C. J. Cadman, G. F. Moolenaar, N. Goosen, and P. McGlynn, "Stimulation of UvrD Helicase by UvrAB," *The Journal of Biological Chemistry*, vol. 284, pp. 9612–9623, Apr. 2009.

[169] F. G. Harmon, J. P. Brockman, and S. C. Kowalczykowski, "RecQ Helicase Stimulates Both DNA Catenation and Changes in DNA Topology by Topoisomerase III," *Journal of Biological Chemistry*, vol. 278, pp. 42668 –42678, Oct. 2003.

[170] F. Harmon, "RecQ Helicase and Topoisomerase III Comprise a Novel DNA Strand Passage FunctionA Conserved Mechanism for Control of DNA Recombination," *Molecular Cell*, vol. 3, pp. 611–620, May 1999.

[171] C. Suski and K. J. Marians, "Resolution of Converging Replication Forks by RecQ and Topoisomerase III," *Molecular Cell*, vol. 30, pp. 779–789, June 2008.

[172] R. D. Shereda, D. A. Bernstein, and J. L. Keck, "A Central Role for SSB in Escherichia coli RecQ DNA Helicase Function," *Journal of Biological Chemistry*, vol. 282, pp. 19247 –19258, June 2007.

[173] S. Cui, D. Arosio, K. M. Doherty, R. M. Brosh, A. Falaschi, and A. Vindigni, "Analysis of the unwinding activity of the dimeric RECQ1 helicase in the presence of human replication protein A," *Nucleic Acids Research*, vol. 32, pp. 2158–2170, Apr. 2004.

[174] R. M. Brosh, D. K. Orren, J. O. Nehlin, P. H. Ravn, M. K. Kenny, A. Machwe, and V. A. Bohr, "Functional and Physical Interaction between WRN Helicase and Human Replication Protein A," *Journal of Biological Chemistry*, vol. 274, pp. 18341 –18350, June 1999.

[175] R. M. Brosh, J.-L. Li, M. K. Kenny, J. K. Karow, M. P. Cooper, R. P. Kureekattil, I. D. Hickson, and V. A. Bohr, "Replication Protein A Physically Interacts with the Bloom's Syndrome Protein and Stimulates Its Helicase Activity," *Journal of Biological Chemistry*, vol. 275, no. 31, pp. 23500 –23508, 2000.

[176] P. L. Opresko, P. A. Mason, E. R. Podell, M. Lei, I. D. Hickson, T. R. Cech, and V. A. Bohr, "POT1 Stimulates RecQ Helicases WRN and BLM to Unwind Telomeric DNA Substrates," *Journal of Biological Chemistry*, vol. 280, no. 37, pp. 32069 –32080, 2005.

[177] P. Mohaghegh, J. K. Karow, R. M. Brosh, V. A. Bohr, and I. D. Hickson, "The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases," *Nucleic Acids Research*, vol. 29, pp. 2843–2849, July 2001.

[178] P. L. Opresko, M. Otterlei, J. Graakjær, P. Bruheim, L. Dawut, S. Kølvraa, A. May, M. M. Seidman, and V. A. Bohr, "The Werner Syndrome Helicase and Exonuclease Cooperate to Resolve Telomeric D Loops in a Manner Regulated by TRF1 and TRF2," *Molecular Cell*, vol. 14, pp. 763–774, June 2004.

[179] K. Lillard-Wetherell, A. Machwe, G. T. Langland, K. A. Combs, G. K. Behbehani, S. A. Schonberg, J. German, J. J. Turchi, D. K. Orren, and J. Groden, "Association and regulation of the BLM helicase by the telomere proteins TRF1 and TRF2," *Human Molecular Genetics*, vol. 13, pp. 1919–1932, Sept. 2004.

[180] Z.-G. He and C. C. Richardson, "Effect of Single-stranded DNA-binding Proteins on the Helicase and Primase Activities of the Bacteriophage T7 Gene 4 Protein," *Journal of Biological Chemistry*, vol. 279, pp. 22190 –22197, May 2004.

