UTILIZING MOLECULAR DYNAMICS' MULTIPOTENT METHODOLOGIES TO MEASURE MICROSCOPIC MOTIONS OF DNA MOLECULES

A Magniloquent Manuscript on DNA's Means and Mannerisms

ADDIE KINGSLAND

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

University of Washington
2017

Reading Committee:
Lutz Maibaum, Chair
David S. Ginger
Stefan Stoll

Program Authorized to Offer Degree:
Chemistry

University of Washington

ABSTRACT

Utilizing Molecular Dynamics' Multipotent Methodologies to Measure Microscopic Motions of DNA Molecules, A Magniloquent Manuscript on DNA's Means and Mannerisms

Addie Kingsland

Chair of the Supervisory Committee:
Prof. Lutz Maibaum
Department of Chemistry

DNA is an amazing molecule which is the basic template for all genetics. It is the primary molecule for storing biological information, and has many applications in nanotechnology. Double-stranded DNA may contain mismatched base pairs beyond the Watson-Crick pairs guanine-cytosine and adenine-thymine. To date, no one has found a physical property of base pair mismatches which describes the behavior of naturally occurring mismatch repair enzymes. Many materials properties of DNA are also unknown, for instance, when pulling DNA in different configurations, different energy differences are observed with no obvious reason why. DNA mismatches also affect their local environment, for instance changing the quantum yield of nearby azobenzene moieties.

We utilize molecular dynamics computer simulations to study the structure and dynamics for both matched and mismatched base pairs, within both biological and materials contexts, and in both equilibrium and biased dynamics. We show that mismatched pairs shift further in the plane normal to the DNA strand and are more likely to exhibit non-canonical structures, including the e-motif. Base pair mismatches alter their local environment, affecting the *trans*- to *cis*- photoisomerization quantum yield of azobenzene, as well as increasing the likelihood of observing the e-motif. We also show that by using simulated

data, we can give new insights on theoretical models to calculate the energetics of pulling DNA strands apart. These results, all relatively inexpensive on modern computer hardware, can help guide the design of DNA-based nanotechnologies, as well as give new insights into the functioning of mismatch repair systems in cancer prevention.

Hey! You look like you're wearing a tuxedo!

— Addie (2007)

PUBLICATIONS

Some ideas and figures have appeared previously in the following publications:

(1) Kingsland, A.; Samai, S.; Yan, Y.; Ginger, D. S.; Maibaum, L. Local density fluctuations predict photoisomerization quantum yield of azobenzene-modified DNA. *J. Phys. Chem. Lett.* **2016**, *7*, 3027–3031. Reprinted with permission, ©2016 American Chemical Society.

(2) Kingsland, A.; Maibaum, L. DNA base pair mismatches induce structural changes and change the free energy landscape of base flip. *Submitted to Biophys. J.* **2017**.

We all have a thirst for wonder. It's a deeply human quality. Science and religion are both bound up with it. What I'm saying is, you don't have to make stories up, you don't have to exaggerate. There's wonder and awe enough in the real world. Nature's a lot better at inventing wonders than we are.

— Carl Sagan, Contact

ACKNOWLEDGEMENTS

So many thanks to all of the folks who have stuck with me on this fantastic journey.

Firstly, thanks to my thesis committee for all of the amazing suggestions, professional development, and kind understanding: Lutz Maibaum, David Ginger, Sarah Keller, Jim Pfaendtner, Stefan Stoll, and Paul Wiggins.

Thanks to my friends from graduate school who understand the struggle. Patrick Lestrange, Nick Montoni, Scott Rayermann, Kayla Sapp, Tracy Stanzel, Eric Stokes, and Sarah Vorpahl are particularly awesome. This experience has shaped us into lifelong friends.

To my family: I wouldn't be here without you. Your unwavering support through all my ups and downs is what has kept me going even during my roughest patches. Special thanks to my parents, John Benson and Leita Kingsland; my favorite sibling, Patrick Benson-Kingsland; and my partners, Bryan Hawley and Kate Miller. I couldn't have done it without you.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AFM	atomic force microscopy
azo-DNA	azobenzene-modified DNA
DFS	dynamic force spectroscopy
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
FES	free-energy surface
LJ	Lennard-Jones
MMR	mismatch repair
NMR	nuclear magnetic resonance
MD	molecular dynamics
NVT	number-volume-temperature
QY	quantum yield
RNA	ribonucleic acid
SASA	solvent accessible surface area
ssDNA	single-stranded DNA
UV	ultra-violet
WCA	Weeks-Chandler-Anderson
WHAM	weighted-histogram analysis method

Part I

PRELUDE

Change is freedom, change is life.

It's always easier not to think for oneself. Find a nice safe hierarchy and settle in. Don't make changes, don't risk disapproval, don't upset your syndics. It's always easiest to let yourself be governed.

There's a point, around age twenty, when you have to choose whether to be like everybody else the rest of your life, or to make a virtue of your peculiarities.

Those who build walls are their own prisoners. I'm going to go fulfil my proper function in the social organism. I'm going to go unbuild walls.

— Ursula K. Le Guin, The Dispossessed

1 INTRODUCTION

1.1 DNA IN SUMMARY

Deoxyribonucleic acid (DNA) is the ubiquitous molecule which describes the essence of our genetic makeup. It is composed of a phosphate-ribose backbone which polymerizes to create an arbitrary number of bases. Because the phosphate groups of the backbone are negatively charged, they repulse each other. This causes DNA to be stiff, with a bending rigidity of 0.188 to 0.224 N nm^2 and persistence length of approximately 50 nm. [1]

Each DNA base covalently bonds to a ribose on the polymer backbone. Bases occur in complementary pairs: adenine (A) matches with thymine (T), and cytosine (C) matches with guanine (G). AT pairs form two hydrogen bonds between strands, and CG pairs form three. This exact pairing allows for DNA duplication during mitosis and cell growth.

DNA bases possess either one or two aromatic rings (pyrimidines and purines, respectively). These rings are flat and sustain attractive *pi-pi* interactions when stacked atop one another. Between *pi*-stacking and hydrogen bonding, double-stranded DNA (dsDNA) binds together and forms a double helix. This double helix possesses two grooves: one larger (major) and one smaller (minor). Each of these grooves experiences a different chemical environment and binds differently to various proteins.

Friedrich Miescher discovered DNA in 1869. [2] In 1944 Avery, MacLeod, and McCary established that this strange molecule encoded genetic information, [2] but little about its structure. James Watson and Francis Crick, bolstered by the appropriated notes of Rosalind Franklin, were the first scientists to publish the double-helix structure for DNA. [3, 4]

1.2 DNA IN BIOLOGY

DNA is one of the primary molecules used for storing biological information. This information propagates from DNA through the processes of transcription (to RNA) and replication (during mitosis). DNA's exact pairing is essential for cellular reproduction and continuation of an organism's genome. Replication, however, is susceptible to errors and mutates the genome at a frequency of 1 per 10^7 base pairs per mitosis cycle; [5] this forms base pair mismatches, such as guanine-thymine (GT). DNA mismatches induce negative conse-

quences on human health, including greater mutation rates, [6] genetic defects, [7–9] and cancer. [9–12]

Mismatch repair (MMR) proteins are a class of proteins which "proofread" and correct any errors during DNA replication. [7, 13] MMR enzymes reduce the rate of errors in replication by 2-3 orders of magnitude, to 1 error per 10^9-10^{10} base pairs per mitosis cycle. [5, 14] The mechanism by which they recognize these mismatches remains unknown. We explore structural and dynamical differences in matched and mismatched DNA in an attempt to elucidate the underlying principles of MMR recognition trends.

1.3 DNA IN MATERIALS

DNA is a promising material for developing novel nanotechnologies, due to its unique properties. Not only can its discrete sequence store information, [15–17] but so-called DNA origami [18–20] applies strand specificity to bend and contort DNA into provocative shapes and structures. Self-assembling nanotechnologies often employ DNA bound to nanoparticles arranged systematically. [21–23] The nanoparticles will join only when in contact with other nanoparticles that display DNA of complementary sequence.

Incorporating additional functional groups into the DNA provides mechanisms to externally control DNA hybridization, enabling the development of stimulus-responsive materials. Photoswitches respond to light exposure, and photoswitch-modified DNA is applied to many sectors, including manipulation of gene expression, [24–27] DNA biosensing, [28–31] and light-controlled DNA nanomachines. [32–38] We explore the effects of azobenzene in DNA, especially as they relate to mismatches in the DNA polymer.

1.4 DNA IN THIS WORK

This dissertation explores three specific domains of DNA dynamics. Chapter 3 analyzes structural and dynamical differences between matched and mismatched DNA strands. We observe mismatched bases shift more than their matched counterparts in the plane perpendicular to the DNA strand. The mismatches possess fewer inter-base hydrogen bonds and are less stable. Several mismatched structures exhibit the e-motif when driven from their ground-state conformation.

Chapter 4 investigates the mechanism by which mismatched DNA bases affect their local environment and, consequently, neighboring azobenzene moieties. We find experimentally-determined quantum yield (QY) correlates with the variance of neighboring DNA bases. As mismatched bases are less stable, their spatial fluctuations increase

and the probability of free volume forming near the azobenzene also increases.

Finally, Chapter 5 adopts a coarse-grained model to explore the energetics of pulling DNA. We fit pulling simulation data to three mathematical models, and evaluate the benefits and drawbacks of each model. We find the Bell-Evans and Friddle-Noy-deYoreo models to interpret the underlying data in non-physical ways. The Hummer-Szabo model confers more realistic energy profiles, especially at slow pulling rates.

1.5 LITERATURE REVIEW

1.5.1 *Base Flip and Mismatch Repair*

Understanding the mechanism by which MMR proteins recognize mismatches is crucial to preventing the health risks mismatches pose. Many sources [13, 39–41] provide a general trend in the ability of MMR proteins to recognize specific mismatches. GT is one of the easiest to recognize, while CC is one of the most challenging:

$$\text{GT} \gg \text{CA} > \text{AC} \approx \text{TG} > \text{CT} \gg \text{TC} > \text{CC}. \tag{1.1}$$

Imhof et. al. [42] performed molecular dynamics (MD) simulations to determine the structural properties of a DNA polymer containing a GT mismatch. Sharma et. al. [43] studied many mismatched pairs and observed a correlation with the overall bend of the DNA polymer and mismatch identity. Sharma et. al. also performed MD simulations of base pair mismatches in the presence of different MMR proteins, showing the effects of protein active sites on the structure of mismatched DNA. [44]

Lau et. al. [45] discovered one of the most fascinating elements of MMR systems in 2000. They observed a crystal structure of an MMR protein not only bound to DNA, but actively flipping one of the DNA bases out of the double helix and into the protein's active site. Theoreticians have explored this phenomenon, known as base flip, since. We examine this and other structural properties for DNA mismatches in Chapter 3.

1.5.2 *Azobenzene in DNA*

Azobenzene is an effective molecular photoswitch due to its chemical stability and reversible photoisomerization. Exposure to UV light converts from the *trans-* to the *cis-* isomer, while blue light or increased heat induce the reverse conformational transformation. The mechanism of the isomerization is subject to debate, but recent experimental and computational evidence indicate it follows a hula-

twist pathway [46–56] that is volume-conserving in the DNA environment. [48] The size and polarity of azobenzene compares to DNA nucleobases, [57–59] making it an exceptional candidate to chemically modify DNA.

Asanuma et. al. [60, 61] demonstrated azobenzene can be incorporated into DNA using a D-threoninol linkage to the phosphate backbone. Despite azobenzene distorting the duplex DNA, the overall stability of the azobenzene-modified polymer is greater than that of the native duplex due to favorable *pi*-stacking interactions between the azobenzene and its neighboring nucleobases. [62–64] Reversible optical control of DNA hybridization and dehybridization is achieved by the photoisomerization of azobenzene: photoswitching *trans*-azobenzene to the non-planar *cis*-form produces steric hindrance which destabilizes the dsDNA, decreasing the melting temperature and unzipping the DNA at typical experimental temperatures. [60]

Recently, Yan et. al. [65] revealed the QY of azobenzene-modified DNA (azo-DNA), that is, the probability of azobenzene isomerizing after absorbing a photon, varies with the nearby sequence of neighboring base pairs. They hypothesize this change is due to the local "free volume" or empty space created by mismatched base pairs next to the azobenzene. We explore this hypothesis in Chapter 4.

1.5.3 *Pulling DNA with Atomic Force Microscopy*

Dynamic Force Spectroscopy (DFS) is used to study single-molecule interactions. [66–69] It can be applied to DNA by employing atomic force microscopy (AFM) [70, 71] wherein the AFM cantilever tip covalently binds to a DNA strand and a collection of corresponding strands bind to a nearby substrate. The AFM tip drops to the substrate, allowing the DNA strands to hybridize. It then retracts, and the resulting force is measured as the DNA melts.

Zhang et. al. [71] recently performed such DFS experiments on DNA in two configurations: pulling the DNA apart from opposite termini of the strand (shear, Fig. 5.1a) and pulling from the same terminus (unzip, Fig. 5.1b). The Second Law of Thermodynamics asserts that because the initial (hybridized) and final (unbound) states are the same, the free-energy difference of this process should be the same for both. In fact, after applying a theoretical model, Zhang et. al. found the two geometries to express different changes in free-energy. We explore in Chapter 5 how the assumptions from various models may lead to such a result.

1.5.4 *Simulation and Force Field Development*

Banavali et. al. [72] defined a DNA base flip reaction coordinate that could be measured during MD simulations; this coordinate described the pseudo-dihedral angle between a base, its ribose, the neighboring ribose, and the neighboring base pair. Song et. al. revealed in 2009 [73] that by updating their reaction coordinate to utilize the base, the flanking phosphates, and all four neighboring bases (Fig. 3.2) they could improve the stability of the restraints and better describe the degree of flipping. We employ this updated coordinate in Chapter 3.

McCullagh et. al. [74, 75] developed force field parameters for an azobenzene-containing DNA strand where the azobenzene was covalently bonded on either side to a DNA strand. We modify their work in Chapter 4 to create a force field for azobenzene adhered to only one strand of the DNA.

Flexible polymer models have been used to model proteins for some time; [76–79] Mishra et. al. [80] apply this model to DNA and we adapt it for use in the coarse-grained pulling simulations explored in Chapter 5.

2 GENERAL METHODS

2.1 BACKGROUND THEORY

One of the alluring benefits of simulating molecules is that we obtain perfectly-resolved detail over the course of a simulation. For Chapters 3 and 4, we employ MD at the atomistic level; that is, we know the position, velocity, and energy of each atom at each timestep. For Chapter 5 we adopt an in-house coarse-grained model which represents each DNA base as a bead in a chain.

We apply these simulation methods to quantify the differences in structure between DNA strands with matched and mismatched base pairs, both with and without azobenzene. For some simulations, we bias the dynamics to flip bases into solution, or to pull the strands apart. These biases allow us to measure the energy associated with such perturbations.

The underlying principle of MD is to apply Newton's laws of motion at the molecular level. A collection of potential energies acting on and between atoms is called a force field. These energies describe everything from the spring-like behavior of bonded atoms and the attraction of oppositely-charged particles to complex dihedral angles. At each timestep, we sum the potentials on each atom, and from this total, we can compute the force:

$$\vec{F}_i = -\vec{\nabla}_i U \tag{2.1}$$

where $\vec{F}_i$ is the force on the ith particle, $\vec{\nabla}_i$ is the three-dimensional gradient with respect to the position of the ith particle, and U is the potential-energy of the system.

Having the force and knowing the mass m_i of a particle (which we acquire by identifying the atom), it is straightforward to calculate the acceleration, $\vec{a}_i$:

$$\vec{F}_i = m_i \cdot \vec{a}_i. \tag{2.2}$$

From there, it is simple to integrate Newton's laws to calculate both the velocity and position of each atom. The forces, velocities, and positions can then be updated. Iterating over these steps reveals emergent dynamics of the whole molecule.

2.2 DETAILED METHODS

2.2.1 *Building Atomistic Models*

We construct atomistic models of B-DNA containing twelve base pairs by employing the 3DNA software. [81] We utilize the sequence CTGA ACXA ATGT, where the placeholder X represents any one of the four nucleotides. Bases 1-6 and 8-12 pair with matched nucleotides, whereas X pairs with either a matched or a mismatched base, denoted Y. We adopt the notation XY to identify our considered sequence. The models employed in Chapter 4 include an azobenzene molecule, which we add manually with VMD. [82] We study a total of twenty systems: eleven without azobenzene (Ch. 3) and nine with (Ch. 4). The non-azobenzene systems studied in Chapter 3 contain four matched pairs (XY $\in$ {AT, TA, CG, GC}) and seven pairs with a single mismatch (XY $\in$ {AA, AC, AG, CC, CT, GT, TT}). We do not include the GG mismatch because steric influences prevent a hybridized equilibrium state. The sequences of azobenzene-containing systems of Chapter 4 are displayed in Table 4.1.

We minimize each model for 200 steps of steepest descent by applying the CHARMM27 Force Field, [83] which we employed throughout. We solvated the structures with explicit TIP3P water [84] in a dodecahedral simulation box extending at least 1 nm beyond the DNA polymer in all directions. We randomly substituted water molecules for 22 Na^+ counter-ions to neutralize the system. The solvent was minimized for 200 steps of steepest descent and equilibrated for 100 ps. We equilibrated each full system for 1 ns.

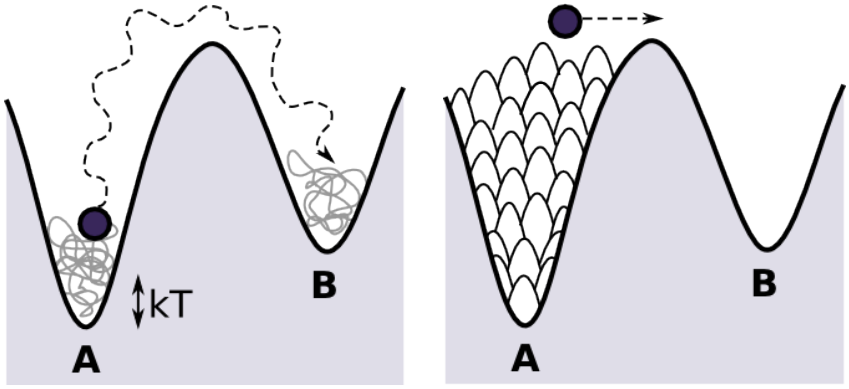
2.2.2 *Atomistic Molecular Dynamics Simulations*

We performed simulations at 300 K in the NVT ensemble with stochastic velocity rescaling, [85] periodic boundary conditions, and a 2 fs timestep employing the Gromacs simulation suite. [86] We utilized the SHAKE algorithm [87] to constrain covalent bonds which include a hydrogen atom, and treated long-range electrostatic interactions with the Particle Mesh Ewald method [88] with a non-bonded cutoff of 1.2 nm. After equilibration, we performed unbiased simulations for 100 ns and saved trajectory coordinates every 10 ps.

2.2.3 *Metadynamics*

When using MD, sometimes the behavior or dynamics one might attempt to measure occur on time scales longer than is currently feasible to simulate. As such, it is necessary to employ biasing methods which distort the potential-energy landscape and drive the system

to less-favorable configurations. We choose to apply well-tempered metadynamics. [89]



(a) unbiased (b) biased

Figure 2.1: A cartoon representation of metadynamics. In an unbiased trajectory (a) the system may not possess enough energy to overcome a barrier and proceed from state A to state B. We would be unable to suggest anything about the energetics of B. Metadynamics (b) adds Gaussian-shaped "hills" to the potential-energy, driving the system out of equilibrium and into unexplored states. The free-energy surface is reconstructed by inverting the added hills.

Metadynamics biases the system by adding Gaussian potentials to the overall energy. These potentials are added along a predefined reaction coordinate at a set rate. (Fig. 2.1) Once the simulation converges, the sum of potentials added to the system describes the inverse of the free-energy surface. An appropriate analogy is "metadynamic sand." The free-energy surface in this metaphor is the bucket and the metadynamics is the sand which is gradually added. By inverting the bucket, the sand holds its shape. We could reconstruct the form of the bucket by observing the surface of the sand. Similarly, we can reconstruct the free-energy profile by observing the added hills.

One of the major obstacles of metadynamics is determining when the trajectory has converged to an appropriate free-energy. A run is converged when it meets three criteria:

A. The difference in free-energy, ΔG, between local energy minima converges to a fixed value.

B. The system explores the entire reaction coordinate, not localizing into any specific place.

C. For well-tempered metadynamics, a variant in which the appended energy decreases as the reaction coordinate is explored, the height of added Gaussian peaks approaches 0.

We display a representative simulation in Figure 2.2. In this example, the ΔG (detailed calculations for which are discussed in Chap-

ter 3) converges at around 400 ns. The reaction coordinate, similarly at 400 ns, ceases to be "stuck" at any one point. All three criteria are satisfied at 600 ns when the height of the added Gaussians becomes < 0.03, $1/100^{\text{th}}$ of its starting value of 3 kJ/mol (not pictured).

2.2.4 *Coarse-graining*

Atomistic molecular dynamics approximates an exact quantum system. The size, charge, and other properties of each atom approximate the position and behavior of its nucleus and electrons. It would be an unfeasible computational problem to model these large systems explicitly.

Similarly, we can incorporate additional approximations to allow for computing larger systems. In Chapter 5 of this work, we approximate each DNA base as one singular bead. In this model, we represent a double-stranded DNA molecule by two chains of beads, which "bond" to the opposite strand with specificity.

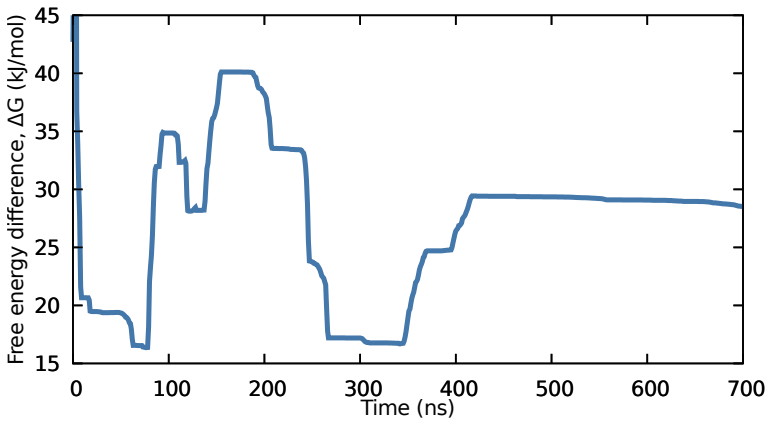
2.3 SIMULATION LIMITATIONS

MD simulations embrace copious advantages, though they struggle to represent DNA in certain contexts. Force field parameters are usually developed to match crystal structures, and can fail to capture the structure and dynamics in solution. Primarily, energy calculations do not always equate to experimentally-determined values. [90] This can lead to artifacts such as over-sampled meta-stable minima. Periodic boundary conditions can introduce artifacts if the simulation box size is too small; [90] even with a large box, an image interacting with itself through long-range water solvation shells may still bias the results.

