Mathematical Methods for studying DNA and Protein Interactions

By

Sarah LeGresley

Submitted to the Department of Physics and Astronomy and the
Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Committee members

Dr. Christopher Fischer, Chairperson

Dr. Barbara Anthony-Twarog

Dr. Michael Murray

Dr. Jack Shi

Dr. Dean Stetler, Department of Molecular Biosciences

Date defended: July 18, 2017

The Dissertation Committee for Sarah LeGresley certifies
that this is the approved version of the following dissertation :

Mathematical Methods for studying DNA and Protein Interactions

Dr. Christopher Fischer, Chairperson

Date approved: July 28, 2017

Abstract

Deoxyribnucleic Acid (DNA) damage can lead to health related issues such as developmental disorders, aging, and cancer. It has been estimated that damage rates may be as high as 10^5 per cell per day. Because of the devastating effects that DNA damage can have, DNA repair mechanisms are of great interest yet are not completely understood. To gain a better understanding of possible DNA repair mechanisms, my dissertation focused on mathematical methods for understanding the interactions between DNA and proteins. I developed a damaged DNA model to estimate the probabilities of damaged DNA being located at specific positions. Experiments were then performed that suggested that the damaged DNA may be repositioned. These experimental results were consistent with the model's prediction that damaged DNA has preferred locations. To study how proteins might be moving along the DNA, I studied the use of the uniform motion "*n-step*" model. The *n-step* model has been used to determine the kinetics parameters (*e.g.* rates at which a protein moves along the DNA, how much energy is required to move a protein along a specified amount of DNA, *etc*.) of proteins moving along the DNA. Monte Carlo methods were used to simulate proteins moving with different types of non-uniform motion (*e.g.* backward, jumping, *etc.*) along the DNA. Estimates for the kinetics parameters in the *n-step* model were found by fitting of the Monte Carlo simulation data. Analysis indicated that non-uniform motion of the protein may lead to over or underestimation of the kinetic parameters of this *n-step* model.

Acknowledgements

I would like to start by thanking Dr. Chris Fischer. Without his help and guidance, I am not sure if I would have ever gotten finished!

I would like to thank the Physics Department at the University of Kansas. It has been an honor and pleasure to be a part of this department. I especially appreciate the support of the Physics Department Chair Dr. Hume Feldman.

I would like to thank my dissertation committee both past and present.

I would like to thank Dr. Matthew Antonik for his mentorship on the Nucleosome Theory and the Nucleosome Experiments.

I would like to thank all those that helped with the nucleosomes experiments: Dr. Dean Stetler (Department of Molecular Biosciences - University of Kansas), Heather Shinogle (Microscopy and Analytical Imaging Laboratory - University of Kansas) and Dr. Fei Philip Gao and Anne Cooper (Protein Production Group - University of Kansas)

I would like to thank my lab mates Koan Briggs and Dr. Gada Al-Ani. It was always a pleasure to work with you both.

I would like to thank Dr. Wade Rush and Dr. Allen Eastlund. Without Wade I would probably have never come to physics. Without Allen (and his great family), Wade and I would not have enjoyed our time here nearly as much.

Finally, I would like to thank my family and friends for their love and support during this endeavor. It seemed like everyone always had total confidence in me to do this.

To my wonderful family -

especially WDR and VISR

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Damaged Deoxyribonucleic Acid (DNA) not located and repaired quickly enough can lead to health related issues such as developmental disorders [2], aging [3, 4], and cancer [3, 5]. Despite its fundamental importance, the DNA damage response is still not completely understood [6]. To fit within the nucleus of a cell, much of our DNA is compacted, starting with the nucleosome, making it difficult for repair proteins to detect, access, and repair DNA damage. Nucleosome breathing is a process that wraps and unwraps DNA from histone proteins [1]. I have modified a model of this nucleosome breathing process to include damaged DNA [7]. The results of the model suggest that the spontaneous wrapping and unwrapping of the nucleosome might allow damaged DNA to be repositioned to locations where it is more likely to be found by repair proteins in a random search. Using a statistical mechanics approach, I modeled damage to the DNA as a change in the energy required to bend the DNA in the nucleosome structure. The change in the bending energy was incorporated in a partition function which ultimately allows for prediction of the preferred location for the damage. Results from the model indicate that the most likely positioning for damage (lesions) is near the dyad (flexible damage) or in the linker DNA (rigid damage). I speculate that these positions potentially serve as rendezvous points where the damage is more likely to be encountered by repair proteins.

In addition to studying how the DNA may be repositioned in the nucleosome structure, I also

studied the movement of proteins along the DNA. I did not focus on repair proteins specifically, but instead took a broader approach that applies to a variety of proteins that translocate (move) along DNA. The methods used to study the movement of proteins along the DNA typically rely on monitoring the total production of adenosine diphosphate (ADP) (described in Section 2.2), the arrival/departure of the motor at/from a particular location (often one end of the DNA), or the dissociation of the protein from the DNA. The kinetic time course data from monitoring can be analyzed assuming a simple sequential "*n-step*" mechanism, to estimate a variety of kinetic parameters (*e.g.*, translocation rate) [8]. The "*n-step*" mechanism assumes uniform and directionally biased motion of the protein along the DNA and explicitly ignores all other non-uniform motion (backward motion, jumping, random pausing, *etc.*) of the protein. Monte Carlo methods were used to generate kinetic times course data assuming that the motion of the protein is no longer constrained to repetition of rate-limited steps and that other non-uniform motion may occur. A simple sequential "*n-step*" model (which assumes uniform directionally biased motion) was then used to analyze the simulated kinetic time courses. Using non linear least square (NLLS) methods, the best fit of the "*n-step*" model to the simulated time course data produced estimates of the kinetic parameters. I was then able to determine which of the kinetic parameters were likely to be over/under estimated due to non-uniform motion of the protein.

Chapter 2

Biological background

My dissertation focuses on the mathematical modeling of both DNA and proteins. The damaged DNA model (Chapter 3) predicts the probability of damage being located at specific locations in equilibrium. The "*n-step*" model (Chapter 5) is used to determine the kinetic parameters of the motion of a protein moving along a template (*e.g.*, DNA). In order for the reader to have a better understanding of the DNA and protein systems discussed in this dissertation, I have provided some relevant biological background information in this chapter.

2.1 DNA, nucleosomes, and chromatin

DNA is often referred to as the instruction manual for our cells. This is because DNA contains the 'codes' for the creation, structure, and maintenance of cells. The particular sequences of the subunits of DNA (called nucleotides) form the 'codes'. There are four types of nucleotides (also called bases): adenine (A), thymine (T), cytosine (C), and guanine (G) which each include deoxyribose (sugar) and a phosphate group. Basepairs (bp) are the hydrogen bonded complementary bases A with T or C with G (Figure 2.1). The basepairs in the double stranded DNA (dsDNA) are covalently bonded (Figure 2.1). Single stranded DNA (ssDNA) has the covalently bonded bases but not the hydrogen bonded complementary bases.

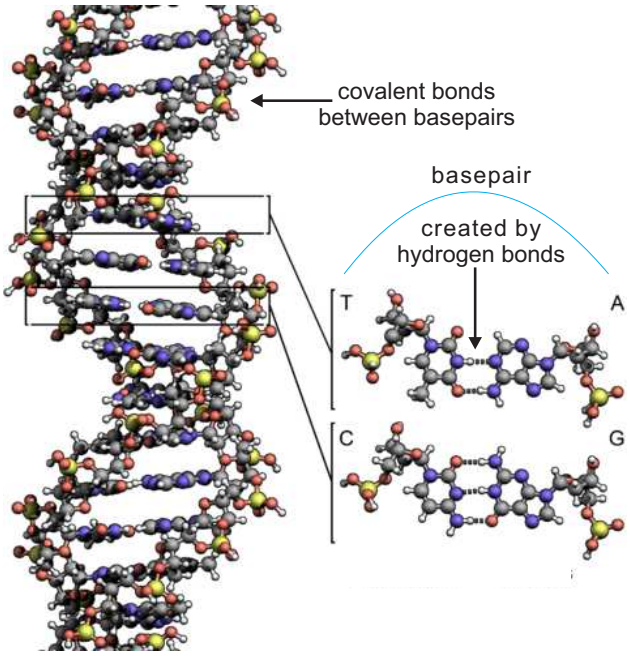


Figure 2.1: Structure of double stranded DNA. Double stranded DNA (dsDNA) is comprised of nucleotides (bases) adenine (A), thymine (T), cytosine (C), and guanine (G). Basepairs are the hydrogen bonded complementary bases of A with T or C with G. Double stranded DNA contains basepairs with covalent bonds between the individual basepairs. Single stranded DNA (ssDNA) (not pictured here) would have the covalent bonds between bases but not the hydrogen bonds to the complementary bases. Image by Zephyris (Own work) [CC BY-SA 3.0 or GFDL].

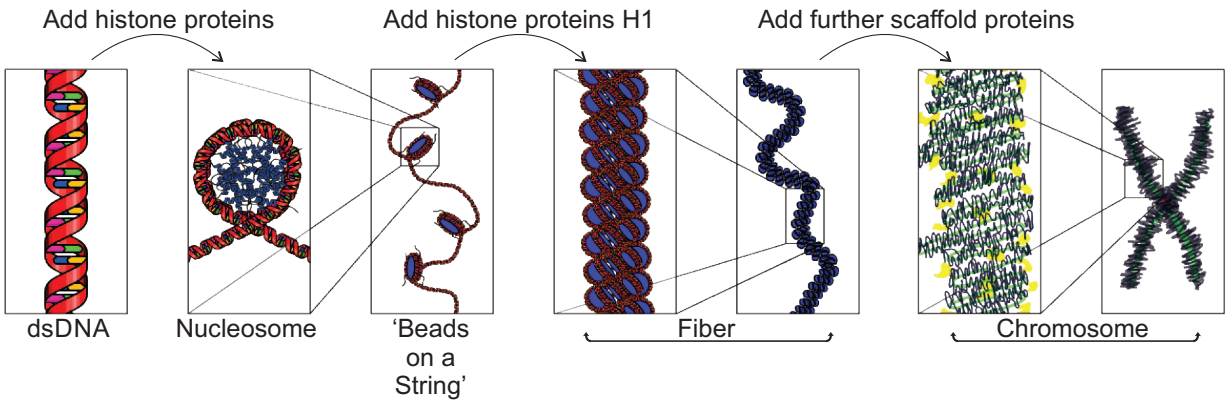


Figure 2.2: Chromatin structure. Approximately 147 basepairs of double stranded DNA wrap 1.7 times around the histone octamer to create the nucleosome structure. The nucleosomes are separated by 10 - 90 bp of linker DNA. Continued folding and packing of the nucleosomes leads to the chromatin structure. Image by Richard Wheeler at English language Wikipedia [CC-BY-SA-3.0 or GFDL], from Wikimedia Commons.

2.1.1 Chromatin structure

Approximately 3 billion basepairs of DNA (which when stretched from end to end would be $\approx$ 2 meters in length) must fit in the cell nucleus which has a diameter of $\approx 6\mu$m [9]. For the DNA to not only fit in the nucleus but be accessible (for reading, copying, and repair), the DNA is packaged and folded into the chromatin structure (Figure 2.2). The most basic level of the chromatin is dsDNA. The first level of packing is when 147 basepairs of dsDNA wrap 1.7 times around a histone octamer (group of 8 histone proteins) to create the nucleosome structure [10–12]. The next level of folding creates "beads on a string" as shown in Figure 2.3 where nuclesomes are closely packed with consecutive nucleosomes separated by linker DNA about 10-90 bp in length [13]. The nucleosomes undergo more packing and folding to eventually become the chromatin structure.

Because of the compactness of the DNA, damage (lesions) can be difficult for repair proteins to find, access, and repair. There is a wide variety of damages that occurs with some examples shown in Figure 2.4. One of the most severe is the double stranded break that separates the dsDNA into two pieces as in Figure 2.4(a). The fact that the DNA is double stranded means that two copies

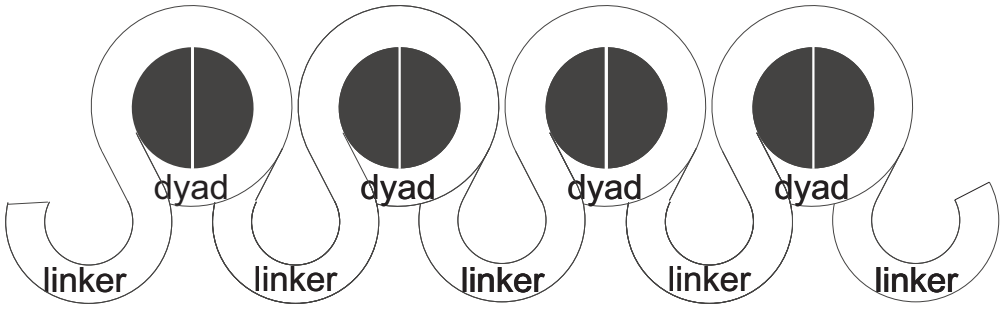


Figure 2.3: Cartoon image of 'beads on a string'. Approximately 147 basepairs of double stranded DNA wrap 1.7 times around the histone octamer (black circle) to create the nucleosome structure. The dyad (axis shown as a white line through histone octamer) is located at base pair 73 and is the center of the wrapped DNA. The linker DNA connecting the nucleosomes is 10 - 90 bp in length.

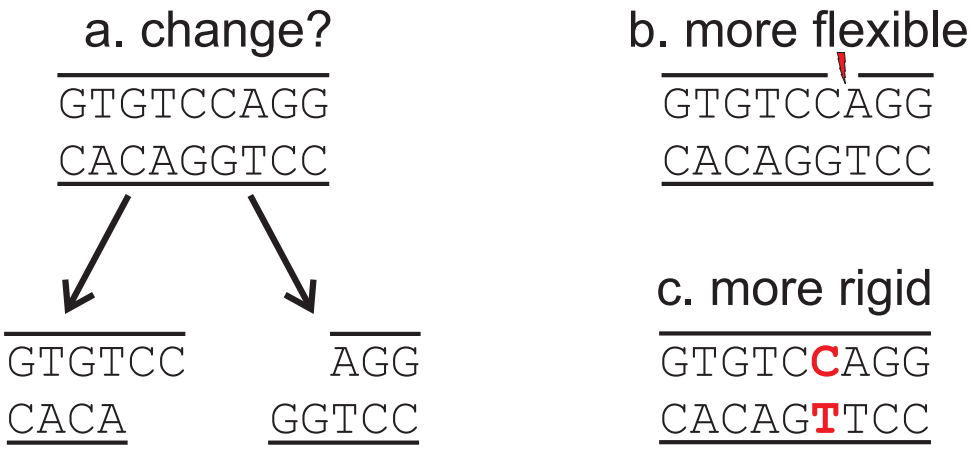


Figure 2.4: Cartoon image of DNA damage. In (a) the doubled stranded break separates the DNA. In (b) the 'nick' separates the phosphate backbone of the DNA leaving the single strand still connected and more flexible. In (c) there is a basepair mismatch with T matched with C and increased rigidity.

exist. However, when a double stranded break occurs, for example in the linker (Figure 2.4(a)), it will be difficult if not impossible, to repair. However, other types of damage, such as "nicks" that break the phosphate backbone of the DNA (Figure 2.4(b)) are much less severe. The final type of damage shown (Figure 2.4(c)) is when a basepair is not matched properly (T is shown mismatched with C).

2.1.2 Nucleosome structure

The nucleosome (Figures 2.3 and 2.5) is the focus of Chapters 3 and 4. Of particular interest will be the DNA wrapped in the nucleosome (Figure 2.5) including the dyad, which is the center of the wrapped DNA, and the linker DNA that connects the nucleosomes (Figure 2.3). The locations of the DNA wrapped in the nucleosome can be described using super helical location (SHL) or twists in the double helical structure of the DNA. The SHL location is the number of turns along the DNA helix away from central base pair at the dyad [11]. In the SHL coordinate system shown in Figure 2.6, the dyad is the center of the wrapped DNA and is located at SHL 0. The nucleosome has 14 binding contact sites where the DNA is "bound" to the histone octamer. The binding sites are located at SHL -6.5, -5.5, ... , +5.5, +6.5. The binding contact sites separate the DNA into 13 segments of wrapped and bent DNA. The 13 segments of ($\approx$10.2 bp in length) wrapped and bent DNA are centered at SHL -6.0, -5.0, ... , +5.0, +6.0 (see Figure 2.6(a) and (b)). That leaves $\approx$7 bp of wrapped DNA located near SHL $\pm$7.0 but structural studies have indicated that the last 10 bp on each end remain unbent [11]. The wrapped segments at SHL $\pm$7.0 being unbent means those segments store no elastic bending energy and thus are not included as a part of my damaged DNA model (Chapter 3).

2.2 Motor proteins

DNA motor proteins are the focus of Chapter 5. Motor proteins move along double or single stranded DNA using the chemical energy obtained from the hydrolysis of Adenosine triphosphate (ATP) (Figure 2.7). When the high energy bond between the three phosphate groups of ATP is broken, adenosine diphosphate (ADP) containing two phosphates and a separate inorganic phosphate P_i are created. The energy released during this reaction can be used by a motor protein to translocate (move) along DNA. Chromatin remodelers are an example of a motor protein that uses the energy from ATP hydrolysis to move and restructure the nucleosome [14]. By restructuring the nucleosome (and chromatin) difference sequences of DNA (potentially damaged DNA) can be

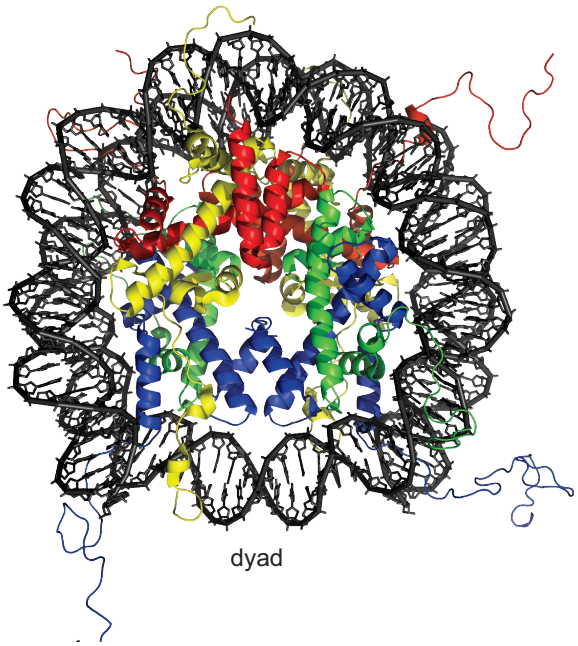


Figure 2.5: Nucleosome structure. The nucleosome structure consists of 147 basepairs of DNA (shown in black) wrapped around a histone octamer. The histone octamer consists of 8 histone proteins (shown as colored ribbons). The DNA is wrapped around the histone octamer 1.7 times. The dyad is the center of the DNA that is wrapped in the nucleosome structure. Note that the dyad has only a single wrap of DNA. Image by Zephyris at the English language Wikipedia [CC-BY-SA-3.0 or GFDL], via Wikimedia Commons.

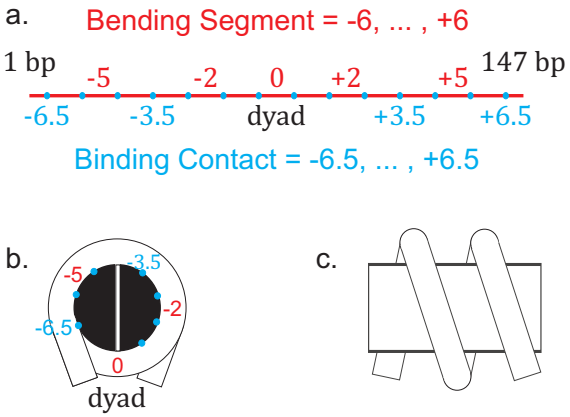


Figure 2.6: Cartoon image of nucleosome using SHL coordinate system. Flattened DNA (a) shown with binding sites (blue) at SHL coordinates -6.5 to +6.5 and bending segments (red) at SHL coordinates -6.0 to +6.0. The bent DNA (b) is wrapped at SHL 0 at the dyad (dyad axis is the white line) and continues to SHL ± 6. The negative SHL are symmetrically located and not shown. A side view of the nucleosome is shown in (c).

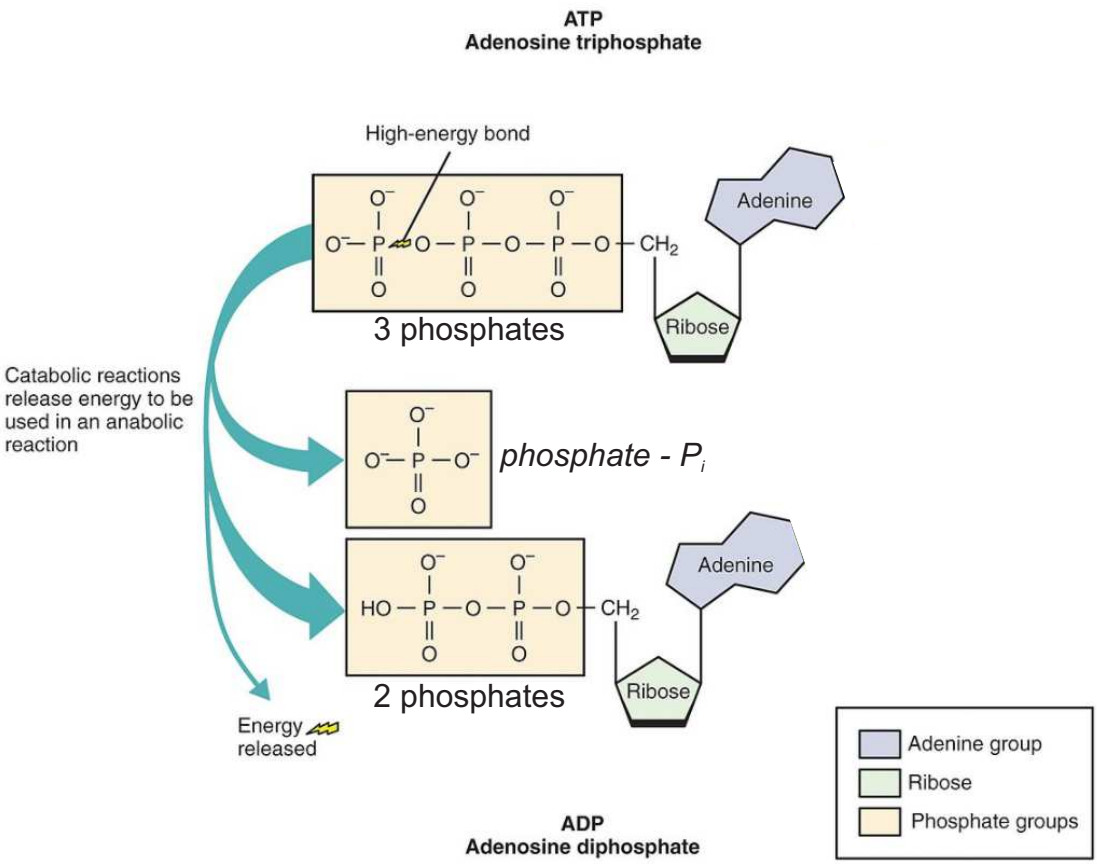


Figure 2.7: ATP hydrolysis. At the top is adenosine triphosphate (ATP) which contains 3 phosphate groups. The bond between the phosphates can be broken, which separates ATP into adenosine diphosphate (ADP), which contains two phosphate groups, and the now separated inorganic phosphate P_i. During this process energy is released. Image by OpenStax College [CC BY 3.0], via Wikimedia Commons.

accessed by repair proteins. Therefore, chromatin remodelers likely play a role in the DNA repair process [15].

When motor proteins bind to DNA they can do so with specific polarity and move along the DNA with directionality. DNA has an asymmetry in that the ends of the single stranded DNA have a different chemical structure. The 5' end of the DNA has an extra phosphate awhile the 3' end does not (see Figure 2.8). Having bound to the DNA with polarity, the proteins then move with directionally biased motion along the DNA. Shown in Figure 2.9, is *Taq* polymerase (a protein used to copy DNA using a technique called polymerase chain reaction (PCR) and described in Section 4.2.1) that binds to single stranded DNA in such a way that it then moves towards the 3'

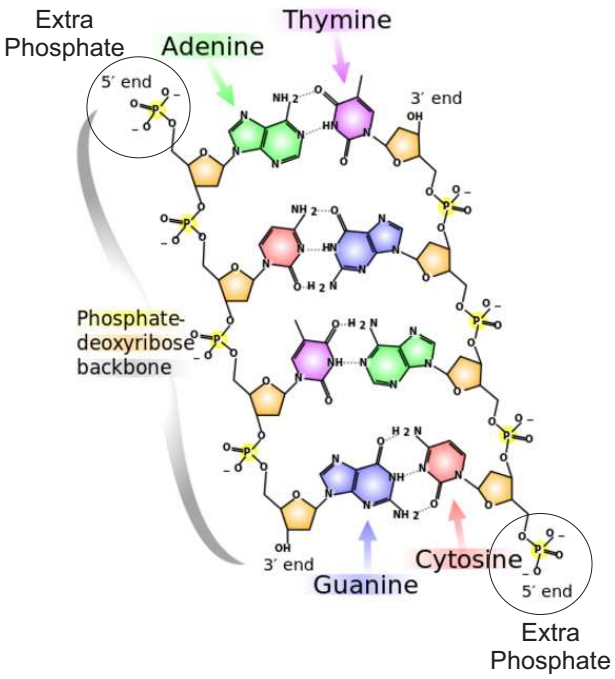


Figure 2.8: Asymmetry in DNA. Each strand of DNA has a 3' and 5' end. The 5' end has an extra phosphate group (circled) not present on the 3' end. Image by Madeleine Price Ball [CC-BY-SA-3.0 or GFDL], via Wikimedia Commons.

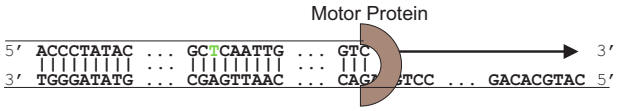


Figure 2.9: Motion of motor protein. A motor protein binds with polarity to the DNA. The protein then translocates along the DNA using the energy from ATP hydrolysis toward a particular end of DNA. This particular protein binds with polarity such that it moves towards the 3' end.

end. Each particular type of motor protein from restriction enzymes that cleave DNA to unwinding proteins that separate double stranded DNA (dsDNA) into single stranded DNA (ssDNA) may have a preferred direction and potentially even a specific DNA sequence to which it binds.

2.3 Experiments

In order to explore the structure of the nucleosome and the motion of motor proteins both single molecule experiments and ensemble experiments can be used. The main difference between the two types of experiments is that single molecule experiments are typically done at much lower concentrations so that each individual molecule (*e.g.*, a single nucleosome) can be analyzed. However,

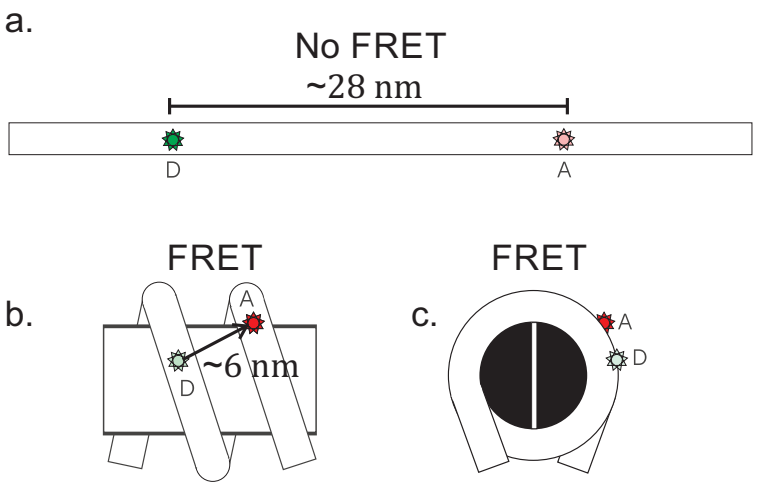


Figure 2.10: FRET with nucleosomes. In part (a) the donor (green) D and acceptor (red) A fluorophores are separated by a distance of ≈ 28 nm so FRET does not occur. In this conformation, the donor (green) is bright and there should be no energy transfer to the acceptor (red). Part (b) is a side view of the DNA wrapped in the nucleosome with the dyes located ≈ 6 nm apart. The donor (green) should be dim due to quenching and there should be a very bright acceptor (red) due to the energy transfer from the donor (green) to the acceptor (red). Part (c) shows the nucleosome rotated 90° from part (b).

ensemble experiments are sometimes easier to develop and perform.

Fluorescence resonant energy transfer (FRET) experiments (Section 4.3) can be done at the single molecule level. One type of single molecule experiment can be used to determine the distances between two fluorescently labeled fluorophores (dyes). In these experiments, only the donor fluorophore is initially excited, possibly using a laser. When the donor and an acceptor fluorophore are in close proximity ($\approx$ 1 - 10 nm), the excited donor fluorophore can transfer energy (through a virtual photon) to the nearby acceptor fluorophore. The efficiency at which this energy transfer occurs is related to the distance by the following equation

$$\text{FRET Efficiency} = \frac{1}{1+(r/R_0)^6} \tag{2.1}$$

where r is the distance between the donor and the acceptor dyes and R_0 = Förster radius is the distance at which the the donor and the acceptor have 50% FRET efficiency [16]. Note that different pairs of fluorophores may have different R_0 values. Higher FRET efficiency occurs when the

fluorophores are closer together. Lower FRET efficiency occurs when the fluorophores are farther apart. The FRET efficiency can be approximated by

$$\text{FRET Efficiency} = I_A/(I_A + I_D) \tag{2.2}$$

where I_D is the fluorescence intensity of the donor and I_A is the fluorescence intensity of the acceptor [17]. The intensity can be determined as described in Section 4.3. Figure 2.10(a) should have a very low (zero) FRET efficiency because the fluorophores are far apart. For fluorophores located more than about 10 nm apart, the green will have a high intensity but the red will have a low intensity ($I_A \approx 0$). In Figure 2.10(b) and (c), there will be energy transfer (FRET) and the acceptor (red) intensity will be greater than zero. The FRET efficiency will depend on the distance and will be higher with fluorophores close together and lower with fluorophores farther apart. Though the FRET efficiency is easily understood in terms of Equations 2.1 and 2.2, the actual calculation of the FRET efficiency will depend on the experimental setup.

Ensemble experiments are typically performed at higher concentrations which usually means the average behavior of the system can be estimated. Electrophoresis is one type of ensemble experiment. During electrophoresis an electric field is applied to a sample. In the electrophoresis experiments described in Chapter 4, DNA and/or nucleosomes samples were loaded into an agarose or polyacrylamide gel. The samples are initially loaded into a gel nearest the negatively charged end of the electric field. The DNA will then move through the gel towards the other end of the gel which is near the positively charged end (Figure 2.11). Shorter DNA moves faster and easier through the gel than longer DNA. On each side of the sample (left and right of Figure 2.11) is a control sample (called a ladder) containing known lengths (in basepairs) and concentrations of DNA. After running a gel, the gel is then removed from the electric field and placed in a solution that contains an intercalating dye. The dye is small enough to move in between the bases of the DNA in a process referred to as 'staining.' After the staining, the gel is then imaged by using a light source (*e.g.*, a laser) that excites the intercalating dyes. By comparing the image results of

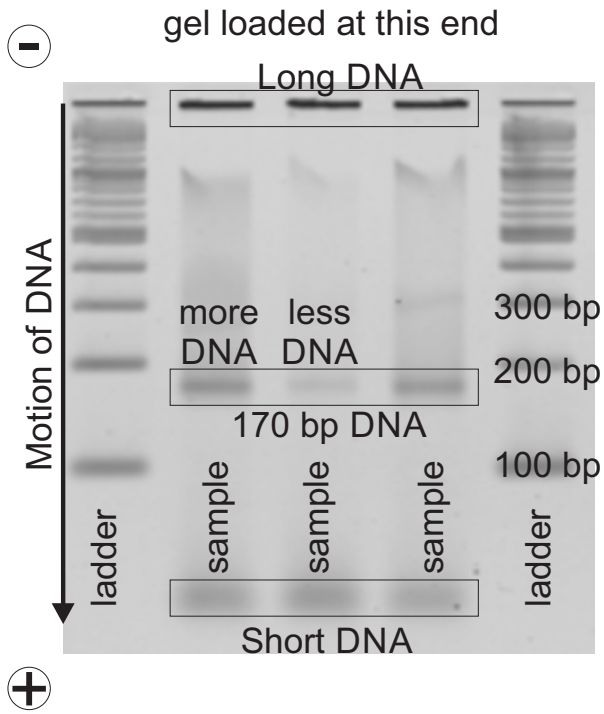


Figure 2.11: Gel experiment. DNA samples and control ladder(s) were loaded into a agarose (or polyacrylamide) gel near the top of this image. The DNA is initially located near the negatively charged end (top of image). Shorter DNA is pushed farther and faster towards the positively charged end (bottom of image) than the longer DNA. The control ladders on the left and right had DNA of 100, 200, 300, ... basepairs in length. In the middle lanes is a sample that contains long DNA (top), 170 bp DNA (middle), and short DNA (bottom). Gel image taken in July 2013 in the Antonik Lab at the University of Kansas.

the known control ladder to the potentially unknown sample, the lengths and concentrations of the unknown sample can be estimated. In the gel image shown in Figure 2.11, the DNA in the very middle sample had a lower concentration than the two adjacent (left and right) samples since the volumes inserted into the gel were the same. This difference in concentration is evident since the band (boxed region of Figure 2.11) in the center is lighter in color (and therefore contains less DNA) than the two adjacent (left and right) bands. Comparison of the samples in the center three lanes to the ladders in the left and right lanes of Figure 2.11 shows the samples contain very long DNA (top of image) and very short DNA (bottom of image). The DNA in the center of the gel is closer in length to 200 basepairs than 100 basepairs (the sample contains 170 bp DNA).

Chapter 3

Nucleosome breathing theory

3.1 Introduction

The health and integrity of the cell can be seriously threatened by DNA damage [5, 18–20]. In order for the cell to remain viable, there must be mechanisms that effectively and effeciently repair DNA damage. Breakdowns of these mechanisms have been linked to developmental disorders [2], aging [3, 4], and cancer [3, 5]. Exposure to ionizing radiation can result in DNA damage (lesions) which can be located anywhere in the DNA [19], leaving the repair mechanism to search a millionfold excess of DNA to find and repair a single lesion. In order to locate the damage, DNA repair proteins search for specific types of DNA lesions [21, 22]. How the repair proteins locate the damage is not completely understood [6]. However, it is widely believed that repair mechanisms work in what is refered to as 'facilitated diffusion' [23–25] where the proteins undergo a cyclical process of searching the DNA locally (1D diffusion) before performing a long range hop or jump to a different location (3D diffusion). This process of switching between 1D and 3D diffusion has been observed experimentally [21, 26]. Not surpisingly, models have shown increased search speeds when the DNA is restricted to a confined space [27] or when proteins "hop" to different locations along the DNA to avoid obstacles [28].

What is not often considered is how little of the DNA may be accessible to binding proteins. With ≈ 2 meters of DNA compacting up into a chromatin structure that fits into the cell nucleus

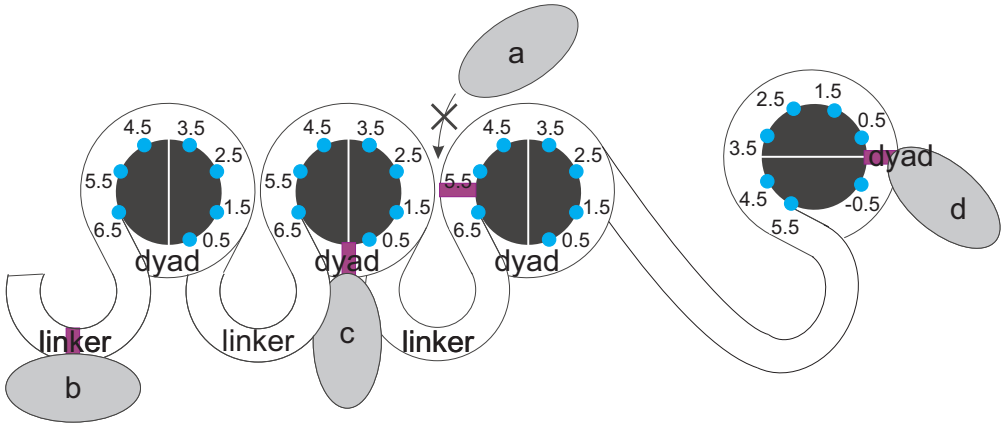


Figure 3.1: Cartoon of the nucleosome organization. 147 bp of DNA wrap ≈1.7 times around a histone core (black circle) and is centered around the dyad. The dyad axis is shown as a white line on the diameter of the nucleosome. Once in the nucleosome structure further organization creates a 'beads on a string' structure. Protein binding sites on the DNA are marked as a purple stripe. The site is inaccessible to the protein (a) due to steric hindrances. Binding sites located in the linker (b), the dyad (c and d), or in unwrapped portions are potentially accessible to proteins.

$\approx 6\mu$m in diameter [9], the chromatin structure will have a large effect on the DNA damage repair process [29–32]. As shown in Figure 3.1, the tight packaging of DNA in the nucleosomes leaves much of the DNA inaccessible to proteins (*e.g.* any proteins that only bind to free DNA) due to steric hindrances. However, experiments have shown that proteins are able to overcome this obstacle due to a process referred to as "breathing" shown in Figure 3.2. During this breathing process, DNA unwraps from the histone octamer core [33–37] allowing proteins to potentially bind to the now accessible DNA. The DNA unwrapping starts near the linkers so the sites that are located near the linker are easily made accessible while sites towards the center of the nuclesome (dyad region) are much less likely to become accessible [38]. DNA sequence [39–44], DNA modifications [45–47], and histone modifications [48–51] all have effects on the specific positioning of the DNA on the histone octamer.

