UNDERSTANDING THE MECHANISMS OF gp32

FILAMENT ASSEMBLY AND SLIDING ON ssDNA

TEMPLATES OF KNOWN SIZE AND POLARITY

by

MEGAN BARNEY

A THESIS

Presented to the Department of Chemistry
and the Robert D. Clark Honors
in partial fulfillment of the requirements for the degree of
Bachelor of Science

Spring 2019

An Abstract of the Thesis of

Megan Barney for the degree of Bachelor of Science
in the Department of Human Physiology to be taken June 2019

Title: Understanding the mechanisms of gp32 filament assembly and sliding
on ssDNA templates of known size and polarity

Approved: _____

Andrew Marcus

DNA replication is a core biological process that rapidly occurs in both
eukaryotic and prokaryotic cells with extreme precision. Gene product 32 (gp32) is a
ssDNA binding protein that is important in the T4 bacteriophage DNA replication
complex. Gp32 is known to bind cooperatively spanning 7 nucleotides of ssDNA. Not
only is it known to bind, but it has the ability to unbind from regions of exposed ssDNA
during DNA synthesis. This thesis reports microsecond single-molecule FRET
(smFRET) measurements on Cy3/Cy5-labeled primer- template (p/t) DNA constructs of
known length and polarity with and without an addition of 0.5uM gp32. The p/t
constructs are characterized by two different lengths, 14-nt poly(deoxythymidine)
$[p(dt)_{14}]$ vs 15-nt poly(deoxythymidine) $[p(dt)_{14}]$ and by the location that the cyanine
dyes(Cy3 and Cy5) are covalently attached to the DNA at the 3' or 5' ends. The
measurements obtained report the distance between the chromophores that are used to
label the ends of 14 and 15 nucleotide segments of ssDNA attached to a p/t DNA
construct. These distance measurements can track the conformational changes seen
between protein bound vs. unbound states on the microsecond time scale. To analyze

the data, a multipoint time correlation function analysis is utilized in order to compare the revealed kinetics of the possible conformational adaptation experienced by the ssDNA of interest. The results of our analysis demonstrate that both length and polarity of the ssDNA influence the way in which gp32 interacts with the ssDNA. The 14-nt templates have slower fluctuations than the 15-nt templates. The 5' ssDNA constructs experience fluctuations faster than the 3' ssDNA constructs. Therefore, this SSB is likely to play a critical role at the replication fork during DNA synthesis.

Acknowledgements

First and foremost, I would like to thank my thesis committee members: Professor Andrew Marcus, Professor Helen Southworth, and Brett Israels. I would like to thank Professors Andrew Marcus and Professor Peter von Hippel for welcoming me into their labs and for helping me to fully examine the topic of single stranded DNA binding proteins. You both have helped me to consider the various perspectives and contexts related to this subject matter. Furthermore, I would like to express my gratitude toward Professor Helen Southworth for being dedicated to helping me achieve success within the honors college. Her passion for teaching has encouraged my pursuit of life-long learning. I am beyond blessed to have had the privilege to work with such excellent mentors who were willing to guide me through the strenuous process. I am honored to have had the opportunity to conduct undergraduate research under their guidance. I am grateful to have had the opportunity to learn firsthand from Brett Israels, your patience in teaching me the experimental procedure and showing me the means of acquiring data is much appreciated. This has been a very rigorous yet rewarding process and I am so thankful for your support every step of the way.

I would additionally like to thank all of my friends who have supported me in my quest for academic excellence. In particular, I'd like to thank Casi Jackson for helping navigate infographic to create several of the figures used in this thesis. Finally, I would like to thank my family for constantly removing barriers for me along this journey, and for encouraging me to excel in all of my pursuits. It means the world to me knowing that you support me unconditionally and have had faith in my successful completion of this thesis.

Table of Contents

List of Accompanying Materials

1.

List of Figures

List of Tables

Introduction

Background

A primer on DNA:

Deoxyribonucleic acid (DNA) is responsible for storing genetic information. DNA is a linear polymer molecule whose structure consists of a sugar-phosphate backbone connecting the sequences of nucleotide(nt) bases that is the genetic "code". DNA is structured as a double-helix and consists of two complementary single DNA strands. There are four nucleotide bases that make up DNA: Thymine, Cytosine, Adenine, and Guanine. These nucleotide bases pair in a predictable fashion. There are three core biological processes in which DNA is involved: replication, repair, and recombination. DNA is preserved and replicated with high fidelity. In all organisms, DNA replication is necessary to allow for growth and reproduction upon cell division. However, as DNA is passed down from one generation to the next, fidelity errors can occur in the replication process. In order to minimize these errors, the DNA is involved in mechanisms known as repair and recombination. It is important to highlight that the human body exhibits these mechanisms of repair and recombination so that the human genome is only altered by a few nucleotides at most upon undergoing cell division. DNA is the key component of our diverse make-up as individual human beings. Thus, it is important that humans pass down non-mutated genetic information from generation to generation to avoid changes in cells that can lead to diseases such as cancer.

DNA replication occurs at rates, "as high as 1,000 nucleotides per second" (Alberts, 2015). The double helix secondary structure of DNA allows for semi-

conservative replication. Because Thymine only binds with Adenine, and Cytosine only binds with Guanine, each strand of DNA can act as a template to specifically duplicate the complementary strand by DNA base-pairing. Thus, in order to replicate DNA, the double stranded helix must be unwound and separated. Several proteins work in collaboration to successfully open up the DNA double helix and effectively create a replication fork of two single stranded regions of DNA, which are available as templates for the synthesis of new daughter strands. An enzyme helicase is responsible for unwinding the double-stranded (ds) DNA and functionally melting its secondary structure by separating the base pairs at the replication fork. At the replication fork, the single-stranded (ss) DNA templates are exposed for the template-directed synthesis of new daughter strands, which are catalyzed by DNA polymerases. Once the replication fork is established, there are two exposed single-stranded DNA. Single-stranded DNA-binding proteins (SSBs) bind to the exposed regions of ssDNA while leaving the bases exposed to the surrounding aqueous environment so that they can still act as templates for DNA synthesis. DNA has a 5' and 3' end which creates the directionality observed in DNA synthesis. The 5' end contains a phosphate group, while the 3'end contains a hydroxyl (-OH) group. DNA is synthesized in the 5' to 3' direction, therefore, this known polarity is important because it allows one strand to be synthesized continuously (leading strand) and explains why the other is synthesized in pieces forming okazaki fragments (lagging strand).
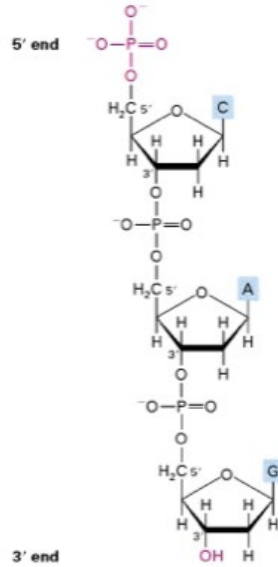
Figure 1(Lodishh, et. al, 2000): This figure shows three nucleotides of DNA connected by two phosphor-diester bonds. The difference in structure that accounts for directionality is depicted.