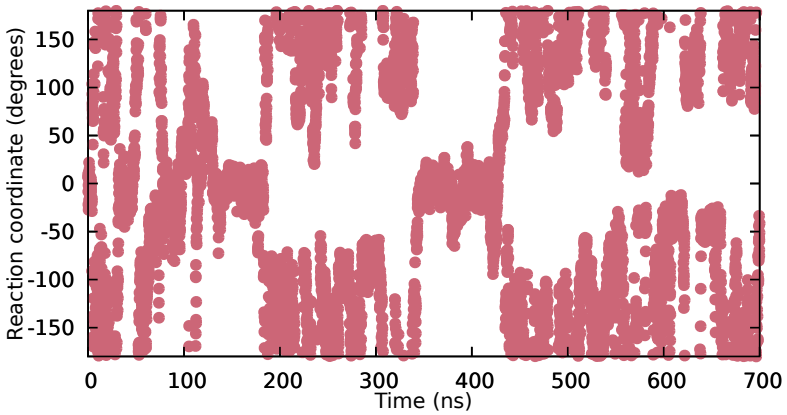
Specifically, the CHARMM27 force field [83] represents DNA in its A-form instead of the more native B-form. [91] The average rise between bases is lower than for experimentally-measured B-DNA [91] and sugar puckering of the backbone is over-represented. [92] Other force fields similarly suffer, [91, 92] highlighting the importance of validating theoretical measurements with experimental data.

Simulation time scales are briefer than experimental time scales, e. g.nanoseconds as opposed to seconds, minutes, or hours, and thus effects such as DNA pulling (explored in Ch. 5) occur at faster speeds in simulation. For methods which are sensitive to velocity, (explored in Ch. 5) these higher speeds can create additional divergence from experimental results.

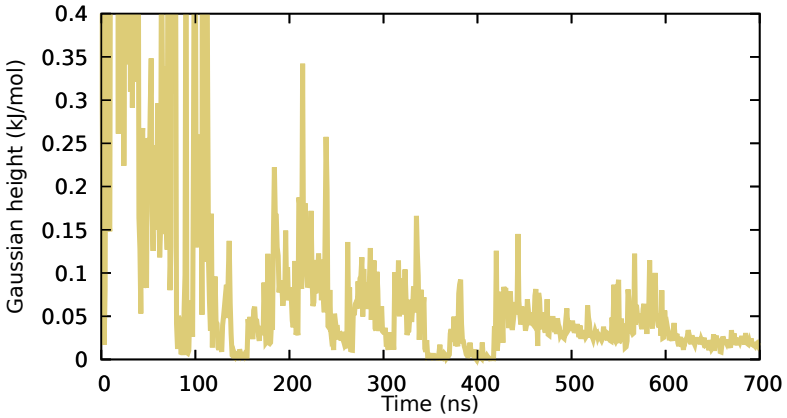
It is noteworthy that nucleic acid force fields are validated on canonical structures, with two hybridized strands in a full-stacked double helix. Any large deformations would alter the electronic distribution of the molecule and engender the underlying force field



(a) ΔG vs. time



(b) reaction coordinate vs. time



(c) Gaussian height vs. time

Figure 2.2: A representative metadynamics calculation, exhibiting the difference in free-energy, ΔG (a), the biased reaction coordinate (b), and the Gaussian hill height (c) over time. The simulation is mostly converged by 400 ns when ΔG has converged to a constant value and the reaction coordinate ceases being "stuck" in one position. The system converges after 600 ns, when the added hill height becomes and remains minute.

less applicable. Polarizable force fields [93] will begin to solve this problem, but until we can perform quantum-level calculations on a macromolecular scale, exact energies and properties of non-canonical nucleic acids will be difficult to simulate.

Coarse-grained simulations are similarly limited, and possess an additional loss of detail when combining atoms into a single bead. The model explored in Chapter 5 does not include any native twist or helix to the polymer and contains twenty unique base types rather than four. What is lost in accuracy is made up for in quantity: with a coarse-grained model we can simulate tens of thousands of trajectories in the time needed to perform one atomistic simulation. Trends extracted from the model are likely to be accurate, but quantitative values should be viewed with a skeptical eye.

Part II

APOLOGUE

Fear and realisation of ignorance, strong medicines against stupid pride.

— Garth Nix, Sabriel

The purpose of a storyteller is not to tell you how to think, but to give you questions to think upon.

— Brandon Sanderson, The Way of Kings

3

INCONGRUITY: DNA MISMATCHES, THEIR STRUCTURE, AND BEHAVIOR

3.1 INTRODUCTION

In this chapter, we explore the differences in structure between matched and mismatched base pairs. We use several strucutral order parameters first defined by Yoon et. al. [94] that we have depicted in Figure 3.1. Six of the parameters (shear, buckle, stretch, propeller, stagger, opening) describe the structure between a base and its pair; the remaining six (shift, tilt, slide, roll, rise, twist) describe the structure between a base pair and its neighboring pair. We show significant difference between matched and mismatched pairs for some bases, especially for those parameters defined in the x-y plane.

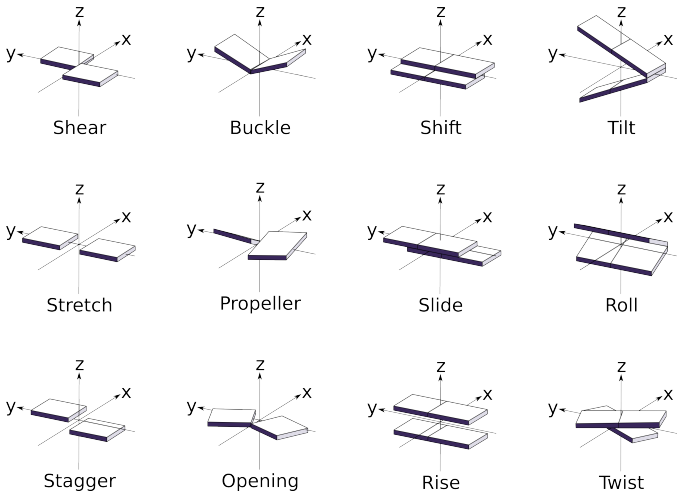


Figure 3.1: The structural parameters employed in this paper. Image adapted from Lu et. al. [81]. Parameters measured between a base and its pair are located in the left two columns, between a base pair and the neighboring pair are in the right two columns. Translational parameters are in the first and third columns, rotational in the second and fourth. Respect to the x-, y-, and z-axis are the first, second, and third row, respectively.

While studying the structure of these mismatched systems, we also calculated information on hydrogen bonding between base pairs. Guanine (G) and Cytosine (C) form three hydrogen bonds when paired with one another. Adenine (A) and Thymine (T) form two hydrogen bonds with one another. As hydrogen bonds are the main driver of stability in the DNA, and are solely responsible for the hybridization of two DNA strands, an AT pair is less stable than a GC

pair. We show decreased hydrogen bonds, and thus decreased stability, for mismatched base pairs.

To further elucidate the physical properties of mismatched base pairs, we compute the energy required to rotate bases out of the DNA double helix. This rotation, called base flip, [73] is a crucial process by which enzymes detect and selectively repair post-replicative errors. Proteins of the mismatch repair (MMR) system proofread the DNA; upon encountering a mismatch, an MMR enzyme flips the offending nucleotide out of the DNA double helix and into its active site before proceeding with the repair pathway. [9, 11, 45, 95, 96] MMR proteins decrease the mutation rate between 50-1000 times, to a rate of 1 per 10^9-10^{10} base pairs per cell division. [5, 14]

Base flip requires an activation energy exceeding thermal fluctuations, and is unlikely to occur on the timescale of typical molecular dynamics (MD) simulations. To rectify this obstacle, we utilize the well-tempered metadynamics algorithm; driving the simulation away from the ground-state. The free-energy profiles in some mismatched pairs demonstrate a metastable state in which both bases flip out of the DNA strand and nest in the minor groove. This may contribute to MMR proteins' ability to recognize mismatches. Our results do not indicate a systematic free-energy difference between the flipped and canonical structure as a quantitative predictor of mismatch recognition, largely due to this unusual minor-groove-nested structure. The presence of the structure does not predict MMR recognition, which highlights the importance of analyzing additional physical properties such as DNA bending. [43, 44, 96–98]

3.2 METHODS

The methods are exactly as specified in Chapter 2, unless as outlined below.

3.2.1 *Structure, Hydrogen Bonding, and Base Flip*

Roll, tilt, twist, slide, rise, shift, buckle, propeller, opening, stagger, shear, stretch, (pictured in Fig. 3.1) and inter-base hydrogen bonding were calculated using 3DNA; [81] probability distributions and free-energy profiles were determined by in-house computer code. We measure bend as by Sharma et. al.; [43] we subsection the DNA into three portions (bases 2-4/21-31, 5-8/17-20, and 9-11/14-16) and calculate the angle between the centers of mass of the heavy atoms. To measure base flip, we adopt a pseudo-dihedral angle first described by Song et. al., [73] defined as the dihedral angle between points P1, P2, P3, and P4 where P1 is the center of mass of the four neighboring base pairs, P2 and P3 are the centers of mass of the phosphates which flank the flipping base, and P4 is the center of mass of the pyrimidine

ring, or the five-membered ring of the purine. (Fig. 3.2) Changes in base flip angle are positive if X crosses into the major groove or if Y crosses into the minor groove, and negative otherwise. To identify representative configurations, we applied the `Biopython` [99] module `cluster.kmedoids`.

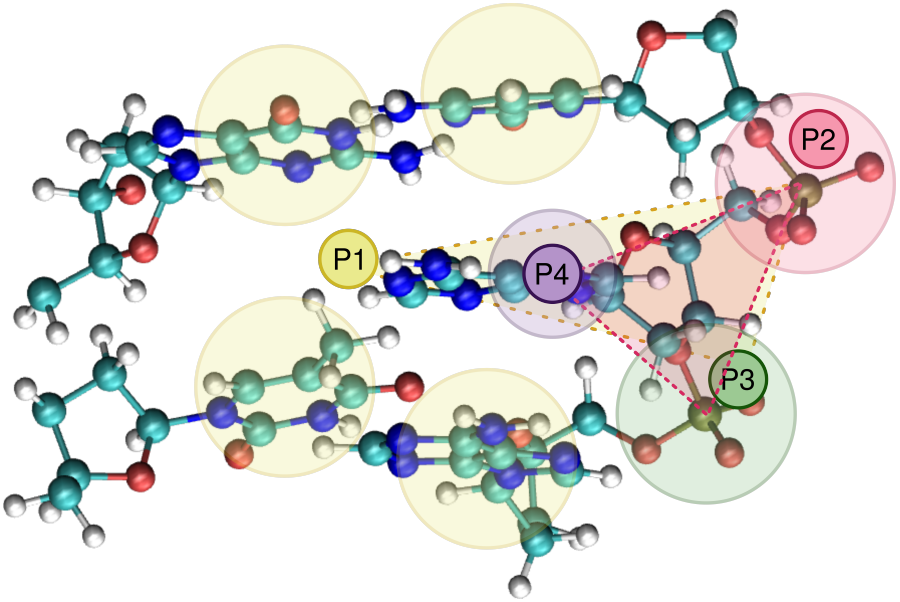


Figure 3.2: Definition of the base flip angle. We define the pseudo-dihedral angle as being between points P1, P2, P3, and P4; P1 is the center of mass of the neighboring base pairs (4 bases in total), P2 and P3 are the centers of mass of the phosphates flanking the flipping base, and P4 is the center of mass of the five-membered ring (of a purine) or the six-membered ring (of a pyrimidine). The two intersecting planes describing the dihedral angle are lightly shaded in yellow and red, and are outlined with dashed lines. For clarity, the opposing base is not pictured.

3.2.2 *Metadynamics Simulations*

We employ well-tempered metadynamics simulations with two collective variables: the base flip angle of X and the base flip angle of Y. For comparison, we performed some simulations in which we adopt only one of these two base flip angles as a collective variable. Using the PLUMED plug-in, [100] we add Gaussian biasing potentials to the collective variable(s) with an initial height of 3 kJ/mol, width of 0.08 radians, a deposition rate of 1 ps, and a well-tempered bias factor of 6, for a total simulation length of at least 700 ns. We reconstruct the 2D free-energy surface as the inverse of the potential-energy added to the system as described by Barducci et. al. [89]. We pinned the base pairs terminating the DNA strand and the base pairs neighboring the flipping bases by their inter-pair hydrogen bonds to prevent the sys-

tem from artificially melting. We constrain each hydrogen bond to a distance of 0.2 nm with a harmonic constant of 40 kJ mol^{-1} nm^{-2}.

3.2.3 *Evaluating Correlations*

For any two order parameters x and y, we calculated the covariance as

$$\text{cov}_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_{xx}\sigma_{yy}}} \tag{3.1}$$

where

$$\sigma_{xy} = \langle xy \rangle - \langle x \rangle \langle y \rangle. \tag{3.2}$$

Here, the angular brackets denote the equilibrium average,

$$\langle x \rangle = \iint_{-\pi}^{\pi} \mathrm{d}x\mathrm{d}y \, xP(x,y), \tag{3.3}$$

where the probability distribution, $P(x,y)$, is proportional to the exponent of the negative free-energy:

$$P(x,y) \propto \exp\left(\frac{-G(x,y)}{k_\text{B}T}\right). \tag{3.4}$$

For well-tempered metadynamics, we reconstructed free-energy diagrams for structural order parameters utilizing the method described by Bonomi et. al. [101]

3.2.4 *Calculating Free-Energy Differences*

We convert two-dimensional free-energy profiles to one-dimensional free-energy profiles by integrating:

$$G(x) = -k_\text{B}T \log\left[\int_{-\pi}^{\pi} \mathrm{d}y \, e^{-G(x,y)/k_\text{B}T}\right], \tag{3.5}$$

where G is the free-energy function, y is the variable integrated out of the profile, x is the remaining variable, and $k_\text{B}T$ is the Boltzmann constant multiplied by the simulation temperature.

We calculate the free-energy difference, ΔG, by subtracting the free-energy of the hybridized conformation from that of the solvent-facing conformation. To calculate the free-energy of each state, we evaluate

$$G_\text{state} = -k_\text{B}T \log\left[\int_{x_\text{min}}^{x_\text{max}} \mathrm{d}x \, e^{-G(x)/k_\text{B}T}\right], \tag{3.6}$$

where x_min and x_max are the left and right boundaries of the state, defined by the flanking global maxima.

3.3 RESULTS AND DISCUSSION

3.3.1 *Structural Order Parameters*

For each system, we calculate the probability distributions of the DNA order parameters buckle, opening, propeller, rise, roll, shear, shift, slide, stagger, stretch, tilt, and twist from unbiased MD simulations. These distributions (and the corresponding free-energy profiles) give valuable information regarding the structures of the hybridized ground-state and of the base pair of interest. Imhof and Zahran [42] calculated these distributions for the AT and GC matches as well as the GT mismatch. For these base pairs, our results agree with theirs, despite our employing different DNA sequences.

For all studied base pairs, we measure the structure of matches and mismatches by calculating the probability overlap:

$$\int_{-\infty}^{\infty} \mathrm{d}x \min\Big(P_{\mathrm{mis}}(x), P_{\mathrm{mat}}(x)\Big), \tag{3.7}$$

which is the integrated area of the minimum of the mismatched probability density (P_{mis}) and the mean of the corresponding matched probability densities (P_{mat}). This measure of overlap ranges from 0 to 1, where 0 indicates complete dissimilarity of structure and 1 indicates exact similarity. We provide an illustrative example in Figure 3.3.

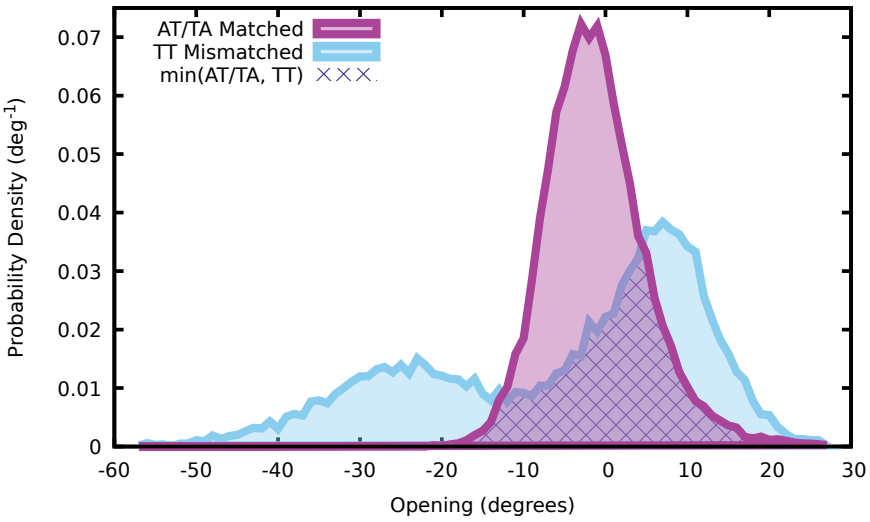


Figure 3.3: Probability overlap for opening. The mean probability density of AT and TA matched base pairs (magenta) relative to the probability density of the TT mismatched system (blue). The area in common (hatched) is the probability overlap; for this example it has a value of 0.47.

We calculate overlap values and present the results in Table 3.1. Our data demonstrate that, on average, GT is the most similar to its

	AA	CC	TT	AG	AC	CT	GT	AVG.
stretch	0.00	0.03	0.01	0.00	0.18	0.05	0.09	0.05
shear	0.00	0.02	0.00	0.76	0.00	0.64	0.00	0.20
opening	0.35	0.38	0.47	0.21	0.53	0.41	0.72	0.44
twist	0.29	0.34	0.23	0.91	0.46	0.96	0.44	0.52
shift	0.54	0.45	0.86	0.59	0.64	0.67	0.77	0.65
tilt	0.29	0.85	0.74	0.52	0.84	0.61	0.93	0.68
slide	0.37	0.62	0.77	0.64	0.90	0.80	0.85	0.71
stagger	0.50	0.78	0.79	0.56	0.80	0.71	0.80	0.71
propeller	0.63	0.80	0.83	0.85	0.78	0.63	0.67	0.74
buckle	0.73	0.76	0.93	0.88	0.86	0.78	0.94	0.84
rise	0.85	0.70	0.77	0.91	0.93	0.92	0.95	0.86
roll	0.86	0.87	0.95	0.84	0.84	0.89	0.93	0.88
bend	0.93	0.92	0.95	0.77	0.88	0.92	0.95	0.90
AVG.	0.49	0.58	0.64	0.65	0.66	0.69	0.70	

Table 3.1: Overlap between order parameter distributions of mismatched and corresponding matched base pairs, Eq. 3.7, ordered both top-to-bottom and left-to-right by increasing overlap (as averaged over rows and columns, respectively). Values range from 0 to 1, where high numbers signify similar structure while low numbers indicate matched and mismatched structures differing.

matched counterparts. This is an unintuitive result, as GT is one of the easiest mismatches for MMR proteins to identify.[13, 39–41] In a similar vein, CC is one of the most difficult for MMR proteins to recognize, [13, 39–41] but it possesses one of the structures most divergent from matched configurations. This trend holds for bend, which has previously been determined to correlate with base flip, [43, 44] and is thus a likely mechanism for MMR recognition.

Our data reveal the distributions of stretch, shear, opening, twist, and shift vary the most when we compare mismatched with matched base pairs. These order parameters are within the plane perpendicular to the DNA. Order parameters presenting negligible differences include roll, rise, buckle, propeller, and stagger; all defined outside of the perpendicular plane. These results indicate that replacing a matched base by a mismatched one distorts the DNA, mostly within the plane of the base pair, not along the length of the molecule. These distortions do not appear to correlate with MMR enzymes' ability to identify mismatches. [13, 39–41]

3.3.2 *Hydrogen Bonding*

We calculate the mean number of hydrogen bonds between the two bases of interest, X and Y. We present our results in the second column of Table 3.2. As expected, the average number of hydrogen bonds for the matched pairs AT and TA is approximately two, and for pairs CG and GC it is approximately three. Mismatched base pairs contain a consistently lower number of hydrogen bonds than their matched counterparts. Because hydrogen bonding between opposing bases determines DNA stability, [102–107] these results indicate decreased stability in the presence of mismatches; a result realized experimentally. [108–111]

According to our results, hydrogen bonds do not systematically correlate with any of the analyzed structural parameters (Fig. 3.4), though weak trends appear. The correlation with stretch, if present, is always negative. This is an expected result; as bases drift from one another, their ability to hydrogen bond decreases. Interestingly, the structures of mismatched pairs CT, GT, and TT correlate most with hydrogen bonding than any other pair.

3.3.3 *Base Flip*

Unbiased simulations are insufficient to determine the free-energy of base flip. The free-energy barrier to flip out of the DNA polymer and into solution is too immense to overcome on the simulation time scale. To ameliorate this complication, we applied the well-tempered metadynamics algorithm [89, 112] which computes the free-energy surface (FES) of base flip over the entire domain. This technique con-

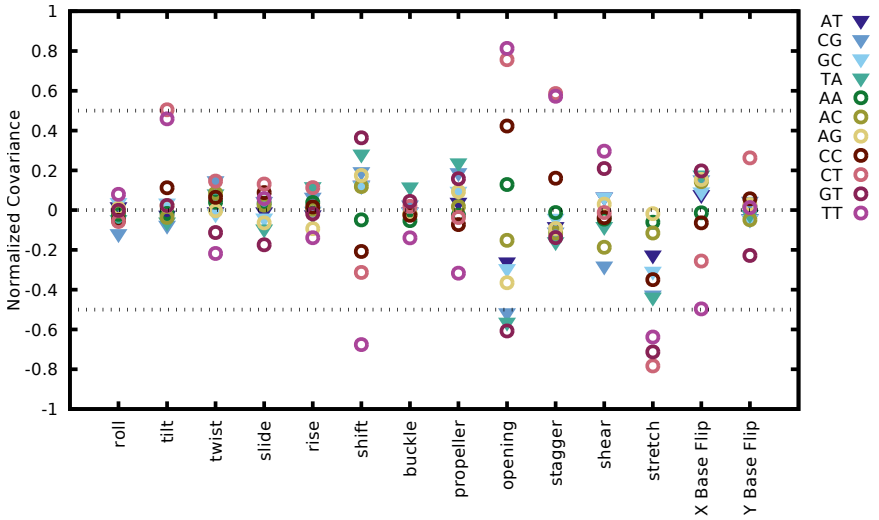


Figure 3.4: The normalized covariance between each structural parameter explored in this study and the number of hydrogen bonds. Solid triangles correspond to matched base pairs; hollow circles correspond to mismatched pairs. "X Base Flip" and "Y Base Flip" refer to the flip angles of bases X and Y, respectively.

	ns UNBIASED	MEAN # HBONDS	ns BIASED	ΔG_X (kJ/mol)	ΔG_Y (kJ/mol)
AT	100	1.99	700	29	26
TA	100	1.95	700	28	71
CG	100	2.97	700	22	68
GC	100	2.98	700	11	7
AA	100	1.01	1000	40	60
AC	100	1.96	700	-1	-1
AG	100	1.08	700	19	56
CC	100	1.08	700	8	8
CT	100	2.41	700	23	16
GT	100	1.97	700	8	21
TT	100	1.62	700	10	3

Table 3.2: First column: the time (in nanoseconds) of unbiased simulation. Second column: The mean number of hydrogen bonds between the bases of interest during unbiased simulation. Third column: the time (in nanoseconds) of biased metadynamics simulation. Fourth and fifth columns: The free-energy of base flip for the first and second (X and Y of XY, respectively) base of each pair.

tinually adds Gaussian-shaped biasing potentials to each reaction coordinate, driving the system toward previously unexplored configurations.

