Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Gustavo J. Borjas April 13, 2020

Transcriptional Regulation by the Bacteriophage 186 Wrapping Protein

by

Gustavo J. Borjas

Laura Finzi

Adviser

Emory University Department of Physics

Laura Finzi

Adviser

David Dunlap

co-Adviser

Arri Eisen

Committee Member

Gordon Berman

Committee Member

2021

Transcriptional Regulation by the Bacteriophage 186 Wrapping Protein

By

Gustavo J. Borjas

Laura Finzi

Adviser

David Dunlap

co-Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Emory University Department of Physics

2021

Abstract

Transcriptional Regulation by the Bacteriophage 186 Wrapping Protein
By Gustavo J. Borjas

DNA provides the information necessary for life. Its code contains the instructions necessary to make various types of RNA molecules during a process known as transcription. The process must be tightly regulated to express only the types of RNA needed at any particular time during the cell cycle. Transcription factors regulate transcription, many of which bind directly to DNA. The aim of this study was to further understand transcriptional regulation by what is a fundamental mode of physical regulation by proteins: DNA wrapping around a wheel-like protein complex. Thus, we used nanometer-resolution imaging by Atomic Force Microscopy to understand the effect of the model bacteriophage 186 CI repressor protein on transcriptional elongation by RNA polymerase (RNAP). In the biological context, 186 CI acts as a genetic switch between the bacteriophage lysogeny/lytic decision by inhibiting RNAP initiation when bound to promoter sites. In this study, 186 CI—which wraps DNA like the eukaryotic histone octamer—was used as a model for the regulatory wheel mode. The topographical images of DNA-protein complexes were processed with MATLAB® to obtain DNA length, protein binding position along DNA as well as protein area and height data. It was found 186 CI and RNAP could be differentiated by their height: 1.5 ± 0.3 nm and 2.5 ± 0.4 nm, respectively. Image analysis showed the wheel mode of roadblocking does not effectively stop elongation by RNAP, but rather controls the genetic switch in bacteriophage DNA via inhibition of RNAP initiation. Furthermore, the reduced 186 CI diameter on DNA in images with active transcription indicates RNAP may partially break the 186 CI wheel by knocking off wheel dimers. Since it has also been shown histone dimers are lost from nucleosomes on the elongation pathway, we speculate loss of units from regulatory protein complexes could be a general mechanism of chromatin remodeling which facilitates roadblock removal from elongation pathways.

Transcriptional Regulation by the Bacteriophage 186 Wrapping Protein

By

Gustavo J. Borjas

Laura Finzi

Adviser

David Dunlap

co-Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Emory University Department of Physics

2021

Acknowledgements

My thanks to both of my advisers, Dr. Laura Finzi and Dr. David Dunlap, for taking me into their lab my freshman year of college and for the past four years of research mentorship and training.

My thanks to all the people who directly helped me in this project, namely Dr. Yue Lu who helped me gather data when I was not able to be in the lab, or Atlanta, because of the COVID-19 pandemic, Dr. Dan Kovari who provided a GUI framework as well as helped me early on in code development and taught me coding paradigms, Dr. Haowei Wang who preceded me in this research and whose dissertation I've studied closely, Dr. Zsuzsanna Vörös who taught me techniques and sample preparation for the AFM, Derrica McCalla and Stefano Martin who helped me with the DNA constructs, and finally, the rest of the Finzi-Dunlap lab for their support.

My thanks to the IMSD group, who provided me with a grant to be able to do engage in this research as well as supported me in my endeavors, especially Dr. Ludy Carmola and Dr. Edward Morgan who have directly mentored me and provided professional advice.

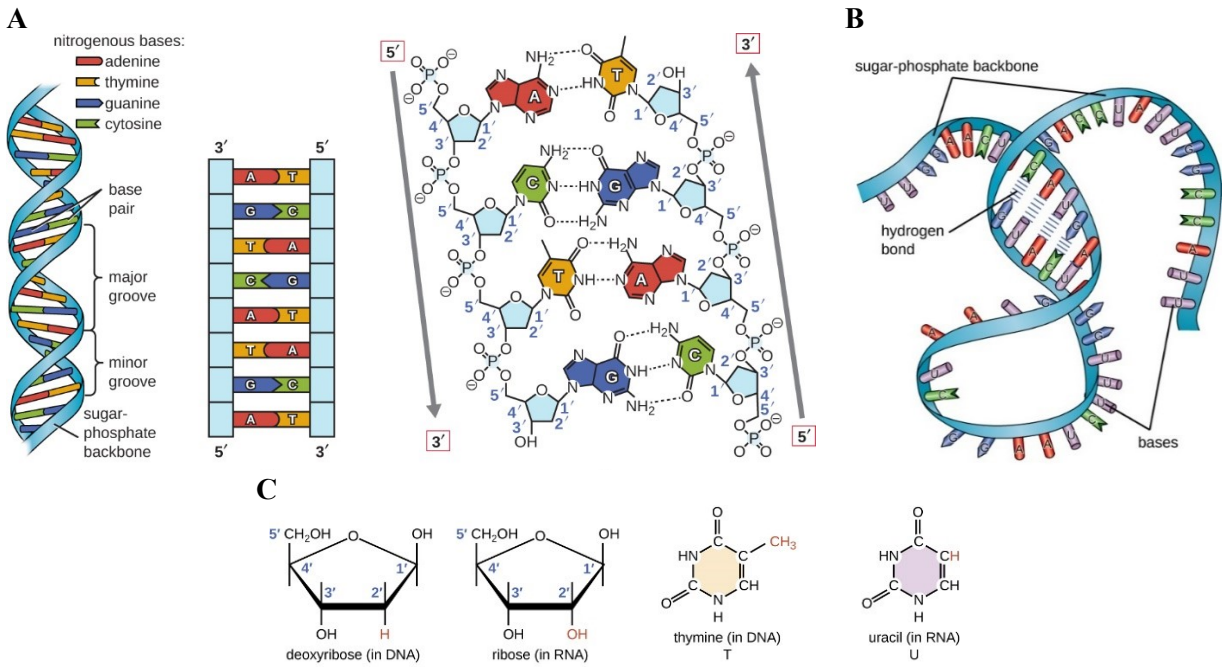
Finally, my thanks to all my friends and family who support me in all my past, present, and future endeavors.

Table of Contents

I. Introduction

A. DNA Transcriptional Regulation

Deoxyribonucleic acid (DNA) is the base code of all living organisms. It contains all the information needed for a cell to carry out its functions. Almost every cell in any given macroorganism shares the same copy of DNA (Fig. 1A) as every other cell in that organism; yet, different cell types have different functions. In part, this is achieved through the regulation of DNA transcription, the process which transcribes DNA into messenger ribonucleic acid (mRNA) molecules (Fig. 1B-C) that can ultimately be translated into functional proteins. Cells can regulate the process of transcription to differentiate into the desired cell type and perform the needed cell functions; thus, differential gene expression will produce differential cell function.

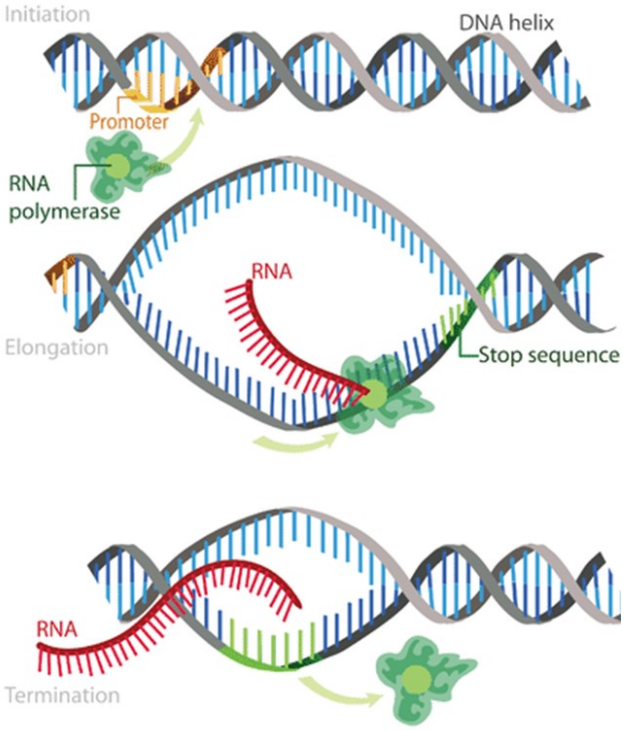


Source: https://openstax.org/books/microbiology/pages/1-introduction

Figure 1. Structure and Form of DNA and RNA. (A) From left to right, the tertiary double helix structure of B-DNA, the primary structure of DNA showing the sequence of nucleotides, and the secondary structure of DNA showing complementary base-pair hydrogen bonding. (B) Single Stranded RNA with complementary hydrogen bonding within the same biomolecule. (C) Differences between DNA and RNA base pairs where ribose sugar contains an extra 2` hydroxyl group (left) and thymine is replaced for uracil in RNA (right). Source: Openstax, Microbiology, under Creative Commons.

The process of transcribing DNA into mRNA is carried out by a protein, the motor enzyme RNA polymerase (RNAP), in three stages: initiation, elongation, and termination. Initiation involves formation of the open complex of DNA and RNAP and the synthesis of the first few-nucleotide-long mRNA. The next phase, elongation, involves promoter escape, where RNAP relinquishes it's hold on the promoter site, and the continued synthesis of mRNA in the 5'->3' direction. During elongation, RNAP lies within the transcription bubble where DNA is locally unwound. This local state causes positive supercoiling in the DNA in front of the transcription bubble and negative supercoiling behind it (Fig. 2). Finally, at termination, the RNAP reaches a terminator sequence and detaches itself from the DNA strand. This process depends on the accessibility of DNA to the RNAP so the latter can bind and initiate transcription first, and then elongate unhindered.

The regulation of the transcription process in relation to DNA's accessibility to transcription proteins has been studied extensively. In eukaryotes, RNAP requires other proteins, known as transcription factors, to come together and facilitate its binding at the transcription start site. This is known as the preinitiation complex and is necessary for initiation. If this complex cannot form, RNAP cannot transcribe DNA into mRNA. In the cell nucleus, DNA exists in a continuum of states between a highly condensed form known as heterochromatin and an uncondensed form known as euchromatin. It is thought the more condensed heterochromatin state impedes the ability of RNAP and its transcription factors to achieve a stable complex and carry out transcription, thus providing a form of transcriptional regulation due to DNA inaccessibility.



Source: https://www.ck12.org/book/ck-12-biology/section/7.2/

Fig 2. DNA Transcription by RNA Polymerase.

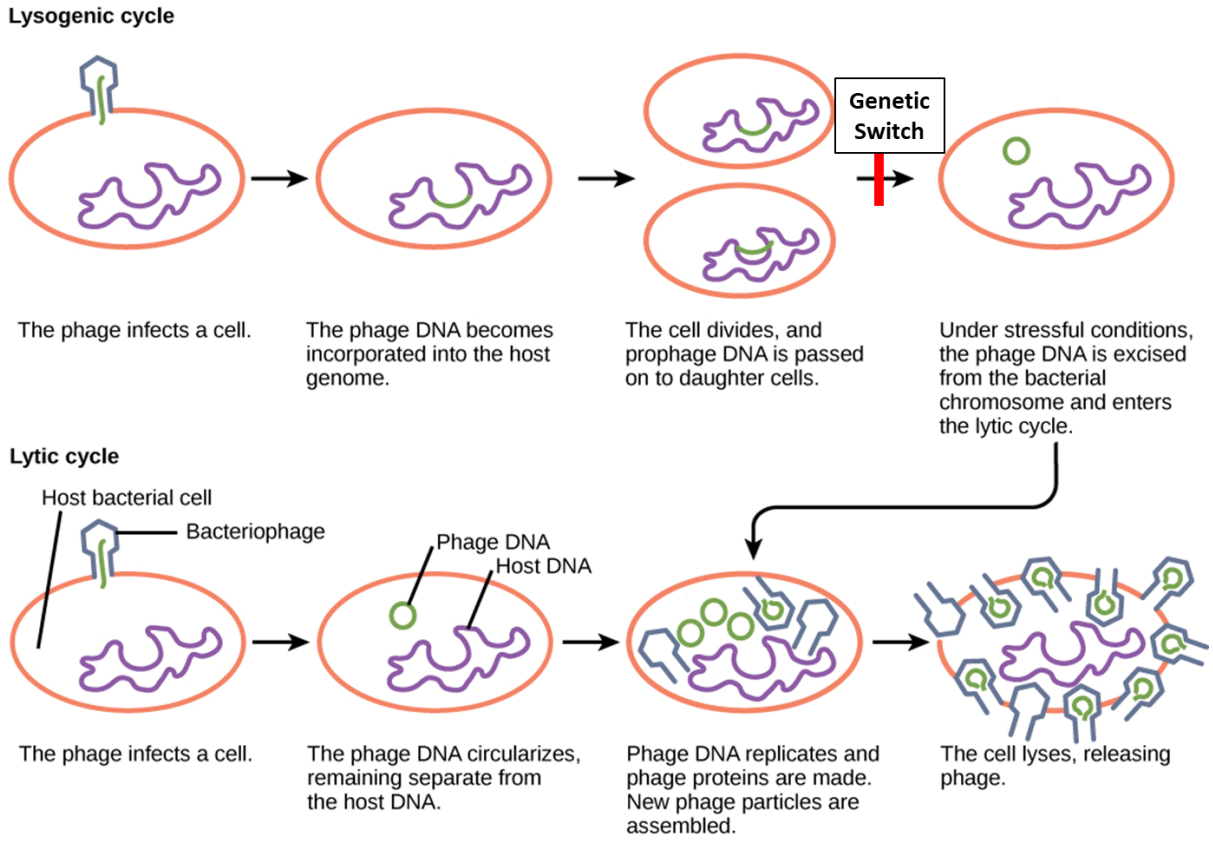
The three steps of DNA transcription where RNAP binds the promoter site during initiation, actively transcribes DNA into RNA during elongation, and reaches the terminator site and diassociates from DNA in termination. A transcription bubble can be seen during elongation. Source: Mariana Ruiz Villarreal (LadyofHats) for CK-12 Foundation, under Creative Commons.