Not only does positioning of the DNA around the octamer factor into DNA repair [52, 53], but DNA damage also impacts nucleosome positioning [29, 32, 54–57]. Published in Physical Review E [7] and discussed here are my theoretical models of how damaged DNA impacts the octamer positioning and how "breathing" may play a role in reorganizing the nucleosome structure so that damage self organizes to locations more easily found by repair proteins. Results of modeling that

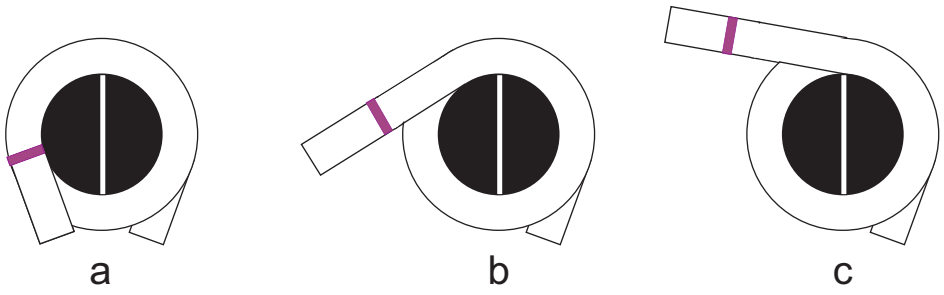


Figure 3.2: Nucleosomal breathing of DNA. A nucleosome that is fulled wrapped is shown in (a). Two partially unwrapped states, (b) and (c), are also shown. DNA binding sites or damage (shown in purple) will become accessible as the nucleosome unwraps and the DNA separates from the histone octamer.

included damaged DNA in the nucleosome breathing process suggest that damaged nucleosomal DNA will thermodynamically reposition itself to either the central axis furthest from the linkers DNA (the dyad) or be expelled out into the linker DNA. Some DNA binding proteins prefer to bind to nucleosomal DNA at the dyad [58] while others prefer free DNA not in the nucleosome. For many binding proteins much of the DNA is not accessible due to the DNA being wrapped in the nucleosome structure. However, if the damage is repositioned to the dyad or the linker DNA, the binding proteins can not only access the DNA, but the detection of the DNA damage should be faster since the search space is greatly reduced. In order to test this hypothesis, I modified an existing statistical mechanical model that determined the probability of healthy nucleosomal DNA being accessible [59], to include damaged DNA. In the modified model, different energies associated with the damage at a variety of elastic bending energies and adsorption binding energies of the DNA were considered.

3.2 Theory

Nucleosomes undergo a "breathing" process where DNA spontaneously unwraps and rewraps from the histone octamer (Figure 3.2). Experimentally, this breathing process has been demonstrated through the use of restriction enzymes [35]. Restriction enzymes typically unable to attach to DNA wrapped in the nucleosomes structure were nevertheless observed to cleave nucleosomal

DNA. This led researchers to believe that "breathing" or unwrapping of the DNA from the histone octamer must occur.

3.2.1 Partition function of healthy DNA

A previously published approach for modeling the unwrapping of the nucleosome structure used a partition function that considered all the possible nucleosome configurations [59]. Shown in Figure 3.3 is a cartoon image of the DNA where the pink region denotes the basepairs that are wrapped in the nucleosome structure. Figure 3.3(a) shows the nucleosome is wrapped from the left at base pair x_L to the right at base pair x_R where $1 \leq x_L \leq x_R \leq 147$. The number of base pairs of DNA that are wrapped is then $x_R - x_L + 1$. The energy required to wrap a single base pair of DNA is defined as the unitless parameter q and is measured in multiples of k_BT. This results in a total complexation energy for the wrapped portion of $-qk_BT(x_R - x_L + 1)$. The partition function sums over all possible nucleosome wrapping configurations. The wrapping configurations could be anything from a single basepair (*e.g.* $x_R = x_L$ which has 147 configurations and would require only qk_BT energy to wrap) to the whole 147 basepairs wrapped (which has only one configuration and would require 147 qk_BT to wrap). Summing over all these configurations with their associated wrapping energies of $-qk_BT(x_R - x_L + 1)$, the partition function is

$$Z = \sum_{x_L=1}^{147} \sum_{x_R=x_L}^{147} e^{-[-q(x_R - x_L + 1)]}$$

which reduces to

$$Z = \sum_{x_L=1}^{147} \sum_{x_R=x_L}^{147} e^{q(x_R - x_L + 1)} \tag{3.1}$$

where $1 \leq x_L \leq x_R \leq 147$.

The probability that a particular target DNA base pair x_B can be accessed by an enzyme is determined by the location of x_B relative to the (pink) wrapped portion of the nucleosome (Figure 3.3(b)-(d)). For the target x_B to be accessible, it is assumed to be located at least δ base pairs outside the wrapping portion, where δ will be determined by comparing the model to experimental data.

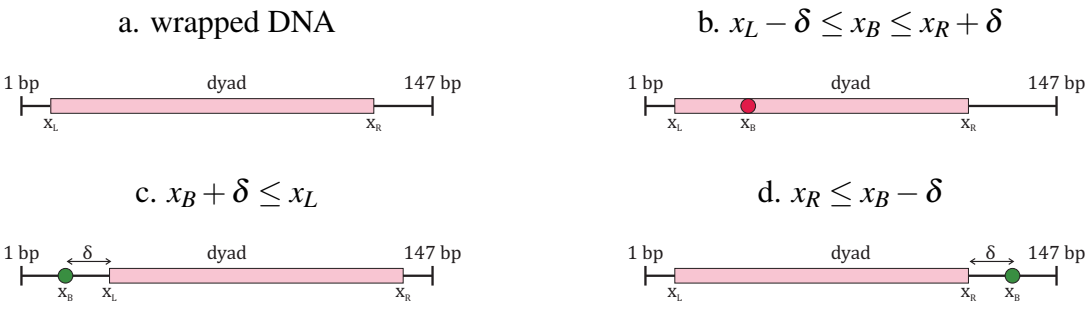


Figure 3.3: Wrapping configurations of the nucleosome. In each figure the pink portion is to denote the basepairs that are wrapped in nucleosome structure. Part (a) shows a nucleosome that is partially wrapped from x_L to x_R where $1 \leq x_L \leq x_R \leq 147$. The partition function is found by summing over all possible nucleosome configurations. Part (b) shows a nucleosome that is wrapped from x_L to x_R where $x_L \leq x_B \leq x_R$. In this case, the target base x_B is not accessible since it is located in the wrapped portion. Parts (c) and (d) show configurations of nucleosomes where the target base x_B is accessible. Part (c) corresponds to the left summation in Equation 3.2 and (d) corresponds to the right summation in Equation 3.2

Shown in Figure 3.3(b) is an example of the target site x_B located within the wrapped portion of the nucleosome and therefore not accessible. In Figure 3.3(c), the target x_B is accessible because it is at least δ basepairs to the left of the wrapped portion with $x_B + \delta \leq x_L$ and in Figure 3.3(d) the target x_B is accessible because it is at least δ basepairs to the right of the wrapped portion with $x_R \leq x_B - \delta$. Summing over all configurations where the target site x_B is located at least δ basepairs outside the wrapped portion then dividing by the partition function gives the probability that a target base pair x_B is accessible. This can be expressed as

$$p_{accessible}(x_B) = \frac{1}{Z}\left[\sum_{x_L=x_B+\delta+1}^{147}\sum_{x_R=x_L}^{147} e^{q(x_R-x_L+1)} + \sum_{x_L=1}^{x_B-\delta}\sum_{x_R=x_L}^{x_B-\delta} e^{q(x_R-x_L+1)}\right] \tag{3.2}$$

where Z is the partition function in Equation 3.1 and the left and right terms are the configurations shown in Figure 3.3(c) and Figure 3.3(d), respectively. As an example, the following equation is used to determine the probability that the target site basepair $x_B = 73$ (located at the dyad) is accessible.

$$p_{accessible}(73) = \frac{1}{Z}\left[\sum_{x_L=73+\delta+1}^{147}\sum_{x_R=x_L}^{147} e^{q(x_R-x_L+1)} + \sum_{x_L=1}^{73-\delta}\sum_{x_R=x_L}^{73-\delta} e^{q(x_R-x_L+1)}\right] \tag{3.3}$$

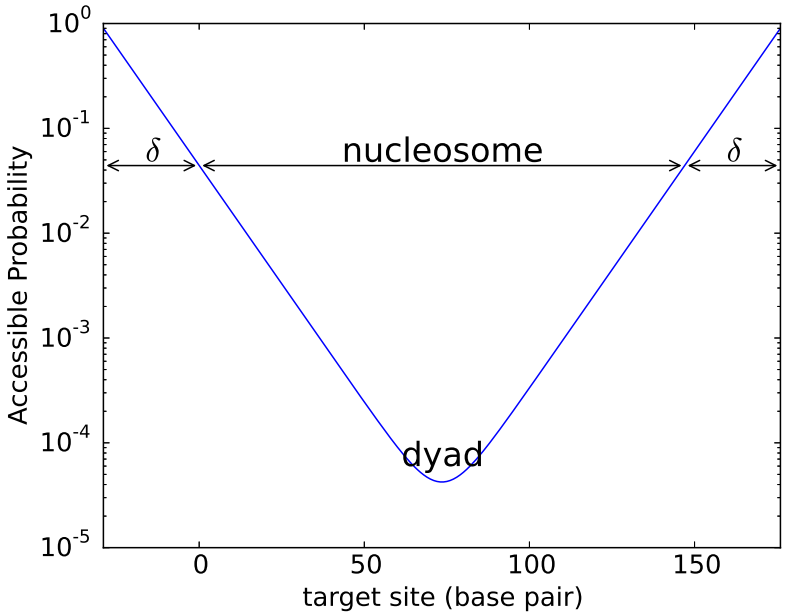


Figure 3.4: Probabilities of target DNA sites being accessible. The nucleosome structure can wrap a maximum of 147 base pairs with the center of the wrapping occurring at base pair 73 which is referred to as the dyad. As the target is located farther from the dyad (towards the left and right edges), the probability of being accessible to a binding protein increases.

Experimental data from nucleosome experiments with restriction enzymes [35] was used to determined the rate at which the restriction enzymes could cleave (cut) DNA in the nucleosome. From those cutting rates, the probability of a DNA target site in the nucleosome being accessible to the restriction enzyme was determined [59]. A reasonable fit to the experimental data was achieved using the parameters $q = 0.104 \pm 0.016$ and $\delta = 30 \pm 12$ bp [59]. Substituting these parameters into Equation 3.2, Figure 3.4 shows that target sites located further from the center base pair of the nucleosome (dyad) have higher probabilities of being accessible. This is likely due to the repeated unwrapping and rewrapping of the DNA around the histone octamer. The target sites near the edges of the wrapping (left and right sides of Figure 3.4) will likely unwrap and rewrap several times before the unwrapping is significant enough to unwrap the DNA all the way at the dyad (half the DNA would be unwrapped). This partial unwrapping results in higher probabilities of accessibility of DNA near the edges of the nucleosome.

3.2.2 Partition function of nicked DNA

In the healthy DNA model described above, q is defined as the net interaction energy per base pair. Experiments have shown large energy barriers must be overcome to unwrap 10 bp segments [1]. Instead of using the net interaction energy per basepair, I used the net interaction energy per ($\approx$10 bp) segment. The net interaction energy can then be separated into the adsorption or binding energy per binding contact ($q_{ad}, negative$) and the bending of elastic energy per wrapped segment ($q_{el}, positive$). Modifying q from above, the net interaction energy per segment is defined as $q_{net} = q_{ad} + q_{el}$. The super helical location (SHL), which is the number of turns along the DNA helix away from central base pair at the dyad [11], is used to denote the locations of the adsorption contact sites and elastic wrapped segments. In the SHL coordinate system (Figure 3.5), the DNA binds to the histone octamer at 14 (binding) contact sites. This results in 147 basepairs of DNA being broken into 15 segments of wrapped DNA separated by the binding sites. The 14 binding contacts are located at SHL -6.5, -5.5, ... , +5.5, +6.5. Since there is no reason to assume that the energies associated with the binding sites are uniform, the adsorption energy per binding site can be defined as $q_{ad,n}$ where n = -6.5, -5.5, ... , +5.5, +6.5. The centers of the bent wrapping segments are located at SHL position -7, -6, ... , +6, +7. However, each helical turn of the DNA has 10.2 bp which means 133 bp of DNA is located between the first and last binding contact sites. That leaves 7 bp of DNA extending beyond the binding site and structural studies have indicated that the last 10 bp on each end remain unbent [11]. Therefore, I assumed 13 wrapped segments that have elastic bending energy and are located at SHL -6, -5, ... , +5, +6. The elastic energy per wrap is not necessarily uniform and is defined as $q_{el,m}$ where m = -6, -5, ... , +5, +6. Since I was not interested in dissociation of the DNA from the histone octamer, it was assumed there was a minimum of one binding contact site which would mean no bending of the DNA. The summing over all healthy DNA wrapping configurations, the partition function in the SHL coordinate system is

$$Z_{healthy} = \sum_{x_L=-6.5}^{+6.5} \sum_{x_R=x_L}^{+6.5} exp\left[-\left(\sum_{n=x_L}^{x_R} q_{ad,n} + \sum_{m=x_L+1/2}^{x_R-1/2} q_{el,m}\right)\right] \tag{3.4}$$

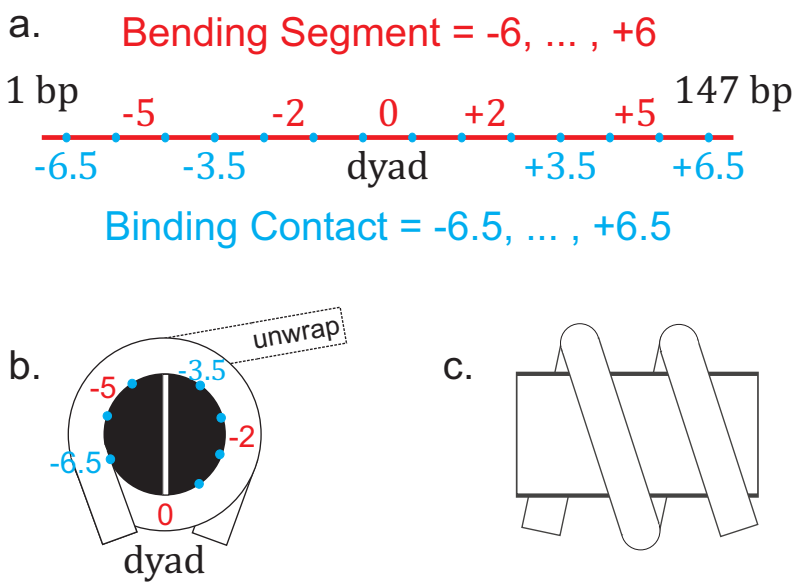


Figure 3.5: Cartoon image of nucleosome using SHL coordinate system. Flattened DNA (a) shown with binding sites (blue) at SHL coordinates -6.5 to +6.5 and wrapping segments (red) at SHL coordinates -6.0 to +6.0. The 147 basepairs of DNA are wrapped from -6 to +6 with the center of the wrapping (basepair 73 or the dyad) located at 0. In (b) the negative SHL are symmetrically located and not shown. A side view of the nucleosome is shown in (c).

where x_L and x_R are SHL -6.5, -5.5, ... , +5.5, +6.5. The left sum in the parentheses in Equation 3.4 is

$$\sum_{n=x_L}^{x_R} q_{ad,n} \tag{3.5}$$

and sums over the adsorption binding contact energies located at SHL -6.5, -5.5, ... , +5.5, +6.5. The right sum in the parentheses in Equation 3.4 is

$$\sum_{m=x_L+1/2}^{x_R-1/2} q_{el,m} \tag{3.6}$$

and sums over the elastic bending energies located at SHL -6.0, -5.0, ... , +5.0, +6.0. The outer summations in Equation 3.4 are

$$\sum_{x_L=-6.5}^{+6.5} \sum_{x_R=x_L}^{+6.5}$$

and is to sum over all possible partially and the one fully wrapped conformation. For example, one conformation is the nucleosome wrapped for x_L = -5.5 to x_R = -3.5 with associated adsorption

energies $q_{ad,-5.5} + q_{ad,-4.5} + q_{ad,-3.5}$ and elastic bending energies $q_{el,-5.0} + q_{el,-4.0}$. Having assumed the DNA does not dissociate from the histone octamer, the summation over the adsorption energies (Equation 3.5) will always have one more term than the summation over the elastic energies (Equation 3.6). When $x_L = x_R$ in the second sum (Equation 3.6), there are no elastic energy terms. In the case that the DNA is only bound to the histone octamer at one site, the DNA is not bent around the histone octamer and corresponds to no elastic (bending) energy.

The location of the DNA damage is defined to be at SHL m_d. Figures 3.6 and 3.7 show an example of damage located at m_d = +5. The damage I considered (Figure 2.4(b) and (c)) is assumed to change the flexibility of the DNA for that segment, so the elastic energy for that wrapping segment is defined to be $q_{el,m_d} = q_{el,m} + \Delta q_{m_d}$. For damage that increases the flexible of the damaged segment m_d, Δq_{m_d} would be negative since the damaged DNA requires less energy to be bent. For damage that increases the rigidity of the damaged segment m_d, Δq_{m_d} would be positive since the damaged DNA requires more energy to be bent. Assuming that the damage is located at fixed SHL m_d, the partition function for nucleosome breathing model is

$$Z_{fix}(m_d) = \sum_{x_L=-6.5}^{+6.5} \sum_{x_R=x_L}^{+6.5} exp\left[-\left(\sum_{n=x_L}^{x_R} q_{ad,n} + \sum_{m=x_L+1/2}^{x_R-1/2} q_{el,m} \right) \right] \tag{3.7}$$

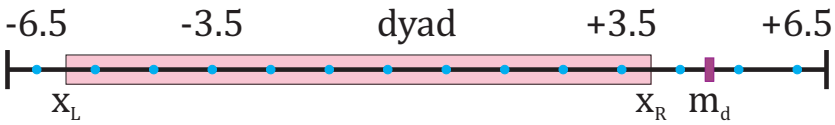
where

$$q_{el,m} = \begin{cases} q_{el,m} & \text{if } m \neq m_d \\ q_{el,m_d} = q_{el,m} + \Delta q_{m_d} & \text{if } m = m_d \end{cases} \tag{3.8}$$

This partition function in Equation 3.7 is the same as in Equation 3.4 for the healthy DNA but with the added constraint of Equation 3.8 to account for the damage.

The partition function in Equation 3.7 is for a static structure where the damage remains fixed at wrapping segment m_d relative to the dyad. In order to include all possible positions of the damaged segment, m_d must be allowed to range from -6, -5, ... , +5, +6. Therefore, the new partition function must use Equation 3.7 but also include a sum over all possible damaged locations. Note that for each iteration of the sum over m_d the elastic energy associated with the damage $q_{el,m}$ where $m = m_d$

a. damage is located outside the wrapped portion



b. damage is located within the wrapped portion

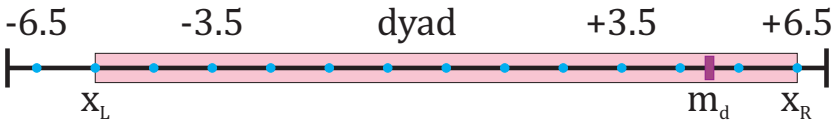


Figure 3.6: Damaged DNA location. The damaged segment m_d located at SHL +5 is between binding contacts (shown in blue) at SHL +4.5 and +5.5. Because of the "breathing" process, the damage may be located outside the wrapped portion (a) or within the wrapped portion (b). When the damage is located in the wrapped portion, the elastic bending energy associated with the damaged segment will be modified.

must be updated. This partition function that allows for the damage to be located at any position is

$$Z_{float} = \sum_{m_d=-6}^{+6} \sum_{x_L=-6.5}^{+6.5} \sum_{x_R=x_L}^{+6.5} exp\left[-\left(\sum_{n=x_L}^{x_R} q_{ad,n} + \sum_{m=x_L+1/2}^{x_R-1/2} q_{el,m}\right)\right] \tag{3.9}$$

where

$$q_{el,m} = \begin{cases} q_{el,m} & \text{if } m \neq m_d \\ q_{el,m_d} = q_{el,m} + \Delta q_{m_d} & \text{if } m = m_d \end{cases} \tag{3.10}$$

and Equation 3.10 depends on the damage location. The probability that the damage is located at SHL m_d is

$$P(m_d) = Z_{fix}(m_d)/Z_{float} \tag{3.11}$$

with the values of the partition functions determined using Equations 3.7 - 3.10.

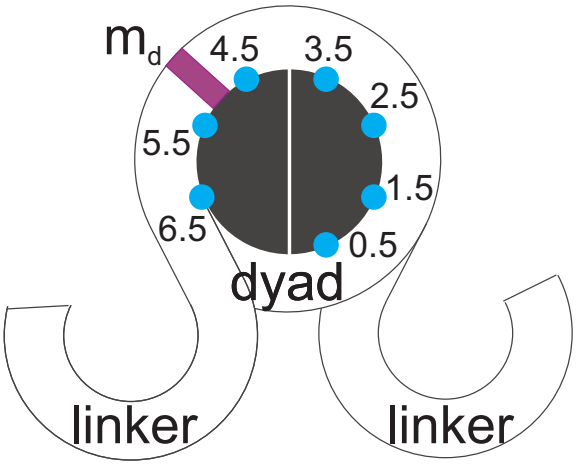


Figure 3.7: Damaged DNA location with histone octamer shown. The damaged DNA segment is shown here at SHL +5 and located midway between binding sites (blue) at SHL coordinates +4.5 and +5.5. Numbering of the wrapping segments begins with 0 at the dyad (dyad axis is the white line) and continues to ± 6. The negative SHL are symmetrically located and not shown.

3.2.3 Estimating the energy

I estimated the elastic energy associated with bending the double stranded DNA using the wormlike-chain approximation [59].

$$E_{bend} = \frac{l_p l}{2R^2} k_B T \tag{3.12}$$

For double stranded DNA the persistence length l_P is ≈ 50 nm and is related to the stiffness of the DNA. The radius of the nucleosome (R) is estimated to be 4.18 nm based on X-ray crystal structure [11, 12]. With the DNA wrapped 1.7 times around the nucleosome each wrapping segment is $\approx 45^\circ$. This estimate of $\approx 45^\circ$ was found by multiplying the approximately 1.7 wraps of DNA by the 360° in a circle and dividing by the 13 wrapped segments. The length of the wrapping segment l is 3.3 nm ($2\pi R \frac{45^\circ}{360^\circ}$). Putting these values into Equation 3.12, the elastic energy required to bend each wrapping segment is $E_{bend} \approx 4.7 k_B T$. With 13 wrapping segments, located from SHL -6 to +6, the nucleosome has a total elastic energy $E_{el} \approx 61.1 k_B T$. The net binding energy of the DNA to the nucleosome $E_{net} = E_{el} + E_{ad}$ and has been estimated as $\approx -15 k_B T$ [59, 60]. Assuming $E_{net} \approx -15 k_B T$ and $E_{el} \approx 61 k_B T$, the expected adsorption energy E_{ad} is $\approx -76 k_B T$. However, these values are just estimates to give some reasonable values to consider in the model.

3.3 Results and discussion

The results shown in Figures 3.8 - 3.11 were generated from Equations 3.4 - 3.11 using the Spyder Integrated Development Environment (IDE) with Python 2.7.9 32bits, Qt 4.8.4, PyQt4 (API v2) 4.9.6 on a Windows 10 workstation.

3.3.1 Uniform E_{el} and E_{ad}

I first considered the binding adsorption energies and the elastic bending energy to be uniform except in the wrapping segment that contains the damage. The damaged wrapping segment, located at SHL m_d, has an associated elastic energy of q_{el,m_d}. I approximated the elastic energy for flexible damage to be $q_{el,m_d} \approx 0k_BT$. An increase in flexiblity is consistent with single-stranded nicks decreasing the rigidity of DNA [61]. Estimates for the energies of the nucleosomes with healthy DNA vary and there appears to be no clear consensus so a range of values found in literature were considered. The elastic bending energies presented here are $q_{el} = 5k_BT$ (as estimated above using the wormlike-chain approximation), $7k_BT$, and $20k_BT$. The net binding energies of the nucleosome have been estimated as $E_{net} \approx -15k_BT$ [59, 60], which when translated to the damaged DNA model coordinate systems gives $q_{net} \approx -15k_BT/14$ binding sites $\approx -1k_BT$ per binding site. Therefore, q_{net} values of $-0.5k_BT$, $-1.0k_BT$ and $-3.0k_BT$ were used. The adsorption energies q_{ad} are determined by the constraint $q_{net} = q_{ad} + q_{el}$ and are shown in Table 3.1. The probabilities of the damage with $q_{el,md} = 0k_BT$ being located at a particular SHL were calculated using Equation 3.11 and are shown in Figure 3.8. It is interestingly that each q_{net} produced very similar probabilities regardless of $q_{ad,n}$ and $q_{el,m}$ (results not shown since the probabilities are so similar). For example $q_{net} = -0.5k_BT$ has similar probabilities for $(q_{el} = 5.0k_BT, q_{ad} = -5.5k_BT)$, $(q_{el} = 7.0k_BT, q_{ad} = -7.5k_BT)$, and $(q_{el} = 20.0k_BT, q_{ad} = -20.5k_BT)$. When both the adsorption and binding energies are uniform, q_{net} is the only parameter that significantly impacts the probabilities. Figure 3.8 (red line) shows that when $q_{net} = -0.5k_BT$ the highest probability for the location of the damage is at SHL 0 which is the dyad. As q_{net} decreases in Figure 3.8 from $-0.5k_BT$ (red

line) to $-1.0k_BT$ (blue line) and finally to $-3.0k_BT$ (green line), the probability of the damage being located that SHL 0 decreases and the damage has less preference for any particular locations. As q_{net} continues to decrease (not shown), the damage no longer has a preference for location and all SHL locations are equally likely.

Table 3.1: Uniform energies. Each of the 13 bending segments has elastic energy $q_{el,m} = 5k_BT$, $7k_BT$, or $20k_BT$ and the net energy of the system is $q_{net} = -0.5k_BT$, $-1.0k_BT$, or $-3.0k_BT$. Since $q_{net} = q_{el} + q_{ad}$ the 14 binding contact sites have energies determined by this constraint (*e.g.* if $q_{el} = 7.0k_BT$ and $q_{net} = -1.0k_BT$, then $q_{ad} = -8.0k_BT$).

q_{net} \ q_{el}	5.0	7.0	20.0
-0.5	-5.5	-7.5	-20.5
-1.0	-6.0	-8.0	-21.0
-3.0	-8.0	-10.0	-23.5

These results are understandable especially when looking at Equations 3.7 and 3.9. When the energies are uniform, the sums within the exponential in Equation 3.5 and Equation 3.6 are multiples of q_{net} plus the adsorption energy. For example, when the nucleosome is wrapped from x_L = -5.5 to x_R = -3.5 the associated adsorption binding energies are $q_{ad,-5.5} + q_{ad,-4.5} + q_{ad,-3.5}$ and elastic bending energies $q_{el,-5.0} + q_{el,-4.0}$. Because the energies are uniform, the sum of the associated energies for this conformation reduces to $2q_{net} + q_{ad}$.

As q_{net} decreases the magnitude of adsorption binding energy q_{ad} is much larger than the magnitude of the elastic energy q_{el}. This means the net binding q_{net} is mostly due to the adsorption energy. Consider an extreme where the q_{ad} is much greater than the q_{el}. When the nucleosome is wrapped from x_L = -5.5 to x_R = -3.5 the associated adsorption binding energies are $q_{ad,-5.5} + q_{ad,-4.5} + q_{ad,-3.5}$ and elastic bending energies are $q_{el,-5.0} + q_{el,-4.0}$. When the energies are uniform and $q_{ad} \gg q_{el}$ this reduces to $3q_{ad}$. Therefore when ($q_{ad} \gg q_{el}$) the adsorption binding energy dominates the elastic energy and changes in elastic energy due to damage should have little impact. In other words, when the adsorption binding energy is high, meaning that DNA is very tightly bound, the damage is unlikely to reposition and therefore all locations have similar proba-

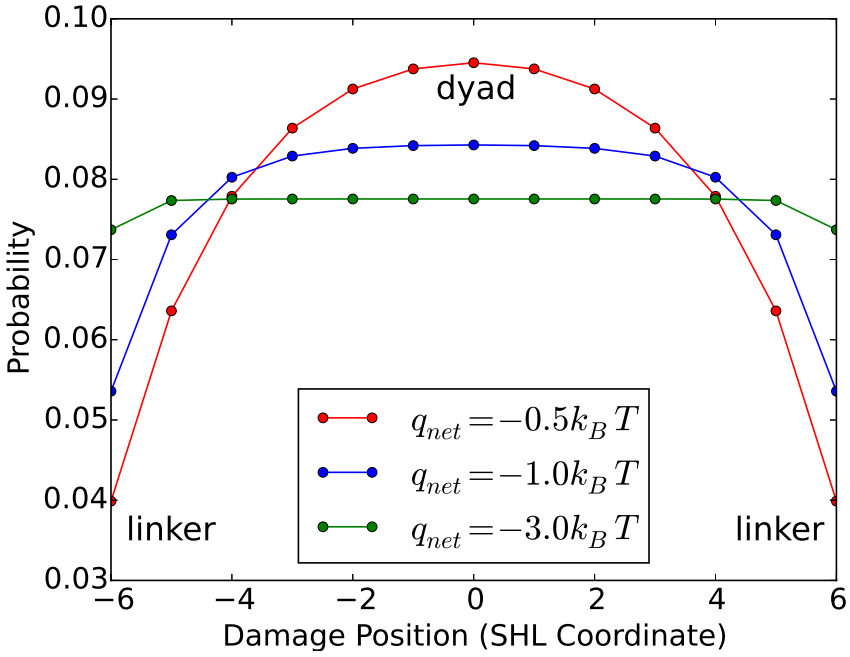


Figure 3.8: Probability of flexible damage being located at a particular segment on the nucleosome with uniform energies. For all plots, the elastic energy of the wrapped segment containing the flexible damage is $q_{el,m_d} = 0$. Red line: when $q_{net} = -0.5 k_B T$, the damage is most likely to be located at SHL 0 which is the dyad and less likely to be located near the linker DNA at SHL ± 6. Blue line: when $q_{net} = -1.0 k_B T$, the damage is slight more likely to be located toward the center of the wrapping from SHL -3 to 3, and less likely to be located near the linker DNA at SHL ± 6. Green line: when $q_{net} = -3.0 k_B T$, the damage is equally likely to be located from SHL -5 to 5 and only slightly less likely to be located at SHL ± 6.

bilities for the damage. Conversely, as q_{net} approaches zero the magnitudes of the elastic bending energy and the absorption binding energy are somewhat neutralized so a change in elastic energy has an impact on the system. The DNA is not so tightly wrapped that repositioning is hindered. In the case of flexible damage or a decrease in the elastic energy, the damaged DNA should be more easily wrapped than healthy DNA. The breathing process will then likely reposition the damage near the dyad.

Since it is not always known how damage will impact the energy of the damaged DNA, we also considered the possibility that the damaged DNA becomes more rigid. Exposure to UV radiation has been shown to increase the rigidity of the DNA due to the changes in the structure of the DNA [62]. More rigid DNA suggests more bending energy, or an increase in the elastic energy, for the wrapping segment that contains the damage. Figure 3.9 shows the results for damage that increases the bending energy $2k_BT$. Specifically, this plot shows an elastic energy of $q_{el,m} = 7k_BT$ with the damaged segment having elastic energy of $q_{el,m_d} = 9k_BT$. Figure 3.9 assumed that $q_{net} = -0.5k_BT$ (red line), $-1.0k_BT$ (blue line), and $-3.0k_BT$ (green line). As q_{net} approaches zero there is increased preference for the location of the damage. However, unlike the flexible damage, rigid damage is more likely to be located towards the edges of the nucleosome. This is understandable because during the breathing process the DNA wraps and unwraps from the histone core. If the damage is sufficiently rigid, assuming that it unwraps, there might not be enough energy to bend and rewrap the damaged segment. This result is consistent with UV radiation causing the damaged DNA to be more rigid [62] and the damage being 6 times more likely to be located in the linker [63]. When the damage increased the flexibility, the segment was more likely to be wrapped, however, the repositioning appears to be more effective for rigid damage than flexible damage.

3.3.2 Rough binding energy landscape

Having already considered uniform adsorption energies, another case to consider is adsorption energies that vary accross the different contact sites [64]. Using previously reported results, I used

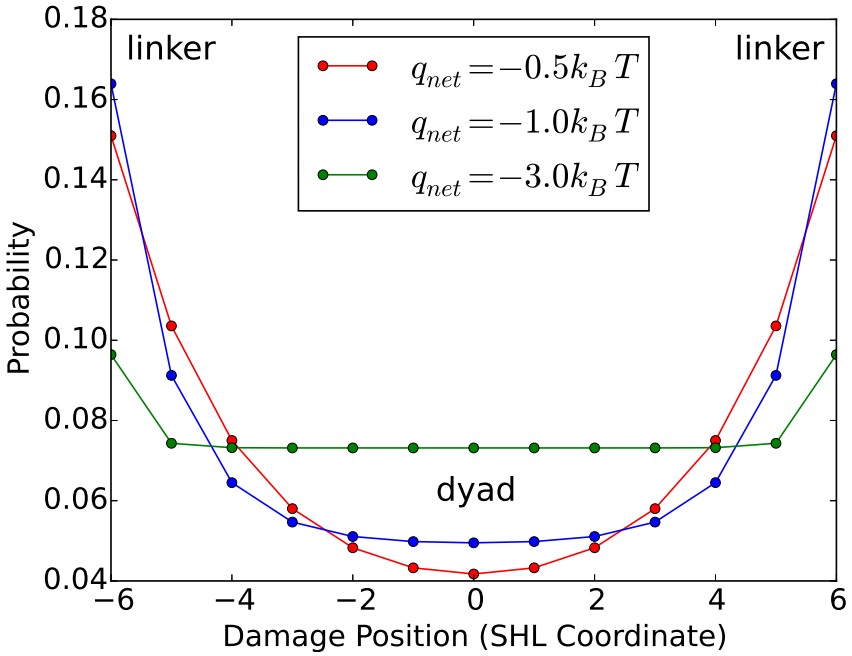


Figure 3.9: Probability of rigid damage being located at a particular segment on the nucleosome with uniform energies. For all plots, the elastic energy of undamaged DNA is $q_{el,m} = 7k_BT$ and the elastic energy of the segment with the damage is $q_{el,m_d} = 9k_BT$. Red line: $q_{net} = -0.5k_BT$ which yields a $q_{ad} = -7.5k_BT$, Blue line: $q_{net} = -1.0k_BT$ ($q_{ad} = -8.0k_BT$), Green line: $q_{net} = -3.0k_BT$ ($q_{ad} = -10.0k_BT$). As in the flexible damage case, q_{net} values closer to zero produce stronger preferences, however the rigid damage prefers the edges of the nucleosome.

a similar approach to estimate $q_{ad,n}$. I began by estimating the average net binding energy assuming

$$\bar{q}_{net} = \frac{1}{2}(f_0 + f_1)d + \frac{E_{el}}{13} \tag{3.13}$$

In Equation 3.13, f_0 and f_1 are the force required to separate the DNA from the histone at a single contact point [64] and $d = 3.4$ nm is the distance between the binding contact sites. There are two estimates for the amount of work done to separate the DNA from the nucleosome structure because of the electrostatic repulsion. Removing the first turn of DNA (with $f_0 = 0.7k_BT/\text{nm}$) leads to an increase in the work needed to remove the second turn (with $f_1 = 1.4k_BT/\text{nm}$) since the electrostatic repulsion of the two strands is no longer present [64]. Though values for the average elastic energy vary, $E_{el} = 70k_BT$ was used to maintain consistency with Reference [64] upon which the average energy Equation 3.13 is based. Using $f_0 = 0.7k_BT/\text{nm}$, $f_1 = 1.4k_BT/\text{nm}$, $d = 3.4$ nm, and elastic energy $E_{el} = 70k_BT$ in Equation 3.13, the estimated average is $\bar{q}_{net} \approx 9.0k_BT$, while the work in reference [64] estimated the average energy to be $12k_BT$, leaving a difference of $\approx 3.0k_BT$. To determine the adsorption energy for each contact site in the SHL coordinate system of the nucleosome breathing model, the estimated energies from Reference [64] (Table 3.2 [64]) were modified to account for the difference of $3.0k_BT$ determined by Equation 3.13 to obtain the adsorption energies per binding site [7] (Table 3.2 [7]). Of note is the binding energies near the dyad being stronger than those towards the edges of the nucleosome.

Table 3.2: Estimated adsorption energy for each binding site. The adsorption energies found in literature and modified by Equation 3.13 with using $f_0 = 0.7k_BT/\text{nm}$, $f_1 = 1.4k_BT/\text{nm}$, $d = 3.4$ nm and elastic energy $E_{el} = 70k_BT$.

Binding Site SHL Coordinate	±6.5	±5.5	±4.5	±3.5	±2.5	±1.5	±0.5
Energy (k_BT) reference [64]	9.2	11.2	12.8	9.2	10.1	14.4	16.2
Energy (k_BT) our model [7]	-6.2	-8.2	-9.8	-6.2	-7.1	-11.4	-13.2

Using $q_{el,m} = 70k_BT/13 \approx 5.4k_BT$ and the $q_{ad,n}$ values from Table 3.2 [7] in Equations 3.7 and 3.9, the position of flexible damage with $q_{el,m_d} = 0k_BT$ is shown in Figure 3.10 (green line). These results are very similar to the flexible damage with $q_{net} = -3.0k_BT$ shown in Figure 3.8 (green

line). This is not surprising because adsorption energy per binding site (defined as $q_{ad,avg}$) from the values in Table 3.2 [7] is $q_{ad,avg} = -8.9k_BT$. Having used $q_{el,m} = 70k_BT/13 \approx 5.4k_BT$ means the average $q_{net,avg} = -3.5k_BT$, and the damage is slightly less likely to be located at SHL ± 6. However, DNA has been shown to interact with other cellular components [65–68]. Interactions could lead to changes in $q_{el,m}$, $q_{ad,n}$, and $q_{net,avg}$ and therefore the preferred location for the damage. For the uniform energies when q_{net} approached zero, the nucleosome could easily unwrap, since the elastic energy and binding energy cancel each other out at every point. However, for this rough energy landscape, increases to q_{el}, or equivalently decreases to q_{ad}, could lead to the average $q_{net,avg}$ approaching zero without the nucleosome completely unwrapping. Figure 3.10 shows that as the elastic energy $q_{el,m}$ (and therefore the average net energy $q_{net,avg}$) increases from $q_{el,m} = 7.4, q_{net,avg} = -1.5k_BT$ (blue line) to $q_{el,m} = 7.9, q_{net,avg} = -1.0k_BT$ (red line), the probability of the damage being located near the center of the nucleosome increases. These results indicate increases to the elastic energy or the decreases to the adsorption energy (*e.g.* with chromatin remodelers) would likely cause flexible damage to be repositioned to the dyad.

The results of a rough energy landscape with adsorption binding energies from Table 3.2 [7] and rigid damage is shown in Figure 3.11. The results are similar to the uniform energies because in both cases, the rigid damage is more likely to be repositioned out towards toward the linker DNA. However, the results of this rigid damage in a rough energy landscape are also different from all three previous cases (rough energy with flexible damage, uniform energy with flexible damage, uniform energy with rigid damage). In the previous three cases (Figures 3.8 - 3.10), the probabilities were flattening out with decreasing q_{net} which means the damage had about the same probability of being located at any location. For these cases, the adsorption binding energy was so high compared to the elastic bending energy that damage was not likely to be repositioned and was almost equally likely to be located anywhere within the nucleosome. As shown in Figure 3.11, with decreasing q_{el}, the damage is most likely located at in the linker DNA. For $q_{el} = 7.4k_BT$ or $7.9k_BT$, there appears to be little preference for the location of the damage in the nucleosome.