Single stranded DNA binding proteins:

At the site of the replication fork, SSB proteins bind both tightly and cooperatively to ssDNA in order to protect and stabilize the ssDNA prior to the synthesis of daughter DNA strands. Immediately after the DNA helicase has melted the dsDNA, the ssDNA must maintain its single-stranded structure free of base pairing until it is able to act as a template for DNA replication. Thus, SSBs bind to ssDNA to stabilize the separated strands and to prevent the DNA from adopting unfavorable conformations. For example, ssDNA will readily form short hairpin helices that can obstruct the desired DNA synthesis. However, when SSBs are bound to ssDNA, such secondary structural motifs cannot form. Once an individual SSB binds, it attracts other SSBs to bind in a cooperative fashion (Watson et. al., 2014). This cooperative binding allows the SSBs to also bind to each other. The interaction between SSBs acts to strengthen the stabilization of subsequent SSB binding to ssDNA.
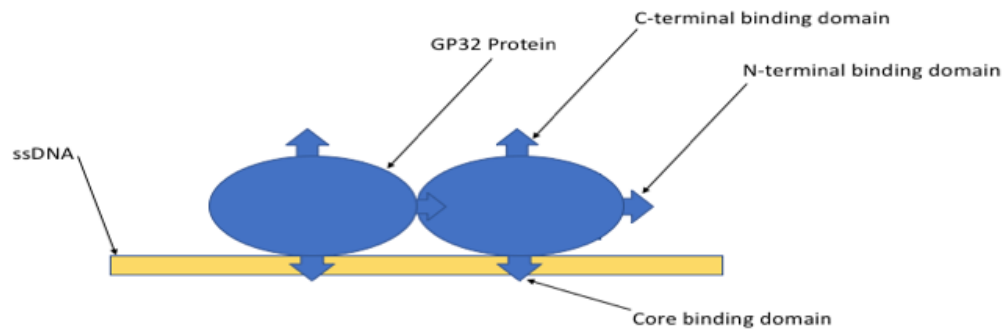
Figure 2: This schematic demonstrates gp32's three binding domains: N-terminal domain, Core domain, C-terminal domain.

 Thus, it is important that these SSBs bind because when ssDNA is coated with SSBs, it takes on an elongated state that facilitates the ssDNA's function for templating replication. As seen in Figure 3, SSB proteins are an integral component of the DNA replication system, allowing the single-stranded regions of DNA to carry out its function as the template for DNA polymerization (von Hippel, 2007). Figure 3 demonstrates a supposed distribution of gp32 molecules which inevitably varies from one complex to the next depending on the lengths of the exposed ssDNA segments. Nonetheless, the SSB proteins have been found to form a "central part of the T4 DNA replication complex" (Albert & Frey, 1970), in addition to playing important roles in DNA recombination and repair mechanisms. Therefore, advancing our understanding of their functional mechanisms is significant to all biological processes involved in gene regulation.
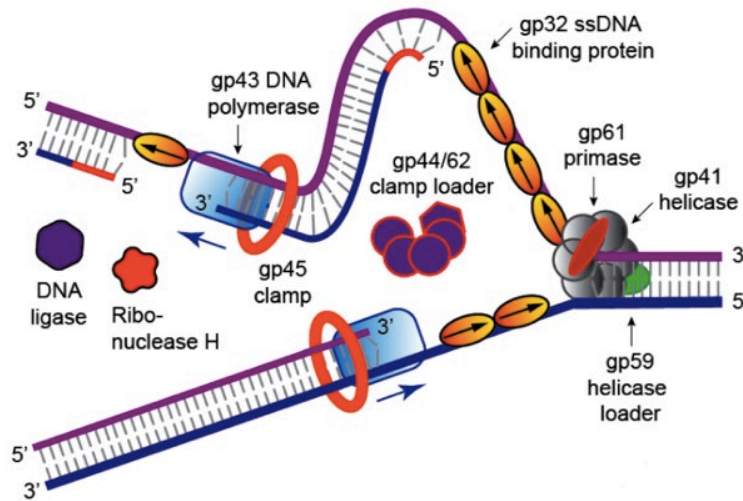
Figure 3(Lee, et. al, 2016): This is a schematic of the primary protein components involved in the replication of DNA within the T4 bacteriophage. The schematic demonstrates the elongating DNA replication complex. It depicts the gp32 proteins bound in a polar fashion to the single-stranded portions of the DNA within the replication complex.

The specific SSB being studied is gene product 32 (gp32). Gp32 is known to cooperatively bind to ssDNA and takes up seven DNA nucleotides (Alberts & Frey, 1970). Because gp32 takes up seven DNA nucleotides on a region of ssDNA, the p(dT)15 DNA construct can only bind up to two gp32 molecules as seen in figure 2. On the phosphate backbone of the ssDNA, the gp32 protein's positive core is bound. Subsequently, when another gp32 protein binds, it will be in an adjacent position and bound to the n-terminal domain of the initially bound gp32. Gp32 is a non-specific SSB with three binding domains: N-terminal domain, core domain, and the C-terminal domain (Jordan & Morrical, 2015). Three binding domains mean that the protein can bind or interact with other components at three different sites. The N-terminal domain is responsible for mediating self-association to encourage bound monomers to cooperatively interact with additional gp32 monomers. The core domain of gp32

5

contains the ssDNA binding site, and the C-terminal domain mediates interactions with other proteins (Jordan & Morrical, 2015). Furthermore, gp32 bound on ssDNA can inhibit it's uncoiling, which indicates that this formation acts as a "stearic regulator" for the components of DNA accessible to the DNA dependent ATPase (Dda) helicase for processing (Jordan & Morrical, 2015). The addition of SSB's help straighten out the DNA, therefore when two gp32 are bound, figure 4 shows a straighter conformation of the ssDNA construct. As a steric regulator, the SSBs help determine which intermediates are accessible for processing by the Dda helicase during the core biological processes of replication and recombination.