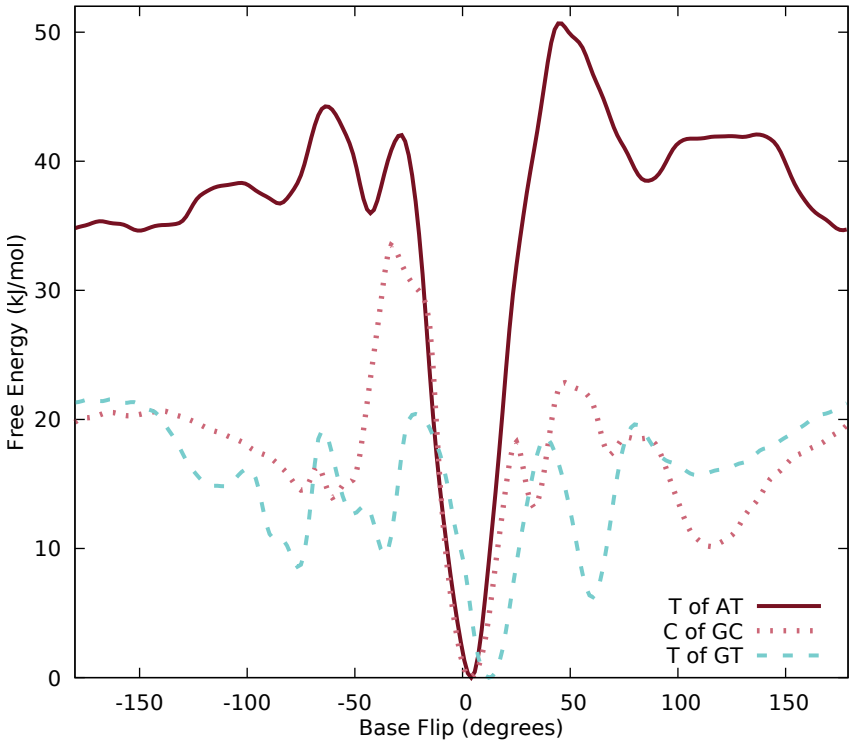


Figure 3.5: The free-energy of flipping the T base out of a mismatched GT pair (dashed blue), compared to the matched counterparts AT (solid red) and GC (dashed red). The mismatched GT pair shows a slight shift in the global minimum.

We compute a FES of base flip for each matched and mismatched pair. We illustrate a representative result in Figure 3.5, which compares the free-energies of flipping a T base in a GT mismatch with those of the corresponding T base in an AT matched pair and the C base of a matched GC pair. We calculate one-dimensional profiles by Eq. 3.5. Each of these bases possess the same local environment as the others. For each system, we identify a global minimum in the FES corresponding to the hybridized ground-state. Extrahelical conformations form local minima separated from the principal minimum by activation barriers. Similar results have been reported in the literature for a subset of the base pairs considered in this study [42, 72, 113].

Base flip may resemble other order parameters when a base rotates into the DNA polymer; indeed, it correlates weakly with twist and shear in unbiased simulation. Once the base rotates out of the DNA strand it does not correlate with any of the other order parameters. (Fig. 3.6) We therefore conclude base flip is a useful descriptor of base pair structure, building on the thirteen order parameters listed

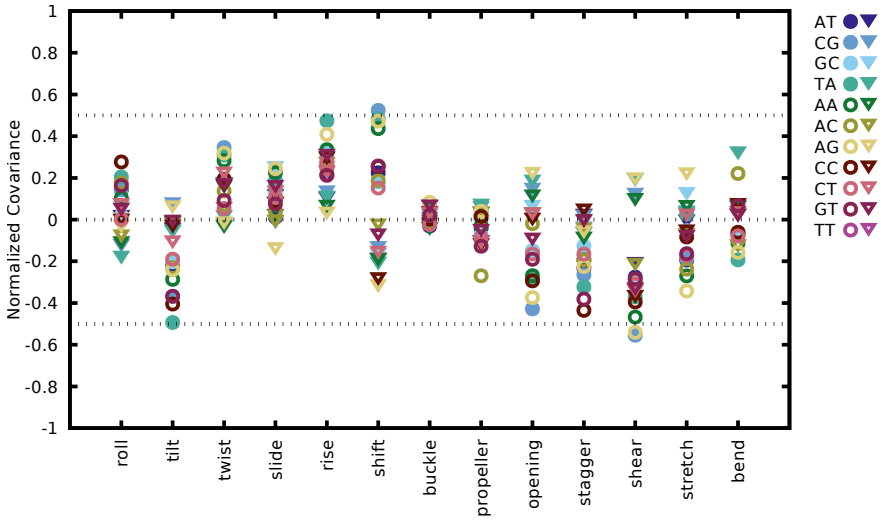
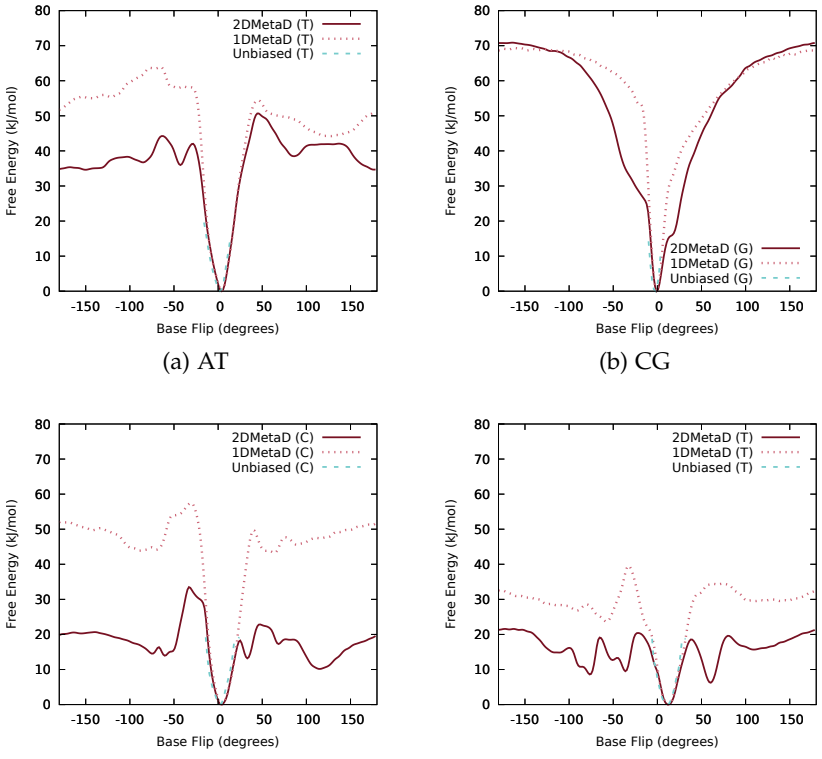


Figure 3.6: The normalized covariance between base flip and each structural order parameter explored in this study, over the entire base flip reaction coordinate. Solid shapes represent matched pairs; hollow shapes represent mismatched pairs. Circles represent the first base of a pair (e. g."A" of "AT"); triangles are the second base.

in Table 3.1. Despite no significant correlations, slight trends appear. For most structural parameters, the first and second bases of a pair have opposing correlations. For example, the parameter "shift" correlates positively with the base flip of the first base and negatively with the second. This is consistent with the opposing symmetries of hybridized DNA strands.

Base flip may be difficult to converge due to its long relaxation timescale; Imhof et. al. [42] observed a notable ($\sim$10°) shift of the minimum from unbiased to biased simulation of the TG mismatched system. They hypothesize the long timescale of flipping and re-stacking causes the free-energy to converge slowly. We notice a similar shift when testing their base flip reaction coordinate, first defined by MacKerell et. al., [72] with 50 ns calculations and biasing only one base, but no such shift when adopting the updated base flip coordinate [73] with a bias potential applied to two bases. This would indicate defining base flip is crucial not only for accuracy, as discussed in ref. [73], but also for convergence.

Simulations which bias two bases exhibit lower free-energies than those which bias only one base. (Fig. 3.7) Because a lower free-energy is favorable, this indicates the 2D biasing potentials possess greater accuracy than their 1D counterparts. The 2D potentials allow the system to explore other favorable states, and better represent the probability landscape of the force field. With our 2D potential we identify configurations we would otherwise not observe biasing only one base; this

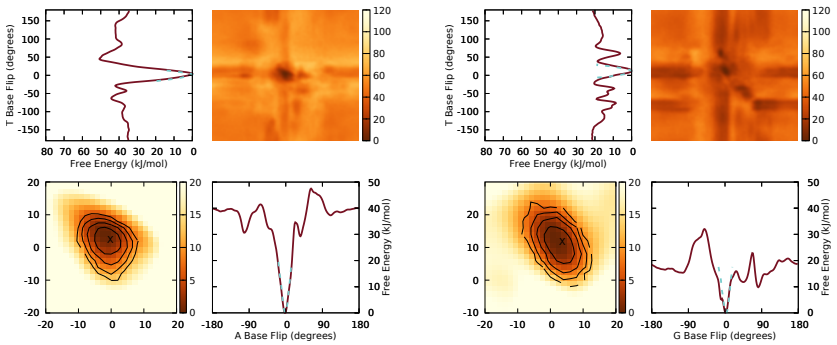


(a) AT (b) CG

(c) GC (d) GT

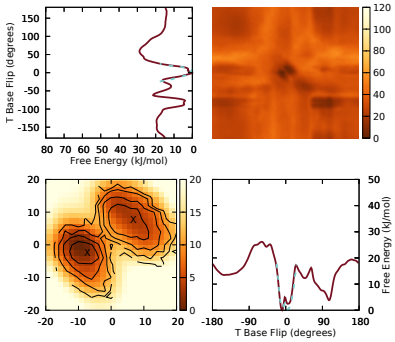
Figure 3.7: Comparison of base flip free-energy profiles for four systems: (a) T of an AT pair, (b) G of a CG pair, (c) C of a GC pair, and (d) T of a GT pair. Shown are results from unbiased MD simulations (dashed blue), metadynamics simulations in which only the flip angle of the T base was biased (dashed red), and metadynamics simulations in which the angles of both bases were biased (solid red). The 1D free-energy profiles are consistently greater than or equal to their 2D counterparts, indicating the 2D biasing potentials are more accurate.

explains why our results disagree with previous studies: they each bias one base where we bias two. [42, 72, 73, 96, 114]

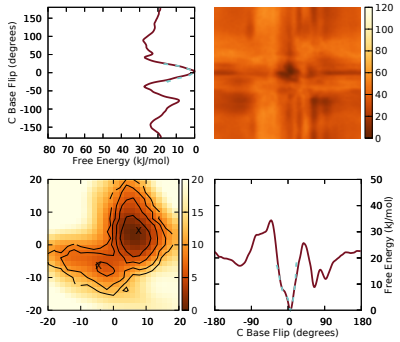


(a) AT

(b) GT



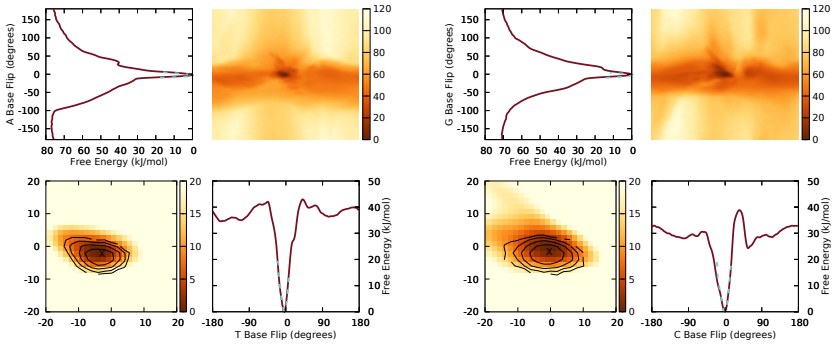
(c) TT



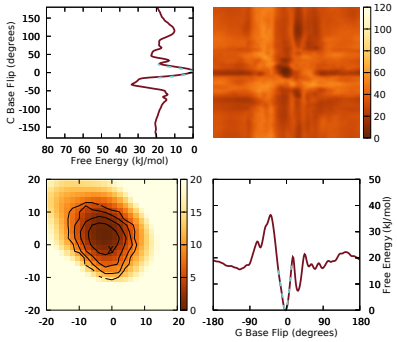
(d) CC

Figure 3.8: Free-energy of base flip for the matched pair (a) AT and (b-d) the mismatched pairs GT, TT, and CC. Shown in the upper right panel of each sub-figure is the two-dimensional free-energy surface as a function of the base flip angles of both bases, as computed from metadynamics simulations. The top left and bottom right panels depict one-dimensional projections of this surface onto the base flip angle of the first (top left) and second (bottom right) base, together with results from unbiased simulations (dashed line). The bottom left panel grants a magnified view of the primary minimum in the two-dimensional free-energy surface to 20 kJ/mol (heat map), together with results from unbiased simulations (contour lines). Flip angles demarcated by crosses correspond to local minimum configurations, as presented in Fig. 3.11. Notably, a secondary minimum appears in multiple simulations at approximately (90°, -100°). This corresponds to the configuration presented by Fig. 3.12.

For most configurations studied, a lower-energy "plus" shape is apparent in the 2D FES. We display representative free-energies in Fig. 3.8; the remaining profiles are depicted in Figs. 3.9 and 3.10. This result indicates it is easier to flip only one base out of the polymer, rather than both simultaneously. The pairs CC and TT both exhibit two minima at the flipped-in ground-states. We clustered the unbiased trajectories to identify molecular geometries corresponding to

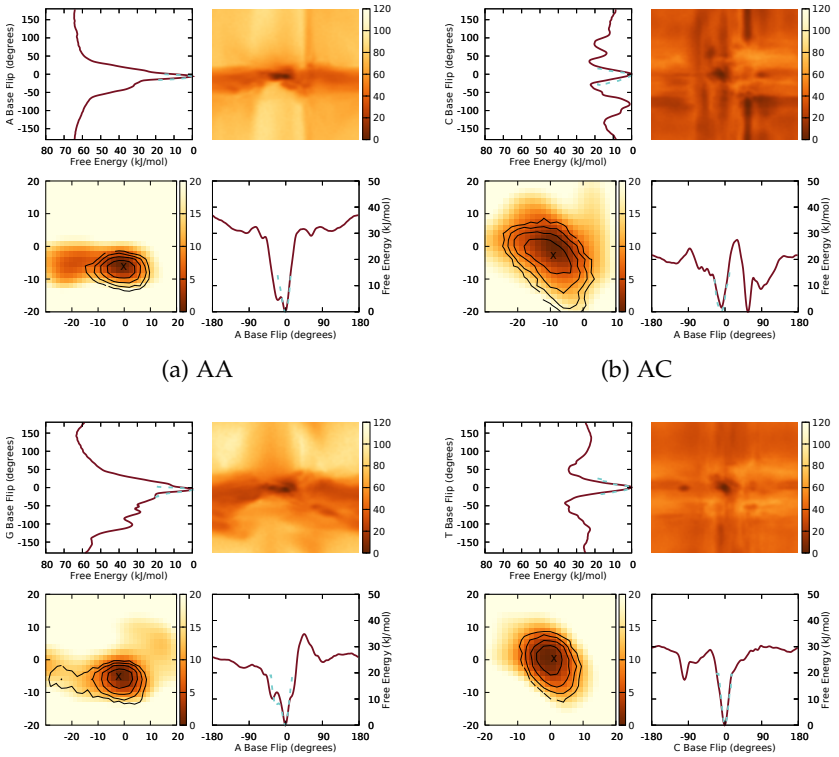


(a) TA (b) CG



(c) GC

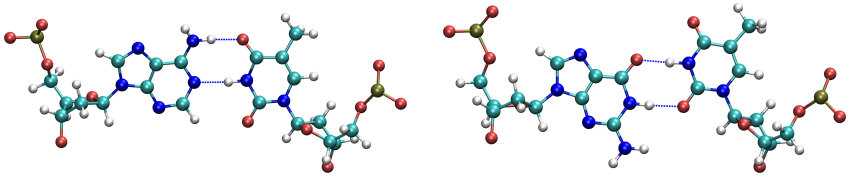
Figure 3.9: Free-energy of base flip for the matched pairs (a) TA, (b) CG, and (c) GC. Figure 3.8 describes each sub-figure.



(a) AA (b) AC

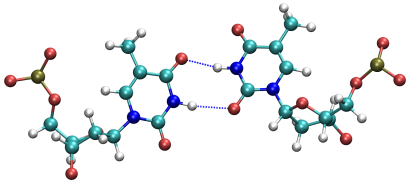
(c) AG (d) CT

Figure 3.10: Free-energy of base flip for the matched pairs (a) AA, (b) AC, (c) AG, and (d) CT. Figure 3.8 describes each sub-figure.

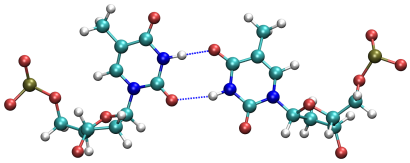


(a) AT minimum

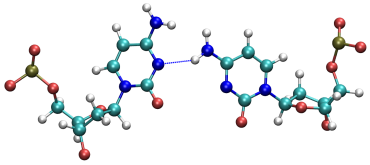
(b) GT minimum



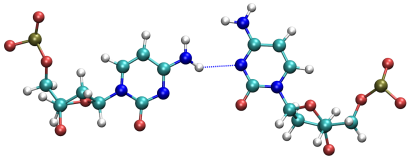
(c) TT primary minimum (∼52%)



(d) TT secondary minimum (∼48%)



(e) CC primary minimum (∼51%)



(f) CC secondary minimum (∼49%)

Figure 3.11: Typical structures for AT, GT, TT, and CC pairs as determined by clustering analysis. Hydrogen bonds as identified by the 3DNA software [81] are expressed as dashed lines. The base flip angles of these configurations are presented in Figure 3.8.

these minima: different hydrogen bonding configurations illustrated in Fig. 3.11. Several of these structures have been observed experimentally. [108, 115–118] In contrast, all matched pairs include a single relevant stable state with a nearly-isotropic free-energy minimum. Unlike previous works [42, 72, 73, 96, 114] which calculated the free-energy profiles of one flipping base, we observe a local minimum at roughly (90°, 100°) in most free-energy profiles. This minimum is largely unexpected, especially as it produces a large impact on the difference in free-energy, ΔG, between the hybridized and solvent-facing states. (Eq. 3.6) We display our results in Table 3.2. For the pair AC this secondary minimum is so prominent that its free-energy is lower than for the hybridized state. We cluster our metadynamics simulations from this minimum and generate a representative image. (Fig. 3.12) Both bases flip out of the DNA polymer and nestle into the minor groove, pointing toward the 5′ termini of the DNA. Because both bases are solvent-facing in this motif, base flip studies biasing only one base could not observe this structure.

This metastable arrangement matches the e-motif proposed by Gao et. al. [119] in their NMR study of mismatched CC pairs. Previous works have investigated the e-motif for mismatch CC pairs in strands with the sequence CCG, typically in triplet repeats i. e.$(CCG)_N$. [119–127] Several of these works identified intrastrand cross-linking sta-

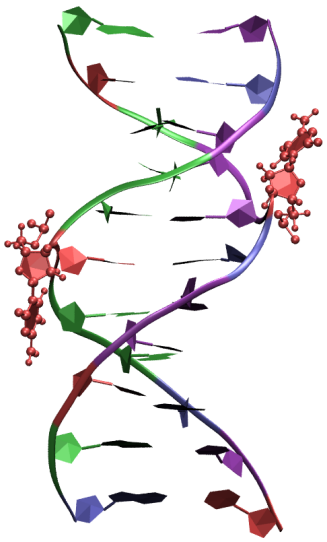


Figure 3.12: Typical configuration of the secondary minimum (e-motif) for the CC pair as demonstrated in Figure 3.8. Both bases of interest are no longer stacked within the DNA polymer, but are flipped to be solvent-facing. The bases lie alongside the DNA, within the minor groove, and pointing toward the 5′ termini.

bilizing the motif, [123–125, 127] suggesting our restraint of neighboring base pairs increases the likelihood of observing this motif. In their MD simulation study, Zhang et. al. [122] demonstrate the e-motif is better stacked, more compact, and allows for better packing than a structure not presenting the motif. We found the e-motif to be stable; a 100 ns unbiased simulation with the e-motif as an initial condition remained in the motif for the duration. Edfeldt et. al. [128] demonstrated intrinsic sequence-dependence of the e-motif; we speculate our DNA sequence affects which pairs express or don't express the motif. CC mismatches are some of the most difficult to be recognized by MMR proteins, [40, 41] but are the most likely to exhibit the e-motif; [121, 122, 124] we hypothesize there is an inversely proportional causal relationship between presence of the e-motif and MMR recognition.

4

DEVIATION: EXPLORING AZOBENZENE'S EFFECTS ON DNA

4.1 INTRODUCTION

Understanding the photoisomerization of azobenzene when it is incorporated in DNA is essential to design and improve the efficiency of optically reprogrammable biosystems and nanosensors. A useful descriptor of this process is the *trans*-to-*cis* photoisomerization quantum yield (QY), which measures the fraction of *trans*-azobenzene molecules that isomerize to the *cis* form upon absorption of a UV photon.

Yan et. al. have shown QY is sensitive to the local environment of the photoswitch. [65] The QY of azobenzene incorporated into single-stranded DNA (ssDNA) is lower than that of free azobenzene in solution, and is further reduced upon hybridization with another strand. In addition, the identity of the surrounding bases in double-stranded DNA (dsDNA) also affects the QY: it is lowest when all bases form matched Watson-Crick pairs, it increases in the presence of base pair mismatches, and it is highest if there is an abasic site immediately adjacent to the azobenzene. In the case of mismatched bases both identity and position of the mismatch influence the QY.

One possible explanation for the observed variation in QY is a correlated variation in the amount of local free space around the azobenzene: one would expect the free volume to be lower in ssDNA than in free solution, and lower still in dsDNA. [65] This hypothesis also rationalizes the experimental data on various dsDNA sequences: base pair mismatches introduce local distortions in the helical structure of DNA, which create larger free volumes than those present in matched dsDNA. The largest increases in QY are caused by mismatches immediately adjacent to the azobenzene site, whereas it is less sensitive to mismatches farther away from the azobenzene.

