REVEALING REMODELER FUNCTION: VARIED AND UNIQUE

By

Allen Eastlund

Submitted to the graduate degree program in 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.

_____

Chairperson Dr. Christopher J. Fischer

_____

Dr. Phillip Baringer

_____

Dr. Steven Hawley

_____

Dr. Carey Johnson

_____

Dr. Adrian Melott

Date Defended: May 1st, 2015

The Dissertation Committee for Allen Eastlund

certifies that this is the approved version of the following dissertation:

REVEALING REMODELER FUNCTION: VARIED AND UNIQUE

_____

Chairperson Dr. Christopher J. Fischer

Date approved: May 13th, 2015

Abstract

Chromatin remodelers perform a necessary and required function for the successful expression of our genetic code. By modifying, shifting, or ejecting nucleosomes from the chromatin structure they allow access to the underlying DNA to the rest of the cell's machinery. This research has focused on two major remodeler motors from major families of chromatin remodelers: the trimeric motor domain of RSC and the motor domain of the ISWI family, ISWI. Using primarily stopped-flow spectrofluorometry, I have categorized the time-dependent motions of these motor domains along their preferred substrate, double-stranded DNA. Combined with collected ATP utilization data, I present the subsequent analysis and associated conclusions that stem from the underlying assumptions and models. Interestingly, there is little in common between the investigated proteins aside from their favored medium. While RSC exhibits modest translocation characteristics and highly effective motion with the ability for large molecular forces, ISWI is not only structurally different but highly inefficient in its motion leading to difficulties in determining its specific translocation mechanics. While chromatin remodeling is a ubiquitous facet of eukaryotic life, there remains much to be understood about their general mechanisms.

Acknowledgements

I would like to thank my graduate committee, Dr. Phil Baringer, Dr. Adrian Melott, Dr. Carey Johnson, and Dr. Steven Hawley for their acceptance of this dissertation as proof that these last 6 years have been worth something.

I would like to thank my supervisor Dr. Chris Fischer for your guidance, help, advice in both research and life, and constant pressure-relieving humor. I now understand that the indentured slavery of a graduate student is worth more than two bags of Doritos and that 100 trillion Zimbabwe dollars is not quite what it seems.

I would like to thank my lab mates Shuja Malik for teaching me all the biochemistry I know, Koan Briggs for reminding and reminding and reminding me how a scientist should look at problems, and Gada Al-Ani for being a model mother to follow in parenting, and incredibly supportive and helpful when nothing was working in the lab. As great friends, you are the foundation upon which this research was conducted.

I would like to thank my mother for allowing me the independence to show that I can.

I would like to thank my father for showing me the responsibility to do.

I would like to thank Wade and Sarah Rush for being the best friends I could have wished for this period of my life. Without hunting, camping, canoeing, games, drinks, hard work, fires, and good company I would have been lost.

Finally, I would like to thank all the friends and family who have directly and indirectly influenced me at any point in this last era of my life. You are legion and I owe you more than I can say.

To my loving wife and extraordinary son.

Everything is for you.

Table of Contents

Table of Figures

Chapter 1 – Introduction

The fullness of human physical possibility is written into our genetic code; a code so complex that it takes 3 billion base pairs of deoxyribonucleic acid (DNA) across 24 chromosomes and 21,000 genes to define what it means to be human[1]. These billions of base pairs of our genetic material require two linear meters of DNA which must fit into the nucleus of a cell which has a volume between 0.9 and 4 cubic nanometers (nm)[2]. To completely package this genetic story into each cell of our body requires multiple levels of compression. The basal level consists of 147 base pair (bp) of double-stranded DNA wrapped around a core complex of eight histone proteins; this complete structure is called a nucleosome core particle[3–5]. Through multiple rounds of the cellular life cycle, the various stages of DNA compression must be done, undone, and redone repeatedly to allow for a litany of required processes including, but not limited to, duplication, translation, and repair. For the proteins responsible for performing these functions, they must have access to specific sections of DNA. In fact, regulation of the compression of DNA is one of the means by which the cell controls the expression of many genes. However, at any given moment the majority of our DNA is inaccessibly woven into cylindrical structures called nucleosome core particles (NCPs). For the necessary repairs and processes to be completed by their respective factors, the NCP structures must somehow be rearranged[6]. To perform this task there are multiple families of proteins called ‘remodelers’ which are responsible for this rearrangement. They are the root mechanism for the unpacking of this structure through modification, movement, or ejection of nucleosomes[7,8]. These remodeler proteins are

ubiquitous across all the major eukaryotic model organisms (*e.g.*, fruit fly *D. melanogaster*, worm *C. elegans*, fungus *S. cerevisiae*, *H. sapiens*) in near-identical copies of each other. Remodeling proteins are subdivided into four major families (SWI/SNF, ISWI, CHD, and INO80); and each family is defined by a conserved core unit that converts chemical energy stored in the bonds of adenosine triphosphate (ATP) into mechanical energy[9]. Each of these families, and even distinct instances from a family, perform their operations through different means[7,10–12]. Due to their vital and important participation in genetic expression, significant problems arise when remodelers are not allowed to complete their role or their function is altered through mistranslation or damage. Abnormal instances of these proteins have been linked to multiple forms of cancer, including brain, breast, and blood, as well as neurodevelopmental disorders[13–19].

Given the importance and ubiquity of these proteins, it is essential to understand their precise microscopic operations. While these remodelers operate exclusively on the combined structure of DNA-wrapped histone complexes, they do so through different mechanisms based on their exact makeup and function[9,20]. Specifically, investigations of the interaction between remodelers and double stranded DNA are necessary to understand their remodeling actions, step-by-step. A grasp of these discrete steps provides a window into how modifications of these proteins affects their DNA-related functions at a truly detailed level; once the specific workings of a protein are elucidated, new avenues of research into detection and correction of aberrant behavior will open. Through a series of energy transduction and double-stranded DNA translocation experiments, I present my findings on the catalytic subunits of two of the major families of remodelers: SWI/SNF and ISWI.

Chapter 2 – Biological Background

2.1 - DNA

Deoxyribonucleic acid (DNA) was originally discovered in 1869 after being the primary genetic material for single and multicellular life for billions of years[21]. Originally termed ‘nuclein’ due to its location within the nucleus of cells, it is comprised of four distinct molecules, termed nucleobases (*i.e.*, cytosine (C), guanine (G), thymine (T), adenine (A)). These nucleobases are planar molecules, either a purine (A and G) or a pyrimidine (T and C), which are then connected to a sugar and a phosphate group. This construct is called a ‘nucleotide’ (nt). Multiple nucleotides are linked together with phosphate bonds to form the backbone of a single long strand of chained nucleotides[22]. This composition is called ‘single-stranded DNA’. Regardless of length, the backbone of a single strand of DNA is not symmetric; at one end is a sugar molecule and at the

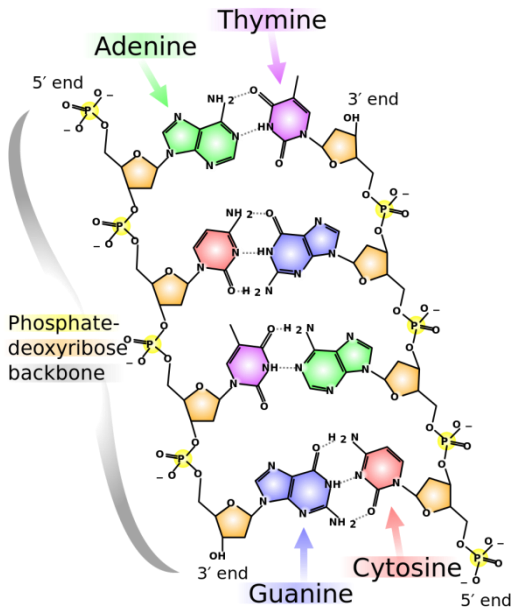


Figure 1 – Double Stranded DNA. An example of double-stranded DNA containing 4 base pairs. Nucleobases (Thymine (T - purple), Adenine (A - green), Guanine (G - blue), and Cytosine (C - red)) are attached to sugar molecules (orange) and phosphate groups (yellow) to form nucleotides. The sugar of one nucleotide then bonds to the phosphate group of another to form the phosphate-deoxyribose backbone. The terminal molecules of each backbone define the 3’ and 5’ end; a sugar at 3’ and a phosphate group at 5’. Image courtesy of Madeleine Price Ball through the CC0 1.0 license.

other, a phosphate group. These ends of the strand are defined as the 3 prime (3') and 5 prime (5') ends; the 3' end ending in a terminal sugar ring and the 5' end having a terminal phosphate group[2], see Figure 1. Due to their highly interlocking shapes, the individual nucleobases of two single stranded chains of nucleotides are paired to specific partners (G to C, T to A) through hydrogen bonding to form double-stranded DNA[23]. In 1953, James Watson and Francis Crick finally identified, through X-ray diffraction experiments, the double helix structure of DNA[24]. All three major forms of complex life (prokaryotic, eukaryotic, and archaea) on Earth rely upon DNA in varying forms to store the information needed to construct every facet of their biological machinery[2]. While prokaryotic life lacks internal structures such as organelles and a nucleus, eukaryotic life is defined by the presence of these structures and frequently has significantly more genetic material than prokaryotic life. This additional material is generally restricted to an organelle within a eukaryotic cell called the nucleus. The nucleus is significantly smaller than the cell itself and therefore requires extra systems for the compaction and storage of such a vast amount of genetic material in a limited volume. Evolution has led eukaryotes to develop 'chromatin', an efficient means of packaging this molecular information. During the process of cellular division eukaryotic DNA is found bundled in *chromosomes*, of which humans have 23. A chromosome is the highest level of organization that the DNA of a cell can achieve and each contain thousands of genes encompassing the information for life. A gene is a region comprised of thousands to tens of thousands of base pairs of DNA. During the majority of the lifetime of a cell, DNA is unstructured within the nucleus with chromosomes only being required during cellular division. The chromosome's purpose is to provide a large-scale structure that can be separated from its clone during division. However, through the majority of a eukaryotic cell's life cycle DNA is in an amorphous, globular structure called *chromatin*.

2.2 - Chromatin

At the lowest level of organization, furthest from the large and orderly chromosome, 146 base pairs of double-stranded DNA is wrapped 1.6 times around a formation of eight histone proteins, two each of histones H2A, H2B, H3, and H4[25] (Figure 3). Histone proteins are highly basic proteins found almost exclusively in eukaryotic cells that are necessary for packaging the nuclear DNA into a compact form[2]. They come in 5 major families: H1/H5, H2A, H2B, H3, and H4[26]. H1 and H5 are linker histones that are responsible for higher order chromatin structure, bundling the beads on a string into tighter structures. Two molecules each of histones H2A, H2B, H3, and H4 form an eight-histone complex called a nucleosome[27]. There are multiple locations where this complex interacts with DNA, leading to the creation of a novel structure. This assembly of DNA and histone proteins is called a '*nucleosome core particle*' (NCP)[28]. The DNA exiting an NCP in either direction is called '*linker* DNA' and connects all the NCPs of an entire chromosome in the manner of "beads on a string", see Figure 2, middle image. Upon the addition of further architectural proteins, this formation is bundled into twisted fibers which, when looped together with various scaffolding proteins, form chromatin (Figure 2, right image). However, at this level of compaction, the DNA is meant for long term storage only to be unfurled and strung out when it is needed. The most biologically active state of cellular DNA is at the level of "beads on a string". This DNA is in a constant

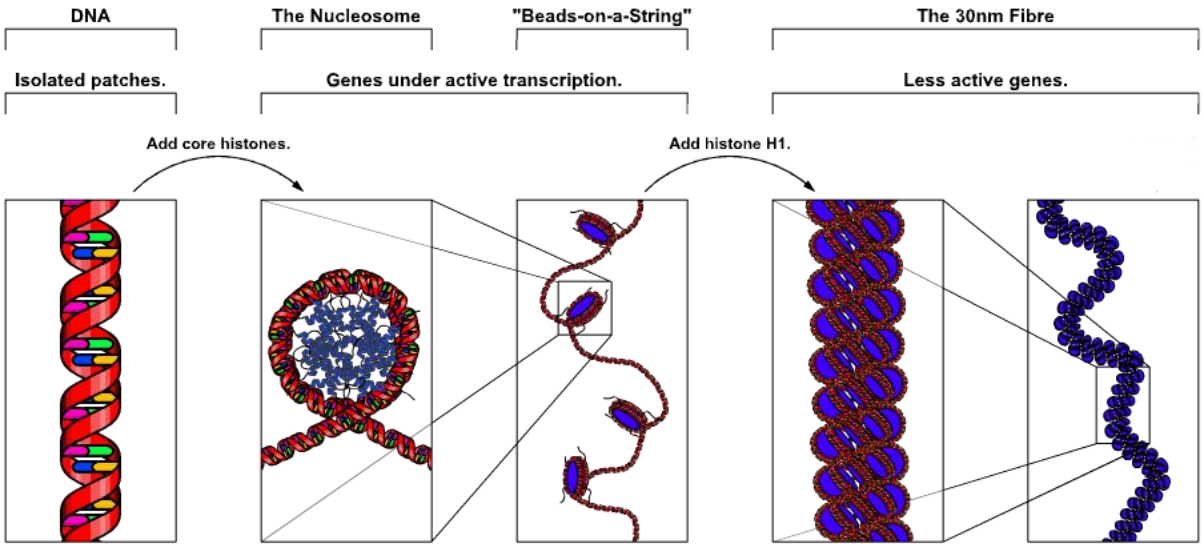
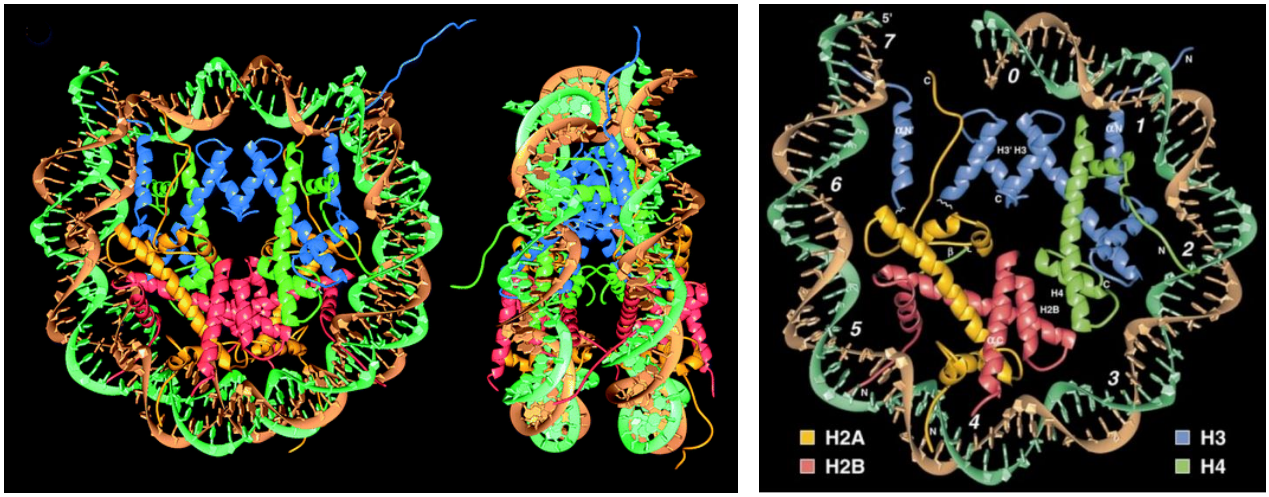


Figure 2 – Chromatin. Chromatin is a highly compact biological data storage mechanism comprised of several hierarchical levels. At the highest level, humans have 24 chromosomes of looped chromatin. Further down, the loops resolve into fibers, the fibers into bundled nucleosomes, and finally the nucleosomes into wrapped DNA. (Image courtesy of Richard Wheeler through the CC-BY-SA-3.0 / GFDL license)

flux of activity consisting of frequent rounds of repair, transcription, and translation[29]. As the average gene requires 10-15 kilobase pair of DNA to be fully coded[30] (1 kilobase pair = 1,000 base pair), and a single nucleosome and associated linker DNA requires essentially 200 ± 40 base pair[27], the information any gene contains is bundled around hundreds of NCPs. When the DNA is wrapped around the nucleosome in this fashion, it has been shown to share 120 protein-DNA interaction points consisting of basic histone residues forming ionic bonds with the acidic phosphate bonds of the backbone of double-stranded DNA[31]. These interactions are not distributed uniformly around the surface, however, but are instead primarily located in two regions named the "histone folds"[5]. While the DNA remains bundled into an NCP it is largely inaccessible to modifications[32], thus mechanisms by which the DNA becomes accessible are required. The device the cell uses to unbundle its nucleosomes and thereby make the information available to the greater proteome is called the *chromatin remodeler.*



A B

Figure 3 – The Structure Of A Nucleosome. A (left) The structure of DNA (brown and turquoise) wrapped around the histone octamer (blue: H3; green: H4; yellow: H2A; red: H2B) to form a nucleosome. The views are down the DNA superhelix axis for the left particle and perpendicular to it for the right particle. Linker DNA would exit from the top of the two forms on the left and right. B (right) Half of a nucleosome core particle labeling the super helix axis locations (SHLs). The view is down the super helix axis with the dyad axis aligned vertically. The central base pair through which the dyad passes is above the SHL0 label, 0 (SHL,). Each SHL label represents one further DNA double helix turn away in a clockwise direction from SHL0 (1–7). Not shown are SHL-1 through -6, which are DNA double helix turns away from SHL0 in a counter-clockwise direction. Adapted by permission from Macmillan Publishers Ltd: Nature, Luger et al., Crystal structure of the nucleosome core particle at 2.8 A resolution. copyright 1997.

2.3 – Chromatin Remodelers

While there are multiple systems by which chromatin is modified, the research herein focuses on a particular family of proteins that are directly responsible for physically altering the chromatin array: 'chromatin remodeler complexes' (remodelers). These motor proteins are related to a larger classification umbrella called 'helicase-like' proteins: proteins that utilize the free-energy obtained from the binding and hydrolysis of ATP to interact with DNA somehow. While helicases themselves track along and separate the two joined backbones of double-stranded DNA, remodelers use linker DNA as a tracking mechanism to systematically alter the NCP array and allow other cellular machinery access to the underlying DNA. Conversely, in the absence of freely available ATP molecules, these motors are unable to perform their function. Due to this reliance on an available chemical energy source, they are titled ATP-dependent chromatin remodelers. As noted earlier, at the heart of these remodeler complexes lies their conserved catalytic motor domain. The helicase-like protein superfamily SF2 contains four major subfamilies: ISWI, SWI/SNF, CHD, and INO80[9,20,33,34]; each subfamily being defined by their distinct catalytic domain. While this motor domain can contain minor structural variations between organisms, its action remains largely the same. Each subfamily is then further divided into specific proteins by any additional subunits with which they are built. These extra subunits confer differing traits to each protein within each subfamily. Even specific remodelers within a subfamily can be differentiated by form and function based upon the utility of the added subunits (*e.g.*, recognition, signaling, recruitment, *etc.*). For example, the model organism *S. cerevisea* (yeast) has two different ISWI family remodelers, ISW1a and ISW1b. These remodelers both contain the Isw1 ATPase as their catalytic subunit and differ only in their added subunits. ISW1a contains the Ioc3 subunit while ISW1b contains Ioc2 and 4. Each of these remodelers is a distinct member of the ISWI family with characteristic remodeling function but they all have the ISWI motor unit at their foundation.

There are multiple mechanisms by which chromatin is remodeled, including directional DNA translocation[35], changing the inter-nucleosomal spacing[36], or complete ejection of the nucleosome from the

NCP[37], though the specific kinetics of how this machinery operates is largely unknown. While static crystal structures of several chromatin remodelers in complex with both DNA and nucleosomes exist[38–41], these are frozen images of protein-NCP complexes in the absence of ATP. They provide no dynamic information of how these proteins specifically remodel the nucleosomes. There are structure- and assay-based theories on how the activity is accomplished, but these remain heavily debated topics in current fields[42–45]. Single-molecule experiments, ensemble assays, and energetic models have all been used to advance mechanistically different paradigms. To further muddy these waters, each fundamental remodeler motor has its own distinct shape and charge configuration which governs the interaction with the nucleosome; any additional subunit can only further complicate these interfaces. The only seeming constant through all the competing models is that DNA translocation seems to be required for remodeling activity to be present. The necessity of clarifying the mechanism of DNA translocation by chromatin remodelers arises from the essential role of this activity in current models of nucleosome remodeling[42,46–48].

Moreover, these remodelers are distinctly important due to their role in the *correct* expression of the information contained in chromatin. DNA damage, incorrect transcription, or unexpected exposure of underlying genes can have profound effects on a cell or organism. While there are many mechanisms by which DNA is repaired during the stages of a cell's life cycle, there are significantly fewer that control

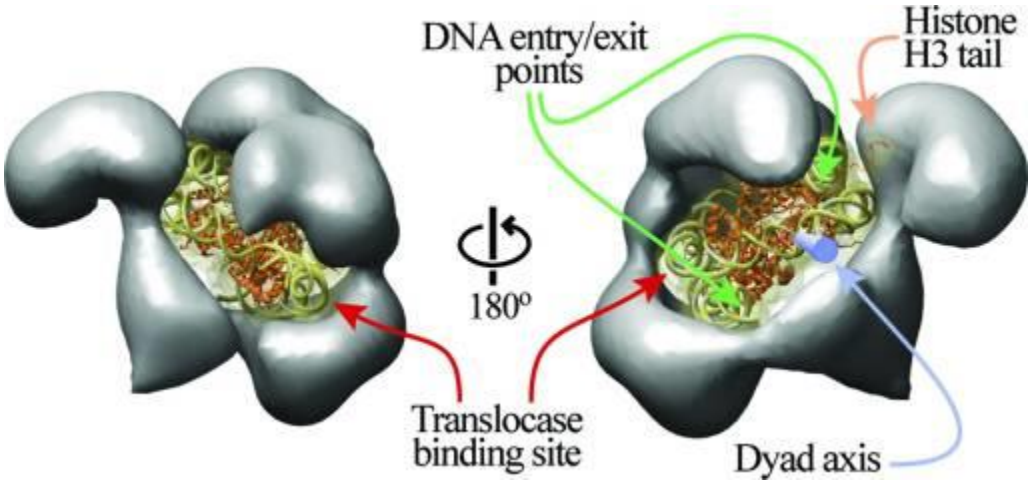


Figure 4 – Model of Nucleosome Binding by RSC. The chromatin remodeler RSC was imaged using electron microscopy (grey) and a known structure of a nucleosome (colored ribbon model) was manually fitted into the binding pocket.[118] Image courtesy of Leschziner *et al.* 'Conformational flexibility in the chromatin remodeler RSC observed by electron microscopy and the orthogonal tilt reconstruction method. *Proc. Natl. Acad. Sci. USA*, **104**, 2007. Copyright (2007) National Academy of Sciences, U.S.A.

chromatin remodelers. Due to the remodeler's important role in genetic expression, aberrant expression, or epigenetic (Appendix A) variation of these proteins can lead to various cancerous cell lines. Specifically, mutations in the human remodeler hSNF5/INI1 (SWI/SNF family) alone have been shown to be present in large numbers in rhabdoid tumors, carcinoma, medulloblastoma, neuroectodermal tumors, and some leukemias[14,15,49]. The ISWI remodeler from the ISWI family is also a target for O-β-D-*N*-acetylglucosamine (O-GlcNAc) signaling (O-GlcNAcylation) shown to be present in high levels in numerous breast, colon, and lung cancer lines[50–52]. In fact, some cancer lines are able to advantageously select for various mutations in remodeler proteins present in their proteome (Appendix A) to further their evolutionary cause[53].

Chromatin has been the dominant method for storing genetic information for billions of years. Chromatin remodelers evolved in parallel with this packaging to resolve the issues that stem from the high density storage of genetic information. Since chromosomal structure must change in response to many different forces (*e.g.*, epigenetics, disease, trauma, *etc.*), remodelers are host to many variations in their behavior. We must understand them at a basal level, to understand their microscopic machinery before we can begin to exert control over the deeply complex expression of our genome. This dissertation attempts to determine the mechanistic interactions of the catalytic domains of the SWI/SNF and ISWI family, RSC and ISWI, respectively, on their substrate of choice: double-stranded DNA. Examining these specific examples may uncover commonalities between these two remodelers which, should they exist, will lead to greater understandings of chromatin remodelers as a family.

Chapter 3 – Modeling and Data Analysis

One of the major stages in the analysis of any collected data is the ability to accurately model the underlying mechanics. This capacity is especially important in terms of biological ensemble experiments containing ~10^{14}-10^{20} simultaneous reactions. Given the sheer number of interactions involved, any experimental output signal will be the averaged mechanical description of all processes involved. Therefore it is important to be able to suitably determine the mechanical steps that create a best fit to the output data. The generalized process through which the remodeler functional mechanisms were interpreted was three-fold: conceive of the underlying assumptions that are present due to experimental conditions, produce a systematic model based on mechanistic ideals and determine a solution to the underlying equations, and globally fit the data across multiple variables to determine the microscopic kinetic parameters that may be in play during processive translocation.

3.1 – Simplifying Assumptions

The models created for any of the data analyses contained herein start from the most simple translocation model with a basic experimental assumption: the protein must be in a concentration range to encourage 1:1 protein:DNA binding. This assumption is important for several reasons. When a DNA-interacting protein like a chromatin remodeler is in solution, its tertiary structure and electrostatic

distribution defines how tightly it binds to the substrate as well as its spatial orientation. If there is a large excess of DNA exiting the binding site of the initially bound molecule, there is a chance that additional proteins may bind to the same DNA fragment. Having more than one protein bound to a DNA segment creates steric complications that interfere with the normal translocation operation of a remodeler. Because of this restriction the experiments must be run in a situation where the number of DNA fragments is in large excess (>4-10x) of the number of proteins in solution. This significantly reduces the probability of two proteins being bound to the same backbone and reduces any related inaccuracy in the final result.