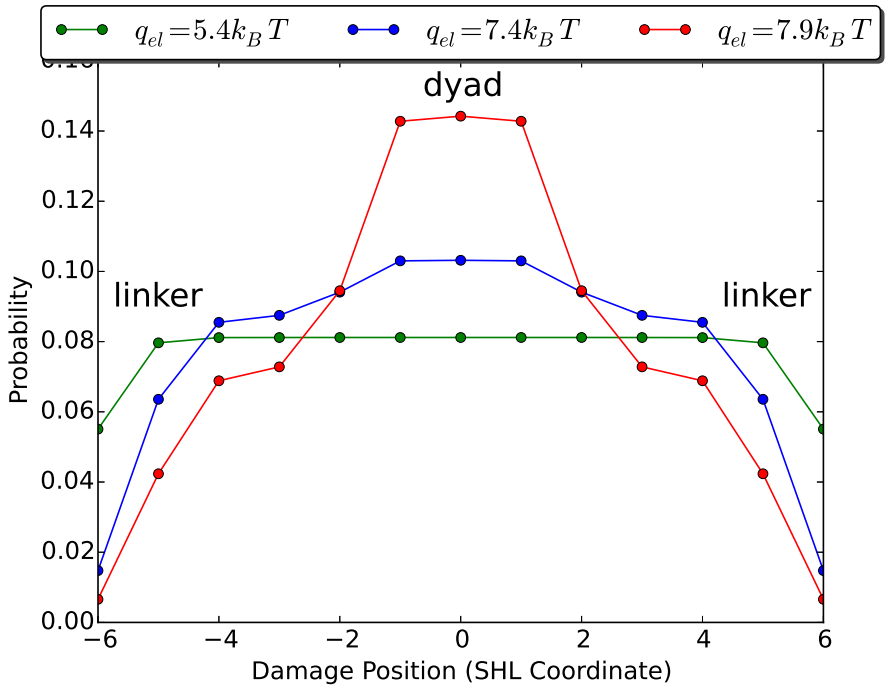


Figure 3.10: Probability of a flexible damage being located at a particular segment on the nucleosome with a rough binding energy landscape. The damage is flexible with $q_{el,m_d} = 0k_BT$ and adsorption energies $q_{ad,n}$ are from Table 3.2. Green line: damage position distribution with $q_{el} = 5.4k_BT$ and $q_{net,avg} = -3.0k_BT$. Blue line: damage position distribution in DNA whose bending energy has been increased by $2.0k_BT$ to $q_{el} = 7.4k_BT$ so $q_{net,avg} = -1.0k_BT$. The same result is achievable if the net binding energy is lowered by $2.0k_BT$ through competitive interactions with nucleosome remodelers. Red line: damage position distribution in DNA with q_{el} increased by $2.5k_BT$ to $7.9k_BT$ and $q_{net,avg} = -0.5k_BT$.

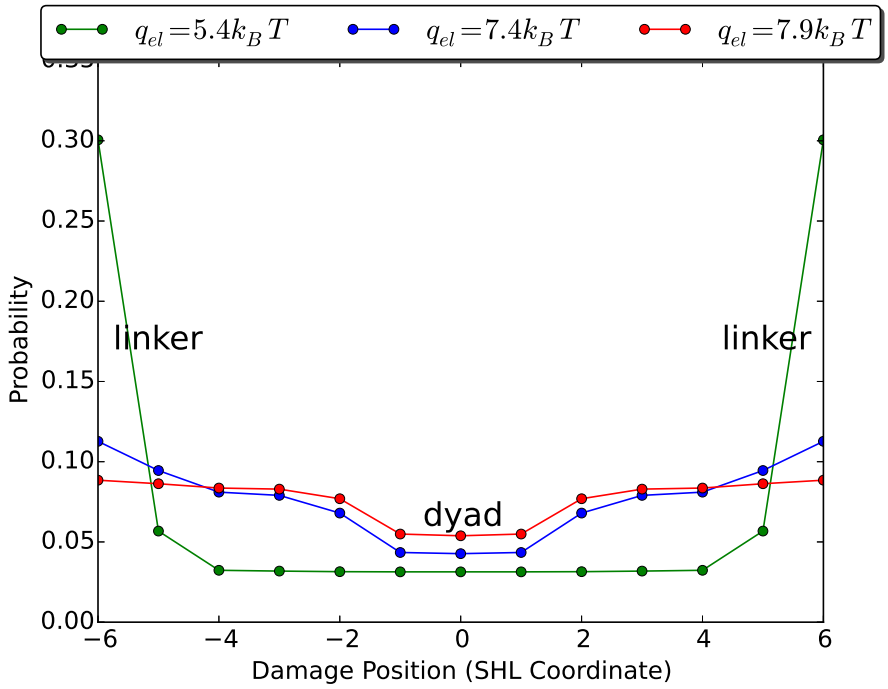


Figure 3.11: Probability of rigid damage being located at a particular segment on the nucleosome with a rough binding energy landscape. $q_{el,m_d} = 8.4k_BT$, $q_{ad,n}$ from Table 3.2. Green line: damage position distribution with $q_{el} = 5.4k_BT$ and $q_{net,avg} = -3.5k_BT$. Blue line: damage position distribution in DNA whose bending energy has been increased by $2.0k_BT$ to $q_{el} = 7.4k_BT$ and $q_{net,avg} = -1.5k_BT$. The same result is achievable if the net binding energy is lowered by $2.0k_BT$ through competitive interactions with nucleosome remodelers. Red line: damage position distribution in DNA with q_{el} increased by $2.5k_BT$ to $7.9k_BT$ and $q_{net,avg} = -1.0k_BT$. These results are different from Figures 3.8 - 3.10 because here, as q_{el} increased, the probability of the DNA being located at any particular site were very similar.

3.4 Conclusions

This work suggests that nucleosome breathing may be a part of the DNA repair process since the breathing results in preferred locations for DNA damage. The structural details of the nucleosome, the damaged DNA, and possibly other cellular interactions (*e.g.* chromatin remodelers) will determine the strength of the preferred location for the damage. Not surprisingly, DNA damage that increases the rigidity of the DNA is most likely located towards the edges of the nucleosome (linkers) by the breathing process, potentially due to the damaged segment requiring too much elastic energy to bend damage around the histone octamer. On the other hand, DNA damage that increases the flexibility of the DNA is most likely located near the dyad, potentially due to the damaged segment being more flexible and therefore requiring less elastic energy to bend than the healthy DNA. Under normal conditions, flexible damage is only weakly repositioned towards the dyad. If non-damaged DNA were to stiffen or nucleosome binding were to weaken due to cellular interactions, the damage can have a strong preference for the dyad region. The essential element for the preferential repositioning of the DNA damage is the nucleosome breathing, or wrapping and unwrapping of the DNA from the histone octamer. Since damaged DNA is more likely to be located at the linker or the dyad and these locations are no longer sterically hindered, the damage is more likely to be found by proteins engaged in a random search. With about 3/4 of DNA wrapped in the nucleosome structure, if repositioning of the damaged DNA occurs, it would greatly reduce the search space. Therefore, I believe that the breathing of the nucleosome, which results in preferred locations for the DNA damage, may be part of the DNA repair process.

Chapter 4

Nucleosome experiments

4.1 Introduction

The nucleosome breathing model presented in Chapter 3 predicts the probability of damage being located at a specific position in equilibrium. The actual probability of the damage being located at a particular site will depend on the details of the nucleosome structure, the energies associated with damaged DNA, and interactions with other molecules in the cell. Furthermore, the model predicts that the equilibrium conformations of the nucleosome containing damaged DNA will be different for nuclesomes containing undamaged DNA. Single molecule fluorescence resonant energy transfer (sm-FRET) measurements can provide direct measurements of the structure of nucleosomes on a molecule-by-molecule basis [1, 36, 69–75] and thus provide a way to study nucleosomes containing damaged DNA. Analysis of FRET experiments have been done previously [1] to estimate the FRET distribution of conformations of the nucleosomes for healthy DNA. Differences in the FRET distribution of nucleosomes containing damaged DNA to those with healthy DNA would be consistent with the nucleosome breathing model (Chapter 3).

In a typical nucleosome FRET experiment (Section 2.3), a donor and an acceptor fluorescent dye pair are attached to the DNA far enough apart that FRET does not occur unless the DNA is wrapped in the nucleosome structure (Figure 4.1). The proximity of the dyes to each other factors into their lifetimes and brightness. The lifetime of a molecule is the amount of time it takes the

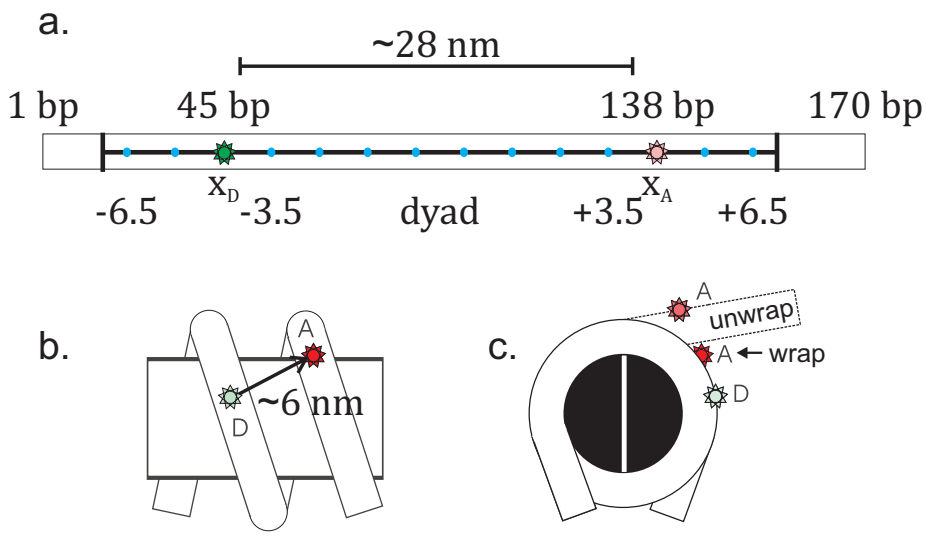


Figure 4.1: DNA and nucleosomes for FRET experiments. In part (a) the DNA is unwrapped and not in the nucleosome structure. The donor x_D and acceptor x_A dyes (located at basepairs 45 and 138 or SHL ± 4.5) are separated by a distance of ≈ 28 nm so FRET does not occur. In this conformation, the donor (green) should be bright and there should be little to no acceptor (red). When in the nucleosome structure, the DNA can be wrapped from SHL -6.5 to SHL +6.5. Part (b) is a side view of the DNA wrapped in the nucleosome structure where the dyes are located ≈ 6 nm apart. The donor (green) should be dim due to quenching and there should be a very bright acceptor (red) due to the high FRET. Part (c) shows the nucleosome (rotated 90° from part(b)) going from wrapped to partially unwrapped (dashed line). As the unwrapping occurs, the distance between the dyes increases leading to decreases in FRET. This decrease in FRET is observed with increases to the brightness of the green (donor) dye and decreases to the brightness of the red (acceptor) dye.

molecule to release a photon while in an excited state. When the donor dye alone in excited, and the dyes are in close proximity, the donor will be quenched (dimmed) while the acceptor will be bright due to FRET (Figure 4.1).

In the FRET experiment for nucleosomes as shown in Figure 4.1, high FRET occurs when the DNA is wrapped at least one contact point lower (in SHL) than the donor (located at basepair x_D) to at least one contact point higher (in SHL) than the acceptor (located at basepair x_A). Figure 4.1(b) shows the donor (green) dye located at SHL -4.5 and the acceptor (red) dye is located SHL +4.5, which means that high FRET should occur when the DNA is wrapped from SHL -5.5 to SHL +5.5. The probability of high FRET is found by summing over all configurations that fulfill these constraints. The probability of high FRET (P_{HF}) is

$$P_{HF} = \frac{1}{Z} \sum_{x_L=-6.5}^{x_D} \left[\sum_{x_R=x_A}^{+6.5} exp \left[- \left(\sum_{n=x_L}^{x_R} q_{ad,n} + \sum_{m=x_L+1/2}^{x_R-1/2} q_{el,m} \right) \right] \right] \tag{4.1}$$

where Z is the partition function from Equation 3.7 and x_D is the basepair location of the donor dye and x_A is the basepair location of the acceptor dye. In Equation 4.1, $q_{ad,n}$ are the adsorption energies for contact sites at SHL -6.5 to 6.5 and $q_{el,m}$ are the elastic bending energies at SHL -6.0 to +6.0. For experiments with damaged DNA, the condition

$$q_{el,m} = \begin{cases} q_{el,m} & \text{if } m \neq m_d \\ q_{el,m_d} = q_{el,m} + \Delta q_{m_d} & \text{if } m = m_d \end{cases} \tag{4.2}$$

should be included.

When the DNA unwraps enough that x_A and/or x_D are no longer bound in the nucleosomes structure, the distance between the dyes increases. Increases in the distances should decrease the FRET efficiency (Section 2.3). In Figure 4.1(c), the DNA is wrapped from SHL -6.5 to SHL +3.5 and x_A is no longer wrapped. This should have lower FRET than the fully wrapped DNA. These non high FRET configurations are referred to as intermediate FRET states. One set of intermediate states occur when the donor dye x_D is still fully wrapped but the acceptor dye x_A is no longer wrapped. As x_A unwraps, the distances between the dyes increases leading to decreases in FRET. By considering all states where the acceptor dye x_A is unwrapped j segments beyond the last high FRET position, the probability of intermediate FRET ($P_{IFA(J)}$) is

$$P_{IFA(j)} = \frac{1}{Z} \sum_{x_L=-6.5}^{x_D} exp \left[- \left(\sum_{n=x_L}^{x_A-j} q_{ad,n} + \sum_{m=x_L+1/2}^{x_A-j-1/2} q_{el,m} \right) \right] \tag{4.3}$$

Figure 4.1(c) shows an example of an intermediate FRET state where the acceptor dye (located at SHL +4.5) is unwrapped. The DNA is wrapped from SHL -6.5 to SHL +3.5 which leaves the acceptor dye one segments outside the wrapped DNA (corresponding to $j = 1$).

Similarly for the donor dye unwrapped i segments, the probability of intermediate FRET

$P_{IFD(I)}$ is

$$P_{IFD(i)} = \frac{1}{Z}\sum_{x_R=x_A}^{+6.5} exp\left[-\left(\sum_{n=x_D+i}^{x_R} q_{ad,n} + \sum_{m=x_D+i+1/2}^{x_R-1/2} q_{el,m}\right)\right] \tag{4.4}$$

If both dyes are unwrapped by some amount i or j, the probability of intermediate FRET $P_{IFD(i)A(j)}$ is

$$P_{IFD(i)A(j)} = \frac{1}{Z} exp\left[-\left(\sum_{n=x_D+i}^{x_A-j} q_{ad,n} + \sum_{m=x_D+i+1/2}^{x_A-j-1/2} q_{el,m}\right)\right] \tag{4.5}$$

where the outermost sum disappears since specifying i and j determines a unique partially wrapped state. For estimating the intermediate FRET with nucleosomes containing damaged DNA, Equations 4.3 - 4.5 would to include the constraints in Equation 4.2.

Previous FRET experiments of Gansen, et al. [1] found 55% of nucleosomes in high FRET configurations. By comparing this value to Equation 4.1, the energies can be estimated for nucleosomes with healthy DNA. To compare the nucleosome breathing model to the FRET experimentally data, I assumed there was no damage to the DNA which made the conditions in Equation 4.2 no longer applicable. In our SHL coordinate system notation, the donor and acceptor fluorescent dyes were experimentally located at SHL -4.5 (x_D) and SHL +4.5 (x_A) as shown in Figure 4.1. By setting 55% equal to Equation 4.1 and using the adsorption energies from Table 3.2, I determined the elastic energy to be $q_{el,m} \approx 7.1 k_B T$. This estimate is high compared to the wormlike chain approximation estimate of $\approx 4.7 k_B T$ (Section 3.2.3) but is much lower than the estimate of $\approx 12 k_B T$ from the course grain model in reference [64]. Using the value of $q_{el,m} \approx 7.1 k_B T$ and assuming uniform adsorption energies, to have 55% high FRET, the nucleosome breathing model predicts a $q_{net} \approx -0.45 k_B T$. This estimate of q_{net} is lower than $\approx -1.1 k_B T$ from reference [59] and much lower than $\approx -6.6 k_B T$ from reference [64]. The differences between the results could be because reference [1] was studying the nucleosome disassembly experimentally, while reference [64] was studying the twisting of nucleosomal DNA through modeling of the nucleosome. In the nucleosome disassembly experiments, the nucleosomes were subjected to different salt concentrations which has an impact on the wrapping structure [1]. Therefore, the conditions of those experiments may have been biased towards nucleosomes unwrapping and the nucleosomes may

have been unable to rewrap. This would result in larger numbers of partially unwrapped nucleosomes and therefore a lower estimate of q_{net} since the binding and bending energies would almost neutralize each other.

Not knowing the complete structural details (*e.g*, energies) of the nucleosome and DNA damage does not mean there are not experiments that can monitor the repositioning of the damaged DNA. Using FRET experiments, the distribution of the FRET efficiency for nucleosomes with healthy DNA can be compared to those with damaged DNA. Differences in the FRET distributions (which is equivalent to differences in the distances between the donor and acceptor dyes), would suggest that the damaged DNA in the nucleosome has a different wrapping structure than the healthy DNA. This would suggest the damaged DNA is being repositioned. However, additional experiments would be needed to determine the specific location of the damage or whether the damage is being repositioned to the dyad or the linker as our damaged DNA model suggests.

4.2 Materials and methods for FRET experiments

The following sections briefly describe the materials and methods I used to create nucleosomes with both healthy and damaged DNA for FRET experiments. Unless otherwise noted, the protocols and instructions provided by the manufacturer were followed.

4.2.1 Creating healthy DNA with polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) is a technique used to exponentially amplify a particular region of template DNA using thermal cycling (shown in Figure 4.2). Denaturation of the double stranded DNA (dsDNA) into single stranded DNA (ssDNA) occurs in the first step of the cycling process. By increasing the temperature (to 95°C) the hydrogen bonds of the basepairs are broken separating the dsDNA into double strands of ssDNA as shown in Figure 4.2(1). Annealing of the primers to the ssDNA occurs in the second step. By reducing the temperature (to typically $45-68$°C depending on the primers) the ssDNA primers (18-22 bases in length) complementary bind to the

target region as shown in Figure 4.2(2). Extension of the ssDNA is the final step in the cycle. By increasing the temperature (to $65-68^{\circ}$C depending on the *Taq* polymerase used), the *Taq* polymerase uses free deoxynucleoside triphosphates (dNTP which is adenine, thymine, cytosine, and guanine containing triphosphate groups) to enzymatically extend the ssDNA to dsDNA as shown in Figure 4.2(3). These steps are repeated for 25 - 35 cycles with each cycle approximately doubling the amount of target DNA. Some particular products may require an initial heat activation period to activate the *Taq* polymerase prior to the cycles and/or an additional extension period after the cycles. For more detailed information on PCR see the New England Biolabs Website.

In order to PCR the 170 basepair sequence of the Widom 601 sequence (shown in Table 4.1) contained in the pGEM-3Z plasmid, the HotStar*Taq* DNA Polymerase Kit (ID:203601 from Qiagen and included an optional Q-Solution) was used. The primers, purchased from Eurofins Genomics, were

Primer 1: which contains Alexa 488 green dye (51 bases total)

5'-ACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGCCGCTCAATTG

Primer 2: which contains Alexa 594 red dye (38 bases total)

3'-CAGAGGTCCGTGCACAGTCTATATATGTAGGACGCATAC-5'

The differences in melting temperatures 76°C (Primer 1) and 65°C (Primer 2) can be a problem in PCR. The PCR reactions that included Q-Solution produced better results, and according to Qiagen, the Q-Solution changes the melting behaviors to facilitate amplification of difficult templates or primers. The template that contained the DNA segment of interest was the pGEM-3Z plasmid (a kind gift from Dr. Timothy J. Richmond). A typical reaction mix is shown in Table 4.2 and the thermal cycling conditions are shown in Table 4.3

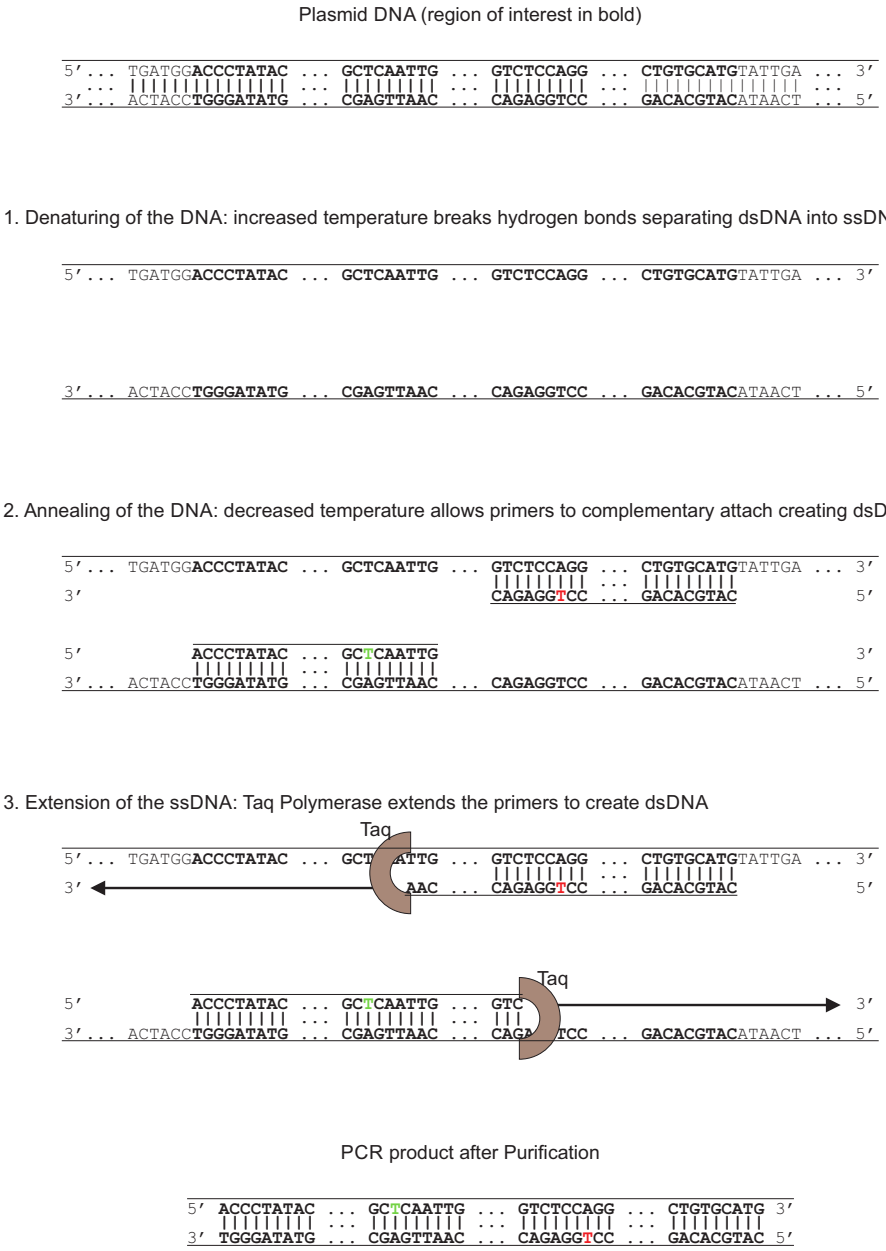


Figure 4.2: Polymerase chain reaction (PCR) of DNA. This shows the thermal cycling process for PCR. The steps of the cycling process are: (1) denaturate at temperature $94-98°$C. The high temperature breaks the hydrogen bonds allowing the dsDNA to separate into ssDNA, (2) anneal at temperature 50 - 65°C. The decrease in temperature allows the short primers to the complementary bind to the single stranded DNA (ssDNA), and (3) extend at temperature $72-80°$. The *Taq* Polymerase (protein) uses the dNTP (free nucleotides A, C, T, and G) to extend the ssDNA which creates dsDNA.

Table 4.1: Widom 601 Sequence - 170 basepairs. 147 basepairs are wrapped in the nucleosomes structure. The basepairs underlined are not wrapped in the nucleosome. The basepair in blue is located at the dyad. The bases shown in green and red are the locations of the donor and acceptor dye, respectively. Finally, the "|" has been used to show the location of the nick from restriction enzyme Nt.bsmAI which targets the sequence in magenta.

bp		Sequence
001-	5'	ACCCTATACGCGGCCGCCCTGGAGAATCCCGGTGCCGAGGCCGCTCAA
048	3'	TGGGATATGCGCCGGCGGGACCTCTTAGGGCCACGGCTCCGGCGAGTT
049-		TTGGTCGTAGACAGCTCTAGCACCGCTTAAACGCACGTACGCGCTGTC
096		AACCAGCATCTGTCGAGATCGTGGCGAATTTGCGTGCATGCGCGACAG
097-		CCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCC|AGGCACG
144		GGGGCGCAAAATTGGCGGTTCCCCTAATGAGGGATCAGAGG TCCGTGC
145-		TGTCAGATATATACATCCTGTGCATG - 3'
208		ACAGTCTATATATGTAGGACACGTAC - 5'

Table 4.2: Reaction mix per sample. Following the protocol for the Qiagen HotStar*Taq* Polymerase Kit, this is a typical reaction mix for the PCR of the 170 bp Widom 601 Sequence from the pGEM-3Z plasmid. The inclusion of the Q-solution was used to mitigate the difference in the melting temperatures of the two primers.

Component	**Volume / Reaction**	**Final Concentration**
Hot Start Buffer (10X)	10.00 μL	1X
Q-Solution (5X)	20.00 μL	1X
dNTPs (10mM)	2.00 μL	200 μM
Primer 1: 488 Primer (10 μM)	2.00 μL	0.2 μM
Primer 2: 594 Primer (10 μM)	2.00 μL	0.2 μM
Hot Start *Taq* Polymerase	0.50 μL	2.5 units / Reaction
H_2O (Molecular Grade)	58.50 μL	
Template DNA: pGEM-3Z ($\approx 1ng/\mu$L)	5.0 μL	$5ng \leq 1\mu g$ / Reaction
Total PCR Volume	100.00 μL	

Table 4.3: Thermal cycling conditions: These are the thermal cycling conditions used for the PCR in a Gene Cycler (Bio-Rad Thermal Cycler).

Step	**Time**	**Temperature**	**Comments**
Initial Heat Activation	95°C	15 minutes	Activates *Taq* Polymerase
3 Steps / Cycle			
Denaturation	94°C	30 seconds	Breaks hydrogen bonds of dsDNA
Annealing	62°C	30 seconds	Reduced temperature allows primers to attach
Extension	72°C	60 seconds	Free dNTP attached to ssDNA
35 Cycles			
Final Extension	72°C	10 minutes	allows any extra ssDNA to become dsDNA

In order to determine if the PCR process was successful, the product is often put into an agarose or polyacrylamide gel for electrophoresis (Section 2.3). Shown in the center three lanes of Figure 4.3 is the product from the PCR (described above) is a 1% agarose gel with 1X Tris/Borate/EDTA (TBE) buffer solution (Thermo-Fischer Scientific), run at 100V in the Hoefer HE-PLUS Horizontal Gel Electrophesis System, stained with 1X GelGreen (Biotium), and imaged on the Typhoon TRIO Imager (GE Healthcare). Control ladders are in the outer (left and right) lanes. The ladder contains DNA that is 100, 200, 300, ... basepairs in length. The center lanes show the lengthy pGEM-3Z plasmid at the top while the short primers (39 and 51 bases in length) pushed down towards the bottom of the gel. Between the pGEM and the primers is the 170 bp Widom 601 sequence of interest. As a way to insure that the final DNA product is of the proper length, it is typical to do a gel extraction. A gel extraction involves cutting the gel that contains the DNA of interest (the box marked W601 in Figure 4.3) and purifying that portion of the gel. For this project, the GenElute Gel Extraction Kit (Catalog Number NA1111 from Sigma-Aldrich) was used and protocols followed. After purification, the NanoDrop 1000 Spectrophotometer (Thermo Scientific) was used to quantify the concentration of the purified DNA.

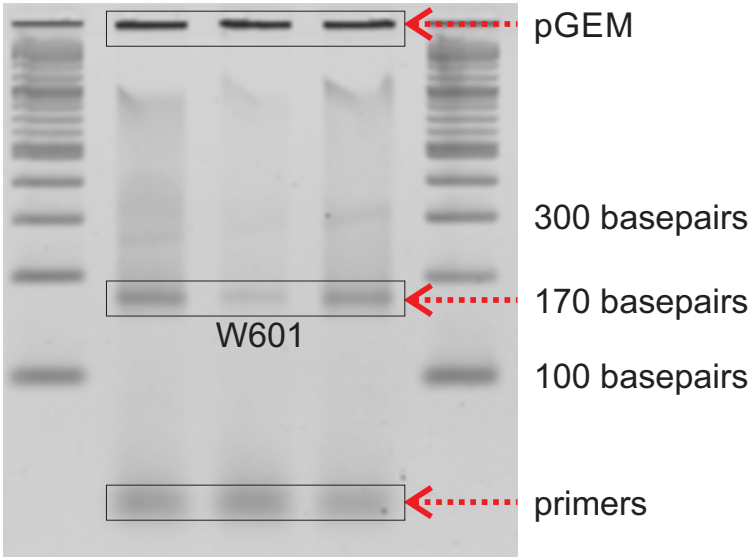


Figure 4.3: Gel image of the PCR of 170 bp Widom 601 sequence. The control ladder is in the outside (left and right) lanes and the PCR product is in the center lanes. The PCR product appears to be located at $\approx$ 170 bp with the short primers near the bottom and the long pGEM-3Z plasmid at the top. Gel image taken in July 2013 in the Antonik Lab at the University of Kansas.

4.2.2 Creating damaged DNA

There are different ways to create damaged DNA. One way is to "nick" the DNA using a restriction enzyme. This "nick" is a break in the phosphate backbone of the DNA strand (Figure 4.4). At the time of these experiments, there was only one restriction enzyme found to create a nick in the wrapping portion of the nucleosome for the 170 bp Widom 601 sequence (shown in Table 4.1). Enzyme Nt.bsmAI (NEB Catalog #R0121S) targets double stranded DNA and cleaves only one strand of the double stranded DNA. The enzymes binds to the target sequence 5'...GTCTCN|N...3' (Table 4.1 and Figure 4.4) where N denotes any base (A, T, C, or G) and the location of the nick is denoted '|'. I followed the protocol provided with the Nt.bsmAI restriction enzyme and used the recreation mix shown in Table 4.4. Another way to achieve the same "nick" is to anneal the individual "fragments" of the DNA. By simply putting the three ssDNA fragments (shown in Figure 4.4) together, heating to 95°C, and allowing to cool at room temperature overnight, the fragments annealed with no connection of the phosphate backbone between bases 137 and 138.

To try to experimentally determine that these approaches were creating the same damage both

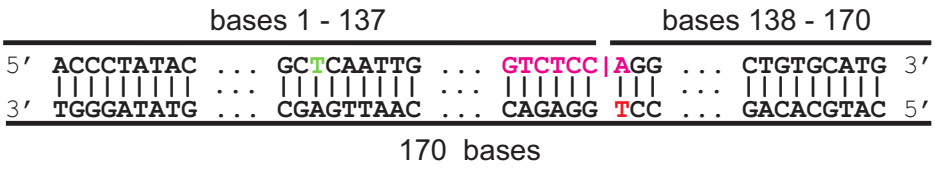


Figure 4.4: Damaged DNA with sequence. Damaged DNA for this project was either "nicked" using restriction enzyme Nt.bsmAI (that targets the sequence GTCTCN|N) or pieced together by annealing the three individual (bases 1 - 137, bases 138 - 170, and bases 1-170) ssDNA fragments. Both techniques result in a disconnect of phosphate backbone between bases 137 and 138.

the restriction enzyme "nicked" DNA and the "annealed" DNA fragments were run through a 2.5% agarose gel (Section 4.1). Analysis of the gel showed two bands (Figure 4.5) for the damaged DNA and only a single band (Figure 4.5(b)(1)) for the healthy DNA. Included in the gels were the other two annealing combinations: 137 bases to 170 bases and 33 bases to 170 bases (Figures 4.5(b)(3) and 4.5(b)(4), respectively) to serve as controls.

Table 4.4: Reaction mix per sample for the W601 DNA sequence: The following is the reaction mix used with the Nt.bsmAI restriction enzymes from NEB. The mix was incubated at 37°C for one hour then heat inactivated at 65°C for 20 minutes.

Component	**Volume / Reaction**	**Final Concentration**
NEBuffer4 (10X)	5.0 μL	1X
W601 DNA	1.0μg	1.0μg
$H_2$0 (Molecular Grade)	variable	
Nt.BsmAI restriction enzyme	2.0 μL	10 U / reaction
Total Reaction Volume	50.0 μL	

Using the Nt.bsmAI restriction enzyme with the intent of doing a FRET experiment can pose a potential issue for this particular sequence. The restriction enzyme target binding site includes the fluorescently labeled green donor dye. There is some concern the restriction enzyme may not be able to bind to the target site or that the cleaving process may be hindered. Because of this issue, annealing of the ssDNA fragments would likely be a better option.

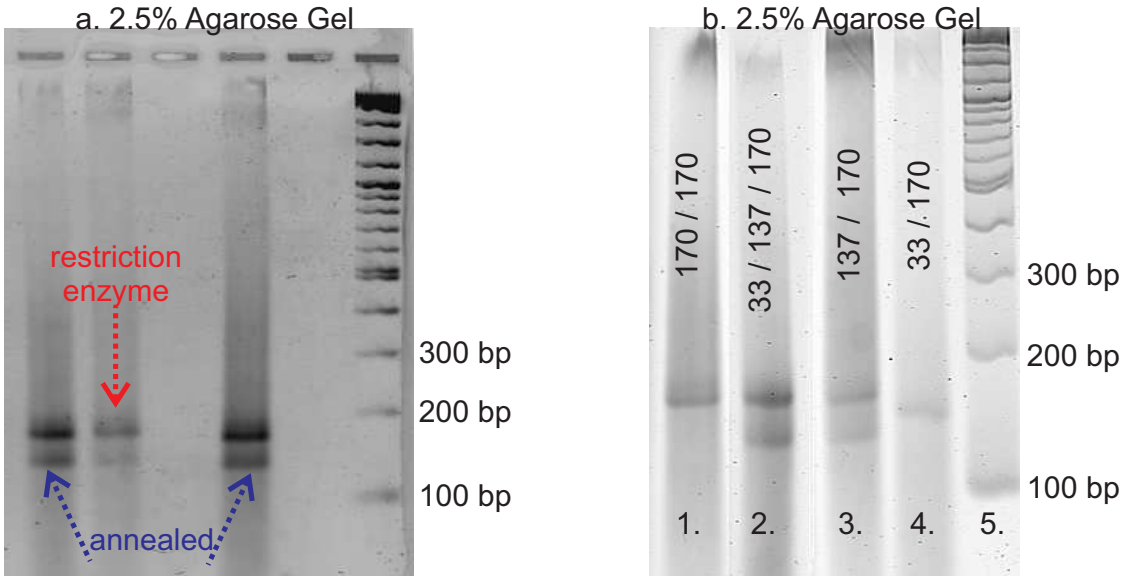


Figure 4.5: Gel of damaged DNA. In (a) DNA nicked using a restriction enzyme and DNA annealed using 33, 137, and 170 base ssDNA are shown in an agarose gel. When the DNA is damaged there are two bands. In (b) DNA annealed in all combinations of 33, 137, and 170 bases. The similarity of double bands for the damaged DNA suggests that either nicking the dsDNA using a restriction enzyme or annealing the ssDNA fragments are equivalent. Gel image taken in January 2014 in the Antonik Lab at the University of Kansas.

4.2.3 Nucleosome reconstitution

The reconstitution of nucleosomes is the process of wrapping DNA around a histone octamer. The reaction starts with the DNA and histone tetramers and histone dimers in a high salt concentration solution. As the salt concentration is reduced by either dialysis or dilution, each histone tetramer associates with two histone dimers to form the histone octamer. Further decreases in the salt concentration allow the negatively charged DNA to wrap around the positively charge histone octamer to form a fully wrapped nucleosome.

4.2.3.1 W601 Sequence

For nucleosome reconstitution, I used the 170 bp W601 sequence (Table 4.1) and the dilution protocol with the EpiMark Nucleosome Assembly Kit (Catalog #E5350S from NEB). A typical reaction mix for this dilution method is shown in Table 4.5.

Table 4.5: Nucleosome reconstitution reaction mix. Following the protocol for the dilution method using the EpiMark Nucleosome Assembly Kit, this is a typical reaction mix for the W601 Sequence. Note the use of 4 molar (M) concentration NaCl when the protocol called for 5M NaCl. 4M NaCl was used with the volumes adjusted so the reaction mix had a initial salt concentration of 2M.

Component	**Volume / Reaction**	**Comments**	**Final Concentration**
NaCl (4M)	3.75 μL		
H_2O (Molecular Grade)	variable		
W601 DNA	8.8 pmol		110 nM
Dimer (20μM)	1.25 μL	supplied in 2M NaCl	156 nM
Tetramer (10μM)	1.25 μL	supplied in 2M NaCl	Octamer
Total Reaction Volume	10 μL		DNA:Octamer = 1:1.4

After reconstitution, a gel shift assay (see Section 2.3) is typically done to verify the nucleosome structure. It is referred to as a gel shift assay because the large molecules (nucleosomes) should not move as far as through the gel as the smaller molecules (free DNA). Successful reconstitution shows a shift of a band upwards to the top of the gel due to the larger, slower moving nucleosomes. An example of a successful nucleosome reconstitution using both the W601 healthy and damaged DNA is shown in Figure 4.6. The nucleosomes (left side) have not traveled as far as the free DNA (right side) and have "shifted" upwards in the gel. The free DNA appears to have the correct base-pair size of $\approx$ 170 bp. The nucleosomes contain the same $\approx$ 170 bp DNA but are larger molecules since wrapped around the histone octamers. The damaged DNA model predicts a repositioning of the damage, which would suggest that nucleosomes with damaged DNA may shift either more or less than the nucleosomes with healthy DNA. This lack of difference between the damaged and healthy DNA nucleosomes could be due to the fact that the W601 sequence is a high binding affinity sequence. High binding affinity (high adsorption binding energy) means the DNA tightly binds to the histone octamer. As previously shown in Section 3.3, when damaged DNA occurs in a nucleosome that has high adsorption energies (which can be due to high binding affinity), there is less preference for the location of the damage. Therefore, repositioning of the damage may not

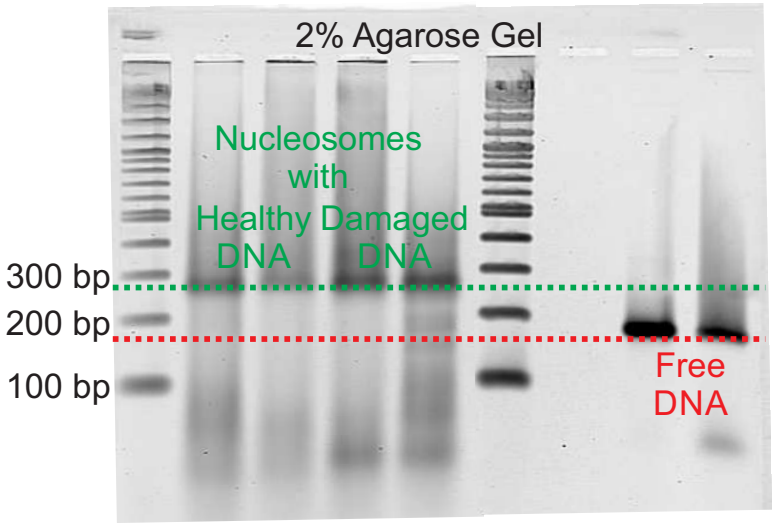


Figure 4.6: Nucleosome reconstitution of W601. This gel shift assay shows that the nucleosomes with both damaged and healthy DNA were successfully reconstituted and little free DNA remains. This lack of free DNA is evident since there is no significant band at the 170 bp location in the lanes containing nucleosomes. Free DNA has been included in the right two lanes as a control to show where free DNA (DNA not wrapped in the nucleosomes) would be located. Both healthy and damaged nucleosomes have a perceived length of $\approx$ 250 bp. Gel image taken in February 2014 in the Antonik Lab at the University of Kansas.

occur with this sequence.