[181] E. E. Biswas, P.-H. Chen, and S. B. Biswas, "Modulation of enzymatic activities of Escherichia coli DnaB helicase by single-stranded DNA-binding proteins," *Nucleic Acids Research*, vol. 30, pp. 2809–2816, July 2002.

[182] M. C. Zittel and J. L. Keck, "Coupling DNA-binding and ATP hydrolysis in Escherichia coli RecQ: role of a highly conserved aromatic-rich sequence," *Nucleic Acids Research*, vol. 33, no. 22, pp. 6982–6991, 2005.

[183] B. Lucic, Y. Zhang, O. King, R. Mendoza-Maldonado, M. Berti, F. H. Niesen, N. A. Burgess-Brown, A. C. W. Pike, C. D. O. Cooper, O. Gileadi, and A. Vindigni, "A prominent β-hairpin structure in the winged-helix domain of RECQ1 is required for DNA unwinding and oligomer formation," *Nucleic Acids Research*, vol. 39, pp. 1703–1717, Mar. 2011.

[184] K. Saikrishnan, B. Powell, N. J. Cook, M. R. Webb, and D. B. Wigley, "Mechanistic Basis of 5'-3' Translocation in SF1b Helicases," *Cell*, vol. 137, pp. 849–859, May 2009.

[185] L. Muzzolini, F. Beuron, A. Patwardhan, V. Popuri, S. Cui, B. Niccolini, M. Rappas, P. S. Freemont, and A. Vindigni, "Different Quaternary Structures of Human RECQ1 Are Associated with Its Dual Enzymatic Activity," *PLoS Biology*, vol. 5, p. e20, Jan. 2007.

[186] A. C. W. Pike, B. Shrestha, V. Popuri, N. Burgess-Brown, L. Muzzolini, S. Costantini, A. Vindigni, and O. Gileadi, "Structure of the human RECQ1 helicase reveals a putative strand-separation pin," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 1039–1044, Jan. 2009.

[187] K. A. Hoadley and J. L. Keck, "Werner helicase wings DNA binding," *Structure*, vol. 18, pp. 149–151, Feb. 2010.

[188] M. Gyimesi, G. M. Harami, K. Sarlós, E. Hazai, Z. Bikádi, and M. Kovács, "Complex activities of the human Bloom's syndrome helicase are encoded in a core region comprising the RecA and Zn-binding domains," *Nucleic Acids Research*, vol. 40, pp. 3952–3963, May 2012.

[189] X. D. Zhang, D. O. U. Shuo-Xing, X. I. E. Ping, W. Peng-Ye, and X. G. Xi, "RecQ helicase-catalyzed DNA unwinding detected by fluorescence resonance energy transfer," *Acta Biochimica et Biophysica Sinica*, vol. 37, no. 9, pp. 593–600, 2005.

[190] C. R. Bagshaw, "ATP analogues at a glance," *Journal of Cell Science*, vol. 114, pp. 459 –460, Jan. 2001.

[191] J. German, M. M. Sanz, S. Ciocci, T. Z. Ye, and N. A. Ellis, "Syndrome-causing mutations of the BLM gene in persons in the Bloom's Syndrome Registry," *Human Mutation*, vol. 28, no. 8, pp. 743–753, 2007.

[192] R. J. Bennett, J. A. Sharp, and J. C. Wang, "Purification and Characterization of the Sgs1 DNA Helicase Activity of Saccharomyces cerevisiae," *Journal of Biological Chemistry*, vol. 273, pp. 9644–9650, Apr. 1998.

[193] S. Lyskov and J. J. Gray, "The RosettaDock server for local protein-protein docking," *Nucleic Acids Research*, vol. 36, pp. W233–W238, May 2008.

[194] M. Sunbul and J. Yin, "Site specific protein labeling by enzymatic posttranslational modification," *Organic & Biomolecular Chemistry*, vol. 7, no. 17, p. 3361, 2009.