3.2 – Model Creation

Translocation model production begins with the simplest arrangement possible: binding, forward movement, and dissociation (unbinding)[54]. Experimentally, this begins with populations of multi-base single- or double-stranded DNA segments and protein in solution. Upon mixing, the DNA becomes bound to the protein binds until equilibrium is reached, which leaves behind both bound and unbound fractions. A protein bound to a DNA substrate will have various parameters in our model that can be constrained, see Figure 5. The length of the DNA substrate is denoted as L and is measured in base pairs. Of this overall length, the protein will have a given number of contact sites that will occlude a quantity of base pairs on the substrate. The total distance between the first and last occluded base pair is labeled d, measured in base pairs. Finally, the size and physical structure of the protein will create a region around the DNA in which it is impossible for additional proteins to bind due to steric hindrance. This is termed the 'occluded site size' and is labeled b [bp]. In this scenario a protein is free to un- and re-bind stochastically while the macroscopic ratio of unbound to bound remains unchanged due to the extreme number of interactions ($>\sim 10^{14}$). Upon

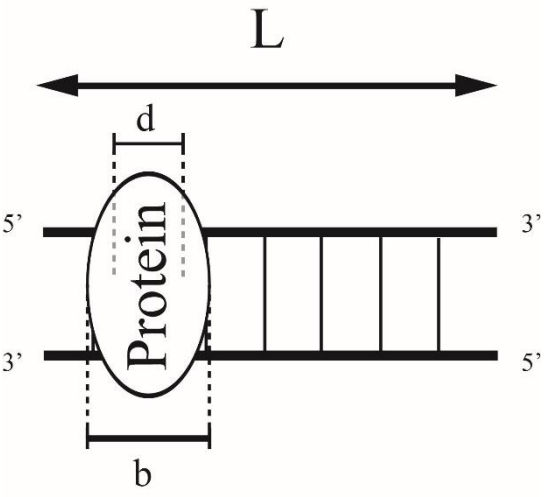


Figure 5 – Bound Protein. After a protein binds to a DNA substrate of length *L* [bp], it begins in a pre-translocative state with certain model parameters. The contact points that a protein shares with the DNA substrate cover a span of the backbone, *d* [bp]. The enzyme also sterically prohibits additional proteins from binding to the DNA within a given range around its own binding site. This distance is listed as *b* [bp].

the addition of ATP, the proteins will then start their process of DNA translocation, see Figure 6. Initially bound populations begin in a pre-translocative state (Figure 6A). In the simplest model, the protein has two possibilities at this point: forward translocation or dissociation. Here we define the rate at which certain processes are performed as 'microscopic rate constants'. These are the tempo at which individual steps in a complex process take place on average across all interactions. For example, we use the microscopic rate constants of k_t and k_d to denote translocation and dissociation respectively. Their units are $[s^{-1}]$ and constrain how frequently the proteins move forward or fall off from the substrate. From these constants, we further define a unitless value *processivity* (P) as $P = \frac{k_t}{k_t + k_d}$. This ratio gives a probability of forward translocation from a pre-translocative state; the value 1-*P* is then the probability of dissociation. If

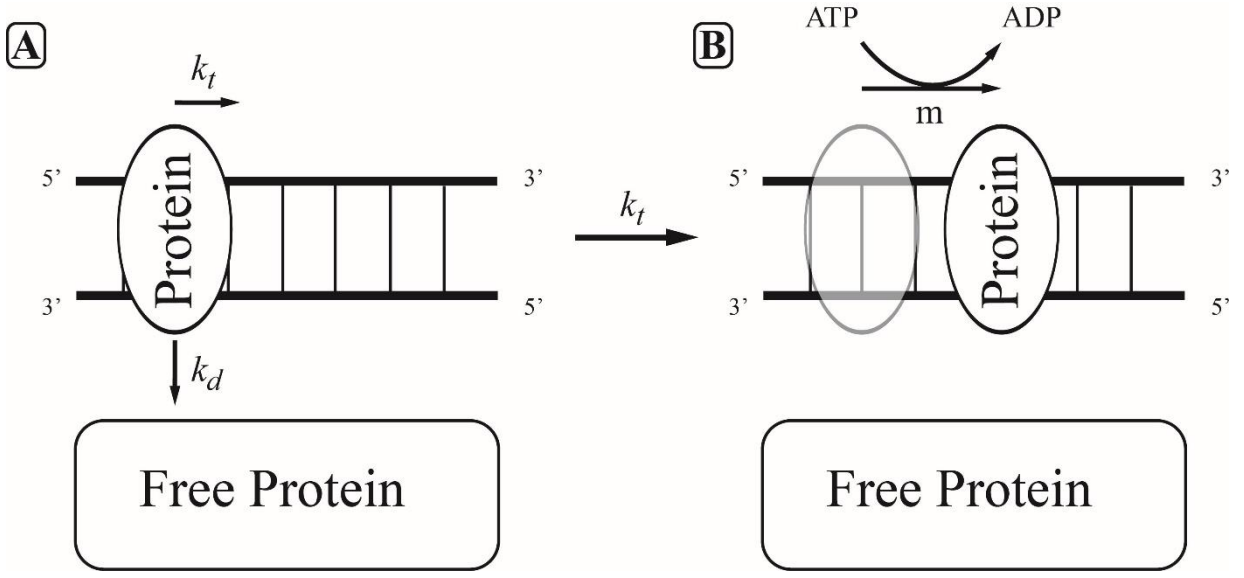


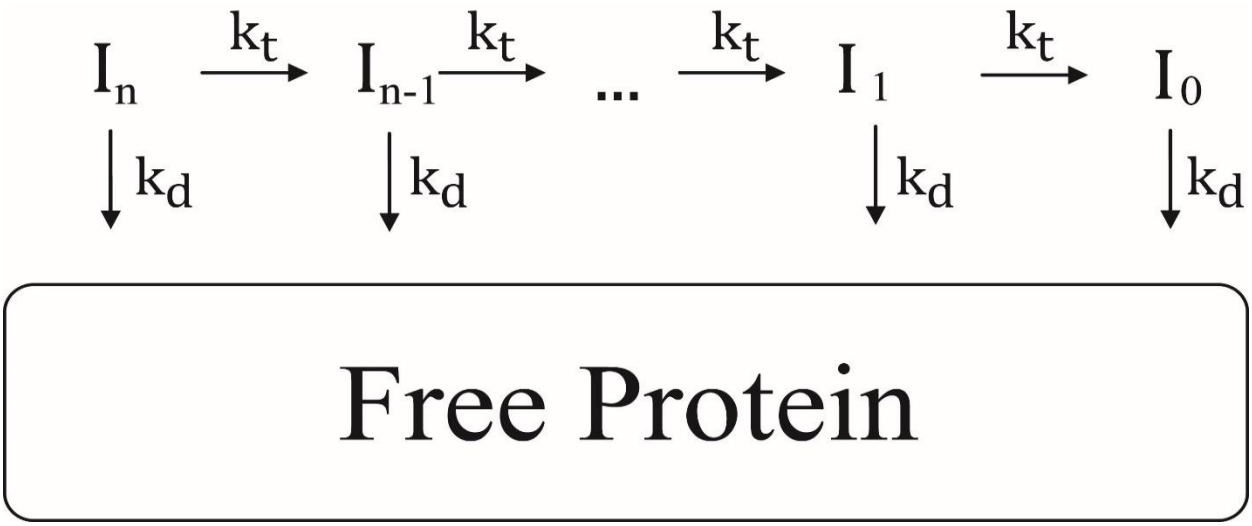
Figure 6 - Translocation Diagrams. Diagrams displaying what happens during rounds of translocation. A) In this pre-translocative state, the protein has been bound to the substrate and is tracking a single backbone in the 3' to 5' direction. There are two possibilities in this state: the protein will move forward with probability equal to the processivity P, or the protein will dissociate with probability 1-P. If the protein is able to bind and utilize and ATP molecule before it fluctuates off the substrate, the protein will then process forward along the backbone. B) The protein has captured a molecule of ATP and hydrolyzed it, releasing the ADP and phosphate group, and used the released free-energy to translocate *m* base pair forward.

the protein instead binds and consumes a given number of molecules of ATP, it will initiate a translocation step and travel towards the 5' end of the DNA backbone with step size m [bp] and rate k_t, see Figure 6B. Once the protein arrives at the new location, it then reverts to a pre-translocative state to begin the cycle anew. These rounds of translocation continue until either of two events occur: the protein dissociates and joins the population of unbound proteins or it reaches the final position at the end of the backbone. If the protein dissociates from the substrate, it then joins the population of free protein and then can bind to a different substrate in the future. However when it reaches the end of the DNA strand, the protein then interacts with a fluorophore commonly placed at this position to detect protein presence. This is described in more detail in later sections.

3.3 – Physical Models to Mathematical

3.3.1 – Simple translocation

While excellent and complete workups of turning physical models into systems of coupled differential equations exist[54–57], herein is provided a summarization for the sake of completeness. Starting with the original model discussed in the previous section, we will begin with a simple translocation model with only



Scheme 1 – Simple Translocation Model.

translocation and dissociation steps. It is also assumed that rebinding can be prevented after the protein dissociates through competitive binding with a preferred substrate. Experimentally, this constraint is commonly accomplished through the addition of heparin in solution. Heparin is a glycosaminoglycan with the highest negative charge density of any known biological molecule. While it is used heavily in medicine as an anti-coagulant, in biochemical experiments it is used as a competitor for DNA binding. Due to a highly negative charged surface, heparin binds rapidly and tightly to any interacting pockets on a protein and prevents rebinding. Given these constraints, we can assume a model such as shown in Scheme 1. As a protein binds to the substrate at a random position, we define this as position I_n where n is the number of positions away from the 5' end of the DNA that the protein has bound. Furthermore, n is defined as $n = \frac{L-d}{m}$ where L is the length of the DNA substrate, d is the binding site size, and m is the step size of the protein during processive translocation. Given these definitions, a particular substrate a fixed number of base pairs in length can appear to have a different quantity of binding positions depending on the unique binding characteristics of differing remodelers. Each of these positions can be modeled as a differential equation:

$$\frac{d[I_n]}{dt} = -k_t[I_n] - k_d[I_n] = -(k_t + k_d)[I_n]$$

$$\frac{d[I_{n<i<0}]}{dt} = k_t[I_{i+1}] - (k_t + k_d)[I_{n<i<0}]$$

$$\frac{d[I_0]}{dt} = k_t[I_1] - k_d[I_0]$$

Where a population at any given position is denoted as $[I_n]$. In this model, the population at any position between n and 0 ($n < i < 0$) is solely dependent on the current location population ($[I_{n<i<0}]$) and the previous ($[I_{i+1}]$). A protein that translocates from $[I_{i+1}]$ at rate k_t will increase the population at $[I_i]$ and any protein that dissociates or translocates from the i^{th} position will decrease the population. The net result of determining these calculations is a list of n coupled differential equations. While this system can be solved numerically through numerous methods, applying the Laplace transform to the system will result in an array

of algebraic equations that can be dealt with much more readily. The Laplace transformation, *F(s)*, of a time-dependent function, *f(t)*, is defined as

$$F(s) = \mathcal{L}[f(t)] = \int_0^\infty f(t)e^{-st}dt$$

where $\mathcal{L}$ denotes the Laplace transformation and s the Laplace variable. The same transformation of a differential function is

$$\mathcal{L}\left[\frac{df(t)}{dt}\right] = sF(s) - f(0)$$

where *f(0)* is the value of the function *f* at $t = 0$. Therefore, applying the Laplace transform to the original set of equations leaves us with the following:

$$\mathcal{L}\left[\frac{d[I_n]}{dt}\right] = sI_n'(s) - [I_n](0) = -(k_t + k_d)I_n'(s) \rightarrow$$

$$(s + k_t + k_d)I_n'(s) = [I_n](0) \rightarrow$$

$$I_n'(s) = \frac{[I_n](0)}{s + k_t + k_d}$$

$$\mathcal{L}\left[\frac{d[I_i]}{dt}\right] = (s + k_t + k_d)I_{(i)}'(s) - k_t I_{i+1}'(s) = [I_i](0)$$

$$\mathcal{L}\left[\frac{d[I_0]}{dt}\right] = (s + k_d)I_0'(s) - k_t I_1'(s) = [I_0](0)$$

Setting all initial non-n population values to zero ($[I_{n>i\geq 0}](0) = 0$) allows only the protein bound at the n^{th}position onto a single piece of DNA substrate to be considered. These can then all be written into an array as such

$$\begin{pmatrix} s+k_t+k_d & 0 & 0 & \dots & 0 \\ -k_t & s+k_t+k_d & 0 & \dots & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & -k_t & s+k_t+k_d & 0 \\ 0 & 0 & 0 & -k_t & s+k_d \end{pmatrix} \begin{pmatrix} I_n' \\ I_{n-1}' \\ \vdots \\ I_i' \\ \vdots \\ I_1' \\ I_0' \end{pmatrix} = \begin{pmatrix} I_n(0) \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \quad (1)$$

which can then be row reduced into solutions for each individual site. This gives us the solution to the Laplace transform of the population at every state through I_0'. The array and subsequent equations are listed below:

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} I_n' \\ I_{n-1}' \\ \vdots \\ I_i' \\ \vdots \\ I_1' \\ I_0' \end{pmatrix} = \begin{pmatrix} \frac{I_n(0)}{s+k_t+k_d} \\ \frac{k_t I_n(0)}{(s+k_t+k_d)^2} \\ \vdots \\ \frac{k_t^{i-1} I_n(0)}{(s+k_t+k_d)^i} \\ \vdots \\ \frac{k_t^{n-2} I_n(0)}{(s+k_t+k_d)^{n-1}} \\ \frac{k_t^{n-1} I_n(0)}{(s+k_d)(s+k_t+k_d)^{n-1}} \end{pmatrix}$$

This final form gives functions for the population of bound protein at every position along the substrate from position n to 0 in Laplace space.

$$I_n'(s) = \frac{I_n(0)}{s+k_t+k_d}$$

$$I_{1<i<n}'(s) = \frac{1}{s+k_t+k_d} \sum_{j=i}^{n} I_n(0) \left(\frac{k_t}{s+k_t+k_d} \right)^{j-i}$$

Experimentally, the fluorophore that the protein interacts with is at a single 5' site along the double-stranded DNA backbone and the resultant signal is assumed to be only from the final position. However, in an ensemble setting the population of bound protein is initially distributed across any possible position along

the DNA. The proper solution to the population at $I_0(t)$ in these experimental conditions is then the summation of all $I_0(t)_n$ from 0 to L where L is the length of the DNA strand in integer step sizes. To cover all population possibilities,

$$I_0(t) = \mathcal{L}^{-1}[I_0'(s)_{L=1} + I_0'(s)_{L=2} + I_0'(s)_{L=3} + \cdots]$$

$$I_0(t) = \mathcal{L}^{-1}\left[\frac{[I_0(0)]}{s+k_d} + \frac{[I_1(0)]}{s+k_d}\frac{k_t}{s+k_t+k_d} + \frac{[I_2(0)]}{s+k_d}\left(\frac{k_t}{s+k_d+k_t}\right)^2 + \cdots\right]$$

The assumption that the protein is bound in equal concentrations to all binding sites is based on the idea that there exists no preference for binding to any specific site over another; the probability of binding to site I_1 is the same as site I_i. Given the extreme number of experimental interactions, the initially bound concentrations of these proteins at each site is equal, *i.e.* $[I_0(0)] = [I_i(0)] = [I_n(0)]$ for all $0 < i < n$. Consequently the total population of bound protein distributed across all sites from 0 to n can then be expressed as $[I(0)] = ([I_0(0)] + [I_1(0)] + [I_2(0)] + \cdots)$ or $[I(0)] = \sum_{i=0}^{n}[I_i(0)]$ or $[I(0)] = [I_n(0)](1+n)$. Our equation then becomes

$$I_0(t) = \mathcal{L}^{-1}\left[\frac{I(0)}{(1+n)(s+k_d)}\left[1 + \frac{k_t}{s+k_t+k_d} + \left(\frac{k_t}{s+k_d+k_t}\right)^2 + \cdots\right]\right]$$

$$I_0(t) = \mathcal{L}^{-1}\left[\frac{I(0)}{(1+n)(s+k_d)}\left[1 + \sum_{i=1}^{n}\left(\frac{k_t}{s+k_d+k_t}\right)^i\right]\right]$$

As the solution to a summation of a power series is

$$\sum_{i=1}^{n} x^i = \frac{x - x^{n+1}}{1-x}$$

That leaves

$$\sum_{i=1}^{n}\left(\frac{k_t}{s+k_d+k_t}\right)^i = \frac{k_t}{s+k_d}\left(1-\left(\frac{k_t}{s+k_t+k_d}\right)^n\right)$$

As a simplified expression, giving us

$$I_0(t) = \mathcal{L}^{-1}\left[\frac{I(0)}{(1+n)(s+k_d)}\left[1+\frac{k_t}{s+k_d}\left(1-\left(\frac{k_t}{s+k_t+k_d}\right)^n\right)\right]\right] \quad (2)$$

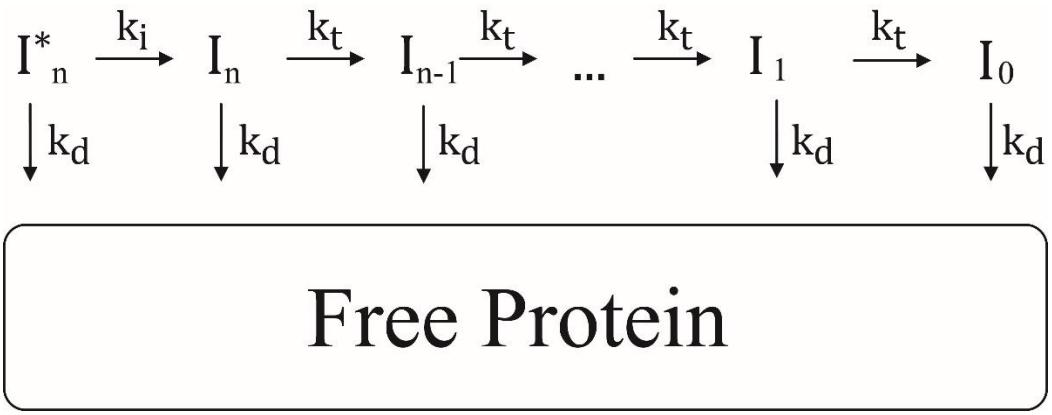
for our final solution of $I_0(t)$ given a simple translocation model with no rebinding.

3.3.2 – Perturbations

The above model assumes little about the underlying mechanics of translocation and is experimentally valid only in a particular case: when the rebinding of protein is blocked, or severely repressed, with the addition of a binding competitor molecule such as heparin. From here, to consider more complex models, we must append additional mechanical steps. This type of modification is not predictive; a good fit to collected data with a base model does not imply additional constraints. Instead, model production must be responsive to different shape traits of the final collected response curve. While there are any number of modifications to these schemes that one could imagine, this section will list several of the major variations that were useful in the course of this thesis.

3.3.2.1 – Initiative step

A common adaptation of simple translocation is the inclusion of an initiative step at the beginning of processive translocation, see Scheme 2. This type of step can have any number of physical correlations including a conformational change in the physical structure of the protein, a transition between differing binding modes dependent on the presence of additional ATP molecules, or any number of other possibilities. When considering any type of modification to a model, it is important to remember that it cannot be used to distinguish between mechanistic possibilities; only that the addition of this step in



Scheme 2 – Initial Pausing Step.

subsequent analyses allows for better fits to varied structures within the kinetic time courses and a better description of what is likely happening at the molecular level. In this instance, there is one supplemental step and the only modification to the original translocation arrays is the addition of the following:

$$\frac{d[I_n^*]}{dt} = -(k_i + k_d)[I_n^*]$$

This changes the arrays as such

$$\begin{pmatrix} s+k_i+k_d & 0 & 0 & \dots & 0 \\ -k_i & s+k_t+k_d & 0 & \dots & 0 \\ 0 & -k_t & s+k_t+k_d & 0 & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & -k_t & s+k_t+k_d & 0 \\ 0 & 0 & 0 & -k_t & s+k_d \end{pmatrix} \begin{pmatrix} I_n^{*\prime} \\ I_n' \\ I_{n-1}' \\ \vdots \\ I_i' \\ \vdots \\ I_1' \\ I_0' \end{pmatrix} = \begin{pmatrix} I_n^*(0) \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

Which, given the same assumptions as the simple translocation model, reduces to

$$I_0(t) = \mathcal{L}^{-1}\left[\frac{k_i I(0)}{(1+n)(s+k_d)(s+k_i+k_d)}\left(1+\sum_{i=1}^{n}\left(\frac{k_t}{s+k_d+k_t}\right)^i\right)\right]$$

Or

$$I_0(t) = \mathcal{L}^{-1}\left[\frac{k_i I(0)}{(1+n)(s+k_d)(s+k_i+k_d)}\left(1+\frac{k_t}{s+k_d}\left(1-\left(\frac{k_t}{s+k_t+k_d}\right)^n\right)\right)\right] \quad (3)$$

which only differs from Equation 2 in the first term with the additional factor of $\frac{k_i}{s+k_i+k_d}$.

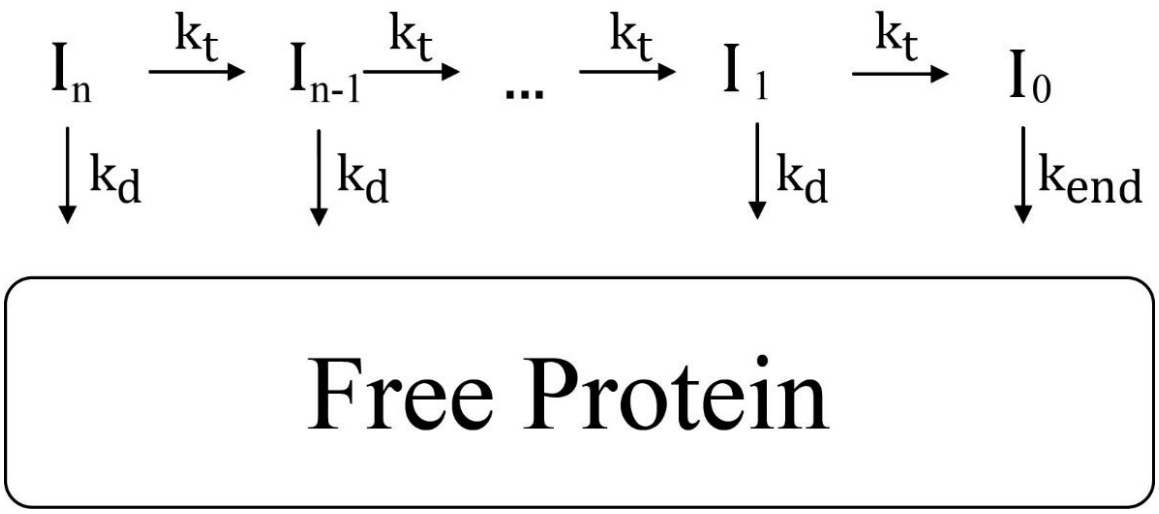
3.3.2.2 – Differential dissociation from site I_0

Similar to an initiative step model, the model can be adapted to include a step that has a differential probability of dissociation from the end of the DNA strand, as shown in Scheme 3. As an example of when this category of step may occur, consider that certain proteins have a means of detecting extraneous DNA that is not bound within the ATPase pocket[58]. If the protein reaches the end of the DNA and does not detect further substrate, it could signal that the time for departure has arrived. Whereas, if there does not exist any way for the protein to sense further room for movement, it could remain at the 5' end of the DNA for additional, futile cycles of ATP utilization, increasing the time spent at the I_0 site and value for the associated rate constant, k_{end}. The sole modification to Equation 1 is only in the bottom right term, when the protein is at position I_0 and is shown below.

$$\begin{pmatrix} s+k_t+k_d & 0 & 0 & \dots & 0 \\ -k_t & s+k_t+k_d & 0 & \dots & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & -k_t & s+k_t+k_d & 0 \\ 0 & 0 & 0 & -k_t & s+k_{end} \end{pmatrix} \begin{pmatrix} I_n' \\ I_{n-1}' \\ \vdots \\ I_i' \\ \vdots \\ I_1' \\ I_0' \end{pmatrix} = \begin{pmatrix} I_n(0) \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

This amends Equation 2 to appear as

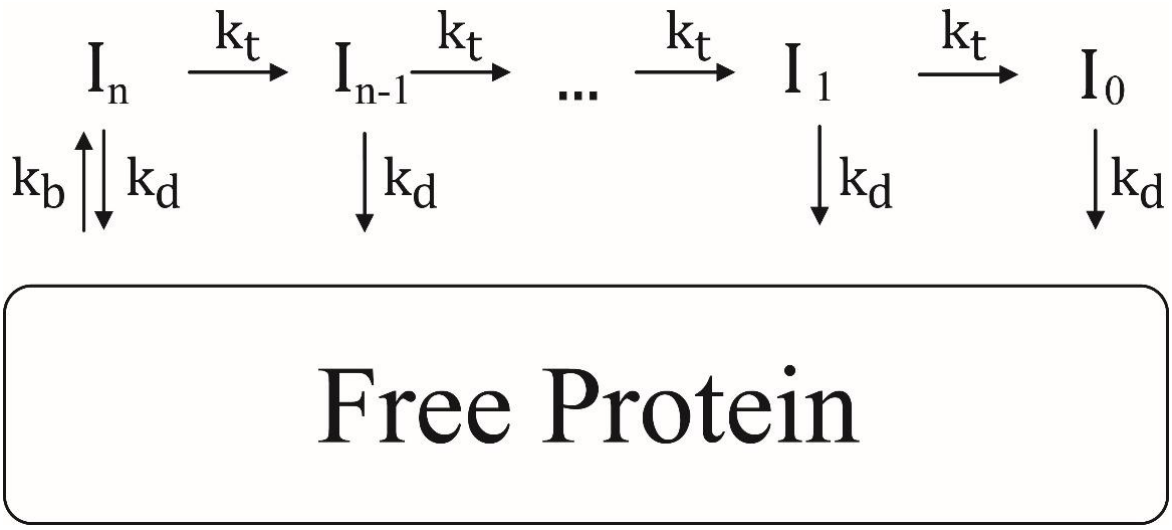
$$I_0(t) = \mathcal{L}^{-1}\left[\frac{I(0)}{(1+n)(s+k_{end})}\left[1+\frac{k_t}{s+k_d}\left(1-\left(\frac{k_t}{s+k_t+k_d}\right)^n\right)\right]\right]$$



Scheme 3 - Differential binding from site I0

3.3.2.3 – Rebinding

During the course of experimentation with the molecular motor of the RSC protein (Chapter 5), it was determined that RSCt was not easily sequestered after translocation without significantly affecting the experimental outcome. This led to the development of translocation equations that were not dependent on the segregation of unbound proteins. Meaning that after repeated rounds of ATP hydrolysis and translocation along the DNA substrate, followed by final dissociation, the protein is then allowed to rebind to a different substrate molecule and continue processive translocation. This approach is shown in Scheme 4.



Scheme 4 – Rebinding of proteins.