0-bound state   1-bound state   2-bound state
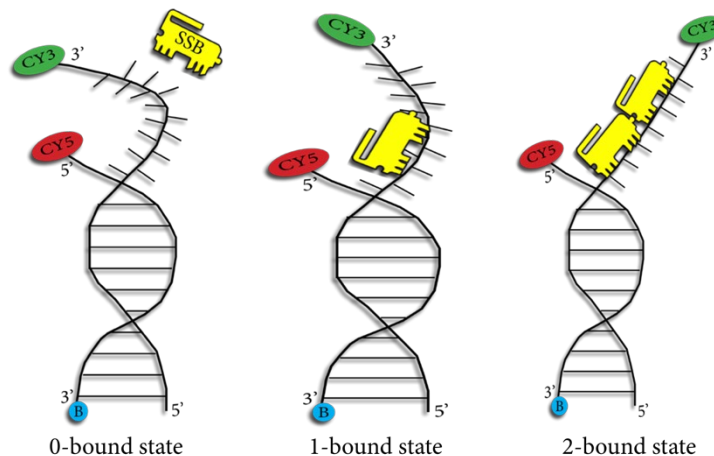
Figure 4: This is a schematic of the unbound vs. bound states of gp32 binding on the p(dT)$_{15}$ DNA construct. Two bound gp32 saturates the area of ssDNA on the p(dT)$_{15}$ DNA construct, thus it can only bind up to two molecules of protein.

The T4 bacteriophage:

There are two types of organisms: eukaryotes and prokaryotes. Eukaryotes are complex multicellular organisms such as human beings, while prokaryotes are viruses and bacteria whose genetic material is packaged into much simpler chromosomes than their eukaryotic counterparts. Therefore, in order to better understand eukaryotic

processes, prokaryotes can serve as simpler models to investigate the core biological processes that regulate genomes. The T4 bacteriophage is one example of a common prokaryotic virus that has not only been thoroughly studied but has also influenced many discoveries in molecular genetics. For instance, the genetic code, and the confirmation that DNA is the genetic substance has been understood more clearly after studying T phages. The T4 Bacteriophage is able to replicate its genome in a simple manner by using a limited number of molecular components, while still using the major replication sub-assemblies that characterize eukaryotic organisms (Lee et. al., 2016). Similar to eukaryotic replication, T4 utilizes a helicase-primase complex to open dsDNA into two single-stranded regions, primes DNA synthesis on the "lagging" strand, has a DNA polymerase, and replication clamps (Lee et. al., 2016). As a result, it is the simplest model for understanding DNA replication in higher organisms and allows us to explore in the mechanistic pathway of SSB assembly and the conformational dynamics between these proteins and their nucleic acid targets.

Previous Research and Rationale:

The structure of DNA plays a major role in determining whether or not a SSB will bind. Segments of DNA that are rich in adenine and tyrosine base pairs are less stable due to segments of DNA that are rich in cytosine and guanine base pairs. Therefore, due to DNA's length with heterogeneous sequences, the genome will melt in a multistate fashion (von Hipple, 2013). This heterogeneity of dsDNA stability suggests that the DNA genome will experience thermally motivated fluctuations in conformation at physiological temperatures. Thus, the behavior of the DNA is sequence dependent and the changes in conformation it experiences may reveal information that influences

the binding of regulatory proteins and protein complexes in reactions that require

dsDNA 'opening'" (von Hipple, 2013). Gp32 is more likely to bind to ssDNA than

dsDNA, therefore this suggests that according to the laws of thermodynamics, this

specific SSB should function as a 'DNA melting protein' (Alberts & Frey, 1970). This

implies that gp32 should, in theory, decrease the melting temperature to allow for

saturation of ssDNA with gp32 (von Hipple, 2013). However, previous experiments did

not find this to be the case in most duplex DNA molecules. Because there wasn't

significant evidence of gp32 acting in this expected manner, it can be predicted that the

melting must be kinetically blocked (von Hipple, 2013).

In 2015, research showed that Gp32 is undeniably a significant SSB that plays a

role in regulating Dda helicase activity which may have influence on the outcomes of

the core biological processes that DNA is involved in (Jordan & Morrical, 2015). The

SSB Gp32 has been found to form cooperatively bound clusters that prefer to assemble

in the 5′ to 3′ direction during DNA replication at the replication fork of ssDNA (Jones

et al., 2004). Understanding this preferential assembly is important because it

encourages the availability of the generated ssDNA for replication prior to full

saturation by the Gp32 protein (Jordan & Morrical, 2015). It was proposed in 1995 that

Gp32 slides freely along ssDNA (Shamoo et al., 1995). If this is true, it is important to

know how assembly and "sliding" affects the conformation of the ssDNA during

replication because it may exert influence over the likeliness that replication errors will

occur. Furthermore, if Gp32 does in fact slide freely, the clusters that Gp32 forms could

be moved by Dda and act as if it were a break on the helicase, which would functionally

slow down the rate at which DNA is replicated (Jordan & Morrical, 2015). Ultimately,

the results of the study suggest the paramount importance of the Gp32 SSB regulatory function. With that being said, understanding the protein's mechanism for regulation of helicase activity in the T4 bacteriophage system with respect to ssDNA is valuable to understanding how it achieves this function and can be indicative of the role SSBs have in the replication of eukaryotic DNA.

In early research on the mechanism of SSB's, bulk solution studies had been used in order to explain binding thermodynamics of gp32. However, since then high-resolution research has been performed in order to indicate a detailed explanation of the structure and interactions of SSBs with ssDNA. After improvements in single-molecule spectroscopic techniques, researchers in the Marcus and von Hippel labs at the University of Oregon presented an approach to further studying the dynamics of gp32-ssDNA interactions. They published an article in 2016 describing that their approach was based on the analysis of single-molecule Förster resonance energy transfer (smFRET) studies of ssDNA backbone fluctuations using fluorescent cyanine dyes as FRET pairs on constructs of DNA replication forks. The study initiated a characterization of the binding dynamics of gp32 as necessary for influence of the binding of sub-assemblies necessary to carry out DNA replication.

Research didn't cease there. Another article was published in 2017 by researchers in the Marcus-von Hippel labs. The publication focuses on how the use of microsecond single-molecule FRET has been used to determine the assembly pathways of T4 ssDNA binding protein onto model DNA replication forks. The experiments allowed for the determination of detailed kinetic pathways about the assembly of gp32 dimer-ssDNA filaments using a 15-nucleotide ssDNA template as the substrate. The

analysis of these experiments led to identification of short-lived gp32 monomer-ssDNA intermediates (Phelps et. al, 2017). However, it was concluded that the observation of the bound states mandates signal fluctuations on the microsecond timescale and their sample of smFRET experiments were far too noisy to undergo analysis using hidden markov modeling (HMM). Therefore, in the future it will be important to collect data with an increased signal to noise ratio in order to perform this desired HMM analysis. Without this modeling, the research was still able to conclude that gp32 monomers do not slide freely along ssDNA like gp32 clusters do as proposed by the previously discussed study that was performed in 2015 by Jordan and Morrical. Thus, it can be deduced that the formation of cooperatively bound clusters of the SSBs must have an initially bound gp32 monomer that's positioned so that it can accommodate additional binding of gp32 proteins (Phelps et. al., 2017). Their research makes it clear that the subsequent binding after an initially bound gp32 is important because otherwise the singly bound protein will dissociate and be unable to act as a regulatory component of DNA replication. Nevertheless, additional research is necessary to determine the limits to this mechanism, and to provide further detail about what determines whether there is a bound gp32 monomer or gp32 clusters.