The maintence of the heterochromatin state depends on the presence of nucleosome complexes formed by the binding of histone proteins to DNA. These histones form an octamer and wrap DNA about themselves. At sufficient concentrations, nucleosomes condense DNA into an inaccessible highly condensed state.

B. Bacteriophage 186 as Model for Transcriptional Regulation by Wrapping

Bacteria and bacteriophages represent excellent model systems to study transcriptional regulatory mechanisms because they are much simpler than eukaryotic systems. As prokaryotic organisms, bacteria are unicellular, do not contain a nucleus, and are less dynamically complex than most eukaryotic organisms.

Bacteriophages are viruses that infect bacteria and archaea to replicate themselves with the host cell's machinery. After host infection, the bacteriophage may follow two developmental pathways for its reproduction, known as the lytic-lysogenic decision (Fig. 3). In the lytic pathway, the bacteriophage DNA is replicated with bacteria machinery and new virions are rapidly synthesized until a critical mass is achieved. Once achieved, specialized viral proteins are used to dissolve the cells walls, and eventually, the cell bursts due to high internal osmotic pressure which can no longer be counteracted by the weakened cell wall. The virions are then released into the environment to where they can potentially infect new cells and repeat the process. In the lysogenic pathway, the bacteriophage DNA, known as the prophage, incorporates itself into the host genome. The host cell then continues to function and reproduce normally while transmitting the prophage down its progeny until an external stimulus triggers the switch to the lytic cycle. Barring any regulatory mechanism, the lytic cycle is the default pathway.



Source OpenStax (Creative Commons): https://openstax.org/books/biology-2e/pages/1-introduction

Fig 3. Diagram of Bacteriophage Reproductive Cycles. Source: Adapted from OpenStax, Biology 2e, under Creative Commons.

In the well-studied lambda phage, the pathway decision is governed by the λ CI protein which represses active transcription of sites that encode proteins necessary for the lytic cycle by steric hindrance reinforced by protein-protein cooperativity via looping (Zurla et al. 2009)(Fig. 4). Bacteriophage 186 uses the 186 CI repressor protein, which regulates the same decision function as the lambda CI protein but utilizes a fundamentally different mechanism (Wang et al. 2013).

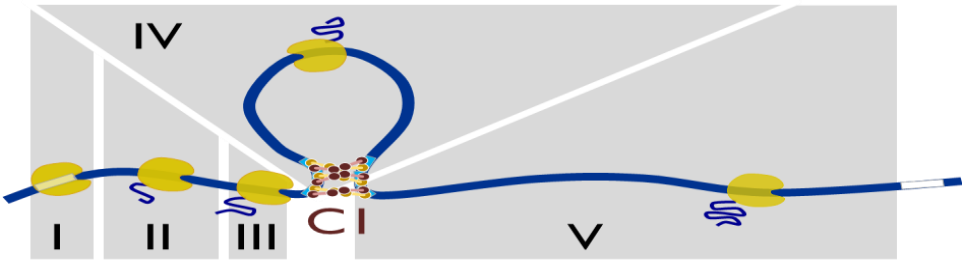


Figure 4. Example of DNA Looping by λ CI Protein. Looping by the λ CI protein that can bring two distant regions of DNA together to form a loop and provide a transcriptional roadblock to an elongating RNAP shown in yellow. Source: Vörös et al. (2017), under Creative Commons.

Most notably, 186 CI acts in a remarkably similar way to the histone proteins found in eukaryotic organisms, that is, by utilizing multiple protein dimers in the formation of a wheel-like protein complex around which DNA is wrapped. 186 CI primarily wraps DNA around itself by binding to different sites to form a complex with DNA (Fig. 5A).

The primary binding site of 186 CI is pR, where the promoter site for lytic expression is located. The FR and FL distal regulatory sites are secondary binding sites for 186 CI and possess lower binding affinities for it. Finally, the pL promoter site, where the promoter site for expression of the 186 CI protein, which maintains lysogeny, is located, is not known to be bound to 186 CI sites in the absence of pR (Shearwin, Dodd, and Egan 2002). Using the distal regulatory sites, 186 CI can both wrap or loop DNA (Wang, et al. 2013) (Fig. 5B). 186 CI acts as a genetic switch in the lytic-lysogenic decision by blocking RNAP from binding to the pR promoter when bound to the pR sites, thereby primarily regulating transcription via the inhibition of initiation.

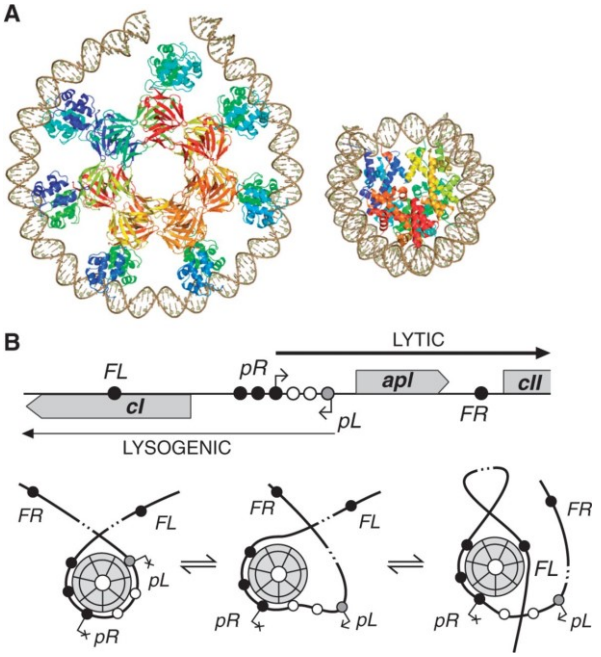


Figure 5. Structure of 186 CI Repressor Protein Complex.

(A) (Left) 186 CI protein is a seven-dimer protein arranged in a wheel complex with DNA. (Right) A histone-DNA complex to scale with respect to the 186 CI protein. (B) The different binding sites of the 186 CI protein which can switch between different conformations. The FL site was removed in this study to reduce the complexity of possible conformations. Source: Wang, et al. (2013), under Creative Commons.

C. DNA Wrapping as a Common Regulatory

Transcriptional regulation is a fundamental process which allows differential protein expression. Thus, DNA is decorated with transcriptional factors and other proteins. These proteins, including those mediating long-range interactions, such as looping and wrapping, are able to act as roadblocks to an incoming RNA polymerase. Previous studies have focused on loops as roadblocks on active transcriptional elongation (Vörös, et al. 2017; Hendrickson 2018)(Fig. 4). Wrapping is also a ubiquitous mode of protein-DNA interaction, mostly seen in the nucleosome

complexes made up of histone-wrapped DNA. While much study has been devoted to the effects of histone-induced repression of the transcription initiation, histones undergo complex post-transcriptional modifications that set them apart from other transcriptional factors and make them more difficult to isolate and study the wrapping mechanism using single molecule techniques. Instead, the 186 CI protein is a simpler model that captures the main features of DNA wrapping proteins.

II. Principles of Atomic Force Microscopy

Atomic Force Microscopy (AFM), a type of scanning probe microscopy, is a microscopy technique which allows imaging beyond the resolution limit of visible light. This high-resolution technique measures the perturbations of a tip running along a sample surface in a raster pattern as it encounters forces created by proximity to the sample surface (Binnig, Quate, and Gerber 1986), and generates a 3D, topographical image (Fig. 6). The AFM is powerful enough to achieve lateral resolutions of ~0.5 nm and vertical resolutions of ~0.1 nm without special treatment of the sample, making it ideal for the imaging of active biomolecules in their native state (Engel and Müller 2000) for single molecule experiments.

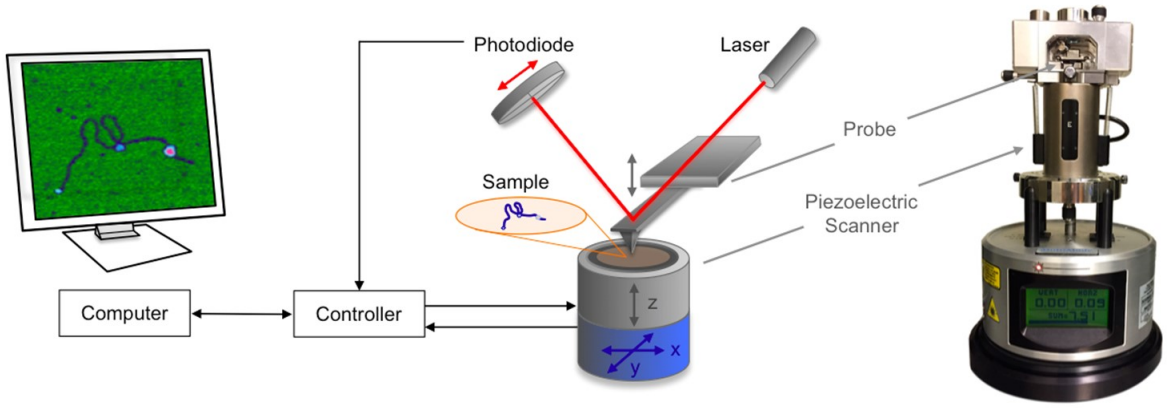


Figure 6. Diagram of Atomic Force Microscopy. As a tip runs over the surface, perturbation data is captured by the photodiode and sent to the controller to adjust the surface-tip separation in real time. Ultimately, the data comes together to make a topographical image seen on the left. Courtesy of Zsuzsanna Vörös.

The tip is located at the end of a flexible cantilever (Fig. 10) which is affixed to a piezoelectric actuator (Jalili and Laxminarayana 2004). The actuator keeps the tip at an oscillatory steady state at or near its resonant frequency using a feedback mechanism, thereby regulating the

separation of the tip and the surface. The tip's deflection from the surface, caused by forces such as Van der Waals forces or coulombic forces, is measured by recording the deflection angle of a laser with a photodiode after the laser is bounced off the cantilever. Feedback from the tip's deflection is not only measured, but ultimately fed back to the actuator to maintain either a constant force or a constant height. A constant force scan allows for the recording of height deviation, while a constant height scan allows for recording of the force on a sample (Jalili and Laxminarayana 2004).

A. Scanning Modes

Three general imaging modes are used with AFM. The first, and most intuitive, is the AFM contact mode where the tip is in constant contact with the surface as it is dragged along the sample. The tip must usually be in strong contact with the surface to be repulsed and counteract the relatively strong attractive forces between it and the surface. Side effects arise from this need for strong contact, including problems with the tip becoming stuck on the surface, sample damage, and the effect the mode has on an AFM tip, namely tips prematurely wearing down and needing replaced sooner relative to other modes.

There is also an AFM non-contact mode where the tip does not actually contact the sample. The AFM cantilever, and the tip on it, is kept at or near its resonant frequency as it experiences Van der Waals forces due to the proximity of the surface. At about 50 – 150 Å above the surface, the forces create the tip deflection necessary for measurement (Jalili and Laxminarayana 2004). This mode is particularly helpful in measuring fluidic surfaces where tip contact would break through the surface. This type of imaging also puts less stress on the tip, so it does not require tip

replacement as often. However, the forces acting on the tip are much smaller so there is a weaker feedback signal which leads to a tradeoff of inferior resolution relative to contact mode.

Finally, there is the AFM tapping mode, a key advance in AFM technology and an intermediate between contact and non-contact mode. In tapping mode, the actuator keeps the tip at or near its resonant frequency while in proximity to, and without contacting, the sample surface. Periodically, the tip also taps and makes contact with the surface, typically between 50 kHz – 500 kHz (Jalili and Laxminarayana 2004). Tapping causes much less damage to both the tip and the surface than in contact mode, thus allowing for imaging of soft materials without damaging the sample. This is because the impulse on both the tip and surface is minimal as the time the tip spends on the surface is minimal. It also has the advantage of providing higher-resolution images than non-contact mode because the tip, in tapping mode, experiences higher forces when it taps the surface, thus allowing bigger perturbations to be picked up by the scanner.

B. Advantages and Disadvantages of AFM

AFM is an extremely powerful technique which offers the opportunity to obtain high-resolution, 3-dimensional topographical images of a sample at the molecular level. It does not require any special sample treatment or environment, unlike other microscopy techniques requiring special vacuum environments, low temperature, or special chemical treatment which may damage the sample. The ability to image in ambient air or fluid makes AFM an advantageous method when studying biological phenomena, especially when paired with the AFM's additional ability to be run in parallel with other microscopy or spectroscopy techniques such as infrared spectroscopy or Raman spectroscopy.

However, compared to other microscopy techniques, AFM can be time intensive and area limited as it requires a tip to go through the entire area, sometimes taking hours for a high-quality scan. Whereas AFM has a maximum scanning area of 150 x 150 μm^2 under certain conditions with specific instruments, an electron microscope can take a scan on the order of millimeters in an instant. AFM also requires extremely well, vibrationally-isolated setups to get high-quality images at the atomic scale because of the nature of the atomic forces being measured by the technique. Any small vibrations on top of these small forces is enough to greatly interfere with image signal.

III. DNA Modification

DNA with the T7A1 promoter and lambda T1 terminator sites as well as the bacteriophage 186 CI binding sites pR, pL, and FR was used for transcription experiments. Molecular cloning was carried out to generate a plasmid with the desired sequence without the FL site to reduce the complexity of possible 186 CI conformations (Fig. 5). Final recombinant DNA used in the measurements (Fig. 7) was amplified from the final recombinant plasmid by PCR (Fig. 9).