4.2.3.2 5SrDNA sequence

Because of the high binding affinity of the W601 sequence, experiments were also done with the 208 bp 5SrDNA (shown in Table 4.6) which conveniently came as a control in the Nucleosome Assembly Kit. The 5SrDNA has lower binding affinity for the histone octamer (so less tightly bound to the histone octamers) and is 38 basepairs longer than the W601 sequence. This additional length may be needed for damaged DNA to be located in specific location relative to the histone octamer. I followed the Nucleosome Assembly Kit protocol for the 5SrDNA (control) to create the nucleosomes. I then put the nucleosomes into the reaction mix described in Table 4.7. Three different volumes of restriction enzyme were used and the incubation time was increased to 16 hours to account for the decreased ability to access the DNA wrapped in the nucleosome structure. As recommended in the manufacturer protocol, the samples were heat inactivated for 20 minutes

at 65°.

Table 4.6: 5SrDNA Sequence - 208 basepairs. 147 basepairs are wrapped in the nucleosomes structure. The basepairs underlined are not wrapped in the nucleosome as determined by nuScore. The "|" has been used to show the location of the nick from restriction enzyme Nt.bsmAI which targets the sequence in magenta. The dyad is the basepair in blue.

bp		Sequence
001-	5'	TATTCGTTGGAATTCCTCGGGGAATTCGGTATTCCCAGGCGGTCTCC|C
048	3'	ATAAGCAACCTTAAGGAGCCCCTTAAGCCATAAGGGTCCGCCAGAGG G
049-		ATCCAAGTACTAACCGAGCCCTATGCTGCTTGACTTCGGTGATCGGAC
096		TAGGTTCATGATTGGCTCGGGATACGACGAACTGAAGCCACTAGCCTG
097-		GAGAACCGGTATATTCAGCATGGTATGGTCGTAGGCTCTTGCTTGATG
144		CTATTGGCCATATTAGTCGTACCAGACCAGCATCCGAGAACGAACTAG
145-		AAAGTTAAGCTATTTAAAGGGTCAGGGATGTTATGACGTCATCGGCTT
192		TTTCAATTCGATAAATTTCCCAGTCCCTACAATACTGCAGTAGCCGAA
193-		ATAAATCCCTGGAAGT - 3'
208		TATTTAGGGACCTTCA - 5'

Table 4.7: Restriction enzyme added to nucleosomes with 5SrDNA reaction mix. Varying amounts of restriction enzyme were tested. Since the DNA is wrapped in the nucleosome structure, the enzyme concentration was increased. However, the enzyme process can be hindered by the amount of glycerol (it comes in 50% glycerol solution), so the incubation time was increased to 16 hours at 37°C to compensate.

Component	**Volume / Reaction**	**Comments**
NEBuffer4 (10X)	2.0 μL	
5SrDNA Nucleosomes	variable	0.5μg of DNA
H_20 (Molecular Grade)	variable	
Nt.BsmAI restriction enzyme	1.0, 2.0, or 4.0 μL	varied enzymes
Total Reaction Volume	20.0 μL	

The healthy DNA nucleosomes showed one band of nucleosomes while the nucleosomes treated with the restriction enzyme showed two bands (Figure 4.7). The two bands were located higher

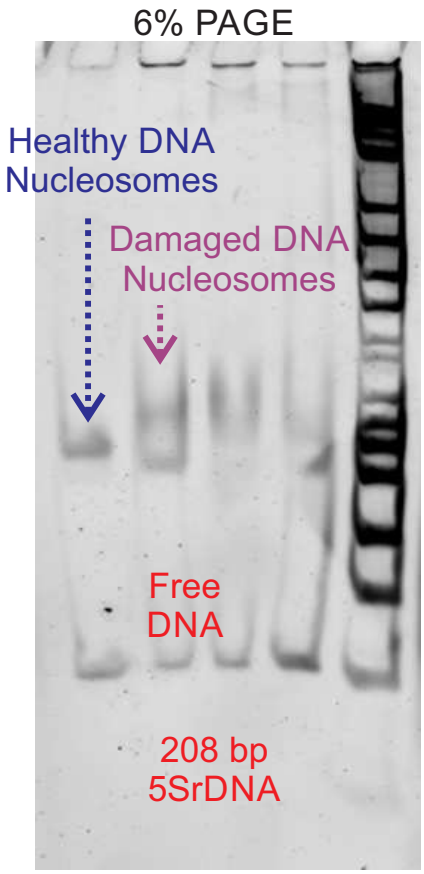


Figure 4.7: Nucleosome reconstitution of 5SrDNA. This is 6% polyacrylamide gel containing nucleosomes wrapped with the 208 bp 5SrDNA. The healthy DNA nucleosomes (left lane) have a single band. The damaged DNA nucleosomes (center three lanes) appear to have a band that shifted up and another band that shifted down. This suggests repositioning of the damage. The 208 bp free DNA in the nucleosome lanes is all located at about the same band. This suggests that the restriction enzyme did not significantly degrade the free DNA. Gel image taken in April 2014 in the Antonik Lab at the University of Kansas.

and lower in the gel than the nucleosomes with healthy DNA. The two bands suggest two different conformations and therefore possibly repositioning of the damage. If the damaged DNA was repositioned, this is consistent with the damaged DNA model in Chapter 3 that predicts damaged DNA has preferred locations.

4.3 Fluorescence resonance energy transfer (FRET)

The gel experiments of the 5SrDNA sequence suggested that damaged DNA may be repositioned. However, as with most ensemble experiments this gives the overall behavior of the system. To

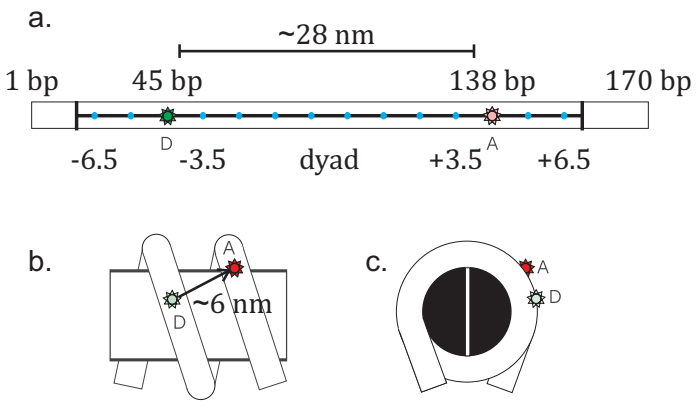


Figure 4.8: FRET using W601 sequence. For the FRET experiments using the W601 sequence, the green donor dye is located at basepair 45 and the red acceptor dye is located at basepair 138 (about 28 nm apart) as shown in (a). When the DNA is fully wrapped in the nucleosome, the dyes are about 6 nm apart as shown in (b) and the rotated view in (c) [1].

determine the possible behavior of the individual nucleosomes, single molecule FRET experiments can be used.

In my FRET experiments (Section 2.3), I used the 170 bp Widom 601 sequence with Alexa 488 (green - donor) and Alexa 594 (red - donor) dyes as shown in Figure 4.8 and used by Reference [1]. Using the Förster radius of ≈ 6 nm for Alexa 488 and Alexa 594 [1] in Equation 2.1, the FRET efficiency as a function of the distance is plotted in Figure 4.9. At the Förster radius $R_0 = 6$ nm the FRET efficiency is 50%. Dyes closer together than R_0 have a higher efficiency and dyes farther apart than R_0 have a lower efficiency.

In the lab where I was doing my experiments, the FRET efficiency was typically determined through the lifetimes associated with each nucleosome. The lifetime is the average amount of time the excited donor spends in the excited state before emitting a photon. The lifetime (τ) and the intensity I are related by $I(t) = \alpha e^{-t/\tau}$ where t is time and α is a normalization term [76]. The FRET efficiency can then be determined by

$$\text{FRET Efficiency} = 1 - \tau_{DA}/\tau_D \tag{4.6}$$

where the lifetime of the donor in the presence of the acceptor is τ_{DA} and in the absence of the acceptor is τ_D [16].

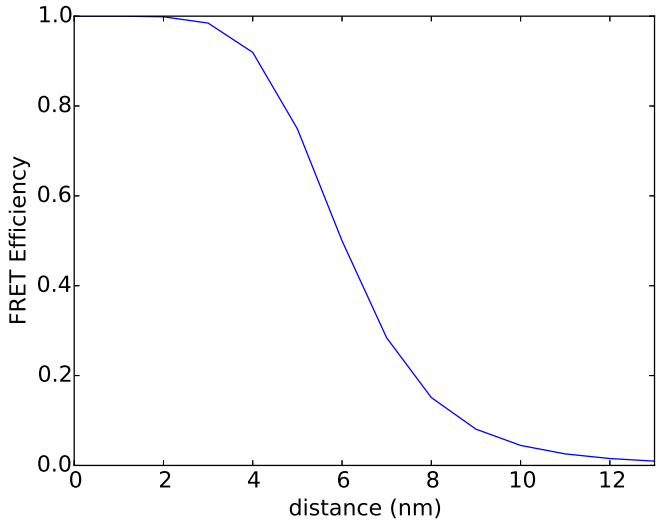


Figure 4.9: FRET efficiency when using Alexa 488 (green - donor) and Alexa 594 (red - acceptor) dyes. According to the Thermo Fischer Scientific Website the Förster radius of Alexa 488 and Alexa 594 is 6 nm. When the dyes are 6 nm apart the FRET efficiency is 50%. The efficiencies increase for shorter distances and decrease for longer distances.

Shown in Figure 4.10 is an example of two nucleosomes that will have different FRET efficiencies. The nucleosome on the left is fully wrapped with dyes close together. The nucleosome on the right is only partially wrapped with dyes farther apart. The nucleosome on the left will have a higher FRET efficiency than the nucleosome on the right.

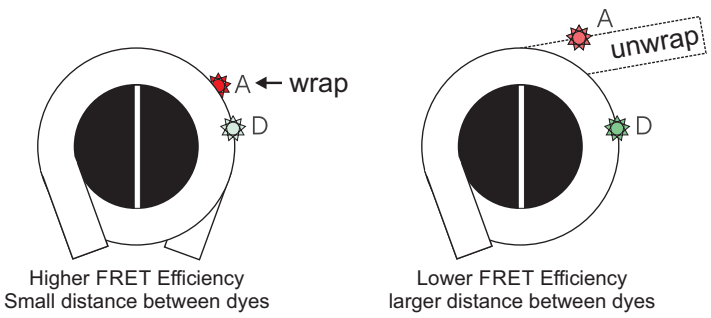


Figure 4.10: The two nucleosomes shown are in different states of wrapping. The nucleosome on the left is fully wrapped and the nucleosomes on the right is partially wrapped. On the left the dyes are close together and should have a high FRET efficiency. On the right the dyes are farther apart so there should be a smaller FRET efficiency.

4.3.1 Nucleosome breathing

Nucleosome breathing is a dynamic process, so experimentally there will be a number of different FRET efficiences. During the FRET experiments, data from individual nucleosomes are collected and analyzed to determine the FRET efficiency distributions. The estimates of the FRET efficiences will vary depending on the wrapping state of the each individual nucleosome. Different wrapping stages will have different distances between the dyes and therefore different FRET efficiencies as shown in Figure 4.10. This will result in a distribution of FRET efficiencies. Based on the damaged nucleosome breathing model, the FRET distribution for healthy DNA should be different from the FRET distribution of damaged DNA (a cartoon image of potential distributions shown in Figure 4.15). The damaged nucleosomes breathing model predicts repositioning of the damaged DNA which should result in repositioning of the fluorescent dyes. The repositioning of the dyes in the damaged DNA nucleosomes should result in a different FRET distribution than the healthy DNA nucleosomes.

4.4 Experimental setup

An example of experimental setup and equipment are shown in Figures 4.11 and 4.12. The beam from the Ti:Sapphire laser (Millennia Pro – 15sJ, Spectra Physics, CA) was frequency doubled to 485 nm using a beta barium borate (BBO) crystal (United Crystals, Port Washington, NY) before traveling into the microscope. The beam then excites only the donor (green) dye located in the focus of the microscope. The light from the laser beam then passes back through the microscope while the light from the green (donor) dye, and possibly the red (acceptor) dye if FRET has occurred, is reflected towards the photon detectors (Figure 4.11). Between the microscope and the photon detectors is a filter (Figure 4.12) that allows the red acceptor photons to pass and reflects the green donor light so that the photons go to their appropriate photon detectors. The arrival time of the photons are then detected by the appropriate avalanche photodiode (APD) detectors.

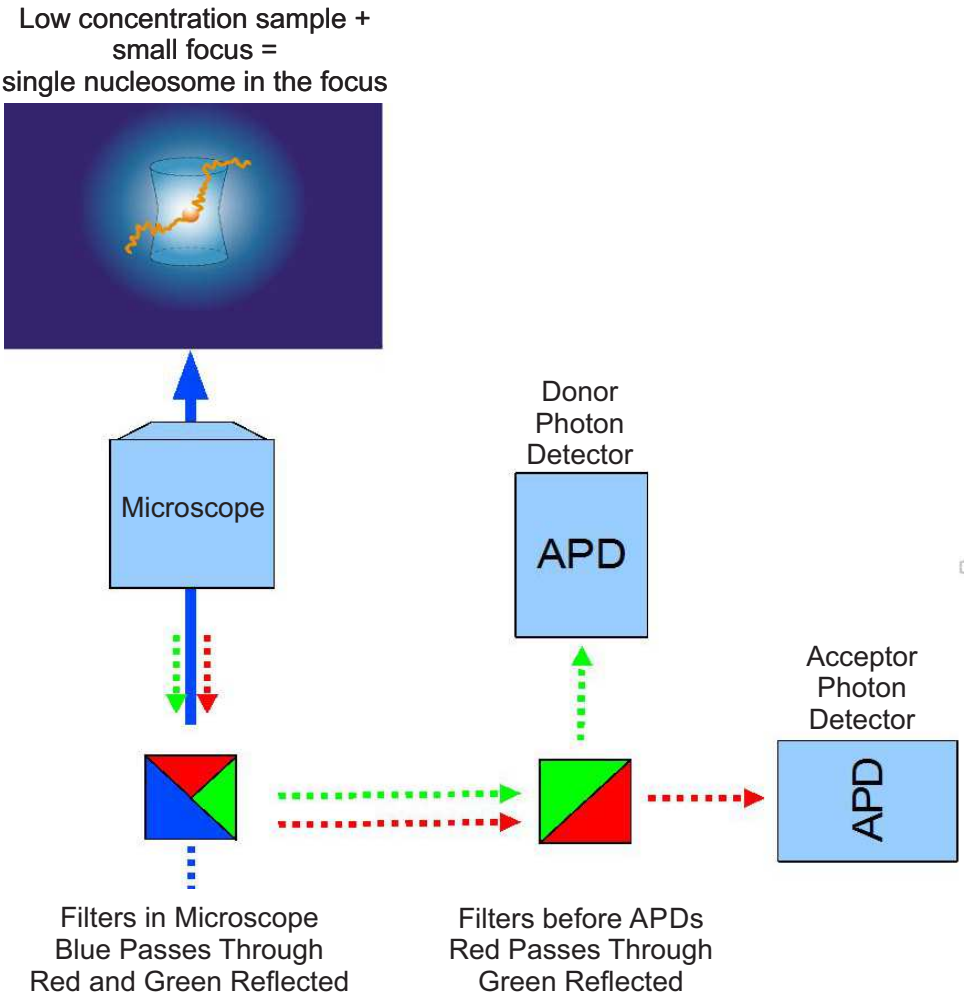


Figure 4.11: Experimental setup of the filters. The laser beam comes into the microscope at 485 nm to excite only the donor dye located in the focus of the microscope. If the donor and acceptor dye are in close enough proximity FRET occurs. The light from the sample is eventually passed through a filter to separate the donor (green) and acceptor (red) light into the appropriate detectors. The photons are then detected in the avalanche photodiode (APD) detectors.

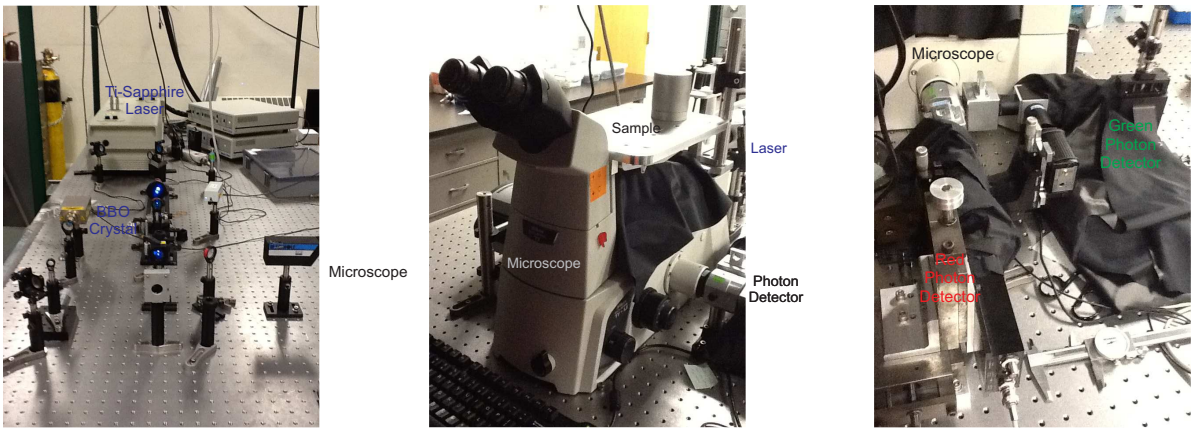


Figure 4.12: Equipment used for FRET experiments. The laser beam is frequency doubled to a wave length of 485 nm using a BBO crystal. The laser beam enters the microscope to excite only the donor (green) dye. Filters allow the light from the donor (green) and acceptor (red) dyes to enter the appropriate avalanche photodiode (APD) detectors.

4.5 Data analysis using custom software

The photon counting data (sample data shown in Figures 4.13 and 4.14) was collected on a SPC-150 TCSCP card (Becker & Hickl, Berlin, Germany) and analyzed using custom software written by Dr. Matthew Antonik (University of Kansas) in LabVIEW (National Instruments, Austin, TX). In Figure 4.13(a) is the raw data from the "counting" of the photons. The photon number (horizontal) and the length of time since the previous photon was collected (vertical). The software then converted to the raw data into binned data. Figure 4.13(b) and (c) shows the time (horizontal) and the number of photons that arrived at that time (vertical). Graph (c) is an enlargement of (b) showing a sample of single molecule concentration. To study the FRET that occurs with each individual nucleosome, the concentration of the fluorescently labeled nucleosomes must be low enough that only one nuleosome is in the microscope focus at a time. The background of the system without any nucleosomes present is ≈ 10 photons. In Figure 4.13(c) there are times when there are no nucleosomes in the focus (so about 10 photons are counted due to the background) and there are times when a nucleosome is in the focus with peaks as high as 130 photons.

The custom software also produces the graphs as in Figure 4.14. The data in the left graph is the number of occurrences for a specific intensity (total number) of photons. For example, in the

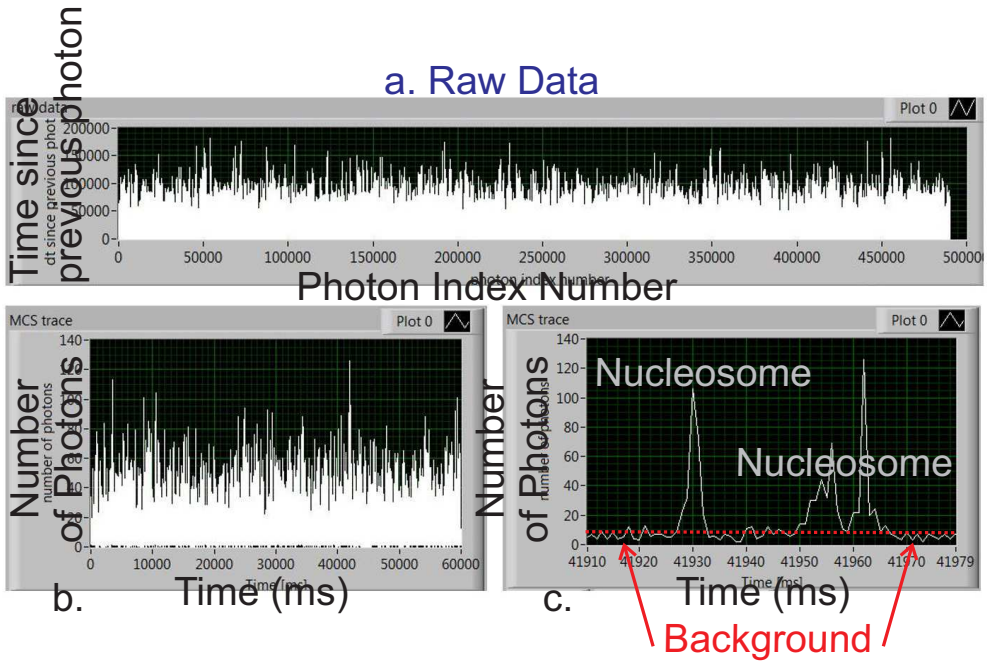


Figure 4.13: The original data (a) collected on the TCSCP card as the photon index number and the time since the previous photon arrived. This data must be modified in order to obtain the ultimate goal of determining the FRET of each nucleosome that passes into the focus. The data in graphs (b) and (c) shows the number of photons as a function of time. Specifically in (c) during the time window shown, two (or three) nucleosomes passed through the focus since the photon count was much higher than the ≈ 10 photons due to the background of the system.

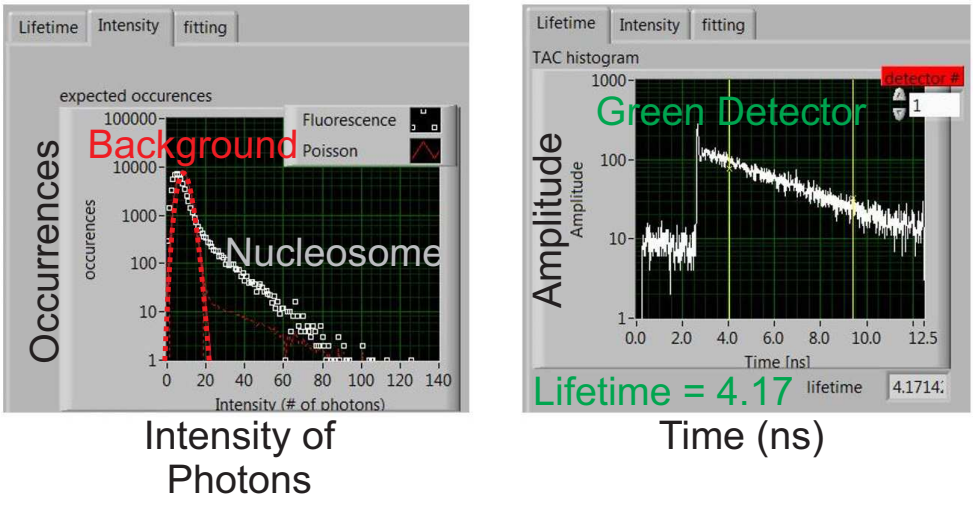


Figure 4.14: More sample data from nucleosome experiments. In the left graph is the number of times that a given number of photons were collected. Approximately 60 photons were collected from about 10 nucleosomes. Most of the occurrences had 15 or fewer photons, due to the background of the system. The high amount of background corresponds to having concentration that is low enough that when the intensity is above the background, the photons detected from a single molecule in the sample. In the graph at right, the estimated lifetime of this particular sample is 4.17 ns.

sample data in Figure 4.14, a total of 60 photons were detected 10 different times. The maximum number of occurrences is for about 7 photons due to "background" of the system which is Poisson Distributed (red curve). The indication that there is fluorescence in the sample is the fact that the photons intensity (white boxes) is higher than the background. If there was only background there would never be more than about 20 photons counted at any given time. The data in the right graph of Figure 4.14 shows the estimate for the average lifetime of the sample collected in the green (donor) detector. This sample was the DNA only (no nucleosomes) shown in Figure 4.1 where the donor and acceptor dyes are ≈ 28 nm apart. The lifetime of Alexa 488 (green) donor dye in the absence of an acceptor dye is about 4.1 ns. Since the data from this sample was roughly estimated to be 4.17 ns, this suggests that no FRET has occurred. The lack of FRET in this sample is expected since the dyes are ≈ 28 nm apart which is well past the range at which FRET occurs. If the lifetime of the sample was estimated to be less than 4.1 ns, this would be an indication that FRET occurred. When the DNA is wrapped in the nucleosome structure, the dyes are closer together so there should be a decrease in the lifetime of the donor dye and an increase in the FRET

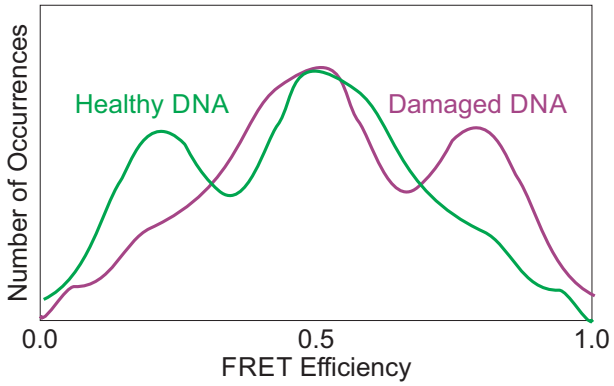


Figure 4.15: Cartoon image of two different FRET distributions. The nucleosomes that contain damaged DNA are shown to have a different distribution of FRET efficiencies than nucleosomes containing healthy DNA. Differences in the FRET distributions would suggest repositioning of the damage.

efficiency.

4.6 Conclusions and future work

The gel image in (Figure 4.7) is from a reaction that used the restriction enzyme Nt.bsmAI on the nucleosomes with 5SrDNA sequence. The single band in the healthy DNA nucleosomes and the two bands in the damaged DNA nucleosomes suggest that repositioning of the DNA damage may have occurred. This repositioning of the damaged DNA is consistent with the damaged DNA model (Chapter 3).

Single molecule FRET experiments are another approach to studying the repositioning of damaged DNA. The FRET distribution for nucleosomes with healthy DNA has been previously determined [1]. I had hoped to use FRET experiments to determine the distribution of the FRET efficiencies of nucleosomes with and without damaged DNA. If the results indicated different distributions (see cartoon image in Figure 4.15), that would suggest that the nucleosomes have different conformations. Different FRET distributions would imply that the nucleosomes that contained damaged DNA have repositioned the damage in some way. However, I was unable to complete these experiments and leave it as possible future work.

Chapter 5

Translocation kinetics of molecular motors

5.1 Introduction

Many "molecular motor" proteins share an ability to translocate processively and in a directionally biased motion along a template (*e,g.* DNA, polypeptides, *etc.*). Some types of motor proteins include polymerases [77], chromatin remodelers [8, 78], and some restriction enzymes [79–81]. In order to translocate along the template, motor proteins use the chemical potential energy obtained from binding and hydrolysis of nucleotide triphosphates (NTP) and in particular adenosine triphosphate (ATP). Shown in Figure 5.1 is an example of translocation. In Figure 5.1, the protein translocates (moves) towards the end of the template at a translocation rate of k_t and separates from the template at a dissociation rate of k_d except at the end where the dissociation rate is k_0. Note that all rates are in units s^{-1}. The kinetics parameters associated with this system include the translocation rate, dissociation rate, processivity ($P \equiv \frac{k_t}{k_t+k_d}$ which is ratio of the rate of translocation to the rate of total motion (translocation and dissociation)), and kinetic step size (distance between steps that occur at the translocation rate k_t) and are all needed to obtain an understanding of the mechanisms involved in the translocation process. A variety of techniques has been used to study the translocation process and determine the kinetic parameters including ensemble and single molecule experiments [82–89].

My approach to gaining information about the translocation process focuses on the sequential

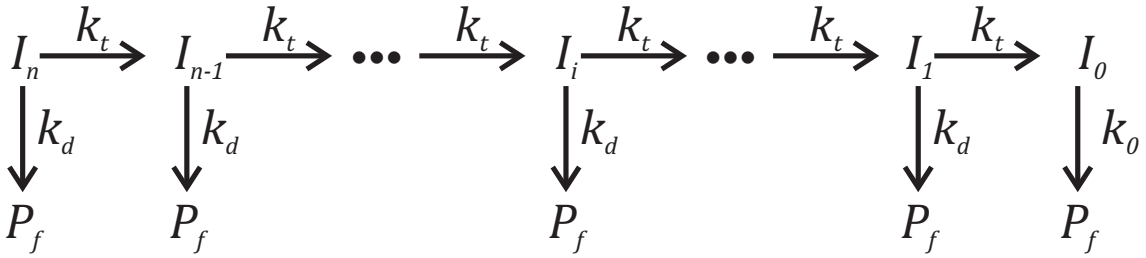


Figure 5.1: Generic scheme for translocation of a protein along a template. The protein concentration at state i is I_i and i is defined as the number of translocation steps from the end. The constraint on i is that $0 \leq i \leq n$ where n is the maximum number of steps from one end to the other end. The proteins translocate at a rate of k_t (unit of s^{-1}) towards the end. The proteins dissociate from the template at a rate of k_d except at the end where dissociation occurs at a rate of k_0. Upon dissociation, the free proteins (P_f) bind to a protein trap which prevents rebinding to the DNA.

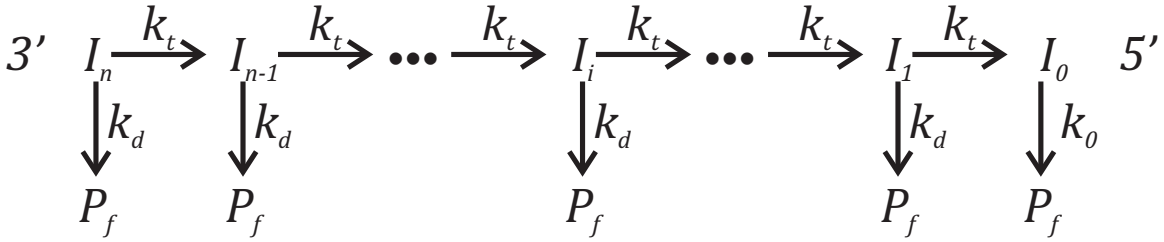


Figure 5.2: Scheme for translocation of the protein along DNA. The protein concentration at state i is I_i and i is defined as the number of translocation steps from the 5’ end of the DNA. The constraint on i is that $0 \leq i \leq n$ where n is the maximum number of steps to get from the 3’ end of the DNA to the 5’ end of the DNA. The proteins translocate at a rate of k_t (unit of s^{-1}) towards the 5’ end. The proteins dissociate from the DNA at a rate of k_d except at the 5’ end where dissociation occurs at a rate of k_0. Upon dissociation, the free proteins (P_f) bind to a protein trap which prevents rebinding to the DNA.

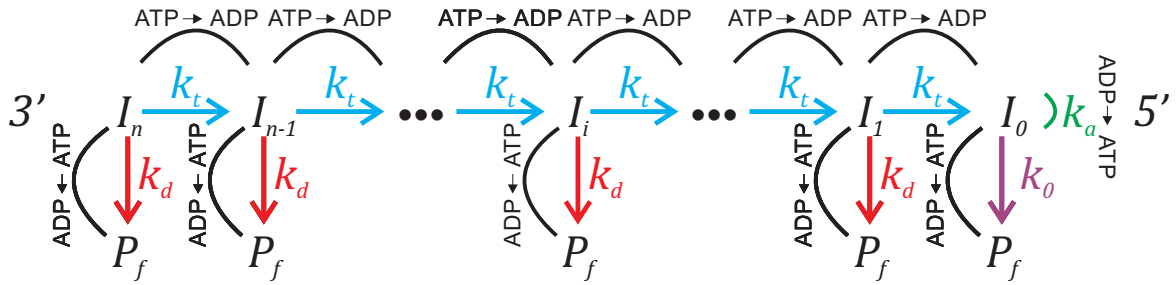


Figure 5.3: Scheme for translocation of the protein and the ATP hydrolysis for the *n-step*. As denoted by the ATP → ADP (described in Figure 2.7), both translocation and dissociation require the energy from the hydrolysis of ATP. In addition, there is the potential for futile hydrolysis at the 5' end. This occurs at a rate of k_a and is not coupled to any movement of the protein.

"*n-step*" model [8, 80, 81, 87, 90–95]. Shown in Figures 5.2 and 5.3 is the sample scheme for the *n-step* model for a DNA translocation protein that moves from 3' to 5' along its nucleic acid template. There are two important things to note about this *n-step* model: translocation is uniform in that it always occurs at rate of k_t and the translocation is directionally biased along the DNA (towards the 5' end in this scheme). This means that any type of non-uniform motion (such as backward motion or occasional pausing) is explicitly neglected from this *n-step* model [8, 87, 92, 94, 96].

However, both ensemble and single molecule experiments have suggested non-uniform motion of the *E. coli* UvrD translocase [91, 96, 97]. Shown in Figure 5.4 is the assumed translocation process from the ensemble experiments described in Reference [96]. The protein undergoes four fast translocation steps followed by a much slower translocation step. The rate at which the protein moves along the DNA is therefore limited by the slow step. The single molecule studies have also suggested that persistent heterogenity (described in Section 5.4.5) of the kinetic translocation rates may occur [91].

Therefore one question to consider is how the fitted kinetic parameters in the *n-step* model are impacted when the data is from proteins that exhibit non-uniform motion. Often, it is not known *a priori* whether non-uniform motion occurs. In order to investigate the impact on the kinetic parameters due to the protein's non-uniform motion (*e.g.* backward motion, random pausing, jumping, and heterogenities in the step size and the translocation rate), I began by using Monte

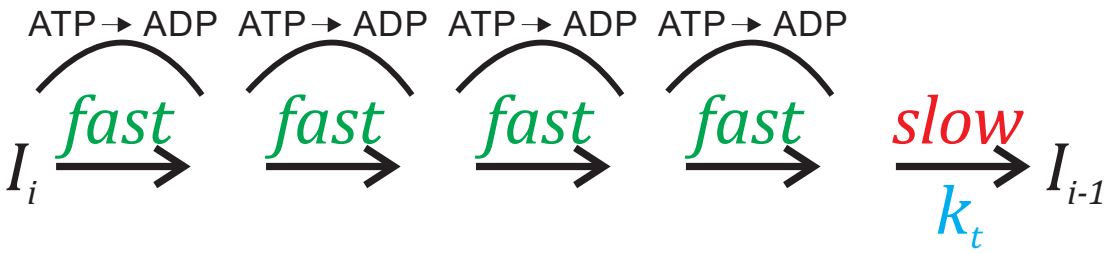


Figure 5.4: Partial scheme for translocation of a protein called *E. coli* UvrD translocase. When translocating from state i to state $i-1$, the protein takes four fast steps each requiring hydrolysis of one ATP molecule, following by one slow step that takes much longer. This slow step is the limited rate that constraints the overall rate of translocation k_t.

Carlo methods to simulate the time course data associated with proteins moving with different types of non-uniform motion.

The simulated time course data was then fit to estimate the kinetic parameters of a *n-step* model. The non linear least squares (NLLS) method used the time course data points from simulations that had non-uniform motion and fit to a model that explicitly neglects non-uniform motion. Through this analysis, I was able to determine how each type of non-uniform motion impacts the estimated kinetic parameters of the *n-step* model. Results of the analysis of the simulated time course data indicate that estimates for both the macroscopic translocation rate (how fast the protein is moving) and the macroscopic ATP coupling stoichiometry (how much energy is being used to move the protein) are reliably obtained using the *n-step* model. However, the microscopic parameters (*e.g.* the kinetic step size m) tended to be over or under estimated. It is also possible in some cases to predict that non-uniform motion of the protein may be occurring based on the kinetic parameters determined using the *n-step* model.

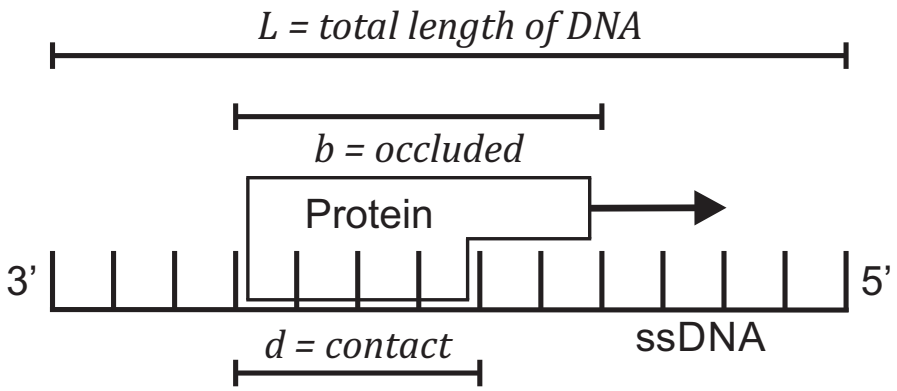


Figure 5.5: Model of the protein bound to the DNA. A cartoon depicting the binding of a protein with a contact size d and occluded site size b to a single stranded DNA of length L. As shown here, the contact size, d, is always less than or equal to the occluded site size, b. The parameters d, b, and L are measured in nucleotides. The protein is initially randomly bound with polarity and will move towards the 5' end.

5.2 *N-step* model for translocation

5.2.1 Bound protein model

The protein was assumed to be bound to single stranded DNA as shown in Figure 5.5. The protein binds with polarity to the ssDNA with a contact size of d which is the number of consecutive nucleotides required to maintain contact with the DNA. Depending on the protein, it is possible to have an occluded site of b nucleotides which is greater than or equal to the contact size d. The protein is assumed to utilize its full contact size even when bound to the ends of the DNA (*i.e.*, no dangling protein) [92]. The total length of the DNA is L nucleotides.

5.2.2 Kinetic motion model

The motion of the protein along the DNA for the *n-step* model is shown in Figures 5.2 and 5.3. Note that in the *n-step* model described and shown in Figures 5.2 and 5.3, translocation is assumed to be directionally biased from 3' to 5'. However, the results are equally valid for translocation directionally biased from 5' to 3'. The protein binds randomly to the DNA at state i which is located i translocation steps way from the 5' end which is state $i = 0$ or the end. A single translocation step

(from state i to state $i-1$) moves the protein a distance of m nucleotides where m is defined as the kinetic step size. A maximum of n translocation steps is needed for a protein initially bound at the 3' end to reach the 5' end of DNA with length L nucleotides. This means that i is constrained such that $0 \leq i \leq n$.