In the present work we use MD computer simulations to obtain a detailed understanding of the effects of DNA sequence on the local environment of embedded azobenzene photoswitches. To the extent that a classical description is appropriate, such calculations can provide microscopic insight into complex and thermally fluctuating systems at the atomistic level. We consider azobenzene-modified DNA (azo-DNA) with a variety of sequences, and compare various microscopic observables extracted from these simulations against the experimental QY data. We find that one such observable, the variance of the occupation number within a small volume enclosing the outer phenyl ring

of the azobenzene group, is strongly correlated with the measured photoisomerization QY. We show that this correlation can be used to predict QYs for additional DNA sequences, and verify the validity of these predictions in subsequent spectroscopic experiments.

4.2 METHODS

The methods are exactly as specified in Chapter 2, unless as outlined below. All sequences are given in Table 4.1.

4.2.1 *Force Field Parameterization, CMAP, and Model Building*

Computer simulations were performed using the CHARMM27 molecular mechanics force field. [83] This parameter set was used to describe all atoms with the exception of the central nitrogens forming the azo double bond. For these nitrogens, nonbonded interaction parameters were taken to be the same as standard CHARMM27 nitrogens. Bonded interactions were taken to be the same as those of the CHARMM27 aromatic atomtype CN3, with the exception of the CCNN and CNNC dihedral angles.

McCullagh et. al. [74] calculated the torsional constants for CNNC and CCNN to be 58.6 kJ/mol and 18.51 kJ/mol respectively. The CNNC dihedral has minima at 0 (*cis*-) and 180 (*trans*-) degrees. The CCNN dihedral has minima at 0 and 180 degrees (*trans*) and -50 and 130 degrees (*cis*). To accomodate the different minima positions for *cis*- and *trans*- geometries, McCullagh et. al. created two different models for *cis*- and *trans*-azobenzene. We have unified the two geometries into one model by employing a custom CMAP [129] correction.

We start with the base equation centering all minima at 0 and 180 degrees:

$$f(x,y) = 58.6(1+\cos(2x-180)) + 18.51(1+cos(2y-180)) \quad (4.1)$$

where x and y are shorthands used to represent the CNNC and CCNN dihedrals, respectively. This potential-energy surface is plotted in Figure 4.1a.

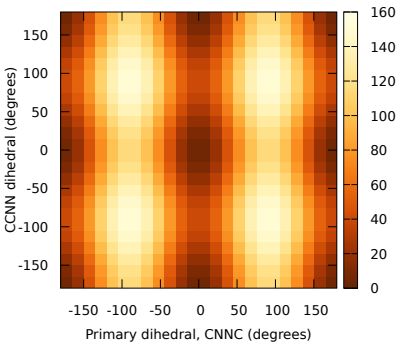
We then shift the minima of the y-values up by 50 degrees when x is close to 0 degrees. We choose a cosine activator function to achieve this shift:

$$\begin{aligned} g(x,y) =& 58.6(1+\cos(2x-180)) \\ &+ 18.51(1+cos(2y-180+50(1+cos(x)))). \end{aligned} \quad (4.2)$$

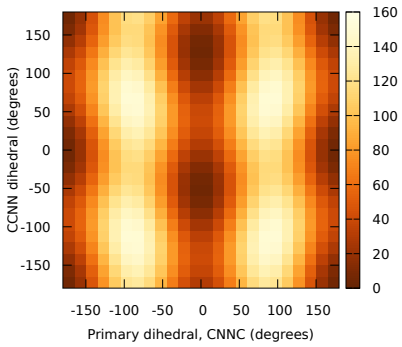
The updated potential-energy surface is depicted in Figure 4.1b. Finally, the CMAP correction is the change between the two surfaces:

$$h(x,y) = g(x,y) - f(x,y) \tag{4.3}$$

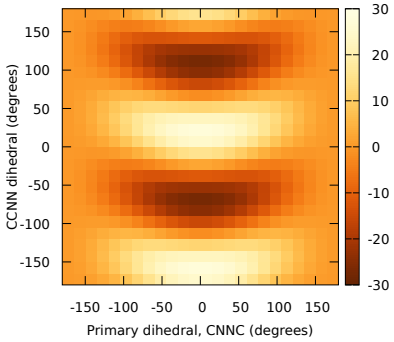
as shown in Figure 4.1c.



(a) initial



(b) final

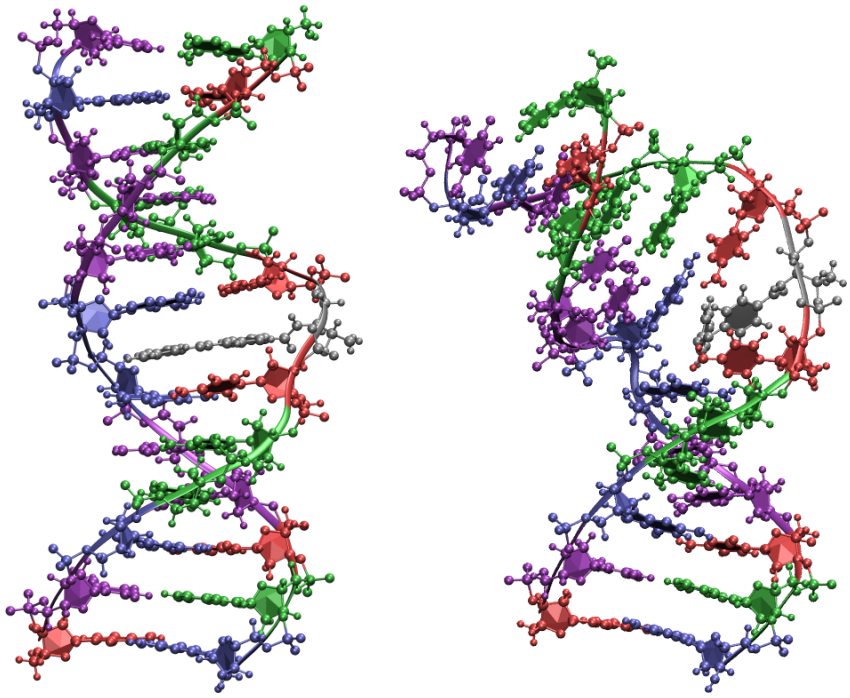


(c) difference

Figure 4.1: Potential-energy functions for the process of CMAP correction for the coupled CNNC and CCNN dihedrals of azobenzene. The initial, symmetrical surface is shown in sub-figure (a). The final, desired surface (b) shifts the location of the y-minima down 50 degrees when x is 0 degrees. The applied CMAP potential (c) is the difference between these two surfaces.

The final result is a single force field which can accurately accomodate azobenzene in both the *cis*- and *trans*-forms. A depiction of azo-DNA in both isomers is shown in Figure 4.2. In the *trans*-form, azobenzene *pi*-stacks with the neighboring bases. There is some backbone distortion in the opposite strand, but the overall integrity of the polymer remains. In the *cis*-isomer, azobenzene disrupts the *pi*-stacking and greatly distorts the DNA polymer. The chosen image is a

depiction of the molecule roughly 100 ns after isomerizing. On longer, experimental timescales, the polymer melts into separate strands.



(a) *trans-* (b) *cis-*

Figure 4.2: Simulation snapshots of azobenzene-modified DNA. The *trans-*isomer (a) stacks with the neighboring DNA bases and preserves the symmetry of the polymer. The *cis-*isomer (b) distorts the polymer, eventually causing it to melt. (Not shown.) Colors: nucleotides A, T, C, and G are shown in green, violet, red, and dark blue, respectively; azobenzene is shown in grey.

Partial charges for the azobenzene and linker atoms were generated using the HF/6-31G* electrostatic potential and fitted using the restrained electrostatic potential module in Amber. [130, 131] We created DNA strands using 3DNA [81] and manually inserted azobenzene using VMD. [82]

4.2.2 *Occupation Number Calculation*

The number of atoms in the immediate vicinity of the azobenzene group was calculated for each simulation frame using the counting function

$$N = \sum_{i \in A} \sum_{j \in B} \frac{1 - \left[(r_{ij} - d_0)/r_0 \right]^6}{1 - \left[(r_{ij} - d_0)/r_0 \right]^{12}} \tag{4.4}$$

as implemented in the PLUMED 1.3.0 plugin. [132] Here A is the set of atoms in the outer phenyl ring of azobenzene, B is the set of re-

SEQUENCE NAME	SEQUENCE	QUANTUM YIELD
ssDNA	5'-AGACTGAACXCAATGTATG-3'	0.022(1)
matched	5'-AGACTGAACXCAATGTATG-3' 3'-TCTGACTTG GTTACATAC-5'	0.0033(2)
mm1A	5'-AGACTGAACXCAATGTATG-3' 3'-TCTGACTTG **A**TTACATAC-5'	0.0046(5)
mm1T	5'-AGACTGAACXCAATGTATG-3' 3'-TCTGACTTG **T**TTACATAC-5'	0.007(1)
mm1C	5'-AGACTGAACXCAATGTATG-3' 3'-TCTGACTTG **C**TTACATAC-5'	0.0069(9)
mm-abasic	5'-AGACTGAACXCAATGTATG-3' 3'-TCTGACTTG **0**TTACATAC-5'	0.012(1)
PM-abasic	5'-AGACTGAACXCAATGTATG-3' 3'-TCTGACTTG0GTTACATAC-5'	0.0027(5)
mm-double-abasic	5'-AGACTGAACXCAATGTATG-3' 3'-TCTGACTT**0** **0**TGACATAC-5'	0.029(1)
ss-polyT	5'-TTTTTTTTTXTTTTTTTTT-3'	0.044(1)

Table 4.1: Computer simulations were performed on the central DNA segment highlighted in gray. Azobenzene is incorporated at the X-position; abasic sites are labeled as 0; mismatched sites are indicated in **bold**.

maining atoms in the DNA, and r_{ij} is the distance between atom i and atom j. d_0 and r_0 are length parameters chosen to be 2.5Å and 0.5Å, respectively. With these choices, the summand is an indicator function that is approximately equal to one if the distance falls within 2 to 3Å. These parameters were optimized to maximize the correlation between experimentally determined QY and the variance σ^2 of the occupation number (see below).

Statistics of the number of close contacts between the outer phenyl ring and the remaining DNA atoms were obtained by calculating the probability distribution $P(N)$, from which the free-energy

$$F(N) = -k_B T \log\left(P(N)\right) \tag{4.5}$$

was calculated. Here, k_B is the Boltzmann constant and T is the temperature of the simulation. The variance of the occupation number was calculated as

$$\sigma^2 = \langle N^2 \rangle - \langle N \rangle^2 \tag{4.6}$$

where the angular brackets denote the time average over the course of the simulation. Correlations between experimental and simulated data were quantified by a linear least-squares fit.

4.3 RESULTS AND DISCUSSION

4.3.1 *Structure*

Our simulations show that DNA modified with *trans*-azobenzene maintains stable duplexes with strands that are either perfectly matched or that contain a single mismatched or abasic site. These duplexes show varying degrees of structural deformations relative to canonical, azobenzene-free dsDNA. The average DNA structure shows relatively small deformations, a result that matches previous studies. [48, 133] The DNA near the azobenzene insertion site shows a structure that diverges from canonical DNA, especially in base stepping parameters (Figure 4.3). This in turn leads to a slight bend of the double helix (Figure 4.4a). The azobenzene is not perfectly stacked within the double helix and is twisted slightly toward the major groove (Figure 4.5), a result matching previous work. [48, 133–135]

The photoisomerization QY of azobenzene has been shown to depend strongly on the sequence of the DNA in which it is embedded. A possible explanation for this behavior are structural changes of the DNA in the vicinity of the azobenzene that affect its ability to isomerize, for example by steric hindrance. [65] Our simulations show that

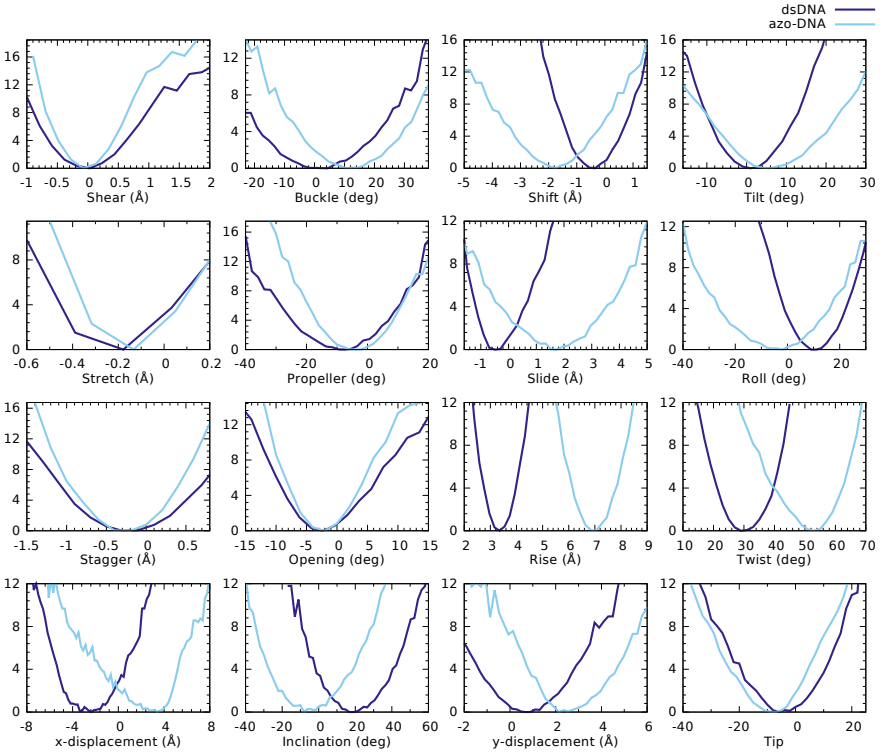
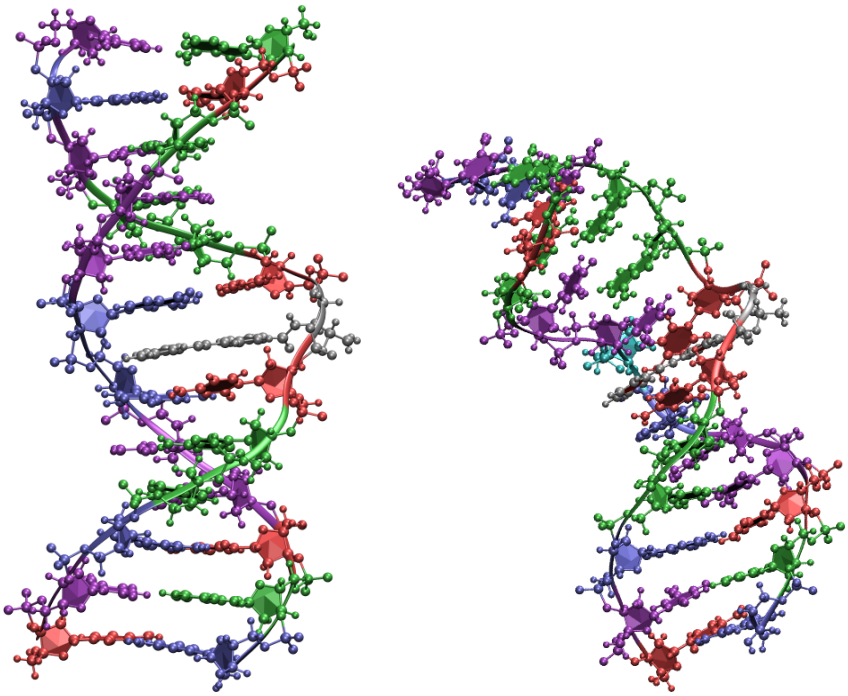


Figure 4.3: FESs of structural order parameters. Shown for comparison are non-modified dsDNA and azo-DNA, both fully matched. The bases considered for this analysis were those four bordering the azobenzene moiety. For the parameters shear, buckle, stretch, propeller, stagger, and opening, not much difference is seen between non-modified and azo-DNA. Interestingly, though the average values of the parameters do not change much, the variance for the parameters is much lower (the free-energy surface is narrower) for the azo-DNA as for the dsDNA. This would indicate that an azobenzene moiety inserted into DNA has the effect of stiffening the polymer. As expected, the step parameters which "step over" azobenzene are substantially different between azo-DNA and dsDNA. This is especially notable in the free-energy profile of rise, whose average value of azo-DNA is almost twice that of dsDNA.



(a) *matched* (b) *mm-ab*

Figure 4.4: Simulation snapshots of *trans*-azobenzene incorporated in ds-DNA. For complementary DNA strands only minor defor mations relative to canonical dsDNA are observed, such as backbone puckering at the azobenzene insertion site (a). Mismatched sequences exhibit larger deformations, in particular sequence mm-abasic (b). Here the azobenzene twists out of the DNA chain, and the AT pair adjacent to the abasic site *pi*-stacks with itself rather than forming hydrogen bonds. Colors: nucleotides A, T, C, and G are shown in green, violet, red, and dark blue, respectively, azobenzene is shown in grey, and the abasic site is shown in turquoise.

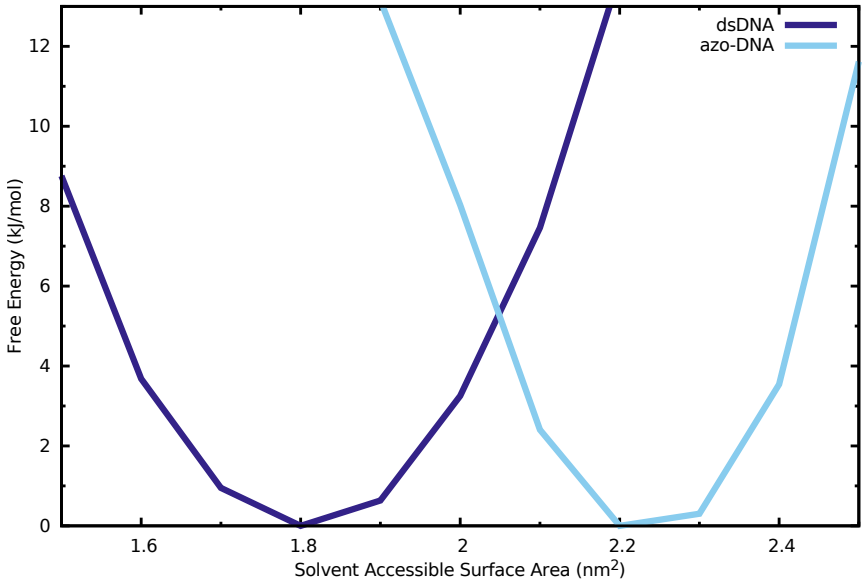


Figure 4.5: FES comparison of dsDNA and azo-DNA with regard to SASA of the bases neighboring the azobenzene moiety. Only the heavy atoms of the six-membered rings were considered. A slight elevation of SASA between azo-DNA and dsDNA indicates that the center of the DNA helix is slightly more exposed to the solvent when azobenzene is present, indicating imperfect stacking between azobenzene and its neighboring bases. The *trans*-azobenzene is slightly oriented toward the major groove of the DNA, though it still remains between and pi-stacks with its neighboring bases

strands containing mismatches and/or abasic sites indeed exhibit larger deformations than the perfectly matched duplex. These deformations are most prominent if the complementary strand contains abasic sites (sequence mm-abasic, Figure 4.4b). In this case the azobenzene twists out of the double helix, disrupting *pi*-stacking interactions with its neighbors. This mutation also affects base pairs that are not adjacent to the azobenzene: the AT pair adjacent to the abasic site melts and *pi*-stacks with itself. Interestingly, this sequence also exhibited the largest QY of all duplexes studied in Ref. [65]. Throughout our simulations the azobenzene remained in its *trans* groundstate due to the high activation barrier of isomerization.

4.3.2 *Energies of Free Volume*

To quantify structural changes that might affect azobenzene isomerization we compute the probability distribution function $P(N)$, and the corresponding free-energy F(N), of the number of DNA atoms that occupy a small region surrounding the outer phenyl ring (see Methods). Intuitively, this number expresses approximately the number of atoms in sufficiently close contact with the azobenzene that they would cause steric clashes if the azobenzene were to undergo isomerization from the *trans* to the *cis-* conformation.

The free-energy profiles obtained from our simulations exhibit a global minimum, corresponding to a most likely occupation number that ranges from approximately 9 to 12 atoms depending on the DNA sequence (Figure 4.6). Single-stranded systems exhibit a seconday minimum at zero occupation number, reflecting the fact that ssDNA can adopt both extended and globular configurations. One might surmise that the average number of contacts is negatively correlated with the experimentally measured QY, since azobenzene isomerization is impeded by the presence of atoms in its immediate vicinity. Surprisingly this is not the case, as shown in the inset of Figure 4.7: the data show no significant correlation between the average occupation number and QY.

4.3.3 *The Role of Fluctuations*

While the average number of contacts is not correlated with QY, the magnitude of fluctuations of this occupation number is correlated, as shown in Figure 4.7. Using a subset of six systems for which QY data were initially available (sequences matched, mm1A, mm1C, mm1T, mm-abasic, ssDNA), we find a nearly linear relationship between the variance σ^2 and the QY, with a correlation coefficient $R^2 = 0.98$. To test whether one can use the resulting linear model to predict the QYs of novel systems, we computed the contact number variance for sequences PM-abasic, mm-double-abasic, and ss-polyT. The model

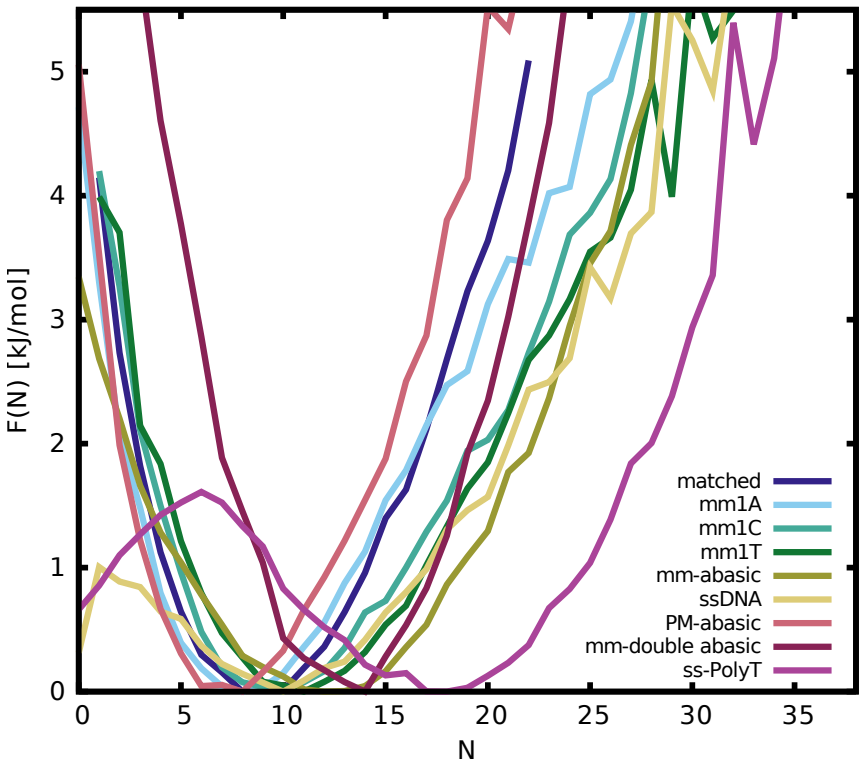


Figure 4.6: Free-energy as a function of the number of close azobenzene-DNA contacts as obtained from computer simulations. For each sequence there is a well-defined global minimum in the free-energy, which corresponds to the contact number with the highest probability. The two single-stranded systems exhibit a secondary minimum at zero contacts.