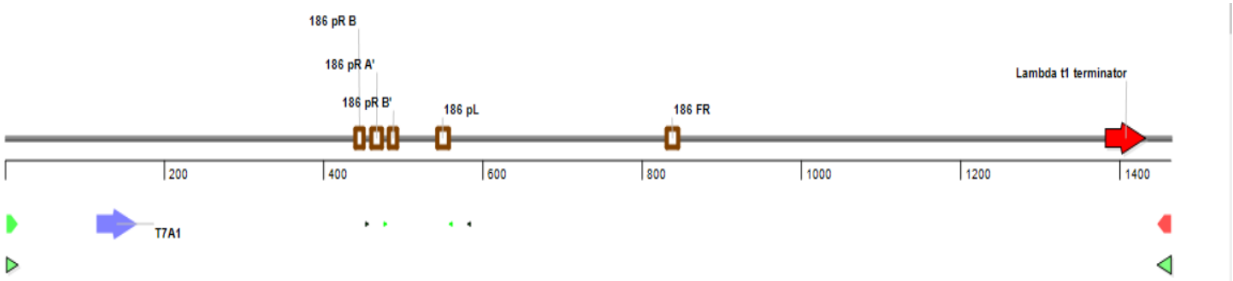


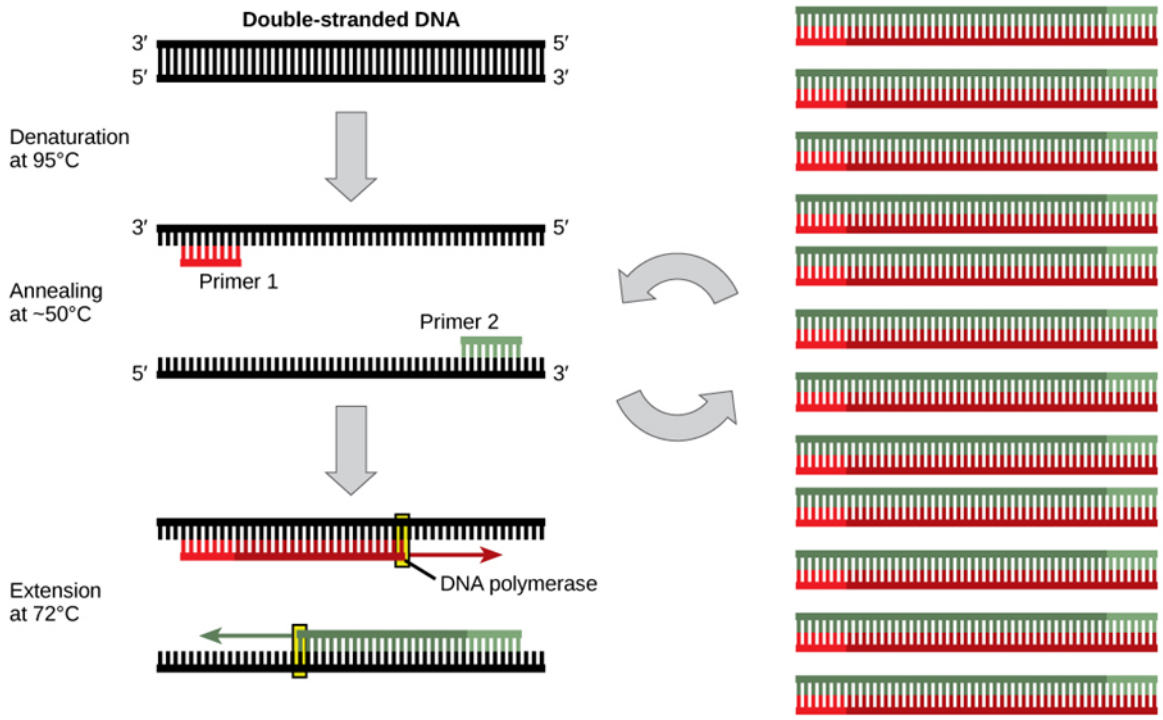
Figure 7. DNA Construct with pR+pL+FR. Final product of molecular cloning yielded this DNA construct used in experiments. Image constructed with DNASTAR Lasergene 17 software from DNASTAR, inc.

A. Polymerase Chain Reaction Extracts Desired Sites

Polymerase Chain reaction (PCR) is a fundamental molecular biology technique used to exponentially amplify and isolate a specific sequence of DNA (Mullis and Faloona 1987). PCR was used both to isolate a DNA insert containing the pR, pL, and FR sites as well as for the amplification of the recombinant DNA molecules used.

The necessary components of a PCR are a DNA template with desired region of DNA, DNA polymerase, a forward and a reverse single-stranded DNA primer, deoxyribonucleotides (dNTPs), and buffer solution (NEBuffer™ 3.1, New England BioLabs). The process is primarily carried out in a series of three temperature-dependent steps inside a thermocycler (Fig. 8). In the first step of PCR, denaturation, the reaction temperature is brought to ~96°C in order to denature

the double-stranded DNA template into two single-stranded DNA molecules by breaking the hydrogen bonds between complementarily pairs with heat. In the next step, annealing, the reaction temperature is lowered to an annealing temperature dependent on the sequence of primers being used. During the annealing phase, the primers attach themselves to either side of the insert on opposite strands of the DNA. Primers are chosen to have a high percentage of triple G-C bounding pairs (Fig. 1C) so bonds are thermodynamically stable at higher temperatures. Finally, in the last step, the temperature is raised to an elongation temperature dependent on the DNA polymerase being used. Typically, *Taq* polymerase is used for its thermostability at biologically high temperatures, so the elongation temperature is ~70°C. While *Taq* has low specificity for initiating DNA for transcription, it requires both sufficiently high temperatures and regions of double stranded DNA. Therefore, it'll have high specificity for sites bound with thermodynamically stable primers at high temperatures.



Source: https://openstax.org/books/concepts-biology/pages/1-introduction

Figure 8. PCR Schematic. Cyclic cycle showing double stranded DNA being denatured at high temperature, then annealed with DNA primers on either end at a low temperature, and finally extended to duplicate each complementary strand at a medium temperature. Multiple cycles result in a massively duplicated DNA sample. Source: Adapted from OpenStax, Concepts of Biology, creative commons.

The whole three-step cycle can be done in less than two minutes and is normally run about forty times. After the main cycling is finished, final elongation and long annealing steps are carried out to ensure association of DNA in its double-stranded form. The finished product is an exponentially amplified sequence of DNA.

For the purposes of this experiment, a 5’ biotinylated primer was used to label the DNA molecule at the end furthest to the T7A1 promoter with a protein marker, streptavidin. Streptavidin is a small molecule, easily discernible from the bigger RNAP or 186 repressor proteins. The two proteins have an incredibly high binding affinity, on the order of femtomolar, making it one of the

strongest organic non-covalent bonds and an invaluable asset for biological assays requiring a simple labeling molecule (Weber et al. 1989).

B. Recombinant Plasmid

After a PCR reaction extracted an insert containing the pR, pL, and FR sites, molecular cloning was used to place the insert inside a vector plasmid containing the T7A1 promoter, lambda T1 terminator, and ampicillin resistance (Fig. 9).

Restriction enzymes are a type of endonuclease that recognize a particular sequence of DNA, bind to it, and cleave the sugar-phosphate backbone. To begin, both the vector and insert were digested by two of these restriction enzymes to generate overhangs on both molecules where they could hydrogen bond to each other before being permanently attached. Then, the now cut plasmid was incubated with a phosphatase to remove the phosphate groups at the cut sites. Tri-phosphate groups are required on the 5'-end of two DNA strands being ligated, or joined, together, so the phosphate removal prevents the original plasmid from being ligated into its original state. Finally, the vector and insert were incubated together with ligase to be joined together.

At this point, the yield of recombinant plasmid is low, so a bacterial transformation was carried out to ultimately use bacterial machinery to massively replicate the recombinant plasmid. Heat-shocking bacteria in warm solution containing divalent cations make their membrane much more permeable and makes it possible for the recombinant plasmid to be taken up. After heat-shock, bacteria were plated on ampicillin-containing plates. Only bacteria with ampicillin resistance from the complete recombinant plasmid survive and form colonies on the plate. Those colonies were then allowed to grow in a high nutrient broth to later lyse and purify to retrieve the recombinant plasmid.

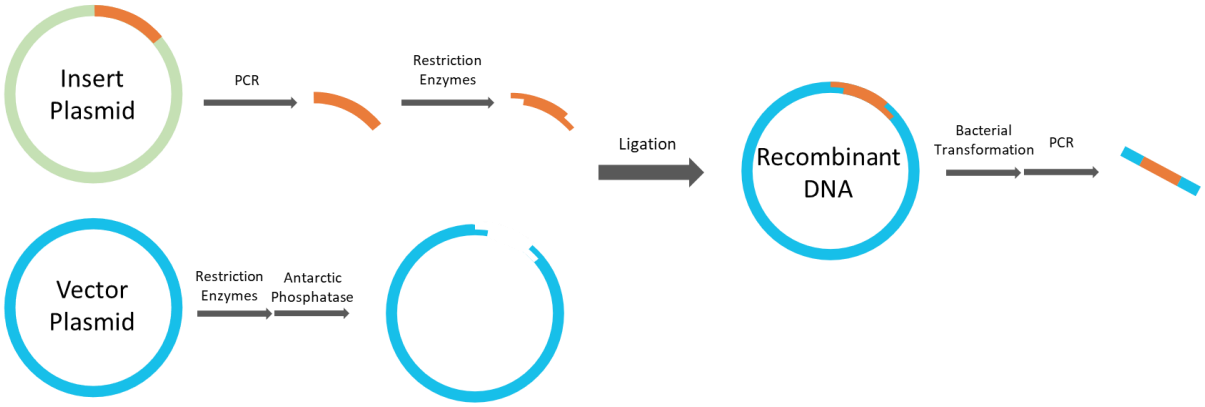


Figure 9. Schematic of Molecule Cloning. A vector plasmid was cut open and dephosphorylated to prevent closure on itself. A DNA insert from an insert plasmid was taking using PCR and ligated with the vector plasmid to form the recombinant DNA which was subsequently massively duplicated in a bacterial transformation. Finally, the desired DNA construct was isolated using PCR.

IV. Materials and Methods

A. Genetic Recombination

PCR was used to generate a DNA fragment for insertion into a plasmid. The insertion was amplified from plasmid (pBS_HSL_FL+pR+pL+FR, Keith Shearwin) containing the FL, pR, pL, and FR binding sites for 186 CI. A fragment containing the pR, pL, and FR sites was amplified and purified. A plasmid (pRS_IN_400, Raven Shah, Finzi-Dunlap Lab) with the T7A1 promoter and lambda T1 terminator sites was digested with restriction enzymes (Acc 65l, Xhol, New England Biolabs, Ipswitch, MA) to isolate a large vector containing the origin of replication, terminator site, and ampicillin resistance features. The plasmid was later dephosphorylated with Antarctic Phosphatase (New England Biolabs, Ipswitch, MA) to prevent ligation of the plasmid fragments. The vector and insertion were ligated to form a recombinant plasmid (Appendix A) with the 186CI binding sites located between the promoter and terminator sites for transcription experiments. A bacterial clone of the plasmid was produced by transformation in E. coli to provide a convenient, abundant source of DNA for experiments. A linear fragment of DNA was amplified from the recombinant plasmid using primers (Appendix B. S/pBR322/2211; A/pBR322/3728) in a PCR to obtain a final sequence length of DNA with 1465 base pairs, approximately 480 nm (Fig. 7).

B. Sample Preparation

Two types of transcription buffers were used. Initially, an acetate buffer was used in experiments without streptavidin (5 mM HEPES, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM DTT, diluted in HPLC). To confirm correct positioning of 186 CI and RNAP, streptavidin was used with a biotin binding site at the downstream end toward which RNAP moves

during transcription. Streptavidin did not bind well in acetate buffer, so a glutamate buffer was substituted as the transcription buffer in samples with streptavidin (20 mM tris-glutamate, 50 mM potassium glutamate, 20 mM magnesium glutamate, 1 mM DTT, diluted in HPLC).

Mica surface was cleaved with adhesive tape, then 20 µL of 0.01 µg/mL of poly-L-ornithine (~1 kDa MW, Sigma-Aldrich, St. Louis, MO) was deposited on surface. After 2 minutes, the poly-L-ornithine was washed off the surface with 1 mL high-performance-liquid-chromatography grade water (HPLC). Treated mica was then gently dried with compressed air.

DNA and transcription buffer were mixed inside PCR tubes. Proteins were added according to experimental designs and were then set to incubate for 23 minutes at 37ºC. Concentrations are in terms of the final 20 µL reaction sample before dilution. Samples were then spiked with 100 µM NTPs and left to incubate for 2 minutes at ~37ºC. 20 mM EDTA was then added to halt transcription without dissociating the RNA polymerase from the DNA. 20 µL of reaction sample was then diluted by 1.5-fold, deposited on prepared poly-L-ornithine-coated mica, and left to sit for 2 minutes. Mica was then washed with 1 mL of HPLC and gently dried with compressed air. While not all samples had all components, the components were used at the following concentrations: DNA at 1 nM, Streptavidin at 100 nM, 186 CI at 500 nM, and RNAP holoenzyme at 6 units/µL (New England Biolabs, Ipswich, Ma). Final RNAP holoenzyme concentration equates to a 167x dilution of the 1000 units/µL stock solution in units of activity as specified by the manufacturer.

C. AFM Image Scan

Images were acquired with a Nanoscope Multimode 8 AFM microscope (Bruker). PeakForce Tapping Mode with automated feedback control, Scanasyst, was used. Probes were

primarily Scanasyst-Air cantilevers (Bruker, Camarillo, CA) with a nominal spring constant of 0.4 N/m, but HQ:NSC18/Al BS cantilevers (Mikromasch, Watsonville, CA) with a nominal spring constant of 0.5 N/m were used in some experiments.

The Bruker cantilevers had a nitride triangular bridge with a silicon tip resting at the end (Fig. 1A-B). The Mikromasch cantilevers had a straight aluminum bridge with a silicon tip at the end (Fig. 1C-D).