The relationship between the length of the DNA L, the maximum number of translocation steps n, the contact size d, and the kinetic step size m can be expressed as

$$L = mn + d \tag{5.1}$$

or

$$n = \frac{L}{m} - \frac{d}{m} \tag{5.2}$$

where L and d are units of nucleotides. The parameter n is in units of steps and m is units of nucleotides per step. As an example, if $L = 17$ nucleotides, $m = 3$ nucleotides, and $d = 2$ nucleotides, then $n = 17/3 - 2/3 = 5$ using Equation 5.1 or 5.2. Assuming the protein binds initially 17 nucleotides from the end, when the protein takes 5 steps of 3 nucleotides each, the protein has moved a total of 15 nucleotides. However, the protein is now located 2 nucleotides from the end. Therefore, this protein is located zero steps from the end, because the distance to the end is less than the step size m. For simplicity, I assumed motion of a protein along single stranded DNA (ssDNA) as shown in Figure 5.5. However, the results are general and would apply to any kinetic processes which display sequential "*n-step*" kinetic mechanisms such as double stranded DNA (dsDNA) unwinding, protein translocation along dsDNA, and polypeptide translocases.

Using the energy from hydrolysis of ATP, the proteins translocate with directional bias at a rate k_t or dissociate from the protein at a rate k_d except at the end where the dissociation rate is k_0 (rates are in units of s^{-1}). The processivity of the protein $P \equiv \frac{k_t}{k_d + k_t}$ is the probability of taking forward steps rather than dissociation. The macroscopic ATP coupling stoichiometry is c/m and is a result of assuming that c molecules of ATP are hydrolyzed for each rate limiting step (which moves the protein from state i to $i-1$ and a distance of m nucleotides). In other words, c/m is the

number of ATP molecules hydrolyzed per nucleotide translocated (*e.g.* If 10 ATP are hydrolyzed to translocate the protein from state i to $i-1$ and there are 2 nucleotides between state i to $i-1$, then $c/m = 5$ ATP are hydrolyzed for each nucleotides translocated). Having moved m nucleotides (nt) during each translocation step, the macroscopic translocation rate is mk_t in units of nt/s (*e.g.* $m = 2$ nt, and the kinetic translocation rate at which the protein translocates from state i to state $i-1$ is $k_t = 60s^{-1}$ then the macroscopic translocation rate is $mk_t = 120$ nt/s). These macroscopic values are a measure of how energy efficient the protein is (c/m) and how fast the protein is moving towards the end (mk_t).

Once the protein has reached the end, it will dissociate from the DNA at a rate of k_0 in units of s^{-1}. While the protein is at the end, ATP hydrolysis that does not result in motion of the protein may occur. This is referred to as futile hydrolysis and occurs at the rate of k_a (Figure 5.3) [82, 96]. Inclusion of a protein trap prevents any free protein that dissociates from rebinding to the DNA [87, 92, 96].

5.2.3 Time dependence of protein concentration and ATP hydrolysis

The time dependence of the protein concentration I_i located at state i for the *n-step* model in Figures 5.3 is shown in the coupled differential Equations 5.3 - 5.5 [92, 94]. In Equations 5.3 - 5.5, I_n is the concentration of protein at the 3' end located n translocations steps for the 5' end, I_i where $1 \leq i \leq n$ is the concentration of proteins located i translocation steps from the 5' end, and I_0 is the concentration of proteins at the 5' end. Note that once dissociation occurs, the protein is assumed not be rebind to the template (*e.g.* DNA).

$$\frac{d}{dt}I_n(t) = -(k_t + k_d)I_n(t) \tag{5.3}$$

$$\frac{d}{dt}I_{1\leq i<n}(t) = -(k_t + k_d)I_i(t) + k_t I_{i+1}(t) \tag{5.4}$$

$$\frac{d}{dt}I_0(t) = -k_0 I_0(t) + k_t I_1(t) \tag{5.5}$$

In Equation 5.3, the change in concentration of proteins I_n depends only on the rate at which the proteins are leaving the state n by either translocation (at rate k_t) or dissociation (at rate k_d). In Equation 5.4, the change in the concentration of proteins I_i at state i (where $1 \leq i < n$) depends not only on the rates at which the proteins are leaving state i due to translocation (k_t) and dissociation (k_t), but also on the rate at which the proteins are leaving state $i+1$ and arriving at state i due to translocation. In Equation 5.5, the change in concentration of proteins I_0 depends on the proteins leaving the state 0 at the dissociation rate (k_0) and on the proteins arriving from the state 1 at the translocation rate (k_t). In other words, the concentration of the proteins at a particular state depends on the rate at which the proteins arrive from another state, or leave due to translocation or dissociation. Equations 5.3 - 5.5 can be transformed using a Laplace Transformation into

$$I_n(t) = \mathscr{L}^{-1}\left[\frac{I_n(0)}{s+k_t+k_d}\right] \tag{5.6}$$

$$I_{1\leq i<n}(t) = \mathscr{L}^{-1}\left[\frac{1}{s+k_t+k_d}\left(\sum_{j=i}^{n} I_j(0)\left[\frac{k_t}{s+k_t+k_d}\right]^{j-i}\right)\right] \tag{5.7}$$

$$I_0(t) = \mathscr{L}^{-1}\left[\frac{1}{s+k_0}\left(I_0(0)+\sum_{i=1}^{n} I_i(0)\left[\frac{k_t}{s+k_t+k_d}\right]^{i}\right)\right] \tag{5.8}$$

where $\mathscr{L}^{-1}$ is the Laplace transform operator and s is the Laplace variable [94]. $I_0(0)$, $I_i(0)$, and $I_n(0)$ represent the initial concentration of the protein initially bound at state i at time $t = 0$. Assuming there is an equal probability of the protein binding at any state, $I_0(0) = I_i(0) = I_n(0) = I(0)$, where $I(0)$ is defined to be the concentration at any state at $t = 0$. By assuming that n is no longer constrained as an integer as done in References [87, 92], Equations 5.6 - 5.8 can be approximated as the following

$$I_n(t) = \mathscr{L}^{-1}\left[\frac{rI(0)}{(s+k_t+k_d)(1+nr)}\right] \tag{5.9}$$

$$I_{1\leq i<n}(t) = \mathscr{L}^{-1}\left[\frac{rI(0)}{(1+nr)}\left(\frac{(s+k_t+k_d)-k_t\left(\frac{k_t}{s+k_t+k_d}\right)^{n-i}}{(s+k_d)(s+k_t+k_d)}\right)\right] \tag{5.10}$$

$$I_0(t) = \mathscr{L}^{-1}\left[\frac{I(0)}{(1+nr)(s+k_0)}\left(1+\frac{k_t r}{(s+k_d)}\left(1-\left(\frac{k_t}{s+k_t+k_d}\right)^n\right)\right)\right] \tag{5.11}$$

where $r \equiv p/p_0$ is the ratio of the probability that the protein binds to any state i where $1 \leq i \leq n$ divided by the probability that the protein binds at the end or state $i = 0$ [87, 92]. Note the following non-integer value example for the number of states n. Assume that a protein moves a total length L = 5 nucleotides along the DNA, with kinetic step size m = 3 nucleotides and contact size d = 1 nucleotide. Then by Equation 5.2, the number of states is $n = (5-1)/3 = 4/3$. In this case, depending on the initial binding location of the protein, one or two steps of length 3 nucleotides may be required for the protein to reach the end.

Equation 5.11 can then be re-expressed as Equations 5.12.

$$f_0(t) = \frac{I(0)}{1+nr} \times \mathscr{L}^{-1}\left[\frac{1}{k_0+s}\left(1+\frac{k_t r}{k_d+s}\left(1-\left(\frac{k_t}{k_d+k_t+s}\right)^n\right)\right)\right] \tag{5.12}$$

In Equation 5.12, $f_0(t)$ is the concentration of the protein at the 5' end and can be determined experimentally. The concentration $I(0)$ can be changed to the scalar A to convert the concentration of the protein at the 5' end into a signal that can be measured experimentally (*e.g.* a spectroscopic change) [87, 92, 96]. For simulated proteins, the concentration $f_0(t)$ is equivalent to the fraction of the total proteins (concentration) at the 5' end.

The time dependence of the ADP produced from ATP hydrolysis for the *n-step* model in Figure 5.3 is

$$\frac{d}{dt}[ADP(t)] = c(k_d+k_t)\sum_{i=1}^{n} I_i(t) + (ck_0+k_a)I_0(t) \tag{5.13}$$

where the ATP hydrolysis that results in translocation is

$$ck_t\sum_{i=1}^{n} I_i(t)$$

and $k_a I_0(t)$ is the ATP hydrolysis that occurs at the end but does not result in translocation (futile hydrolysis k_a is assumed to have a single ATP hydrolyzed at rate k_a) [92, 94]. This rate of futile hydrolysis k_a that occurs at the 5' end will also add to the concentration of ADP. The terms

$$ck_d \sum_{i=1}^{n} I_i(t) + ck_0 I_0(t) \tag{5.14}$$

in Equation 5.14 are due to the ATP hydrolysis that results in dissociation of the protein.

The Laplace transformation of Equation 5.13 is

$$[ADP(t)] = \mathscr{L}^{-1}\Big[\frac{1}{s}\Big(c(k_d+k_t)\sum_{i=1}^{n}\mathscr{L}(I_i(t)) + (ck_0+k_a)\mathscr{L}(I_0(t))\Big)\Big] \tag{5.15}$$

and by substituting in Equations 5.9 - 5.11 can be approximated as Equation 5.16

$$\begin{aligned}[ADP(t)] = \mathscr{L}^{-1}\Bigg[\frac{I(0)}{s(1+nr)}\Bigg(&\frac{rc(k_d+k_t)\Big(n(s+k_d)+k_t\Big(\Big(\frac{k_t}{s+k_t+k_d}\Big)^n-1\Big)\Big)}{(s+k_d)^2}+\\ &\frac{(ck_0+k_a)}{s+k_0}\Big(1+\frac{rk_t}{s+k_d}\Big(1-\Big(\frac{k_t}{s+k_t+k_d}\Big)^n\Big)\Big)\Bigg)\Bigg]\end{aligned} \tag{5.16}$$

to approximate the concentration of ADP at time t. The concentration of ADP can be determined experimentally.

The simulated time course data for the protein translocation and ATP hydrolysis (as described in Section 5.3) were then used to estimate the *n-step* model kinetic parameters using Equation 5.12 and Equation 5.16.

5.3 Materials and methods

Monte Carlo simulations of proteins with non-uniform motion were used to generate time course data for the fraction (concentration) of the protein at the 5' end (I_0) and the concentration of ADP resulting from ATP hydrolysis occuring during translocation. An example of non-uniform motion that occurs at rate k_b in the backwards direction is depicted in Figure 5.6 with sample simulated

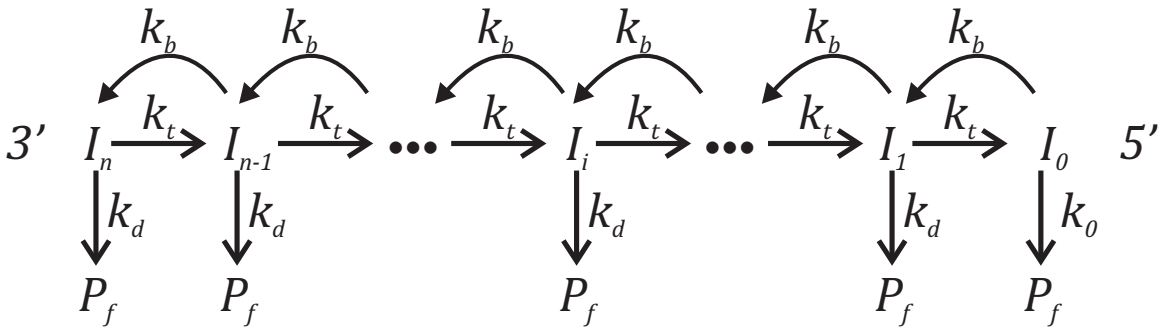


Figure 5.6: Scheme for translocation of the protein with backwards motion. The time course data from Monte Carlo simulations of a protein that has backward motion (5' to 3') at a rate of k_b nt/step with a step size of m_b nt/step. The analysis of the time courses was performed assuming the *n-step* model as in Figure 5.3.

time course data of the proteins in Figure 5.7 and of the ADP concentration in Figure 5.8. The simulated time course data was then used to estimate the kinetic parameters in the "*n-step*" model (Section 5.2) [92, 96].

Specifically, the time course data for the protein translocation (sample for backward motion in Figure 5.7) was fit to Equation 5.12. Using the kinetic parameters found for Equation 5.12, the time course data for the ADP concentrations (sample for backward motion in Figure 5.8) was fit to Equation 5.16 to estimate parameters c/m and k_a.

5.3.1 Monte Carlo simulations

Monte Carlo simulation programs were written using the Spyder Integrated Development Environment (IDE) with Python 2.7.9 32bits, Qt 4.8.4, PyQt4 (API v2) 4.9.6 on a Windows 10 workstation. The kinetics parameters for simulations with uniform motion are: $k_{t,0}$ = 60 step/s, $k_{d,0} = 1s^{-1}$, $k_{0,0} = 30s^{-1}$ and m_0 = 1 nt/step. Using other kinetic parameters for k_t, k_d and k_0 produced qualitatively similar results (data not shown). Time courses were simulated for 5 lengths of DNA (6 nt, 11 nt, 16 nt, 21 nt, 26 nt).

In all simulations the motion of the protein was determined by the probability of a particular outcome occurring (forward motion, backward motion, jumping, dissociation, *etc.*). The probabil-

ity of a particular outcome was determined from the microscopic rate constants ($k_t, k_d, k_0, etc.$). The relationship between the two sets of parameters is (in the case of backward motion) $k_b/k_t = p_b/p_t$.

For all the Monte Carlo simulations, assumptions were made regarding the kinetic rate parameters. The first assumption is the total rate of the ATP hydrolysis is held constant, which then constrains the rate of the motions to also be constant. This occurs since ATP hydrolysis gives the proteins the required energy to translocate or dissociate. The first assumption can be described mathematically as

$$k_{d,0} + k_{t,0} = k_d + k_t + k_\theta \tag{5.17}$$

where the presence of the subscript 0 is to denote uniform motion and the absence of the subscript is to denote non-uniform motion. Specifically, the term k_θ is the rate at which the non-uniform motion θ occurs where $\theta = b$ (backward motion), j (jumping), b, j (backward and jumping), p (random pausing), *etc*. The second assumption is that the ratio of the dissociation rate to translocation rate is held constant, which keeps the ratio of the probability of translocation to probability of dissociation equivalent. This is achieved by constraining the kinetic rate parameters so that

$$k_{d,0}/k_{t,0} = k_d/k_t \tag{5.18}$$

and the ratio of the rates of dissociation to translocation with uniform motion (subscript 0) is equal to the ratio of the rates of dissociation to translocation with non-uniform motion (absence of subscript 0).

An example of the kinetic rate constants used for the Monte Carlo simulations of proteins simulated with backward motion are shown in Table 5.1. Notice that the kinetic parameters fulfill the constraints shown in Equations 5.17 and 5.18. The kinetic parameters $k_{t,0} = 60$ nt/step, $k_{d,0} = 1s^{-1}$ were used the case of uniform motion. With backwards motion, the constraints are $k_d + k_t + k_b = 61$ by Equation 5.17 and $k_d/k_t = 1/60$ by Equation 5.18. When the ratio of $k_b/k_t = p_b/p_t = 0.2$, the translocation rate is $k_t = 50.14$, the dissociation rate $k_d = 0.84$, and the backward motion rate is $k_b = 10.03$, so that the two constraints are fulfilled. The time course data using the

parameters from Table 5.1 with $n = 5$ is shown in Figures 5.7 and 5.8.

Table 5.1: Sample of kinetic parameters for backward motion. The parameters are constrained by Equations 5.17 and 5.18. The columns sum to 61 and $k_d/k_t = 1/60$. Notice that when $p_b/p_t = 0$ there is no backward motion and therefore the motion of the protein is uniform.

$k_b/k_t = p_b/p_t$	0.0	0.1	0.2	0.3	0.4
k_t	60.00	54.63	50.14	46.33	43.06
k_d	1.00	0.91	0.84	0.77	0.72
k_b	0.00	5.46	10.03	13.90	17.22

5.3.2 Sample data from Monte Carlo simulations

5.3.2.1 Fraction of protein located at I_0

The Monte Carlo simulations included time course data for 5 lengths of DNA (6 nt, 11 nt, 16 nt, 21 nt, 26 nt). Displayed in Figure 5.7 is the simulated time course data for $n = 5$ (6 states) using the rates from Table 5.1. Assuming a uniform distribution, the proteins bind to random locations on the DNA, with $1/(n+1)$ of the total concentration located at I_0 when time $t = 0$. As time increases, the concentration of protein located at I_0 eventually decreases to ≈ 0 due to all proteins having dissociated from the DNA. For the case of uniform motion where $p_b/p_t = 0.0$, or no backward motion occurs, the concentration of protein increases to a maximum of ≈ 0.27 with all proteins having dissociated by approximately ≈ 0.3 seconds. For the case of non-uniform motion where $p_b/p_t = 0.2$, the concentration of protein increases to a maximum of ≈ 0.19 with all proteins having dissociated by ≈ 0.45 seconds. For the case of non-uniform motion where $p_b/p_t = 0.4$, the concentration begins at $\approx 1/6$ and decreases with all proteins having dissociated by ≈ 0.6 seconds. As expected, with increasing probability of backward motion to forward motion, the average length of time required for the protein to have dissociated (or equivalently for the concentration at I_0 to have been reduced to ≈ 0) increases.

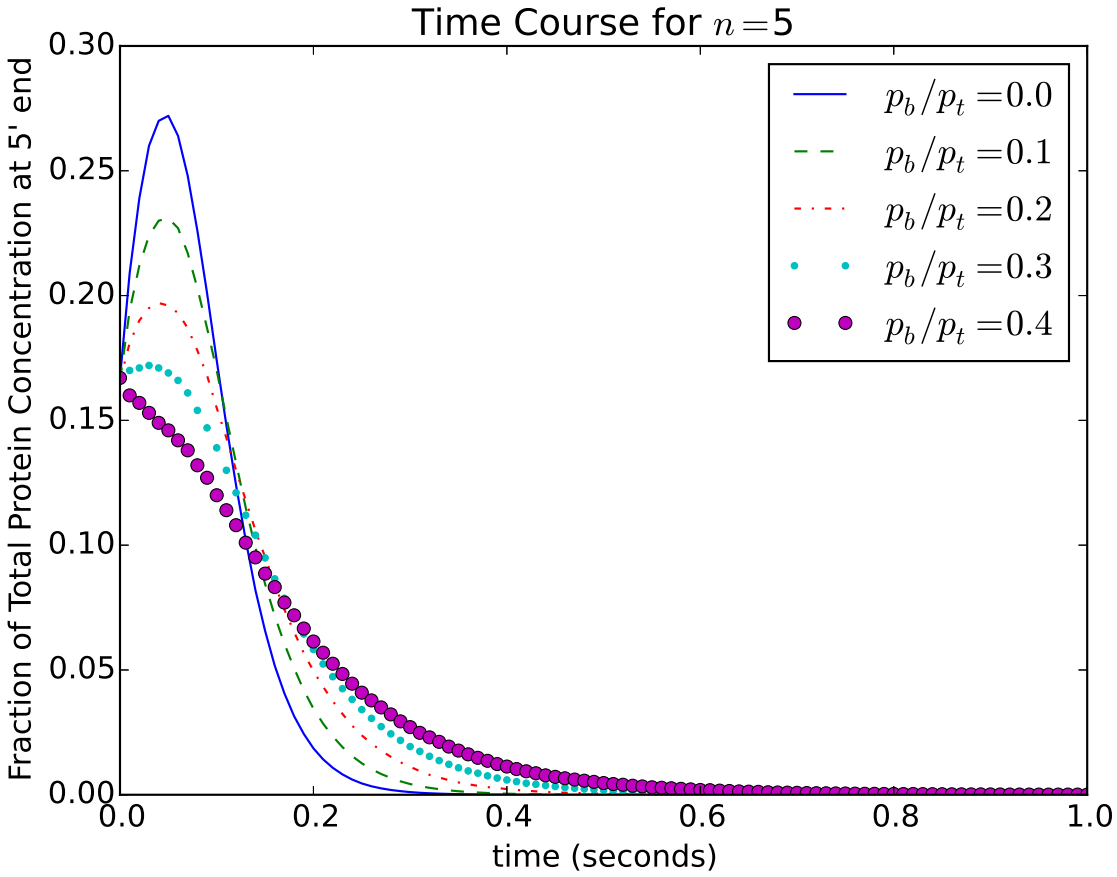


Figure 5.7: Monte Carlo simulated protein concentration time course data for backward motion for n = 5. This shows the fraction of the total protein concentration located at the 5' end (state $i = 0$) as a function of time. The kinetic parameters used to generate the data are shown in Table 5.1. Since the proteins are uniformly randomly bound, approximately 1/(n+1) or 1/6 (for n = 5) of the proteins are initially located at the 5' end at time $t = 0$. As the ratio of backward motion to translocation increases, there is an increase in the amount of time needed for the fraction of total protein concentration located at the 5' end (state $i = 0$) to be reduced to zero. As the probability of backward motion increases (p_b/p_t = 0.4) there is no longer an accumulation of protein at the 5' end before all the protein have dissociated and the concentration goes to zero.

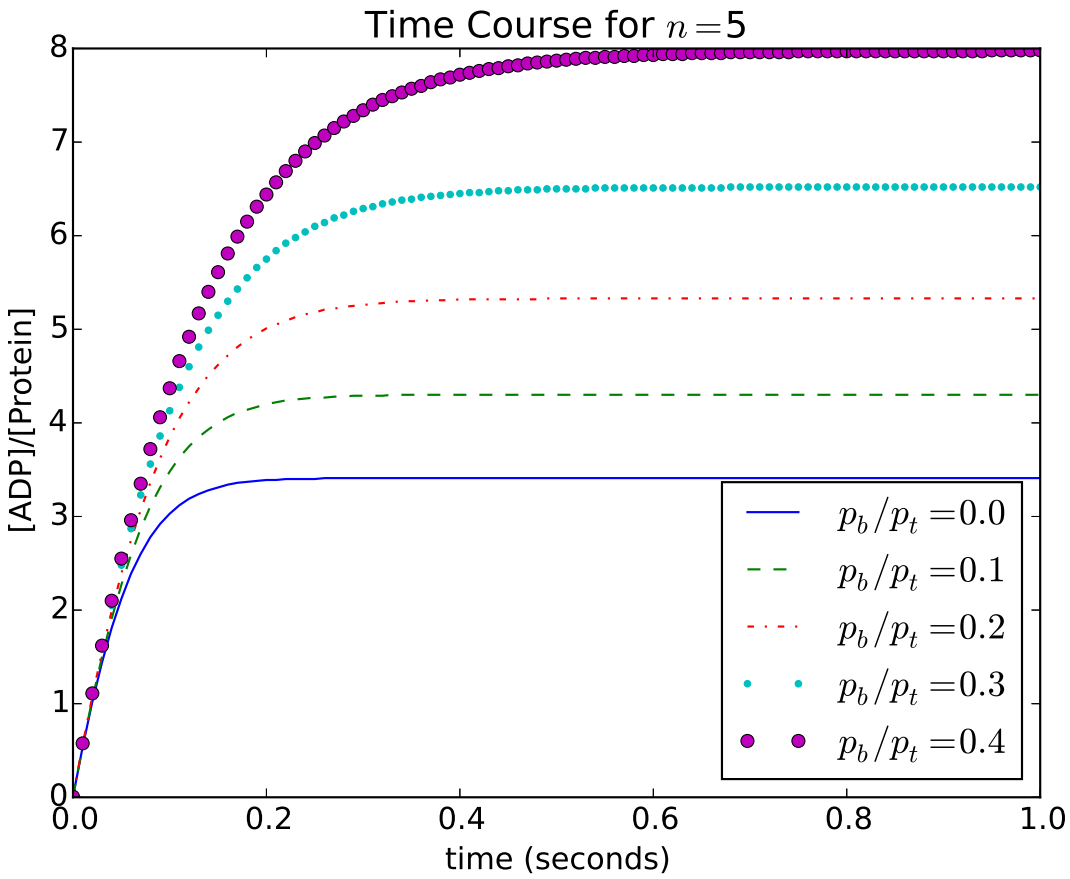


Figure 5.8: Monte Carlo simulated ATP hydrolysis time course data for backward motion for *n* = 5. The kinetic parameters used to generate this data are shown in Table 5.1. The amount of ATP hydrolyzed is equivalent to the amount of ADP. The concentration of ADP per protein is normalized by dividing the total amount of ADP in the system by the total number of proteins. As the ratio of backward motion to translocation increases, the amount of ADP produced increases. A comparison of this Figure to Figure 5.7 shows that the amount of ADP produced stops increasing roughly the same time that the fraction of protein at the 5' end goes to zero.

5.3.2.2 ATP hydrolysis

The Monte Carlo simulated time course data for the amount of ATP hydrolyzed for DNA with $n = 5$ is shown in Figure 5.8. The amount of ADP produced (as a result of the hydrolysis of the ATP) is coupled to the amount of motion (forward, backward, jumping, and pausing). When the ratio of backward motion to translocation (p_b/p_t) is zero there is no backward motion. The proteins require hydrolysis of ≈ 3.4 ATP per protein to reach the end (state $i = 0$) and/or dissociate. When $p_b/p_t = 0.2$ or $p_b/p_t = 0.4$, it takes hydrolysis of approximate ≈ 5.3 ATP per protein or ≈ 8.0 ATP per protein, respectively, for the proteins to reach to reach the end (state $i = 0$) and/or dissociate. As expected, when increasing the ratio of backward motion to forward motion, there is an increase in the amount of ATP hydrolyzed (ADP produced) because some of the energy is going into backward motion.

5.3.3 Analysis of the kinetic data

The Monte Carlo simulated time course data for proteins undergoing non-uniform motion (at a rate of k_θ) for five different lengths of DNA was generated using Monte Carlo methods. To estimate the *n-step* model kinetics parameters in Equations 5.12 and 5.16, a non linear least squares (NLLS) method was used to find the parameters that best fit the *n-step* model equations to the appropriate time course data. Specifically, the kinetic parameters in Equation 5.12 were found by fitting the simulated time course data for the fraction of proteins at the 5' end (state $i = 0$). The kinetic parameters estimates from the fitting of Equation 5.12 were then used in Equation 5.16 leaving only two unknown parameters which are associated with the ATP hydrolysis (ADP concentration). The parameters c and k_a in Equation 5.16 where estimated from the NLLS fitting to the ATP hydrolysis (ADP concentration) time course data. A standard NLLS analysis [87, 92, 96] was used to determine the global parameters k_t, k_d, k_0, c, and r which are independent of DNA length and the local parameter n which is dependent on DNA length. A linear least squares methods was used to determine m through the dependence of n on L according to Equation (5.2) [87, 92, 96]. All NLLS analyses were performed using the Conlin package [98] kindly provided by Dr. Jeremy

Williams. The software library CNL50 (Visual Numerics Incorporated, Houston, TX) was used for the numerical calculation of the inverse Laplace transform.

5.4 Results

All simulations and data analysis were performed as described in the Materials and methods Section 5.3. The results of the analysis of the simulated time course data is shown in Figures 5.9 - 5.32. The "apparent value" of the parameter in the presence of the non-uniform motion (as determined from the global NLLS analysis) is defined as X_{app}. The value of the kinetic parameter in the absence of any non-uniform motion is defined X_0. The relative change in the estimates of the parameters $m, mk_t, c/m$, and r are plotted as $\Delta X/X_0$ where $\Delta X = X_{app} - X_0$. The values of the parameters in the absence of non-uniform motion are m_0 = 1 nt/step, $m_0 k_{t,0}$ = 60 nt/second, c_0/m_0 = 1 ATP/nt and r_0 = 1. The change in the estimate for parameter d is plotted as a linear shift in d with $\Delta d = d_{app} - d_0$ rather than a relative change. As shown in Figure 5.5, d is the contact size and therefore a linear offset of the DNA length (Equations 5.1 and 5.2). The futile hydrolysis parameter k_a is also plotted as a linear shift because the Monte Carlo simulations had a futile hydrolysis rate of $k_{a_0} = 0$ in all simulations. That means $\Delta k = k_{a,app} - k_{a_0} = k_{a,app}$. The uncertainties associated with the kinetic parameters are small (sample uncertainties in Tables A.9 and A.10). Due to the uncertainties being smaller than the point size on the plots, the uncertainties are not included in Figures 5.9 - 5.32.

The simulated time course data for proteins moving with non-uniform motion (sample time course data in Figure 5.7) were almost always well described using the simple uniform "*n-step*" model in Equation 5.12. A well described fit was determined by considering the variance (and the sum of the squared residuals (SSR)) between the Monte Carlo simulated data and Equation 5.12 or Equation 5.16 (see Tables A.1 - A.4). For two types of random pausing the variance was higher suggesting the model better fit different cases of random pausing (Section 5.4.4).

5.4.1 Case 1: Backward motion

The first non-uniform motion explored was inclusion of occasional backward steps (reversal of directionality) during translocation as shown in the Scheme in Figure 5.6. In these simulations it was assumed that each hydrolysis of ATP was coupled to one step of the protein either in the forward (3' to 5') direction at the translocation rate $m_0 = 1$ nt/step (with probability p_t) or in the backward (5' to 3') direction at $m_b = 1$ nt/step (with probability p_b). Note that qualitatively similar results were obtained from the analysis of time courses generated from simulations in which backward motion of the translocase was not coupled to ATP hydrolysis.

The results of the NLLS analysis of proteins that moved with backward motion are shown in Figure 5.9 for the simulated protein time course data and in Figure 5.10 for the simulated ATP (ADP) time course data. As shown in Figure 5.9, as the ratio of the probability of backward motion to forward motion (p_b/p_t) increases, there are increases to the changes in both m and r. It was also observed that the estimate of Δd increases, though only slightly, with increasing p_b/p_t. Finally, as expected since the protein is making slower net forward progress towards the 5' end, the estimate of the macroscopic translocation parameter mk_t decreased with increasing p_b/p_t.

NLLS analysis of the simulated ATP time course data (Figure 5.10) showed that the estimate of c/m increases as p_b/p_t increases; this is consistent with the fact that ATP hydrolysis does not always result in forward motion of the translocase. Even though the simulations did not incorporate any futile hydrolysis ($k_a = 0$) by a protein bound at the 5' end of the DNA, the best fit of the *n-step* model results in a negative estimate of k_a. As shown in Figure 5.10, the magnitude of the estimate of k_a decreased significantly with increasing p_b/p_t and is likely a result of the larger than expected ADP production (Figure 5.8).

5.4.2 Case 2: Jumping motion

The second type of perturbation considered was "jumping" of the protein along the DNA. In these simulations, each hydrolysis of ATP resulted in the protein moving forward (3' to 5') either one step of step size m_0 with probability p_t or one step of step size m_j with probability p_j. Having

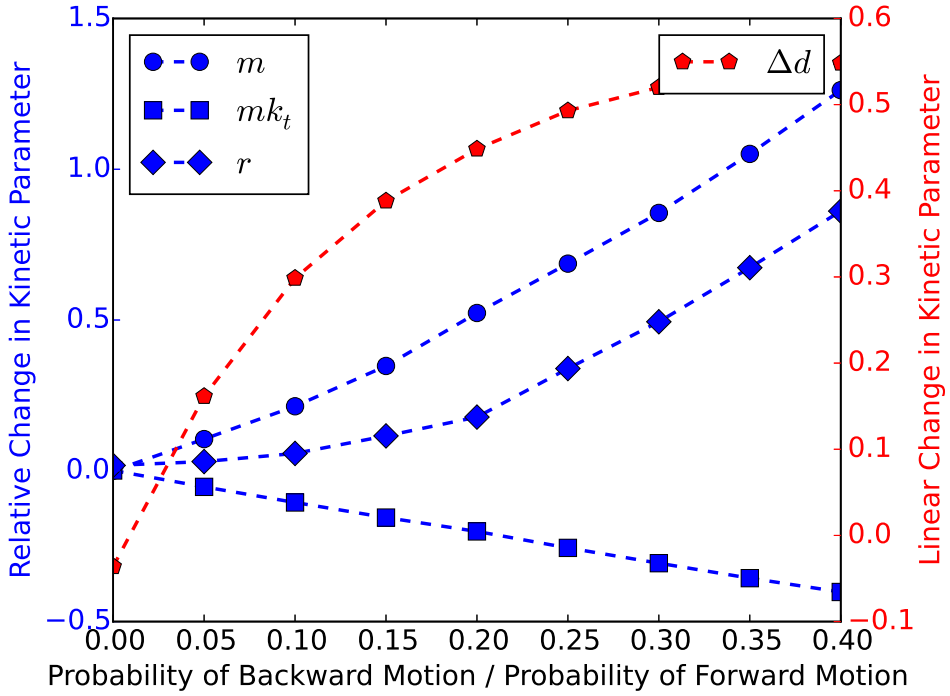


Figure 5.9: Backward motion. The estimates of the kinetic parameters were found using a NLLS analysis of Equation 5.12 with the simulated time course data for proteins. The data points in blue represent the relative change in the parameter (*e.g.* the circles are $\Delta m/m_0$) for the probability of non-uniform motion to translocation (*e.g.* p_b/p_t). The data points in red are the linear changes in the estimated contact size d (Figure 5.5). With increases to the ratio of the probability of backward motion to forward motion (p_b/p_t), there is decreases to the macroscopic translocation rate mk_t. Parameters m and r increased with p_b/p_t. There is little change in the linear parameter Δd.

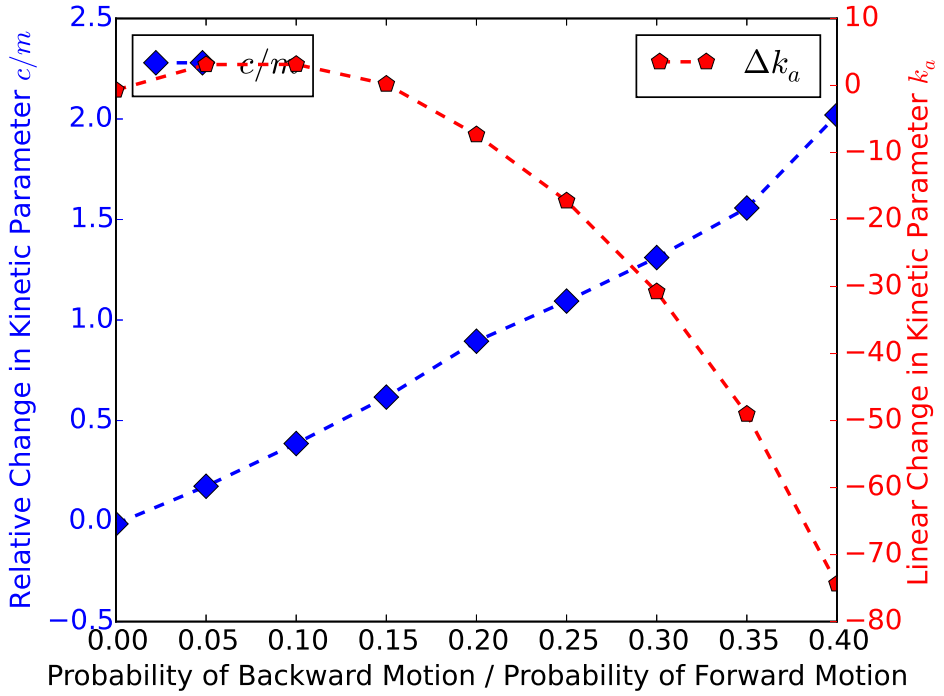


Figure 5.10: Backward motion. The estimates of the kinetic parameters were found using a NLLS analysis of Equation 5.16 with the simulated time course data for ATP hydrolysis. The data points in blue represent the relative change in the parameter c/m for the probability of non-uniform motion to translocation (*e.g.* p_b/p_t). There was a significant increase in the estimate of c/m with increasing p_b/p_t. The data points in red are the linear changes in the futile hydrolysis parameter k_a. Even though no futile hydrolysis was included in the simulations, the results of the analysis show significant decreases (negative values) for the rate of k_a.

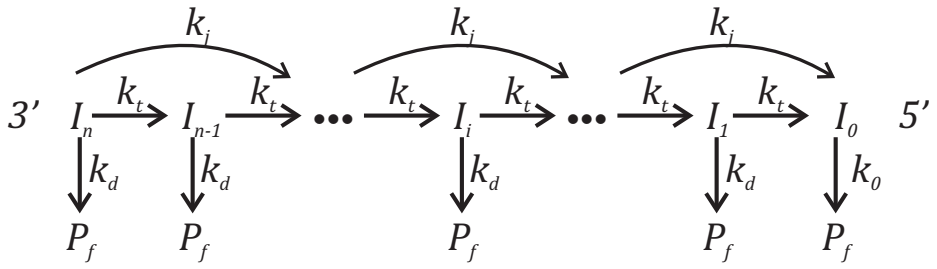


Figure 5.11: Scheme for translocation of the protein with jumping. The Monte Carlo simulations assume the protein jumps forward (3' to 5') at a rate of k_j nt/step and translocates forward (3' to 5') at a rate of k_t nt/step. The jumping step size (m_j is greater than the translocation step size (m_0). The protein will dissociated at rate k_d except at the 5' end where it dissociated at rate k_0.

assumed that $m_j > m_0$, moving a step size m_j which is greater than m_0 is defined as "jumping". Shown in Figures 5.12 and 5.13 are the results of the analysis of the data from simulations that assumed the translocation step size $m_0 = 1$ nt/step and the jumping step size was $m_j = 2$ nt/step. The NLLS analysis of Monte Carlo simulation data with increased jumping step sizes (m_j = 3 or 4 nt/step) showed qualitatively similar changes (though larger increases or decreases) in the kinetic parameter estimates of the *n-step* model and are not shown here.

The analysis of the simulated kinetic time courses protein data (Figure 5.12) shows that when increasing the ratio of the probability of jumping to the probability of translocating (p_j/p_t) both the step size m and the macroscopic translocation rate mk_t increase. This is not surprising since the jumping occurs in the forward direction and therefore leads to faster net forward motion. There were increases to r when increasing p_j/p_t which is most likely to account for the increases in both m and mk_t. There appears to be a slight decrease in the estimates of Δd with increasing p_j/p_t though this changes is much less significant than changes to the other parameters m, mk_t, and r.