The system of equations is largely the same, with the exception of the initial binding position. The n^{th} position will have an additional term where the protein binds. This reaction is limited by the binding rate constant, k_b, multiplied by the concentration of free DNA, $[D]$, and the concentration of free protein, $[FP]$. However, because the experimental conditions are such that the DNA substrate concentration is in large excess over the protein concentration, we are able to consider this reaction as pseudo-first order and consider the concentration $[D]$ a constant (See Appendix A). This means that the $k_b$rate constant is actually pseudo-first order approximation of $k_b = k_b'[D]$. As the Laplace transform of a constant is $\mathcal{L}[C] = \frac{C}{s}$, the modification is shown below:

$$\frac{d[I_n]}{dt} = -k_t[I_n] - k_d[I_n] + k_b * [FP] = -(k_t + k_d)[I_n] + k_b * [FP]$$

$$\mathcal{L}\left[\frac{d[I_n]}{dt}\right] = sI_n'(s) - [I_n](0) = -(k_t + k_d)I_n'(s) + k_bFP'(s) \rightarrow$$

$$(s + k_t + k_d)I_n'(s) = [I_n](0) + k_bFP'(s) \rightarrow$$

$$I_n'(s) = \frac{[I_n](0) + k_bFP'(s)}{(s + k_t + k_d)}$$

As a protein becomes unbound from a DNA substrate, it joins the larger population of free protein. From this state, it is then capable of rebinding another substrate with rate constant k_b. Because of this, the concentration of free protein is time dependent and requires an additional equation.

$$\frac{d[FP]}{dt} = k_d([I_n] + [I_{n-1}] + \cdots + [I_0]) - k_b[FP]$$

$$\mathcal{L}\left[\frac{d[FP]}{dt}\right] = sFP'(s) - [FP](0) = k_d\left(I_n'(s) + I_{n-1}'(s) + \cdots + I_0'(s)\right) - k_b FP'(s)$$

$$FP'(s) = \frac{[FP](0) + k_d\left(I_n'(s) + I_{n-1}'(s) + \cdots + I_0'(s)\right)}{s + k_b}$$

making our matrices

$$\begin{pmatrix} s+k_t+k_d & 0 & 0 & \dots & 0 & -k_b \\ -k_t & s+k_t+k_d & 0 & \dots & 0 & 0 \\ & \vdots & & \ddots & \vdots & 0 \\ 0 & 0 & -k_t & s+k_t+k_d & 0 & 0 \\ 0 & 0 & 0 & -k_t & s+k_d & 0 \\ -k_d & -k_d & -k_d & -k_d & -k_d & s+k_b \end{pmatrix} \begin{pmatrix} I_n' \\ I_{n-1}' \\ \vdots \\ I_i' \\ \vdots \\ I_1' \\ I_0' \\ FP' \end{pmatrix} = \begin{pmatrix} I_n(0) \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

whose solution is

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} I_n' \\ I_{n-1}' \\ \vdots \\ I_i' \\ \vdots \\ I_1' \\ I_0' \\ FP' \end{pmatrix} = \begin{pmatrix} \frac{(s+k_d)(s+k_b)I_n(0)}{s(s+k_b+k_d)(s+k_t+k_d)} \\ \frac{k_t(s+k_d)(s+k_b)I_n(0)}{s(s+k_b+k_d)(s+k_t+k_d)^2} \\ \vdots \\ \frac{k_t^{i-1}(s+k_d)(s+k_b)I_n(0)}{s(s+k_b+k_d)(s+k_t+k_d)^i} \\ \vdots \\ \frac{k_t^{n-1}(s+k_d)(s+k_b)I_n(0)}{s(s+k_b+k_d)(s+k_t+k_d)^n} \\ \frac{(s+k_b)I_n(0)k_t^n}{s(s+k_b+k_d)(s+k_t+k_d)^n} \\ \frac{k_d + FP(s+k_d)}{s(s+k_b+k_d)} \end{pmatrix}$$

For the total population at I_0, the summation over all n being

$$I_0(t) = \mathcal{L}^{-1}\left[\frac{I(0)(s+k_b)}{(1+n)(s+k_b+k_d)}\left[1+\frac{k_t}{s+k_t+k_d}+\left(\frac{k_t}{s+k_d+k_t}\right)^2+\cdots\right]\right]$$

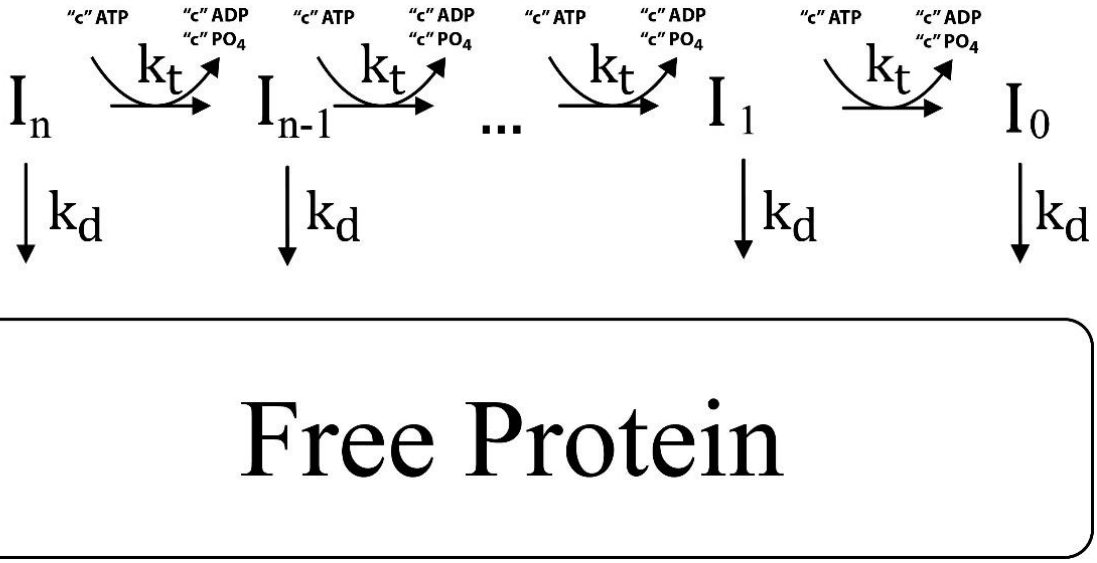
Which equals

$$I_0(t) = \mathcal{L}^{-1}\left[\frac{I(0)(s+k_b)}{(1+n)(s+k_b+k_d)}\left[1+\frac{k_t}{s+k_d}\left(1-\left(\frac{k_t}{s+k_t+k_d}\right)^n\right)\right]\right]$$

3.3.3 – ADP production

It is often significant to quantify the ADP production versus time in a particular set of experiments. This inclusion of this information can frequently provide additional constraints on parameter estimates. As the protein undergoes processive rounds of translocation, it must also undergo the process of hydrolysis: to consumption of ATP molecules and subsequent production of the same number of ADP and phosphate fragments, see Scheme 5. Generally, only a single ATP is consumed at a time but this does not necessarily mean that a single hydrolysis event will be followed by processive translocation. When a given motor uses more than one ATP molecule per translocation step, it is referred to as 'futile hydrolysis' and has been observed in many systems[59,60]. This futile hydrolysis can inflate the 'coupling efficiency' of a protein which is defined as the number of ADP molecules produced, C, per base pair translocated. In this simple-translocation ADP production model, the population of free ADP in solution will then change in time as

$$\frac{d[ADP(t)]}{dt} = Ck_t\sum_{i=1}^{n}[I_i(t)]$$



Scheme 5 – ADP production cycles in translocation modeling

which is simple and generalizable to varying models. The main difference between this and what has been shown previously is that the production of ADP happens at every site whereas the previous models have largely focused only on the population at I_0. This leads to the summation for ADP production being over all $[I_i(t)]$.

$$[ADP(t)] = \mathcal{L}^{-1}\left\{\frac{1}{s}\left[\frac{Ck_tI(0)}{(1+n)(s+k_d)}\left(n - \frac{k_t}{s+k_d}\left(1 - \left(\frac{k_t}{s+k_t+k_d}\right)^n\right)\right)\right]\right\}$$

Similar to the derived translocation equations in previous sections, there exist any number of perturbations to this representation. By devising new schemes that include novel steps in the mechanistic sequence, the resultant arrays of equations lead to distinct solutions. For instance, if you were to find evidence of futile hydrolysis when the protein is at the I_0 position, that can easily be modeled by the addition of the end term with a different ADP production rate constant, k_a.

$$\frac{d[ADP(t)]}{dt} = Ck_t\sum_{i=1}^{n}[I_i(t)] + k_a[I_0(t)]$$

Which would equal

$$[ADP(t)] = \mathcal{L}^{-1}\left\{\frac{I(0)}{s(1+n)(s+k_d)}\left(Ck_t\left(n + \frac{k_t}{s+k_d}\left(\left(\frac{k_t}{s+k_t+k_d}\right)^n - 1\right)\right) + k_a\left(1 + \frac{k_t}{s+k_d}\left(1 - \left(\frac{k_t}{s+k_t+k_d}\right)^n\right)\right)\right)\right\}$$

Chapter 4 – Materials and Methods

4.1 - Buffers

All buffers were prepared with reagent grade chemicals using twice-distilled water that was deionized with a Milli-Q purification system (Millipore Corp., Bedford, MA). The storage buffer used for both RSC and ISWI is 25 mM Tris HCl (pH 7.5@25°C), 150 mM KOAC, 10% (v/v) glycerol, and 0.5 mM 2-mercaptoethanol. All experiments were conducted in the reaction buffer consisting of 10 mM HEPES (pH 7.0@25°C), 5 mM MgCl2, 100 mM KOAC, 4% (v/v) Glycerol, 2.5 mM DTT, and 0.1 mg/mL BSA.

4.2 - Oligonucleotide substrates

All labeled and unlabeled oligonucleotides were purchased from IDT Technologies (Coralville, IA) and were HPLC purified. Each oligonucleotide was extensively dialyzed against Milli-Q water. Double-stranded DNA substrates were annealed by mixing together equal concentrations of complimentary strands in a DNA annealing buffer consisting of 10mM HEPES (pH 7.0@25°C) and 40 mM KAC. The mixture is then heated to 95°C and allowed to cool to room temperature over the course of 12 hours. For fluorophore-labeled DNA, a 5% overabundance of unlabeled DNA was included to ensure complete annealing of all fluorescein-labeled DNA strands. For all fluorophore-labeled substrates, the fluorophore was attached only

at the 5’ end of a single strand of the duplex, leaving the opposing single-stranded mated substrate unlabeled. This allows sampling of only the population of protein translocating along one backbone.

4.3 - RSCt Purification

RSCt was over-expressed and purified from an *Escherichia coli* bacterial over-expression system. The CDF Duet-1 vector (Novagen) bearing a Sth1 construct (301-1097aa) with a 10X histidine tag at the N-terminus was transformed into BL21(DE3) codon plus strain along with the RSF Duet vector (Novagen) containing Arp9 and Arp7 constructs. These were selected on streptomycin and kanamycin plates. The cells were grown in nutrient-rich auto-inducible media at 37°C for 4 hours, 30°C for 12 hours and at 22°C for 24 hours, harvested by centrifuging at 6000g at 4°C, resuspended in lysis buffer (50 mM phosphate buffer pH 7.5, 300 mM NaCl, 10% glycerol, 0.5 mM β-mercaptoethanol and 1X protease inhibitors), sonicated at 30 % duty cycle for 30 seconds 10-15 times. The cells were kept on ice at all times. The lysate was then centrifuged at 20,000g for 30 minutes at 4°C. The supernatant was run on a pre-packed Ni-NTA column pre-equilibrated with 20 mM Tris pH 7.5, 100 mM NaCl, 10% glycerol, 0.5 mM β-mercaptoethanol and 30 mM imidazole. The protein was eluted in a gradient using buffer containing 20 mM Tris pH 7.5, 100 mM NaCl, 10% glycerol, 0.5 mM β-mercaptoethanol and 500 mM imidazole. The protein started eluting at 50% of the gradient. The purified protein was run on 12% polyacrylamide gel containing SDS. The purified fractions were pooled, concentrated and then run through a gel filtration column equilibrated with a sizing buffer (20 mM Tris pH 7.5, 200 mM NaCl, 10 % glycerol, 0.5 mM β-mercaptoethanol and 1X PMSF). The fractions that eluted corresponding to RSCt were pooled and concentrated further and then flash frozen in liquid nitrogen to be stored at -80°C. Protein concentration was determined using a Bradford assay comparison to a BSA standard at 595 nm.

4.4 - Recombinant ISWI expression and purification

cDNA coding for *Xenopus Laevis* ISWI was amplified using PCR with primers containing BglII restriction site at the 5' end and EagI restriction site at the 3' end into pCR4-TOPO vector (Invitrogen). The insert was then further subcloned into BamHI-NotI digested pPIC3.5-CBPXpress-zz yeast expression vector. All recombinant constructs were confirmed by sequencing. The ISWI containing recombinant construct was then transformed into GS115 strain of Pichia pastoris yeast through electroporation followed by recombinant ISWI expression according to manufacturer protocol (Invitrogen). Briefly, cells were grown in buffered glycerol complex media (Invitrogen) until O.D.600= 10, the cells were then resuspended in buffered methanol media and allowed to shake for 6 hours at 30 ºC to induce protein expression. Cells were harvested by centrifugation at 3000x g for 5 minutes at 4 ºC. Cell paste was loaded into syringe, dispensed into liquid nitrogen and stored at -80 ºC until ready for use. Mixture of dry ice and frozen yeast cells were mechanically lysed followed by addition of lysis buffer (50 mM Tris (pH 8.0), 300 mM NaCl, 0.1 % Triton® X-100, 2 mM CaCl2, 2 mM MgCl2, 10% glycerol, 10 mM PMSF, 1.3 mM β-Mercaptoethanol) and centrifugation at 15,000x g for 30 minutes at 4 ºC. Supernatant containing 1 mM CaCl2 was then incubated with Calmodulin Sepharose 4b resin (GE Healthcare) for 4 hours. CBP tagged ISWI was then eluted using 10 mM EGTA-containing buffer. Collected elutions were examined using a 8% SDS-PAGE analysis and ISWI-containing fractions were further purified using heparin column (GE Healthcare) followed by buffer exchange (20 mM HEPES (pH 7.8), 100 mM NaCl, 10 mM MgCl2, 5% glycerol, 0.5 mM DTT) and stored at -80 ºC. Protein concentration was determined through measurements of the A280 and extinction coefficient and further confirmed using a Bradford assay.

4.5 - ATPase assay

ATPase activity was measured at 25°C (ISWI: 30°C) by monitoring the hydrolysis of [α-32P] ATP. The reaction was carried out in a reaction buffer containing 10mM HEPES pH 7.0, 1mM ATP, 5mM

MgCl2, 20mM potassium acetate, 4% glycerol, 0.5mM DTT and 0.1 mg/ ml BSA. Remodeler at 50 nM concentration was pre-incubated with a series of concentrations of double-stranded DNA, ranging from 10 to 1000 nM, and the reaction was initiated by the addition of ATP. Aliquots of the reaction samples were removed at fixed time points and quenched with an equal volume of 0.5M EDTA to stop the reaction. The reaction products were separated by TLC on PEI Cellulose F sheets in 0.6M potassium phosphate buffer, pH 3.4 and quantized with a PhosphorImager (GE healthcare). The ATPase time course was analyzed using ImageJ to determine the amount of ATP hydrolyzed.

4.6 - RSCt Streptavidin Displacement Assay

100 picomole of each DNA length was radio-labeled using 5 units polynucleotide kinase from phage T4 infected *E. coli* (T4 PK), and 50 uCi ^{32}P γ-labeled ATP. DNA labeling buffer is 40 mM Tris HCl (pH 7.5 @ 25°C), 10 mM MgCl2, and 0.5 mM DTT. The solution was incubated @ 37°C for 1 hour before being spun on a GE Illustra ProbeQuant G-50 micro column to remove any remaining γ-labeled ATP. Prior to initializing the reaction, 20 nM of biotinylated DNA lengths were incubated with 300 nM streptavidin for 30 minutes to ensure total saturation of all DNA-biotin sites. After the addition of 20 nM RSCt, 3 µM biotin was then included to act as a trap for any free streptavidin or streptavidin that will be displaced by the RSC and was incubated with DNA for 40 minutes at 30°C. Reactions were started upon addition of 5 mM ATP and carried out in 100 µL solution volume with 10 µL aliquots being removed every 5 minutes for the initial 30 minutes and then at 10 minute intervals until 1 hour had been reached. Reaction aliquots were quenched in equal volumes of quench buffer consisting of 0.6% (w/v) SDS, 200 mM EDTA, 0.08% (w/v) xylene cyanol, 0.08% (v/v) bromophenol blue, 10% (v/v) glycerol. To ensure a constant concentration of ATP, a regeneration system was integrated by including 15 units/mL rabbit muscle pyruvate kinase (PK), and 18 units/mL lactate dehydrogenase (LDH). 2 lug fractions of each time point were analyzed using 6% TBE Invitrogen Novex gels and electrophoresis was performed using an Invitrogen X-Cell SureLock

minicell system with Novex TBE Running Buffer. Voltage was applied with an Invitrogen PowerSure500 power supply and run at constant voltage. Gels were imaged on a GE HealthCare phosphorimager and resulting images were quantitatively analyzed with the ImageJ software package. Streptavidin was stored in a storage buffer consisting of 10 mM HEPES (pH 7.0 @ 25°C), 20% (v/v) glycerol, and 5 mM NaCl.

4.7 - Stopped-flow translocation assay

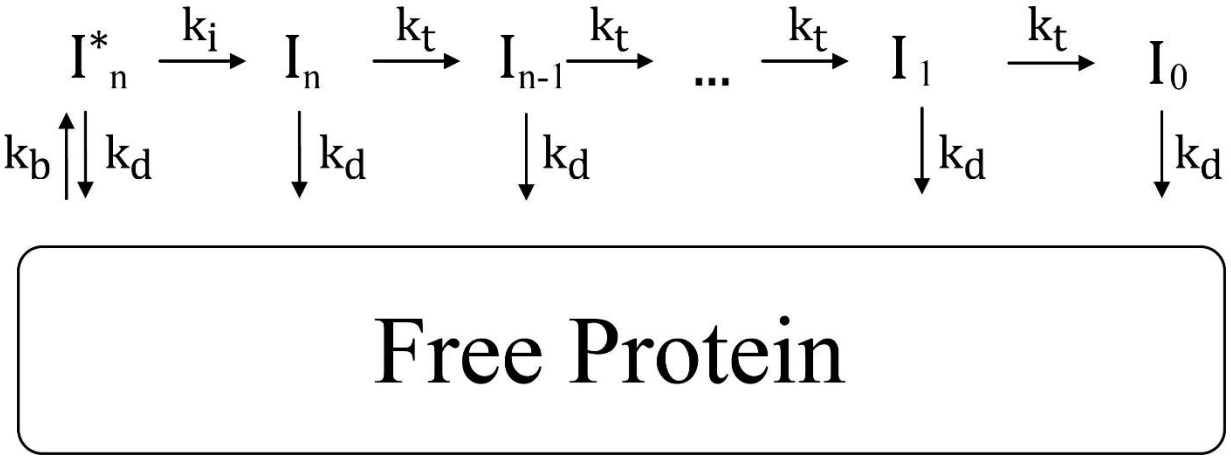
All RSCt fluorescence stopped-flow experiments were carried out in the same reaction buffer as ATPase assays, at 25°C (ISWI: 30°C) using an Applied Photophysics SX.18MV-R Stopped-Flow spectrophotometer (Applied Photophysics Ltd., Leatherhead, UK). Optical filters were purchased from ThorLabs (Newton, NJ). Fluorescein-labeled single-stranded DNA was purchased from IDT technologies (Coralville, IA) along with complementary unlabeled single-stranded DNA. The individual strands were then annealed together as described above. The fluorescein label was excited at 495 nM, and its fluorescence emission was monitored at wavelengths >515 nM using a long-pass filter. 80 nM ISWI was incubated with fluorescein-labeled DNA of varying lengths for 5 minutes before loading onto one of the syringes of the stopped-flow instrument. These experiments were performed under conditions where the total concentration of the DNA was in excess of the total concentration of RSCt to ensure that the predominant bound species were singly- bound DNA-RSCt complexes. A solution composed of the same reaction buffer listed above and 20mM ATP was loaded on the other syringe of the stopped-flow instrument. Reaction mixtures in both syringes were incubated for 2 minutes at 25°C to allow for temperature equilibration prior to mixing. Following the 1:1 mixing of the two solutions, the final reaction concentrations were as follows: 100nM RSCt (40 nM ISWI), 1.5 µM DNA base pair concentration (ISWI: 12µM), and 10mM ATP. The kinetic traces used for analysis are an average of a minimum of 15 individual traces.

4.8 - RSCt Data Analysis

All analysis performed was NLLS fits using Conlin[61]. All fitting models were written in C and compiled using Microsoft Visual Studio .NET 2003 compiler on a Windows XP workstation. The software library CNL 6.0 (Visual Numerics Incorporated, Houston, TX) was used for the numerical calculation of the inverse Laplace transform. The uncertainties in all fitted parameters reported in this manuscript are the standard deviation of the mean.

During its translocation along double-stranded DNA RSC tracks along a single strand of the duplex with a 3' to 5' directional bias[35,62]. Because of this we can model double-stranded DNA translocation by RSCt as a 3' to 5' directionally biased single-stranded DNA translocation process[54,56]. Initially, we analyzed our translocation time courses using a single-step, multiple-turnover model[55,63,64], but were unable to find a satisfactory description of the data as determined by the variance of fit (data not shown). Since previous work has hinted that possible secondary steps are required for competent RSC translocation[65,66], we derived and implemented another model that included these ancillary steps. This model, shown in Scheme 6, upon binding at a random location, I_n^*, involves an initiation step prior to processive translocation[54,56]. This step could be a conformational change in the protein, a transition between different binding modes[67], *etc*. This model makes no distinction as to what the physical mechanism could be; only that it is required prior to active movement by RSC. The equation corresponding to Scheme 6 that we used for analysis is given as

$$[I_0(t)] = \mathcal{L}^{-1}\left\{\frac{k_i I(0)(s + Dk_b)(s + k_t + k_d)}{s(n+1)(s + k_d + Dk_b)(s + k_i + k_d)(s + k_d)}\left(1 - \left(\frac{k_t}{s + k_t + k_d}\right)^n\right)\right\} + \frac{C}{s} \qquad (4)$$



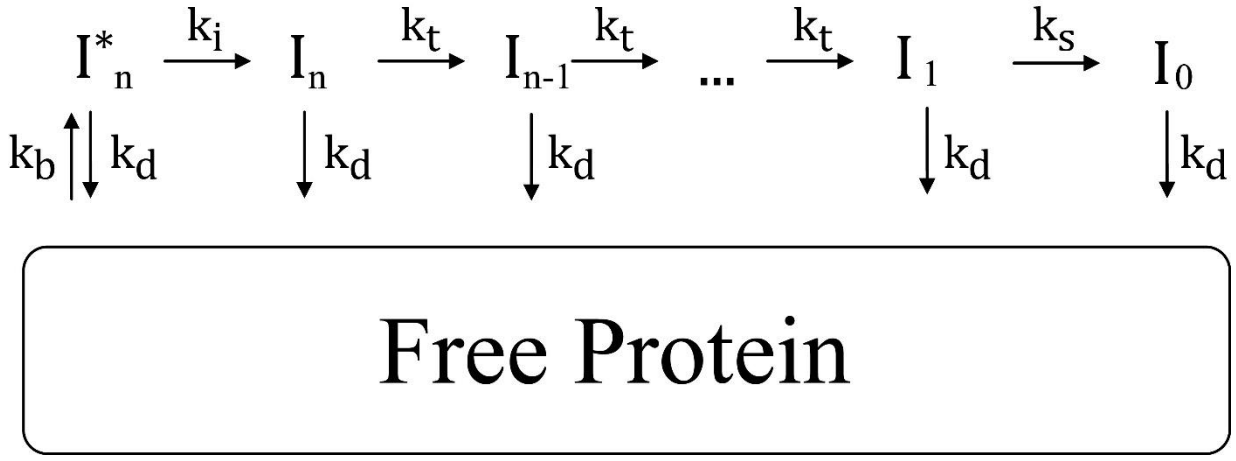
Scheme 6 - Paused Step Model. Before initiating translocation, the protein binds with rate constant k_b to a random location along the DNA substrate and begins the process in state I_n^*. In this state the protein can either dissociate with rate constant k_d or go through some form of conformational or energetic change that precedes translocation and is governed by rate constant k_i. Upon completing any slow-step process that could be concealed in this initiative activity, the protein is then in state I_n and prepared for processive translocation. Any further translocation is governed by the rate constant k_t and at each available state I_i can either proceed at a rate k_t or dissociate with a rate k_d.

In this equation $\mathcal{L}^{-1}$ denotes the inverse Laplace operator. The microscopic rate constants k_t, k_d, and k_i are the same as previously defined. *C* is a scalar offset term that is included for any arbitrary offsets in system signal that is not taken out of the data upon normalization. Because C is not a global variable, each data set is able to have its own arbitrary scaling factors independent of the other lengths. The variable *n* is the number of possible steps the protein can take on a given nucleotide as defined in Chapter 3.

The model we used for analysis of the time courses collected using the streptavidin displacement assay included the initiation step from Scheme 6 as well as a final step, occurring immediately prior to reaching the end of the DNA, which is associated with the displacement of the streptavidin molecule from its biotin bond; this step is associated with the microscopic rate constant k_s. The equation corresponding to this model is too cumbersome to reproduce here, but its steady-state limit is shown here:

$$f(t) = A * \frac{k_i}{(k_d + k_i)} \frac{k_s^2}{(k_d + k_s)} \frac{k_b}{(k_d + k_b)} \frac{1}{1 + r(n-1)} \left(1 + \frac{r k_t}{k_d}\left(1 - \left(\frac{k_t}{k_d + k_t}\right)\right)\right) t \qquad (5)$$

where the scalar variable *A* is proportional to the population of initially bound protein.



Scheme 7 - Paused-Step Streptavidin Displacement Model. A scheme built upon Scheme 6 considering an additional kinetic step where upon reaching the biotinylated end of the DNA strand, the protein encounters the streptavidin molecule and must overcome the bond strength. This results in a different microscopic rate constant at that end of the DNA, labeled k_s, and therefore includes a different controlling variable in the time-course data.

The equation utilized to analyze the data collected from the ATPase assay was derived in Khaki *et al.*[68] and is given as

$$k_{cat} = \left[ck_t r \frac{\left(n(1-P) - P\left(1-P^n\right)\right)}{(1+nr)(1-P)} + c_i k_d \right] \left(\frac{k_i}{k_d + k_i} \right) \quad (6)$$

where k_{cat} is the maximum ATPase rate under saturating conditions of nucleic acid. The constants c and c_i are the ATP coupling stoichiometries; c is defined as the number of ATP consumed per translocative step while c_i refers to the ATP consumed per initiation step. P is the processivity defined in Chapter 3 and is a unitless measure of a protein's translocation ability[68–71] The processivity can also be used to determine the average number of base pairs translocated by RSCt before dissociation from the DNA

$$m * \frac{P}{1-P}$$

This calculation is equivalent to the processivity as defined in Sirinakis *et al.*[66]

4.9 - ISWI Data analysis

Linear fits for all ATP analyses were performed in Mathematica 5.2 using the Regress[] function from the LinearRegression package. All analyses of translocation data were identical to the analysis of RSCt data.