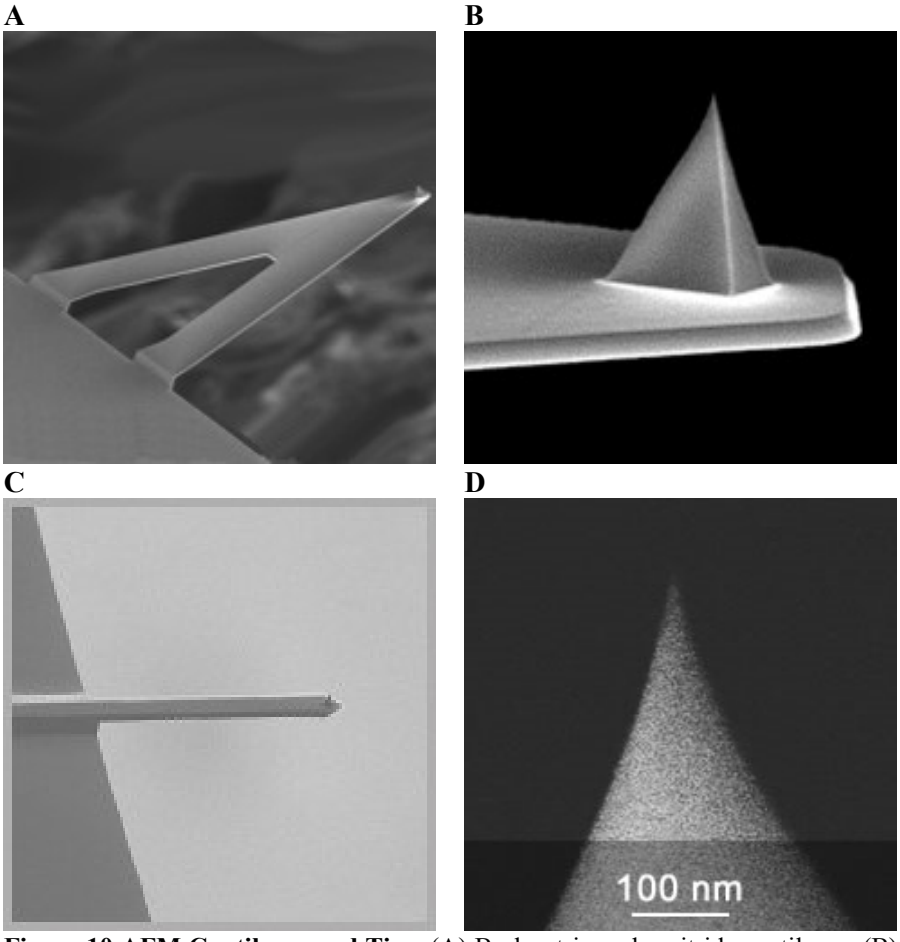


Figure 10 AFM Cantilever and Tips. (A) Bruker triangular nitride cantilever. (B) Bruker Scanasyst-Air Silicon Tip: T = 650 nm, L = 115 µm, W = 25 µm, f_0 = 70 kHz, k = 0.4 N/m. (C) Mikromasch aluminum cantilever. (D) Mikromasch Silicon tip: f_o = 65 kHz, k = 0.5 Source: Images reprinted from Bruker Nano Surfaces and Metrology; Nanoandmore with permission.

Multiple areas for each sample, ranging from 1 x 1 μm^2 to 5 x 5 μm^2, were scanned. Resolution was set for 512 x 512 pixels per μm^2 scanned, i.e. a 512 x 512 resolution for a 1 x 1 μm^2 area and a 2048 x 2048 resolution for a 4 x 4 μm^2 area. For the purposes of preserving image quality, the tip speed was generally maintained low at 1.25~3.00 μm/s depending on image scan and were generally higher for smaller images. For a 3 x 3 μm^2, tip speed was kept at ~2.5 μm/s, while for a 4 x 4 μm^2, tip frequency was kept at ~1.75 μm/s. An example image is shown (Fig. 11).

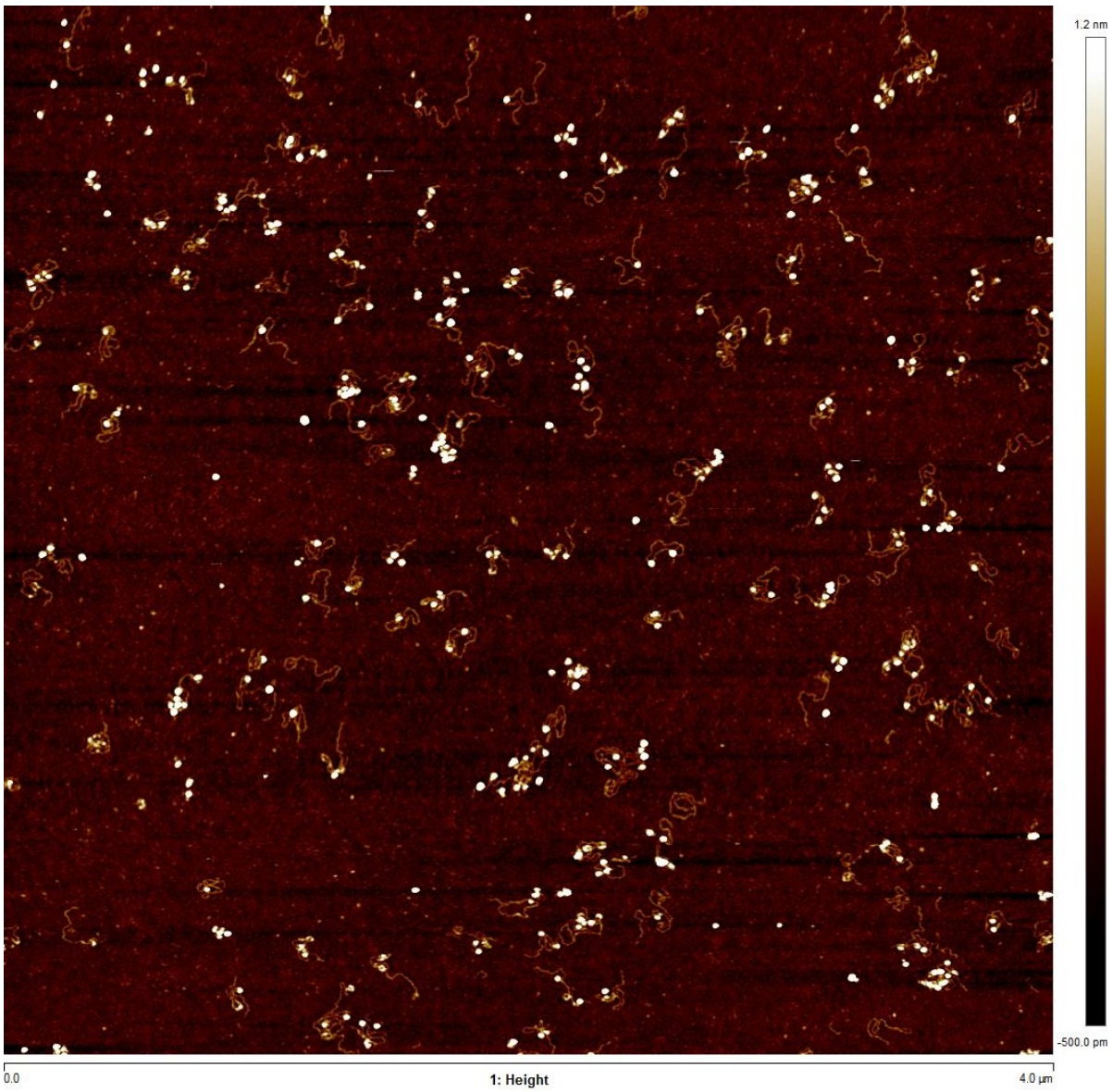


Figure 11. Example AFM Image. Example of an AFM image containing both 186 CI and RNAP proteins without streptavidin. Coloring of individual pixels, as denoted by the color scale on the right, denotes height of the pixel in the image.

V. Image Processing and Analysis

Images were analyzed using MATLAB® scripts. The goal of programming was to automatically identify proteins, process data statistics of individual proteins, and detect the position of protein on a DNA strand. All data was originally in terms of pixel units and was converted to physical metric units using a pixel scale obtained from image data. GitHub information for code can be found in Appendix C.

A. Tracer GUI

Previously, an algorithm coded in MATLAB® was developed to trace curvilinear features in topographical images of molecules on a surface, in this case, AFM images of DNA on polyornithine-treated mica (Wang 2011). Later, a graphical user interface (GUI), the program titled *DNAtracer*, was assembled to let a user manually edit the automatic traces and add new ones on an image surface (Dan Kovari, Ph.D., Finzi lab, Emory University, unpublished). Segments below a minimum length are discarded to avoid including background artifacts; the default threshold used was 200 nm. The algorithm itself performs well on molecular contours with relatively smooth topography along their surfaces; however, abrupt height changes along the contour cause the algorithm to stop the trace and break a molecule, or several molecules, into segments labeled as one whole broken molecule. The most common abrupt height change along the DNA contour were those associated with a relatively large protein, as compared to the rest of the sample. Since it was tedious to reconnect the segments of each molecule together, the GUI was used to manually trace out contours of molecules without the automated DNA tracing algorithm.

The GUI itself allows the user to place control points along the molecule. Curves are then interpolated between control points via the Catmull-Rom method where a polynomial is fitted to

four local control points in a recursive fashion until the entire trace is drawn (Barry and Goldman 1988). This method produces an accurate representation of the trace along the contour which yields the total arc length used for each DNA molecule (Fig. 12). Also used was a simpler trace for quick calculations based on the MATLAB® function *pchip* which interpolates between control points using the piecewise cubic hermite interpolating polynomial method (Fig. 12). While simple to implement because of the built-in MATLAB®, it is only used to get a weight position because actual length of *pchip* trace is inaccurate by up to ~100 nm, however, error is evenly distributed throughout the trace.

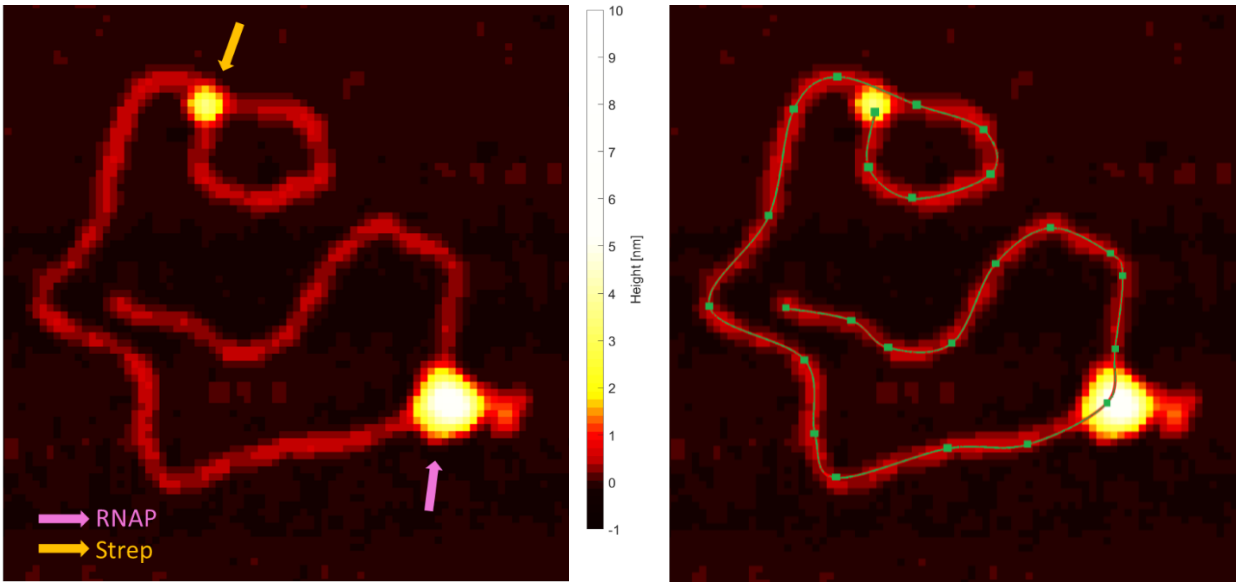


Figure 12. Molecular Trace. Tracing of a DNA molecule where image on the left has a yellow arrow denoting streptavidin and a pink one denoting RNAP with nascent RNA hanging off it. Image on the right is the molecular trace where trace is wholly defined by the control points along the molecule.

B. Particle Analysis

Locating protein particles along DNA contours required image processing in several, sequentially performed steps. First, using Bruker Nanoscope Analysis 1.9 software, a raw AFM image with file extension "SPM" is flattened (Fig. 12A). The flattening process removes low

frequency noise, tilt, and bow from each image line by individually fitting a polynomial of nth order using a least-squares fit method to each line (Bruker Bruker_Corporation 2011). The polynomial is then subsequently subtracted from its corresponding line. Each increasing nth order polynomial subtracts a different feature from the image (Table 1). Higher-order polynomials work to fit and subtract the bow with higher nth polynomials.

nth Order	Polynomial Form	Feature Removed
0	$z(x) = a$	• Centers each line at z = 0, removing image offset
1	$z(x) = ax + b$	• Centers each line at z = 0, removing image offset • Centers the slope of each line, removing the image tilt
2	$z(x) = ax^2 + bx + c$	• Centers each line at z = 0, removing image offset • Centers the slope of each line, removing the image tilt • Removes bow of each line

Table 1. Image Flattening Methods.

The flattened AFM image is then passed over to MATLAB® where an algorithm creates a binary mask of the image (Wang 2011) to find curvilinear features. A radial blur is then applied to expand the effective area of each feature (Fig. 12B). In this process, all pixels connected to each expanded feature are converted into ones and any others are converted into zeros. When this new image is multiplied element-wise by the original image, the zeros cancel out the background noise so any proteins or other artifacts not associated with a DNA molecule in the original AFM image are discarded. A curvilinearly masked AFM image is created in this fashion to isolate the data on curvilinear features (Fig. 12C).