The NLLS analysis of the simulated ATP time course data (Figure 5.13), shows that c/m decreases as p_j/p_t increases. This decrease is expected since jumping occurs in the forward direction and gives the appearance that ATP is not required for every step forward and therefore the total amount of ATP required to progress m steps is smaller. As with the simulations with backward motion, no futile hydrolysis occurred in the Monte Carlo simulations with jumping motion, however, unlike in the backward motion results, increases in the p_j/p_t resulted in a slight decrease in

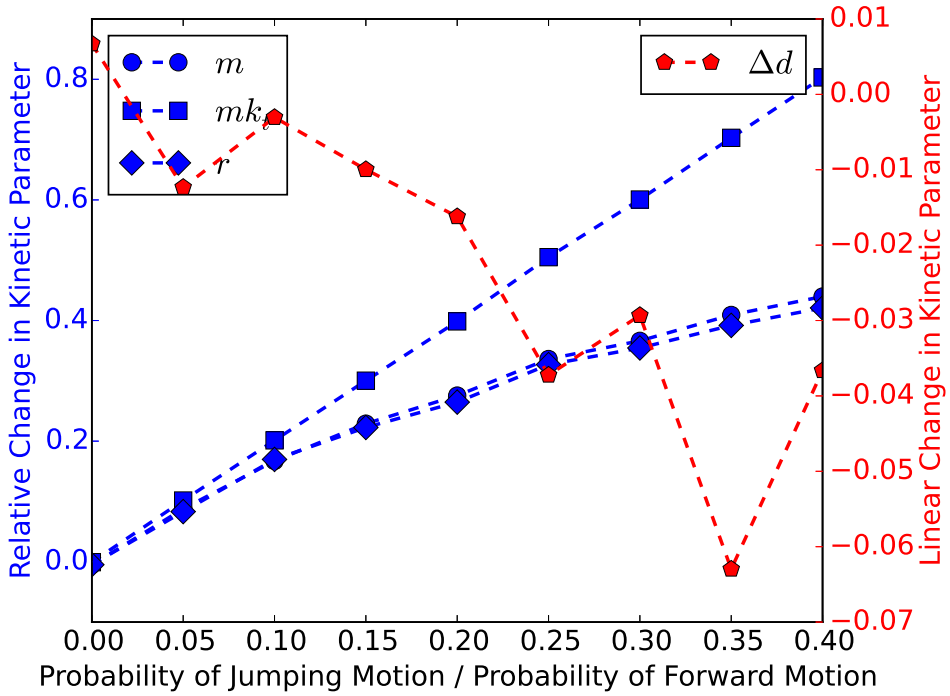


Figure 5.12: Jumping forward a distance of 2 nt/step. The estimates of the kinetic parameters were found using a NLLS analysis of Equation 5.12 with the simulated time course data for proteins. The data points in blue represent the relative change in the parameter (*e.g.* the circles are $\Delta m/m_0$) for the probability of non-uniform motion to translocation (*e.g.* p_j/p_t). The data points in red are the linear changes in the estimated contact size d (Figure 5.5). Shown here are changes in the estimates for a protein that not only translocates forward (1 nt/step) but also moves forward by "jumping" (2 nt/step). As with all types of non-uniform motion, increasing the ratio of the probability of jumping to the probability of translocation (p_j/p_t) increases both m and r. The increasing estimates to macroscopic translocation rate mk_t are expected with increasing p_j/p_t. There is a slight decrease in Δd with increasing p_j/p_t.

Δk_a. Because the simulations included no futile hydrolysis, this decrease results in a negative Δk_a which is not physical.

5.4.3 Case 3: Backward and jumping motion

In addition to simulations that incorporated either of backward motion or jumping motion, Monte Carlo simulations that included both backward motion and jumping as described in Sections 5.4.1 and 5.4.2 were performed. The intent of this case was to determine how the kinetics parameters where impacted when the non-uniform motion did not change the expected kinetic step size. When

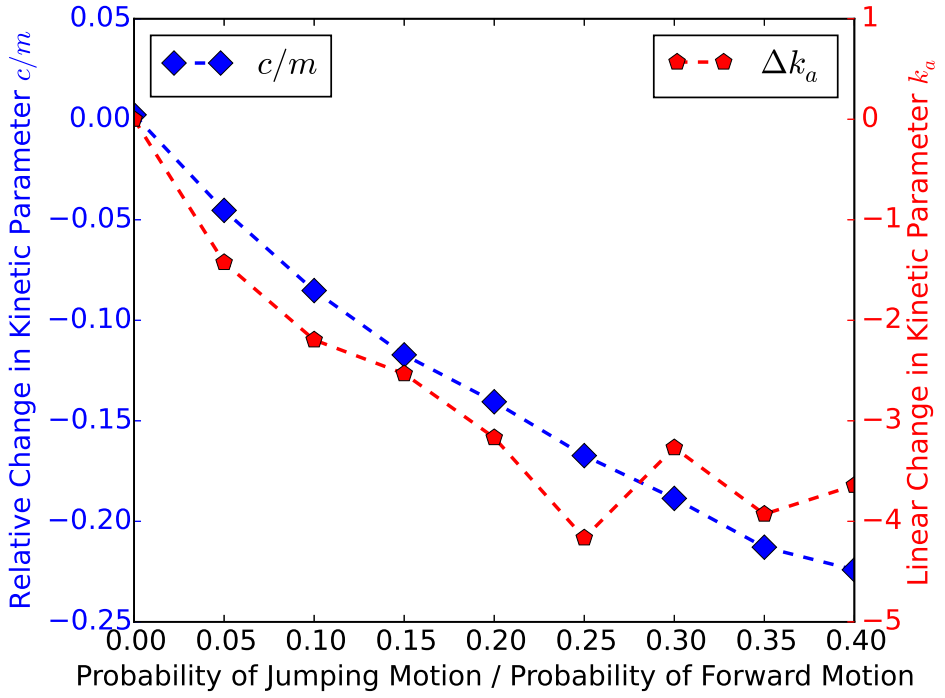


Figure 5.13: Jumping forward a distance of 2 nt/step. The estimates of the kinetic parameters were found using a NLLS analysis of Equation 5.16 with the simulated time course data for ATP hydrolysis. The data points in blue represent the relative change in the parameter c/m for the probability of non-uniform motion to translocation (*e.g.* p_j/p_t). There was a slight decrease to the estimates of ATP coupling c/m. The data points in red are the linear changes in the futile hydrolysis parameter k_a which showed a slight decrease with increasing p_j/p_t.

using the following constraints on the Monte Carlo simulated proteins, the expected step size should not increase.

1. backward motion (5' to 3') occurred at a distance of m_b nt/step
2. jumping occurred in the forward direction (3' to 5') at a distance of $m_j = \lambda m_b$ nt/step
3. the probability of jumping is p_j
4. the probability of backwards motion is $p_b = \lambda p_j$ or $\lambda = p_b/p_j$

where λ is a constant greater than one. The expected step size is

$$E[M] = \sum_i m_i p_i(m_i) \tag{5.19}$$

where m_i is the step size m_b, m_j, or m_0 and $p_i(m_i)$ is the probability of that particular step size. Using the constraints above and a negative for the backward motion step size m_b

$$E[M] = -m_b p_b + m_j p_j + m_0 p_t \tag{5.20}$$

and applying constraint (2)

$$E[M] = -m_b p_b + \lambda m_b p_j + m_0 p_t \tag{5.21}$$

and applying constraint (4)

$$E[M] = -m_b p_b + \frac{p_b}{p_j} m_b p_j + m_0 p_t \tag{5.22}$$

and reduces to

$$E[M] = -m_b p_b + p_b m_b + m_0 p_t \tag{5.23}$$

and reduces further to

$$E[M] = m_0 p_t \tag{5.24}$$

so the jumping motion and backwards motion average out to zero.

In the simulations presented here, $\lambda = 2$, $m_b = 1$ nt/step which due to the above constraints means $m_j = 2$ nt/step. The distance jumped (3' to 5') is twice the distance of backward motion (5' to 3') and the probability of backward motion was twice the probability of jumping ($p_b = 2p_t$). These results are of interest since as with the other non-uniform motion there was increases to both m and r, but there were no significant changes to the estimates of mk_t or Δd even with increasing the ratio of the probabilities of backward and jumping motion to the probability of translocation ($p_{b,j}/p_t$) (Figure 5.14). The changes (or lack of) to these parameters are likely due to the constraints resulting in the protein moving backward twice as often but with half the distance of the jumping (for $\lambda = 2$) and negating changes to both mk_t and d. As with the case of backward motion, the NLLS analysis of the ATP time course data showed increases to c/m and negative rates of futile hydrolysis k_a (Figure 5.15) when increasing $p_{b,j}/p_t$.

5.4.4 Case 4: Random pausing

To determine the effects of random pausing, the hydrolysis of each ATP molecule was assumed to result in either motion of the protein translocating along the DNA in the forward (3' to 5') direction or a pause (or stall) in the motion of the protein. An arbitrary chosen probability of recovering from the pause was held constant, but the time at which the ATP hydrolysis occurred in association with the pause varied in each of the three simulations. In the first set of simulations it was assumed that only recovery from the pause was associated with ATP hydrolysis. As shown in Figure 5.16, as the ratio of the probability of pausing to the probability of translocation (p_p/p_t) increases the estimate of mk_t correspondingly decreases. Increases in p_p/p_t also resulted in increases to m and r, although the increase was slightly smaller than the change due to backward motion. There appears to be only a slight dependence of the estimate of Δd on the probability of random pausing. The increase to c/m was lower than backward motion but still consistent with the fact that ATP hydrolysis does not always result in forward motion of the translocase (Figure 5.16). The increases to m and r were initially much larger than c/m but then leveled out.

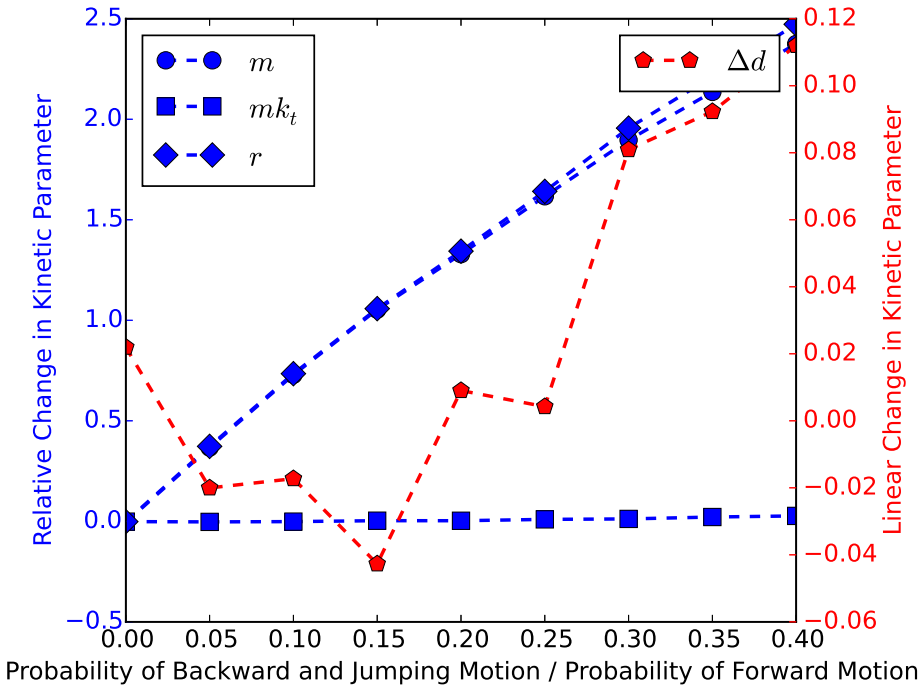


Figure 5.14: Both backward and jumping motion occur with the probability of backward motion equal to twice the probability of jumping. The backward motion step size ($m_b = 1$ nt/step) is equal to half the jumping motion step size ($m_b = 2$ nt/step). Just as the other cases of non-uniform motion, there are increases to both m and r when increasing the ratio of the probability of backward and jumping motion to the probability of translocation ($p_{b,j}/p_t$). There are no changes in mk_t and a slightly increasing trend for Δd.

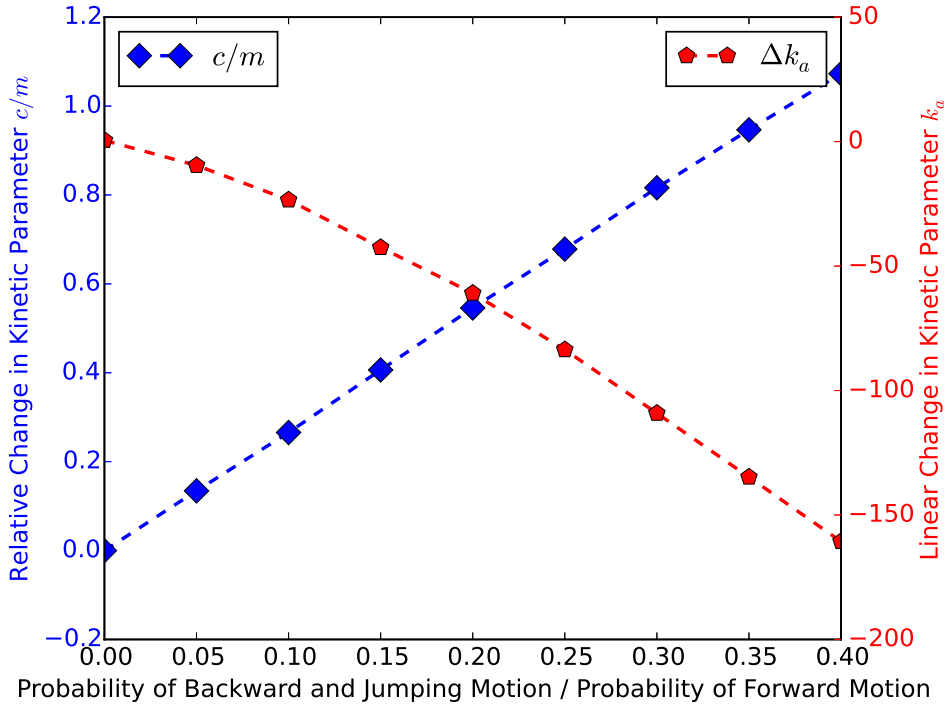


Figure 5.15: Both backward and jumping motion occur with the probability of backward motion equal to twice the probability of jumping. There is an increase in c/m with increasing $p_{b,j}/p_t$. No futile hydrolysis was included in the simulations but analysis showed an negative rates of futile hydrolysis k_a and the increase was more than with only backward motion.

The changes in the parameters with random pausing where qualitatively similar to the changes with backwards motion in that m, mk_t, r, and d all increased while c/m decreased. However, a comparison of Figures 5.10 and 5.17 shows that increasing the non-uniform motion decreases Δk_a for both backwards motion and random pausing. Since the simulations did not include futile hydrolysis $k_a = 0$, the estimates for the $k_{a,app}$ was negative. This negative $k_{a,app}$ might be an indicator of non-uniform motion that is physically unreal.

In our second set of simulations, it was assumed both the start of a pause (or entering the pause) and the recovery of a pause (or exiting the pause) resulted in ATP hydrolysis. As p_p/p_t increased, the differences between the simulated data and the model increased (see the variance in Table A.2). For the third set of simulations, it was assumed ATP hydrolysis was coupled to the start of the pause (and none was required for pause recovery). Including the ATP hydrolysis only at the start of the pause again resulted in greater differences between the simulated data and the model (Table A.2). This *n-step* model appears to better fit the case of ATP hydrolysis being associated with the recovery of the protein (or exiting) from the stall. Due to the model no longer providing as good of a fit to the data, results of the analysis for the second and third sets of simulations have not been shown.

5.4.5 Case 5: Persistent heterogeneity in translocation rates among translocase motors

The rates of DNA unwinding by the *E.coli* UvrD helicase from a single-molecule study have suggested significant and persistanct variation among individual proteins [97]. In other words, protein A would unwind the DNA at a rate of $k_{t,A}$ and protein B would unwind the DNA at a rate of $k_{t,B}$, so that each individual protein is unwinding the DNA at its own specific, yet constant rate. In order to achieve this persistent heterogeneity in the translocation rates, it is likely that proteins with different rates would have to be chemically or conformationally different which does not seem particularly likely. However, these perturbations were explored by simulating the time course data for proteins that have a microscopic translocation that is normally (Gaussian) distributed about the

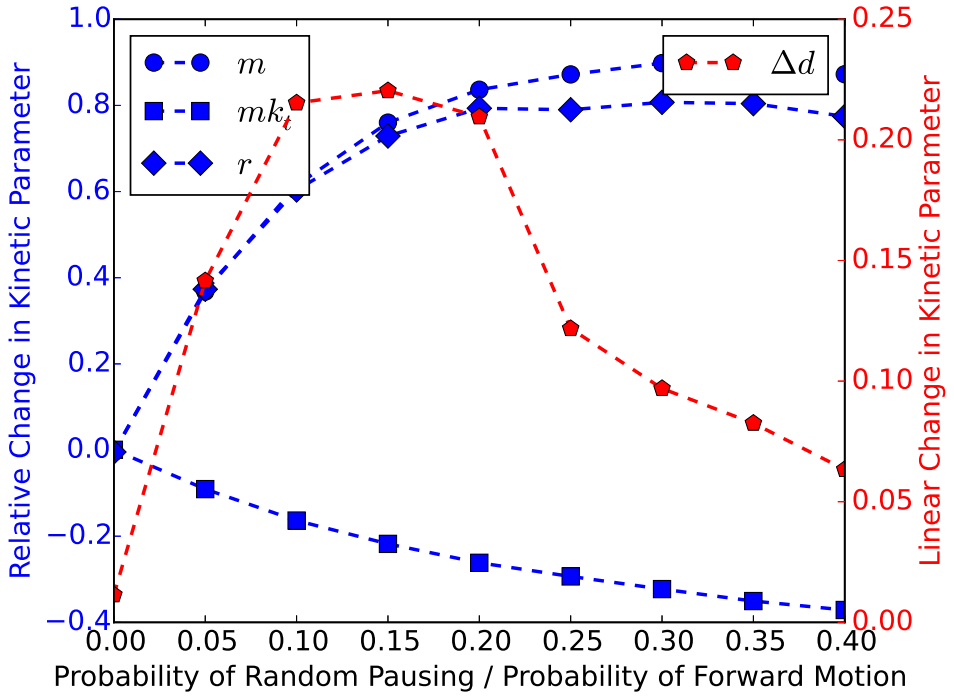


Figure 5.16: ATP hydrolysis of a molecule results in either forward motion of the protein (3' to 5') or recovery from a random pause or stall. As with all non-uniform motion, there are increases to both m and r when increasing p_p/p_t. Just as in the case of backward motion, there are decreases to mk_t. There are no significant changes in Δd.

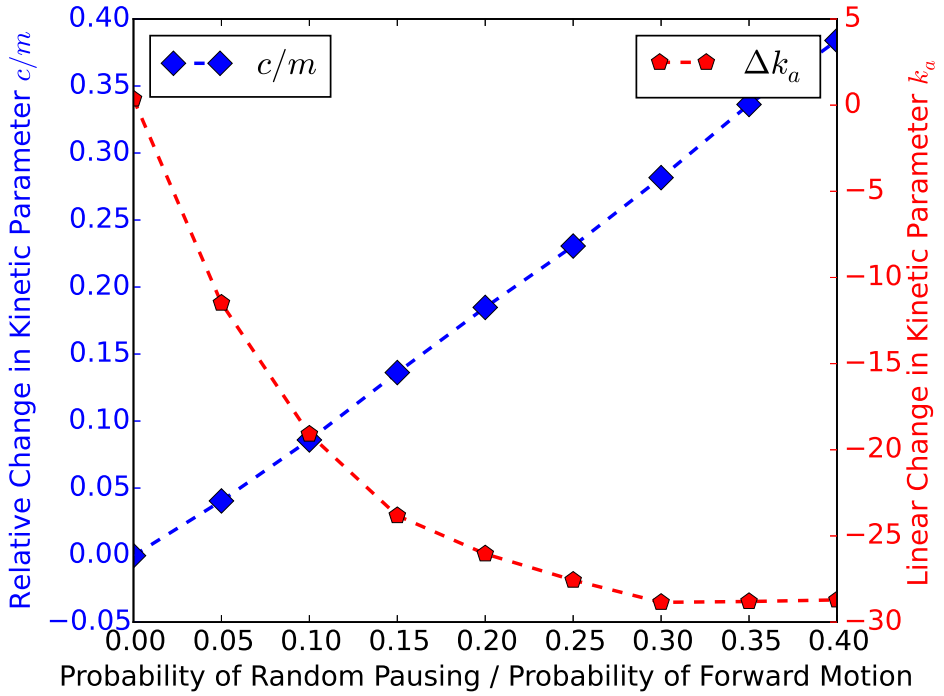


Figure 5.17: ATP hydrolysis of a molecule results in either forward motion of the protein (3' to 5') or recovery from a random pause or stall. Just as in the case of backward motion, there are increases to c/m with increases to p_p/p_t. Similarly, backward motion decreases the estimated futile hydrolysis as does random pausing. This decrease in Δk_a means a negative kinetic parameter which is not physically real.

mean translocation rate $k_{t,0}$. Each (simulated) protein moved with a constant translocation rate that was randomly generated assuming a normal distribution with mean $\mu = k_{t,0}$ and standard deviation σ. The changes in the kinetic parameters are plotted as a function of the standard deviation of the translocation rate divided by the expected value of the translocation rate (σ/μ).

Using the initial parameters $k_{t,0} = 60$ and $k_{d,0} = 1$ (Materials and methods Section 5.3), Equation 5.17 reduces to $k_d + k_t = 61$ and Equation 5.18 reduces to $k_d/k_t = 1/60$. This system with two equations and two unknowns has a single unique solution for uniform motion. Therefore, the constant sum (Equation 5.17) and the constant ratio (Equation 5.18) were considered as separate independent cases.

5.4.5.1 Case 5a: Sum of $k_{d,0} + k_{t,0} = k_d + k_t$ = constant

The results of NLLS analysis of simulated time course for persistent heterogeneity in the translocation rate with kinetic parameters constrained by Equation 5.17 are shown in Figures 5.18 and 5.19. There is an increase in the macroscopic translocation rate mk_t (Figure 5.18) when increasing σ/μ. As with other perturbations, there are increases to both m and r (Figure 5.18). There is no significant decrease to Δd (Figure 5.19). There was a decrease to c/m (Figure 5.19) when increasing σ/μ. These results are similar to the case with "jumping" of the protein. However, unlike when "jumping", persistent heterogeneity in the translocation rate constrained by Equation 5.17 showed negative rates of futile hydrolysis k_a (Figure 5.19).

5.4.5.2 Case 5b: Ratio of $k_{d,0}/k_{t,0} = k_d/k_t$ = constant

Comparing the results of Figures 5.18 and 5.19 with Figures 5.20 and 5.21, there are differences in the changes to the kinetic parameters when using the different constraints from Equation 5.17 and 5.18. As with all perturbations, there are increases to both m and r with increases to σ/μ. Unlike in Case (5a) there are only slight decreases to mk_t. Parameters $\Delta d, m$, and r increase with increasing σ/μ much higher more than Case (5a). Increasing σ/μ has lead to decreases in Δk_a and estimated negative rates of futile hydrolysis.

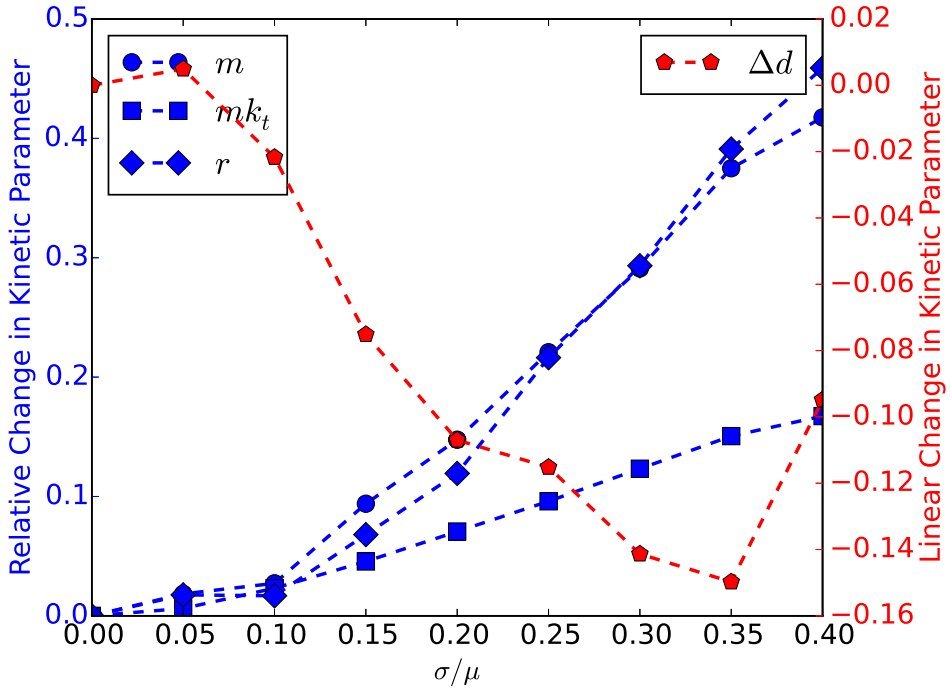


Figure 5.18: Case 5a: Persistent heterogeneity in translocation rates among translocase motors with $k_d + k_t$ = constant. Changes are plotted as a function of the standard deviation of the translocation rate (σ/μ). As with all non-uniform motion, there are increases to both m and r. A slight increase in mk_t occurs with increasing ratios of σ/μ. There is a slight decrease in Δd with increasing σ/μ.

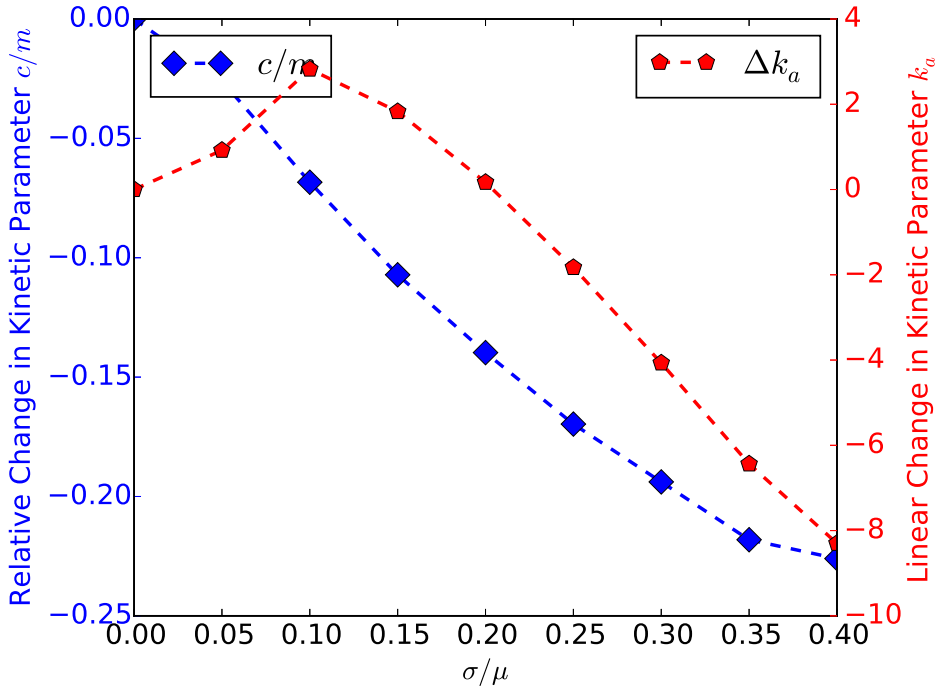


Figure 5.19: Case 5a: Persistent heterogeneity in translocation rates among translocase motors with $k_d + k_t$ = constant. Changes are plotted as a function of the standard deviation of the translocation rate divided by the mean of the translocation rate (σ/μ). When increasing the ratio of σ/μ, there are negative rates for k_a even though the simulations had a futile hydrolysis rate of $k_a = 0$. There are decreases to c/m with increases to σ/μ.

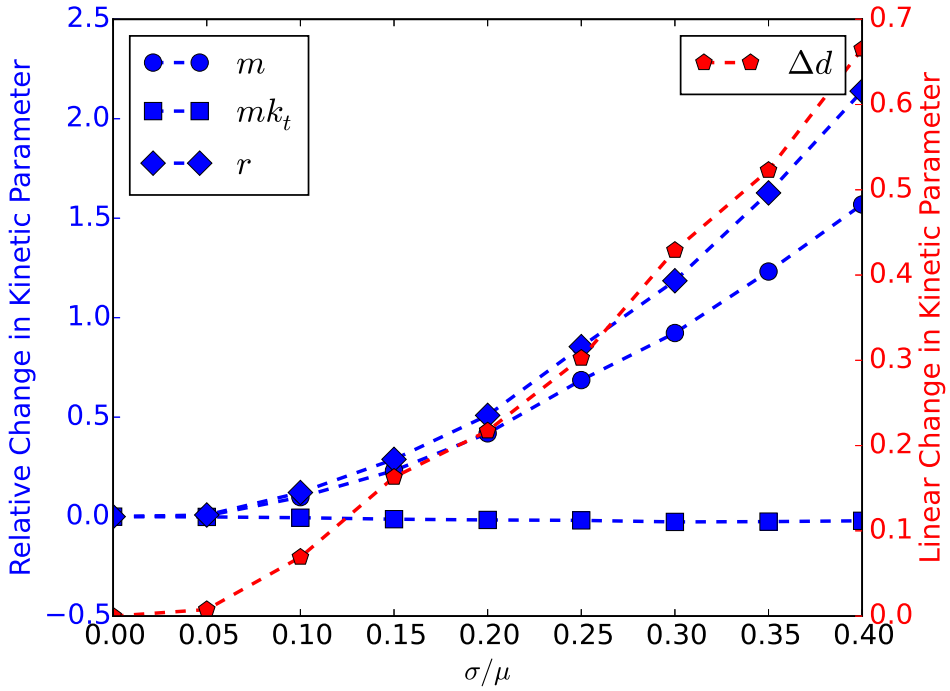


Figure 5.20: Case 5b: Persistent heterogeneity in translocation rates among translocase motors with k_d/k_t = constant. Changes are plotted as a function of the standard deviation divided by the mean (σ/μ). There are increases to $m, \Delta d$, and r with increasing σ/μ and only a slight decreases in mk_t.

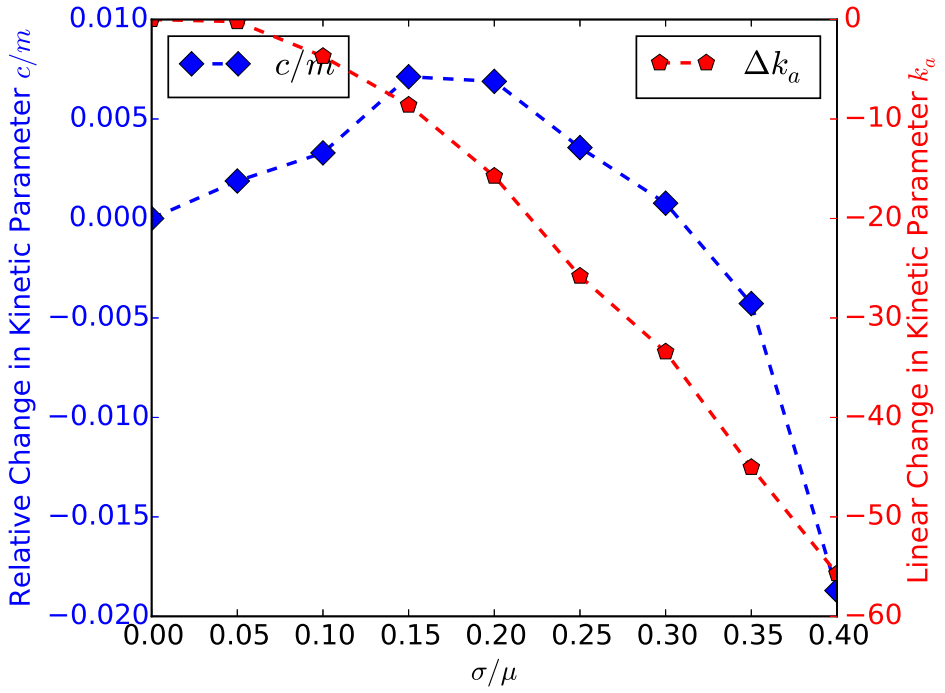


Figure 5.21: Case 5b: Persistent heterogeneity in translocation rates among translocase motors with k_d/k_t = constant. Changes are plotted as a function of the standard deviation divided by the mean (σ/μ). There are large decreases to Δk_a and slight increases to c/m when increasing σ/μ.

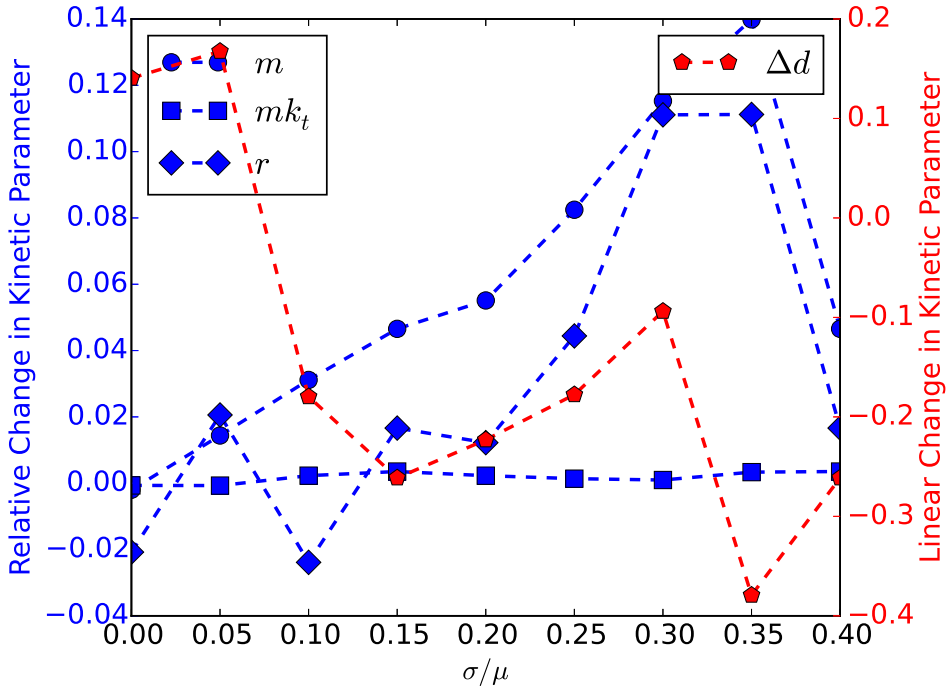


Figure 5.22: Case 6a: Each translocase samples from a distribution of step-sizes for each individual step. Changes are plotted as a function of the standard deviation divided by the mean (σ/μ). There appears to be no significant changes with increasing (σ/μ).

5.4.6 Case 6: Variations in the step-size

5.4.6.1 Case 6a: Each translocase samples from a distribution of step-sizes for each individual step

In addition to considering heterogeneities in the translocation rate, heterogeneities in the translocation step size m were considered. For one case it was assumed the protein moves so that each translocation step size was determined from a normal (Gaussian) distribution assuming the mean $\mu = m_0 = 10$ and the standard deviation was $\sigma = 0.0, 0.5, 1.0, ..., 3.5, 4.0$. In other words, a protein might move a distance of m_A nucleotides, then a distance of m_B nucleotides, etc. The analysis of both the protein and the ATP time course data (Figures 5.22 and 5.22) showed no significant changes. However, it is interesting to note that this type of non-uniform motion had the most variance between the simulated data and the model (Table A.2).

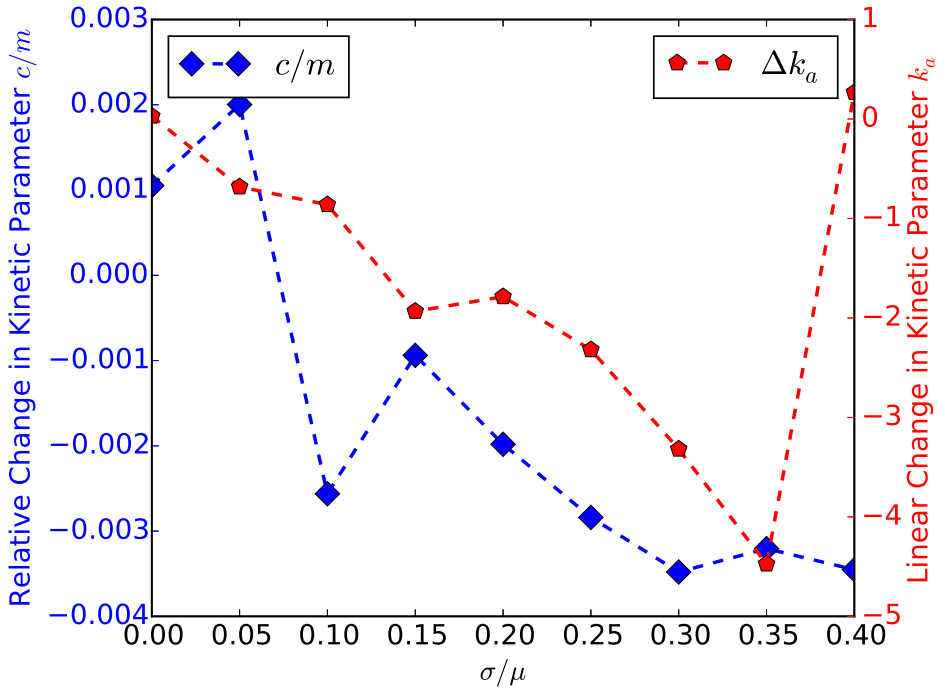


Figure 5.23: Case 6a: Each translocase samples from a distribution of step-sizes for each individual step. Changes are plotted as a function of the standard deviation divided by the mean (σ/μ). There appears to be no significant changes with increasing (σ/μ).

5.4.6.2 Case 6b: Persistent heterogeneity in step-sizes among translocase motors

The case where each individual protein always moved with a (potentially different among proteins) translocation step-size m that was determined from a normal (Gaussian) distribution was considered with $\mu = m_0 = 10$ and the standard deviation $\sigma = 0.0, 0.5, 1.0, ..., 3.5, 4.0$. In other words, protein A would always translocate a distance of m_A nucleotides and protein B would always translocate a distance of m_B, so that each individual protein is moving at their own specific, yet constant distance. As mentioned in Section 5.4.5, this type of persistent heterogeneity in the behavior of the protein would suggest differences in the chemical properties or conformation of the proteins. However, I did explore this option.

As shown in Figure 5.24, there are no significant changes to mk_t. However, Figure 5.24 shows increases to m and Δd and even larger increases to r when increasing the standard deviation to mean ratio σ/μ. Increasing (σ/μ) slightly increased c/m and significantly decreased Δk_a (Figure 5.25).

5.4.7 Summary of results

Quantitative estimates of the kinetic parameters from the fitting of Equation (5.12) to the protein time course data and Equation (5.13) to the ATP hydrolysis (ADP concentration) time course data were almost always impacted by the non-uniform motion (perturbations). A summary of the changes to the kinetic parameters is given in Table 5.2.