ISWI has been shown to have 3'-5' directionally biased ATP-stimulated translocation[72] through triplex oligonucleotide displacement assays[73,74]. Because of these results, we can model the double-stranded ISWI movement as single-stranded DNA translocation along a single backbone[56,65]. We analyzed our results using a single-step, multiple-turnover model[55,63,64] shown in Scheme 4. This is the simplest model of processive translocation that includes rebinding and does not account for any secondary steps that may occur during successive rounds of translocation. Given that DNA is in large excess of ISWI, the time required to find and rebind a new substrate is very, very small; as soon as an ISWI molecule dissociates from the DNA backbone it almost immediately rebinds to another. From this assumption, the rate constant related to rebinding, k_b, can be assume to be very large with regard to the other rate constants. From this, we can take the limit as k_b goes to ∞. Using the methods described in Chapter 3 and taking the $\lim_{k_b \to \infty} f_{5'}(t)$, the equation used to describe the population of protein positioned at the 5' end of the DNA with respect to time is

$$f_{5'}(t) = \frac{A}{1+n}\mathcal{L}^{-1}\left[\frac{(s+k_t+k_d)}{s(s+k_d)}\left(1-\left(\frac{k_t}{s+k_t+k_d}\right)^n\right)\right] \quad (7)$$

where $\mathcal{L}^{-1}$ is the symbol for the inverse Laplace transform. *A* is a scalar constant that incorporates both the initial concentration of the protein bound to the single-stranded oligonucleotide and the fluorescence signal change associated with the protein bound at the fluorophore-labeled 5' end of the DNA. The microscopic rate constants k_t and k_d are associated with the translocation and dissociation, respectively. The variable *n* is defined as the number of possible steps the protein can take on a given nucleotide and is given by:

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

Here, the variable *L* is the total DNA length in base pairs; *m*, the kinetic step size in base pairs; and *d*, the binding site size of ISWI in base pairs. Future references to Scheme 8 will be referencing Scheme 8B solely.

For simulation analysis of ATP curves, the equation for simple ATP production used was

$$ADP(t) = \mathcal{L}^{-1}\left\{\frac{ck_t}{ms^2(n+1)}\left(n - \frac{k_t}{k_d + s}\left(1 - \left(\frac{k_t}{k_d + k_t + s}\right)^n\right)\right)\right\}$$

All terms are defined above with the exceptions of m and c. The former, referred to as the kinetic step size, is a measure of the distance that a protein moves along the DNA between rate-limiting steps during translocation. The latter is the number of ATP molecules converted to ADP between rate-limiting steps during translocation. The ratio of the two, c/m , is a measure of the overall efficiency with which the protein processes down the substrate backbone. High values of c/m correspond to low efficiencies requiring large numbers of ATP to travel short distances. Conversely, low values of c/m relate to high efficiency; few ATP molecules for large distances covered.

We also report several calculated parameters to describe the motion and efficiency of DNA translocation. Processivity, *P*, is a unitless value ranging between $0 < P < 1$ relating the probability of a protein to move forward along the DNA to the probability of dissociating from the DNA. Processivity is calculated by $P = \frac{k_t}{k_t + k_d}$ [55,71]. It follows that $\frac{P}{1-P}$ is the average number of forward steps a protein will complete before dissociating. The parameter $m * k_t$is the macroscpoic translocation rate and $\frac{m*P}{1-P}$ is the average number of base pair that the protein will translocate before dissociating from the DNA.

4.10 – Simulations

While the bulk of this dissertation is experimental, there were points when the experiments were unable to give tight constraints on kinetic parameters included in the translocation models. In these instances, it was necessary to perform additional theoretical work in an effort to uncover what may have been happening to produce the results that were collected. This work was performed through *in silico* experiments, simulating up to $1x10^7$ proteins as they moved along their substrates. Initially, several arrays were created to contain multiple values pertinent to translocation. An arrays for each probability in the chosen model with a range of values that were pertinent to the investigation at hand; for example, a model where it was allowable for the protein to perform hops on the DNA backbone was considered. In this setup, the protein may have dissociated from the substrate and rebound further towards the 5' end, effectively 'hopping' the distance with no additional energy input. This simulation included probabilities for translocation, dissociation, and jumping. The arrays would been constructed in this manner:

$$p_{translocation} = [10, 15]$$

$$p_{jumping} = [1, 3, 5, 15]$$

$$p_{dissociation} = [.1, 1.5, 5]$$

An array for the predetermined DNA lengths, similar to the lengths utilized in experimentation would be created, similarly to those shown above; and arrays for the free protein in solution and ADP produced which were the same size as the number of DNA lengths being investigated. The number of proteins to be simulated and the total experiment time were left as fixed values for a given round of simulation.

The first step was to populate the free protein to the DNA segments. First, the probability of binding must be determined for the given substrate and protein pair. When simulating ISWI, dissociation constants of 42 nM were used from Al-Ani *et al.*[75] and the concentration of bound protein determined using the equation

$$p_{bound} = \frac{\left(\frac{l}{K_D}\right)}{1 + \left(\frac{l}{K_D}\right)}$$

where l is the concentration of substrate in molar units and K_D the dissociation constant in molar units (see Appendix A). Simulating the experimental conditions and the substrate concentration therein, gave a probability of bound protein at .998. This is in line with the assumption that for a tight binding protein and large excess of substrate molecule, the binding agent will be nearly completely bound in solution.

As an example, for 1000 simulated proteins, it was first determined how many proteins would be bound by multiplying the number of proteins in solution by the binding probability, $n_{bound} = n_{total} * p_{bound}$. A loop was created that would iterate through the total number of bound proteins and randomly select a position along the DNA array, bind a protein at that position by adding one to the current value, and then performing the same function to the next available protein. For a DNA segment of with five available binding sites, the results may have looked similar to

$$DNA[\,] = [[213][229][186][174][196]]$$

This particular method for determining binding site does not describe a single molecule of DNA with 998 proteins attached. Instead the array represents all possible DNA segments and the total number of instances with a protein attached to site 1 (213), site 2 (229), *etc*. This is a computationally more efficient method of protein simulation than creating 998 individual DNA segments with 5 sites, randomly binding a protein to each, and mimicking movement individually. After the requisite number of proteins had been bound, the remainder were stored as in an array with a position for each DNA length.

Once all protein had been bound and the free protein logged, the simulated experiment began. A time counter was created at zero seconds and iterated forward for a given time step that was chosen for the specific simulation being ran. Common time steps were on the order of 0.001 to 0.0005 seconds. As the time scale of translocation for an individual protein was on the order of seconds, moving the simulation

forward at small fractions of seconds assured that a small segment of the bound population would move at any given time and would produce seemingly continuous results. For each translocation stage, the number of proteins at each site was determined and an empty array of identical size to the DNA segment was initiated. A loop would begin at position one, repeat for each protein present at the site (213 in our example), randomly select a number between zero and one, and compare it to the value necessary for translocation. As each probability was initially set with the units of seconds for all possible steps in the model, *e.g.* translocation, dissociation, pause, *etc.* they were first reduced by the same factor that determined the time step size, generally 1000 to 2000. The empty array would then have one added to the site if a protein was chosen for movement and one subtracted if chosen for dissociation. These steps would repeat for each site until all bound proteins were counted and a resultant movement array would have looked similar to

$$DNA_{move}[\,] = [[-1][0][2][1][1]]$$

For proteins bound at the final site, position five in this case, if a protein was chosen to translocate forward, it instead was subtracted from the population and one added to the free protein counter, similar to a dissociation event at any site. After successfully accounting for all bound proteins, the values in DNA_{move} were then added to the initial DNA array to finalize the movement phase of the time step.

Once all proteins had been given the opportunity to undergo motion and any movement resolved, the population of unbound proteins then underwent a similar procedure to determine if they bound to a DNA site. A fixed-value probability of binding was initialized at the beginning of the simulation and was modified by the time-scale factor. A loop then counted through each unbound protein, generated a random number, and compared that number against the modified probability value. If a rebinding event was determined to take place, it would be randomly assigned to a position along the simulated DNA, one added to that position in the DNA array, and one removed from the free protein population.

The DNA final position value along with the total ADP produced was then written to a data file and charts created detailing each population with respect to time. This output formed the basis of the

comparison to physical data and guided the formation of the mechanistic model. All simulation code was written and compiled in Python using the Spyder IDE and performed on a Windows XP workstation. The Python 2.7 release was used in conjunction with NumPy 1.8.0.

Chapter 5 – RSC

5.1 - Introduction

The chromatin remodeler RSC (Remodels the Structure of Chromatin), from the *Saccharomyces cerevisiae* genome, is an essential member of the SWI/SNF-family of ATP-dependent chromatin remodelers[9]. RSC is a large protein complex consisting of 15 distinct subunits and has a total molecular weight of approximately 1 MDa[76]. It is capable of repositioning nucleosomes from the center of DNA fragments toward the ends without disrupting the integrity of the nucleosomes[76,77]; Indeed, in conditions of high concentrations of RSC, the complex has been shown to be capable of ejecting nucleosomes from the DNA[11]. Similar to many other remodelers, the ability of RSC to translocate along DNA is essential for its nucleosome repositioning activity[35,62]. At a microscopic level, determining the kinetic steps that this remodeler undergoes contributes to a fundamental understanding of this remodeler family. Current estimates for the kinetic translocation step size of RSC, the number of base pair translocated in a single kinetic step, are around one or two base pair[66]. The results of recent studies have also suggested a two-step mechanism for the repositioning of nucleosomes by RSC; the formation of a stable complex of DNA, RSC, and the histone octamer serves as an intermediate for the repositioning reaction[10,78,79]. According to this model after the initial binding of RSC to the nucleosome, competition exists between RSC and the histones for binding to the nucleosomal DNA. This then leads to the formation of an intermediate complex

characterized by the nucleosomal DNA being associated primarily with RSC and largely dissociated from the histones[10,79]. This partial unwrapping of the DNA from the histones provides an environment in which the small loops of DNA formed by the DNA translocation of RSC can more easily propagate around the surface of the octamer and subsequently reposition the nucleosome. However, despite the significance of DNA translocation to the nucleosome repositioning activity of RSC and other remodelers[80], details about the mechanism that couples the binding and hydrolysis of ATP to mechanical work remains unclear. In this study we utilize a truncated construct of RSC, termed 'trimeric minimal RSC' (RSCt), which only contains three of the original 15 subunits of full length RSC[66,81]; the subunits are ARP7, ARP9, and a truncated version of STH1. STH1 is the DNA-binding ATPase and translocation motor[81,82] of the construct whereas ARP7 and ARP9 are nuclear actin related proteins whose interaction with STH1 greatly improves the stability and solubility of the complex[76,82,83]. The truncated construct was utilized because it can be over expressed in *Escherichia coli*, unlike the full complex. This allows for increased yields during the purification process. This construct has been used previously by Sirinakis *et al*. to characterize the translocation properties of the RSC motor domain as well as Malik *et al*. to study DNA binding[66,81]. In our experiments we monitored the translocation of RSCt along double-stranded DNA using a stopped-flow assay that monitors changes in fluorescence intensity associated with the translocation of RSCt along fluorophore labeled DNA[55,69,84]. Through global analysis of associated time courses of translocation we were able to determine the macroscopic rate of DNA translocation and associated kinetic step size. Through further analysis of the ATPase activity associated with DNA translocation we were also able to determine the efficiency at which RSCt couples the binding and hydrolysis of ATP to its physical movement along the DNA. This in turn provides a limit on the efficiency at which RSC could couple ATP hydrolysis to nucleosome repositioning. Finally, in a separate assay we monitored the ability of the RSC trimer to displace streptavidin from biotinylated DNA[84,85]. The ability of RSC to actively displace the streptavidin demonstrates the large forces that this molecular motor is capable of exerting during translocation and thus, by extension, during nucleosome repositioning.

5.2 – Results

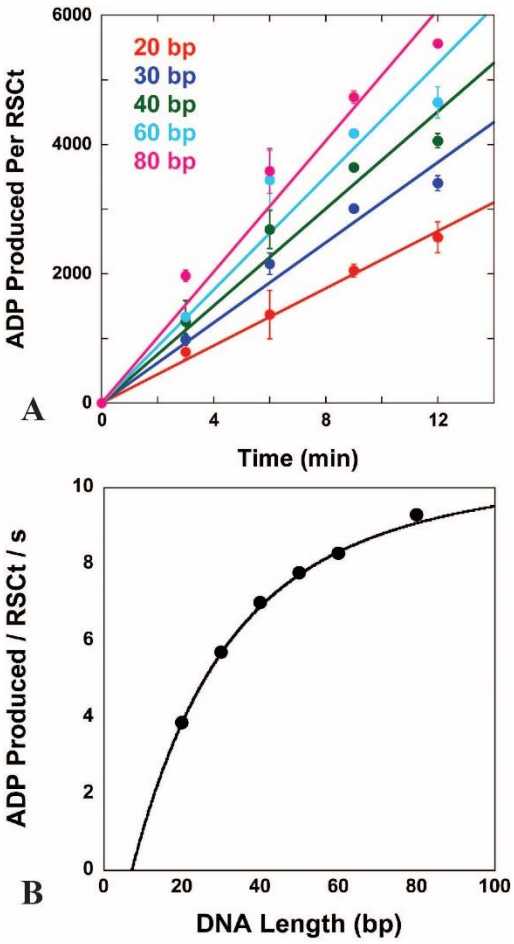


Figure 6 - RSCt Is A DNA Translocase. A) Time course of ADP production for RSCt translocating along DNA of varying lengths. The solid lines are linear fits of the time courses used to determine the value of k_{cat} for each length. B) Dependence of the k_{cat} of the DNA stimulated ATPase activity of RSCt on DNA length. The solid line in the figure is a fit of the data using Equation 6.

The results of several previously published independent experiments have demonstrated that RSC is an ATP-dependent double-stranded DNA translocase[62]. During its translocation along double-stranded DNA, RSC tracks along only one strand of the duplex with a 3' to 5' directional bias[35,62]. Furthermore, the ability of RSC to translocate along DNA is fundamental to its ability to reposition nucleosomes [35]. In order to gain further insight into the physical process of DNA translocation including the associated mechanism through which the energy obtained from the binding and hydrolysis of ATP by RSCt are coupled to the physical movement of the enzyme along the DNA we conducted a series of independent experiments designed to determine the kinetic mechanism of DNA translocation by the enzyme.

5.2.1 - RSCt Is a Viable Translocase

In our first set of experiments we determined the dependence of the steady-state rate of the DNA stimulated ATPase activity of RSCt on DNA length. Using an estimated 10 bp binding site[62], we began by establishing the concentration of DNA binding sites that corresponded to maximal stimulation of RSCt's ATPase activity (data not shown) and found it to be 10.2 μM; since these experiments are multiple turnover experiments in which RSCt experiences several rounds of DNA binding, DNA translocation, and dissociation all DNA concentrations are presented in terms of binding sites[65]. All subsequent experiments were performed under identical saturating DNA concentrations. The results of a

series of DNA-stimulated ATPase assays with RSCt and different lengths of DNA are shown in Figure 6A; the solid lines in Figure 6A are linear fits of this data used to determine the apparent maximal rate of ADP catalysis, $k_{cat,}$

$$k_{cat} = ck_t\left(\frac{n(1-P) - P(1-P^n)}{(1+n)(1-P)}\right) \quad \textbf{(8)}$$

for each DNA length. As shown in Figure 6A and Figure 6B, the steady-state ATPase rate increases with increasing length of DNA, consistent with RSCt being a DNA translocase[65]. Indeed, since all of these experiments were conducted under conditions of equal RSCt binding site concentrations, the DNA length dependent differences in the ATPase rate are not a trivial consequence of a DNA length dependent change in the binding rate of the enzyme[65]; *i.e.*, a faster ATPase rate for longer DNA substrates is not a result of those substrates having more RSCt binding sites and thus a faster rebinding rate for dissociated RSCt.

A determination of the efficiency at which this ATPase activity is coupled to DNA translocation requires an additional determination of the kinetic mechanism of DNA translocation by RSCt. Since RSC tracks along a single strand of the duplex during DNA translocation we monitored the translocation of RSCt along DNA using a modified form of a single-stranded DNA translocation assay[56,63,64]. In these experiments, a pre-incubated solution containing RSCt and double-stranded fluorophore-labeled DNA is rapidly mixed with ATP and the subsequent time course of fluorophore florescence intensity is monitored. Since the interaction of RSCt with the fluorophore results in a quenching of its fluorescence intensity (Figure 7A and B) the resulting time course is directly correlated with the translocation of RSCt along DNA[55,63].

As shown in Figure 7B the time courses of fluorescence observed in these experiments consisted of several phases: the first phase is a rapid decrease in the fluorescence intensity of the fluorophore. Given that this signal change was present even in the absence of ATP (Figure 7A), we attribute this phase to

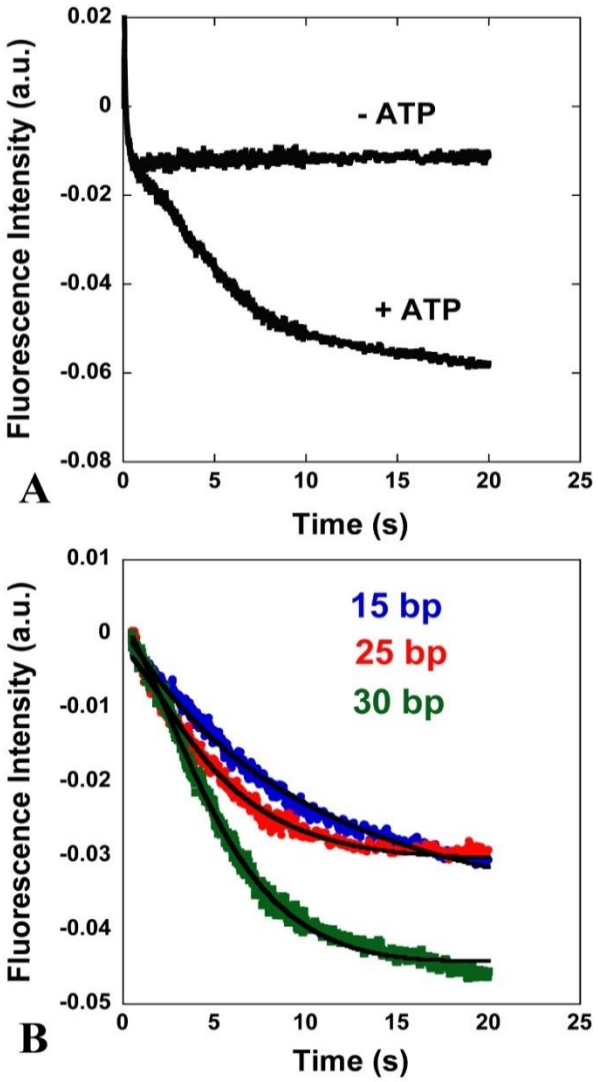


Figure 7 - Monitoring DNA Translocation By RSCt Using A Fluorescence Stopped-Flow Assay. A) ATP +/- Controls. A stopped-flow chart showing the spectrographic signal involved in binding and translocation in the presence of ATP as well as binding with no translocation in the absence of ATP. These results show that ATP is a required component in translocation stopped-flow experiments. Without the presence of ATP, we see no change in fluorescence signal indicative of molecular translocation. Only upon the addition of ATP to the mixing solution do we see any signal change characteristic of translocation by RSC. In very short time-courses (> 0.5 s), we detect a signal change that is present in both –ATP and +ATP experiments. Past this timescale is when we are able to determine defining characteristics between the two sets of data. This data set was taken using 25 bp double-stranded fluorescein-labeled DNA. B) Time courses for translocation of RSCt along 15 (red), 25 (blue), and 30 (green) bp DNA substrates. The data set for each DNA length is an average of 25 independent experiments. The black lines show the results generated by a global NLLS fit of the data using Equation 4 which describes an initiation step model with the DNA length as a defining parameter.

processes not associated with the translocation of RSCt along the DNA (*e.g.*, a redistribution of free and bound RSCt following the mixing and dilution of the sample in the stopped-flow instrument). Additional phases are observed only in the presence of ATP and are associated with further decreases in fluorescence intensity of the fluorophore. Since time courses of these changes depend upon the length of the DNA (Figure 7B), they are associated with the ATP-dependent directionally-biased translocation of RSCt along the DNA [62–64]; In order to avoid the presence of ATP-independent processes affecting the analysis of the ATP-dependent translocation, we restricted our analysis to the portions of the time courses occurring after the first 0.5 seconds of the mixing reaction. There is no longer any change in the fluorescence time courses observed in the absence of ATP in this region (Figure 7A).

In previous applications of this assay to monitor the translocation of helicases along single-stranded DNA, a protein trap was included with the ATP to prevent rebinding of dissociated protein to DNA[60]; this ensures that only single-turnover DNA-binding kinetics were observed. The inclusion of a protein trap in our experiments resulted in rapid and

total dissociation of RSCt from the DNA because of the ability of the trap to actively displace RSCt from the DNA (data not shown); active displacement of DNA translocases by protein traps has been previously observed[56,60,64]. Because of this, we were unable to monitor DNA translocation by RSCt in the presence of a protein trap and thus the time courses that we observe are influenced by both pre-steady state and steady-state kinetics. To minimize the complications associated with analyzing these data we conducted all experiments using equal concentrations of total DNA binding sites for RSCt for each DNA length[65] and thus the kinetics of DNA binding by initially free and dissociated RSCt are the same for each length of DNA. Interestingly, we still see length-dependent changes in the data indicating a contribution from pre-steady state kinetics. If these time courses were completely dominated by the steady-state signal, the traces would necessarily be the same for all lengths of DNA.

In our first attempt to analyze these translocation time courses we used a multiple turnover translocation model as derived in Fischer *et al.*[55,63,64] In this model RSCt is assumed to bind randomly, but uniformly, to the DNA; thus, we ignore any potential variations in binding affinity of RSCt near the ends of the DNA associated with differences in counter ion condensation at the ends[68,86,87]. RSCt is initially bound in a random location, *i* translocation occurs steps away from the end of the DNA, where *i* is constrained between the values of 1 and *n*, the maximum number of translocation steps required for RSCt to move from one end of the DNA to the other. Upon addition of ATP to the system, RSCt translocates along the DNA, tracking along one strand of the duplex with 3' to 5' directionality[62,88], through a series of repeated rate-limiting translocation steps with associated microscopic rate constant k_t. During each rate-limiting step RSCt moves *m* base pairs along the DNA and hydrolyzes *c* ATP molecules; the microscopic parameter *m* is referred to as the kinetic step size[54,65]. The microscopic rate constant for dissociation during translocation is k_d. All initially free and dissociated RSCt can bind the DNA with a rate constant k_b.

This model was unable to provide a constrained fit to the time courses (data not shown). This suggests that an additional process, not included in this model, is occurring during DNA translocation[63]. Based upon previous studies of other DNA translocases the two likely candidates for this additional process

are either an initiation step preceding processive translocation[65] or a dissociation event from the end of the DNA that is different from the other dissociation process[64,68]. Unfortunately, we were unable to derive time-dependent expressions for the population of protein at the end of the DNA associated with these models for arbitrary lengths of the DNA (*i.e.*, for arbitrary values of *n*). Since determining the kinetic step size of translocation requires analyzing the dependence of *n* on the length of the DNA[54,56] we instead analyzed these time courses using Equation 3. In this analysis, effects of all processes not associated with the processive translocation of the enzyme (initiation, rebinding of dissociated RSCt, *etc*.) are related within the single parameter k_i. Analysis of computer simulated time courses of DNA translocation using this approach confirmed that this approach resulted in correct estimates of *m*, k_t, and k_d for these data sets, but that estimated values of k_i were dependent upon the rates for all the processes not associated with processive translocation (the rate of rebinding, the rate of initiation, *etc*.). For this reason, estimates of k_i determined from the analysis of our time courses are influenced by both initiation and rebinding processes of RSCt. The results of simultaneous global NLLS of all kinetic time courses to Equation 3 are shown in Figure 7B and Table 1. This analysis provides a good description of the time courses as judged by the variance of the fit, and returns a value of 2.3 ± 0.1 bp/s for the macroscopic rate of DNA translocation with a processivity and kinetic step size of 0.92 ± 0.01 and 1.24 ± 0.18 bp, respectively.

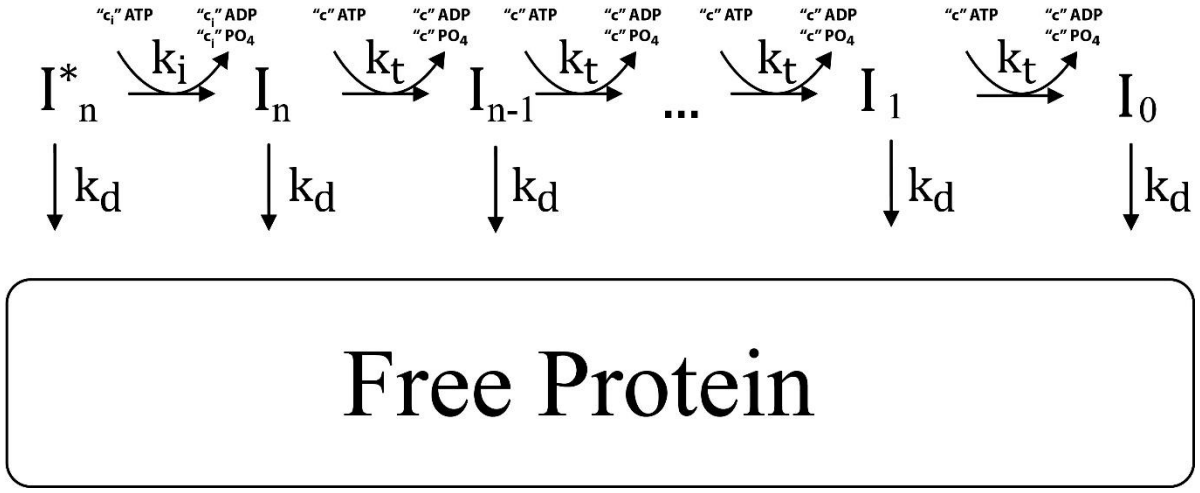
5.2.2 - ATP Hydrolysis Is Weakly Coupled To DNA Translocation

To determine the stoichiometry of ATP binding and hydrolysis associated with DNA translocation by RSCt we also analyzed the ATPase activity of RSCt during translocation (Figure 6) according to Scheme 1. In each round of translocation, as the protein moves from position I_i to I_{i-1}, (rate constant k_t from Scheme 1) it uses the energy from ATP hydrolysis[76]. Similarly, during the combined slow-step phases modeled in Scheme 1 (rate constant k_i) we allow for ATP hydrolysis to also occur in combination with the slow-step phases. Utilizing an ATPase model based upon the possibility of equal probabilities of binding along the double-stranded substrate along with a pausing step derived in Khaki *et al.*[68] and shown in Scheme 8, we were able to fit to this model (Equation 6) and thereby distinguish contributions to ADP production from

either translocation or the aggregated slow processes. In our analysis of the data in Figure 7B using Equation 6 we fixed the values of k_t, k_d, k_i, r, m, and d to those determined from our previous NLLS analysis of our fluorescence translocation time courses (Figure 8B and Table 1). The solid line in Figure 6B is the result of this analysis which returned a value of $c = 3.77 \pm 0.02$ ATP per step which corresponds to an ATP utilization rate of 3.0 ± 0.4 ATP/bp for DNA translocation by RSCt. Furthermore, it was also determined that during any aggregated pausing-step phases associated with the effective rate constant k_i step in Scheme 2, it utilizes 11.69 ± 0.11 ATP per step. As shown in Scheme 8, this constant corresponds to the value of c_i in Table 1 which is a measure of the number of ATP utilized in the aggregate slow-step phases. However, since we are unable to determine estimates of the individual kinetic steps that are absorbed into k_i, we are unable to determine the amount of ATP utilized per individual step within that group of processes.