Research Question

 The topic of research I am investigating in my thesis is to further understand the mechanisms of gp32 filament assembly and sliding on ssDNA templates of known size and polarity. In order to achieve this purpose, my exploration consists of asking "What is the binding mechanism of gp32 on single stranded DNA, what conformational changes are seen, and how does polarity and length affect gp32 protein sliding?"

Methodology

In order to determine the binding mechanism of gp32 we take advantage of single-molecule FRET. The FRET efficiency reflects changes in distance and is useful for determining conformational changes of molecules. The phenomenon of FRET spectroscopy involves the distance-dependent transfer of energy between two fluorophores with spectral overlap. FRET is calculated by dividing the intensity of the donor molecule by the sum of the intensity of both the donor and acceptor over a certain amount of time. Therefore, by site-specifically labeling the DNA with FRET chromophores Cy3 and Cy5, the change in FRET signal is interpreted as a distance on the DNA polymer. Cy3 and Cy5 are utilized because they have a high florescence quantum yield and they do not disturb the DNA backbone. They are able to attach to DNA templates without disruption of the DNA structure because they are small chromophores characterized by a structure very similar to that of the DNA backbone as demonstrated in figure 5.
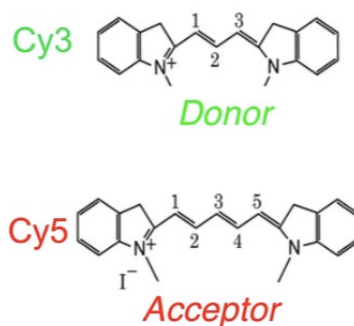

Figure 5: This is the structure of the Cy 3 and Cy 5 chromophores being utilized during the single molecule experimentation to record fluorescent signals and calculate FRET.

In this work, four different fluorescently labeled primer-template(p/t) DNA
constructs are being used to collect single molecule data. These constructs are
characterized by two different lengths, 14-nt poly(deoxythymidine) $[p(dt)_{14}]$ vs 15-nt
poly(deoxythymidine) $[p(dt)_{14}]$. In addition, the cyanine dyes(Cy3 and Cy5) are
covalently attached to DNA on either the 3' or 5' side of the fork junction. Thus, the
constructs allow us to study different polarities, 3' Cy3 attachment vs 5' Cy3
attachment. The four constructs utilized are depicted by a schematic in figure 6.

$3'\text{-}p(dT)_{14}$	$3'\text{-}p(dT)_{15}$	$5'\text{-}p(dT)_{14}$	$5'\text{-}p(dT)_{15}$
S18S2	S1S2	S19S5	S4S5
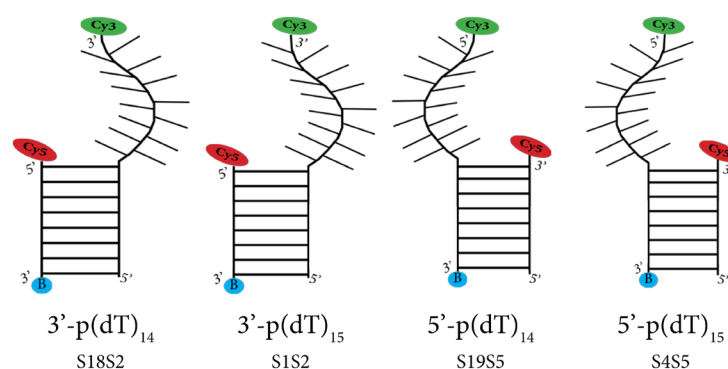
Figure 6: This is a schematic of the four constructs used during the single molecule
experimentation. The blue circle at the base of the dsDNA region represents the biotin-
chemistry that is used to bind the DNA to the quartz slide that is being imaged. The red
circle depicts where the cy5 chromophore is attached and the green shows where the
cy3 chromophores are attached for each of the four constructs.

Traditional ensemble measurements are limited to determining an average signal
of many molecules and thus can't reveal the transition between short-lived intermediate
states. Therefore, single molecule spectroscopy is being utilized instead to allow the
observation of one molecule over time. In doing so, it allows us to learn about the
dynamics and identify short-lived intermediate states. To perform the single molecule
experiment, we use a microscope to image labeled DNA chemically attached to a quartz
slide. The DNA is attached to the quartz slide using biotin neutravidin chemistry. The

12

DNA is attached to the slide to ensure that the DNA does not move around during imaging which allows the imaging of the single molecule over extended time. The microscope sends the image to the electron-multiplying charged coupled device (emCCD) or avalanche photo-diodes (APD). emCCD's image the molecules at millisecond(ms) resolution, while the APDs measure fluorescence at a faster resolution (μsec, or one-millionth of a second). The emCCDs image many molecules simultaneously(figure 8B), whereas the APD has a single element for measuring fluorescence intensity which can only be used to monitor a single molecule at a time (figure 8A).

Donor (Cy3)   Acceptor (Cy5)
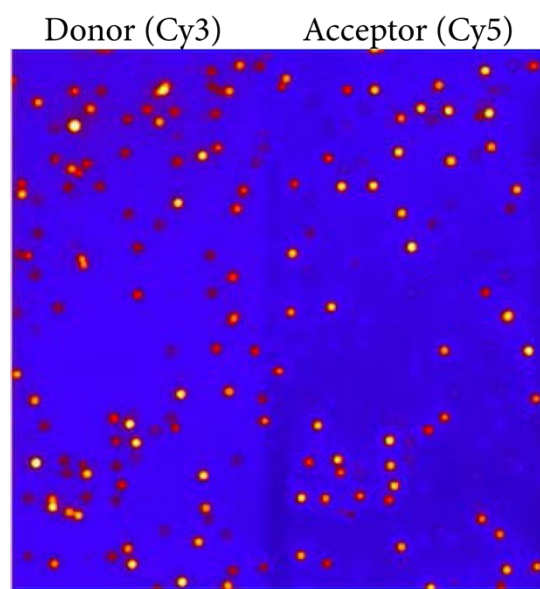
Figure 7: This is a screenshot of the emCCD split screen. The donor is detected on the left side while the acceptor is detected on the right side of the image. From this image, the donor is mapped to the acceptor to determine the location of the corresponding chromophore pair.

Because many of the important intermediates only last for milliseconds (one-thousandth of a second) and less, we need the faster detection resolution in order to visualize them. The single-molecule set up includes a 100µm pinhole that limits the fluorescence to one

molecule and focused onto the APDs (figure 8A). The fluorescence from the donor and acceptor is spectrally separated in order to monitor both fluorophores independently so that the relative fluorescence intensity can be used to calculate FRET. Single-molecule data is taken on all four constructs that consist of only DNA, and on constructs of DNA with an addition of gp32 protein.