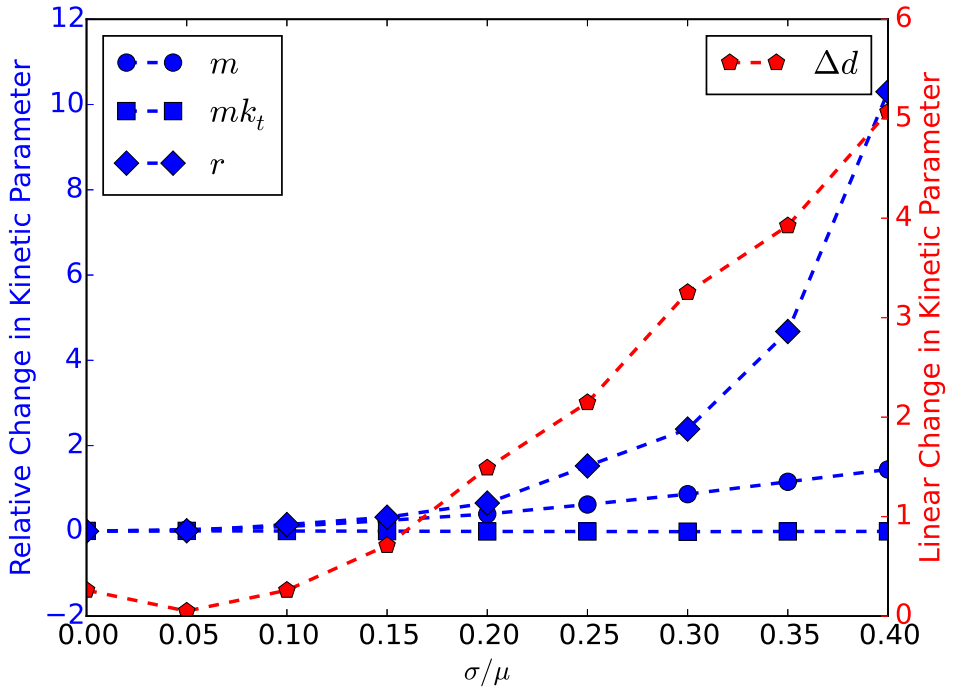


Figure 5.24: Case 6b: Persistent heterogeneity in step-sizes among translocase motors. Changes are plotted as a function of the standard deviation divided by the mean (σ/μ) of the step size. There is a significant increase in r and small increases to both Δd and m. There was no changes to mk_t.

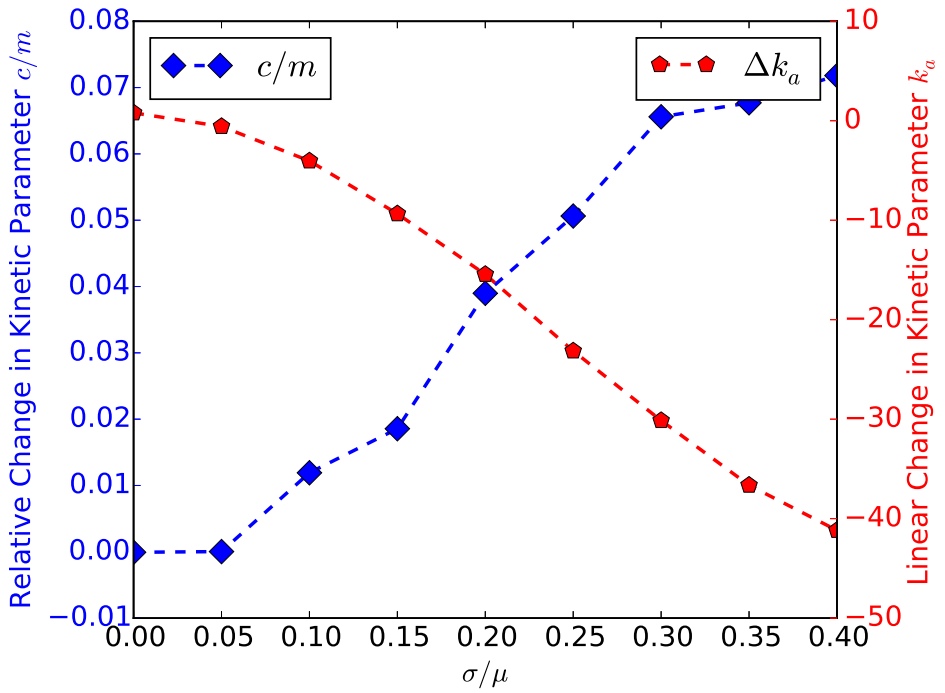


Figure 5.25: Case 6b: Persistent heterogeneity in step-sizes among translocase motors. Changes are plotted as a function of the standard deviation divided by the mean (σ/μ) of the standard deviation. There are slight increase to c/m with increases to σ/μ. Increasing σ/μ lead to decreases in $Deltak_a$ and negative rates of futile hydrolysis.

Table 5.2: A summary of the changes to kinetic parameters due to increasing probabilities of non-uniform motion or stepping rates. In (almost) all cases, increasing probabilities of non-uniform motion leads to increases in both m the apparent kinetic step size and r the ratio of the probability that the protein is bound to any binding site other than the 5' end at time $t = 0$. Also, increasing non-uniform motion often lead to decreases in Δk_a.

Non-Uniform Motion	m	mk_t	r	$\Delta d(nt)$	c/m	Δk_a
Backward	increase	decrease	increase	increase?	increase	decrease
Jumping	increase	increase	increase	decrease?	decrease	$\approx$
Backward and Jumping	increase	$\approx$	increase	$\approx$	increase	decrease
Random Pausing	increase	decrease	increase	increase?	increase	decrease
Case 5a	increase	increase	increase	decrease	decrease	decrease
Case 5b	increase	decrease	increase?	increase	increase	decrease
Case 6a	$\approx$	$\approx$	$\approx$	decrease?	$\approx$	$\approx$
Case 6b	increase	$\approx$	increase	increase	increase	$\approx$

It is interesting to note that when increasing any type of non-uniform motion, there is (almost) always an increase to the apparent step size (m) and the ratio of the probability of the translocase binding to any one binding site on the DNA other than the 5' end to the probability of the translocase binding to the 5' end (kinetic parameter r). The coupling between m and r becomes apparent when analyzing the time course data for simulations that followed the Scheme in Figure 5.3 but increased the step size. In Figures 5.26 and 5.27 are the results for increasing step size. As the step size increase, the only change in the kinetic parameters is the increase in r. In addition, the coupling of the parameters m and r is also apparent when considering the correlation matrices (Tables A.5 - A.8).

5.5 Discussion

The kinetic parameters (step-size, translocation or unwinding rates, *etc.*) associated with translocation of proteins along the DNA or unwinding of the DNA have been estimated [82, 83, 86, 87, 89, 91, 95–97, 99–107]. Estimates of the kinetic parameters where often found by fitting models

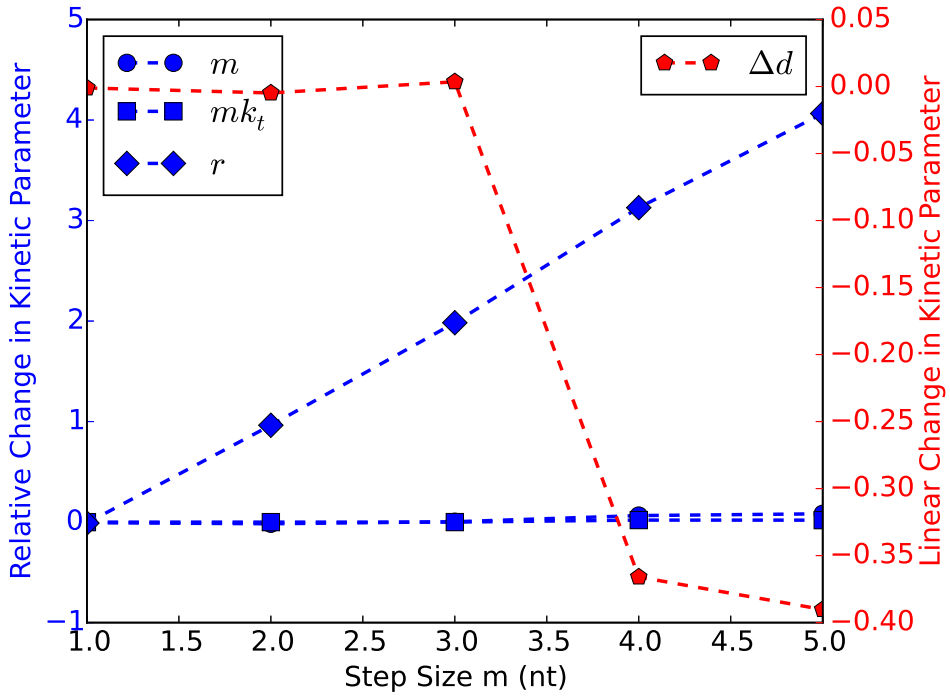


Figure 5.26: Increases to the step size. Data analysis of the simulated time course data for uniform motion as shown in Figure 5.3 with different step sizes. The only significant change in the kinetic parameters is to r which increases with the step size m.

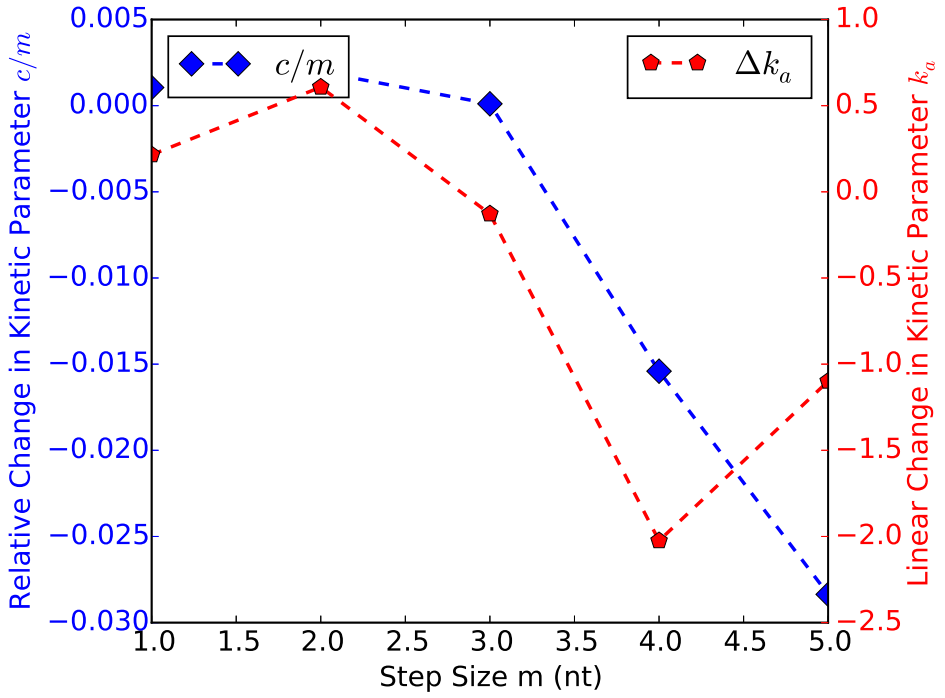


Figure 5.27: Increases to the step size. Data analysis of the simulated time course data for uniform motion as shown in Figure 5.3 with different step sizes. There are no significant changes when increasing the step size *m*.

that assume uniform motion of the proteins. A variety of non-uniform motions were considered for the simulations and the time course data analyzed assuming a uniform sequential "*n-step*" model [87, 92, 96]. By understanding the impact that specific perturbations have on the kinetic parameters when analyzed using the *n-step* model, a better understanding of how the proteins are translocating along the DNA can be determined. Though as mentioned previously, these results are not specific to translocation of protein along single stranded DNA as modeled here but are general and applicable to the translocation of other proteins whose analysis can be done using the *n-step* model.

5.5.1 Estimates of the macroscopic kinetic parameters are least affected by perturbations

The macroscopic parameters of the *n-step* model are the macroscopic translocation rate (mk_t) and ATP coupling to the motion (c/m). These parameters are representative of the overall behavior of the system. In order to understand the impact of the macroscopic translocation rate mk_t, I considered the mean (average) time it takes for a protein to translocate n steps along the DNA. For a protein initially bound n steps from the the 5' end and moving at a rate of k_t, the expected time τ for the protein to arrive at the 5' end that is Gamma distributed $\Gamma(n,k_t)$ is

$$\mathbf{E}[X] = \tau = \frac{n}{k_t} = \frac{L-d}{mk_t} \tag{5.25}$$

where **E**[X] is the expected value. Equation 5.25 can be re-written as a function of τ

$$mk_t = \frac{L-d}{\tau} \tag{5.26}$$

By considering mk_t as function of τ, it seems reasonable that any perturbations that increase the expected time of the proteins arrival at the 5' end (τ) should have an impact on the the macroscopic translocation rate mk_t. The decrease in the estimate of mk_t due to the backward and random

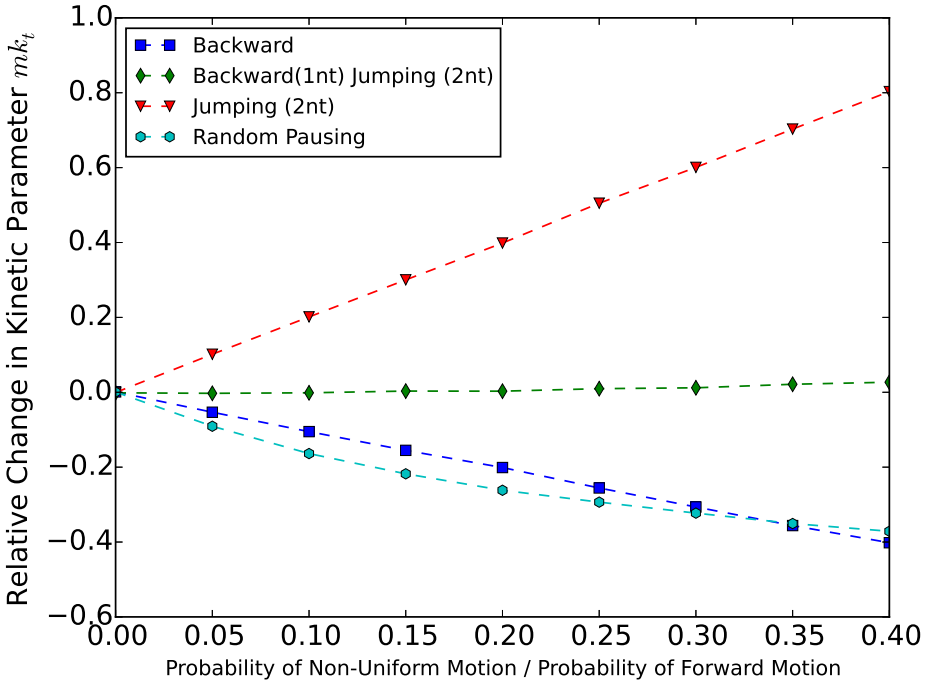


Figure 5.28: Relative changes in the macroscopic translocation rate mk_t. Increasing the probability of jumping increases mk_t. Increasing the probability of backward and random pausing motion decreases mk_t. Then backward motion occurs at twice the probability of jumping motion but with half the distance, increasing the probabilities had no impact on mk_t.

pausing motion and the increase in the estimate of mk_t due to jumping are expected since net forward progress has been modified. The lack of changes in mk_t due to backward with jumping and heterogeneous step size are also expected since there should be no overall change in the forward progress. Overall, any perturbations that change the mean arrival time should have an impact on the estimate of the macroscopic translocation rate mk_t. Equation 5.26 shows an inverse relationship between mk_t and τ so perturbations that decrease the average arrival time should increase the estimate of mk_t and perturbations that increase the average arrival time will decrease the estimate of mk_t.

The rate of ATP associated with net forward motion (c/m) increased for proteins with non-uniform motion that does not result in net forward progress of the protein (*e.g* backward and random pausing). The ATP associated with net forward motion decreased when "apparent" forward motion of the protein occurred without ATP hydrolysis (*e.g* jumping). With jumping motion, the

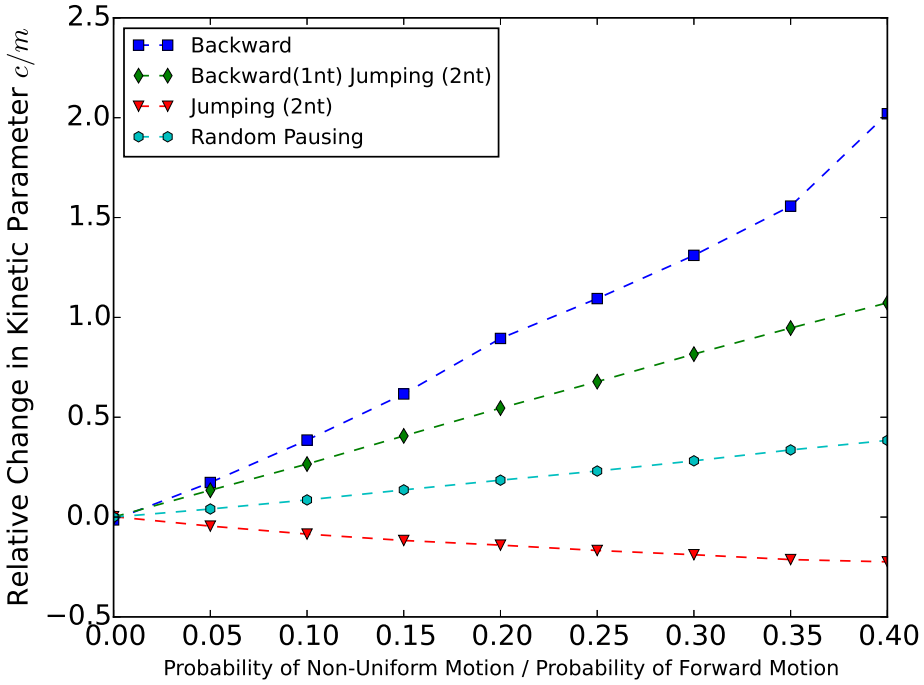


Figure 5.29: Relative changes in the macroscopic rate of ATP associated with net forward motion c/m. Increasing the probability of jumping decrease c/m. Increasing the probability of backward and random pausing motion increases c/m. When backward motion occurs at twice the probability of jumping motion but with half the distance, increasing the probabilities lead to increases to c/m.

protein could move forward two steps with the hydrolysis of only one ATP. This two step "jump" would normally require two ATP and gives the appearance that net forward progress occurred without ATP hydrolysis. For backwards motion with jumping, increasing probabilities of non-uniform motion are not expected to impact c/m but data analysis shows an increase to c/m likely due to the increase in m (see below).

5.5.2 Perturbations that increase estimates of the kinetic step-size

In order to understand the impact of the microscopic step size m, the variance in the time it takes for a protein to translocate n steps along the DNA was considered. For a protein initially bound n steps from the the 5' end and moving at a rate of k_t, the variance σ in the time of the protein

arriving at the 5' end is Gamma distributed $\Gamma(n, k_t)$ with

$$\mathbf{Var}[X] = \sigma^2 = \frac{n}{k_t^2} \tag{5.27}$$

where **Var**[X] is the variance. Using Equations 5.2 and 5.25, Equation 5.27 can be re-written as

$$\sigma = \frac{\sqrt{n}}{k_t} = \frac{n}{\sqrt{n}k_t} = \frac{\tau}{\sqrt{n}} = \sqrt{m}\frac{\tau}{\sqrt{L-d}} \tag{5.28}$$

where σ is a function of m.

By considering m as function of σ, it seems that any perturbations that do not change the expected arrival time τ but increase the standard deviation (σ) of the proteins arrival at the 5' end should have an impact on the microscopic step size m. Because all perturbations considered should increase the standard deviation of the arrival times, increases to the step size m for all perturbations are expected. As shown in Figure 5.30, increasing the probabilities of perturbation increases the step size. This leads to an overestimation in the step size m.

5.5.3 Significance of the r parameter as an indicator of non-uniformity

In Equations 5.12 and 5.16 of the "*n-step*" model, the parameter r is defined as the ratio (at time $t = 0$) of the probability that the protein is bound to any one binding position other than the 5' end to the probability that the protein is bound at the 5' end [87, 92, 96]. When the protein is equally likely to bind at any position along the DNA, $1 \leq r \leq m$ (see Section A.1 and [92]). If NLLS analysis of the simulated time course data estimated a value of r that is greater than m, this may be an indication that the model is not correctly describing the behavior of the system. Therefore, when results of the fitting of the model find $r > m$, this may be a indication that non-uniform motion is occurring.

In order to gain an understanding of the coupling between r and m, consider the amount of

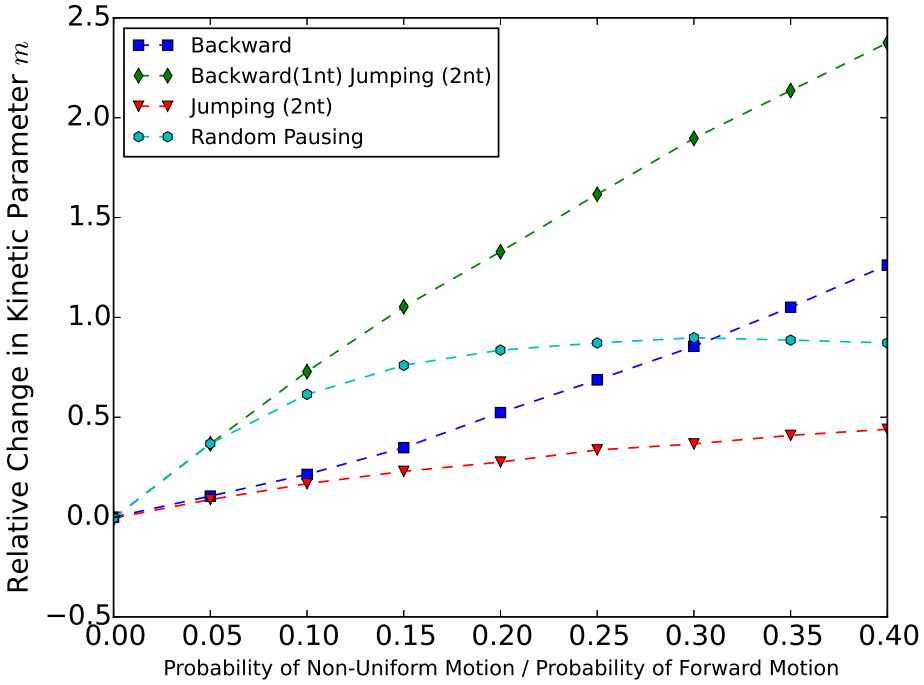


Figure 5.30: Relative changes in the microscopic step size m. When increasing the probabilities of non-uniform motion, the microscopic translocation rate m always increased. With both backward and jumping motion, the largest increases occurred. This is likely due to the depend between variance and the step size. Having two types of non-uniform motion should increase the variance more than only one type of motion and leads to the largest increase.

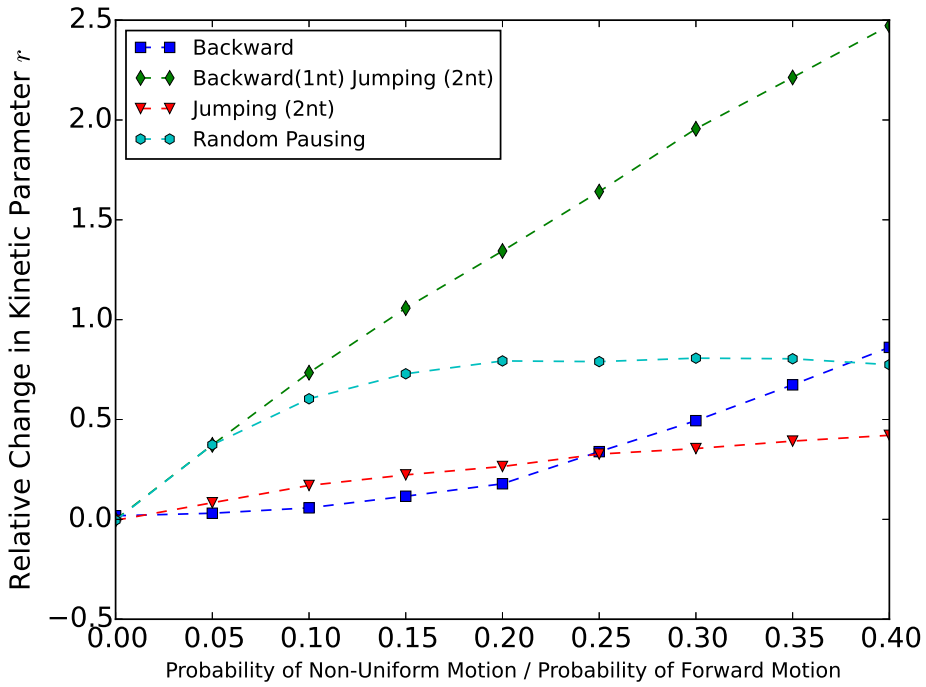


Figure 5.31: Relative changes in the r parameter. Due to the coupling of r with the step size m, the results of the changes in r are similar to those seen in m. With almost all increases to uniform motion, there is increases in r with the most significant occurring with backward and jumping motion.

protein located at the 5' end at time $t = 0$. To start, Equation 5.12 can be expressed as

$$f_0(t) = \mathscr{L}^{-1}\left[\frac{A}{1+nr} \times \frac{1}{k_0+s}\right] + \mathscr{L}^{-1}\left[\frac{1}{k_0+s}\left(\frac{k_t r}{k_d+s}\left(1-\left(\frac{k_t}{k_d+k_t+s}\right)^n\right)\right)\right]. \quad (5.29)$$

The left term in Equation 5.29 can be simplified using the Inverse Laplace Transform to

$$f_0(t) = \left[\frac{A}{1+nr} \times e^{-k_0 t}\right] + \mathscr{L}^{-1}\left[\frac{1}{k_0+s}\left(\frac{k_t r}{k_d+s}\left(1-\left(\frac{k_t}{k_d+k_t+s}\right)^n\right)\right)\right]. \quad (5.30)$$

At time $t = 0$ the left term in Equation 5.30 reduces to

$$f_0(0) = \frac{A}{1+nr} = \frac{A}{1+\frac{L-d}{m}r} = \frac{A}{1+\frac{r}{m}(L-d)} \quad (5.31)$$

when substituting Equation 5.2 in for n (the right term in Equation 5.30 was ignored since parameters k_t, k_d and k_0 are all time dependent). This suggests that at time $t = 0$ there is a coupling of r and m.

Going back and considering Equation 5.30, the left term is related to the amount of protein initially bound at the 5' end and continues to contribute even after time $t = 0$ due to exponential term which is a function of the time t. The right term is also a time dependent function which is a function of both r and $n = (L-d)/m$. Therefore the coupling between r and m changes for $t > 0$.

5.5.4 Futile ATP hydrolysis at the DNA end

As described in the Materials and Methods Section 5.3, none of the Monte Carlo simulations explicitly incorporated any futile hydrolysis at the 5' end of the DNA (state $i = 0$). However, when fitting the ATP data using Equation 5.16, futile hydrolysis k_a was included as a fitting parameter for the *n-step* model. This model that includes the potential for futile hydrolysis was based upon previously published mechanisms for single stranded DNA translocation by helicases [82, 96]. The results from the NLLS fit analysis of the ATP data in Figure 5.32, show when perturbations of backward motion, backward motion and jumping, or random pausing of the protein occurs, the

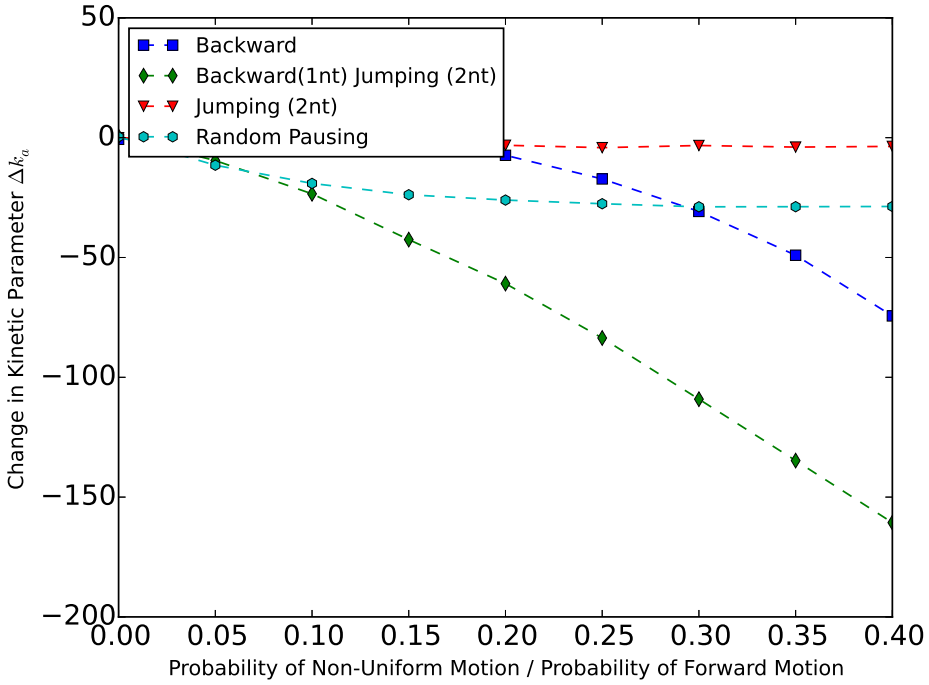


Figure 5.32: Linear changes in the futile hydrolysis rate k_a. None of the simulations included futile hydrolysis. However, when increasing the ratio of the probability of backward motion to the probability of forward motion, backward motion both with and without jumping, and random pausing all resulting in increasing negative rates of futile hydrolysis. Jumping motion had no impact on futile hydrolysis.

ATP time course data is best described by a *n-step* model that includes futile hydrolysis. However, the estimates of the rate of futile hydrolysis were all negative suggesting that non-uniform motion is present. Therefore, when ATP time course data is best described with a *n-step* model that predicts negative futile hydrolysis, additional experiments are needed to determine if non-uniform motion is occurring.

5.6 Conclusions

In order to have a more complete understanding of how proteins translocate along DNA, estimates of the kinetic parameters are needed. To find estimates of the kinetic parameters, time course data can be analyzed using the "*n-step*" model. In general, the macroscopic kinetic parameters (mk_t and c/m) are reliable obtained and give accurate measures of the net forward motion of the protein

and the net ATP hydrolysis coupled to the net forward motion. However, since the "*n-step*" model assumes only uniform directionally biased motion, the microscopic kinetic parameters (m, r, and Δd) associated with proteins that undergo non-uniform are not always accurate.

Shown in Figures 5.9 through 5.32 are detectable signature changes in the kinetic parameterse estimates associated with non-uniform motion of the protein when using a *n-step* model. The estimate of the kinetic step-size (m) is always increased when non-uniform motion occurs. However, it is difficult, or potentially impossible, to know *a priori* if non-uniform motion of the protein occurs. Therefore, the potentially inaccurate estimation of m is concerning given the importance of the magnitude and periodicity of the rate limiting steps in *n-step* enzyme reactions [92, 95, 96, 99, 100, 106]).

To mitigate possible inaccuracy in the estimated kinetic parameters determined from a *n-step* model on time course data, a two-step approach can be used. The first step is to determine what type of experiment and associated kinetic parameters are needed to best describe the observed kinetics of the system. As an example, experiments to determine the translocation of a protein along the DNA include considering the arrival or departure of the protein from a particular point along the DNA [80, 81, 87, 92, 96], the dissociation of the protein from the DNA [87, 92, 96], or the ATP hydrolysis associated with motion of the protein along the DNA [92, 96]. When multiple types of experiments are possible, global analysis of the time courses for each experiment should be used. This type of simultaneous global analysis using the time course data from multiple types of experiments has been done to study the motion of the protein *E. coli* UvrD translocase along single stranded DNA [96]. Another way of studying a protein from multiple aspects would be to use both single molecule and ensemble type experiments.

The second step would require additional experiments when non-uniform motion is suspected. If for example, the global analysis of the time course data using the *n-step* model predicted futile hydrolysis then non-uniform motion might be suspected. Therefore, experiments to determine if indeed futile hydrolysis is occurring should be developed and performed. If those experiments were to suggest no futile hydrolysis is occurring, then non-uniform motion is likely and the microscopic

parameters are potentially inaccurate. Additional experiments might include single molecule experiments. The time course data is from an ensemble experiment. A single molecule approach allows for determining the individual behaviors of a molecule not just the overall behavior of the system.

As a final note, the estimates for the changes in the kinetics parameters are quantitative and will vary depending on the kinetic parameters of the system. The system that I studied was for a protein moving along single stranded DNA, however the results of this study are equally valid for double stranded DNA unwinding, translocation of proteins along ssDNA, and polypeptide translocation.

Chapter 6

Conclusions and future work

DNA has been estimated to be as high as 10^5 per cell per day [108] and repair mechanisms are not completely understood [6]. However, by applying simple statistical mechanics models to biological systems, the physics of the system can be investigated. For example, by incorporating the damage as changes in the energy of the DNA, the damaged DNA model predicts where the damage in the nucleosome is most likely located. DNA damage that increases the elastic energy (increases rigidity) is likely to be located at the linker DNA. DNA damage that decreases the elastic energy (increases flexibility) is likely located near the dyad. These results suggest that the damaged DNA is possibly repositioned in the nucleosome to locations more easily found by repair proteins. The gel experiments with the 5SrDNA sequence suggested that nucleosomes with damaged DNA have different conformations than nucleosomes with healthy DNA. These different conformations are likely due to the damaged DNA being located at different locations in the nucleosome. To more rigorously test the nucleosome breathing model and determine where the damage is located, additional work is needed.

Another approach to understanding repair mechanisms is to consider how the repair proteins are moving along the DNA to find the damage. However, this work did not consider repair proteins in particular, but instead considered the motion of the more general category of motor proteins. One method to determine the kinetic parameters (*e.g.* how fast the protein moves along the DNA) of

motor proteins is to analyze time course data assuming a *n-step* mechanism. This *n-step* model has been used successfully for proteins that exhibit uniform directionally biased motion, but recent work [109] has suggested that protein motion may not be uniform and directionally biased. Therefore, Monte Carlo methods were used to simulate the time course data for motor proteins with non-uniform motion. The kinetic parameters were then estimated by fitting the *n-step* model to the simulated time course data. Results of the analysis show that certain kinetic parameters are likely to be over and underestimated.

For future work, a combination of these two methods could be studied. Repair mechanisms that incorporate both repositioning of the damage to specific locations and proteins that exhibit non-uniform motion would reduce the amount of time required for repair proteins to find, access, and repair DNA damage. If the nucleosome breathing were to reposition the damage to the dyad and/or linker (as predicted by Chapter 3) and the protein were to "jump" along the DNA as shown in Figure 6.1 from linker to dyad to the next linker, repairing damage before "jumping", this would significantly reduce the search region and therefore the time needed for repair proteins to find and repair damage.

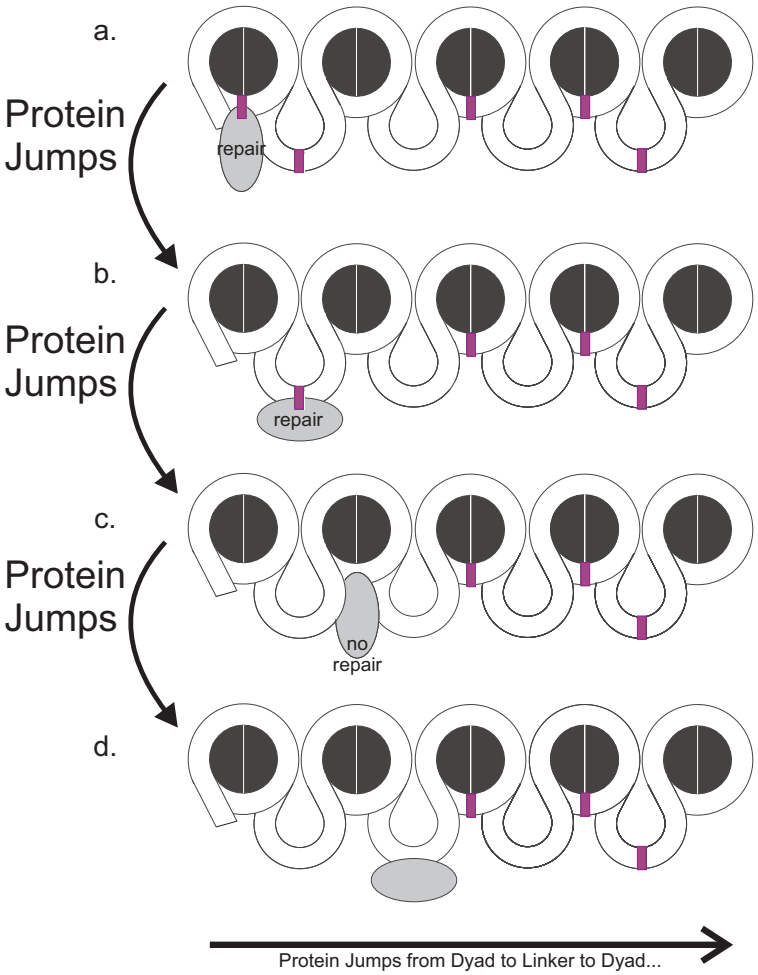


Figure 6.1: Cartoon image of a hypothetical repair mechanism incorporating nucleosome breathing and "jumping". The nucleosome breathing predicts the damage will be repositioned to specific locations and the protein is only those locations. As the protein moves in a directionally biased motion (toward the right), the protein repairs damage at the dyad (a) then "jump" to the linker and repairs damage (b). The protein "jumps" to the next linker (c) but finds no damage to repair. The protein then "jumps" to the next linker (d). This process repeats.

References

[1] A. Gansen, A. Valeri, F. Hauger, S. Felekyan, S. Kalinin, K. Tóth, J. Langowski, and C. A. Seidel, "Nucleosome disassembly intermediates characterized by single-molecule FRET," *Proceedings of the National Academy of Sciences*, vol. 106, no. 36, pp. 15308–15313, 2009.

[2] K. K. Khanna and S. P. Jackson, "DNA double-strand breaks: signaling, repair and the cancer connection," *Nature Genetics*, vol. 27, no. 3, pp. 247–254, 2001.

[3] J. H. Hoeijmakers, "DNA damage, aging, and cancer," *New England Journal of Medicine*, vol. 361, no. 15, pp. 1475–1485, 2009.

[4] T. Kaneo, S. Tahara, and M. Matsuo, "Non-linear accumulation of 8-hydroxy-2'-deoxyguanosine, a marker of oxidized DNA damage, during aging," *Mutation Research/DNAging*, vol. 316, no. 5, pp. 277–285, 1996.

[5] J. Rouse and S. P. Jackson, "Interfaces between the detection, signaling, and repair of DNA damage," *Science*, vol. 297, no. 5581, pp. 547–551, 2002.

[6] S. P. Jackson and J. Bartek, "The DNA-damage response in human biology and disease," *Nature*, vol. 461, no. 7267, p. 1071, 2009.

[7] S. E. LeGresley, J. Wilt, and M. Antonik, "DNA damage may drive nucleosomal reorganization to facilitate damage detection," *Physical Review E*, vol. 89, no. 3, p. 032708, 2014.