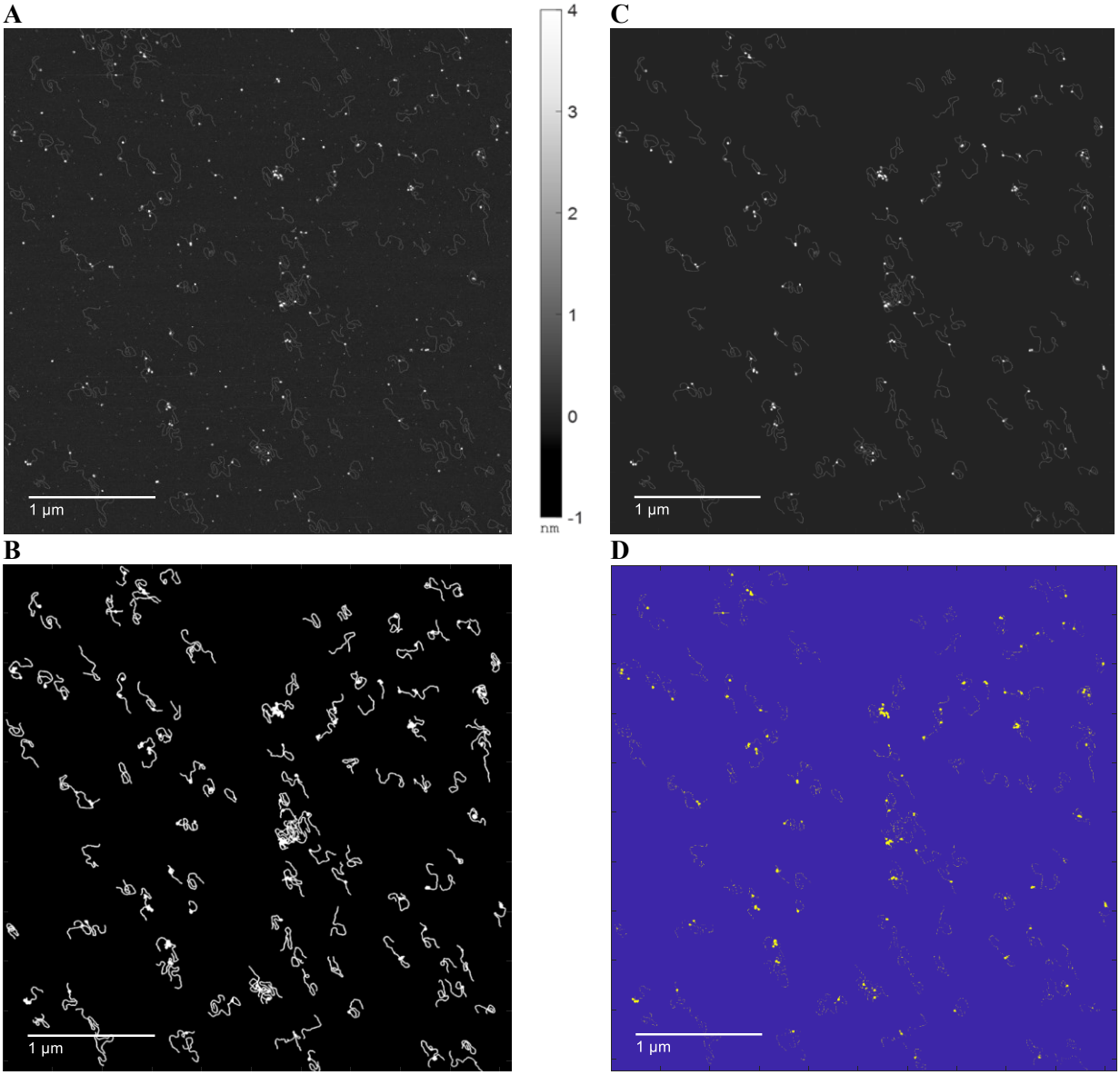


Figure 13. Particle Masks. (A) AFM image before masking. (B) Binary mask image using Dr. Haowei Wang's algorithmn. (C) AFM image with binary mask applied. (D) Binary image of particles.

A second binary mask is created by applying a height threshold, by default set to 0.9 nm, to the curvilinearly masked AFM image (Fig. 12D). All points above this threshold are converted to one, and all points below to zero. Only protein particles surpass this threshold and are included

in the binary mask as double-stranded DNA height is reported to be 0.4 ± 0.1 nm when using AFM tapping mode (Hansma et al. 1996). The MATLAB® function *regionprops* is applied directly to the binary height mask to automatically calculate statistics (Table 2) of continuous regions, particles, identified in the mask. Additionally, using the '*PixelIdxList'* property of *regionprops*, we can directly pull the index of particle regions from the original AFM image. This lets us calculate the maximum, median, and standard deviation of the height for individual particles and add them to the list of statistics of their respective particles.

Statistic	Description
Area	Total area of the region of interest (nm^2)
PixelIdxList	Linear pixel list that indexes the area, i.e., which pixels in the image make it up
Eccentricity	Eccentricity of the region of interest
EquivDiameter	Diameter calculated by $d = (4 * Area)/\pi$ (nm)
Centroid	Location of the center of mass of the region (x,y)
MaxFeretDiameter	Maximum ferret diameter of the region (nm)
MinFeretDiameter	Minimum ferret diameter of the region (nm)

Table 2. Values calculated by '*Regionprops*'

The statistics are then cleaned through thresholding: the particles must be within a minimum and maximum '*Area'* property of *regionprops* as well as below the max height calculated from the original AFM image and masks. This ensures large and small features overlapping with DNA molecules will not be considered in the final list of particle statistics. After this, final particle data is kept (Fig. 13)

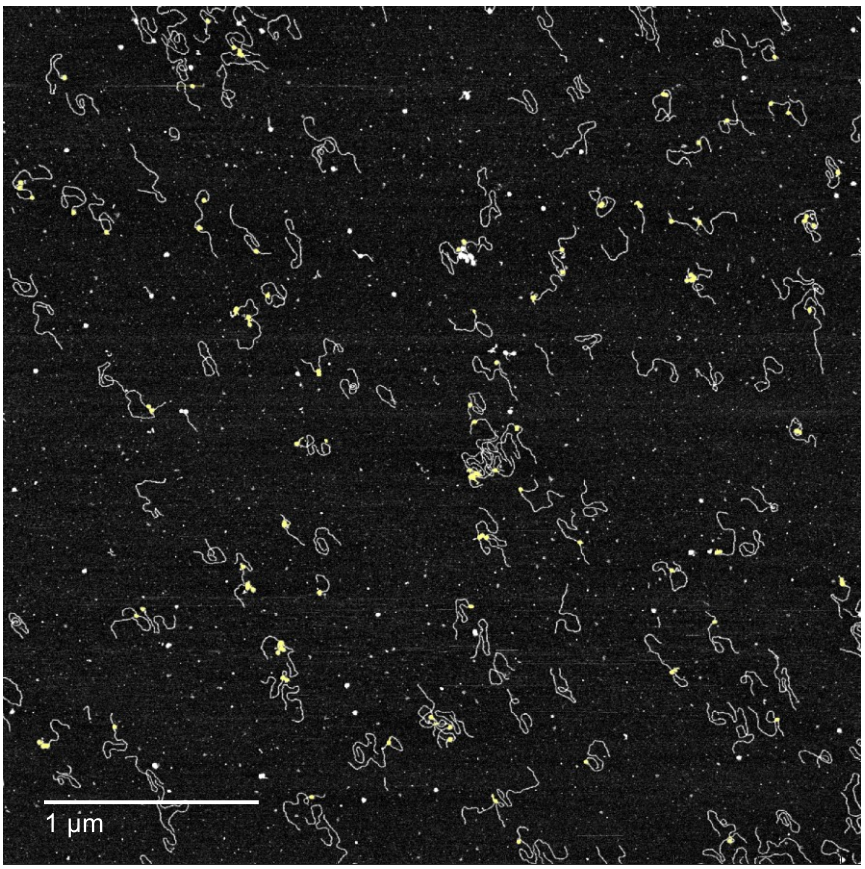


Figure 13. Protein Overlay on Original AFM image. Original image with proteins denoted by yellow blobs. Background noise, defined as anything that is not a DNA molecule, is excluded from image analysis using the methods outlined.

C. Linking Particle and Molecule Data

To determine the location of a particle along a DNA molecule, I considered the centroid of a particle. The original control points of molecules were used to create a bounding box to minimally circumscribe each DNA molecule (Fig. 14). If the centroid of a particle fell within a box, the particle was identified as a possible interaction with the molecule. Then, the minimum distance between the centroid and a point on the simple *'pchip'* trace of the molecule was found. The fractional position of the closest point with respect to the total length of the simple trace was then converted to a position along the Catmull-Roll spline curve found earlier. This position was reported as the position of the protein on DNA from the upstream start point.

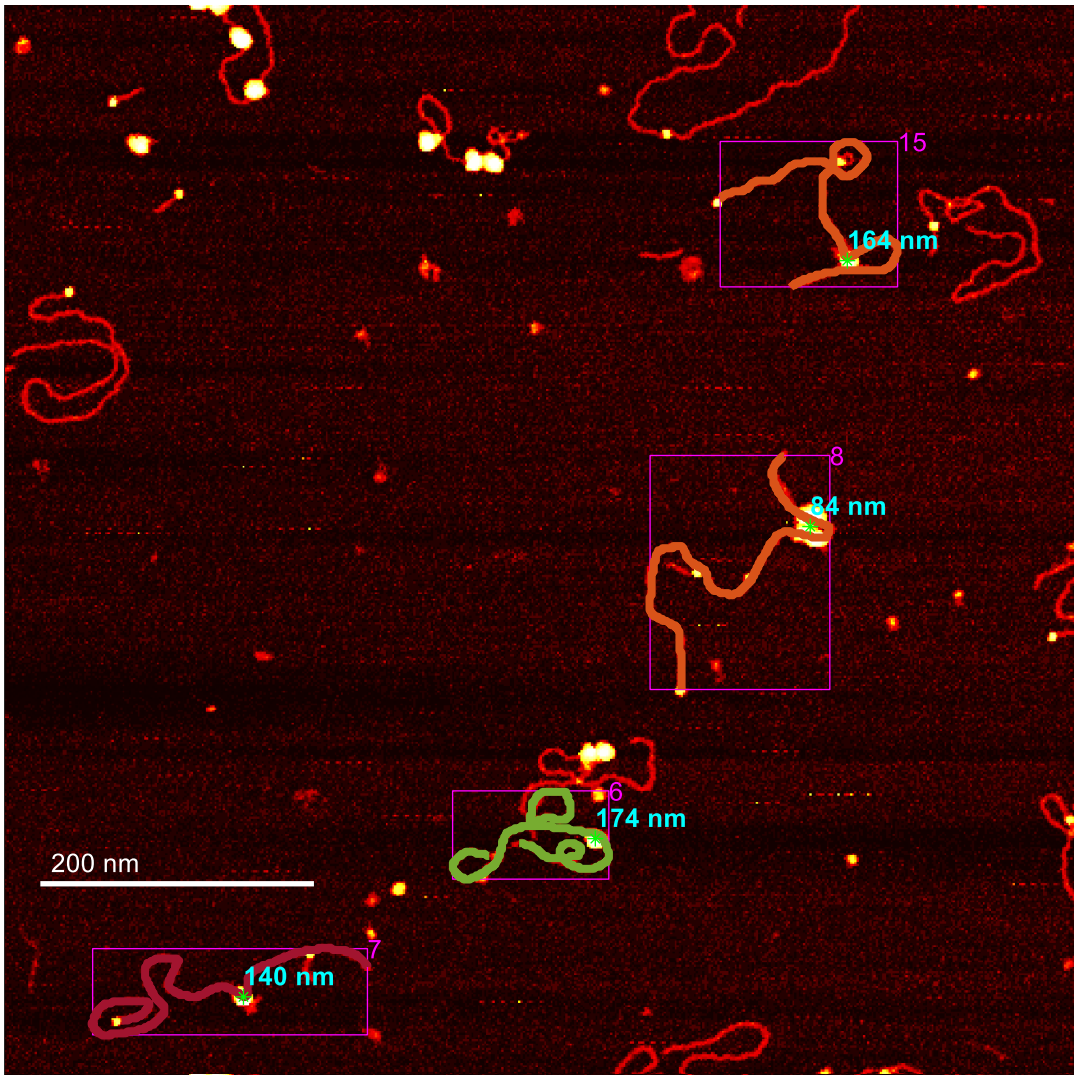


Figure 14. Boundary Boxes Ascertain Protein Position. Representation of algorithm used to link a protein particle to a DNA molecule where each DNA molecule has a bounding box which is compared to the location of the center of mass of each protein. The protein location is then compared against the trace to check for overlay between protein and DNA molecule since bounding boxes can overlap. Finally, the minimum distance from a trace pixel to protein location is denoted as the location of the protein along DNA.

VI. Results and Analysis

A. Control Imaging

186 CI and RNAP were imaged independently both in the presence and absence of NTPs and Streptavidin to establish reference points in the data and to act as controls for expected outcomes, such as proper binding position along DNA and transcriptional activity.

Binding position of 186 CI along DNA molecules showed a peak at 150 nm, as expected, which corresponds to 186 CI bound at the pR binding site, confirming proper binding of 186 CI to DNA (Fig. 15). Binding of RNAP to DNA in the absence of NTPs confirmed proper binding of RNAP to the promoter site T7A1, with a peak of ~50 nm.

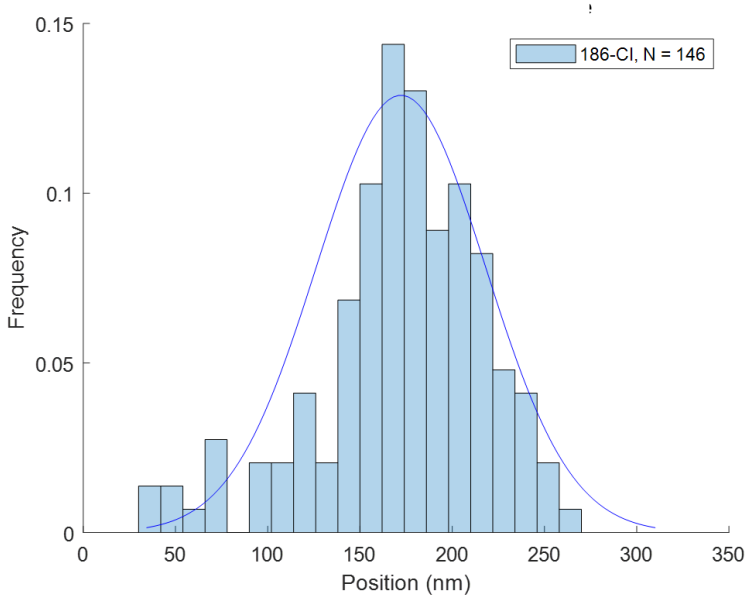


Figure 15. 186 CI Position along DNA. 186 CI position is confirmed to lie with a gaussian peak set at 150 nm which roughly corresponds to the pR binding site. Data was obtained from images of only 186 CI and DNA in the sample.