The purpose of the single molecule experiment is to determine the quantity and nature of conformational states observed, as well as the rates associated with transitions between states from the analysis of single-molecule trajectories.
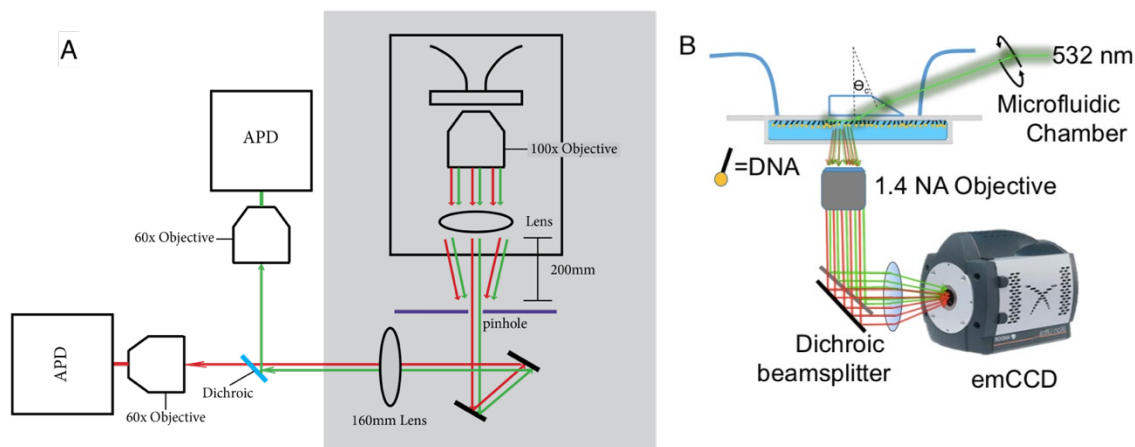
Figure 8: (A) A 100-micron pinhole blocks the fluorescence from all but one molecule. Subsequent to the pinhole, the light that passes through is separated and focused onto two avalanche photo-diodes (APDs). APDs are utilized for μsec resolution. (B) Biotin-labeled DNA constructs are chemically attached to quartz slides. The slide is then imaged with a high magnification microscope onto a detector. The electron-multiplying charge coupled device (emCCD) is used for millisecond resolution.

Results

The primer template strands of DNA being used are 14- and 15- nucleotides long. Because a gp32 protein occupies a binding site of seven nucleotides long, both of these ssDNA templates are limited to supporting a maximum of two gp32 proteins at the same time. When a gp32 protein binds to ssDNA, it is known to straighten out the conformation of the region to which it is bound. This interaction increases the separation between the Cy3 and Cy5 labels added to the regions of ssDNA being studied. In turn, when a gp32 protein is bound it results in a decreased smFRET efficiency. Therefore, gp32 binding to varied ssDNA templates induces stochastic transitions between conformational states of the primer template strands under investigation.

Figure 9 highlights the differences in raw data retrieved between DNA only and DNA with 0.5uM gp32. Looking at the smFRET trajectories of donor(Cy3) and acceptor(Cy5) it becomes apparent that the two intensities experience a dramatic increase in correlated fluctuations in the presence of 0.5uM gp32. This is also demonstrated in the graphical representation of FRET(blue). In the raw data from a DNA only recording, there is a relatively stable calculation for FRET. The sudden decrease is representative of a photo bleach that occurred during the duration of the recording. The DNA with 0.5uM gp32 data, in contrast, shows clear fluctuations in FRET values. These FRET value fluctuations are predictive of the DNA undergoing conformational changes throughout the recording as a result of the binding and unbinding of gp32.

15
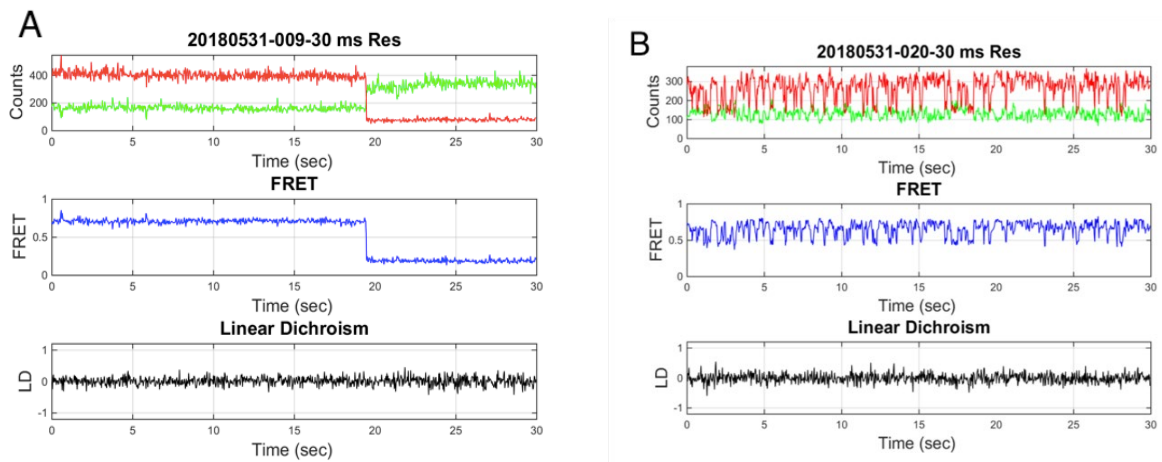
Figure 9: Representative single-molecule donor Cy3 (green), acceptor Cy5 (red), and smFRET trajectories (blue) taken from the 3' -p(dT)$_{14}$ DNA construct, and linear dichroism (black). (A) Microsecond Resolution Single-Molecule FRET on the 3`-p(dT)$_{14}$ with 0.0uM gp32. (B) Microsecond Resolution Single-Molecule FRET on the 3`-p(dT)$_{14}$ with 0.5uM gp32.

Furthermore, the histograms seen in figure 10 provides insight to all the possible conformational states taken on by the DNA at a given resolution. In addition, they clearly depict the distribution of these conformations for the constructs of varied length and polarity. There is clearly a wider distribution of observed FRET in the data taken with 0.5μM gp32 in comparison to the FRET values calculated from the data taken of DNA only. This wider distribution implies that the DNA is undergoing more fluctuations and taking on different confirmations when gp32 is present at 0.5μM. Examples of these conformations are depicted by the 0-bound, 1-bound, or 2-bound state demonstrated in figure 3.