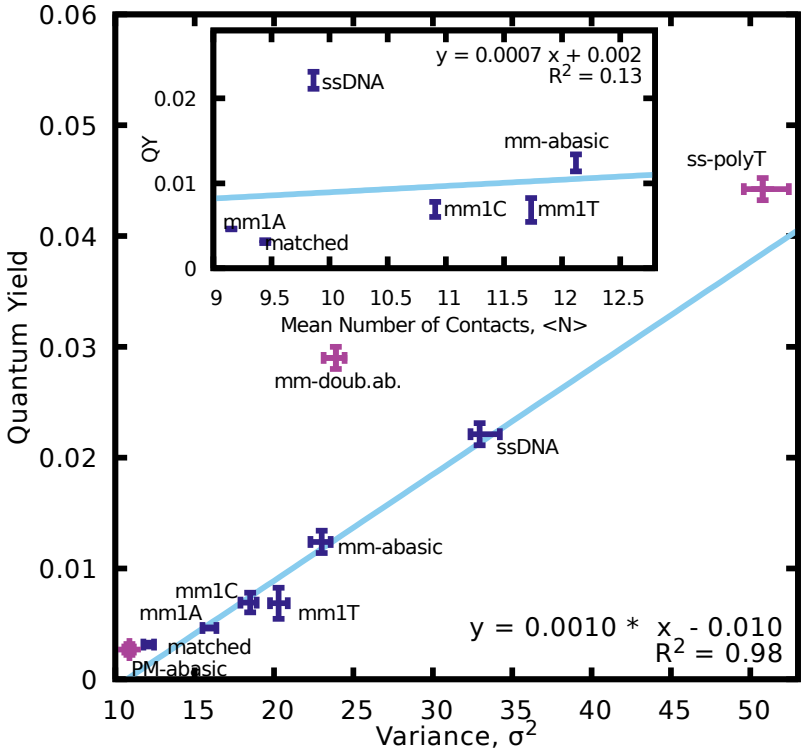


Figure 4.7: Linear regression reveals a significant correlation between the variance of the number of contacts, as obtained from computer simulations, and experimentally measured QY. Only the six sequences shown in blue were used in the regression analysis; QYs for sequences shown in purple were predicted based on the resulting model. No significant correlation is found between the average number of contacts and QY (inset).

predicts that the QY of PM-abasic is the lowest of all systems studied, that of mm-double-abasic falls in between mm-abasic and ssDNA, and that of ss-polyT is the highest. Spectroscopic measurements confirm these predictions qualitatively: PM-abasic and ss-polyT indeed have the lowest and highest QY, respectively. The predicted position of mm-double-abasic in this sequence was off by one, as its QY is in between that of ssDNA and ss-polyT.

That contact number variance can be used to predict the QY of azo-DNA at least qualitatively is surprising. Photoisomerization is a process that involves nuclear motion in an excited electronic state, [136, 137] and yet classical simulations of *trans*-azobenzene appear to capture the dominant contribution to the QY. The fact that it is the variance, rather than the average, of the azobenzene-DNA contact number that correlates with QY indicates that fluctuations in this quantity are particularly important. While neither photon absorption nor isomerization actually occur in our simulations, occupation number fluctuations in the *trans*-azobenzene ground-state are accurately sampled.

There are multiple physical explanations that could explain the observed correlation. The first is based on the notion that azobenzene isomerization can only occur if at the time of photon absorption there is a fluctuation that effectively empties out the region in the vicinity of the outer phenyl ring so that it is free to move without being hindered by steric clashes. The second considers the reversible work required to empty out this surrounding volume: If ϕ is a generalized force that acts on the occupation number N in this region, the equilibrium response to that force is related the contact number variance in linear response theory by

$$\frac{d\langle N\rangle}{d\phi} = \frac{-\sigma^2}{k_B T}. \tag{4.7}$$

In other words, the susceptibility of the average occupation number to a force acting to decrease it is directly proportional to the contact number variance. In this equilibrium picture the azobenzene can effectively use the energy from the absorbed photon to push other atoms out of the way to clear a path for isomerization. Despite the stark difference between those two pictures (instantaneous fluctuation vs. reversible work), our simulations cannot discriminate between them. To the contrary, they are intimately linked: a more frequent occurrence of fluctuations with low occupation numbers implies a larger susceptibility to an applied force.

5

DISPLACEMENT: STRETCHING DNA TO ITS BREAKING POINT

5.1 INTRODUCTION

Many domains require an understanding of the DNA unbinding pathway; biologically, it is essential for molecular recognition, storage and retrieval of biological information, cell replication, and DNA transcription. [138, 139] It is relevant in developing novel materials: the fields of programmable assembly and molecular sensing have both employed DNA hybridization. [140–143]

DNA was historically believed to melt by an "all-at-once" mechanism, [144] where every base in the polymer separates from its partner at about the same time. In their dynamic force spectroscopy (DFS) study, Ginger et. al. recently showed this mechanism might not always hold true. [71, 145] They revealed DNA may melt "piece-wise," or base-by-base when stressing the polymer in a specific fashion. When they pull the DNA in the shearing geometry, (Fig. 5.1a) their data are consistent with an "all-at-once" melting pattern. When they pull their DNA in the unzipping geometry, (Fig. 5.1b) their data are consistent with the "piece-wise" model.

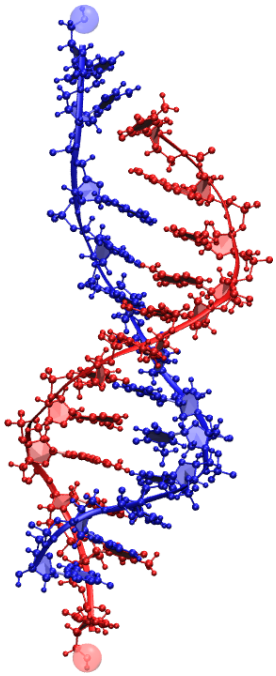
They identify several key differences between their shearing and unzipping geometries. They must apply a greater force to the shearing geometry to achieve rupture. The values for the dissociation rate of dsDNA, k_u^0, are 5-7 orders of magnitude lower for the shearing geometry than for the unzipping geometry. Most intriguingly, by employing state-of-the-art technologies for non-equilibrium pulling, their analysis predicts different values for ΔG in the two geometries. The different ΔGs which they calculate thus indicate an incorrect interpretation by the underlying model.

Free-energy *must* be the same for shearing and for unzipping. Because the initial (hybridized) and final (melted) states are the same for both mechanisms, the Second Law of Thermodynamics states ΔG should be the same for both geometries; the only mechanism by which a measured energy can be greater is if irreversible work has been performed on the system.

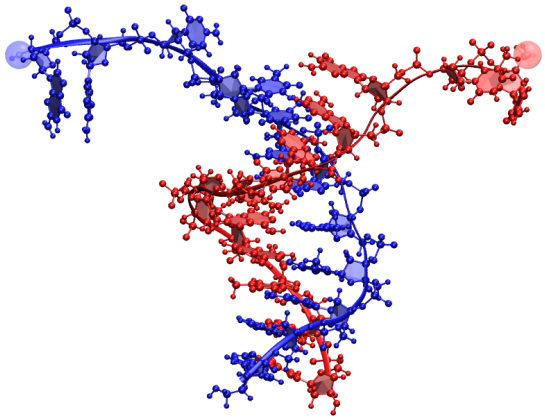
$$W_{\text{observed}} \geq \Delta G \tag{5.1}$$

or,

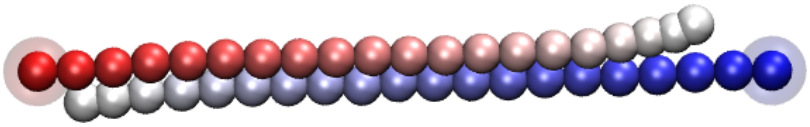
$$W_{\text{observed}} = \Delta G + W_{\text{irrev}} \tag{5.2}$$



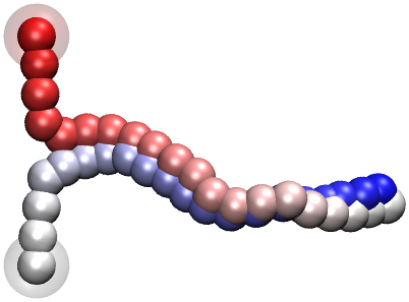
(a) shear



(b) unzip



(c) shear (coarse-grained)



(d) unzip (coarse-grained)

Figure 5.1: Simulation snapshots of DNA in the shearing (a) and unzipping (b) geometries. Strands are depicted in blue and red to differentiate them. The semi-transparent beads indicate the pull sites for the simulations. As depicted, shearing would pull vertically and unzipping would pull horizontally. Coarse-grained models of shearing (c) and unzipping (d) geometries. The semi-transparent beads indicate the pull sites for the simulations. As depicted, shearing would pull horizontally and unzipping would pull vertically.

where W_{observed} is the observed difference in energy and W_{irrev} is the irreversible work performed on the system.

In this chapter, we attempt to elucidate the results observed by Ginger et. al. [71, 145] by exploring two primary questions:

A. When pulled in different geometries, does DNA melt all-at-once or piece-wise?

B. Which theoretical fittings confer ΔG values which are most consistent with our intuitive physical understanding?

5.2 METHODS

5.2.1 *Force Field Creation*

We adopt the Weeks-Chandler-Anderson (WCA) potential [146] defined as:

$$V_{\text{repulsive}}(r) = \begin{cases} V_{\text{LJ}}(r) + \epsilon & \text{if } r < 2^{1/6}\sigma \\ 0 & \text{if } r \geq 2^{1/6}\sigma \end{cases} \tag{5.3}$$

$$V_{\text{attractive}}(r) = \begin{cases} -\epsilon & \text{if } r < 2^{1/6}\sigma \\ V_{\text{LJ}}(r) & \text{if } r \geq 2^{1/6}\sigma \end{cases} \tag{5.4}$$

to describe the potential-energy, V between two particles distance r apart. ϵ is the energy scale of the system, σ is the length scale, and V_{LJ} is the Lennard-Jones-potential described by:

$$V_{\text{LJ}} = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]. \tag{5.5}$$

We apply a coarse-grain polymer model adapted from Mishra et. al., [80] where we define two chains, each of twenty beads (Fig. 5.1c and 5.1d). Beads bond to each other along the chain with a harmonic potential. Every non-bonded pair experiences the repulsive component of the WCA potential (Eq. 5.3). Complementary beads on opposing strands (e. g.beads 1 and 1) experience the attractive portion of

the WCA potential (Eq. 5.4). Together, this creates the following system potential-energy:

$$\begin{aligned} V_{\text{tot}} = &\sum_{c=1}^{2}\sum_{j=1}^{19}\frac{1}{2}k(r_{j,j+1}^{c}-d_0)^2 \\ &+\sum_{c=1}^{2}\sum_{i=1}^{18}\sum_{j>i+1}^{20}V_{\text{repulsive}}(r_{i,j}^{c}) \\ &+\sum_{i=1}^{20}\sum_{j\geq i}^{20}V_{\text{repulsive}}(|\vec{r}_i^{\,1}-\vec{r}_j^{\,2}|) \\ &+\sum_{i=1}^{20}\sum_{j=1}^{20}V_{\text{attractive}}(|\vec{r}_i^{\,1}-\vec{r}_j^{\,2}|)\delta_{ij} \end{aligned} \tag{5.6}$$

where $r_{i,j}^{c}$ indicates the distance between beads i and j on the cth chain. We adopt $k = 100\epsilon/\sigma^2$ for our harmonic bonded spring constant, and $d_0 = 1\sigma$ for our equilibrium bonded distance. We implement our potential employing the Gromacs [86] simulation suite.

5.2.2 *Pulling Simulations*

We perform pulling simulations in two different geometries: shearing and unzipping. We attach a harmonic potential to each strand; the same terminus for unzipping and the opposite terminus for shearing. We shift the trap potential at a velocity $v \in \{0.01, 0.03, 0.05, 0.07, 0.1, 0.3, 0.5, 0.7, 1\}$ and with trap spring constant $k_c \in \{0.1, 1, 10, 100\}$. For each geometry, velocity, and spring constant, we performed 1000 trajectories. We define "rupture force" as the maximum force exerted on the system, excluding the first 25% of each trajectory. "Rupture" is then defined as the time at which the rupture force occurs.

5.2.3 *Calculating Work*

We calculate work employing the method derived by Park et. al. [147] They begin with the principal definition of work:

$$W_{0\to t} = \int_0^t \mathrm{d}t' \frac{\delta\lambda_{t'}}{\delta t'}\left[\frac{\delta\tilde{\mathcal{H}}}{\delta\lambda_{t'}}\right] \tag{5.7}$$

where $W_{0\to t}$ is the work performed from time 0 to time t, λ_t is the position of the trap, and $\tilde{\mathcal{H}}$ is the Hamiltonian characterized by:

$$\tilde{\mathcal{H}}(\vec{r},\vec{p},\lambda_t) = \mathcal{H}(\vec{r},\vec{p}) + u(\vec{r},\lambda_t) \tag{5.8}$$

where $\mathcal{H}(\vec{r}, \vec{p})$ is the unperturbed Hamiltonian and $u(\vec{r}, \lambda_t)$ is the guiding potential described by:

$$u(\vec{r}, \lambda_t) = \frac{k_c}{2}\left(r_{ij} - \lambda_t\right)^2 \tag{5.9}$$

which constrains the distance between points i and j, r_{ij}, to remain near λ_t with spring constant k_c. λ_t evolves in time at a constant rate by:

$$\lambda(t) = \lambda_0 + vt \tag{5.10}$$

where v is the constant pulling velocity and λ_0 is the trap position at time 0.

Given that $\frac{\delta\lambda}{\delta t} = v$, and that

$$\frac{\delta\tilde{\mathcal{H}}}{\delta\lambda_t} = \frac{\delta u}{\delta\lambda_t} = k_c\left(\lambda_t - r_{ij}\right), \tag{5.11}$$

we can define the work from Equation 5.7 as:

$$W_{0\to t} = \int_0^t \mathrm{d}t'\, v \cdot k_c\left(\lambda_{t'} - r_{ij}(t')\right) \tag{5.12}$$

which is Eq. 13 in Park et. al.'s work. The force for a harmonic potential can be described as

$$F(t) = k_c\left(\lambda_t - r_{ij}\right); \tag{5.13}$$

thus Equation 5.12 can simplify to

$$W_{0\to t} = v\int_0^t F(t')\,\mathrm{d}t'. \tag{5.14}$$

5.2.4 *Differing Models*

5.2.4.1 *Bell, Evans*

A model introduced by Bell [148] and refined by Evans et. al.[149] proposes the following relationship between loading rate, r, and mean rupture force, $\langle F\rangle$ (Eq. 2 in Evans' paper):

$$\frac{1}{r} = \sum_n \frac{x_t(n)}{k_u^0(n) k_B T}\exp\left[\frac{-\langle F\rangle x_t(n)}{k_B T}\right] \tag{5.15}$$

where n represents the n^{th} barrier in the FES, $x_t(n)$ is the distance from the n^{th} barrier with respect to the global minimum, $k_u^0(n)$ is the intrinsic unbinding rate of that barrier, k_B is the Boltzmann constant,

and T is the temperature. For a single barrier, the equation simplifies to:

$$\frac{1}{r} = \frac{x_t}{k_u^0 k_B T} \exp\left[\frac{-\langle F \rangle x_t}{k_B T}\right], \tag{5.16}$$

which can be re-arranged to produce

$$\langle F \rangle = \frac{k_B T}{x_t} \ln\left(\frac{r x_t}{k_B T k_u^0}\right). \tag{5.17}$$

In this model the mean rupture force increases as $\ln(r)$.

5.2.4.2 *Friddle, Noy, deYoreo*

Friddle, Noy, and deYoreo [150] developed a model to extract FES parameters from incomplete pulling data. It derives the distance from the bound state to the transition state, x_t, force required to maintain the bound and unbound states in equilibrium, F_{eq}, and the unperturbed unbinding rate constant, k_u^0.

They begin by adopting the unbinding rate supplied by Eq. 16 of the Bell model: [148]

$$k_u(F) = k_u^0 \exp\left[\frac{1}{k_B T}\left(F x_t - \frac{1}{2} k_c {x_t}^2\right)\right] \tag{5.18}$$

where $k_u(F)$ is the force-dependent rate constant for the unbinding reaction, k_B is the Boltzmann constant, T is the temperature, F is the force, and k_c is the spring constant of the pulling trap.

Similarly, the force-dependent binding rate, $k_b(F)$, can be defined as:

$$k_b(F) = k_b^0 \exp\left[\frac{-k_c}{2 k_B T}\left(\frac{F}{k_c} - x_t\right)^2\right]. \tag{5.19}$$

The difference in free-energy between the bound and unbound states, $\Delta G_{bu} = \Delta G_u - \Delta G_b$, can be calculated from the ratio of the intrinsic binding and unbinding rates by:

$$\exp\left[\frac{\Delta G_{bu}}{k_B T}\right] = \frac{k_b^0}{k_u^0}. \tag{5.20}$$

Rearranging Eq. 5.19 and applying the definition provided by Eq. 5.20 yields:

$$\begin{aligned} k_b(F) &= k_b^0 \frac{k_u^0}{k_u^0} \exp\left[\frac{-k_c}{2k_\mathrm{B}T}\left(\frac{F}{k_c} - x_t\right)^2\right] \\ &= k_u^0 \exp\left[\frac{\Delta G_{bu}}{k_\mathrm{B}T}\right] \exp\left[\frac{-k_c}{2k_\mathrm{B}T}\left(\frac{F}{k_c} - x_t\right)^2\right] \\ &= k_u^0 \exp\left[\frac{\Delta G_{bu}}{k_\mathrm{B}T} - \frac{k_c}{2k_\mathrm{B}T}\left(\frac{F^2}{{k_c}^2} - 2\frac{Fx_t}{k_c} + {x_t}^2\right)\right] \\ &= k_u^0 \exp\left[\frac{\Delta G_{bu}}{k_\mathrm{B}T} - \frac{1}{k_\mathrm{B}T}\left(\frac{F^2}{2k_c} - Fx_t + \frac{k_c {x_t}^2}{2}\right)\right] \\ &= k_u(F) \exp\left[\frac{1}{k_\mathrm{B}T}\left(\Delta G_{bu} - \frac{F^2}{2k_c}\right)\right] \end{aligned} \tag{5.21}$$

The equilibrium force F_eq occurs when the binding and unbinding rates are equal:

$$\begin{aligned} k_u(F_\mathrm{eq}) &= k_b(F_\mathrm{eq}) \\ &= k_u(F_\mathrm{eq}) \exp\left[\frac{1}{k_\mathrm{B}T}\left(\Delta G_{bu} - \frac{{F_\mathrm{eq}}^2}{2k_c}\right)\right]. \end{aligned} \tag{5.22}$$

Solving for ΔG yields:

$$\Delta G_{bu} = \frac{{F_\mathrm{eq}}^2}{2k_c} \tag{5.23}$$

which is Eq. 4 in Friddle, Noy, and deYoreo's paper.

As the rebinding rate dominates the dynamics for forces less than F_eq and it decays at forces larger than F_eq, Friddle et. al. simplify their master equation by starting bond loading processes at $F = F_\mathrm{eq}$ instead of at $F = 0$. With this approximation, they define their Eq. 5 as:

$$\int_1^{p_\mathrm{b}} \frac{\mathrm{d}p'_\mathrm{b}}{p'_\mathrm{b}} \approx \frac{-1}{r} \int_{F_\mathrm{eq}}^{F} k_u(F')dF' \tag{5.24}$$

where p_b is the probability of being bound at a specific force and $r = \frac{\mathrm{d}F}{\mathrm{d}t}$ is the loading rate (not to be confused with v, the pulling

rate), assumed to be a constant in the model. This equation simplifies to:

$$\begin{aligned}
\ln(p_b')\Big|_1^{p_b} &= \frac{-1}{r}\int_{F_{eq}}^{F} k_u^0 \exp\left[\frac{1}{k_B T}\left(F' x_t - \frac{1}{2}k_c {x_t}^2\right)\right] dF' \\
\ln(p_b) &= \frac{-k_u^0}{r}\exp\left[\frac{-k_c {x_t}^2}{2k_B T}\right]\int_{F_{eq}}^{F}\exp\left[\frac{F' x_t}{k_B T}\right] \mathrm{d}F' \\
&= \frac{-k_u^0}{r}\exp\left[\frac{-k_c {x_t}^2}{2k_B T}\right]\cdot\frac{k_B T}{x_t}\exp\left[\frac{F' x_t}{k_B T}\right]\Big|_{F_{eq}}^{F} \\
&= \frac{-k_B T}{r x_t} k_u^0 \exp\left[\frac{1}{k_B T}\left(F' x_t - \frac{1}{2}k_c {x_t}^2\right)\right]\Big|_{F_{eq}}^{F} \\
&= \frac{-k_B T}{r x_t}\big(k_u(F) - k_u(F_{eq})\big) \\
&= \frac{k_B T}{r x_t}\big(k_u(F_{eq}) - k_u(F)\big),
\end{aligned} \tag{5.25}$$

therefore:

$$p_b(F) = \exp\left[\frac{k_B T}{r x_t}\big(k_u(F_{eq}) - k_u(F)\big)\right]. \tag{5.26}$$

If we define the probability of rupture at a certain force as:

$$p_r(F) = \frac{-\mathrm{d}p_b(F)}{\mathrm{d}F} \tag{5.27}$$

and utilize the definition of mean rupture force:

$$\langle F\rangle = \int_{F_{eq}}^{\infty} F p_r(F)\mathrm{d}F = -\int_{F_{eq}}^{\infty} F\frac{\mathrm{d}p_b(F)}{\mathrm{d}F}\mathrm{d}F, \tag{5.28}$$

we can apply integration by parts to obtain:

$$\langle F\rangle = -F p_b(F)\Big|_{F_{eq}}^{\infty} + \int_{F_{eq}}^{\infty} p_b(F)\mathrm{d}F. \tag{5.29}$$

Assuming $p_b(F)$ converges to 0 as $F \to \infty$ and knowing $p_b(F_{eq}) = 1$, Eqs. 5.26 and 5.29 combine to produce:

$$\begin{aligned}
\langle F\rangle &= F_{eq} + \int_{F_{eq}}^{\infty}\exp\left[\frac{k_B T}{r x_t}\big(k_u(F_{eq}) - k_u(F)\big)\right]\mathrm{d}F \\
&= F_{eq} + \exp\left[\frac{k_B T}{r x_t}k_u(F_{eq})\right]\int_{F_{eq}}^{\infty}\exp\left[\frac{-k_B T}{r x_t}k_u(F)\right]\mathrm{d}F.
\end{aligned} \tag{5.30}$$

Integrating yields:

$$\langle F\rangle = F_{eq} + \exp\left[\frac{k_B T}{r x_t}k_u(F_{eq})\right]\cdot\frac{k_B T}{x_t}E_i\left(\frac{-k_B T}{r x_t}k_u(F)\right)\Big|_{F_{eq}}^{\infty} \tag{5.31}$$

where E_i is the exponential integral. Given $E_i(\infty) = 0$,

$$\langle F \rangle = F_{\text{eq}} + \exp\left[\frac{k_{\text{B}}T}{rx_t}k_u(F_{\text{eq}})\right] \cdot \frac{-k_{\text{B}}T}{x_t}E_i\left(\frac{-k_{\text{B}}T}{rx_t}k_u(F_{\text{eq}})\right). \quad (5.32)$$

Applying definition $E_1(y) = -E_i(-y)$, we come upon:

$$\langle F \rangle = F_{\text{eq}} + \exp\left[\frac{k_{\text{B}}T}{rx_t}k_u(F_{\text{eq}})\right] \cdot \frac{k_{\text{B}}T}{x_t}E_1\left(\frac{k_{\text{B}}T}{rx_t}k_u(F_{\text{eq}})\right). \quad (5.33)$$

We define $R(F_{\text{eq}}) = \frac{rx_t}{k_{\text{B}}Tk_u(F_{\text{eq}})}$ which is Eq. 6 in the Friddle et. al. paper, and arrive at the final form of the model:

$$\langle F \rangle = F_{\text{eq}} + \frac{k_{\text{B}}T}{x_t}\exp\left[\frac{1}{R(F_{\text{eq}})}\right]E_1\left(\frac{1}{R(F_{\text{eq}})}\right) \quad (5.34)$$

where E_1 is the exponential integral defined by $E_1(y) = \int_y^\infty \frac{e^{-x}}{x}\,\mathrm{d}x$.