[8] C. J. Fischer, A. Saha, and B. R. Cairns, "Kinetic model for the ATP-dependent transloca-

tion of saccharomyces cerevisiae RSC along double-stranded DNA," *Biochemistry*, vol. 46, no. 43, pp. 12416–12426, 2007.

[9] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "Molecular Biology of the Cell , Chapter 4," *New York: Garland Science*, 2002.

[10] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond, "Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution," *Journal of Molecular Biology*, vol. 319, no. 5, pp. 1097–1113, 2002.

[11] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, "Crystal structure of the nucleosome core particle at 2.8 Å resolution," *Nature*, vol. 389, no. 6648, pp. 251–260, 1997.

[12] T. J. Richmond and C. A. Davey, "The structure of DNA in the nucleosome core," *Nature*, vol. 423, no. 6936, pp. 145–150, 2003.

[13] R. Buning and J. van Noort, "Single-pair FRET experiments on nucleosome conformational dynamics," *Biochimie*, vol. 92, no. 12, pp. 1729–1740, 2010.

[14] C. R. Clapier and B. R. Cairns, "The biology of chromatin remodeling complexes," *Annual Review of Biochemistry*, vol. 78, pp. 273–304, 2009.

[15] S. E. Polo, "Reshaping chromatin after DNA damage: the choreography of histone proteins," *Journal of Molecular Biology*, vol. 427, no. 3, pp. 626–636, 2015.

[16] R. M. Clegg, "Fluorescence resonance energy transfer," *Current Opinion in Biotechnology*, vol. 6, no. 1, pp. 103–110, 1995.

[17] R. Roy, S. Hohng, and T. Ha, "A practical guide to single-molecule FRET," *Nature Methods*, vol. 5, no. 6, pp. 507–516, 2008.

[18] M. Löbrich and P. A. Jeggo, "The impact of a negligent G2/M checkpoint on genomic instability and cancer induction," *Nature Reviews Cancer*, vol. 7, no. 11, pp. 861–869, 2007.

[19] Y. Shiloh, "ATM and related protein kinases: safeguarding genome integrity," *Nature Reviews Cancer*, vol. 3, no. 3, pp. 155–168, 2003.

[20] D. C. van Gent, J. H. Hoeijmakers, and R. Kanaar, "Chromosomal stability and the DNA double-stranded break connection," *Nature Reviews Genetics*, vol. 2, no. 3, pp. 196–206, 2001.

[21] P. C. Blainey, A. M. van Oijen, A. Banerjee, G. L. Verdine, and X. S. Xie, "A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA," *Proceedings of the National Academy of Sciences*, vol. 103, no. 15, pp. 5752–5757, 2006.

[22] Y. Qi, K. Nam, M. C. Spong, A. Banerjee, R.-J. Sung, M. Zhang, M. Karplus, and G. L. Verdine, "Strandwise translocation of a DNA glycosylase on undamaged DNA," *Proceedings of the National Academy of Sciences*, vol. 109, no. 4, pp. 1086–1091, 2012.

[23] O. G. Berg, R. B. Winter, and P. H. Von Hippel, "Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. models and theory," *Biochemistry*, vol. 20, no. 24, pp. 6929–6948, 1981.

[24] S. E. Halford and J. F. Marko, "How do site-specific DNA-binding proteins find their targets?," *Nucleic Acids Research*, vol. 32, no. 10, pp. 3040–3052, 2004.

[25] T. Hu, A. Y. Grosberg, and B. Shklovskii, "How proteins search for their specific sites on DNA: the role of DNA conformation," *Biophysical Journal*, vol. 90, no. 8, pp. 2731–2744, 2006.

[26] M. Hedglin and P. J. O'Brien, "Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein," *ACS Chemical Biology*, vol. 5, no. 4, pp. 427–436, 2010.

[27] G. Foffano, D. Marenduzzo, and E. Orlandini, "Facilitated diffusion on confined DNA," *Physical Review E*, vol. 85, no. 2, p. 021919, 2012.

[28] A. Marcovitz and Y. Levy, "Obstacles may facilitate and direct DNA search by proteins," *Biophysical Journal*, vol. 104, no. 9, pp. 2042–2050, 2013.

[29] B. C. Beard, J. J. Stevenson, S. H. Wilson, and M. J. Smerdon, "Base excision repair in nucleosomes lacking histone tails," *DNA Repair*, vol. 4, no. 2, pp. 203–209, 2005.

[30] D. R. Chafin, J. M. Vitolo, L. A. Henricksen, R. A. Bambara, and J. J. Hayes, "Human DNA ligase I efficiently seals nicks in nucleosomes," *The European Molecular Biology Organization Journal*, vol. 19, no. 20, pp. 5492–5501, 2000.

[31] D. Lydall and S. Whitehall, "Chromatin and the DNA damage response," *DNA Repair*, vol. 4, no. 10, pp. 1195–1207, 2005.

[32] B. Suter and F. Thoma, "DNA-repair by photolyase reveals dynamic properties of nucleosome positioning *in Vivo*," *Journal of Molecular Biology*, vol. 319, no. 2, pp. 395–406, 2002.

[33] J. Anderson, A. Thåström, and J. Widom, "Spontaneous access of proteins to buried nucleosomal DNA target sites occurs via a mechanism that is distinct from nucleosome translocation," *Molecular and Cellular Biology*, vol. 22, no. 20, pp. 7147–7157, 2002.

[34] J. Anderson and J. Widom, "Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites," *Journal of Molecular Biology*, vol. 296, no. 4, pp. 979–987, 2000.

[35] K. Polach and J. Widom, "Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation," *Journal of Molecular Biology*, vol. 254, no. 2, pp. 130–149, 1995.

[36] G. Li, M. Levitus, C. Bustamante, and J. Widom, "Rapid spontaneous accessibility of nucleosomal DNA," *Nature Structural & Molecular Biology*, vol. 12, no. 1, pp. 46–53, 2004.

[37] H. Schiessel, "The nucleosome: A transparent, slippery, sticky and yet stable DNA-protein complex," *The European Physical Journal E*, vol. 19, no. 3, pp. 251–262, 2006.

[38] K. Voltz, J. Trylska, N. Calimet, J. C. Smith, and J. Langowski, "Unwrapping of nucleosomal DNA ends: a multiscale molecular dynamics study," *Biophysical Journal*, vol. 102, no. 4, pp. 849–858, 2012.

[39] L. Bai and A. V. Morozov, "Gene regulation by nucleosome positioning," *Trends in Genetics*, vol. 26, no. 11, pp. 476–483, 2010.

[40] F. Battistini, C. A. Hunter, E. J. Gardiner, and M. J. Packer, "Structural mechanics of DNA wrapping in the nucleosome," *Journal of Molecular Biology*, vol. 396, no. 2, pp. 264–279, 2010.

[41] P. Lowary and J. Widom, "New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning," *Journal of Molecular Biology*, vol. 276, no. 1, pp. 19–42, 1998.

[42] A. Thåström, J. Gottesfeld, K. Luger, and J. Widom, "Histone-DNA binding free energy cannot be measured in dilution-driven dissociation experiments," *Biochemistry*, vol. 43, no. 3, pp. 736–741, 2004.

[43] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, I. K. Moore, J.-P. Z. Wang, and J. Widom, "A genomic code for nucleosome positioning," *Nature*, vol. 442, no. 7104, pp. 772–778, 2006.

[44] B. Wu, K. Mohideen, D. Vasudevan, and C. A. Davey, "Structural insight into the sequence dependence of nucleosome positioning," *Structure*, vol. 18, no. 4, pp. 528–536, 2010.

[45] J. S. Choy, S. Wei, J. Y. Lee, S. Tan, S. Chu, and T.-H. Lee, "DNA methylation increases nucleosome compaction and rigidity," *Journal of the American Chemical Society*, vol. 132, no. 6, pp. 1782–1783, 2010.

[46] A. Pérez, C. L. Castellazzi, F. Battistini, K. Collinet, O. Flores, O. Deniz, M. L. Ruiz, D. Torrents, R. Eritja, M. Soler-López, *et al.*, "Impact of methylation on the physical properties of DNA," *Biophysical Journal*, vol. 102, no. 9, pp. 2140–2148, 2012.

[47] V. Subramanian, R. M. Williams, D. L. Boger, and K. Luger, "Methods to characterize the effect of DNA-modifying compounds on nucleosomal DNA," in *Drug-DNA Interaction Protocols*, pp. 173–192, Springer, 2010.

[48] K. Polach, P. Lowary, and J. Widom, "Effects of core histone tail domains on the equilibrium constants for dynamic DNA site accessibility in nucleosomes," *Journal of Molecular Biology*, vol. 298, no. 2, pp. 211–223, 2000.

[49] T. Jenuwein and C. D. Allis, "Translating the histone code," *Science*, vol. 293, no. 5532, pp. 1074–1080, 2001.

[50] B. Fierz, C. Chatterjee, R. K. McGinty, M. Bar-Dagan, D. P. Raleigh, and T. W. Muir, "Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction," *Nature Chemical Biology*, vol. 7, no. 2, pp. 113–119, 2011.

[51] P. Voigt, G. LeRoy, W. J. Drury III, B. M. Zee, J. Son, D. B. Beck, N. L. Young, B. A. Garcia, and D. Reinberg, "Asymmetrically modified nucleosomes," *Cell*, vol. 151, no. 1, pp. 181–193, 2012.

[52] H. A. Cole, J. M. Tabor-Godwin, and J. J. Hayes, "Uracil DNA glycosylase activity on nucleosomal DNA depends on rotational orientation of targets," *Journal of Biological Chemistry*, vol. 285, no. 4, pp. 2876–2885, 2010.

[53] A. Scrima, R. Koníčková, B. K. Czyzewski, Y. Kawasaki, P. D. Jeffrey, R. Groisman, Y. Nakatani, S. Iwai, N. P. Pavletich, and N. H. Thomä, "Structural basis of UV DNA-damage recognition by the DDB1–DDB2 complex," *Cell*, vol. 135, no. 7, pp. 1213–1223, 2008.

[54] J. M. Gale, K. A. Nissen, and M. J. Smerdon, "UV-induced formation of pyrimidine dimers in nucleosome core DNA is strongly modulated with a period of 10.3 bases," *Proceedings of the National Academy of Sciences*, vol. 84, no. 19, pp. 6644–6648, 1987.

[55] M.-R. Duan and M. J. Smerdon, "UV damage in DNA promotes nucleosome unwrapping," *Journal of Biological Chemistry*, vol. 285, no. 34, pp. 26295–26303, 2010.

[56] G. Längst and P. B. Becker, "ISWI induces nucleosome sliding on nicked DNA," *Molecular Cell*, vol. 8, no. 5, pp. 1085–1092, 2001.

[57] C. Suquet and M. J. Smerdon, "UV damage to DNA strongly influences its rotational setting on the histone surface of reconstituted nucleosomes.," *Journal of Biological Chemistry*, vol. 268, no. 32, pp. 23755–23757, 1993.

[58] J. V. Kosmoski, E. J. Ackerman, and M. J. Smerdon, "DNA repair of a single UV photoproduct in a designed nucleosome," *Proceedings of the National Academy of Sciences*, vol. 98, no. 18, pp. 10113–10118, 2001.

[59] P. Prinsen and H. Schiessel, "Nucleosome stability and accessibility of its DNA to proteins," *Biochimie*, vol. 92, no. 12, pp. 1722–1728, 2010.

[60] J. M. Gottesfeld and K. Luger, "Energetics and affinity of the histone octamer for defined DNA sequences," *Biochemistry*, vol. 40, no. 37, pp. 10927–10933, 2001.

[61] H. Qu and G. Zocchi, "The complete bending energy function for nicked DNA," *Europhysics Letters*, vol. 94, p. 18003, apr 2011.

[62] H. Yokoyama and R. Mizutani, "Structural biology of DNA (6-4) photoproducts formed by ultraviolet radiation and interactions with their binding proteins," *International Journal of Molecular Sciences*, vol. 15, no. 11, pp. 20321–20338, 2014.

[63] D. L. Mitchell, T. D. Nguyen, and J. E. Cleaver, "Nonrandom induction of pyrimidine-pyrimidone (6-4) photoproducts in ultraviolet-irradiated human chromatin.," *Journal of Biological Chemistry*, vol. 265, no. 10, pp. 5353–5356, 1990.

[64] A. Fathizadeh, A. B. Besya, M. R. Ejtehadi, and H. Schiessel, "Rigid-body molecular dynamics of DNA inside a nucleosome," *The European Physical Journal E*, vol. 36, no. 21, p. 21, 2013.

[65] M. S. Shukla, S. H. Syed, F. Montel, C. Faivre-Moskalenko, J. BeDNAr, A. Travers, D. Angelov, and S. Dimitrov, "Remosomes: RSC generated non-mobilized particles with approximately 180 bp DNA loosely associated with the histone octamer," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 1936–1941, feb 2010. PMID: 20080697.

[66] L. Siman, I. S. S. Carrasco, J. K. L. da Silva, M. C. de Oliveira, M. S. Rocha, and O. N. Mesquita, "Quantitative assessment of the interplay between DNA elasticity and cooperative binding of ligands," *Physical Review Letters*, vol. 109, p. 248103, dec 2012.

[67] M. J. McCauley, E. M. Rueter, I. Rouzina, L. J. Maher, and M. C. Williams, "Single-molecule kinetics reveal microscopic mechanism by which high-mobility Group B proteins alter DNA flexibility," *Nucleic Acids Research*, vol. 41, pp. 167–181, jan 2013. PMID: 23143110PMCID: PMC3592474.

[68] W. Li, Z.-Q. Sun, P. Xie, S.-X. Dou, W.-C. Wang, and P.-Y. Wang, "Elastic response and length change of single DNA molecules induced by a combination of cisplatin and transplatin," *Physical Review E*, vol. 85, p. 021918, feb 2012.

[69] W. Koopmans, A. Brehm, C. Logie, T. Schmidt, and J. van Noort, "Single-pair FRET microscopy reveals mononucleosome dynamics," *Journal of Fluorescence*, vol. 17, no. 6, pp. 785–795, 2007.

[70] W. Koopmans, R. Buning, T. Schmidt, and J. van Noort, "spFRET using alternating excita-

tion and FCS reveals progressive DNA unwrapping in nucleosomes," *Biophysical Journal*, vol. 97, no. 1, pp. 195–204, 2009.

[71] D. Lovullo, D. Daniel, J. Yodh, D. Lohr, and N. Woodbury, "A fluorescence resonance energy transfer-based probe to monitor nucleosome structure," *Analytical Biochemistry*, vol. 341, no. 1, pp. 165–172, 2005.

[72] M. Tomschik, H. Zheng, K. van Holde, J. Zlatanova, and S. H. Leuba, "Fast, long-range, reversible conformational fluctuations in nucleosomes revealed by single-pair fluorescence resonance energy transfer," *Proceedings of the National Academy of Sciences*, vol. 102, no. 9, pp. 3278–3283, 2005.

[73] K. Tóth, N. Brun, and J. Langowski, "Trajectory of nucleosomal linker DNA studied by fluorescence resonance energy transfer," *Biochemistry*, vol. 40, no. 23, pp. 6921–6928, 2001.

[74] K. Tóth, N. Brun, and J. Langowski, "Chromatin compaction at the mononucleosome level," *Biochemistry*, vol. 45, no. 6, pp. 1591–1598, 2006.

[75] H. S. Tims and J. Widom, "Stopped-flow fluorescence resonance energy transfer for analysis of nucleosome dynamics," *Methods*, vol. 41, no. 3, pp. 296–303, 2007.

[76] J. R. Lakowicz, "Fluorescence anisotropy," in *Principles of fluorescence spectroscopy*, pp. 291–319, Springer, 1999.

[77] R. D. Kornberg, "The molecular basis of eukaryotic transcription," *Proceedings of the National Academy of Sciences*, vol. 104, no. 32, pp. 12955–12961, 2007.

[78] P. B. Becker, "Nucleosome remodelers on track," *Nature Structural and Molecular Biology*, vol. 12, no. 9, pp. 732–734, 2005.

[79] M. D. Szczelkun, "Kinetic models of translocation, head-on collision, and DNA cleavage by type I restriction endonucleases," *Biochemistry*, vol. 41, no. 6, pp. 2067–2074, 2002.

[80] K. Firman and M. D. Szczelkun, "Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement," *The European Molecular Biology Organization Journal*, vol. 19, no. 9, pp. 2094–2102, 2000.

[81] S. E. McClelland, D. T. Dryden, and M. D. Szczelkun, "Continuous assays for DNA translocation using fluorescent triplex dissociation: application to type I restriction endonucleases," *Journal of Molecular Biology*, vol. 348, no. 4, pp. 895–915, 2005.

[82] M. S. Dillingham, D. B. Wigley, and M. R. Webb, "Demonstration of unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed," *Biochemistry*, vol. 39, no. 1, pp. 205–212, 2000.

[83] M. S. Dillingham, D. B. Wigley, and M. R. Webb, "Direct measurement of single-stranded DNA translocation by PcrA helicase using the fluorescent base analogue 2-aminopurine," *Biochemistry*, vol. 41, no. 2, pp. 643–651, 2002.

[84] K. M. Brendza, W. Cheng, C. J. Fischer, M. A. Chesnik, A. Niedziela-Majka, and T. M. Lohman, "Autoinhibition of Escherichia coli Rep monomer helicase activity by its 2B subdomain," *Proceedings of the National Academy of Sciences*, vol. 102, no. 29, pp. 10076–10081, 2005.

[85] B. Sikora, R. L. Eoff, S. W. Matson, and K. D. Raney, "DNA unwinding by Escherichia coli DNA helicase I (TraI) provides evidence for a processive monomeric molecular motor," *Journal of Biological Chemistry*, vol. 281, no. 47, pp. 36110–36116, 2006.

[86] M. K. Levin, M. Gurjar, and S. S. Patel, "A Brownian motor mechanism of translocation and strand separation by hepatitis C virus helicase," *Nature Structural & Molecular Biology*, vol. 12, no. 5, pp. 429–435, 2005.

[87] C. J. Fischer, N. K. Maluf, and T. M. Lohman, "Mechanism of ATP-dependent translocation of E. coli UvrD monomers along single-stranded DNA," *Journal of Molecular Biology*, vol. 344, no. 5, pp. 1287–1309, 2004.

[88] A. N. Kapanidis, E. Margeat, S. O. Ho, E. Kortkhonjia, S. Weiss, and R. H. Ebright, "Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism," *Science*, vol. 314, no. 5802, pp. 1144–1147, 2006.

[89] S. Dumont, W. Cheng, V. Serebrov, R. K. Beran, I. Tinoco, A. M. Pyle, and C. Bustamante, "RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP," *Nature*, vol. 439, no. 7072, pp. 105–108, 2006.

[90] T. M. Lohman, J. Hsieh, N. K. Maluf, W. Cheng, A. L. Lucius, C. J. Fischer, K. M. Brendza, S. Korolev, and G. Waksman, "DNA helicases, motors that move along nucleic acids: lessons from the SF1 helicase superfamily," *The Enzymes*, vol. 23, pp. 303–VII, 2003.

[91] J. A. Ali and T. M. Lohman, "Kinetic measurement of the step size of DNA unwinding by Escherichia coli UvrD helicase," *Science*, vol. 275, no. 5298, pp. 377–380, 1997.

[92] C. J. Fischer and T. M. Lohman, "ATP-dependent translocation of proteins along single-stranded DNA: models and methods of analysis of pre-steady state kinetics," *Journal of Molecular Biology*, vol. 344, no. 5, pp. 1265–1286, 2004.

[93] A. L. Lucius, C. J. Wong, and T. M. Lohman, "Fluorescence stopped-flow studies of single turnover kinetics of E. coli RecBCD helicase-catalyzed DNA unwinding," *Journal of Molecular Biology*, vol. 339, no. 4, pp. 731–750, 2004.

[94] A. L. Lucius, N. K. Maluf, C. J. Fischer, and T. M. Lohman, "General methods for analysis of sequential "*n-step*" kinetic mechanisms: application to single turnover kinetics of helicase-catalyzed DNA unwinding," *Biophysical Journal*, vol. 85, no. 4, pp. 2224–2239, 2003.

[95] A. L. Lucius, A. Vindigni, R. Gregorian, J. A. Ali, A. F. Taylor, G. R. Smith, and T. M. Lohman, "DNA unwinding step-size of E. coli RecBCD helicase determined from single

turnover chemical quenched-flow kinetic studies," *Journal of Molecular Biology*, vol. 324, no. 3, pp. 409–428, 2002.

[96] E. J. Tomko, C. J. Fischer, A. Niedziela-Majka, and T. M. Lohman, "A nonuniform stepping mechanism for E. coli UvrD monomer translocation along single-stranded DNA," *Molecular Cell*, vol. 26, no. 3, pp. 335–347, 2007.

[97] M.-N. Dessinges, T. Lionnet, X. G. Xi, D. Bensimon, and V. Croquette, "Single-molecule assay reveals strand switching and enhanced processivity of UvrD," *Proceedings of the National Academy of Sciences*, vol. 101, no. 17, pp. 6439–6444, 2004.

[98] D. J. Williams and K. B. Hall, "Monte Carlo applications to thermal and chemical denaturation experiments of nucleic acids and proteins," *Methods in Enzymology*, vol. 321, pp. 330–352, 2000.

[99] A. L. Lucius and T. M. Lohman, "Effects of temperature and ATP on the kinetic mechanism and kinetic step-size for E. coli RecBCD helicase-catalyzed DNA unwinding," *Journal of Molecular Biology*, vol. 339, no. 4, pp. 751–771, 2004.

[100] S. Myong, M. M. Bruno, A. M. Pyle, and T. Ha, "Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase," *Science*, vol. 317, no. 5837, pp. 513–516, 2007.

[101] A. Saha, J. Wittmeyer, and B. R. Cairns, "Chromatin remodeling through directional DNA translocation from an internal nucleosomal site," *Nature Structural & Molecular Biology*, vol. 12, no. 9, pp. 747–755, 2005.

[102] T. T. Perkins, H.-W. Li, R. V. Dalal, J. Gelles, and S. M. Block, "Forward and reverse motion of single RecBCD molecules on DNA," *Biophysical Journal*, vol. 86, no. 3, pp. 1640–1648, 2004.

[103] M. M. Martinez-Senac and M. R. Webb, "Mechanism of translocation and kinetics of DNA unwinding by the helicase RecG," *Biochemistry*, vol. 44, no. 51, pp. 16967–16976, 2005.

[104] D.-E. Kim, M. Narayan, and S. S. Patel, "T7 DNA helicase: a molecular motor that processively and unidirectionally translocates along single-stranded DNA," *Journal of Molecular Biology*, vol. 321, no. 5, pp. 807–819, 2002.

[105] J. Kawaoka, E. Jankowsky, and A. M. Pyle, "Backbone tracking by the SF2 helicase NPH-II," *Nature Structural & Molecular Biology*, vol. 11, no. 6, pp. 526–530, 2004.

[106] R. Galletto, M. J. Jezewska, and W. Bujalowski, "Unzipping mechanism of the double-stranded DNA unwinding by a hexameric helicase: quantitative analysis of the rate of the dsDNA unwinding, processivity and kinetic step-size of the Escherichia coli DnaB helicase using rapid quench-flow method," *Journal of Molecular Biology*, vol. 343, no. 1, pp. 83–99, 2004.

[107] M. K. Levin, Y.-H. Wang, and S. S. Patel, "The functional interaction of the hepatitis C virus helicase molecules is responsible for unwinding processivity," *Journal of Biological Chemistry*, vol. 279, no. 25, pp. 26005–26012, 2004.

[108] T. Lindahl, "Instability and decay of the primary structure of DNA," *Nature*, vol. 362, no. 6422, pp. 709–715, 1993.

[109] A. Eastlund, S. S. Malik, and C. J. Fischer, "Kinetic mechanism of DNA translocation by the RSC molecular motor," *Archives of Biochemistry and Biophysics*, vol. 532, no. 2, pp. 73–83, 2013.

Appendix A

Translocation kinetics

A.1 Conditions for the *r* parameter

The conditions for the *r* parameter can thought of in the following way.

Consider a system with only 8 binding sites and a protein that can only move in full steps of $m = 3$ nucleotides. If a protein binds 0,1, or 2 nucleotides from the end, the protein is essentially zero steps from the end with concentration $I_{end}(0) = 3/8$. For a protein that binds 3, 4, or 5 nucleotides from the end, the protein is one step from the end with concentration $I_1(0) = 3/8$. For a protein that binds 6 or 7 nucleotides from the end, the protein is two steps from the end with concentration $I_2(0) = 2/8$. The ratio of binding 0, 1, or 2 steps away from the end at time $t = 0$ is 1:1:1.5. So, the protein is 1.5 times more likely to bind not at the end and $r = 1.5$.

Assume now that the protein can move in less than full steps. If the protein binds at the end, the protein is zero steps from the end with concentration $I_{end}(0) = 1/8$. If the protein binds 1, 2, or 3 nucleotides from the end, the protein is one step from the end with concentration $I_1(0) = 3/8$. If the protein binds 4, 5, or 6 nucleotides from the end, the protein is two steps from the end with concentration $I_2(0) = 3/8$. And finally, if the protein binds 7 nucleotides from the end, the protein is three steps from the end with concentration $I_3(0) = 1/8$. The ratio of binding 0, 1, 2, or 3 steps from the end at time $t = 0$ is 1:3:3:1. So, the protein is 3 times more likely to bind not at the end and $r = 3$. Combining the previous two scenarios, one can assume that *r* should be between 1 and

m. See Reference [92] for more details.

A.2 Variance in *n-step* model kinetic parameters

Variance from the fitting of the Monte Carlo simulated time course data for protein concentration using Equation 5.12 and ADP concentration using Equations 5.16 are shown below. Note that in each case, the *n-step* model was fit to 505 data points from the Monte Carlo simulations.

Table A.1: Variance from the fitting of the time course data for protein concentration using Equation 5.12. The probability of non-uniform motion (*e.g* backward motion, jumping, *etc*.) to the probability of translocation is 0.0, 0.1, 0.2. 0.3, and 0.4. Not surprisingly, as the probability of non-uniform motion increased, there were slight increases to the variance.

p(non-uniform)/p(translocation)	0.00	0.10	0.20	0.30	0.40
Backward	2.29E-08	9.92E-08	1.50E-07	1.63E-07	1.75E-07
Backward and Jumping	2.09E-08	1.75E-07	1.62E-07	1.45E-07	1.38E-07
Jumping	2.06E-08	2.42E-08	2.68E-08	2.83E-08	2.42E-08
Random Pausing	1.79E-08	4.17E-07	5.04E-07	4.92E-07	4.37E-07
Case 5a	2.36E-08	2.47E-07	4.37E-07	5.27E-07	5.96E-07
Case 5b	2.36E-08	7.77E-08	7.07E-07	2.22E-06	4.31E-06
Case 6a	2.36E-08	7.62E-08	4.69E-07	1.35E-06	2.44E-06
Case 6b	2.36E-08	2.18E-08	2.48E-08	2.80E-08	2.08E-08

Table A.2: Variance from the fitting of the time course data of ATP hydrolylsis (ADP concentration) data using Equation 5.16. The fitting with the ADP time course data had much larger variances than the fitting of the protein time course data. The increased variance is likely a result of having held the kinetic parameters from the protein time data constant and only fitting c (ATP hydrolysis per step) and k_a (futitle hydrolysis). For the random pausing, the best fit to the ATP hydrolysis data occurred when assuming the ATP hydrolysis occurred when recovering (exiting) the pause.

p(non-uniform)/p(translocation)	0.00	0.10	0.20	0.30	0.40
Backward	5.51E-05	6.96E-05	3.74E-04	8.27E-04	1.08E-03
Backward and Jumping	2.45E-04	1.06E-03	1.56E-03	2.16E-03	2.07E-03
Jumping	5.16E-05	9.86E-06	2.64E-05	6.94E-05	2.47E-05
Random Pausing (exit)	2.29E-05	2.66E-03	4.77E-03	4.54E-03	4.00E-03
Random Pausing (enter)	2.29E-05	8.70E-03	2.01E-02	2.78E-02	3.35E-02
Random Pausing (enter and exit)	2.29E-05	3.64E-03	8.61E-03	1.12E-02	1.33E-02
Case 5a	2.72E-05	1.96E-03	4.33E-03	6.88E-03	8.51E-03
Case 5b	2.72E-05	5.79E-05	7.28E-04	3.53E-03	7.70E-03
Case 6a	2.72E-05	4.00E-04	5.12E-03	3.32E-02	1.22E-01
Case 6b	2.72E-05	2.52E-05	5.99E-05	9.64E-05	2.50E-04

A.3 Sum of the squared residuals (SRR) in *n-step* model kinetic parameters

Sum of the squared residuals from the fitting of the Monte Carlo simulated time course data for protein concentration using Equation 5.12 and ADP concentration using Equations 5.16 are shown below. Note that in each case, the *n-step* model was fit to 505 data points from the Monte Carlo simulations.

Table A.3: Sum of the squared residuals (SSR) from the fitting of the time course data for protein concentration using Equation 5.12. The probability of non-uniform (*e.g* backward motion, jumping, *etc.*) to probability of translocation is 0.0, 0.1, 0.2. 0.3, and 0.4. Not surprisingly, as the probability of non-uniform motion increased, there were slight increases to the SSR.

p(non-uniform)/p(translocation)	0.00	0.10	0.20	0.30	0.40
Backward	1.16E-05	5.01E-05	7.59E-05	8.24E-05	8.85E-05
Backward and Jumping	1.06E-05	8.84E-05	8.18E-05	7.32E-05	6.95E-05
Jumping	1.04E-05	1.22E-05	1.36E-05	1.43E-05	1.22E-05
Random Pausing	9.01E-06	2.10E-04	2.54E-04	2.48E-04	2.21E-04
Case 5a	1.19E-05	1.25E-04	2.21E-04	2.66E-04	3.01E-04
Case 5b	1.19E-05	3.93E-05	3.57E-04	1.12E-03	2.18E-03
Case 6a	1.19E-05	3.85E-05	2.37E-04	6.82E-04	1.23E-03
Case 6b	1.19E-05	1.10E-05	1.25E-05	1.42E-05	1.05E-05

Table A.4: Sum of the squared residuals (SSR) from the fitting of the time course data of ATP hydrolylsis (ADP concentration) data using Equation 5.16. The fitting with the ADP time course data had much larger SSR values than the fitting of the protein time course data. The increased SSR values is likely a result of having held the kinetic parameters from the protein time data constant and only fitting c (ATP hydrolysis per step) and k_a (futitle hydrolysis). For the random pausing, the best fit to the ATP hydrolysis data occurred when assuming the ATP hydrolysis occurred when recovering (exiting) the pause.

p(non-uniform)/p(translocation)	0.00	0.10	0.20	0.30	0.40
Backward	2.78E-02	3.50E-02	1.88E-01	4.16E-01	5.43E-01
Backward and Jumping	1.23E-01	5.34E-01	7.83E-01	1.08E+00	1.04E+00
Jumping	2.60E-02	4.96E-03	1.33E-02	3.49E-02	1.24E-02
Random Pausing (exit)	1.15E-02	1.34E+00	2.40E+00	2.29E+00	2.01E+00
Random Pausing (enter)	1.15E-02	4.37E+00	1.01E+01	1.40E+01	1.68E+01
Random Pausing (enter and exit)	1.15E-02	1.83E+00	4.33E+00	5.65E+00	6.67E+00
Case 5a	1.37E-02	9.86E-01	2.18E+00	3.46E+00	4.28E+00
Case 5b	1.37E-02	2.91E-02	3.66E-01	1.78E+00	3.87E+00
Case 6a	1.37E-02	2.01E-01	2.57E+00	1.67E+01	6.15E+01
Case 6b	1.37E-02	1.27E-02	3.01E-02	4.85E-02	1.26E-01

A.4 Correlation matrices

The correlation between the kinetic parameters for different types of non-uniform motion was shown below. The rates k_t and k_d are always highly positively correlated likely due to using the constraints in Equations 5.17 and 5.18. Kinetic parameters r and m are also very highly positively correlated as seen in Equations 5.29 - 5.31. For example, the correlation between m and r for backward motion (Table A.5) is 0.98 and was determined by considering the estimated m and r values for p_b/p_t = 0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, and 0.40.

Table A.5: Correlation matrix for backwards motion. This matrix shows the correlation between the estimated kinetic parameters of the *n-step* model found by fitting of the simulated time course data for proteins that with rate k_b would undergo backward motion. The correlation was determined by considered the estimated kinetic parameters such as those found in Table 5.1.

Backward	k_t	k_d	k_{end}	r	m	d	mk_t	P
k_t	**1.00**	0.99	0.96	-0.86	-0.95	-0.98	0.99	0.91
k_d	0.99	**1.00**	0.93	-0.89	-0.96	-0.95	0.99	0.93
k_{end}	0.96	0.93	**1.00**	-0.70	-0.83	-0.99	0.92	0.77
r	-0.86	-0.89	-0.70	**1.00**	0.98	0.76	-0.92	-0.99
m	-0.95	-0.96	-0.83	0.98	**1.00**	0.88	-0.98	-0.99
d	-0.98	-0.95	-0.99	0.76	-0.98	**1.00**	-0.96	-0.83
mk_t	0.99	0.99	0.92	-0.92	-0.98	-0.96	**1.00**	0.00
P	0.91	0.93	0.77	-0.99	-0.99	-0.83	0.00	**1.00**

Table A.6: Correlation matrix for jumping motion. This matrix shows the correlation between the estimated kinetic parameters of the *n-step* model found by fitting of the simulated time course data for proteins that with rate k_j would undergo jumping motion.

Jumping	k_t	k_d	k_{end}	r	m	d	mk_t	P
k_t	**1.00**	0.93	-0.22	-0.96	-0.96	0.77	-0.93	-0.85
k_d	0.93	**1.00**	-0.34	-0.99	-0.99	0.88	-0.99	-0.98
k_{end}	-0.22	-0.34	**1.00**	0.23	0.23	-0.55	0.23	0.39
r	-0.96	-0.99	0.23	**1.00**	1.00	-0.85	1.00	0.95
m	-0.96	-0.99	0.23	1.00	**1.00**	-0.86	1.00	0.95
d	0.77	0.88	-0.55	-0.85	1.00	**1.00**	-0.87	-0.89
mk_t	-0.93	-0.99	0.23	1.00	1.00	-0.87	**1.00**	0.96
P	-0.85	-0.98	0.39	0.95	0.95	-0.89	0.96	**1.00**

Table A.7: Correlation matrix for backwards and jumping motion. This matrix shows the correlation between the estimated kinetic parameters of the *n-step* model found by fitting of the simulated time course data for proteins that with rates k_b and k_j would undergo backward and jumping motion.

Backward and Jumping	k_t	k_d	k_{end}	r	m	d	mk_t	P
k_t	**1.00**	0.99	-0.91	-0.91	-0.92	-0.44	0.97	0.89
k_d	0.99	**1.00**	-0.89	-0.95	-0.96	-0.52	0.99	0.93
k_{end}	-0.91	-0.89	**1.00**	0.71	0.72	0.14	-0.82	-0.65
r	-0.91	-0.95	0.71	**1.00**	1.00	0.75	-0.98	-1.00
m	-0.92	-0.96	0.72	1.00	**1.00**	0.74	-0.99	-0.99
d	-0.44	-0.52	0.14	0.75	-0.99	**1.00**	-0.63	-0.77
mk_t	0.97	0.99	-0.82	-0.98	-0.99	-0.63	**1.00**	0.97
P	0.89	0.93	-0.65	-1.00	-0.99	-0.77	0.97	**1.00**

Table A.8: Correlation matrix for random pausing. This matrix shows the correlation between the estimated kinetic parameters of the *n-step* model found by fitting of the simulated time course data for proteins that with rate k_p randomly pause (stall).

Random pausing	k_t	k_d	k_{end}	r	m	d	mk_t	P
k_t	**1.00**	0.97	0.94	-0.99	-0.99	-0.32	0.95	0.99
k_d	0.97	**1.00**	0.99	-0.94	-0.96	-0.08	1.00	0.95
k_{end}	0.94	0.99	**1.00**	-0.91	-0.94	0.00	1.00	0.93
r	-0.99	-0.94	-0.91	**1.00**	1.00	0.41	-0.92	-1.00
m	-0.99	-0.96	-0.94	1.00	**1.00**	0.33	-0.95	-1.00
d	-0.32	-0.08	0.00	0.41	-0.95	**1.00**	-0.03	-0.36
mk_t	0.95	1.00	1.00	-0.92	-0.95	-0.03	**1.00**	0.94
P	0.99	0.95	0.93	-1.00	-1.00	-0.36	0.94	**1.00**

A.5 Uncertainties in estimated kinetic parameters

Samples of the uncertainties in the kinetic parameters estimates are shown below. The uncertainties were calculated using a bootstrapping method. Note: Figures 5.9 - 5.32 do not include the uncertainties because the uncertainties were smaller than the size of the points.

Table A.9: Uncertainties in estimates of the kinetic parameters for proteins with backward motion.

p_b/p_t	0.00	0.10	0.20	0.30	0.40
k_t	5.93E-03	7.46E-03	6.19E-03	3.90E-03	2.47E-03
k_d	2.20E-04	3.76E-04	4.51E-04	4.24E-04	3.59E-04
k_{end}	3.25E-03	6.90E-03	1.23E-02	2.97E-02	8.03E-02
m	1.08E-04	2.54E-04	4.19E-04	5.12E-04	5.33E-04
d	2.98E-04	6.74E-04	1.22E-03	2.25E-03	4.13E-03
r	2.06E-04	4.56E-04	8.45E-04	2.13E-03	6.67E-03
c	6.14E-06	4.62E-06	1.53E-05	5.35E-05	6.31E-05
k_a	1.83E-03	9.99E-04	2.43E-03	7.05E-03	7.47E-03
P	2.75E-06	6.86E-06	1.28E-05	1.92E-05	3.09E-05
mk_t	1.22E-03	2.19E-03	2.88E-03	3.11E-03	2.95E-03
c/m	6.17E-06	3.81E-06	1.01E-05	1.85E-05	1.87E-05

Table A.10: Uncertainties in estimates of the kinetic parameters for proteins with backward and jumping motion.

$p_{b,j}/p_t$	0.00	0.10	0.20	0.30	0.40
k_t	5.93E-03	4.14E-03	2.24E-03	2.11E-03	2.27E-03
k_d	2.20E-04	4.94E-04	4.16E-04	3.34E-04	4.21E-04
k_{end}	3.25E-03	1.51E-02	2.65E-02	6.03E-02	1.04E-01
m	1.08E-04	3.34E-04	3.78E-04	5.52E-04	1.20E-03
d	2.98E-04	1.18E-03	1.59E-03	2.52E-03	2.97E-03
r	2.06E-04	1.04E-03	2.05E-03	5.54E-03	1.14E-02
c	6.14E-06	1.91E-05	3.41E-05	5.35E-05	6.31E-05
k_a	1.83E-03	3.67E-03	5.24E-03	7.05E-03	7.47E-03
P	2.75E-06	1.62E-05	2.43E-05	3.00E-05	3.97E-05
mk_t	1.22E-03	3.52E-03	3.52E-03	3.22E-03	3.96E-03
c/m	6.17E-06	1.11E-05	1.46E-05	1.85E-05	1.87E-05