	Pausing Step Translocation Model	Streptavidin-Displacement Model	ATPase Model
k_t [s^{-1}]	2.3 ± 0.4	*	*
k_d [s^{-1}]	(2.03 ± 0.02) x 10^{-1}	*	*
k_i [s^{-1}]	(7.47 ± 0.13) x 10^{-2}	*	*
k_s [s^{-1}]	-	(7.16 ± 0.05) x 10^{-3}	-
m [bp]	1.24 ± 0.18	*	*
d [bp]	14.3 ± 0.2	*	*
C [$ATP/step$]	-	-	3.77 ± 0.02
C_i [$ATP/step$]	-	-	11.69 ± 0.11
C/m [ATP/bp]	-	-	3.0 ± 0.4
$m*k_t$ [bp/s]	2.9 ± 0.1	-	-
P	0.92 ± 0.01	-	-
$\frac{P}{1-P}$	12 ± 2	-	-
$\frac{m*P}{1-P}$ [bp]	14.3 ± 0.6	-	-

Table 1 - Derived and Fitted Parameters For All RSCt Data Analyses. The model utilized is denoted in the column header and are listed in Chapter 4. Starred cells are fixed parameters from the fitting of fluorescence translocation data (Column 1) and dashed marks denote parameters not included in the fitting equation.



Scheme 8 - Paused-Step ADP Production Model. This model is based on Scheme 6 and is used to represent the physical understanding of Equation 6. As the protein begins the k_i step before processive translocation, it utilizes c_i molecules of ATP. However, during each forward step with a rate constant k_t, the protein utilizes c molecules of ATP.

5.2.3 - RSCt Is Capable Of Disrupting the Streptavidin-Biotin Interaction in A DNA Length-Dependent Manner

Finally, in an effort to qualitatively assess the forces that RSCt exerts during DNA translocation, we monitored the ability of translocating RSCt to displace streptavidin from biotinylated DNA. A similar assay has been used to quantify the DNA translocation activity of viral helicases[85]. The streptavidin-biotin interaction has one of the strongest binding affinities in nature, with an affinity on the order of 10^{-14} molar (M)[89] (see Appendix A) and thus the ability of RSCt to disrupt this interaction can be used as a qualitative measurement of the minimum potential force that RSCt can apply during DNA translocation. We performed assays involving 40 and 80 bp double stranded DNA molecules that had been biotinylated at the 5' end of one strand while the other end of the complimentary strand was left available for radiolabeling. As described in Chapter 4, we then monitored the streptavidin-biotin-DNA concentrations versus free DNA over the course of an hour to examine the change in populations. In the absence of RSCt or ATP there was no change in the population of streptavidin-bound complexes. However, when ATP was added to the reaction there was a significant change in the amount of free DNA in solution (Figure 8A).

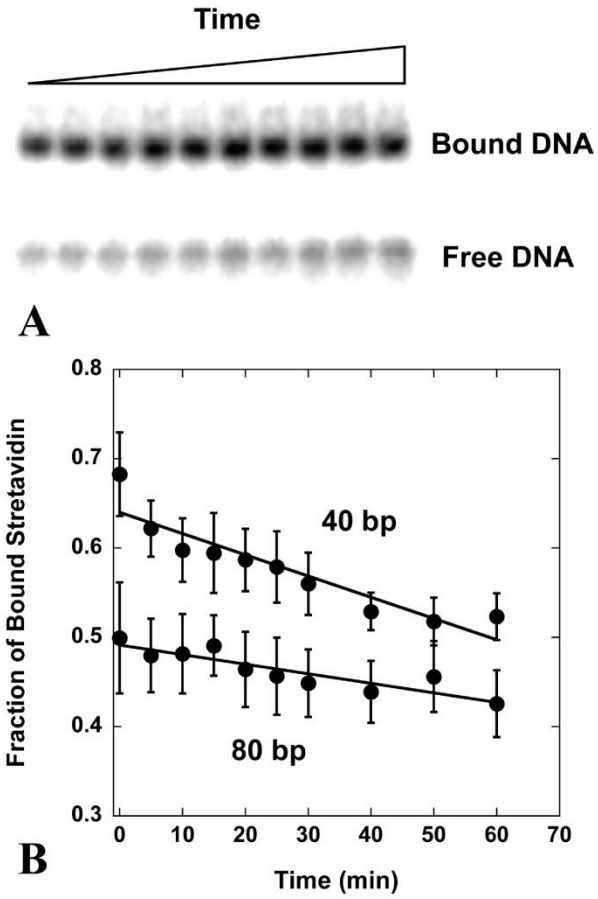


Figure 8 - Streptavidin Displacement Due To RSCt. A) Native gel example showing streptavidin displacement due to RSC. At increasing time intervals (increasing lane numbers), we see larger signal from free DNA indicating streptavidin displacement from the DNA-biotin linkage. This control experiment was conducted using 40bp DNA substrate at the listed concentrations of streptavidin, RSCt, and DNA as described in Appendix B. B) Streptavidin displacement versus time along 40 and 80 bp DNA substrates. The data sets are averages of three runs and the standard deviation is shown. The solid lines are linear fits to the data weighted by their respective errors. Note that RSCt displaces more overall streptavidin on the 40 bp substrate than on the 80 bp substrate.

Another interesting characteristic of this experiment is that the displacement activity of RSCt is also length-dependent, as shown in Figure 8B. We chose two disparate lengths of double-stranded DNA to monitor, one of 40 bp and another of 80 bp, and compared their respective rates of streptavidin displacement over time. As shown in Figure 8B, the steady-state of streptavidin displacement by RSCt is faster for the 40 bp DNA than for the 80 bp DNA. This is consistent with the 40 bp DNA having approximately 55% of binding sites lying within the DNA translocation processivity range of the 5' biotinylated end of the chain whereas on the 80 bp DNA substrate, only ~22% fall within that range. RSCt bound to the 40 bp DNA will thus have a higher probability of being initially located within translocation distance of the biotin and consequently will more frequently come into contact with the streptavidin and thus have more opportunity to disrupt the bond than RSCt bound to the 80 bp DNA. Using this data across two lengths of DNA, we globally fit the data to a steady-state model of translocation and streptavidin displacement, shown in Equation 5. From this analysis we came up with a value for the rate of streptavidin displacement of $(7.16 \pm 0.05) \times 10^{-3}$ displacements per second. All other rate constants in this analysis were constrained to values determined from previous analyses.

5.3 – Discussion

Since DNA translocation by chromatin remodelers is required for these enzymes to reposition nucleosomes, a complete understanding of nucleosome repositioning by these enzymes requires a quantitative characterization of their mechanisms of DNA translocation. This includes not only a determination of the associated rate constants, but also an estimate of the efficiency at which these enzymes can couple the binding and hydrolysis of ATP to the physical work of DNA translocation. Here we have demonstrated a straightforward and generally applicable ensemble approach to obtaining both this information and an associated qualitative measurement of the forces exerted by the enzyme during DNA translocation.

Based upon the analysis of our results, we propose that RSCt must undergo an initiation process following its binding to DNA before becoming competent for DNA translocation. Furthermore, we propose a very weak coupling of ATP binding and hydrolysis to DNA translocation by RSCt. After considering the results of additional computer simulations, we propose that binding and possibly subsequent hydrolysis of ATP by RSCt can resort either in forward translocation or in a "paused" state of the enzyme. The presence of these periodic paused states further decreases processivity of DNA translocation by the enzyme. Finally, we also find that the molecular motor of RSCt has sufficient motive force to dissociate streptavidin from its biotin interaction during DNA translocation, indicating the complex's inherent capabilities for remodeling nucleosome structures in chromatin.

5.3.1 - Determination of Microscopic Translocation Parameters

The time courses of double-stranded DNA translocation by RSCt are consistent with the presence of additional processes that both follow DNA binding and precede processive DNA translocation or that occur when RSCt has reached the end of the DNA. Further analysis of these time courses favored the former and therefore that corresponding model was used to determine estimates of the microscopic rate constants

associated with DNA translocation by RSCt. It is worth noting, however, that estimates of macroscopic parameters are independent of the choice of model (data not shown)[68].

Through a global NLLS analysis of data collected using our fluorescence stopped-flow based DNA translocation assay using Scheme 2 we have determined that RSCt translocates along double-stranded DNA through an ATP-dependent reaction characterized by a kinetic step size of 1.24 ± 0.18 bp and a macroscopic rate constant of 2.3 ± 0.4 bp/s. These results are consistent with previously published studies of double-stranded DNA translocation by the full RSC complex which have suggested a physical step size on the order of 1 nucleotide[35,62]. Our estimate of the processivity of DNA translocation, P = 0.92 ± 0.01, corresponds to a molecule of RSCt moving an average of 12 ± 2 steps before dissociation[65]. This is similar to but slightly smaller than what has been previously reported for the full RSC complex[66,90], suggesting that the other subunits of the full complex may also contribute to the DNA binding affinity of the enzyme.

The macroscopic translocation rate of DNA translocation by RSCt, 2.3±0.4 bp/s, is similar to the macroscopic rate of single-stranded DNA translocation by the NS3h helicase from hepatitis C virus (3 nt/s)[68], but much slower than the macroscopic rates of single-stranded DNA translocation by the UvrD (190 nt/s)[64], Rep (300 nt/s)[91]and T7 (132 nt/s)[92] helicases. This suggests a fundamental underlying structure-function difference between helicase superfamilies with SF2 helicases (NS3h and RSC) translocating along single- or double-stranded DNA much more slowly than SF1 (UvrD and Rep) or DNAB-like (T7) helicases. Of course, while these numbers are interesting to make general comparisons, it is difficult to draw distinctive conclusions from them as studies of these enzymes were performed under varying conditions.

5.3.2 - Computer Simulation Results

We performed simulations in which hydrolysis of each ATP molecule resulted in forward motion of protein along the DNA, in backward motion along the DNA, or in a pause (or stall) in the motion of the

protein[93]. Remarkably, the resulting translocation time courses could be modeled using Scheme 2 even in the presence of non-uniform motions. We found that the inclusion of either backward motion or pausing events resulted in an increase in the estimate of the ATP coupling efficiency (*c*/*m*) and a decrease in the estimate of the macroscopic translocation rate ($m*k_t$). Although the estimate of the kinetic step size (*m*) was increased in the presence of backward motion it was unaffected by random pausing. Since the value of *m* determined from the analysis of the time courses of RSCt translocation along double-stranded DNA is 1.24 ± 0.18 bp (and thus is unlikely to be significantly over-estimated) we believe that the most likely explanation for the seemingly large value of *c/m* observed for RSCt is that the enzyme undergoes random pausing during its translocation along the DNA. The occurrence of random pausing, especially when dissociation from the DNA substrate might occur during the paused state, would also be consistent with the relatively low processivity of DNA translocation that we observe for RSCt when compared to DNA translocation by helicases such as UvrD or NS3h, for which there was no indication of such non-uniformity in the translocation process. Furthermore, this would also be consistent with the observation that the ATP coupling stoichiometry for DNA translocation by UvrD (1 ATP/nt) and NS3h (0.5 ATP/nt) is lower than for RSCt (3 ATP/bp). Other remodelers in the ISWI family display low template commitment, releasing the substrate almost as fast as it is bound. It has been suggested SNF2 and hACF undergo several rounds of substrate sampling before a successful remodeling event takes place[94]. While the experiments with RSCt were not carried out on a nucleosome substrate, it is still worth mentioning that other remodeling proteins experience events that could 'pause' processive translocation.

It is worth noting that Sirinakis *et al*. report seeing backwards motion along the double-stranded DNA in their tethered StART molecule experiments[66]. However, it is difficult to determine if the backwards motion seen in those experiments is specific to the StART complex or a more general feature of the RSC motor; for example, since the StART complex is strongly tethered to the DNA initially, it is possible that backwards translocation may be enhanced when tension forces are applied.

5.3.3 - Relationship Between DNA Binding And Translocation

Because DNA translocation is an ATP-dependent process, it involves repeating cycles of ATP binding, ATP hydrolysis, and product release[95]. These individual steps affect the apparent microscopic translocation properties as well as the macroscopic values that we determine from our analysis. It has been shown that RSCt binds DNA tightly in the absence of nucleotide but has a significant decrease in binding affinity in the presence of ADP, ATP or ATP-like analogues, on the order of 1000 fold[81]. Other helicases display similar characteristics, such as Rep, RepA, and RecQ[96–98]. In this system it would seem natural that RSCt would have a higher probability of dissociation during ATP-, or ADP-bound populations. However, if the subsequent hydrolysis event occurred on a short enough timescale, the impact of this reduction of binding affinity would not greatly affect the translocation characteristics of RSCt. It is nevertheless still possible that this slightly increased chance to dissociate would result in a lower processivity value in comparison to other translocases which do not show a similar allosteric regulation of their binding affinity. Indeed, the low processivity of DNA translocation by RSCt may result from the fact that its affinity for binding DNA is reduced in the presence of both ADP and ATP-analogs, regulation that is unique among previously reported studies of DNA binding by helicases. For example, the affinity of the Rep helicase for DNA binding is reduced in the presence of ATP, but unaffected by the presence of ADP[97], and the processivity of DNA translocation by the Rep helicase is ~800 nt[91]. Furthermore, data also suggest an inverse relationship between the affinity of DNA binding and the macroscopic rate of DNA translocation. For example, the affinity of the Rep helicase for single-stranded DNA is 480 ± 70 nM[97] with a corresponding macroscopic translocation rate of 530 ± 10 nt/s[91] whereas the affinity and macroscopic translocation rate for RSCt and double-stranded DNA are 99 ± 5 nM[81] and 2.3 ± 0.1 bp/s. A similar relationship has been observed for single-stranded nucleic acid translocation by the SF2 helicase NS3h; the affinity and macroscopic translocation rate for oligomeric thymine are 139 ± 6 nM and 3.1 ± 0.3 nt/s, respectively, and for oligomeric uracil are 530 ± 60 nM and 34.2 nt/s[68], respectively. Direct comparisons

between translocases are naturally problematic because of differences in solution conditions and assays, but these trends are nevertheless noteworthy.

5.3.4 - RSCt Utilizes Several ATP Molecules per Translocative Step

It is important to remember that the kinetic step size determined for double-stranded DNA translocation by RSCt is a measure of the average number of nucleotides translocated between two repeated rate-limiting steps in Scheme 2. Thus, both m and k_t may not reflect the step size or rate constant for the physical movement of the enzyme, but rather the step that is rate-limiting, a protein conformational change, the release of ADP or inorganic phosphate, *etc*.[63,64]. Further insight into the physical mechanism of DNA translocation by RSCt can be determined from an analysis of the DNA-stimulated ATPase activity of the enzyme[68]. Indeed, through a global analysis of the dependence on DNA length of the maximum DNA-stimulated ATPase rate (*i.e.,* the rate obtained under saturating DNA concentrations) we determined that RSCt hydrolyzes 3.77 ± 0.02 ATP per second per RSCt molecule per rate-limiting step. This corresponds to RSCt hydrolyzing 3.0 ATP per bp translocated, which is less efficient than the previously reported ATP coupling stoichiometries for single-stranded DNA translocation by other helicases[68,92]. Furthermore, this result suggests that a large amount of futile ATP hydrolysis (*i.e.*, hydrolysis not connected with the translocation of RSCt along the DNA) occurs during the translocation reaction. This may correspond to hydrolysis occurring during the additional kinetic process observed for the kinetic mechanism of DNA translocation. It is difficult to consider how these sorts of pausing, futile occurrences would affect the results of our modeled pausing events. It could be the case that it is not an individual event that would directly relate to the modeled pausing step but a collection of cases whose sum would be the value found in our results. Over the course of an average of 12 steps in a round of translocation, any pauses taken, involving ATP-binding, hydrolysis, movement, and DNA rebinding, would show up in our data fitting as a single microscopic parameter. Nevertheless, without direct experimental results on the rapidity of ATP hydrolysis and release during translocation, this remains only a conjecture in regards to our results. It is still not

possible to determine if the modeled pausing event is a single, long event, or the accumulation of multiple fast steps in each round of translocation.

Another explanation for the poor coupling stoichiometry and the large amount of futile hydrolysis could be that RSCt experienced non-uniform motion during its translocation along the DNA; possible examples of non-uniform motion would be backward motion or random pausing[56]. Indeed, based upon results of computer simulated DNA translocation time courses we believe that RSCt undergoes frequent random pausing during its translocation along the DNA. During these pauses RSCt can dissociate from the DNA and/or undergo rounds of futile ATP hydrolysis. As discussed above, a model for DNA translocation by RSCt which includes ATP-dependent random pausing would also be consistent with the relatively low processivity of DNA translocation by RSCt since the enzyme can dissociate from the paused state. Thus, together our kinetic data favor a model in which the binding and/or hydrolysis of ATP is associated either with forward motion or pausing. This, in turn, argues in favor of the pausing being a stalled or failed translocation of the enzyme along the DNA.

As determined above, the maximal rate of DNA-stimulated ATPase for RSCt is 3.77 ± 0.02 ATP per RSCt per second. In our DNA experiments with 50 nM RSCt and 1 mM ATP, approximately 9% of the ATP was consumed during the first 8 minutes of the reaction and approximately 14% consumed over the entire experiment. Since a linear fit was consistent with our ATPase time-courses (Figure 7), this suggests that the kinetics of ATP binding are not rate-limiting for the DNA-stimulated ATPase of RSCt. Thus, this decrease in ATP concentration appears to not significantly affect our results; we nevertheless wished to determine how this decrease in ATP concentration would affect our estimate of k_{cat}. Based upon model equations derived in Fischer *et al.*[63] which model the apparent rate constant of translocation, k_t, in terms of the concentration of ATP, the equation

$$k_t = \frac{K_T [ATP]}{1 + K_T [ATP]} k_f$$

shows the ATP concentration dependence of k_t. The variable K_T is the equilibrium binding constant, which is a ratio of the microscopic ATP-binding rate constant k_1 and the microscopic ATP dissociation rate constant k_{-1}, and the variable k_f is the rate constant governing processive forward translocation after nucleotide binding. The equation makes the assumption that the dissociation rate constant is much faster than the forward rate of translocation ($k_{-1} >> k_f$). From Malik *et al.*[81], we know that the equilibrium constant for protein/DNA interaction with ATP is 9×10^3 M^{-1} which is also equal to the ratio k_1/k_{-1}. From previously published estimates of the association rate constant k_1 being on the order of 10^7 $M^{-1}s^{-1}$[99], implying a dissociation rate, k_{-1}, of 10^3 s^{-1}. From the results of our global analysis, we know that processive translocation occurs for RSCt at a rate of ~3 s^{-1} which puts us well within the regime $k_{-1} >> k_f$. After affirming this underlying assumption, we continued with the analysis of the above equation. Considering a reduction of 14% from the initial 1 mM ATP concentration, the forward translocation constant k_t only loses 1.6% of its value. Due to this very small drop in k_t, the processivity remains within 0.5% of its original value. Finally, the equation for k_{cat}, the maximum ATPase rate under saturating nucleic acid conditions, (Equation 6) is heavily dependent on k_t and P, this reduction of overall ATP concentration gives a negligible change in the value of k_{cat}.

5.3.5 - RSCt Is a Relatively Slow and Short Processive Translocase

From our analyses we determined the processivity of double-stranded DNA translocation by RSCt to be 0.92 ± 0.01 which corresponds to a low average number of 14.3 ± 0.6 base pairs traveled by RSCt before dissociation. This estimate is similar to a previously reported value of 20 ± 1 bp by Fischer *et al* [65] for the full RSC complex[76,83]. The small difference can most likely be attributed to the full complex of RSC having a different interaction with its double-stranded substrate than the smaller RSC trimer. In comparison to other translocase proteins, RSCt is a relatively poor translocase. The HCV helicase NS3h has been shown

to have a processivity value between 0.993 and 0.98 corresponding to an average of 230 base pairs traveled before dissociation[68,84]. The *E. coli* helicase UvrD has an even higher value of 0.997 and 2400 bp covered [64].

The kinetic step size that we determined from our analyses is 1.24 ± 0.18 base pair and agrees with previous determinations in which, during experiments measuring the capability of RSC to displace triple-helix forming oligonucleotides, a single base pair gap in the tracking strand was sufficient to obstruct the displacement activity [35,62,76].

5.3.6 - RSCt Is Capable of Exerting Large Forces during DNA Translocation

The fact that RSCt can dislocate streptavidin from biotin is significant. It demonstrates that the motor domain possesses sufficient strength to remove a bond which has an affinity on the order of 10^{-14} M. Furthermore, this information can provide an estimate of the power available for the protein's necessary task of chromatin remodeling. A protein's capability of disrupting this strong bond points to similarities in the forces required for dislocating the linkages between a histone octamer and its associated DNA. Although RSCt is clearly capable of displacing streptavidin in our assays, the reaction is extremely slow, happening every 2 minutes on average. A possible scenario for this situation would be that when RSCt translocates to the end of the DNA strand and comes into contact with the biotin-streptavidin linkage, it primarily dissociates and displaces the streptavidin with a much low probability. Sirinakis *et al.* provides an upper-bound estimate of the forces involved in single-molecule DNA-translocation of 30 pN[66]. Considering that nucleosomal structure has been shown to be disrupted at a mechanical force of 23 pN[100], these results lend further evidence that the DNA-RSCt interaction strength is large enough to displace the contacts at even high-strength binding locations such as the nucleosomal dyad[25].

5.3.7 - Comparison to Single Molecule Studies

A similar minimalist RSC construct was utilized by Sirinakis *et al.* in a single molecule DNA translocation assay [66]. In those experiments, however, the STH1 component of the trimeric complex was modified to include a fusion to the site specific DNA binding protein TetR[66]. Although the results of that study do not allow for a determination of the kinetic mechanism of DNA translocation and associated microscopic parameters the authors did provide an estimate of the macroscopic kinetic parameters which are noticeably larger than those we present here. Sirinakis *et al.* report that RSCt translocates an average of ~35 bp per round of translocation [66], which is double our estimate of 14.3 ± 0.6 bp. Interestingly, in studies of the DNA translocation activity of full-length RSC larger estimates of kinetic parameters were observed for single-molecule studies than for ensemble studies[65,88,90]. In order to resolve this discrepancy it was proposed that RSC may have two distinct processive modes[65]; a common, less processive mode that is not readily observable in single-molecule experiments and provides the basis for ensemble method results, and a more rare, highly processive mode, which dominates data recorded in single-molecule experiments. In ensemble experiments this latter mode would not have a large effect on the averages found due to the infrequent nature of the particular mode. Conversely, in single-molecule experiments it would play a significant role owing to the limited resolution of the associated instruments. A typical minimum loop detection size limit of 10 bp in single-molecule experiments would necessarily exclude the common mode from being detected, thereby increasing the apparent average distance translocated.

However, ensemble experiments sample the modalities in their natural frequencies, providing more complete averaging bases. Coincident with a loop size bias inherent within the detection resolution of the instrument, the reported average distance traveled of 35 bp may also be affected by numerous extremely large measurements, several orders of magnitude larger than the stochastic average, such as the single instance of 1500 bp translocated[66]. It is also possible that the high affinity of the TetR tag on the StART complex for the DNA itself is also artificially increasing the processivity of translocation and thereby making direct comparisons between the two sets of experiments problematic. Other dissimilarities become apparent in the comparisons between RSCt and StART which could potentially be explained by the

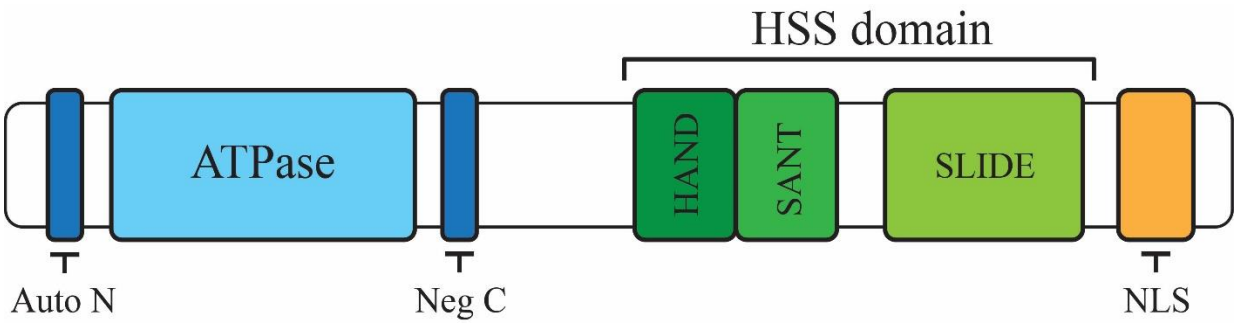
difference in structure between the two proteins. In Sirinakis *et al*. when discussing the properties of their StART complex, it is mentioned that it has the same DNA-dependent ATPase activity as the full RSC protein which corresponds to ~5 ATP/sec[66]. Combined with their reported translocation speed of ~25 bp/sec, this would imply a rate of 5 bp/ATP, which seems inconsistent with the estimate of 2 bp/step for a kinetic step size. However, there is good agreement between the high-force-exerting capabilities of the core RSC complex. While we have shown that RSCt has extreme motile forces in translocation by showing biotin-streptavidin displacement during translocation, the single-molecule assay points to forces generated in their experiments of up to 30 pN. Both of these results suggest that during translocation and nucleosome repositioning, large forces may be required to complete the complex's role in the nucleus.