5.2.4.3 *Hummer, Szabo*

Hummer and Szabo [151] base their method for extracting FESs from Jarzynski's identity: [152]

$$\exp\left[-\frac{1}{k_{\text{B}}T}\Delta G(t)\right] \equiv \left\langle \exp\left[-\frac{1}{k_{\text{B}}T}W_t\right]\right\rangle \quad (5.35)$$

where the angular brackets, $\langle\rangle$, denote the ensemble average, k_{B} is the Boltzmann constant, and T is the system temperature. This identity allows for us to transform from work, W_t, to free-energy, ΔG.

Equation 5.35 can re-arrange to give:

$$G(x) \equiv -k_{\text{B}}T\ln\left\langle \delta(x - x_t)\exp\left[-\frac{1}{k_{\text{B}}T}\Delta W_t\right]\right\rangle \quad (5.36)$$

which is Eq. 7 in Hummer and Szabo's work [151] where $G(x)$ is the unperturbed free-energy at position x.

Implementation is similar to umbrella sampling. [153] For each step in time, t, we create a histogram of positions weighted by their respective work profiles, W_t. In this way, we create numerous ensemble averages across our pulling coordinate. Hummer and Szabo adapt the weighted-histogram analysis method (WHAM) [154] to reconstruct the free-energy surface (Eq. 8 in their paper):

$$G(x) = -k_{\text{B}}T\ln\left(\frac{\sum_t \frac{\langle \delta(x - x_t)\exp[-W_t/k_{\text{B}}T]\rangle}{\langle \exp[-W_t/k_{\text{B}}T]\rangle}}{\sum_t \frac{\exp[-u(x,\lambda_t)/k_{\text{B}}T]}{\langle \exp[-W_t/k_{\text{B}}T]\rangle}}\right) \quad (5.37)$$

where $u(x, \lambda_t)$ is the time-dependent biasing potential defined by Eq. 5.9.

We can be represent this equation in a discrete form:

$$G\left(\left[l - \frac{1}{2}\right]\Delta x\right) = -k_B T \ln\left(\frac{\sum_{i=0}^{N} h_i(l)/\eta_i}{\sum_{i=0}^{N} \exp[-u_{il}/k_B T]/\eta_i}\right) \tag{5.38}$$

which is Eq. 10 in the Hummer/Szabo paper and where l represents the lth interval of size Δx, i is each discrete time step to N time steps, $h_i(l)$ is a weighted-histogram defined by:

$$h_i(l) = \frac{1}{K}\sum_{k=1}^{K} \exp\left[-W_{ik}/k_B T\right]\theta_l(x_{ik}) \tag{5.39}$$

which is Eq. 9 in Hummer and Szabo's work and which sums over all K (in this dissertation, 1000) trajectories and where $\theta_l(x)$ tests whether x is in the l^{th} interval:

$$\theta_l(x) = \begin{cases} 1 & \text{if } x \in \left[(l-1)\Delta x,\, l\Delta x\right) \\ 0 & \text{else} \end{cases}, \tag{5.40}$$

and η_i is calculated by:

$$\left\langle \exp[-W_t/k_B T]\right\rangle \approx \eta_i \equiv \frac{1}{K}\sum_{k=1}^{K} \exp\left[-W_{ik}/k_B T\right]. \tag{5.41}$$

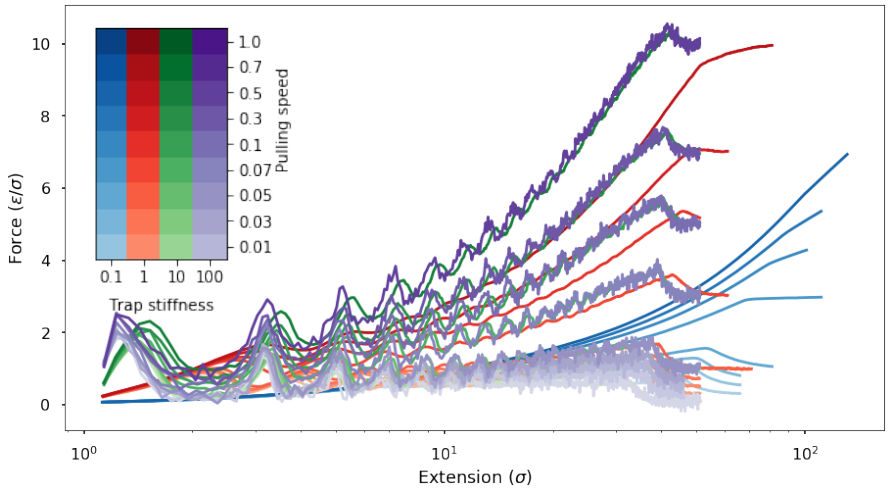
5.3 RESULTS AND DISCUSSION

Measuring the data thousands of times creates a distribution which can be statistically analyzed.

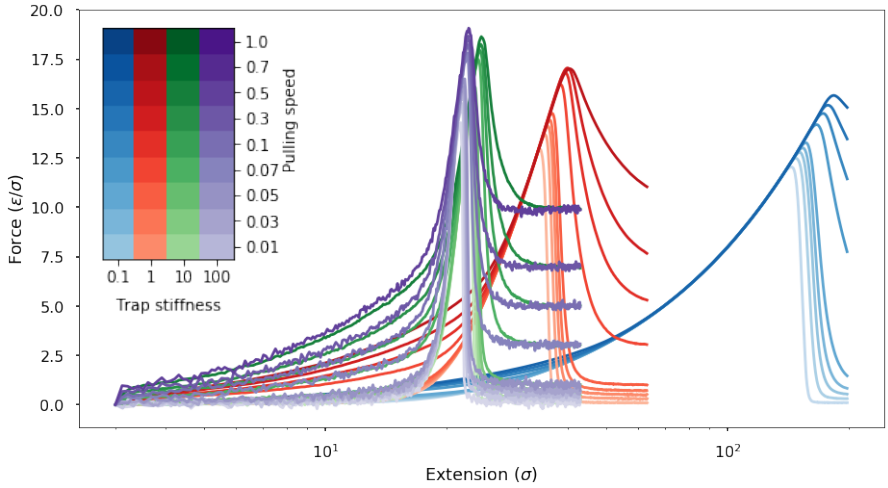
5.3.1 *Force-Extension Curves*

In experiment, loading rate and rupture force are calculated from force-extension curves. [71] The force as measured by atomic force microscopy (AFM) is plotted against the distance between the AFM tip and the substrate, the extension. Similarly, we plot the mean force vs. λ_t, the "extension" distance of our biasing potential, in Figure 5.2.

Several trends are apparent. For both shearing and unzipping geometries, as the pulling speed increases, so does the force. This matches experimental results. [71] There also exists an obvious trend regarding trap strength; as the spring constant k_c decreases, the extension required to cause a rupture event increases. Finally, the force at large extensions is consistently greater than the initial force. This is due to the Langevin dynamics of the system: there is an energy cost associated with pulling the polymers through



(a) unzip



(b) shear

Figure 5.2: Mean force vs. extension for both unzipping (a) and shearing (b) geometries. Four different spring constants are depicted: k_c=0.1 (blue), k_c=1 (red), k_c=10 (green), and k_c=100 (purple). The curves are pale for slow pulling speeds and dark for fast pulling speeds. Unzipping exhibits clear piece-wise melting, with forces increasing and decreasing as each base pair stretches and then breaks, respectively. Shearing manifests a single rupture event. As with experiment, faster pulling speeds lead to increased forces. Smaller spring constants require pulling the system farther to observe a rupture event.

the solution. The friction of this process manifests in these profiles through non-zero forces at large extensions.

There is a marked difference in the force-extension profiles between the unzipping and shearing geometries. Shearing geometries exhibit a smoothly-increasing profile with a single maximum which rapidly decays. This is consistent with an all-at-once mechanism, as hypothesized. Unzipping geometries display increased variety in profiles. There are copious local maxima along the curve, in some instances (for slow pulling rates) exceeding the rupture force. Each of these maxima correspond to individual base pair rupture events, consistent with the hypothesized piece-wise mechanism.

5.3.2 *The Relationship Between Pulling Rate and Rupture Force*

We plot in Figure 5.3 the complex relationships between work, rupture force, trap stiffness, and pulling speed. As Ginger et. al. demonstrated in their experiments, [71, 145] shearing geometries require larger average forces to rupture. Our results replicate this finding. We illustrate the work performed on the system is on average greater for the shear geometries than for unzipping. This matches their findings that the measured free-energies are larger for shearing than for unzipping.

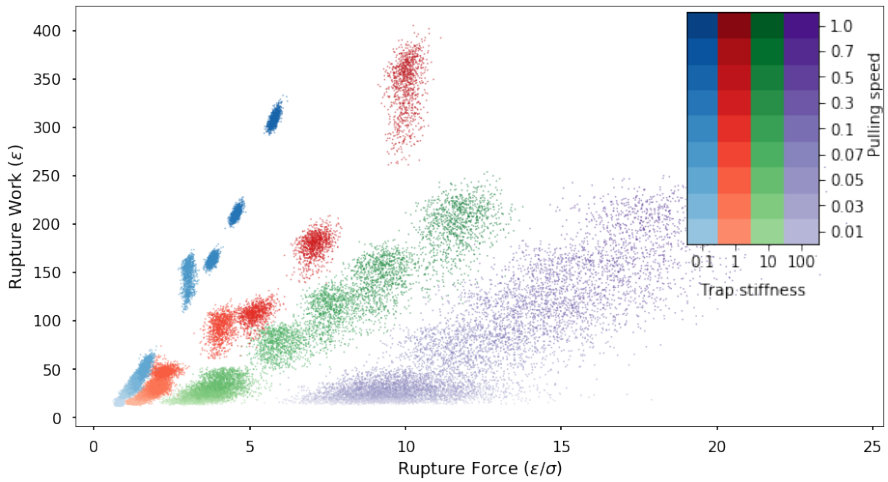
Rupture forces increase with pulling rate as thermal fluctuations permit less time to interact with the system and "bump" it over the edge into melting. Thermal fluctuations explain why simulations can express negative rupture forces: if a thermal fluctuation melts the system, the trap must exert less force overall. Statistically, some of these forces would be negative.

The spread of rupture forces increases as the trap stiffness. While a spring constant $k_c = 0.1$ affects a tight distribution for both unzipping and shearing geometries, a spring constant $k_c = 100$ affects a wide distribution. We can explain why this may be by approximating the system with a quadratic potential with spring constant k_s. We add a quadratic pulling potential with spring constant k_c. Together, they produce a quadratic potential with spring constant $k_{\text{tot}} = k_s + k_c$. In equilibrium, the variance of the position, $\langle \Delta x^2 \rangle$, is T/k_{tot}. The biasing force goes as $k_c x$, so its variance should go as ${k_c}^2 T/k_{\text{tot}} = {k_c}^2 T/(k_s + k_c)$. Therefore, variance in rupture force should increase with the stiffness of the trap, which we observe.

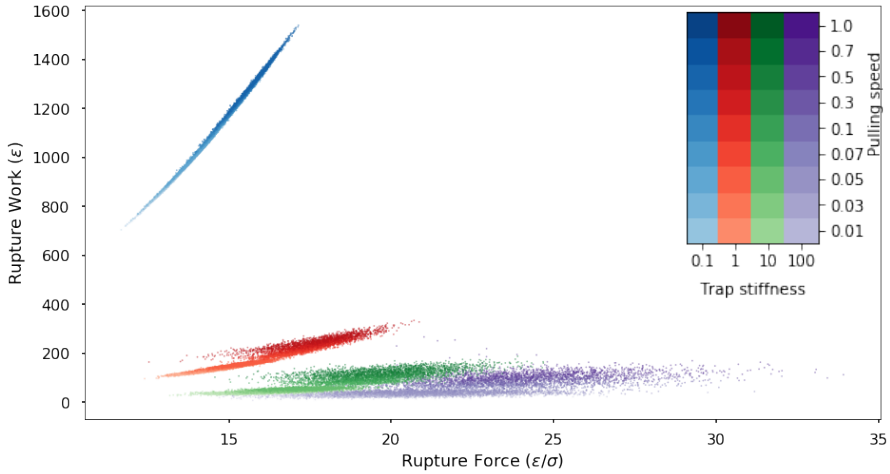
5.3.3 *Comparison of Models*

5.3.3.1 *Bell, Evans*

We extracted the loading rate $r = \frac{\mathrm{d}F}{\mathrm{d}t}$ for each set of k_cs and vs by fitting the linear portion of the mean-force profile and extracting



(a) unzip



(b) shear

Figure 5.3: Rupture work vs. rupture force for both unzipping and shearing geometries at all trap stiffnesses and pulling speeds studied. Higher pulling speeds (darker) exhibit both greater rupture forces and rupture work. Higher trap stiffnesses (e. g.purple) exhibit less rupture work, larger rupture force, and a greater variance in their distributions.

the slope. We fit our average rupture force, loading rates, and spring constants to Eq. 5.17 employing the `scipy` library routine `optimize.curvefit`. [155] Results are exhibited in Table 5.1 and Figure 5.4.

	$k_c(\epsilon/\sigma^2)$	$x_t(\sigma)$	$k_u^0(\epsilon/\sigma^2)$
unzip	0.1	1.55×10^{-1}	6.60×10^{-5}
	1	1.19×10^{-1}	5.58×10^{-5}
	10	1.06×10^{-1}	7.47×10^{-5}
	100	9.56×10^{-2}	8.26×10^{-5}
shear	0.1	1.24×10^{-1}	6.26×10^{-10}
	1	9.79×10^{-2}	3.52×10^{-8}
	10	8.68×10^{-2}	2.07×10^{-7}
	100	8.33×10^{-2}	3.42×10^{-7}

Table 5.1: Fit parameters to our simulation data adopting the Bell-Evans model. x_t is the distance between the primary minimum and transition state, and k_u^0 is the intrinsic unbinding rate constant.

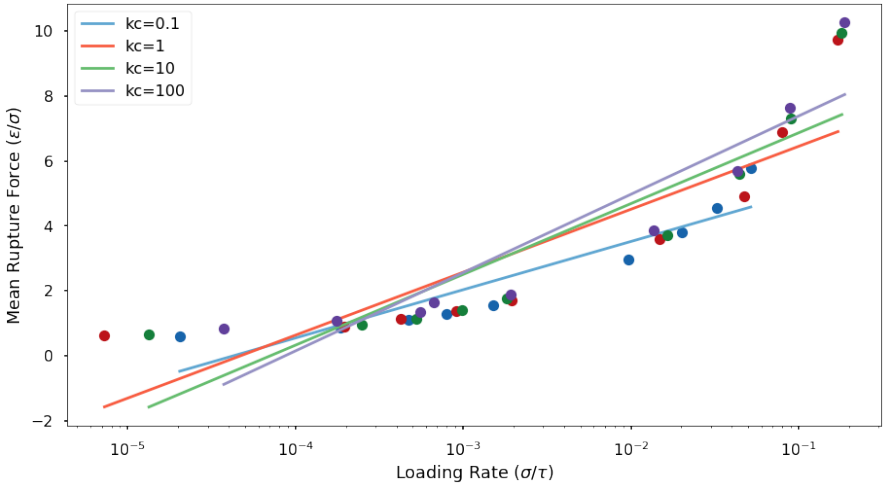
The model does not fit any system particularly well, and the fit parameters do not represent the underlying physics. Specifically, x_t appears to be too minute; the calculated numbers imply the transition state is within $1/10^{\text{th}}$ of a bead width from the ground-state. It is likely x_t would be on the order of the polymer length, i. e. 20σ. A more descriptive model is necessary to fit the system.

5.3.3.2 *Friddle, Noy, deYoreo*

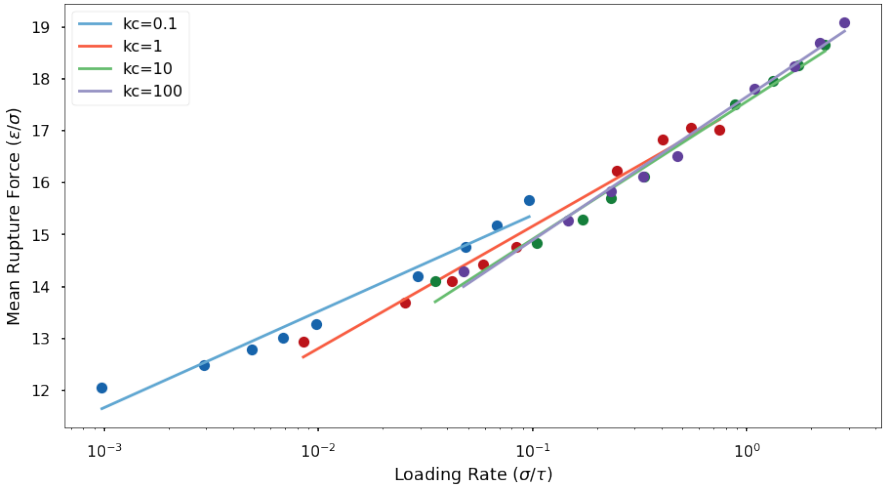
We extracted the loading rate as in Section 5.3.3.1 and fit our average rupture force, loading rates, and spring constants to Eq. 5.34 applying `scipy` library `optimize.curvefit`. [155] Results are exhibited in Table 5.2 and Figure 5.5.

In each of these systems, ΔG is calculated by Equation 5.23 from the fitted values of F_{eq}. The model fits all systems relatively well, though the fit parameters do not match the underlying physics. The value for ΔG should, in theory, be the same for all simulations. We would expect the value to be on the order of 20ϵ, as it is the energy gained by the system for the 20 hybridized beads; instead the values range from 6.76×10^{-3} to $6.98 \times 10^{2}\epsilon$.

Each of the fits express consistent values for F_{eq}, x_t, and, in the unzipping geometry, k_u^0; this is expected, given the similarity of mean rupture force vs. loading rate curves in Figure 5.5. We believe the predicted values of x_t to be far too small; the calculated numbers imply the transition state is within $1/100^{\text{th}}$ of a bead length from the fully bound state. x_t should be on the order of the polymer length, i. e. 20σ.

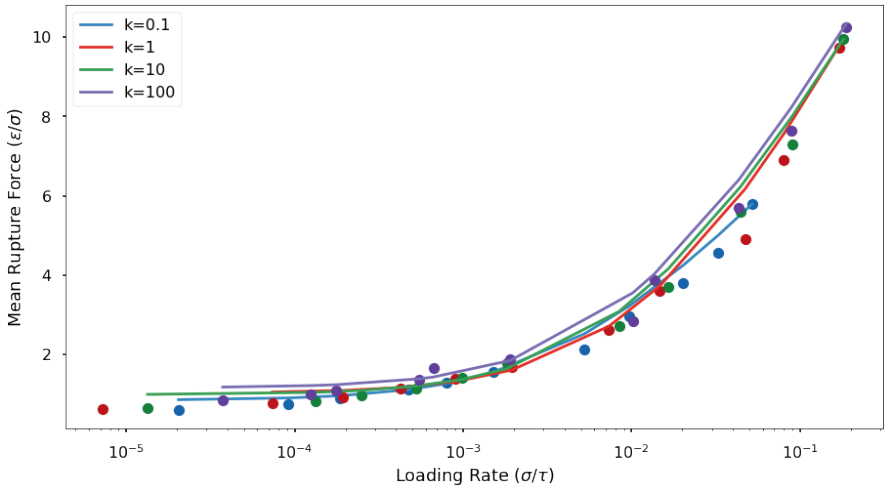


(a) unzip

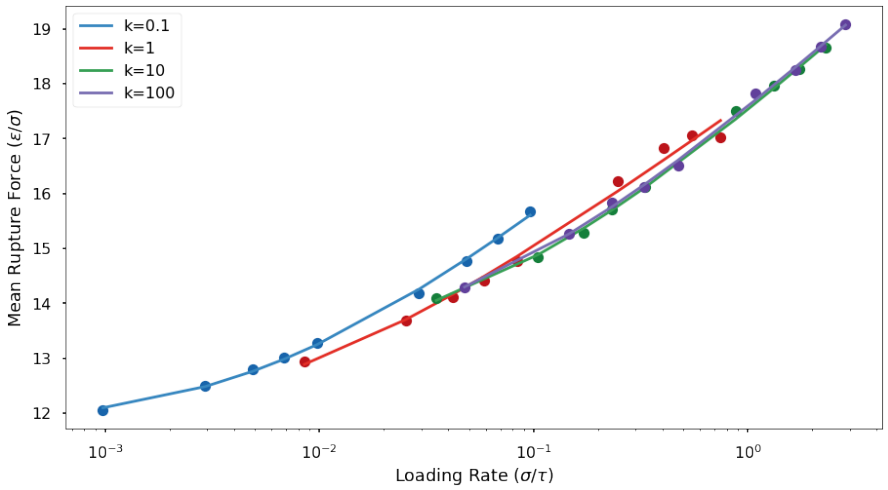


(b) shear

Figure 5.4: Mean rupture force vs. loading rate for both unzipping (a) and shearing (b) geometries. Simulation data are represented with solid points. Lines represent the Bell-Evans theoretical fit. Four different spring constants are depicted: k_c=0.1 (blue), k_c=1 (red), k_c=10 (green), and k_c=100 (purple).



(a) unzip



(b) shear

Figure 5.5: Mean rupture force vs. loading rate for both unzipping (a) and shearing (b) geometries. Simulation data are represented with solid points. Lines represent the Friddle et. al. theoretical fit. Four different spring constants are depicted: k_c=0.1 (blue), k_c=1 (red), k_c=10 (green), and k_c=100 (purple).