In control images of RNAP with NTPs and Streptavidin, we observed expected formation of elongation complexes (Fig. 16). In each instance of the complex, RNAP was seen halted at a different point in its progression along the molecule. In considering single instances as a continuous function, we can learn more about RNAPs progression along the Molecule on DNA.

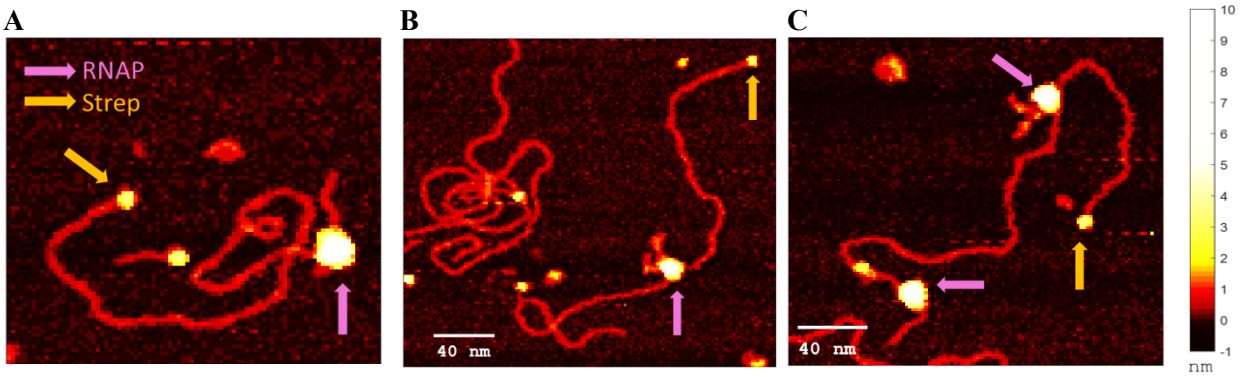


Figure 16. RNAP Progression along DNA under AFM. Example images of RNAP, denoted by pink arrow, progression along DNA with terminator side of DNA denoted by streptavidin, denoted by yellow arrow, captured in transcription samples without 186 CI. (A) RNAP, with a small amount of nascent RNA, near the promoter site just after initiation. (B) RNAP halfway through elongation with nascent RNA hanging off it. (C) Two RNAP on the same DNA molecule where one is nearing the terminator site and has a sizable amount of nascent RNA and the other is near the promoter site with no nascent RNA visible.

By considering the position of RNAP on a DNA molecule across different instances of the elongation complex, we can construct a cumulative probability curve demonstrating the probability of RNAP to progress past a point on a DNA molecule (Fig. 17). We use this graph as a reference in later analysis. Unsurprisingly, RNAP has the highest chance to be found near the start of its pathway, at the T741 promoter site.

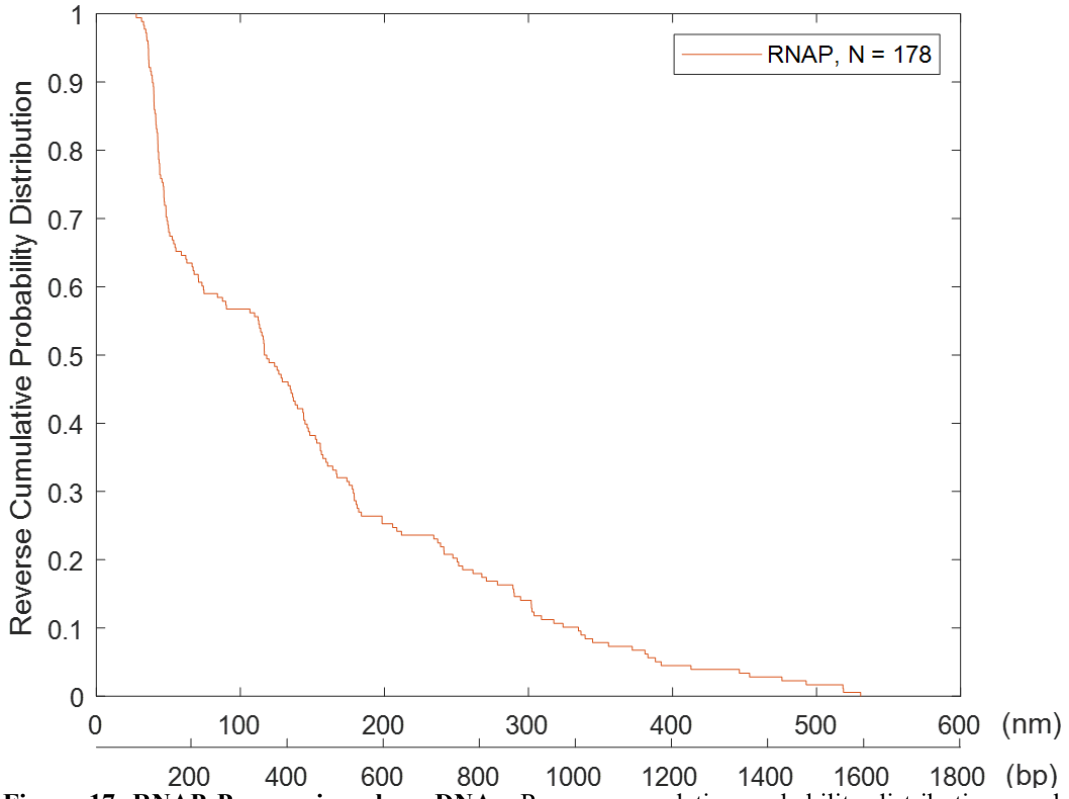


Figure 17. RNAP Progression along DNA. Reverse cumulative probability distribution graph showing the probability of RNAP progression past a point on DNA. Data was obtained from images of transcription without 186 CI as shown earlier.

B. Differentiating 186 CI and RNAP

Data on the height of protein particles along DNA attributed to RNAP and 186 CI were processed and analyzed to determine if the two proteins could be distinguished in subsequent analysis. Using the image processing and analysis methods described earlier, the mean height of each individual particle region was used to construct probability distributions (Fig. 18). The mean height was chosen because it showed the least variance in height data.

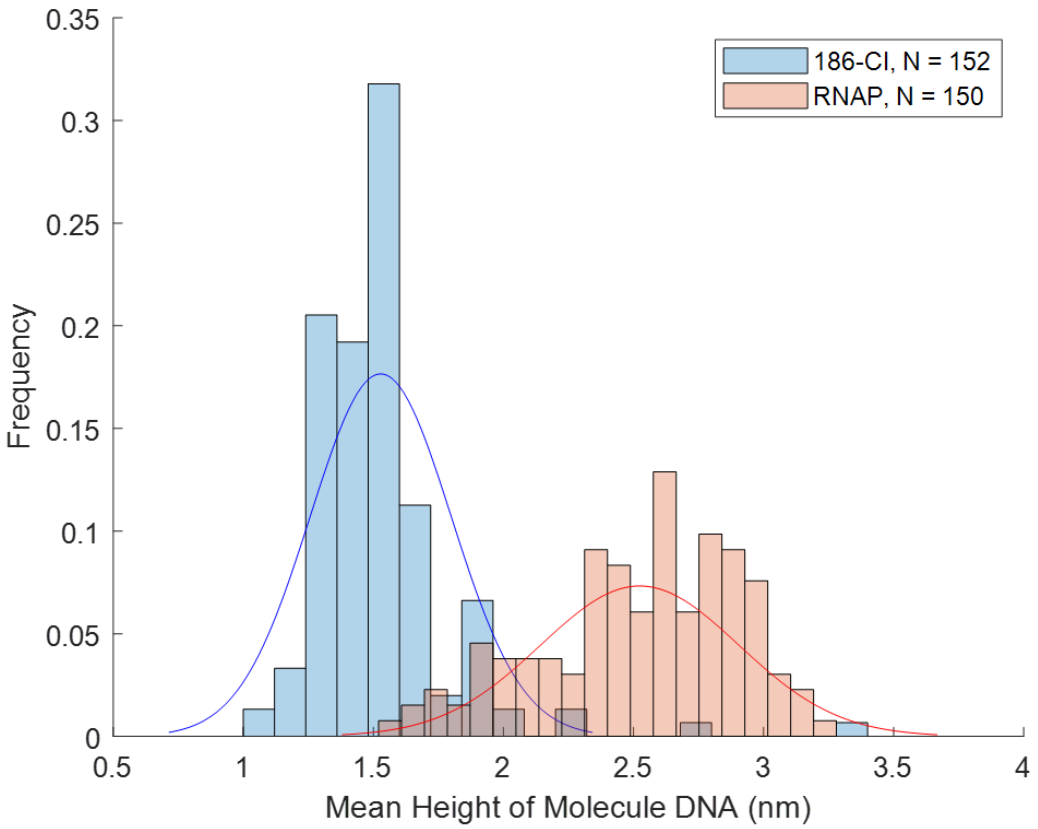


Figure 18. Mean Height Distribution of RNAP and 186 CI. Height distribution of the mean height of individual protein particles. Mean height of 186 CI, denoted in blue, had a distribution with $\mu = 1.528$ nm, $\sigma = 0.2711$ and RNAP had a distribution with $\mu = 2.5$ nm, $\sigma = 0.4$. Data was obtained from separate sample images of RNAP and 186 CI.

A gaussian curve was fitted to the individual probability distributions. 186 CI was shown to have a lower height, with its gaussian curve sitting at $\mu = 1.5$ nm, $\sigma = 0.3$. RNAP was shown to have the higher height, with a gaussian at $\mu = 2.5$ nm, $\sigma = 0.4$. When looking at high-contrast imaging of the two proteins, most samples could be differentiated based on the mean height (Fig. 19). However, when overlap occurred, secondary methods to confirm protein identity were presence of RNAP nascent RNA tail (Fig. 19A-B) and the wheel structure of 186 CI (Fig. 19C-D).

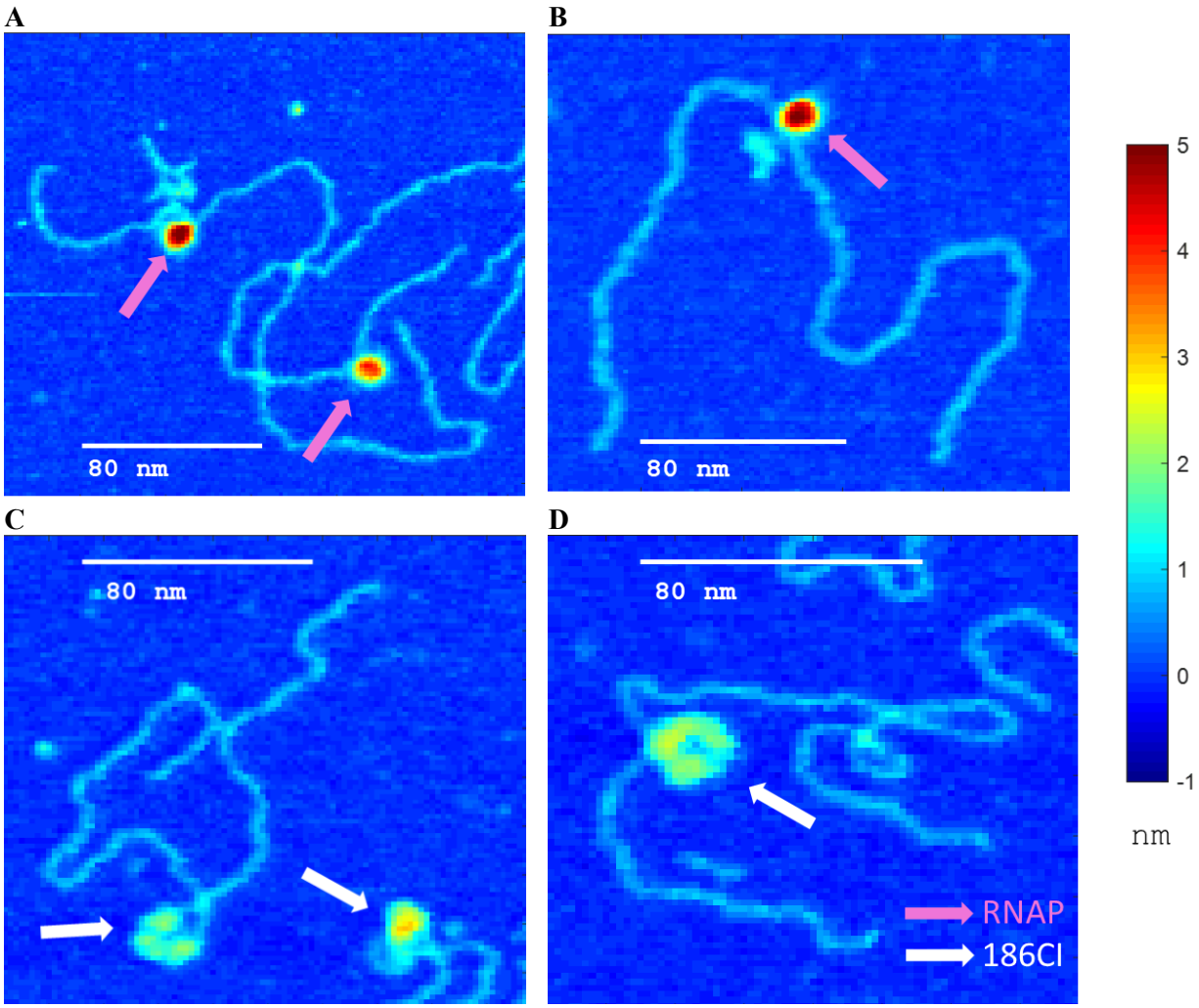


Figure 19. Differentiation of RNAP and 186 CI. Representative images of RNAP, denoted by pink arrows, and 186 CI, denoted by white arrows, showing differences between the two proteins. (A)(B) RNAP with nascent RNA sitting at a higher height as denoted by darker coloring based on the color scale on the right. (C)(D) 186 CI sitting at a lower height as denoted by lighter coloring. The wheel structure of 186 CI can also be made out based on the empty space within the protein-DNA complex.

While not always visible, 186 CI displayed the characteristic wheel structure as denoted by the formation of a loop around the protein by DNA and the empty space/drop in height within the complex.