Chapter 6 – ISWI

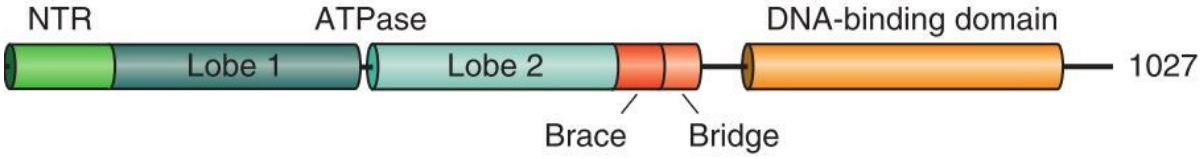
6.1 Introduction

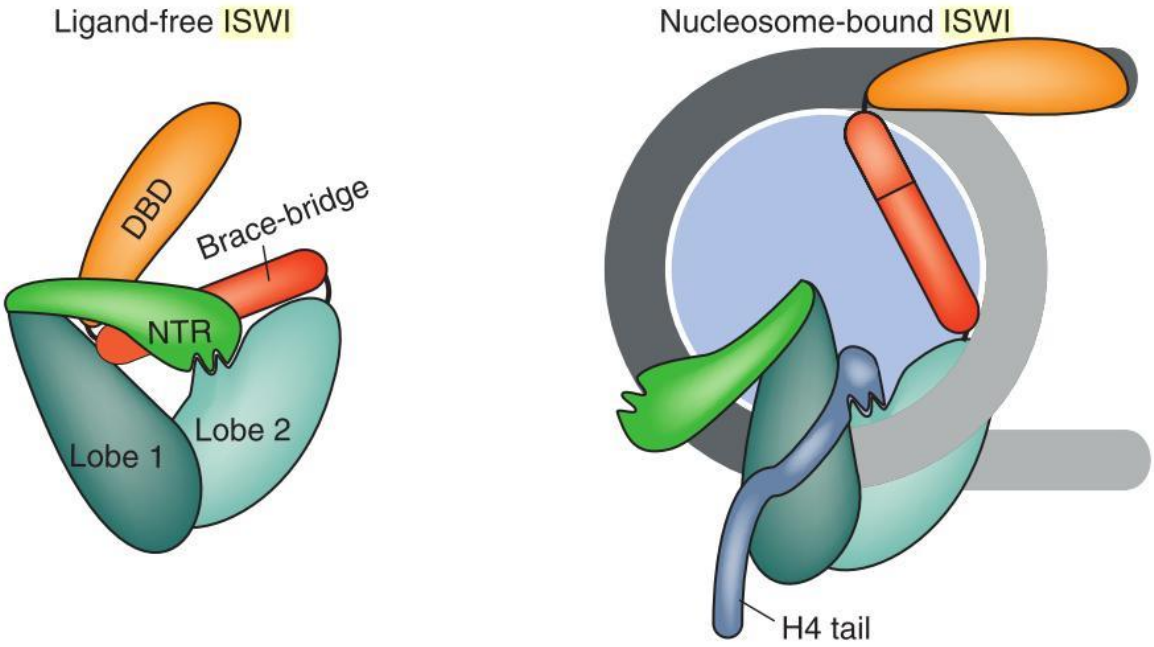
While all chromatin remodeler families share a distinct, highly conserved ATPase domain within themselves, members of specific families are differentiated based upon additional non-catalytic domains that their members contain. The conserved catalytic domain of the ISWI family, likewise termed ISWI, was initially discovered through sequence similarity to Snf2p, a member of the SWI/SNF subfamily[101,102]. Given this relationship, it was named Imitation SWI/SNF (ISWI)[103]. This motor has been shown to localize on the nucleosome two helical turns from the dyad. This is consistent with the remodeler using helicase-like DNA backbone translocation to mobilize nucleosomes[80,104]. Indeed, ISWI utilizes ATP to reposition histone octamers along the chromatin structure to promote repression during chromatin regulation[44,105,106]. Some complexes in this family are even capable of replacing specific histones within a nucleosome[12].

There are currently two proposed models by which the position of the nucleosome is altered: the first is a 'power-stroke' model, in which the HAND-SANT-SLIDE (HSS) domain of the ISWI complex (see Figure 10) pulls multiple base pairs of linker DNA into the nucleosome (the 'power stroke'), causing torsion along the contacts between the DNA and histones[46–48]. The ATPase domain resides on the opposite end of the dyad and subsequent translocation events are responsible for resolving the association torsional strain. The second model is centered on the action of the ATPase rather than the HSS domain. In this



A



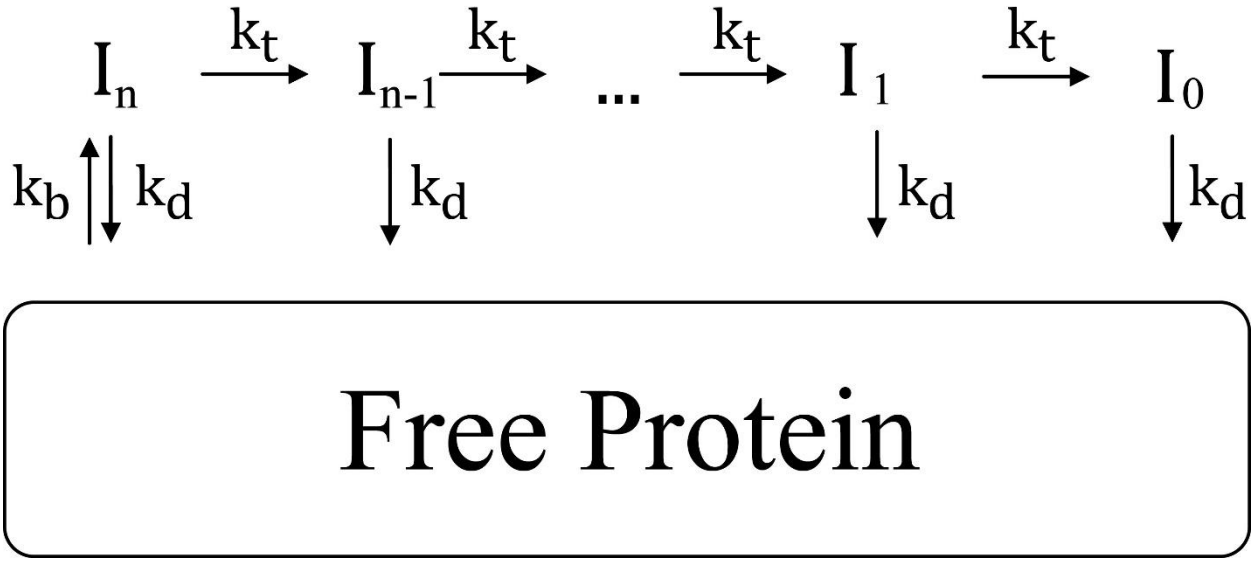


B

Figure 9 – ISWI interaction with nucleosomes. A) The ISWI protein is comprised of several domains (rectangles) and two main motifs: the first is the ATPase motif containing two additional regions, AutoN and NegC. AutoN is a region that negatively regulates ATP hydrolysis and NegC decouples ATP hydrolysis to productive DNA translocation[117]. This region is then followed by a NegC region which, when not in complex with a nucleic acid substrate, binds within the ATPase cleft between lobes one and two (see B). Finally, furthest towards the N-terminus, is the DNA recognition and binding motif, HAND-SANT-SLIDE (HSS)[58]. **B)** The N-terminal region (NTR) of the ISWI protein contains the Auto N motif, a negatively charged region that mimics a section of the H4 tail (seen right). When not in complex with the nucleosome, the NTR interferes with the second lobe of the ATPase domain and does not allow strong binding with double-stranded DNA. When an NCP is present, the H4 tail displaces the NTR and binds more efficiently to the ATPase domain, allowing increased catalytic efficiency[122]. Reprinted by permission from Macmillan Publishers Ltd: Mueller-Planitz *et al. Nat. Struc. Mol. Bio.*, **20**, (2012)

'ratchet' model, the ATPase domain pulls 7bp of flanking DNA through the nucleosome causing internal tensions that must be resolved[42] through dis- and reconnection of the DNA/histone linkages. The HSS domain is then responsible for feeding multiple base pairs of flanking DNA into the nucleosome to resolve the overall tension. These steps coincide with multiple rounds of ATP hydrolysis to continue repositioning of the nucleosome.

The isolated ATPase subunit of the ISWI chromatin remodeler family has been shown to be an ATP-dependent DNA translocase as well as an functional chromatin remodeler[72,107] capable of both constructing nucleosomes on linear DNA as well as repositioning existing nucleosomes towards the ends of a linear section of DNA[108]. In this study we determined the mechanism of DNA translocation and associated ATPase activity for the ISWI molecular motor. This information is important for understanding how the ISWI motor converts the chemical potential energy it derives from ATP hydrolysis to the mechanical work of translocation along DNA. For our experiments we employed an established stopped-flow fluorescence assay for DNA translocation that monitors the interaction between the protein and a



Scheme 9 – The individual microscopic steps of a simple translocation model. The protein binds at a random position along the DNA backbone, denoted as I_n, with a binding constant of k_b. The subscript denotes the position of the protein along the DNA backbone, position I_i, and can range from 0 to n, where 0 is the final position at the 5' end of the DNA strand and n is an integer number, ranging from 0 to L/m where L is the length of the DNA in basepair and m is the number of basepair per translocated step. From there the protein is capable of either translocating forward at a rate of k_t or dissociating with a rate of k_d. Upon successful translocation, the protein then moves m base pair to position I_{i-1}. After multiple rounds of forward translocation, the protein will both fall off and become available as a free protein to rebind, or will reach the end of the DNA at location I_0. From here it can process no further and will eventually dissociate from the substrate.

fluorophore attached to the end of the DNA substrate[55,60,84]. In addition to defining macroscopic rates of DNA translocation and ATP utilization, the analysis of this data also demonstrated that ISWI is a poorly processive DNA translocase. These combined results are closely aligned with the prediction of the ratchet model, showing a very similar translocation step size to the proposed initial 'ratchet' step. Lastly, this characterization of ISWI function is crucial to the further refinement of models of nucleosome repositioning by this important class of enzymes.

6.2 Results

There are currently multiple competing models for nucleosome repositioning by ISWI and its homologs based upon specific interactions between domains of ISWI and the nucleosome[42,46,47,109,110]. However, there have been few proposals on how the protein translocates along the DNA substrate when not in the presence of the histone octamer. To address this, we performed several assays and associated analyses to determine microscopic parameters describing double-stranded DNA translocation by ISWI and performed additional computer simulations to form a complete picture of the translocation mechanism.

6.2.1 – ISWI Translocation On DNA Is Not A Processive Mechanism

Time courses of the changes in fluorescence intensity associated with DNA translocation by ISWI consisted of two phases: An initial large decrease in fluorescence followed by a much slower fluorescence increase (see Figure 11A) that then approaches a steady-state value. The initial phase is present in controls as well as previous studies[57] and is thought to be a product of the rapid mixing within the flow cell. The results of additional control experiments conducted in the absence of ATP demonstrated that this latter phase was associated with translocation, also shown in Figure 11B. We consequently limited our data collection and analysis to only the first 15 seconds of activity following ATP addition for all subsequent studies of DNA translocation. The simplest model associated with repeated DNA rapid rebinding and translocation by ISWI is shown in Scheme 1. Based upon this scheme, we derived Equation 2 to describe

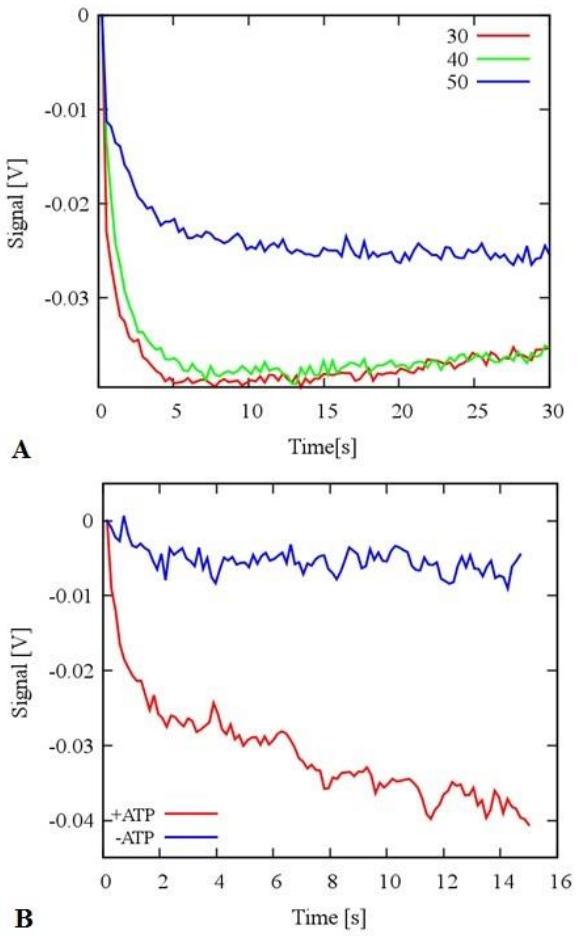


Figure 10 – ISWI translocation. A) An example of raw translocation data showing ATP-dependent translocation characteristics of the signal. Each trace is an average of a minimum of 15 traces and have had the data at very early times (<.05 s) removed (see text). This data is further normalized to a common origin at 0 V. As the data showed solely linear behavior after 15s, analysis was restricted to the initial 15s to determine the pre-equilibrium kinetics. **B)** A control experiment demonstrating that ATP is required for processive translocation.

the time dependence of the population of ISWI bound at the end of the DNA and thus in contact with the fluorophore. We began our initial fitting of the data against a simple translocation model including only forward translocation and dissociation from the substrate. The results of the simultaneous global analysis of time courses of DNA translocation by ISWI collected using three different DNA substrates (30 bp, 40 bp, and 50 bp) are shown in Table 2. These results are consistent with ISWI exhibiting both a slow macroscopic rate of DNA translocation and a very low processivity. We subsequently attempted to fit this data with different translocation models, each with additional kinetic steps (*e.g.* pausing steps, initiative steps, differential end binding)[56,111] in an effort to further refine our estimate of the mechanism of DNA translocation by ISWI (data not shown). Still, the lack of significant curvature in the fluorescence time courses in the pre-steady-state time scale and similarity in the shapes of the signals measured in experiments with the different DNA substrates prevented us from obtaining constrained estimates of any additional parameters added to the translocation model shown in Scheme 8. Meaning the presence of these additional parameters did not improve the quality of the fit as compared to the simple translocation model (data not shown). Therefore all data analysis was conducted using the simple translocation model shown in Scheme 8; the normalized, fitted data is shown in Figure 12. However, since the possibility nevertheless exists that additional processes are occurring

during DNA translocation by ISWI (*e.g.* non-uniform motion[54,63]) the macroscopic parameters obtained from this analysis are the best descriptions of the translocation process[56,111].

As mentioned previously, the shape of the fluorescence time courses associated with DNA translocation (Figure 11A) did not show a strong dependence on DNA substrate length. We believe the most probable explanation for this is that ISWI has very low processivity for translocation along free double-stranded DNA. If a protein has a very low rate of forward translocation (k_t in Scheme 8 and Equation 7) and/or a very high rate of dissociation (k_d in Scheme 8 and Equation 7) the processivity of DNA translocation will be much less than one, indicating that ISWI initially bound far from the fluorophore labeled end of the DNA will have a very small chance of reaching the fluorophore labeled end of the DNA, where its position can be detected. Indeed, only ISWI that binds

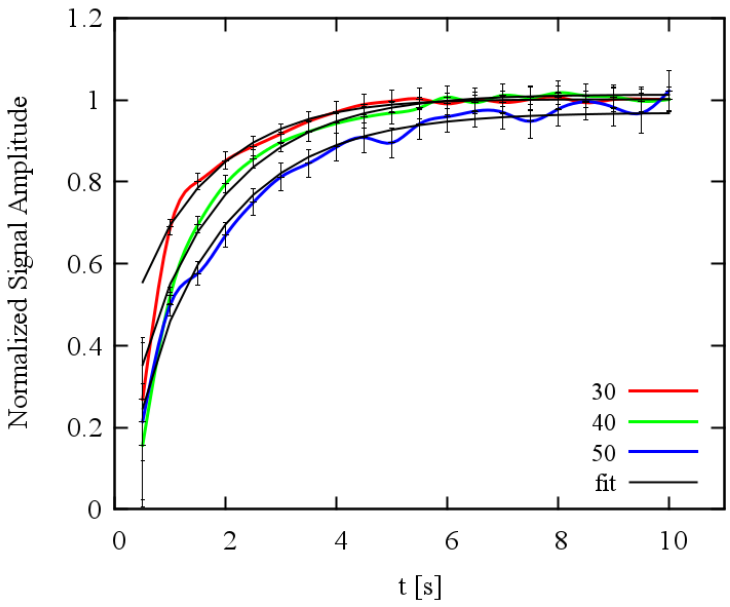


Figure 11 – Monitoring DNA translocation by ISWI translocation using a fluorescence stopped-flow assay. The time courses for DNA translocation by ISWI (colored lines) were smoothed as described in Appendix B; the error bars shown represent standard error. The solid black lines are fits of these time courses according to Scheme 8. The high level of signal homogeneity across the DNA substrate lengths is an indication that these lengths are above the processivity limit for DNA translocation by ISWI.

Parameter	Estimate	Parameter	Estimate
$k_t\ [s^{-1}]$	$1.00 \pm .18$	$\frac{C}{[ATP/ISWI\ /step]}$	0.33 ± 0.06
$k_d\ [s^{-1}]$	$0.60 \pm .03$	P	0.63 ± 0.04
$d\ [bp]$	12.9 ± 0.7	$\frac{P}{1-P}$	1.7 ± 0.3
$m\ [bp]$	4.9 ± 0.4	$\frac{m * P}{1-P}\ [bp]$	8.1 ± 1.5

Table 2 – Results of global fitting across multiple DNA substrate lengths.

in close proximity to the 5' end of the DNA will be able to translocate to the terminal base and thus affect the fluorescence of the fluorophore before dissociation. This means the observed fluorescence time courses will be largely length-independent of any substrate longer than this processivity 'limit'.

6.2.2 – A Highly Efficient Motor

Alongside fluorescence assays to establish translocation parameters, DNA-stimulated ATPase assays provide an additional window into the mechanisms by which ISWI interacts with DNA[60]. Ascertaining the coupling efficiency, as defined in chapter 3, enables the further identification of the specific mechanisms that may occur within the translocation cycle. Since uncovering the efficiency at which ISWI couples ATP binding and hydrolysis to DNA translocation requires a global analysis of the DNA-stimulated ATPase activity across a series of substrate of lengths, it is first necessary to verify the DNA concentration that saturates the reaction. As the results from Figure 13A show, we determined that a DNA concentration twice that of ISWI was sufficient for saturating the ATPase activity for all DNA lengths.

Since these ATPase reactions are associated with multiple rounds of DNA translocation, dissociation, and rebinding, a length-dependence in apparent rate of DNA-stimulated ATPase activity might result from an affiliated increase in the rate of binding due to increased substrate length. A longer DNA strand will have more sites to bind and subsequently a protein will have more sites to bind to, increasing the chances of being bound. To avoid this bias all subsequent ATPase experiments were conducted under conditions of constant base pair concentration of DNA[60]. While we concluded that a 2-fold increase in substrate concentration was saturating, we performed all experiments at a 4-fold increase to ensure that all ISWI bound were preferentially in a 1:1 stoichiometry with the DNA substrate. After fixing this ratio, we measured the DNA-stimulated ATPase activity of ISWI using multiple lengths of DNA (20, 30, and 40 base pairs). We found that there was little length dependence to the ATPase rates listed in Table 3. Previous

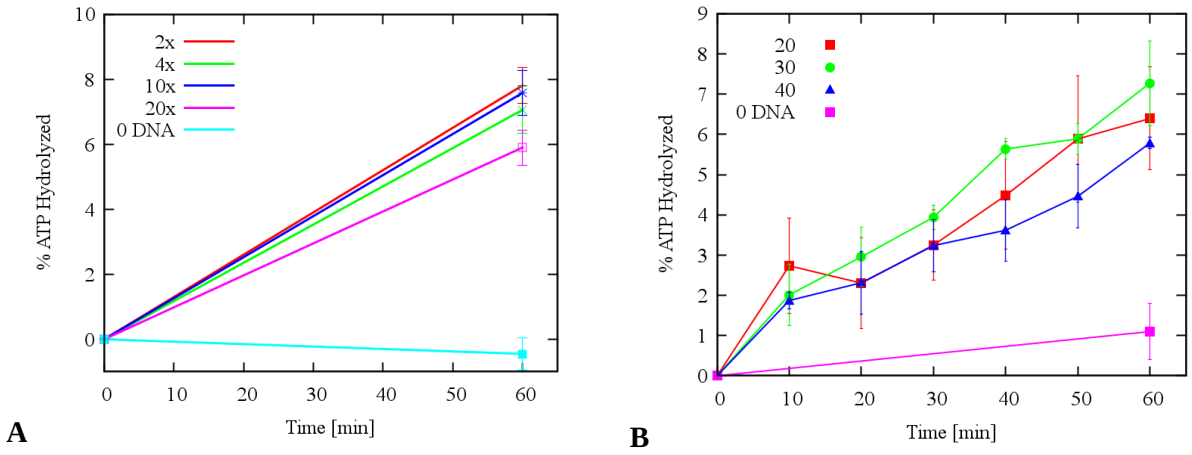


Figure 12 – Determining the DNA concentration required for maximal stimulation of ISWI ATPase. A) ISWI's DNA-stimulated ATPase activity was monitored across a wide range of DNA concentrations, reported here in x-fold excess of the ISWI concentration. DNA availabilities of 2x-10x showed no significant difference and 20x a slight reduction in the ATP hydrolyzed in comparison. The lines connect the normalized start point to the data point at 60 min., and error bars denote standard error of the mean. B) Data points are an average of three assays and the error bars denote the associated standard error. There is little to no length-dependence in the ATPase rate across these lengths of DNA. Lines are provided for visual distinction between data points and are unrelated to analysis. C) The dashed lines are linear fits to the data in Figure 4A. The slope of these fits give us the results listed in Supplementary Table 1. As shown, across a variety of double-stranded DNA substrate lengths, there is no significant difference between the rates of ATP hydrolysis.

derivations have shown that even in simple translocation models, such as Scheme 8, k_{cat} is heavily dependent on the processivity of the underlying motor as seen in Equation 8[68]. If a protein has very low processivity, P, the length (n) dependence of k_{cat} is reduced. Variables c and k_t are the ATP coupling constant and the microscopic translocation rate constant, respectively, as defined in Chapter 4. This provides further evidence that ISWI's inherent processivity is very modest on a free DNA substrate. Given this limitation, the analyses performed on the data were linear fits taking into account the standard error in each point. Figure 13C shows the data with associated fits. We determined that the ISWI complex utilizes 20 ± 4 ATP/ISWI/minute (0.33±.06 ATP/ISWI/s) on average across the shown DNA lengths with length-specific rates being listed in Table 3.

Length of Substrate [bp]	Rate of ADP production [ADP/ISWI/m]	ADP/ISWI/s
20 bp	19 ± 2	0.31 ± 0.07
30 bp	23.4 ± 0.9	0.39 ± 0.07
40 bp	15.7 ± 1.3	0.26 ± 0.05
0 bp	4 ± 2	0.07 ± 0.04

Table 3 - Length Dependence of ISWI

6.2.3 – Simulations Add to Larger Translocation Picture

Very low processivity coupled with apparently high motor efficiency, defined as a low value of C (0.33 ± 0.06 ATP/ISWI/step), initially seem to be opposing results. In an effort to find an explanation for these numbers, we performed multiple simulations (as described in Chapter 4) of protein translocation at varying processivities and DNA substrate lengths to establish what expected translocation time courses should look like, given constraints on the parameters involved. Indeed, the presence of additional steps in the translocation time courses can affect the shape of the fluorescence curve[57,64]. We began by modeling proteins with varying values of processivity, from P = 0.993 to P = 0.75, and determined that over this range the DNA translocation time courses from the differing lengths of DNA substrate became increasingly homogenous, with little difference in the shapes of the curves, as shown in Figure 13. These results demonstrate that our fluorescence data were consistent with ISWI displaying poor processivity on double-stranded DNA, though the simulations could not explain the low apparent ATP coupling efficiency (data not shown).

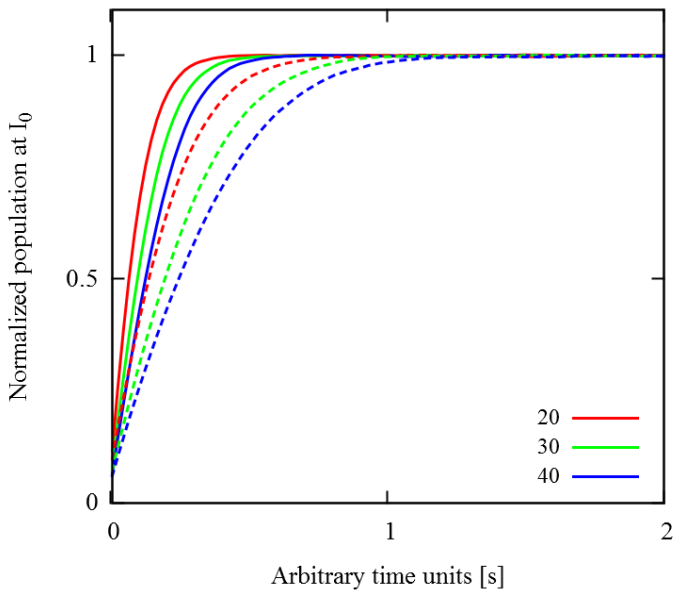
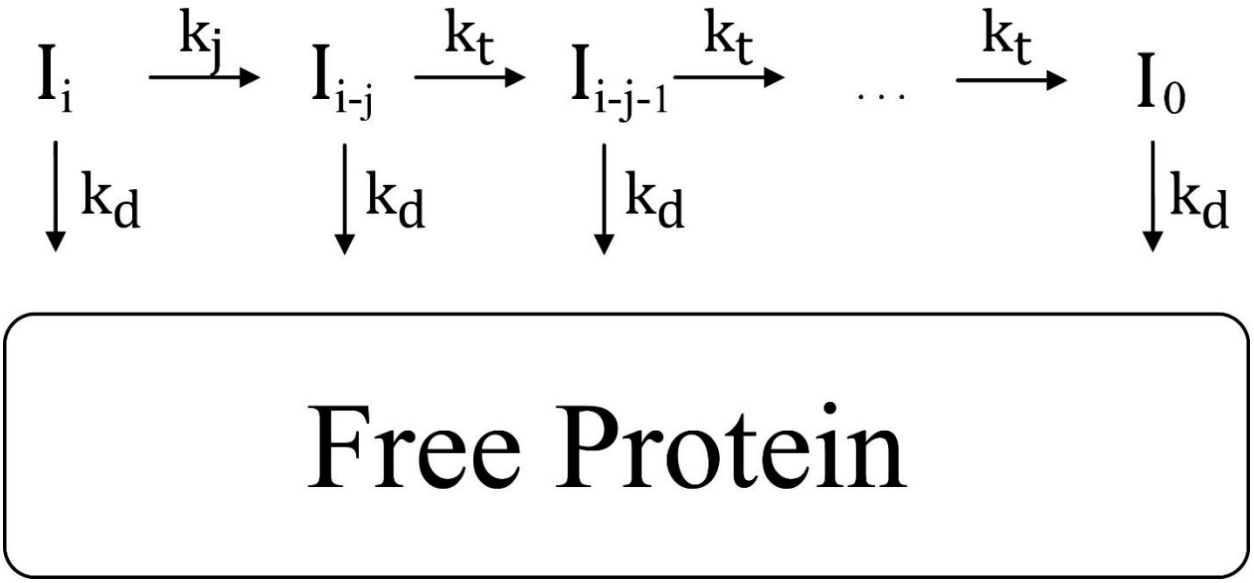


Figure 13 – Simulated time courses of ISWI translocation with jumping. In these simulations the lengths of DNA are 20 bp (red), 30 bp (green), and 40 bp (blue), and the relative probability of jumping, p_j, is 5 (dashed lines) or 15 (non-dashed lines). A higher relative probability of jumping decreases the duration of the burst phase, creating high levels of similarity between the curves.