DNA Only   DNA + 0.5 μM gp32

3'-p(dT)$_{14}$

3'-p(dT)$_{15}$

5'-p(dT)$_{14}$

5'-p(dT)$_{15}$
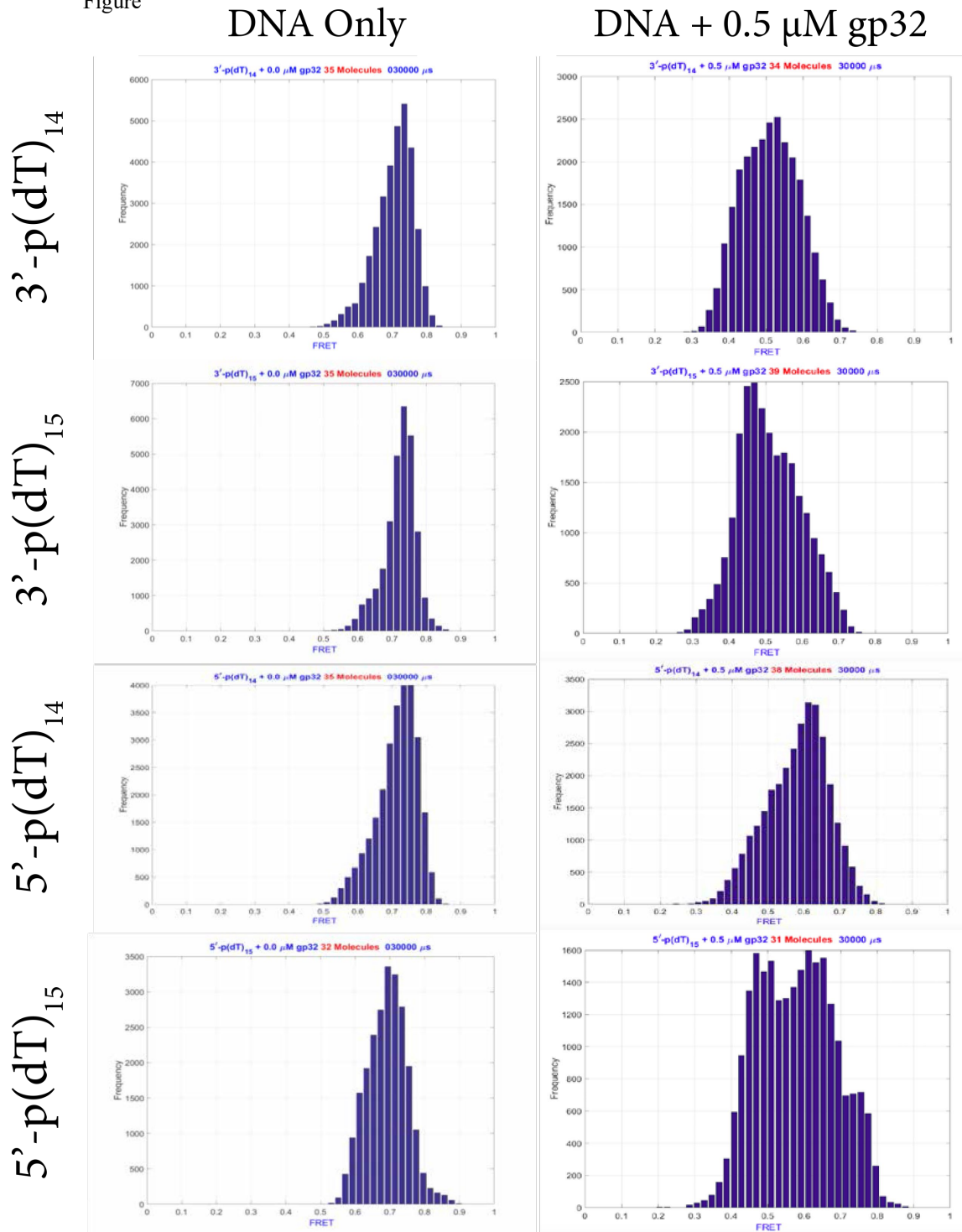
10: These are the histograms associated with each of the DNA constructs that were imaged with DNA only and with DNA + 0.5μM gp32. Each of the histograms reveals the observed distribution in FRET states.

In order to interpret the data that was collected during the smFRET experiment, the application of time correlation functions (TCFs) was utilized. TCFs demonstrate the rate at which the system experiences changes between states. Furthermore, TCF's allow the study of kinetics in addition to the mechanisms of protein binding, unbinding, and the possible sliding of gp32. The second-order TCF of the signal from the FRET at a particular instant is the average product of two consecutive measurements made at times t1 and t2, which are separated by the interval $\tau = t2 - t$.
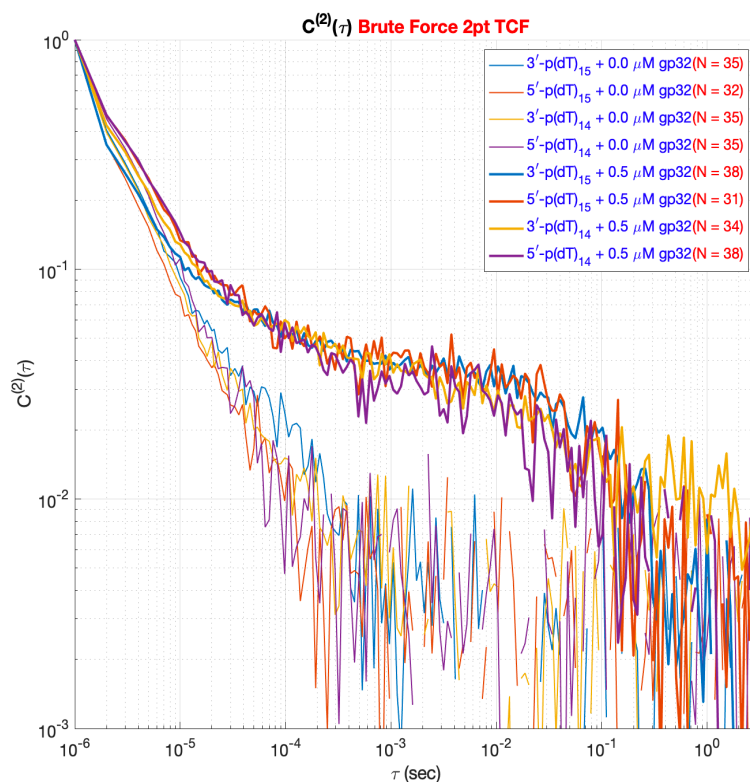
Figure 11: Calculated, normalized C(2)(τ) (2-point time correlation functions) at 1 μsec resolution. The decays are plotted using a linear-log axes scale. The graph shows that the templates without an addition of gp32 the C(2)(τ) can be fit with three exponentials, however, when gp32 is added, to properly fit the C(2)(τ), it required at least four exponentials.