C. Transcription Experiments

Obtaining images of active transcription with 186 CI proved to be more difficult than other imaging conditions. The images obtained were lower in quality than those of the controls; however, we were able to gather sufficient data to draw a conclusion on the effect of the 186 CI roadblock on RNAP elongation (Fig. 20). The lower quality of the images was the result of protein-DNA complex aggregations (fig. 20D) and dispersed fragments of unknown origin.

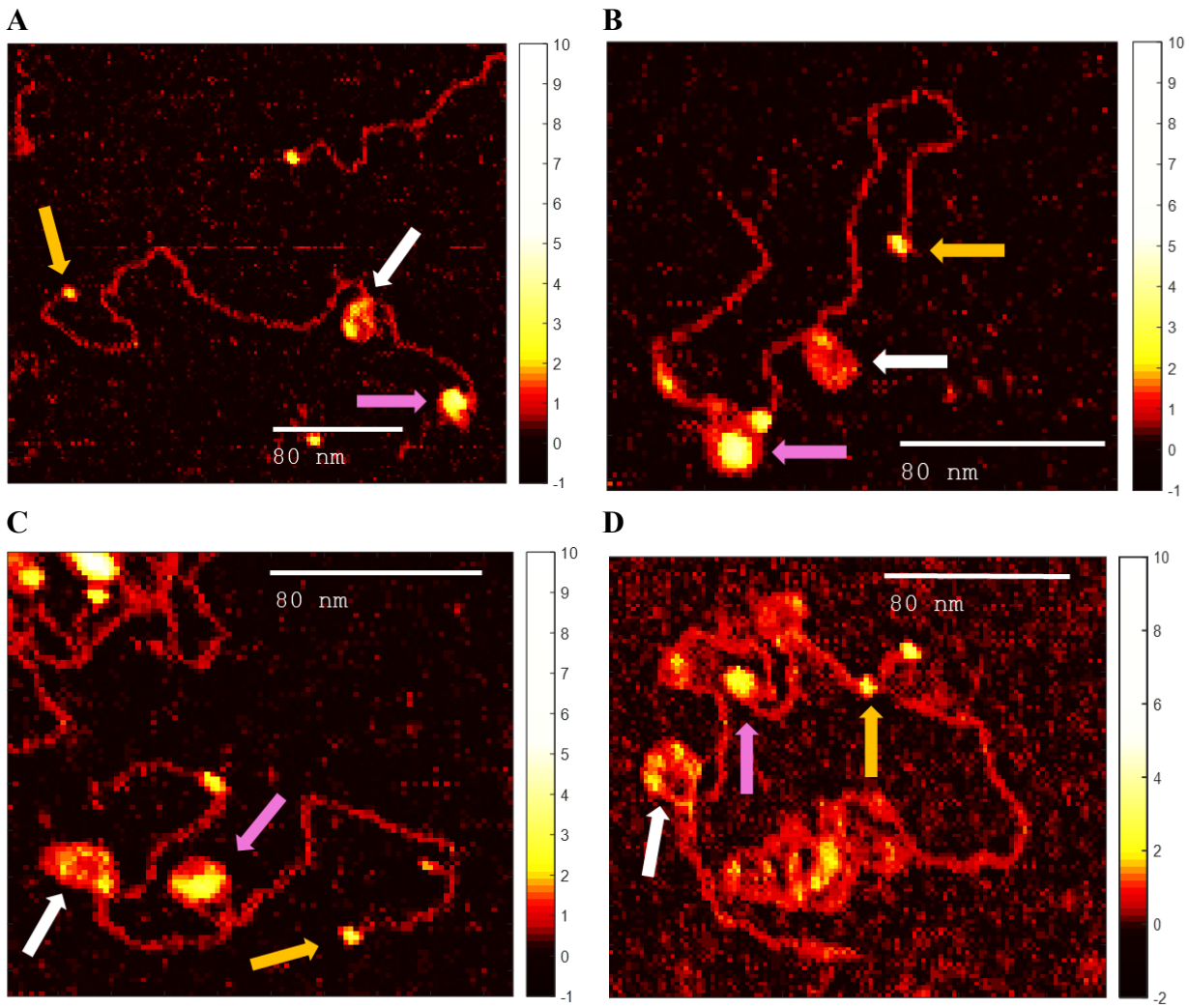


Figure 20. AFM Images of RNAP Progression with 186 CI. (A) RNAP near initiation site before hitting the 186 CI obstacle. (B) RNAP further along DNA before it encounters the 186 CI obstacle. (C) RNAP past the 186 CI on its way towards the terminator. (D) Two DNA molecules where the left molecule shows RNAP past 186 CI and the right one shows aggregation of a protein-DNA complex.

Progression of RNAP along DNA in the presence of 186 CI was analyzed. It was determined 186 CI is not an effective transcriptional roadblock for an elongating RNAP, as evidenced by the lack of decreased cumulative probability distribution along DNA by RNAP (Fig. 21).

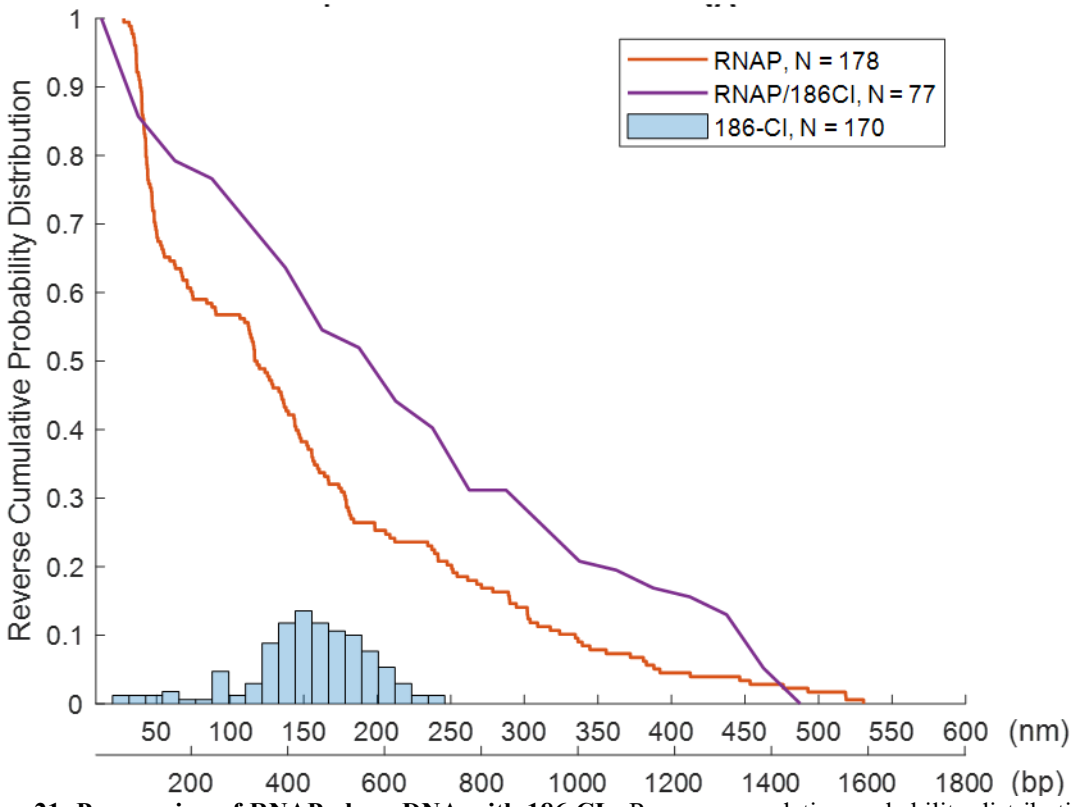


Figure 21. Progression of RNAP along DNA with 186 CI. Reverse cumulative probability distribution graph showing the probability of RNAP progression past a point on DNA. Orange line represents same progression as Fig. 17 and blue histograms are not on the same data scale and represent same data from Fig. 15. Purple line denotes the progression of RNAP on DNA molecules that showed 186 CI. Purple data was provided courtesy of Yue Lu.

Furthermore, a shift in the diameters of 186 CI particles was found when measured in transcriptional conditions (Fig. 22B). Overall, the 186 CI diameters were found to be smaller in experiments with active transcriptions, yet binding position along DNA stayed on the higher affinity pR site (fig. 22A), rather than the lower affinity, regulatory FR site.

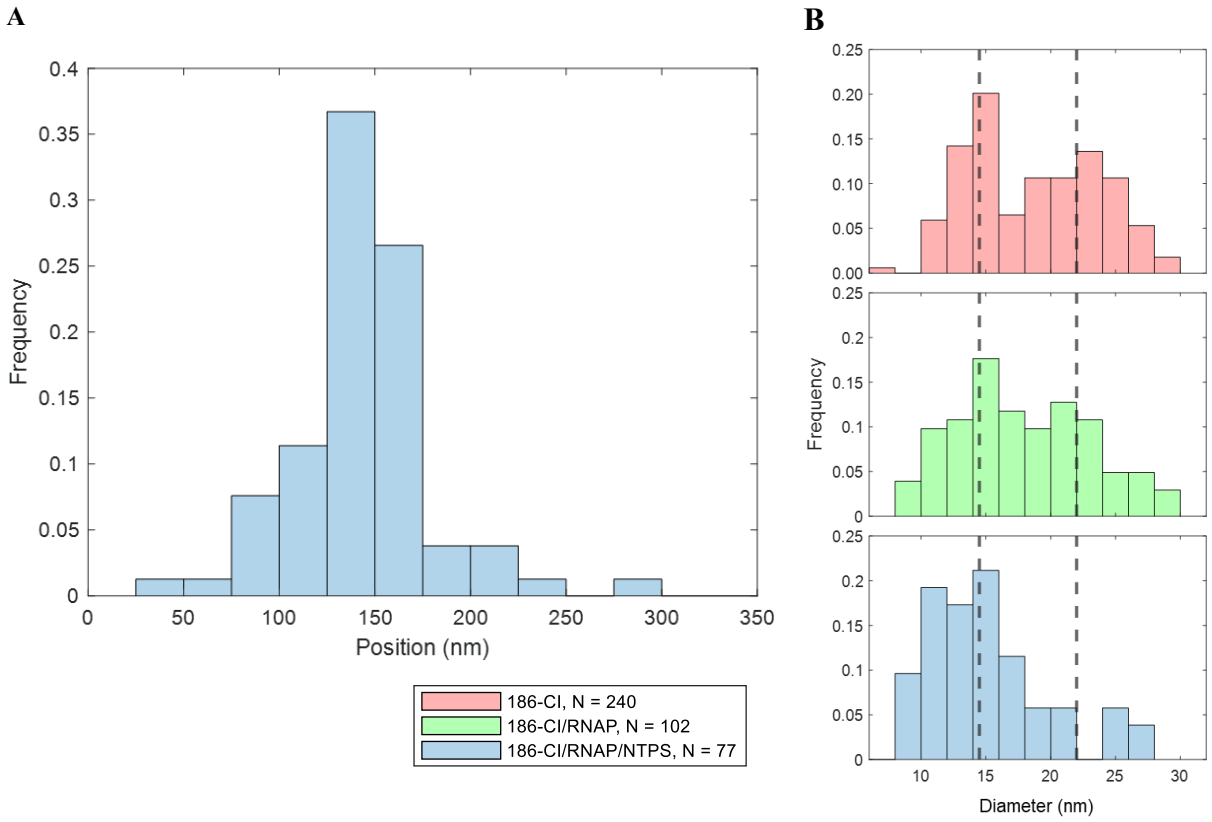


Figure 22. Active Transcription Alters 186 CI Profile. Sample conditions denoted in legend which corresponds to all graphs. (A) Probability histogram of 186 CI position on DNA in the presence of transcription, where majority of 186 CI bind to pR site. (B) Probability histogram of the diameter of 186 CI in different conditions show a difference in the 186 CI profile where dashed lines represent peaks of a fitted bimodal gaussian distribution. Data for this figure provided courtesy of Yue Lu.

VII. Discussion

A. Transcriptional Control through Chromatin Remodeling

Results indicate 186 CI does not act as an effective roadblock to RNAP elongation. Although we observed this in an artificial model system, the results may apply also to the wild-type situation in which 186 CI operates to repress pR and autoregulate its own expression. Indeed, 186 CI regulates transcription by changing the accessibility of promoter sites to RNAP. The lysogenic pL promoter is more accessible than the pR promoter since the former is a weaker binding site for the 186 CI protein. The lytic pR promoter is repressed because it is also a strong binding site for 186 CI. Thus, a thermodynamic equilibrium between DNA fully and partially wrapped around 186 CI regulates accessibility of RNAP to pL and, consequently, the amount of 186 CI produced as it is the promoter for expression of its own regulatory protein. Our evidence suggests once RNAP manages to bind pL, it can transcribe through the pR site even if the site is bound by 186 CI, since the repressor protein does not seem to be a significant roadblock. This mechanism is advantageous in the biological context because it allows for maintenance of lysogeny, which in turn allows replication of the bacteriophage genome along with the intertwined host genome and viral transmission to the bacterium progeny. Thus, DNA wrapping around a transcriptional factor as a mode of transcriptional regulation is an efficient mechanism to allow multifunctionality, such as repression of pR along with controlled expression of pL in the 186 bacteriophage with the 186 CI repressor protein, or expression of euchromatin along with compaction of the genome in eukaryotes with the histone octamer.

B. RNAP Breaks the Wheel

Previous studies have shown proteins can represent an effective roadblock to transcriptional elongation (Vassylyev 2009; King, Sen, and Weisberg 2003; Vörös, et al. 2017). Furthermore, to be propagated down progeny, 186 CI must allow DNA polymerase through during replication of the bacterium genome. Given these two facts, along with our experimental conclusion, there must be a mitigating factor which allows polymerases through 186 CI obstacles.

RNAP has been shown to not only break off histone dimers from nucleosomes complexes, but to translocate the complex down the DNA strand (Bintu et al. 2011; Kulaeva, Gaykalova, and Studitsky 2007). In a proposed mechanism (Fig. 23), as the transcription bubble collides with the nucleosome complex the disk-shaped protein complex shifts through its binding site conformations as forces push it off. As RNAP makes it through the obstacle, the disk-shaped protein complex "rotates" on the DNA and ends up being translocated behind the polymerase.