	$k_c(\epsilon/\sigma^2)$	$F_{eq}(\epsilon/\sigma)$	$x_t(\sigma)$	$k_u^0(\epsilon/\sigma^2)$	$\Delta G(\epsilon)$
unzip	0.1	8.48×10^{-1}	4.86×10^{-2}	1.05×10^{-3}	3.59
	1	1.03	2.77×10^{-2}	2.18×10^{-3}	5.28×10^{-1}
	10	9.89×10^{-1}	3.08×10^{-2}	1.78×10^{-3}	4.89×10^{-2}
	100	1.16	3.04×10^{-2}	2.55×10^{-3}	6.76×10^{-3}
shear	0.1	1.18×10^{1}	7.36×10^{-2}	4.79×10^{-7}	6.98×10^{2}
	1	1.19×10^{1}	8.18×10^{-2}	2.57×10^{-7}	7.02×10^{1}
	10	1.32×10^{1}	6.56×10^{-2}	5.36×10^{-6}	8.68
	100	1.32×10^{1}	6.59×10^{-2}	3.43×10^{-5}	8.72×10^{-1}

Table 5.2: Fit parameters to our simulation data using the Friddle et. al. model. F_{eq} is the equilibrium force, x_t is the distance between the primary minimum and transition state, k_u^0 is the intrinsic unbinding rate constant, and ΔG is the free-energy difference between bound and unbound states.

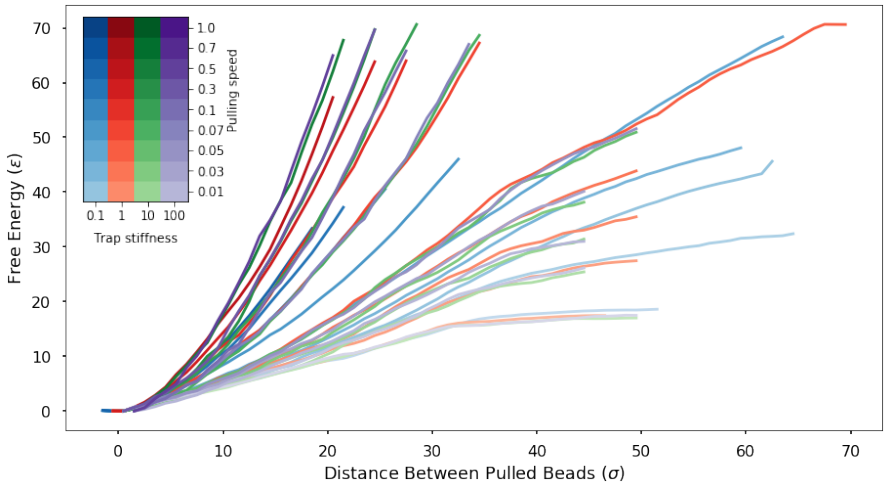
Because F_{eq} displays consistent values, and because ΔG depends on F_{eq} and k_c, mathematically, the values of ΔG would diverge by orders of magnitude for each system. Theoretically, the strength of the trap should not produce any effect on the difference in free-energy. It is worth noting Friddle, Noy, and deYoreo propose a second model which assumes multiple bond rupture events; [150] this model gives similarly poor results for our data (not pictured).

In the supplementary information of their paper, Friddle, Noy, and deYoreo [150] provide an analysis of previous experimental data after applying their model. They display striking fits to the data, but do not comment on the scale of their fit parameters. We show in Table 5.3 that for at least three of the works studied, the calculated values for ΔG do not match other experimentally-determined values. For the three systems, the ΔG predicted by Friddle et. al. differs from other experimental results by 1-2 orders of magnitude.

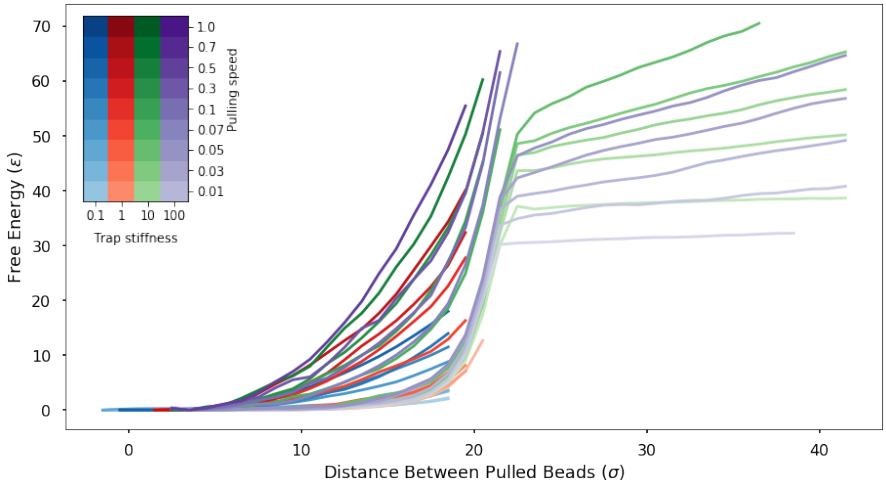
5.3.3.3 *Hummer, Szabo*

Unlike the Friddle-Noy-deYoreo model, the Hummer-Szabo model does not rely on defining mean rupture force, and is instead related to the work profiles of each trajectory. Free-energy profiles are calculated by Equation 5.38 and results are presented in Figure 5.6.

The unzipping geometry registers a steady rise in free-energy as the distance between pulled particles increases. This is consistent with our understanding; as each bead separates one at a time, the energy cost should be proportional to the separation distance. Conversely, shear expresses a flat energy profile until a distance of approximately 20σ, the length of the polymer, when there is a sharp increase in energy. This is consistent with an "all-at-once" melting mechanism,



(a) unzip



(b) shear

Figure 5.6: Free-energy profiles for both unzipping (a) and shearing (b) geometries. Four different spring constants are depicted: k_c=0.1 (blue), k_c=1 (red), k_c=10 (green), and k_c=100 (purple). The curves are pale for slow pulling speeds and dark for fast pulling speeds. Unzipping geometries exhibit a steady rise in free-energy with pulling distance, while shearing geometries present a flatter energy profile with a marked increase at approximately 20σ which again flattens at a greater energy. In each geometry, the free-energy surface differs little with respect to spring constant, but dramatically steepens as pulling speed increases.

BOND	F_{eq} (pN)	k_c (pN/Å)	ΔG^F (kJ/mol)	ΔG^E (kJ/mol)
His6/Ni-NTA	189	1 [156]	2.22×10^3	30 [157]
Digoxigenin/ Anti-Digoxigenin	23.2	6-8.8 [158]	3.8-5.6	33 [159]
Biotin/ Streptavidin	105	1-5 [160]	1.4×10^2-6.9×10^2	10-70 [161]

Table 5.3: Calculations for ΔG based on Friddle et. al.'s work [150]. Values for equilibrium force, F_{eq}, originated from Friddle et. al. k_c values were gathered from the cited works. ΔG^F is the value for free energy calculated by the Friddle-Noy-deYoreo model applying Eq. 5.23. ΔG^E is reported from the cited experimental works. In all three calculations, ΔG^F and ΔG^E differ by 1-2 orders of magnitude.

where first the polymer aligns with the pulling axis, then strain increases on the system until each bead separates from its partner at the same time. After this sharp increase in free-energy, the profile again flattens, as the two strands no longer interact.

Pulling speed appears to generate a large effect on the calculated FESs. As pulling speed increases, the slope of each FES increases, across both geometries. FESs for higher pulling speeds terminate at shorter distances than their slow-pulling counterparts. As the pulling speed increases, and time increases, the work, W_t, performed on the system becomes excessively large, driving η_i (Eq. 5.41 to be effectively 0. This prevents the analysis of $G(x)$ (Eq. 5.34) at large distances.

In their paper describing the model, [151] Hummer and Szabo state pulling speed likely affects its accuracy. The faster the pulling speed, the greater the variance in observed work at a given time step, and the larger the associated error. Thus, we believe the FESs at low pulling velocities to be the most accurate. Trap stiffness produces no observable effect on these FESs, as expected: the strength of the trap should not influence the final free-energy surface.

For those stiffnesses and pulling speeds which display a clear ΔG, the value is similar between the two geometries, as expected by the Second Law of Thermodynamics. In the slow-pulling limit, which is closest to equilibrium, the ΔG value is 18 for the unzipping geometry, and 32 for the shearing geometry. Each bead pair provides 1 potential-energy unit of stability to the system; we would expect the free-energy to be close to 20ϵ, which it is for both systems.

5.3.4 *Recommendations*

We believe the Hummer-Szabo model to provide the most desirable fit for our data, as it constructs the entire FES and not properties of it, produces consistent results between geometries and across spring constants, and the values for ΔG and x_t are closest to the hypothesized values. In practice, the Hummer-Szabo model is limited to slow pulling speeds, where the entire FES can be calculated, and variance in the work distributions is limited.

Both the Bell-Evans and the Friddle model are highly sensitive to the definition of mean rupture force, $\langle F \rangle$, which can be estimated several different ways. Both models produce results which depend on the trap strength, k_c, which isn't allowed by the Second Law of Thermodynamics. The Hummer-Szabo model supplies the same results for different trap strengths, and does not depend on defining mean rupture force. The model does generate different FESs for different pulling speeds; we recommend pulling as slowly as possible, which will decrease the variance in the work profiles and produce more accurate free-energy surfaces.

Part III

DENOUEMENT

To absent friends, lost loves, old gods, and the season of mists; and may each and every one of us always give the devil his due.

— Neil Gaiman, Season of Mists

6 CONCLUSIONS

We have explored unique ways in which mismatched base pairs in DNA affect their surrounding environment: in structural properties, energetics, and effects on neighboring azobenzene moieties. Using molecular dynamics (MD) simulations, both atomistic and coarse-grained, we explored and contrasted varied theoretical models to calculate energetic and kinetic properties. These contribute to the greater understanding of DNA, both in the biological properties of mismatch repair (MMR) enzymes and in interpreting and designing DNA-containing materials.

6.1 KEY FINDINGS

Mispaired DNA bases display increased conformational fluctuations, especially in the plane perpendicular to the length of the DNA. These structures, which diverge from canonical Watson-Crick pairs, exhibit fewer hydrogen bonds and thus less overall stability.

Base pair mismatches alter their local environment. *trans*-to-*cis* photoisomerization quantum yield (QY) of azobenzene incorporated into DNA depends on the identities of neighboring base pairs and their relative stability. Mispairs are inclined to adopt the e-motif, which induces large structural distortions in the surrounding polymer.

Many biasing techniques and models exist to extract the free-energy from a system. Selecting parameters is critical for the model to succeed: in the metadynamics of base flip, we revealed that selecting two order parameters provides more accurate results than one. Similarly, both choosing the model and defining mean rupture force affect the results of the dynamic force spectroscopy (DFS) free-energies. We explored the importance of grasping the underlying systems and methods and choosing proper inputs for each model.

6.2 ANALYSIS

MD simulations remain a useful tool for describing DNA behavior. The photoisomerization of azobenzene requires both an electronic excitation and subsequent nuclear motion, but we demonstrate that a relatively simple observable, the variance in occupation number, is strongly correlated with QY. This value can be computed from *classical* MD computer simulations, and yet correlates with an experimentally-determined *quantum* process.

This provides contemporary insight into the physics of QY trends. QY of azobenzene is affected by both the greater probabilities of DNA fluctuations which allow the azobenzene's phenyl ring to shift unhindered and a larger susceptibility to an effective thermodynamic force which decreases the local occupation number. This latest understanding will aid future designers of materials who wish to incorporate light-sensitive photoswitches into DNA.

MD has allowed us to observe an e-motif in some mismatched structures, a result previously unexplored as a possible indicator for MMR recognition. We established that DNA structure differs for mispaired bases without the presence of an MMR enzyme, but that these structural differences do not match MMR recognition trends and are thus not sufficient to predict the pathway for MMR recognition. If the e-motif is a real, physical phenomenon of mispaired DNA and not a simulation artifact, it could be an influence in the detection capabilities of MMR systems.

With MD, we were able to compare and contrast theoretical models and recommend the optimum model to apply to experimentally-gathered data. We identified the Hummer-Szabo model as the most accurate for resolving free-energy profiles for DFS experiments. This is because it does not depend on defining mean rupture force, is consistent between differing trap strengths, and reconstructs the whole free-energy profile, not properties of it. For the model to function most effectively, it is vital to experiment at the slowest possible speed to reduce the variance in generated work profiles.

6.3 FUTURE RESEARCH

We recommend several further studies, especially for experimental researchers. Many considerations contribute to the specific interactions between DNA and azobenzene. Samai et. al. [162] recently indicated temperature is a primary contributor to QY effects. Environment effects such as pressure, pH, and ionic strength should be studied in-depth.

We greatly urge experimentalists researching DNA structure, especially in the context of MMR systems, to seek further evidence of the e-motif. This would be especially telling for traditional sequences which do not retain repeats.

Finally, for theoreticians studying DFS pulling dynamics, we recommend systematic study of pulling speeds' effects on free-energy calculation, including an implementation of cumulant expansion corrections to the variance in work profiles. Ideally such research would establish a means for detecting the ideal pulling speed, both in terms of accuracy and resource management.

6.4 IMPLICATIONS

These results demonstrate how to adopt inexpensive computer simulations to guide the design of azobenzene-modified DNA (azo-DNA) sequences and linkers [134, 163] to attain specific objectives, for example, maximizing QY. We have demonstrated that MD simulations can provide structural insight into mispaired DNA bases, with possible consequences regarding the detection and repair of mismatches by MMR systems. Finally, simulations remain a powerful tool in comparing and validating models to fit experimental data.

The range of applications, especially for materials, is vast: from triggering structural remodling in aptamers [31] to designing probe sequences [164, 165] optimized to discriminate specific DNA sequences. We expect our results to be of immediate practical use in those development efforts.

BIBLIOGRAPHY

(1) Geggier, S.; Vologodskii, A. Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15421–15426.

(2) Dahm, R. Friedrich Miescher and the discovery of DNA. *Dev. Biol.* **2005**, *278*, 274–288.

(3) Sayre, A., *Rosalind Franklin and DNA*; W. W. Norton & Company: 2000.

(4) Watson, J. D.; Crick, F. H. C. The Structure of DNA. *Cold Spring Harb. Sym.* **1953**, *18*, 123–131.

(5) Hu, C.; Zhao, Y.; Sun, H.; Yang, Y. Synergism of Dam, MutH, and MutS in methylation-directed mismatch repair in Escherichia coli. *Mutat. Res.-Fund. Mol. M.* **2017**, *795*, 31–33.

(6) De Wind, N.; Dekker, M.; Claij, N.; Jansen, L.; van Klink, Y.; Radman, M.; Riggins, G.; van der Valk, M.; van't Wout, K.; te Riele, H. HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions. *Nat. Genet.* **1999**, *23*, 359–362.

(7) Kunkel, T. A.; Erie, D. A. DNA mismatch repair. *Annu. Rev. Biochem.* **2005**, *74*, 681–710.

(8) Iyer, R. R.; Pluciennik, A.; Burdett, V.; Modrich, P. L. DNA mismatch repair: Functions and mechanisms. *Chem. Rev.* **2006**, *106*, 302–323.

(9) Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **2008**, *18*, 85–98.

(10) Kolodner, R. D. et al. Human mismatch repair genes and their association with hereditary non-polyposis colon cancer. *Cold Spring Harb. Sym.* **1994**, *59*, 331–338.

(11) Harfe, B. D.; Jinks-Robertson, S. DNA mismatch repair and genetic instability. *Annu. Rev. Genet.* **2000**, *34*, 359–399.

(12) Stivers, J. T.; Jiang, Y. L. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **2003**, *103*, 2729–2760.

(13) Kunkel, T. A.; Erie, D. A. Eukaryotic mismatch repair in relation to DNA replication. *Annu. Rev. Genet.* **2015**, *49*, 291–313.

(14) Drake, J. W. The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. *Ann. NY. Acad. Sci.* **1999**, *870*, 100–107.

(15) Bancroft, C.; Bowler, T.; Bloom, B.; Clelland, C. T. Long-term storage of information in DNA. *Science* **2001**, *293*, 1763–1765.

(16) Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E. M.; Sipos, B.; Birney, E. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **2013**, *494*, 77–80.

(17) Church, G. M.; Gao, Y.; Kosuri, S. Next-generation digital information storage in DNA. *Science* **2012**, 1226355.

(18) Han, D.; Pal, S.; Nangreave, J.; Deng, Z.; Liu, Y.; Yan, H. DNA origami with complex curvatures in three-dimensional space. *Science* **2011**, *332*, 342–346.

(19) Alemany, A.; Ritort, F. Force-dependent folding and unfolding kinetics in DNA hairpins reveals transition-state displacements along a single pathway. *J. Phys. Chem. Lett.* **2017**, *8*, 895–900.

(20) Thubagere, A. J. et al. A cargo-sorting DNA robot. *Science* **2017**, *357*, eaan6558.

(21) Chen, J. I. L.; Chen, Y.; Ginger, D. S. Plasmonic nanoparticle dimers for optical sensing of DNA in complex media. *J. Am. Chem. Soc.* **2010**, *132*, 9600–9601.

(22) Demers, L. M.; Ginger, D. S.; Park, S.-J.; Li, Z.; Chung, S.-W.; Mirkin, C. A. Direct patterning of modified oligonucleotides on metals and insulators by dip-pen nanolithography. *Science* **2002**, *296*, 1836–1838.

(23) Ding, B.; Deng, Z.; Yan, H.; Cabrini, S.; Zuckermann, R. N.; Bokor, J. Gold nanoparticle self-similar chain structure organized by DNA origami. *J. Am. Chem. Soc.* **2010**, *132*, 3248–3249.

(24) Yamazawa, A.; Liang, X.; Asanuma, H.; Komiyama, M. Photoregulation of the DNA Polymerase reaction by oligonucleotides bearing an azobenzene. *Angew. Chem.-Ger. Ed.* **2000**, *112*, 2446–2447.

(25) Matsunaga, D.; Asanuma, H.; Komiyama, M. Photoregulation of RNA digestion by RNase H with azobenzene-tethered DNA. *J. Am. Chem. Soc.* **2004**, *126*, 11452–11453.

(26) Liu, Y.; Sen, D. Light-regulated catalysis by an RNA-cleaving deoxyribozyme. *J. Mol. Biol.* **2004**, *341*, 887–892.

(27) Liang, X.; Fujioka, K.; Asanuma, H. Nick sealing by T4 DNA Ligase on a modified DNA template: Tethering a functional molecule on D-threoninol. *Chem.-Eur. J.* **2011**, *17*, 10388–10396.

(28) Hara, Y.; Fujii, T.; Kashida, H.; Sekiguchi, K.; Liang, X.; Niwa, K.; Takase, T.; Yoshida, Y.; Asanuma, H. Coherent quenching of a fluorophore for the design of a highly sensitive in-stem molecular beacon. *Angew. Chem. Int. Ed.* **2010**, *49*, 5502–5506.

(29) Asanuma, H.; Osawa, T.; Kashida, H.; Fujii, T.; Liang, X.; Niwa, K.; Yoshida, Y.; Shimada, N.; Maruyama, A. A polycation-chaperoned in-stem molecular beacon system. *Chem. Commun.* **2012**, *48*, 1760–1762.

(30) Kim, Y.; Phillips, J. A.; Liu, H.; Kang, H.; Tan, W. Using photons to manipulate enzyme inhibition by an azobenzene-modified nucleic acid probe. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6489–6494.

(31) Phillips, J. A.; Liu, H.; O'Donoghue, M. B.; Xiong, X.; Wang, R.; You, M.; Sefah, K.; Tan, W. Using azobenzene incorporated DNA aptamers to probe molecular binding interactions. *Bioconjugate Chem.* **2011**, *22*, 282–288.

(32) Tanaka, F.; Mochizuki, T.; Liang, X.; Asanuma, H.; Tanaka, S.; Suzuki, K.; Kitamura, S.-i.; Nishikawa, A.; Ui-Tei, K.; Hagiya, M. Robust and photocontrollable DNA capsules using azobenzenes. *Nano Lett.* **2010**, *10*, 3560–3565.

(33) Liang, X.; Nishioka, H.; Takenaka, N.; Asanuma, H. A DNA nanomachine powered by light irradiation. *ChemBioChem* **2008**, *9*, 702–705.

(34) Nishioka, H.; Liang, X.; Kato, T.; Asanuma, H. A photon-fueled DNA nanodevice that contains two different photoswitches. *Angew. Chem. Int. Ed.* **2012**, *51*, 1165–1168.

(35) Yuan, Q.; Zhang, Y.; Chen, Y.; Wang, R.; Du, C.; Yasun, E.; Tan, W. Using silver nanowire antennas to enhance the conversion efficiency of photoresponsive DNA nanomotors. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9331–9336.

(36) Yuan, Q.; Zhang, Y.; Chen, T.; Lu, D.; Zhao, Z.; Zhang, X.; Li, Z.; Yan, C.-H.; Tan, W. Photon-manipulated drug release from a mesoporous nanocontainer controlled by azobenzene-modified nucleic acid. *ACS Nano* **2012**, *6*, 6337–6344.

(37) Tang, Z.; Lim, C.-K.; Palafox-Hernandez, J. P.; Drew, K. L. M.; Li, Y.; Swihart, M. T.; Prasad, P. N.; Walsh, T. R.; Knecht, M. R. Triggering nanoparticle surface ligand rearrangement via external stimuli: Light-based actuation of biointerfaces. *Nanoscale* **2015**, *7*, 13638–13645.

(38) Qian, Z.; Ginger, D. S. Reversibly reconfigurable colloidal plasmonic nanomaterials. *J. Am. Chem. Soc.* **2017**, *139*, 5266–5276.

(39) Kramer, B.; Kramer, W.; Fritz, H.-J. Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli. *Cell* **1984**, *38*, 879–887.

(40) Su, S. S.; Lahue, R. S.; Au, K. G.; Modrich, P. Mispair specificity of methyl-directed DNA mismatch correction in vitro. *J. Biol. Chem.* **1988**, *263*, 6829–6835.

(41) Thomas, D. C.; Roberts, J. D.; Kunkel, T. A. Heteroduplex repair in extracts of human HeLa cells. *J. Biol. Chem.* **1991**, *266*, 3744–3751.

(42) Imhof, P.; Zahran, M. The effect of a G:T mispair on the dynamics of DNA. *PLoS ONE* **2013**, *8*, e53305.

(43) Sharma, M.; Predeus, A. V.; Mukherjee, S.; Feig, M. DNA bending propensity in the presence of base mismatches: Implications for DNA repair. *J. Phys. Chem. B* **2013**, *117*, 6194–6205.

(44) Sharma, M.; Predeus, A.; Kovacs, N.; Feig, M. Differential mismatch recognition specificities of eukaryotic MutS homologs, MutSα and MutSβ. *Biophys. J.* **2014**, *106*, 2483–2492.

(45) Lau, A. Y.; Wyatt, M. D.; Glassner, B. J.; Samson, L. D.; Ellenberger, T. Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13573–13578.