In the graph of the second-order TCF seen in figure 11, it is clear that when data was taken on the constructs of DNA only, the rate of decay is faster than the data taken

on the constructs of DNA with an added $0.5\mu M$ of gp32. Furthermore, the DNA only

data required 3 exponentials, while the DNA + $0.5\mu M$ gp32 data required at least 4

exponentials. It is important to recognize that this difference in fit exists. The

significance of the decay components in second-order TCF's are explained by the

theory of Markov chains (Phelps et al., 2016). Markov chain theory is responsible for

characterizing a chemical system's potential kinetics in equilibrium and states that often

times stochastic transitions between identified N states occur (Reichl, 1998). According

to the Markov chain theory, second-order TCF's contain N-1 decay components.

Therefore, the data taken on all four constructs of DNA only TCF suggests that $N \geq 4$,

and the data taken on all four constructs of DNA $+0.5\mu M$ gp32.suggests that when

SSB's are added $N \geq 5$.

Cy3	Cy5	Template Description	[gp32] (μM)	Fast Fluctuation Rate (μs)	Medium Fluctuation Rate (μs)	Slow Fluctuation Rate (μs)	Fast Amplitude	Medium Amplitude	Slow amplitude	R^2
S1	S2	3'-p(dT)15	0	1.7	127	-	0.93	0.11	-	0.974
S1	S2	3'-p(dT)15	0.5	1.4	102	115	0.92	0.06	0.04	0.985
S18	S2	3'-p(dT)14	0	1.8	75	-	0.97	0.06	-	0.967
S18	S2	3'-p(dT)14	0.5	1.7	100	50	0.92	0.07	0.03	0.985
S4	S5	5'-p(dT)15	0	1.5	149	-	0.94	0.08	-	0.978
S4	S5	5'-p(dT)15	0.5	2.1	82	78	0.83	0.11	0.06	0.983
S19	S5	5'-p(dT)14	0	1.9	65	-	0.93	0.11	-	0.967
S19	S5	5'-p(dT)14	0.5	2	81	35	0.87	0.12	0.04	0.978

Table 1: This table organizes the results of fitting the microsecond resolution 2-point
TCFs. The reported fluctuation rates are inverse rate constants.

According to the data represented in table 1, the 14-nt templates have

slower fluctuations than the 15-nt templates. In addition, the 5' ssDNA

constructs experience fluctuations faster than the 3' ssDNA constructs. In all

four constructs, the smFRET trajectories qualitatively reveal the same behavior

as the experimental trajectories, therefore the data is able to provide evidence

that describes the dynamics of the ssDNA-(gp32)$_n$ system. It is important, however, to keep in mind that the indicated intermediates are incredibly short lived.

Discussion

The results of the analysis above provides indication of the basic assembly mechanism of the cooperatively bound T4 bacteriophage ssDNA and gp32. In addition, the results can contribute insight to the way in which the replication cofactors are attracted by the binding of SSB's during replication, recombination and repair. In each of the constructs, with and without gp32, a fast fluctuation occurs around $1.8\mu s$. The consistency across the constructs both with and without gp32 suggest this could potentially be the timescale in which nucleotides fluctuate. The slow fluctuation rate is only observed in the DNA + $0.5\mu M$ gp32 data. Therefore, it can be inferred that this slowest fluctuation is the time scale that the protein is interacting with the DNA forks.

It is important to note that the data described in the results demonstrates something about both length and polarity. When there is DNA only, the length of the single stranded region dictates the timescale of the fluctuations between intermediates. The ssDNA with a length of 14 nucleotides displays faster fluctuations than the ssDNA with a length of 15 nucleotides. However, when protein is present in solution, the polarity appears to take precedence and determines the timescale of the characteristic fluctuations. The constructs where the single stranded region's 5' end is exposed demonstrates faster fluctuations than when the 3' end is exposed.

Why do the 14-nt constructs fluctuate faster than the 15-nt constructs?

The faster fluctuations observed in the 14-nt constructs can be a result of the decreased distance for the Cy3 and Cy5 molecules to travel in order to become close in proximity. With a decreased distance between the fluorescent molecules, it allows for a greater transfer of energy quicker due to the molecules having to travel a smaller distance to become within close proximity. In addition, in the presence of protein, the 15-nt templates can have a gp32 dimer bound in two ways (1-14 or 2-15). This is because gp32 takes up 7-nt and therefore, a bound dimer would saturate the 14-nt templates leaving it with only one possible location to bind (1-14). With that being said, because the 15-nt has more potential orientations for the protein of interest to bind, it may experience slower fluctuations.

Why do the 5' constructs fluctuate faster with protein?
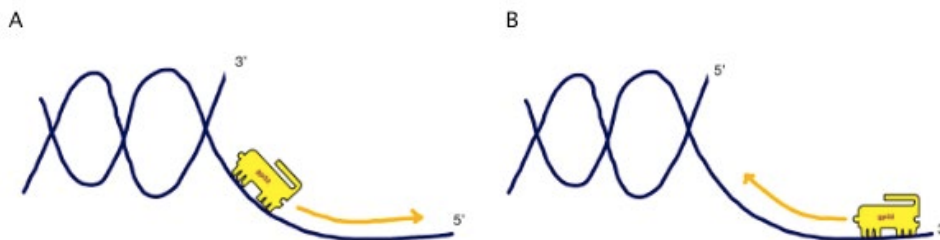
Figure 12: (A) This schematic demonstrates the proposed mechanism of sliding on the 5' templates. It demonstrates that because analysis suggests gp32 slides from the 3' to 5' direction, on the templates being studied the gp32 could "fall off."(B) This schematic demonstrates the proposed mechanism of sliding on the 3' templates. It demonstrates that because analysis suggests gp32 slides from the 3' to 5' direction, on the templates being studied the gp32 could be inhibited from unbinding due to interaction with the DNA replication fork.

The analysis suggests that the dimer slides from the 3' to 5' of end of the ssDNA. This proposed direction of "sliding" is consistent with the direction in

which DNA polymerase reads DNA which would make physiological sense due to gp32's regulatory role in both potentially attracting DNA polymerase, and in coordinating repair mechanisms to ensure DNA is replicated with precision. With that being said, this proposed sliding mechanism is consistent with our data because it would mean that dimers bound to the 3' constructs will slide toward the fork, but dimers bound to the 5' construct will slide away from the fork as seen in figure 12. On the DNA constructs being studied, if the dimer is sliding away from the fork, it may "fall off" the single stranded region of DNA (figure 12A). The potential for the dimer to "fall off" would promote a quicker rate of fluctuation between the bound and unbound conformations, which is consistent with our data (table 1). However, for the 3' constructs being studied, the dimer would slide toward the DNA fork (table 12B). If the protein slides toward the DNA fork, it's interaction with the double-stranded region of DNA may encourage the dimer to stay bound for a longer duration of time and hence lead to the observed slower rate of fluctuation consistent with the data in table 1.