[195] J. B. Hunt, S. H. Neece, H. K. Schachman, and A. Ginsburg, "Mercurial-promoted Zn2+ release from Escherichia coli aspartate transcarbamoylase.," *Journal of Biological Chemistry*, vol. 259, pp. 14793–14803, Dec. 1984.

[196] J. B. Hunt, S. H. Neece, and A. Ginsburg, "The use of 4-(2-pyridylazo)resorcinol in studies of zinc release from Escherichia coli aspartate transcarbamoylase," *Analytical Biochemistry*, vol. 146, pp. 150–157, Apr. 1985.

[197] P. W. Riddles, R. L. Blakeley, and B. Zerner, "Reassessment of Ellman's reagent," in *Enzyme Structure Part I*, vol. Volume 91, pp. 49–60, Academic Press, 1983.

[198] M. Taki, M. Shiota, and K. Taira, "Transglutaminase-mediated N- and C-terminal fluorescein labeling of a protein can support the native activity of the modified protein," *Protein Engineering Design and Selection*, vol. 17, pp. 119 –126, Feb. 2004.

[199] M. Jäger, E. Nir, and S. Weiss, "Site-specific labeling of proteins for single-molecule FRET by combining chemical and enzymatic modification," *Protein Science*, vol. 15, no. 3, pp. 640–646, 2006.

[200] F. G. Harmon and S. C. Kowalczykowski, "RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination," *Genes & Development*, vol. 12, pp. 1134–1144, Apr. 1998.

[201] J. Sambrook and D. W. Russell, *Molecular cloning: a laboratory manual*. CSHL Press, Jan. 2001.

[202] M. Cooper, A. Ebner, M. Briggs, M. Burrows, N. Gardner, R. Richardson, and R. West, "Cy3b™: Improving the Performance of Cyanine Dyes," *Journal of Fluorescence*, vol. 14, no. 2, pp. 145–150, 2004.

[203] B. Rad and S. C. Kowalczykowski, "Translocation of E. coli RecQ Helicase on Single-Stranded DNA," *Biochemistry*, vol. 51, pp. 2921–2929, Apr. 2012.

[204] F. G. Harmon and S. C. Kowalczykowski, "Biochemical Characterization of the DNA Helicase Activity of theEscherichia coli RecQ Helicase," *Journal of Biological Chemistry*, vol. 276, pp. 232 –243, Jan. 2001.

[205] N. Li, E. Henry, E. Guiot, P. Rigolet, J.-C. Brochon, X.-G. Xi, and E. Deprez, "Multiple Escherichia coli RecQ Helicase Monomers Cooperate to Unwind Long DNA Substrates," *Journal of Biological Chemistry*, vol. 285, pp. 6922 –6936, Mar. 2010.

[206] B. Y. Pan, S. X. Dou, Y. Yang, Y. N. Xu, E. Bugnard, X. Y. Ding, L. Zhang, P. Y. Wang, M. Li, and X. G. Xi, "Mutual Inhibition of RecQ Molecules in DNA Unwinding," *Journal of Biological Chemistry*, vol. 285, pp. 15884–15893, Mar. 2010.

[207] J. E. Bronson, J. Fei, J. M. Hofman, R. L. Gonzalez, and C. H. Wiggins, "Learning Rates and States from Biophysical Time Series: A Bayesian Approach to Model Selection and Single-Molecule FRET Data," *Biophysical Journal*, vol. 97, pp. 3196–3205, Dec. 2009.

[208] S. J. Holden and A. N. Kapanidis, "Alternating-Laser Excitation and Pulsed-Interleaved Excitation of Single Molecules," in *Single Particle Tracking and Single Molecule Energy Transfer* (C. Bräuchle, D. C. Lamb, and J. Michaelis, eds.), pp. 131–162, Wiley-VCH Verlag GmbH & Co. KGaA, 2009.

[209] A. N. Kapanidis, T. A. Laurence, N. K. Lee, E. Margeat, X. Kong, and S. Weiss, "Alternating-Laser Excitation of Single Molecules," *Accounts of Chemical Research*, vol. 38, pp. 523–533, July 2005.