Scheme 10 – Translocation with jumping parameter. In this scheme, during processive translocation, the protein has a probability, k_j, of jumping forward j binding sites, moving from arbitrary position I_i to I_{i-j}. At this point the protein can continue processive translocation, dissociate, or perform another jump. As noted in Scheme 1b, the rebinding constant k_b has been assumed to be very, very large and is not shown in the scheme.

ISWI's apparent low efficiency for coupling ATP hydrolysis to DNA translocation suggests that the enzyme can experience large displacements (*i.e.*, multiple base pair changes in position) associated with the hydrolysis of a single ATP. To test this possibility, in our next set of simulations we therefore added a mechanism for 'jumping': instead of direct dissociation from the bound DNA or single-step translocation, we added the capability of ISWI to displace itself over significant, directionally-biased distances see Scheme 10. Physically, this might occur if the inherent flexibility of the DNA brings a different section of into proximity with ISWI binding site during a round of ATP hydrolyzation. During the associated period of weakened association with the DNA, ISWI could then bind to the newly proximal DNA and thereby would appear to have 'jumped' along the DNA with only a single ATP hydrolysis. We would expect that such behavior would manifest itself in the analysis of the associated experimental data as a significant increase in coupling efficiency (*i.e.*, decrease in the value of c/m). We performed simulations under the conditions of large jumping distances (10-15bp) and across a range of jumping probabilities ($0 < p_j < 15$) and confirmed that a suppression of coupling efficiency, c/m, across all processivities does occur when jumping is possible (see Figure 14)[93].

6.3 – Discussion

Several previously published results show that ISWI is capable of both ATP-dependent chromatin-remodeling[80,112] and DNA translocation[80,113]. The results presented here agree with these findings and further suggest that ISWI displays poor processivity for translocation along free DNA. This may result from the allosteric effect of nucleotide binding on DNA binding by ISWI. Specifically, that the binding of ATP reduces the DNA binding affinity of ISWI[75]. However in the presence of nucleosome core particles or structures of multiple nucleosomes, ISWI has been shown to have a much higher binding affinity and that nucleotide binding causes no allosteric weakening of the affinity of nucleosome binding[75,107]. These results indicate that the HSS domain shares additional histone recognition responsibilities[43]. Notably, despite being a poorly processive DNA translocase on double-stranded DNA, ISWI remains remarkably efficient, requiring little energy from ATP hydrolysis for translocation. Our results are not consistent with the proposed hypothesis that DNA translocation is the energetically rate-limiting process during nucleosome repositioning[114].

6.3.1 - ISWI Translocation and Repositioning

Our results show that ISWI is a slow and non-processive motor for double-stranded DNA translocation (see Table 1). In our experiments, the base pair concentration is the same for each DNA substrate and due to the binding site size for ISWI being assumed constant across substrates, the concentration of DNA binding sites is also the same. Consequently, the rate at which ISWI rebinds free DNA following dissociation is identical across all lengths of DNA. This is important because if the processivity of DNA translocation is small, ISWI's steady-state ATPase rate for DNA-stimulated translocation should be equal across all substrates due to the majority of ISWI undergoing translocation will dissociate before reaching the end of the DNA. This would make each substrate above a minimum threshold length appear infinitely long and eliminate any length-dependence in the translocation parameters.

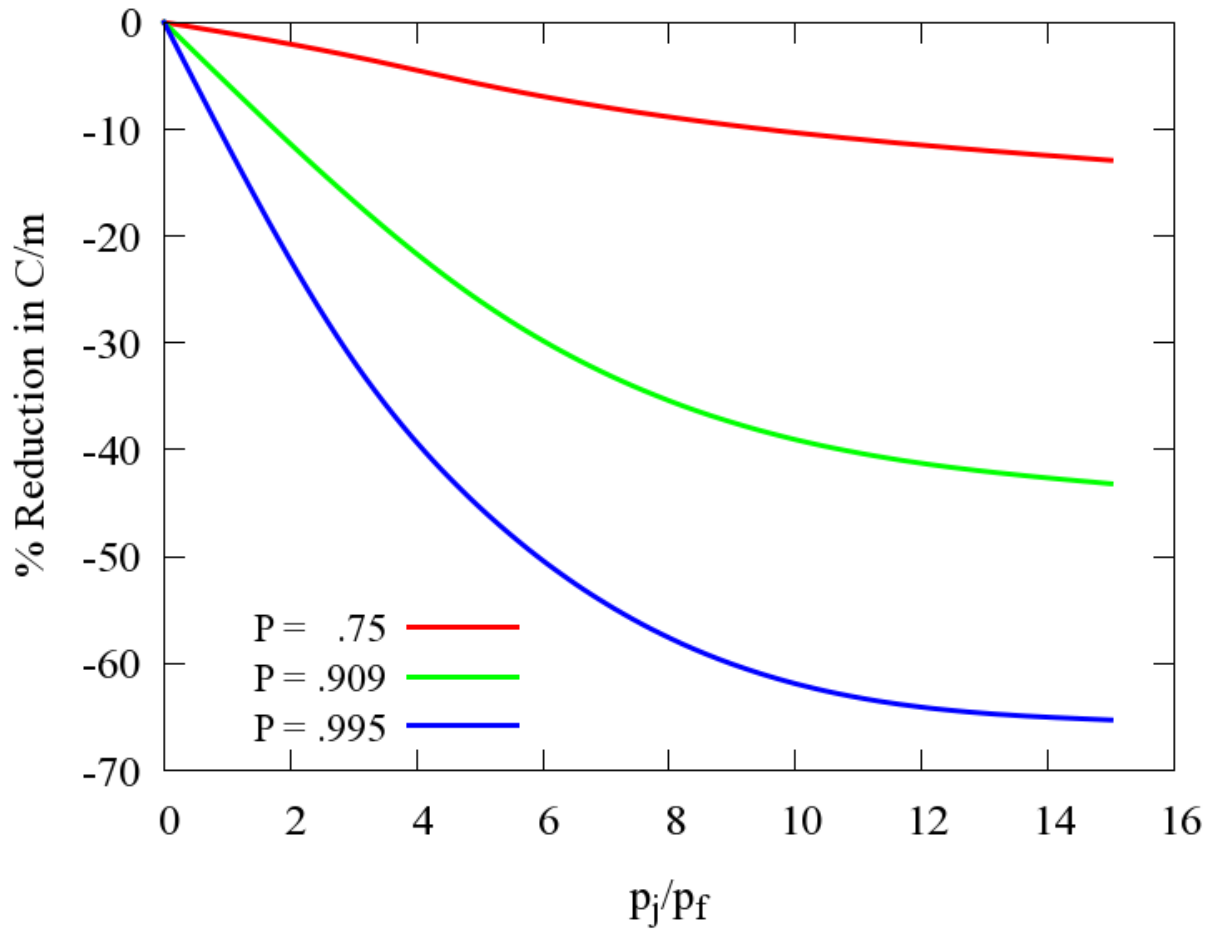


Figure 14 – Suppression of ATP coupling efficiency with increased relative probability of jumping. The curves were generated by producing simulations of the ATP curves at the listed processivity of P = 0.993,0 .909, and 0.75, using substrate lengths of 20, 30, and 40 base pairs, and a randomly chosen jump distance between 10 and 15 base pairs. The ATP curves were then globally analyzed as described in Chapter 4, at p_j values of 0,1,3,5, and 15 to determine the associated coupling efficiency. As shown above, increasing the probability of jumping suppresses the value for c/m; furthermore, the extent of the suppression is dependent upon the processivity of translocation. Enigmatically, the simulations indicate that the suppression of coupling efficiency is more length-dependent at low processivities than it is at high processivities.

Indeed, other remodelers in the ISWI family display binding with DNA that is short-lived and frequent, experiencing several rounds of substrate dissociation and rebinding before a successful translocation event. takes place[94]. If the ISWI motor domain detaches from the bound substrate and simply rebinds closer to the 5' base, this could be seen as a perfectly efficient translocation event requiring no energy input and consuming no ATP as a result. This would deflate the apparent ATP/step value, causing the ISWI motor to look falsely efficient in its motion.

Given processivity estimates that are as low as we have determined, this leads to the conclusion that while the ISWI molecular motor is capable of translocation along the DNA, it likely can form additional contacts with histone complexes to increase its affinity for nucleosome binding. This is also consistent with recent findings on the binding affinity of the ISWI motor with nucleosome core particles[75]. As a chromatin remodeling enzyme, the primary function of this protein is binding with the nucleosome to perform remodeling reactions. As such, if the principal method of contact to a nucleosome core particle is attachment to the histones, strong bonds with the DNA substrate may not be necessary for stabilizing the interaction between ISWI and the nucleosome.

6.3.2 - ISWI as an ATP-Efficient Motor

By combining the results from the ATPase reactions with the translocation kinetic parameters, we established that the ISWI consumes 0.067 ± 0.018 ATP/bp translocated or 0.33 ± .09 ATP/step, which makes this a very efficient process. When compared to ISWI ADP production rates during nucleosome repositioning (~900 ATP utilized per 12 bp shift in the nucleosomal position)[115], our results point to the fact that the energy requirement for DNA translocation is significantly lower and drastically more efficient. This makes sense from an energetic standpoint since far less energy would be required to walk along the back bone of a segment than to separate at the multitude of water-mediated bonds across the surface of the entire nucleosome, including 14 sites where nucleosomal arginine intercalates into the DNA minor groove[31]. While much of the energy of these bonds may go to distorting the surrounding DNA, completely disrupting these interfaces would require significant additional energy input by ISWI and indicates that the vast majority of the energy released from ADP production would be required to break the bonds and move the nucleosome, thus this energy would not be associated with DNA translocation. While it has been suggested previously that DNA interaction and translocation is the rate limiting process for nucleosome repositioning[114], energetically efficient DNA translocation by ISWI suggests that this is not the case. Given that ISWI capable of using less than 1 ATP per translocated base pair, it is very energetically efficient in its motion along free DNA. Additionally, it exhibits poor free DNA binding due to domain interference[116].

This implies that perhaps when the ISWI substrate is in complex with a nucleosome core particle, the energetic interaction between the DNA and histone core is the limiting factor in nucleosome repositioning, not an interaction between ISWI and the DNA.

6.3.3 - Potential Mechanistic Additions

Simulations that we have performed indicate that low processivity values lead to homogenous DNA translocation time courses, a common trait of the data that we have collected. This evidence together with the very low values of ATP coupling efficiency led us to test a secondary method of movement that included a 'jumping' mechanism. During an ATPase cycle the ISWI shares only a tenuous bond with double-stranded DNA[75]. In this system, this potentially causes rapid dissociation from the substrate and subsequent re-association at a binding site closer to the 5' end of the DNA strand. The end result is an apparent large translational movement of ISWI along the DNA which occurs with little to no associated ADP production. Our computer simulations agree with the results that we observe, namely that high levels of signal similarity in the modeled time course shapes occur due to both low processivity as well as an additional jumping mechanism. Furthermore, the low ATP coupling efficiency can also be explained by this jumping mechanism. Taken together, it would appear that the translocational mechanism on double-stranded DNA is not a controlling factor in the nucleosome repositioning activity of ISWI. Rather, other factors such as domain interactions or energetic limitations would present a much larger obstacle to the repositioning.

Chapter 7 – Conclusions

Chromatin remodeling is at the heart of proper expression of our genome. Without the proper function of these molecular motors, cellular life is unable to accurately carry out internal directives involved in reproduction and maintenance. These families of proteins have been copied and propagated for millions of years and we still understand little about their microscopic mechanisms. The research presented herein used existing methods of experimentation alongside novel data analysis techniques and simulation to draw new conclusions about the microscopic processes that these motors incorporate into their function. Through the use of fluorescent stopped-flow analysis, ATP utilization assays, simulation, and global data analysis, their operations have been made more transparent.

Both domains studied are by nature ATP-dependent DNA translocases. They require and utilize the free-energy stored in the phosphate bonds of ATP to perform work and move along the backbone of double-stranded DNA. In the absence of this energy source they perform no processive translocation, as shown in Figures 8 and 11. By performing ATP-utilization assays and the associated global data fitting, we determined the characteristics that RSCt and ISWI had in common and what makes each of the two remodelers unique.

RSCt was shown to be a slow, inefficient DNA translocase. It displayed very strong length-dependence to its ATP consumption, as shown in Figure 7, but a relatively slow translocation rate (2.3 $\pm$

0.4 bp/s) and low processivity ($P = 0.92 \pm 0.01$) when compared to similar helicase proteins[63,64,68,84,91,92]. In contrast, ISWI was determined to have very low length-dependence to its rate of ATP utilization, as shown in Figure 13. Using double-stranded DNA lengths of 20 to 80 base pairs, it was shown that RSCt's values of k_{cat} had a hyperbolic relationship to length, pointing to a saturating length of over 100 base pair. Interestingly, this was not a trait shared by ISWI. In Figure 13, RSCt showed strong length dependence in its ATP usage, whereas in Figure 6 ISWI showed almost none. The experiments conducted on both RSCt and ISWI reveal a similar conclusion: the saturation level of RSCt was >10 μM binding site concentration whereas the saturating concentration for ISWI was <1 μM. The processivity of ISWI (0.63 ± 0.04) is significantly lower than that of RSCt leading to a greatly reduced length dependence of k_{cat} as noted in Chapter 6.2.

The results from those experiments also determined ISWI uses an order of magnitude less ATP per protein per second than RSCt (0.33 ± 0.06 vs 4-8) on a double-stranded substrate. This is an interesting result to note that ISWI is such an efficient motor, seemingly requiring little energy input to perform its function. However, ISWI has self-interfering domains which are fully opened to ATP consumption only when in the presence of an NCP. The NTR of the protein contains the Auto N motif, a basic H4-tail mimic. When unbound from an NCP, this region of the protein cooperates with the H4-interacting patch on Lobe 2 of the ATPase domain (Figure 10). Additionally, the NegC domain found at the end of Lobe 2 interferes with both lobes by binding in the cleft between the two[117]. When in the presence of an NCP however, the HSS domain contacts the opposite side of the NCP[113], drawing NegC out of the cleft and removing one antagonist to processive translocation. In this regard, ISWI has several roadblocks to processive translocation when given only DNA upon which to operate. Unfortunately, there is little structural information for RSC, or any minimalist construction thereof, and what exists is generally low resolution or not in specific context with a nucleosome or DNA[38,118]. This limited information makes it difficult to draw the same structural arguments for explanation of translocation that is possible with the ISWI remodeler.

Through multiple length-dependent analyses of the ATP utilization rates of RSCt, it was shown to utilize 3.77 ± 0.02 ATP per step, or 3.00 ± 0.04 ATP per basepair translocated. Additionally, during the associated initiative step prior to processive translocation, the complex utilizes 11.69 ± 0.11 ATP molecules. This is in distinct contrast to the ISWI catalytic domain on the same substrate: ISWI consumes a mere 0.33 ± 0.09 ATP per forward step during translocation. It could be considered that ISWI is not, in fact, performing processive translocation to achieve such efficiencies. Given that there are several hindrances to free DNA binding[116], it could well be that the ISWI molecule is more often than not slowly processing ATP while performing short 'hops' along the DNA: micro-instances of unbinding and rebinding in extremely close intervals along the same DNA strand. If the binding affinity fluctuates wildly at very short time intervals such as during ATP binding and hydrolysis, then perhaps the energy efficient manner of ISWI is achieved through a faux-step process instead of direct translocation along the backbone. Evidence to support this idea was put forth showing that in the presence of double-stranded DNA only, ISWI displays a binding affinity of 18 ± 2 nM. When in the presence of an unhydrolyzable ATP analog, ISWI binding affinity to DNA jumped to 390 ± 70 μM[75]. Additionally, other remodelers in the ISWI family display binding with DNA that is short-lived and frequent, experiencing several rounds of substrate dissociation and rebinding before a successful translocation event takes place[94]. Along these lines, DNA binding of RSC in the presence of ADP was shown to be 370 ± 14 nM, significantly tighter than ISWI[81]., This points to RSC possibly requiring multiple rounds of energetic input from hydrolysis before the interaction is disrupted enough for the motor to progress.

Another interesting avenue of research indicated RSCt showed such a strong association to DNA that it was able to displace the streptavidin-biotin linkage whereas ISWI is known to self-interfere with non-nucleosomal substrates. Similarly, the full RSC complex appears to exert large molecular forces in optical tweezer experiments performed by Sirinakis *et al.*[66], a determination which agrees with the presented streptavidin displacement experiments showing the RSCt complex as able to interrupt one of the strongest non-covalent chemical bonds in nature.

It was also found that these two remodelers do not perform their function in the same manner. In both cases, the translocation parameters were determined through fluorescence stopped-flow DNA translocation assays and associated modeling coupled with global analysis. The data collected across several DNA lengths and in the absence of a protein trap was analyzed using NLLS techniques and each set was fit to various translocation models. The production of these models was covered in Chapter 3 and the specific models used were detailed in Chapters 5 and 6. Each data set was initially fit with a simple translocation model, shown in Scheme 1 and Equation 1. From here, additional mechanistic modifications were made to the simplest representation in an effort to minimize the sum of the square residuals (SSR) of the fit. While this technique is not an explicit description of microscopic parameters of motion, it has been used broadly to provide direction, support, and comparison to other, more specific experiments[55,60,63].

Through this technique it was determined that RSCt is best described by having an initiative process before beginning processive translocation. The rate constant determined for this initial period included all steps that may be occurring prior to movement, such as rebinding, ATP binding, physical changes in structure, *etc*. but cannot be a comprehensive, specific description of the physical happenings. It is worth noting that the smaller RSCt complex presented similar step characteristics to the full RSC complex even when comparing across ensemble and single-molecule experimental results[66,90]. In atomic force microscopy experiments, Lia *et al.* presented data pointing to a burst-like model of DNA translocation[90], further reinforcing the idea that perhaps there is a short initial period in which the RSCt motor undergoes before translocating forward. Though there were additional experiments detailing pauses between DNA distortion events[90], it remains a plausible description of what happens at a molecular level when RSCt interacts with the substrate. ISWI, on the other hand, proved more difficult to describe when using the same method. While there is AFM and optical tweezer evidence that ISWI also undergoes looping when attached to DNA and circumstantially suggests that it does not rapidly dissociate from the end of the DNA[119], any model more complex than simple translocation with rebinding were unable to return a better fit parameter to the data. Unfortunately, all the data collected were relatively noisy and too similar across substrate lengths to

provide additional detail. This led to further simulation work which lent evidence to the idea that a 'jumping' event is taking place, causing the high levels of signal homogeneity and very efficient motion. With a parameter included in the simulation for a jumping mechanism, the suppression of c/m can be explained. As noted in Chapter 6, this sort of mechanism is similar to other ISWI family members.

The data analysis techniques used in this research were novel continuations of existing equations provided largely through previous studies. These techniques use the Laplace transform to take in sets of differential equations and turn them into first-order algebraic equations that can be solved through array row reduction. However, the result remains in Laplace space and requires expensive, proprietary numerical packages to solve. In an effort to remedy this bottleneck and make this technique more open and available to additional researchers, work was begun on straightforward numerical solutions to the original differential equations. Using Mathematica's built-in DSolve function, it is more straightforward to input an array of differential equations and return an output of exponential equations that are easily solved numerically by any number of open source software packages. Unfortunately, there was work yet to be done on the production and streamlining of this technique and it could not be included in this final document. Future work on this technique should be pursued to allow interested parties access to this powerful method.

While this research uncovered novel and stimulating understandings of these two specific chromatin remodelers, it is only the beginning of exposing the mechanisms of chromatin remodelers as a whole. The conclusions drawn herein are not the complete picture of either motor, only a snapshot into specific frameworks in which they operate. They function only in conjuncture with other bodies of work, both supporting and being supported by these additional efforts. Chromatin remodeling remains one of the key topics to unlocking the heart of our genetic and epigenetic expressions. Lifetimes of further work must be done to fully unravel the mysteries of this family.

References

1. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–51 (2001).

2. Alberts, B. *et al. Molecular Biology Of The Cell*. (Garland Science, 2008). at <http://medcontent.metapress.com/index/A65RM03P4874243N.pdf>

3. Klevan, L., Armitage, I. M. & Crothers, D. M. 31P NMR studies of the solution structure and dynamics of nucleosomes and DNA. *Nucleic Acids Res.* **6,** 1607–16 (1979).

4. Flaus, A. & Richmond, T. J. Base-pair resolution mapping of nucleosomes in vitro. *Methods Mol. Biol.* **119,** 45–60 (1999).

5. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389,** 251–60 (1997).

6. Luger, K. Dynamic nucleosomes. *Chromosom. Res. An Int. J. Mol. Supramol. Evol. Asp. Chromosom. Biol.* **14,** 5–16 (2006).

7. Fan, H.-Y., Narlikar, G. J. & Kingston, R. E. Noncovalent modification of chromatin: different remodeled products with different ATPase domains. *Cold Spring Harb. Symp. Quant. Biol.* **69,** 183–92 (2004).

8. Slawson, C., Housley, M. P. & Hart, G. W. O-GlcNAc cycling: how a single sugar post-translational modification is changing the way we think about signaling networks. *J. Cell. Biochem.* **97,** 71–83 (2006).

9. Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78,** 273–304 (2009).

10. Lorch, Y., Maier-Davis, B. & Kornberg, R. D. Mechanism of chromatin remodeling. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 3458–62 (2010).

11. Lorch, Y., Zhang, M. & Kornberg, R. D. Histone Octamer Transfer by a Chromatin-Remodeling Complex. *Cell* **96,** 389–392 (1999).

12. Toto, M., D'Angelo, G. & Corona, D. F. V. Regulation of ISWI chromatin remodelling activity. *Chromosoma* **123,** 91–102 (2014).

13. Mellor, J. Imitation switch complexes. *Ernst Schering Res. Found. Workshop* 61–87 (2006). at <http://www.ncbi.nlm.nih.gov/pubmed/16568949>

14. Versteege, I. *et al.* Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* **394,** 203–6 (1998).

15. Sévenet, N. *et al.* Spectrum of hSNF5/INI1 somatic mutations in human cancer and genotype-phenotype correlations. *Hum. Mol. Genet.* **8,** 2359–68 (1999).

16. Wong, A. K. *et al.* BRG1, a component of the SWI-SNF complex, is mutated in multiple human tumor cell lines. *Cancer Res.* **60,** 6171–7 (2000).

17. Slawson, C. & Hart, G. O-GlcNAc signalling: implications for cancer cell biology. *Nat. Rev. Cancer* **11,** 678–684 (2011).

18. Wolffe, A. P. Chromatin remodeling: why it is important in cancer. *Oncogene* **20,** 2988–2990 (2001).

19. Bochar, D. A. *et al.* BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. *Cell* **102,** 257–265 (2000).

20. Bao, Y. & Shen, X. SnapShot: chromatin remodeling complexes. *Cell* **129,** 632 (2007).

21. Dahm, R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* **122,** 565–81 (2008).

22. Levene, P. The structure of yeast nucleic acid. *J Biol Chem* **40,** 415–24 (1919).

23. Chargaff, E. Structure and function of nucleic acids as cell constituents. *Fed. Proc.* **10,** 654–9 (1951).

24. WATSON, J. D. & CRICK, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171,** 737–8 (1953).

25. Luger, K. & Richmond, T. J. DNA binding within the nucleosome core. *Curr. Opin. Struct. Biol.* **8,** 33–40 (1998).

26. Bhasin, M., Reinherz, E. L. & Reche, P. A. Recognition and classification of histones using support vector machine. *J. Comput. Biol.* **13,** 102–12

27. McGhee, J. D. & Felsenfeld, G. Nucleosome structure. *Annu. Rev. Biochem.* **49,** 1115–56 (1980).

28. Wang, J., Hogan, M. & Austin, R. H. DNA motions in the nucleosome core particle. *Proc. Natl. Acad. Sci. U. S. A.* **79,** 5896–900 (1982).

29. Ehrenhofer-Murray, A. E. Chromatin dynamics at DNA replication, transcription and repair. *Eur. J. Biochem.* **271,** 2335–49 (2004).

30. Strachan, T. & Read, A. in *Human Molecular Genetics* (Wiley-Liss, 1999). at <http://www.ncbi.nlm.nih.gov/books/NBK7587/>

31. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J. Mol. Biol.* **319,** 1097–113 (2002).

32. Cosgrove, M. S., Boeke, J. D. & Wolberger, C. Regulated nucleosome mobility and the histone code. *Nat. Struct. Mol. Biol.* **11,** 1037–43 (2004).

33. Vignali, M., Hassan, A. H., Neely, K. E. & Workman, J. L. ATP-dependent chromatin-remodeling complexes. *Mol. Cell. Biol.* **20,** 1899–910 (2000).

34. Fairman-Williams, M. E., Guenther, U.-P. & Jankowsky, E. SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* **20,** 313–24 (2010).

35. Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodeling through directional DNA translocation from an internal nucleosomal site. *Nat. Struct. Mol. Biol.* **12,** 747–55 (2005).