Conclusion

Correlation functions can now be fit to smFRET experiments on DNA at microsecond resolution. The two-point time correlation functions have assigned preliminary values that are able to describe the decay. With respect to the four primer templates we studied, there is evidence that the DNA templates with 14 nucleotides experience fluctuations slower than the DNA templates with 15 nucleotides. In addition, the 5`-constructs fluctuate at faster rates than the 3`-constructs.

Relevance

Based on previous research, it becomes clear that SSBs play a significant role in the regulation of helicase activity (Jordan & Morrical, 2015). Given that helicase is the primary enzyme responsible for initiating DNA replication, it is important to understand the mechanism and regulation of ssDNA binding proteins for advancement in the fields of oncology, and drug development. Cancer, although extremely complex, can be simply defined as a genetic disease that is developed as a result of mutation in genes that exert influence over cell birth and cell death. When there is a balance between cell birth and death the tissues are healthy and at a state of homeostasis. Proto-oncogenes are responsible for promoting cell division or cell survival in eukaryotes, while tumor suppressor genes are in charge of inhibiting unnecessary cell division or cell survival. Mutating either of these could cause inappropriate activity that leads to a disruption in the cell ultimately leading to cancer. Therefore, it is extremely important to understand the core biological processes associated with DNA in order to understand what specific mutants have possible relation to cancer.

Existing research provides evidence of the central roles played by SSB's in DNA replication, the restart of stalled replication forks, DNA damage repair, cell cycle-checkpoint activation, and telomere maintenance (Ashton et. al., 2013). Therefore, because of SSB's importance in mechanisms used to maintain a homeostatic balance between cell birth and cell death, drugs have been developed to inhibit certain SSBs from binding ssDNA in humans preventing the progress of replication (Wu et. al., 2016). With that being said, it is of increased importance to fully understand their mechanisms of interaction with DNA in order to continue to develop the necessary pharmaceuticals to aid in inhibiting the progression of unwanted DNA replication in addition to encouraging the mechanisms of DNA repair when necessary to avoid unwanted mutations that can lead to malignancy.

Future Directions

Although there have been preliminary values associated with the TCF's reported according to the three-exponential fit to the DNA only data and at lease a four-exponential fit to the DNA with 0. 5µM gp32, it is possible that there is an additional component to these reported protein-mediated decays. A reason this component may have gone undetected includes the possibility that it occurred over a longer period of time that wasn't able to be captured with the microsecond resolution data. A future direction of this research includes determining if this additional component does indeed exist.

In addition, time density plots (TDP's) and hidden markov models (HMM) can be utilized as a method of analyzing the single molecule data collected during the

experiment. HMM could help determine which conformations are statistically significant, and TDP's could help determine the ways in which the conformational states are connected with one another.

Other SSB's that are known to play a direct role in DNA replication can be studied in a similar manner to see how the data retrieved compares or contrasts to the data yielded as a result of the experiment with gp32. Furthermore, in the future human SSB's can be analyzed to identify differences between the simple T4 bacteriophage and higher order organisms.

Bibliography

Alberts, B.M., Frey, L. (1970). T4 bacteriophage gene 32: a structural protein in the replication and recombination of DNA. Nature, 277, 1313-1318.

Ashton N.W., Bolderson E., et.al. (2013) Human single-stranded DNA binding proteins are essential for maintaining genomic stability. BMC Mol Biol. 14:9.

Jordan, C. S., & Morrical, S. W. (2015). Regulation of the Bacteriophage T4 Dda Helicase by Gp32 Single-stranded DNA-Binding Protein. DNA Repair, 25, 41–53.

Lee, W., Jose, D. et.al. (2013). A Single-Molecule View of the Assembly Pathway, Subunit Stoichiometry and Unwinding Activity of the Bacteriophage T4 Primosome (helicase-primase) Complex. Biochemistry,52, 3157-3170.

Lee, W., von Hippel, P. H., & Marcus, A. H. (2014). Internally labeled Cy3/Cy5 DNA constructs show greatly enhanced photo-stability in single-molecule FRET experiments. Nucleic Acids Research, 42(9), 5967–5977. http://doi.org/10.1093/nar/gku199

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000). Molecular Cell Biology, 4th edition. New York: W.H. Freeman.

Meisburger, S. P., Sutton, J. L., et. al. (2013). Polyelectrolyte properties of single stranded DNA measured using SAXS and single molecule FRET: beyond the wormlike chain model. Biopolymers, 99(12).

Murphy, M. C., Rasnik, I., et. al. (2004). Probing Single-Stranded DNA Conformational Flexibility Using Fluorescence Spectroscopy. Biophysical Journal, 86, 2530-2537.

Noble, E., Spiering, M. M., & Benkovic, S. J. (2015). Coordinated DNA Replication by the Bacteriophage T4 Replisome. *Viruses*, *7*(6), 3186–3200. http://doi.org/10.3390/v7062766

Phelps C, Israels B., et. al. (2016) Using multiorder time-correlation functions (TCFs) to elucidate biomolecular reaction pathways from micro- second single-molecule fluorescence experiments. J Phys Chem B 120:13003–13016.

Phelps, C., Israels, B., et. al. (2017). Using microsecond single-molecule FRET to determine the assembly pathways of T4 ssDNA binding protein onto model DNA replication forks. Proceedings of the National Academy of Sciences of the United States of America, 114, E3612–E21.

Plumridge, A., Meisburg, S. P., et.al. (2017). The impact of base stacking on the conformations and electrostatics of single-stranded DNA. Nucleic Acids Research,45, 3932-3943.

Reichl LE (1998) A Modern Course in Statistical Physics (Wiley, New York), 2nd Ed.

Roy, R., Hohng, S., & Ha, T. (2008). A Practical Guide to Single Molecule FRET. *Nature Methods*, *5*(6): 507–516. http://doi.org/10.1038/nmeth.1208

Shamoo Y, Friedman AM, et. al. (1995). Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA. Nature. 376(6538):362–366.

von Hippel, P. H., (2007). From "Simple" SNA-Protein Interactions to the Macromolecular Machines of Gene Expression. The Annual Review of Biophysics and Biomolecular Structure, 36, 79-105.

Wu Y, Lu J, Kang T. (2016) Human single-stranded DNA binding proteins: guardians of genome stability. Acta Biochim Biophys Sin. 2016;48:671–7.

Yap, M. L., & Rossmann, M. G. (2014). Structure and function of bacteriophage T4. Future Microbiology, 9, 1319–1327.

Yuanzhong Wu, Jinping Lu, Tiebang Kang. (2016). Human single-stranded DNA binding proteins: guardians of genome stability. *Acta Biochimica et Biophysica Sinica*, 48(7): 671-677.