(46) Steinwand, S.; Halbritter, T.; Rastädter, D.; Ortiz-Sánchez, J. M.; Burghardt, I.; Heckel, A.; Wachtveitl, J. Ultrafast spectroscopy of hydroxy-substituted azobenzenes in water. *Chem.-Eur. J.* **2015**, *21*, 15720–15731.

(47) Yu, L.; Xu, C.; Zhu, C. Probing the $\pi \rightarrow \pi*$ photoisomerization mechanism of cis-azobenzene by multi-state ab initio on-the-fly trajectory dynamics simulation. *Physical Chemistry Chemical Physics* **2015**, *17*, 17646–17660.

(48) Biswas, M.; Burghardt, I. Azobenzene photoisomerization-induced destabilization of B-DNA. *Biophys. J.* **2014**, *107*, 932–940.

(49) Quick, M.; Dobryakov, A. L.; Gerecke, M.; Richter, C.; Berndt, F.; Ioffe, I. N.; Granovsky, A. A.; Mahrwald, R.; Ernsting, N. P.; Kovalenko, S. A. Photoisomerization dynamics and pathways of trans- and cis-azobenzene in solution from broadband femtosecond spectroscopies and calculations. *J. Phys. Chem. B* **2014**, *118*, 8756–8771.

(50) Gao, A.-H.; Li, B.; Zhang, P.-Y.; Liu, J. Photochemical dynamics simulations for trans-cis photoisomerizations of azobenzene and bridged azobenzene. *Comput. Theor. Chem.* **2014**, *1031*, 13–21.

(51) Gámez, J. A.; Weingart, O.; Koslowski, A.; Thiel, W. Cooperating dinitrogen and phenyl rotations in trans-azobenzene photoisomerization. *J. Chem. Theory Comput.* **2012**, *8*, 2352–2358.

(52) Böckmann, M.; Marx, D.; Peter, C.; Site, L. D.; Kremer, K.; Doltsinis, N. L. Multiscale modelling of mesoscopic phenomena triggered by quantum events: Light-driven azo-materials and beyond. *Phys. Chem. Chem. Phys.* **2011**, *13*, 7604–7621.

(53) Jiang, C.-W.; Xie, R.-H.; Li, F.-L.; Allen, R. E. Comparative studies of the trans-cis photoisomerizations of azobenzene and a bridged azobenzene. *J. Phys. Chem. A* **2011**, *115*, 244–249.

(54) Böckmann, M.; Doltsinis, N. L.; Marx, D. Nonadiabatic hybrid quantum and molecular mechanic simulations of azobenzene photoswitching in bulk liquid environment. *J. Phys. Chem. A* **2010**, *114*, 745–754.

(55) Böckmann, M.; Doltsinis, N. L.; Marx, D. Azobenzene photoswitches in bulk materials. *Phys. Rev. E* **2008**, *78*, 036101.

(56) Sampedro Ruiz, D.; Cembran, A.; Garavelli, M.; Olivucci, M.; Fuß, W. Structure of the conical intersections driving the cis-trans photoisomerization of conjugated molecules. *Photochem. Photobiol.* **2002**, *76*, 622–633.

(57) Lee, S.; Kang, H. S.; Park, J.-K. Directional photofluidization lithography: Micro/nanostructural evolution by photofluidic motions of azobenzene materials. *Adv. Mater.* **2012**, *24*, 2069–2103.

(58) Bandara, H. M. D.; Burdette, S. C. Photoisomerization in different classes of azobenzene. *Chem. Soc. Rev.* **2012**, *41*, 1809–1825.

(59) Rau, H. In *Photoreactive Organic Thin Films*, Sekkat, Z., Knoll, W., Eds., DOI: 10.1016/B978-012635490-4/50002-0; Academic Press: San Diego, 2002, pp 3–47.

(60) Asanuma, H.; Matsunaga, D.; Liu, M.; Liang, X.; Jhao, J.; Komiyama, M. Photo-regulation of DNA function by azobenzene-tethered oligonucleotides. *Nucl. Acid S.* **2003**, *3*, 117–118.

(61) Asanuma, H.; Liang, X.; Nishioka, H.; Matsunaga, D.; Liu, M.; Komiyama, M. Synthesis of azobenzene-tethered DNA for reversible photo-regulation of DNA functions: hybridization and transcription. *Nature Protocols* **2007**, *2*, 203–212.

(62) Liang, X.; Asanuma, H.; Kashida, H.; Takasu, A.; Sakamoto, T.; Kawai, G.; Komiyama, M. NMR study on the photoresponsive DNA tethering an azobenzene. Assignment of the absolute configuration of two diastereomers and structure determination of their duplexes in the trans-form. *J. Am. Chem. Soc.* **2003**, *125*, 16408–16415.

(63) Asanuma, H.; Ito, T.; Yoshida, T.; Liang, X.; Komiyama, M. Photoregulation of the formation and dissociation of a DNA duplex by using the cis-trans isomerization of azobenzene. *Angew. Chem. Int. Ed.* **1999**, *38*, 2393–2395.

(64) Asanuma, H.; Liang, X.; Yoshida, T.; Komiyama, M. Photocontrol of DNA duplex formation by using azobenzene-bearing oligonucleotides. *ChemBioChem* **2001**, *2*, 39–44.

(65) Yan, Y.; Wang, X.; Chen, J. I. L.; Ginger, D. S. Photoisomerization quantum yield of azobenzene-modified DNA depends on local sequence. *J. Am. Chem. Soc.* **2013**, *135*, 8382–8387.

(66) Alsteens, D.; Pfreundschuh, M.; Zhang, C.; Spoerri, P. M.; Coughlin, S. R.; Kobilka, B. K.; Müller, D. J. Imaging G protein-coupled receptors while quantifying their ligand-binding free-energy landscape. *Nat. Methods* **2015**, *12*, 845–851.

(67) Rief, M.; Clausen-Schaumann, H.; Gaub, H. E. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Mol. Biol.* **1999**, *6*, 346–349.

(68) Merkel, R.; Nassoy, P.; Leung, A.; Ritchie, K.; Evans, E. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* **1999**, *397*, 50–53.

(69) Bizzarri, A. R.; Cannistraro, S. The application of atomic force spectroscopy to the study of biological complexes undergoing a biorecognition process. *Chem. Soc. Rev.* **2010**, *39*, 734–749.

(70) Strunz, T.; Oroszlan, K.; Schäfer, R.; Güntherodt, H.-J. Dynamic force spectroscopy of single DNA molecules. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11277–11282.

(71) Zhang, J.; Yan, Y.; Samai, S.; Ginger, D. S. Dynamic melting properties of photoswitch-modified DNA: Shearing versus unzipping. *J. Phys. Chem. B* **2016**, *120*, 10706–10713.

(72) Banavali, N. K.; MacKerell Jr., A. D. Free energy and structural pathways of base flipping in a DNA GCGC containing sequence. *J. Mol. Biol.* **2002**, *319*, 141–160.

(73) Song, K.; Campbell, A. J.; Bergonzo, C.; de los Santos, C.; Grollman, A. P.; Simmerling, C. An improved reaction coordinate for nucleic acid base flipping studies. *J. Chem. Theory Comput.* **2009**, *5*, 3105–3113.

(74) McCullagh, M.; Franco, I.; Ratner, M. A.; Schatz, G. C. DNA-based optomechanical molecular motor. *J. Am. Chem. Soc.* **2011**, *133*, 3452–3459.

(75) McCullagh, M.; Franco, I.; Ratner, M. A.; Schatz, G. C. Defects in DNA: Lessons from molecular motor design. *J. Phy. Chem. Lett.* **2012**, *3*, 689–693.

(76) Kouza, M.; Chang, C.-F.; Hayryan, S.; Yu, T.-h.; Li, M. S.; Huang, T.-h.; Hu, C.-K. Folding of the protein domain hbSBD. *Biophys. J.* **2005**, *89*, 3353–3361.

(77) Li, M. S. Secondary structure, mechanical stability, and location of transition state of proteins. *Biophys. J.* **2007**, *93*, 2644–2654.

(78) Kumar, S.; Jensen, I.; Jacobsen, J. L.; Guttmann, A. J. Role of conformational entropy in force-induced biopolymer unfolding. *Phys. Rev. Lett.* **2007**, *98*, 128101.

(79) Schluttig, J.; Bachmann, M.; Janke, W. Comparative molecular dynamics and Monte Carlo study of statistical properties for coarse-grained heteropolymers. *J. Comput. Chem.* **2008**, *29*, 2603–2612.

(80) Mishra, R. K.; Mishra, G.; Li, M. S.; Kumar, S. Effect of shear force on the separation of double-stranded DNA. *Phys. Rev. E* **2011**, *84*, 032903.

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

(82) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(83) MacKerell, A. D.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **2000**, *56*, 257–265.

(84) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(85) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(86) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(87) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(88) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(89) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(90) Cheatham, T. E.; Young, M. A. Molecular dynamics simulation of nucleic acids: Successes, limitations, and promise. *Biopolymers* **2000**, *56*, 232–256.

(91) Feig, M.; Pettitt, B. M. Experiment vs force fields: DNA conformation from molecular dynamics simulations. *J. Phys. Chem. B* **1997**, *101*, 7361–7363.

(92) Feig, M.; Pettitt, B. M. Structural equilibrium of DNA represented with different force fields. *Biophys. J.* **1998**, *75*, 134–149.

(93) Cao, L.; Ren, H.; Miao, J.; Guo, W.; Li, Y.; Li, G. Validation of polarizable force field parameters for nucleic acids by intermolecular interactions. *Front. Chem. Sci. Eng.* **2016**, *10*, 203–212.

(94) Yoon, C.; Privé, G. G.; Goodsell, D. S.; Dickerson, R. E. Structure of an alternating-B DNA helix and its relationship to A-tract DNA. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 6332–6336.

(95) Stivers, J. T. Extrahelical damaged base recognition by DNA glycosylase enzymes. *Chem.-Eur. J.* **2008**, *14*, 786–793.

(96) Law, S.; Feig, M. Base-flipping mechanism in postmismatch recognition by MutS. *Biophys. J.* **2011**, *101*, 2223–2231.

(97) Erie, D. A.; Weninger, K. R. Single molecule studies of DNA mismatch repair. *DNA Repair* **2014**, *20*, 71–81.

(98) Vafabakhsh, R.; Ha, T. Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science* **2012**, *337*, 1097–1101.

(99) Cock, P. J. A. et al. Biopython: Freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423.

(100) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(101) Bonomi, M.; Barducci, A.; Parrinello, M. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.* **2009**, *30*, 1615–1621.

(102) Turner, D. H.; Sugimoto, N.; Kierzek, R.; Dreiker, S. D. Free energy increments for hydrogen bonds in nucleic acid base pairs. *J. Am. Chem. Soc.* **1987**, *109*, 3783–3785.

(103) Šponer, J.; Leszczynski, J.; Hobza, P. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers* **2001**, *61*, 3–31.

(104) Kool, E. T. Hydrogen bonding, base stacking, and steric effects in DNA replication. *Annu. Rev. Bioph. Biom.* **2001**, *30*, 1–22.

(105) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **2006**, *34*, 564–574.

(106) Khakshoor, O.; Wheeler, S. E.; Houk, K. N.; Kool, E. T. Measurement and theory of hydrogen bonding contribution to isosteric DNA base pairs. *J. Am. Chem. Soc.* **2012**, *134*, 3154–3163.

(107) Shankar, A.; Jagota, A.; Mittal, J. DNA base dimers are stabilized by hydrogen-bonding interactions including non-Watson–Crick pairing near graphite surfaces. *J. Phys. Chem. B* **2012**, *116*, 12088–12094.

(108) Peyret, N.; Seneviratne, P. A.; Allawi, H. T.; SantaLucia, J. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A·A, C·C, G·G, and T·T mismatches. *Biochemistry* **1999**, *38*, 3468–3477.

(109) SantaLucia Jr., J.; Hicks, D. The thermodynamics of DNA structural motifs. *Annu. Rev. Bioph. Biom.* **2004**, *33*, 415–C–2.

(110) Harris, N. C.; Kiang, C.-H. Defects can increase the melting temperature of DNA-nanoparticle assemblies. *J. Phys. Chem. B* **2006**, *110*, 16393–16396.

(111) Geyer, C. R.; Battersby, T. R.; Benner, S. A. Nucleobase pairing in expanded Watson-Crick-like genetic information systems. *Structure* **2003**, *11*, 1485–1498.

(112) Laio, A.; Gervasio, F. L. Metadynamics: A method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **2008**, *71*, 126601–126623.

(113) Yin, Y.; Yang, L.; Zheng, G.; Gu, C.; Yi, C.; He, C.; Gao, Y. Q.; Zhao, X. S. Dynamics of spontaneous flipping of a mismatched base in DNA duplex. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8043–8048.

(114) O'Nei, L. L.; Wiest, O. Structures and energetics of base flipping of the thymine dimer depend on DNA sequence. *J. Phys. Chem. B* **2008**, *112*, 4113–4122.

(115) Kouchakdjian, M.; Li, B. F. L.; Swann, P. F.; Patel, D. J. Pyrimidine · pyrimidine base-pair mismatches in DNA. *J. Mol. Biol.* **1988**, *202*, 139–155.

(116) Gervais, V.; Cognet, J. A. H.; Le Bret, M.; Sowers, L. C.; Fazakerley, G. V. Solution structure of two mismatches A·A and T·T in the K-ras gene context by nuclear magnetic resonance and molecular dynamics. *Eur. J. Biochem.* **1995**, *228*, 279–290.

(117) Boulard, Y.; Cognet, J. A. H.; Fazakerley, G. V. Solution structure as a function of pH of two central mismatches, C·T and C·C, in the 29 to 39 K-ras gene sequence, by nuclear magnetic resonance and molecular dynamics. *J. Mol. Biol.* **1997**, *268*, 331–347.

(118) He, G.; Kwok, C. K.; Lam, S. L. Preferential base pairing modes of T·T mismatches. *FEBS Lett.* **2011**, *585*, 3953–3958.

(119) Gao, X.; Huang, X.; Smith, G. K.; Zheng, M.; Liu, H. New antiparallel duplex motif of DNA CCG repeats that is stabilized by extrahelical bases symmetrically located in the minor groove. *J. Am. Chem. Soc.* **1995**, *117*, 8883–8884.

(120) Zheng, M.; Huang, X.; Smith, G. K.; Yang, X.; Gao, X. Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding. An NMR and UV spectroscopic study. *J. Mol. Biol.* **1996**, *264*, 323–336.

(121) Yu, A.; Barron, M. D.; Romero, R. M.; Christy, M.; Gold, B.; Dai, J.; Gray, D. M.; Haworth, I. S.; Mitas, M. At physiological pH, d(CCG)15 forms a hairpin containing protonated cytosines and a distorted helix. *Biochemistry* **1997**, *36*, 3687–3699.

(122) Zhang, Y.; Roland, C.; Sagui, C. Structure and dynamics of DNA and RNA double helices obtained from the GGGGCC and CCCCGG hexanucleotide repeats that are the hallmark of C9FTD/ALS diseases. *ACS Chem. Neurosci.* **2017**, *8*, 578–591.

(123) Chou, S.-H.; Chin, K.-H.; Wang, A. H.-J. Unusual DNA duplex and hairpin motifs. *Nucleic Acids Res.* **2003**, *31*, 2461–2474.

(124) Romero, R. M.; Mitas, M.; Haworth, I. S. Anomalous crosslinking by mechlorethamine of DNA duplexes containing C–C mismatch pairs. *Biochemistry* **1999**, *38*, 3641–3648.

(125) Edfeldt, N. B. F.; Harwood, E. A.; Sigurdsson, S. T.; Hopkins, P. B.; Reid, B. R. Solution structure of a nitrous acid induced DNA interstrand cross-link. *Nucleic Acids Res.* **2004**, *32*, 2785–2794.

(126) Renčiuk, D.; Kypr, J.; Vorlíčková, M. CGG repeats associated with fragile X chromosome form left-handed Z-DNA structure. *Biopolymers* **2011**, *95*, 174–181.

(127) Rojsitthisak, P.; Romero, R. M.; Haworth, I. S. Extrahelical cytosine bases in DNA duplexes containing d[GCC]n·d[GCC]n repeats: Detection by a mechlorethamine crosslinking reaction. *Nucleic Acids Res.* **2001**, *29*, 4716–4723.

(128) Edfeldt, N. B. F.; Harwood, E. A.; Sigurdsson, S. T.; Hopkins, P. B.; Reid, B. R. Sequence context effect on the structure of nitrous acid induced DNA interstrand cross-links. *Nucleic Acids Res.* **2004**, *32*, 2795–2801.

(129) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(130) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(131) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W., et al. AMBER 2015. *University of California, San Francisco* **2015**.

(132) Bonomi, M. et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.

(133) Rastädter, D.; Biswas, M.; Burghardt, I. Molecular dynamics study of the controlled destabilization of an rna hairpin structure by a covalently attached azobenzene switch. *J. Phys. Chem. B* **2014**, *118*, 8478–8488.

(134) Mondal, P.; Biswas, M.; Goldau, T.; Heckel, A.; Burghardt, I. In search of an efficient photoswitch for functional RNA: Design principles from a microscopic analysis of azobenzene-linker-RNA dynamics with different linkers. *J. Phys. Chem. B* **2015**, *119*, 11275–11286.

(135) Goldau, T.; Murayama, K.; Brieke, C.; Steinwand, S.; Mondal, P.; Biswas, M.; Burghardt, I.; Wachtveitl, J.; Asanuma, H.; Heckel, A. Reversible photoswitching of RNA hybridization at room temperature with an azobenzene C-nucleoside. *Chem.-Eur. J.* **2015**, *21*, 2845–2854.

(136) Duchstein, P.; Neiss, C.; Görling, A.; Zahn, D. Molecular mechanics modeling of azobenzene-based photoswitches. *J. Mol. Model.* **2012**, *18*, 2479–2482.

(137) Crecca, C. R.; Roitberg, A. E. Theoretical study of the isomerization mechanism of azobenzene and disubstituted azobenzene derivatives. *J. Phys. Chem. A* **2006**, *110*, 8188–8203.

(138) Lee, G. U.; Chrisey, L. A.; Colton, R. J. Direct measurement of the forces between complementary strands of DNA. *Science* **1994**, *266*, 771–773.

(139) Noy, A.; Vezenov, D. V.; Kayyem, J. F.; Maade, T. J.; Lieber, C. M. Stretching and breaking duplex DNA by chemical force microscopy. *Chem. Biol.* **1997**, *4*, 519–527.

(140) Albrecht, C.; Blank, K.; Lalic-Mülthaler, M.; Hirler, S.; Mai, T.; Gilbert, I.; Schiffmann, S.; Bayer, T.; Clausen-Schaumann, H.; Gaub, H. E. DNA: A programmable force sensor. *Science* **2003**, *301*, 367–370.

(141) Rahbani, J. F.; Hariri, A. A.; Cosa, G.; Sleiman, H. F. Dynamic DNA nanotubes: Reversible switching between single and double-stranded forms, and effect of base deletions. *ACS Nano* **2015**, *9*, 11898–11908.

(142) Ririe, K. M.; Rasmussen, R. P.; Wittwer, C. T. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal. Biochem.* **1997**, *245*, 154–160.

(143) Kamiya, Y.; Takagi, T.; Ooi, H.; Ito, H.; Liang, X.; Asanuma, H. Synthetic gene involving azobenzene-tethered T7 promoter for the photocontrol of gene expression by visible light. *ACS Synth. Biol.* **2015**, *4*, 365–370.

(144) Severin, P. M. D.; Zou, X.; Schulten, K.; Gaub, H. E. Effects of cytosine hydroxymethylation on DNA strand separation. *Biophys. J.* **2013**, *104*, 208–215.

(145) Sengupta, E.; Yan, Y.; Wang, X.; Munechika, K.; Ginger, D. S. Dynamic force spectroscopy of photoswitch-modified DNA. *ACS Nano* **2014**, *8*, 2625–2631.

(146) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* **1971**, *54*, 5237–5247.

(147) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* **2003**, *119*, 3559–3566.

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

(149) Evans, E.; Leung, A.; Hammer, D.; Simon, S. Chemically distinct transition states govern rapid dissociation of single L-selectin bonds under force. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3784–3789.

(150) Friddle, R. W.; Noy, A.; de Yoreo, J. J. Interpreting the widespread nonlinear force spectra of intermolecular bonds. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 13573–13578.

(151) Hummer, G.; Szabo, A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 3658–3661.

(152) Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.

(153) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(154) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(155) Jones, E.; Oliphant, T.; Peterson, P., et al. SciPy: Open source scientific tools for Python., 2001–.

(156) Berquand, A.; Xia, N.; Castner, D. G.; Clare, B. H.; Abbott, N. L.; Dupres, V.; Adriaensen, Y.; Dufrêne, Y. F. Antigen binding forces of single antilysozyme Fv fragments explored by atomic force microscopy. *Langmuir* **2005**, *21*, 5517–5523.

(157) Lata, S.; Reichel, A.; Brock, R.; Tampé, R.; Piehler, J. High-affinity adaptors for switchable recognition of histidine-tagged proteins. *J. Am. Chem. Soc.* **2005**, *127*, 10205–10215.

(158) Neuert, G.; Albrecht, C.; Pamir, E.; Gaub, H. Dynamic force spectroscopy of the digoxigenin–antibody complex. *FEBS Lett.* **2006**, *580*, 505–509.

(159) Yangyuoru, P. M.; Dhakal, S.; Yu, Z.; Koirala, D.; Mwongela, S. M.; Mao, H. Single-molecule measurements of the binding between small molecules and DNA aptamers. *Anal. Chem.* **2012**, *84*, 5298–5303.

(160) Yuan, C.; Chen, A.; Kolb, P.; Moy, V. T. Energy landscape of streptavidin–biotin complexes measured by atomic force microscopy. *Biochemistry* **2000**, *39*, 10219–10223.

(161) Wright Lucas, S.; Harding, M. M. Detection of DNA via an ion channel switch biosensor. *Anal. Biochem.* **2000**, *282*, 70–79.

(162) Samai, S.; Bradley, D. J.; Choi, T. L. Y.; Yan, Y.; Ginger, D. S. Temperature-dependent photoisomerization quantum yields for azobenzene-modified DNA. *J. Phys. Chem. C* **2017**, *121*, 6997–7004.

(163) Kou, B.; Guo, X.; Xiao, S.-J.; Liang, X. Highly efficient room-temperature photoresponsive dna tethering azobenzene through backbone-inserted glycerol via ether bond. *Small* **2013**, *9*, 3939–3943.

(164) Yan, Y.; Samai, S.; Bischoff, K. L.; Zhang, J.; Ginger, D. S. Photocontrolled DNA hybridization stringency with fluorescence detection in heterogeneous assays. *ACS Sens.* **2016**, *1*, 566–571.

(165) Yan, Y.; Chen, J. I. L.; Ginger, D. S. Photoswitchable oligonucleotide-modified gold nanoparticles: Controlling hybridization stringency with photon dose. *Nano Lett.* **2012**, *12*, 2530–2536.