36. Kagalwala, M. N., Glaus, B. J., Dang, W., Zofall, M. & Bartholomew, B. Topography of the ISW2-nucleosome complex: insights into nucleosome spacing and chromatin remodeling. *EMBO J.* **23,** 2092–104 (2004).

37. Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodelling: the industrial revolution of DNA around histones. *Nat. Rev. Mol. Cell Biol.* **7,** 437–47 (2006).

38. Chaban, Y. *et al.* Structure of a RSC-nucleosome complex and insights into chromatin remodeling. *Nat. Struct. Mol. Biol.* **15,** 1272–1277 (2008).

39. Tosi, A. *et al.* Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell* **154,** 1207–19 (2013).

40. Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467,** 562–6 (2010).

41. McDonald, S. M., Close, D., Xin, H., Formosa, T. & Hill, C. P. Structure and biological importance of the Spn1-Spt6 interaction, and its regulatory role in nucleosome binding. *Mol. Cell* **40,** 725–35 (2010).

42. Ludwigsen, J., Klinker, H. & Mueller-Planitz, F. No need for a power stroke in ISWI-mediated nucleosome sliding. *EMBO Rep.* **14,** 1092–7 (2013).

43. Mueller-Planitz, F., Klinker, H. & Becker, P. B. Nucleosome sliding mechanisms: new twists in a looped history. *Nat. Struct. Mol. Biol.* **20,** 1026–32 (2013).

44. Bowman, G. D. Mechanisms of ATP-dependent nucleosome sliding. *Curr. Opin. Struct. Biol.* **20,** 73–81 (2010).

45. Flaus, A. & Owen-Hughes, T. Mechanisms for ATP-dependent chromatin remodelling: the means to the end. *FEBS J.* **278,** 3579–95 (2011).

46. Deindl, S. *et al.* ISWI remodelers slide nucleosomes with coordinated multi-base-pair entry steps and single-base-pair exit steps. *Cell* **152,** 442–52 (2013).

47. Hota, S. K. *et al.* Nucleosome mobilization by ISW2 requires the concerted action of the ATPase and SLIDE domains. *Nat. Struct. Mol. Biol.* **20,** 222–9 (2013).

48. Strohner, R. *et al.* A 'loop recapture' mechanism for ACF-dependent nucleosome remodeling. *Nat Struct Mol Biol* **12,** 683–690 (2005).

49. Grand, F. *et al.* Frequent deletion of hSNF5/INI1, a component of the SWI/SNF complex, in chronic myeloid leukemia. *Cancer Res.* **59,** 3870–4 (1999).

50. Caldwell, S. A. *et al.* Nutrient sensor O-GlcNAc transferase regulates breast cancer tumorigenesis through targeting of the oncogenic transcription factor FoxM1. *Oncogene* **29,** 2831–42 (2010).

51. Slawson, C., Pidala, J. & Potter, R. Increased N-acetyl-beta-glucosaminidase activity in primary breast carcinomas corresponds to a decrease in N-acetylglucosamine containing proteins. *Biochim. Biophys. Acta* **1537,** 147–57 (2001).

52. Krzeslak, A., Pomorski, L. & Lipinska, A. Elevation of nucleocytoplasmic beta-N-acetylglucosaminidase (O-GlcNAcase) activity in thyroid cancers. *Int. J. Mol. Med.* **25,** 643–8 (2010).

53. Ellis, L., Atadja, P. W. & Johnstone, R. W. Epigenetics in cancer: targeting chromatin modifications. *Mol. Cancer Ther.* **8,** 1409–20 (2009).

54. Fischer, C. J., Wooten, L., Tomko, E. & Lohman, T. Kinetics of motor protein translocation on Single-Stranded DNA. *Helicases* **587,** 45–56 (2010).

55. Lucius, A. L., Maluf, N. K., Fischer, C. J. & Lohman, T. M. General methods for analysis of sequential 'n-step' kinetic mechanisms: application to single turnover kinetics of helicase-catalyzed DNA unwinding. *Biophys. J.* **85,** 2224–2239 (2003).

56. Tomko, E. J., Fischer, C. J. & Lohman, T. M. Ensemble methods for monitoring enzyme translocation along single stranded nucleic acids. *Methods* **51,** 269–276 (2010).

57. Eastlund, A., Malik, S. S. & Fischer, C. J. Kinetic mechanism of DNA translocation by the RSC molecular motor. *Arch. Biochem. Biophys.* **532,** 73–83 (2013).

58. Grüne, T. *et al.* Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol. Cell* **12,** 449–60 (2003).

59. Dillingham, M. S., Wigley, D. B. & Webb, M. R. Direct measurement of single-stranded DNA translocation by PcrA helicase using the fluorescent base analogue 2-aminopurine. *Biochemistry* **41,** 643–51 (2002).

60. Tomko, E. J., Fischer, C. J., Niedziela-Majka, A. & Lohman, T. M. A nonuniform stepping mechanism for E. coli UvrD monomer translocation along single-stranded DNA. *Mol. Cell* **26,** 335–347 (2007).

61. Williams, D. J. & Hall, K. B. Monte Carlo applications to thermal and chemical denaturation experiments of nucleic acids and proteins. *Methods Enzymol.* **321,** 330–52 (2000).

62. Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes Dev.* **16,** 2120–2134 (2002).

63. Fischer, C. J. & Lohman, T. M. ATP-dependent translocation of proteins along single-stranded DNA: models and methods of analysis of pre-steady state kinetics. *J. Mol. Biol.* **344,** 1265–1286 (2004).

64. Fischer, C. J., Maluf, N. & Lohman, T. Mechanism of ATP-dependent translocation of E. coli UvrD monomers along single-stranded DNA. *J. Mol. Biol.* **344,** 1287–1309 (2004).

65. Fischer, C. J., Saha, A. & Cairns, B. R. Kinetic model for the ATP-dependent translocation of Saccharomyces cerevisiae RSC along double-stranded DNA. *Biochemistry* **46,** 12416–12426 (2007).

66. Sirinakis, G. *et al.* The RSC chromatin remodelling ATPase translocates DNA with high force and small step size. *EMBO J.* (2011). doi:10.1038/emboj.2011.141

67. Fischer, C. J., Yamada, K. & Fitzgerald, D. J. Kinetic mechanism for single-stranded DNA binding and translocation by Saccharomyces cerevisiae Isw2. *Biochemistry* **48,** 2960–8 (2009).

68. Khaki, A. R. *et al.* The macroscopic rate of nucleic acid translocation by hepatitis C virus helicase NS3h is dependent on both sugar and base moieties. *J. Mol. Biol.* **400,** 354–78 (2010).

69. Lohman, T. M. & Bjornson, K. P. Mechanisms of helicase-catalyzed DNA unwinding. *Annu. Rev. Biochem.* **65,** 169–214 (1996).

70. Von Hippel, P. H. & Delagoutte, E. A general model for nucleic acid helicases and their 'coupling' within macromolecular machines. *Cell* **104,** 177–90 (2001).

71. McClure, W. R. & Chow, Y. The kinetics and processivity of nucleic acid polymerases. *Methods Enzymol.* **64,** 277–97 (1980).

72. Whitehouse, I., Stockdale, C., Flaus, A., Szczelkun, M. D. & Owen-Hughes, T. Evidence for DNA translocation by the ISWI chromatin-remodeling enzyme. *Mol. Cell. Biol.* **23,** 1935–1945 (2003).

73. Gowers, D. M. & Fox, K. R. Towards mixed sequence recognition by triple helix formation. *Nucleic Acids Res.* **27,** 1569–77 (1999).

74. Firman, K. & Szczelkun, M. D. Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement. *EMBO J.* **19,** 2094–102 (2000).

75. Al-Ani, G. *et al.* Quantitative determination of binding of ISWI to nucleosomes and DNA shows allosteric regulation of DNA binding by nucleotides. *Biochemistry* **53,** 4334–4345 (2014).

76. Cairns, B. R. *et al.* RSC, an Essential, Abundant Chromatin-Remodeling Complex. *Cell* **87,** 1249–1260 (1996).

77. Lorch, Y., Zhang, M. & Kornberg, R. D. RSC unravels the nucleosome. *Mol. Cell* **7,** 89–95 (2001).

78. Lorch, Y., Cairns, B. R., Zhang, M. & Kornberg, R. D. Activated RSC–Nucleosome Complex and Persistently Altered Form of the Nucleosome. *Cell* **94,** 29–34 (1998).

79. Shukla, M. S. *et al.* Remosomes: RSC generated non-mobilized particles with approximately 180 bp DNA loosely associated with the histone octamer. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 1936–1941 (2010).

80. Zofall, M., Persinger, J., Kassabov, S. R. & Bartholomew, B. Chromatin remodeling by ISW2 and SWI/SNF requires DNA translocation inside the nucleosome. *Nat. Struct. Mol. Biol.* **13,** 339–46 (2006).

81. Malik, S., Rich, E., Viswanathan, R., Cairns, B. R. & Fischer, C. J. Allosteric interactions of DNA and nucleotides with S. cerevisiae RSC. *Biochemistry* **50,** 7881–90 (2011).

82. Cairns, B. R., Erdjument-Bromage, H., Tempst, P., Winston, F. & Kornberg, R. D. Two actin-related proteins are shared functional components of the chromatin-remodeling complexes RSC and SWI/SNF. *Mol. Cell* **2,** 639–51 (1998).

83. Du, J., Nasir, I., Benton, B. K., Kladde, M. P. & Laurent, B. C. Sth1p, a Saccharomyces cerevisiae Snf2p/Swi2p homolog, is an essential ATPase in RSC and differs from Snf/Swi in its interactions with histones and chromatin-associated proteins. *Genetics* **150,** 987–1005 (1998).

84. Matlock, D. L. *et al.* Investigation of translocation, DNA unwinding, and protein displacement by NS3h, the helicase domain from the hepatitis C virus helicase. *Biochemistry* **49,** 2097–2109 (2010).

85. Morris, P. D. *et al.* Hepatitis C virus NS3 and simian virus 40 T antigen helicases displace streptavidin from 5'-biotinylated oligonucleotides but not from 3'-biotinylated oligonucleotides: evidence for directional bias in translocation on single-stranded DNA. *Biochemistry* **41,** 2372–8 (2002).

86. Zhang, W., Bond, J. P., Anderson, C. F., Lohman, T. M. & Record, M. T. Large electrostatic differences in the binding thermodynamics of a cationic peptide to oligomeric and polymeric DNA. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 2511–6 (1996).

87. Zhang, W. *et al.* The importance of coulombic end effects: experimental characterization of the effects of oligonucleotide flanking charges on the strength and salt dependence of oligocation (L8+) binding to single-stranded DNA oligomers. *Biophys. J.* **76,** 1008–17 (1999).

88. Zhang, Y. *et al.* DNA translocation and loop formation mechanism of chromatin remodeling by SWI/SNF and RSC. *Mol. Cell* **24,** 559–568 (2006).

89. Green, N. M. Avidin. *Adv. Protein Chem.* **29,** 85–133 (1975).

90. Lia, G. *et al.* Direct observation of DNA distortion by the RSC complex. *Mol. Cell* **21,** 417–25 (2006).

91. Brendza, K. M. *et al.* Autoinhibition of Escherichia coli Rep monomer helicase activity by its 2B subdomain. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10076–81 (2005).

92. Kim, D.-E., Narayan, M. & Patel, S. S. T7 DNA helicase: a molecular motor that processively and unidirectionally translocates along single-stranded DNA. *J. Mol. Biol.* **321,** 807–19 (2002).

93. Fischer, C. J., Wooten, L., Tomko, E. J., Eastlund, A. & Lohman, T. M. Effects of Non-uniform motion, Static and Dynamic Disorder on the Kinetic Parameters obtained from Ensemble Studies of Nucleic Acid Translocases and Helicases.

94. He, X., Fan, H.-Y., Narlikar, G. J. & Kingston, R. E. Human ACF1 alters the remodeling strategy of SNF2h. *J. Biol. Chem.* **281,** 28636–47 (2006).

95. Levin, M. K., Gurjar, M. M. & Patel, S. S. ATP binding modulates the nucleic acid affinity of hepatitis C virus helicase. *J. Biol. Chem.* **278,** 23311–6 (2003).

96. Andreeva, I. E., Roychowdhury, A., Szymanski, M. R., Jezewska, M. J. & Bujalowski, W. Mechanisms of interactions of the nucleotide cofactor with the RepA protein of plasmid RSF1010. Binding dynamics studied using the fluorescence stopped-flow method. *Biochemistry* **48,** 10620–36 (2009).

97. Wong, I. & Lohman, T. M. Allosteric effects of nucleotide cofactors on Escherichia coli Rep helicase-DNA binding. *Science* **256,** 350–5 (1992).

98. Dou, S.-X., Wang, P.-Y., Xu, H. Q. & Xi, X. G. The DNA binding properties of the Escherichia coli RecQ helicase. *J. Biol. Chem.* **279,** 6354–63 (2004).

99. Hsieh, J., Moore, K. J. & Lohman, T. M. A two-site kinetic mechanism for ATP binding and hydrolysis by E. coli Rep helicase dimer bound to a single-stranded oligodeoxynucleotide. *J. Mol. Biol.* **288,** 255–74 (1999).

100. Brower-Toland, B. D. *et al.* Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1960–5 (2002).

101. Smith, C. L. & Peterson, C. L. A conserved Swi2/Snf2 ATPase motif couples ATP hydrolysis to chromatin remodeling. *Mol. Cell. Biol.* **25,** 5880–92 (2005).

102. Dürr, H., Körner, C., Müller, M., Hickmann, V. & Hopfner, K.-P. X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121,** 363–73 (2005).

103. Elfring, L. K., Deuring, R., McCallum, C. M., Peterson, C. L. & Tamkun, J. W. Identification and characterization of Drosophila relatives of the yeast transcriptional activator SNF2/SWI2. *Mol. Cell. Biol.* **14,** 2225–34 (1994).

104. Dang, W., Kagalwala, M. N. & Bartholomew, B. The Dpb4 subunit of ISW2 is anchored to extranucleosomal DNA. *J. Biol. Chem.* **282,** 19418–19425 (2007).

105. Längst, G. & Becker, P. B. Nucleosome mobilization and positioning by ISWI-containing chromatin-remodeling factors. *J. Cell Sci.* **114,** 2561–8 (2001).

106. Yadon, A. N. & Tsukiyama, T. SnapShot: Chromatin remodeling: ISWI. *Cell* **144,** 453–453.e1 (2011).

107. Mueller-Planitz, F., Klinker, H., Ludwigsen, J. & Becker, P. B. The ATPase domain of ISWI is an autonomous nucleosome remodeling machine. *Nat. Struct. Mol. Biol.* **20,** 82–9 (2012).

108. Längst, G., Bonte, E. J., Corona, D. F. & Becker, P. B. Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. *Cell* **97,** 843–52 (1999).

109. Bartholomew, B. ISWI chromatin remodeling: one primary actor or a coordinated effort? *Curr. Opin. Struct. Biol.* **24C,** 150–155 (2014).

110. Yadon, A. N. & Tsukiyama, T. DNA looping-dependent targeting of a chromatin remodeling factor. *Cell Cycle* **12,** 1809–1810 (2013).

111. Fischer, C. J., Tomko, E. J., Wu, C. G. & Lohman, T. M. Fluorescence methods to study DNA translocation and unwinding kinetics by nucleic acid motors. *Methods Mol. Biol.* **875,** 85–104 (2012).

112. Corona, D. F. *et al.* ISWI is an ATP-dependent nucleosome remodeling factor. *Mol. Cell* **3,** 239–45 (1999).

113. Dang, W. & Bartholomew, B. Domain architecture of the catalytic subunit in the ISW2-nucleosome complex. *Mol. Cell. Biol.* **27,** 8306–17 (2007).

114. Partensky, P. D. & Narlikar, G. J. Chromatin remodelers act globally, sequence positions nucleosomes locally. *J. Mol. Biol.* **391,** 12–25 (2009).

115. Al-Ani, G., Malik, S. S., Eastlund, A., Briggs, K. & Fischer, C. J. ISWI remodels nucleosomes through a random walk. *Biochemistry* **53,** 4346–4357 (2014).

116. Manning, B. J. & Peterson, C. L. Releasing the brakes on a chromatin-remodeling enzyme. *Nat. Struct. Mol. Biol.* **20,** 5–7 (2013).

117. Clapier, C. R. & Cairns, B. R. Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature* **492,** 280–4 (2012).

118. Leschziner, A. E. *et al.* Conformational flexibility in the chromatin remodeler RSC observed by electron microscopy and the orthogonal tilt reconstruction method. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 4913–4918 (2007).

119. Lia, G. *et al.* ATP-dependent looping of DNA by ISWI. *J. Biophotonics* **1,** 280–6 (2008).

120. Bird, A. Perceptions of epigenetics. *Nature* **447,** 396–8 (2007).

121. Carey, N. *The Epigenetics Revolution: How Modern Biology is Rewriting our Understanding of Genetics, Disease and Inheritance*. (Icon Books, 2011). at <https://books.google.com/books?id=2hGUwTlHNH0C&pgis=1>

122. Hota, S. K. *et al.* Nucleosome mobilization by ISW2 requires the concerted action of the ATPase and SLIDE domains. *Nat. Struct. Mol. Biol.* **20,** 222–9 (2013).

Appendix A – Glossary

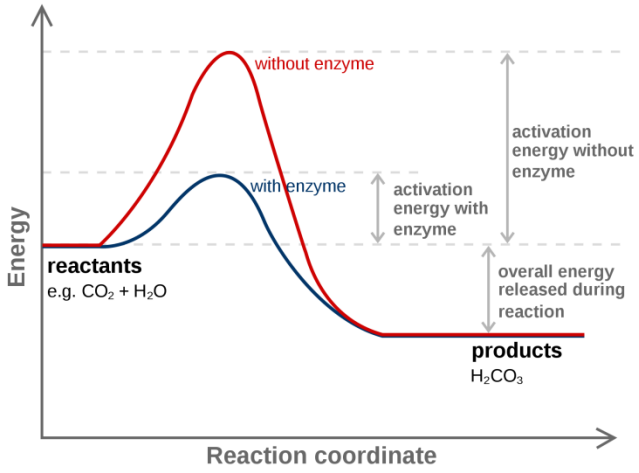
Due to the interdisciplinary nature of this body of work, there are many terms and topics that are not generally found in a physics dissertation. In this section you will find a collection of common words that may need additional explanation for those lacking a background in biochemistry. Barring that, it will suffice as a simple method of reference for the abundance of field-specific terminologies.

Adenosine triphosphate (ATP) – an adenosine molecule connected through bonds to three inorganic phosphates. By breaking the bond between the second and third phosphate groups of an ATP molecule, an enzyme obtains ~7.8 kcal/mol of free-energy to perform work. The products of this reaction are Adenosine diphosphate (ADP) and a free phosphate group (PO_4). ATP is a common and ubiquitous energy source for all known life.

Binding affinity – Two molecules in solution that can bind with each other will do so with respect to their shared intermolecular forces. When the two units interact, these influences are what determine the binding affinity, K_B. If strong forces exist, the molecules share high binding affinity; if low forces, a similar low binding affinity will be measured. This value is the opposite metric of the dissociation constant, K_D, a measurement of the concentration molarity at which half of the populations are bound. As the unit of K_D is molarity (M), the unit of a binding affinity is inverse molarity, M^{-1}.

Dalton (Da) – A common measure of mass at the atomic scale equal to one unified atomic mass unit (symbol: u). This is approximately the mass of one nucleon (proton or neutron).

Dissociation constant – For a general two part reaction $A_xB_y \leftrightarrow xA + yB$ in which a complex of two parts is disassembled, the dissociation constant is defined as $K_D = \frac{[A]^x[B]^y}{[A_xB_y]}$. Commonly in biochemistry and biology, a single instance of *A* and *B* interact to form the product, *AB* ($x = y = 1$). This leads to $K_D$being equal to the concentration of free *A* at which half the total population of B are in complex. The dissociation constant has the same units as concentration and is measured in molarity (M).



Appendix Figure 1 – Enzyme catalyzed reaction.

Enzyme - An enzyme is a highly-selective biological catalyst capable of lowering the activation energy ($\Delta G^{\ddagger}$) of a substrate-to-product transition and greatly increasing the rate of a reaction. It cannot, however, affect the equilibrium position of the reaction.

Epigenetics – While the majority of our genes are directly expressed through the base pair code of our genes, epigenetics is the study of cellular and physiological variations in genetic expression that are not directly caused by changes in the sequence. Rather, how external pressures can alter the translation of this code and cause amplification or suppression of these variations[120,121].

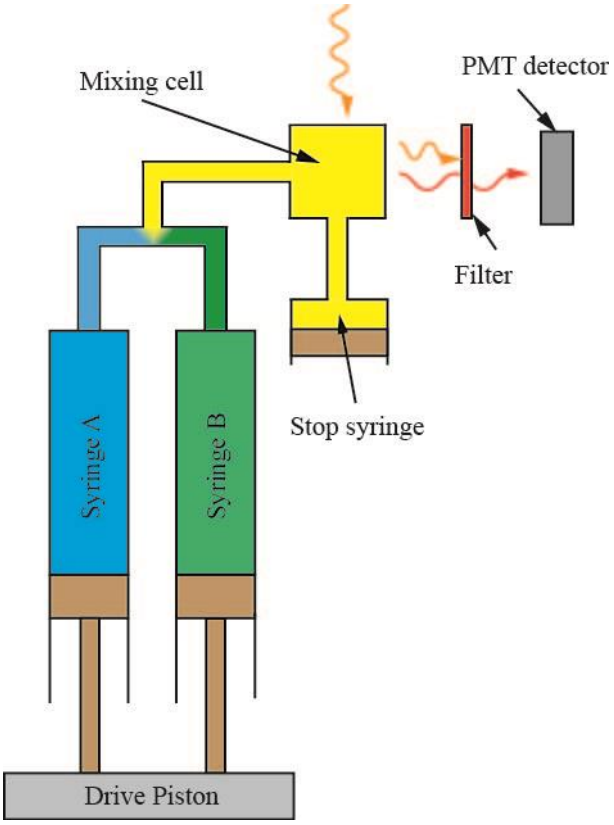
Molarity – Molarity is defined as the number of moles (6.022×10^{23}) of a substance per liter of solution, or $c = n/V$ where c is the concentration in Molar, n is the amount of solute in moles, and V the volume of solution in liters.

Proteome – The proteome is the entirety of the set of proteins expressed by a genome or cell at any given moment. It is a modular set, defined by environmental conditions, point in the cell cycle, or other external and internal factors.

Pseudo-first order reaction – Many biochemical reactions are considered 'second-order'; this means the reaction requires two components to occur. For example, let us begin with the assumption that reactants A and B are required to complete a reaction, forming a product AB. A solution containing both A and B is created with each reactant being at known concentrations, denoted as $[A]$ and $[B]$. The rate at which A will react with B is denoted by the rate constant k_1 whose units are s^{-1}. The equation that describes the rate at which AB is formed is $[AB] = k_1[A][B]$. This can be tricky to experimentally determine since accurate measurements of each concentration must be taken simultaneously and in real time. We can make simplifying assumptions if an extreme excess of A is included: during the course of the experiment the concentration of A can be considered to be unchanging, a constant. There is simply not enough B available in solution to make a significant change in the population of A. This allows us to suppose that the second-order reaction is *pseudo-first order* instead and dependent on the concentration of a single reactant only. We then denote a new rate constant, k_1', which is the product of the initial rate constant k_1 multiplied by the assumed-constant concentration of our reactant, $k_1' = [A]k_1$. The pseudo-first order rate equation R' is then $R' = k_1'[B]$. This is also true if the opposite ratio is desired, with B in large excess of A. The pseudo-first order rate equation would then be $R' = k_1'[A]$ instead.

Appendix B – Stopped Flow Spectroscopy

The majority of the work for this thesis was performed using stopped-flow spectrophotometric methods. A major benefit of this technique is that very fast timescales can be monitored. The two solutions are considered mixed within several microseconds and manufacturers often cite 1 millisecond as the lower limit for kinetic signal detection. It is a very useful experimental tool used to determine the reaction kinetics of ensemble solutions when you have reactants at sub-molar concentrations. Generally, there are two reactants that interact in some manner to form a product. They may form a product population either by creating a new compound, combining to generate a novel product, or one reactant may modify the other into a new form. In this technique, two solutions are held in separate reservoirs before being forced into a mixing chamber by the drive piston (Appendix Figure C). Usually, there are two reactants: one held in each reservoir before mixing. In general, the reactants by themselves are inert in solution. The reaction can then be monitored in the mixing cell by tagging one or more reactants with fluorescent molecules. The cell is then bathed in monochromatic light at a specific wavelength to incite a change in the fluorophore and by use of filters, the resultant emitted light is detected by photo multiplier tube (PMT) excitation. The resultant output is monitored as a change in the reference voltage from the PMT in Volts (V).



Appendix Figure 2 – Stopped Flow Spectrophotometer. Upon an initial signal, the drive piston simultaneously pushes solution from syringes A and B into the mixing cell. The amount driven is larger than the combined volume of tubing leading up to the cell, the mixing cell itself, and an additional amount necessary to push the stop syringe to its resting location. The cell is bathed in monochromatic light while the reaction takes place. The resulting emission is monitored with one or more detection tools at right angles to the inbound light.

In the experiments performed, the reactants were the individual chromatin remodelers, RSCt or ISWI, and a DNA substrate. Both being ATP-dependent remodeling proteins, a third compound was required for proper interaction with DNA: ATP. As explained in previous sections, without this available energy source the remodelers will bind to the DNA without having any available free energy to overcome the bonds and will be unable to undergo processive translocation. In the first syringe we placed the remodeler in a stabilizing solution, described in Chapter 4, and in the second we placed ATP and fluorescently-labeled double- or single-stranded DNA. When the solutions are mixed, the protein has access to both ATP and substrate. This allows them to rapidly bind available DNA molecule and ATP molecules, and began processive translocation.

Example collected data can be seen in Figure 8 and the general collected data curve shape can be seen most clearly in Figure 11B. When there is an interaction between the fluorophore and the protein in solution, there is a slight change in overall signal voltage versus time. This can be recorded as a positive or a negative change from the initial reference voltage, depending on the specific reaction between the two molecules. For these experiments, it was a predominantly a reduction in the output voltage of the PMT that was recorded. In control assays monitoring the binding of DNA with no present ATP, the sole interaction is the binding of protein to a specific site on the substrate. This causes a minor change that approaches equilibrium and once rebinding and unbinding are balanced, the signal will remain constant as

long as the protein complex is stable in solution. However when there is ATP present in solution and processive translocation can occur, we see different characteristics in shape. Figure 11 shows large signal change with the addition of ATP due to the increased interaction when ISWI is capable of reaching the fluorophore at the 5’ end of the substrate.