[210] J. G. Yodh, B. C. Stevens, R. Kanagaraj, P. Janscak, and T. Ha, "BLM helicase measures DNA unwound before switching strands and hRPA promotes unwinding reinitiation," *The EMBO Journal*, vol. 28, pp. 405–416, Feb. 2009.

[211] K. Sarlós, M. Gyimesi, and M. Kovács, "RecQ helicase translocates along single-stranded DNA with a moderate processivity and tight mechanochemical coupling," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 9804–9809, June 2012.

[212] M. Gu and C. M. Rice, "Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 521–528, Jan. 2010.

[213] L. Le Reste, J. Hohlbein, K. Gryte, and A. N. Kapanidis, "Characterization of Dark Quencher Chromophores as Nonfluorescent Acceptors for Single-Molecule FRET," *Biophysical Journal*, vol. 102, pp. 2658–2668, June 2012.

[214] H. Balci, S. Arslan, S. Myong, T. M. Lohman, and T. Ha, "Single-molecule nanopositioning: structural transitions of a helicase-DNA complex during ATP hydrolysis," *Biophysical Journal*, vol. 101, pp. 976–984, Aug. 2011.

[215] W. H. Taylor and P. J. Hagerman, "Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure: II. NaCl-dependence of DNA flexibility and helical repeat," *Journal of Molecular Biology*, vol. 212, pp. 363–376, Mar. 1990.

[216] J. Lapham, J. P. Rife, P. B. Moore, and D. M. Crothers, "Measurement of diffusion constants for nucleic acids by NMR," *Journal of Biomolecular NMR*, vol. 10, pp. 255–262, Oct. 1997.

[217] K. Tyagarajan, E. Pretzer, and J. E. Wiktorowicz, "Thiol-reactive dyes for fluorescence labeling of proteomic samples," *Electrophoresis*, vol. 24, pp. 2348–2358, July 2003.

[218] N. K. Lee, A. N. Kapanidis, H. R. Koh, Y. Korlann, S. O. Ho, Y. Kim, N. Gassman, S. K. Kim, and S. Weiss, "Three-Color Alternating-Laser Excitation of Single Molecules: Monitoring Multiple Interactions and Distances," *Biophysical Journal*, vol. 92, pp. 303–312, Jan. 2007.

[219] S. J. Holden, S. Uphoff, J. Hohlbein, D. Yadin, L. Le Reste, O. J. Britton, and A. N. Kapanidis, "Defining the Limits of Single-Molecule FRET Resolution in TIRF Microscopy," *Biophysical Journal*, vol. 99, pp. 3102–3111, Nov. 2010.

[220] J. Shi, A. Gafni, and D. Steel, "Simulated data sets for single molecule kinetics: some limitations and complications of data analysis," *European Biophysics Journal*, vol. 35, pp. 633–645, Oct. 2006.

[221] G. E. Healthcare, "Cy3 Maleimide Mono-Reactive Dye 5-pack," tech. rep.

[222] G. E. Healthcare, "Amersham Cy3b mono-reactive dye," tech. rep.

[223] G. E. Healthcare, "Cy5 Maleimide Mono-Reactive Dye 5-pack," tech. rep.

[224] ATTO-TEC GmbH, "Fluorescent Labels and Dyes," tech. rep., 2011.

[225] K. Büttner, S. Nehring, and K.-P. Hopfner, "Structural basis for DNA duplex separation by a superfamily-2 helicase," *Nature Structural & Molecular Biology*, vol. 14, no. 7, pp. 647–652, 2007.

[226] Terpetschnig, Ewald A. ; et al., "Luminescent compounds (US Patent Application 0040166515)."

[227] K. Kolmakov, V. N. Belov, C. A. Wurm, B. Harke, M. Leutenegger, C. Eggeling, and S. W. Hell, "A Versatile Route to Red-Emitting Carbopyronine Dyes for Optical Microscopy and Nanoscopy," *European Journal of Organic Chemistry*, vol. 2010, no. 19, pp. 3593–3610, 2010.