While this mechanism has yet to be fully confirmed, the loss of a histone dimer from the octamer nucleosome complex, leaving behind a hexamer-DNA complex, is substantiated by considerable experimental evidence. As noted earlier, the images attained from transcriptional experiments in the presence of 186 CI were of particular low quality compared to previous imaging due to protein aggregation and showed evidence of protein fragmentation. However,

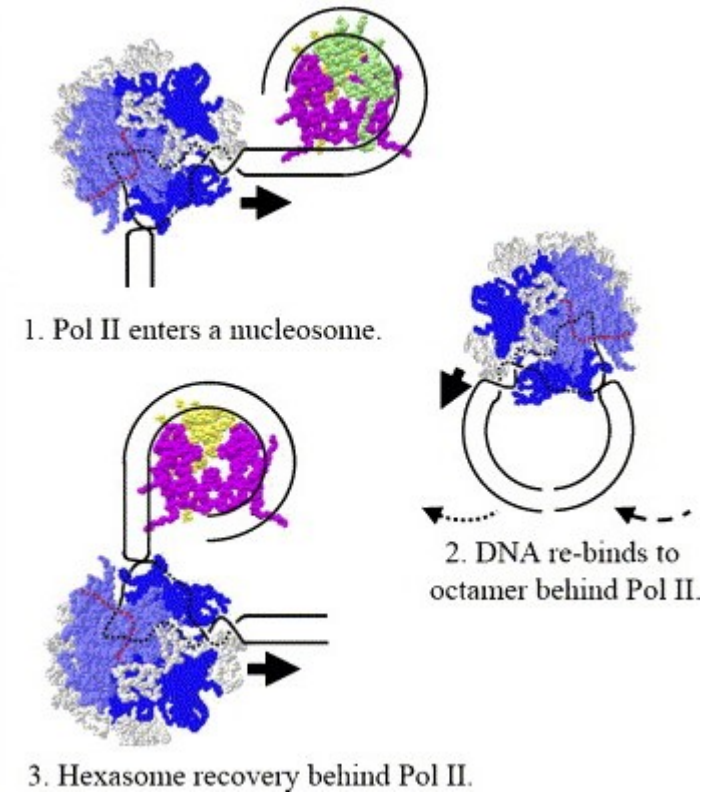


Figure 23. Proposed Mechanism of Wheel Translocation. Blue Protein is a polymerase approaching the nucleosome complex shown in pink. As figure illustrates, a possible mechanism for RNAP to bypass a histone is for histone to shift through its binding sites and bind behind the polymerase. Source: Adapted with permission from Kulaeva, Gaykalova, and Studitsky (2007).

it seems reasonable something similar to what is hypothesized for nucleosome translocation occurs in the case of elongation through 186 CI where the 186 CI disk may partly dissociate and lose some of the dimers which compose it. In our data, we not only showed most 186 CI, in the presence of transcription, lies at the pR site (Fig. 22A), but that the diameter of 186 CI in the presence of transcription is lowered (Fig. 22B). We therefore suggest 186 CI is being both translocated behind the RNAP during transcription and that the 186 CI wheel loses a dimer or more during this translocation process. This process would lead to an increase in the number of 186 dimers in solution which would decrease the quality of AFM images. Disaggregation and loss of units could possibly be a characteristic feature of proteins that form disk-like complexes when approached by

molecular motors and DNA helicases. In the case of transcription, the torque, from positive supercoiling generated by RNAP, on 186 CI in the downstream DNA may be the driving force for the remodeling of potential protein complex roadblocks, such as the 186 CI heptamer.

VIII. Conclusion

Using Atomic Force Microscopy and computational analysis, this study was able to gain insight into the mechanism at the basis of the 186 CI epigenetic switch and, more generally, of the roadblock activity of disk-shaped protein complexes. The results suggest the bacteriophage 186 CI repressor protein does not act as an effective roadblock to RNAP during transcriptional elongation. In addition, our results suggest the 186 CI heptameric wheel loses a dimer or more as RNAP elongates through during transcription. Further study and data analysis is needed to make quantitative estimates of the number of dimers effectively lost as a consequence of the encounter with an elongating RNAP and to test the hypothesis that 186 dimers may be translocated behind RNAP as suggested for histone octamers.

IX. References

Barry, Phillip J., and Ronald N. Goldman. 1988. "A Recursive Evaluation Algorithm for a Class of Catmull-Rom Splines." Paper presented at the Proceedings of the 15th annual conference on Computer graphics and interactive techniques. https://doi.org/10.1145/54852.378511.

Binnig, G., C. F. Quate, and Ch Gerber. 1986. "Atomic Force Microscope." *Physical Review Letters* 56, no. 9 (03/03/): 930-933. https://dx.doi.org/10.1103/PhysRevLett.56.930.

Bintu, Lacramioara, Marta Kopaczynska, Courtney Hodges, Lucyna Lubkowska, Mikhail Kashlev, and Carlos Bustamante. 2011. "The Elongation Rate of Rna Polymerase Determines the Fate of Transcribed Nucleosomes." *Nature Structural & Molecular Biology* 18, no. 12 (2011/12/01): 1394-1399. https://dx.doi.org/10.1038/nsmb.2164.

Bruker_Corporation. 2011. "Nanoscope Software 8.15 User Guide."

Engel, Andreas, and Daniel J. Müller. 2000. "Observing Single Biomolecules at Work with the Atomic Force Microscope." *Nature Structural Biology* 7, no. 9 (2000/09/01): 715-718. https://dx.doi.org/10.1038/78929.

Hansma, Helen G., Irene Revenko, Kerry Kim, and Daniel E. Laney. 1996. "Atomic Force Microscopy of Long and Short Double-Stranded, Single-Stranded and Triple-Stranded Nucleic Acids." *Nucleic Acids Research* 24, no. 4: 713-720. Accessed 3/28/2021. https://dx.doi.org/10.1093/nar/24.4.713.

Hendrickson, Cristin. 2018. "Transcriptional Elongation Roadblocking by the Lambda Bacteriophage Ci Protein." Accessed 2019-11-21 13:52:06 -0500. Emory Emory Theses and Dissertations.

Jalili, Nader, and Karthik Laxminarayana. 2004. "A Review of Atomic Force Microscopy Imaging Systems: Application to Molecular Metrology and Biological Sciences." *Mechatronics* 14, no. 8 (2004/10/01/): 907-945. https://dx.doi.org/https://doi.org/10.1016/j.mechatronics.2004.04.005.

King, Rodney A., Ranjan Sen, and Robert A. Weisberg. 2003. "Using a Lac Repressor Roadblock to Analyze the E. Coli Transcription Elongation Complex." In *Methods in Enzymology*, vol 371, 207-218: Academic Press.

Kulaeva, Olga I., Daria A. Gaykalova, and Vasily M. Studitsky. 2007. "Transcription through Chromatin by Rna Polymerase Ii: Histone Displacement and Exchange." *Mutation research* 618, no. 1-2: 116-129. https://dx.doi.org/10.1016/j.mrfmmm.2006.05.040.

Mullis, Kary B., and Fred A. Faloona. 1987. "[21] Specific Synthesis of DNA in Vitro Via a Polymerase-Catalyzed Chain Reaction." In *Methods in Enzymology*, vol 155, 335-350: Academic Press.

Shearwin, Keith E, Ian B Dodd, and J Barry Egan. 2002. "The Helix-Turn-Helix Motif of the Coliphage 186 Immunity Repressor Binds to Two Distinct Recognition Sequences." *Journal of Biological Chemistry* 277, no. 5: 3186-3194.

Vassylyev, Dmitry G. 2009. "Elongation by Rna Polymerase: A Race through Roadblocks." *Current Opinion in Structural Biology* 19, no. 6 (2009/12/01/): 691-700. https://dx.doi.org/https://doi.org/10.1016/j.sbi.2009.10.004.

Vörös, Zsuzsanna, Yan Yan, Daniel T Kovari, Laura Finzi, and David Dunlap. 2017. "Proteins Mediating DNA Loops Effectively Block Transcription." *Protein Science* 26, no. 7: 1427-1438. https://dx.doi.org/10.1002/pro.3156.

Wang, Haowei. 2011. *A Single Molecule Study of Two Bacteriophage Epigenetic Switches*: Emory Theses and Dissertations. https://etd.library.emory.edu/concern/etds/707958495?locale=en.

Wang, Haowei, Ian B. Dodd, David D. Dunlap, Keith E. Shearwin, and Laura Finzi. 2013. "Single Molecule Analysis of DNA Wrapping and Looping by a Circular 14mer Wheel of the Bacteriophage 186 Ci Repressor." *Nucleic Acids Research* 41, no. 11: 5746-5756. Accessed 10/27/2020. https://dx.doi.org/10.1093/nar/gkt298.

Weber, P. C., D. H. Ohlendorf, J. J. Wendoloski, and F. R. Salemme. 1989. "Structural Origins of High-Affinity Biotin Binding to Streptavidin." *Science* 243, no. 4887 (Jan 6): 85-8. https://dx.doi.org/10.1126/science.2911722.

Zurla, C., C. Manzo, D. Dunlap, D. E. Lewis, S. Adhya, and L. Finzi. 2009. "Direct Demonstration and Quantification of Long-Range DNA Looping by the Lambda Bacteriophage Repressor." *Nucleic Acids Res* 37, no. 9 (May): 2789-95. https://dx.doi.org/10.1093/nar/gkp134.

X. Appendix

A. Recombinant Plasmid Sequence

```
   1 tatcacagtt aaattgctaa cgcagtcagg caccgtgtat gaaatctaac aatgcgctca
  61 tcgtcatcct cggcaccgtc accctggatg ctgtaggcat aggcttggtt atgccggtac
 121 tgccgggcct cttgcgggat atcgtccatt ccgacagcat cgccagtcac tatggcgtgc
 181 tgctagcgct atatgcgttg atgcaatttc tatgcgcacc cgttctcgga gcactgtccg
 241 accgctttgg ccgccgccca gtcctgctcg cttcgctact tggagccact atcgactacg
 301 cgatcatggc gaccacaccc gtcctgtgga tcctctacgc cggacgcatc gtggccggca
 361 tcaccggcgc cacaggtgcg gttgctggcg cctatatcgc cgacatcacc gatggggaag
 421 atcgggctcg ccacttcggg ctcatgagcg cttgtttcgg cgtgggtatg gtggcaggcc
 481 ccgtggccgg gggactgttg ggcgccatct ccttgcatgc accattcctt gcggcggcgg
 541 tgctcaacgg cctcaaccta ctactgggct gcttcctaat gcaggagtcg cataagggag
 601 agcgtcgacc gatgcccttg agagccttca acccagtcag ctccttccgg tgggcgcggg
 661 gcatgactat cgtcgccgca cttatgactg tcttctttat catgcaactc gtaggacagg
 721 tgccggcagc gctctgggtc attttcggcg aggaccgctt tcgctggagc gcgacgatga
 781 tcggcctgtc gcttgcggta ttcggaatct tgcacgccct cgctcaagcc ttcgtcactg
 841 gtcccgccac caaacgtttc ggcgagaagc aggccattat cgccggcatg gcggccgacg
 901 cgctgggcta cgtcttgctg gcgttcgcga cgcgaggctg gatggccttc cccattatga
 961 ttcttctcgc ttccggcggc atcgggatgc ccgcgttgca ggccatgctg tccaggcagg
1021 tagatgacga ccatcaggga cagcttcaag gatcgctcgc ggctcttacc agcctaactt
1081 cgatcattgg accgctgatc gtcacggcga tttatgccgc ctcggcgagc acatggaacg
1141 ggttggcatg gattgtaggc gccgccctat accttgtctg cctccccgcg ttgcgtcgcg
1201 gtgcatggag ccgggccacc tcgacctgaa tggaagccgg cggcacctcg ctaacggatt
1261 caccactcca agaattggag ccaatcaatt cttgcggaga actgtgaatg cgcaaaccaa
1321 cccttggcag aacatatcca tcgcgtccgc catctccagc agccgcacgc ggcgcatctc
1381 gggcagcgtt gggtcctggc cacgggtgcg catgatcgtg ctcctgtcgt tgaggacccg
1441 gctaggctgg cggggttgcc ttactggtta gcagaatgaa tcaccgatac gcgagcgaac
1501 gtgaagcgac tgctgctgca aaacgtctgc gacctgagca acaacatgaa tggtcttcgg
1561 tttccgtgtt tcgtaaagtc tggaaacgcg gaagtcagcg ccctgcacca ttatgttccg
1621 gatctgcatc gcaggatgct gctggctacc ctgtggaaca cctacatctg tattaacgaa
1681 gcgctggcat tgaccctgag tgatttttct ctggtcccgc cgcatccata ccgccagttg
1741 tttaccctca caacgttcca gtaaccgggc atgttcatca tcagtaaccc gtatcgtgag
1801 catcctctct cgtttcatcg gtatcattac ccccatgaac agaaatcccc cttacacgga
1861 ggcatcagtg accaaacagg aaaaaaccgc ccttaacatg gcccgcttta tcagaagcca
1921 gacattaacg cttctggaga aactcaacga gctggacgcg gatgaacagg cagacatctg
1981 tgaatcgctt cacgaccacg ctgatgagct ttaccgcagc tgcctcgcgc gtttcggtga
2041 tgacggtgaa aacctctgac acatgcagct cccggagacg gtcacagctt gtctgtaagc
2101 ggatgccggg agcagacaag cccgtcaggg cgcgtcagcg ggtgttggca ggtgtcgggg
```

2161 cgcagccatg acccagtcac cccatggtgc agtatgaagg cggcggagcc gacaccacgg

2221 ccaccgatat tatttgcccg atgtacgcgc gcgtggatga acaccagccc ttcccggctt

2281 tatcaaaaag agtattgact taaagtctaa cctataggat acttacagcg atggagaggt

2341 gtagtggtaa ccagaagata agatggcttt cgctacctgg agagacgcgc ccgctgatcc

2401 tttgcgaata cgcccacgcg atgggtaaca gtcttggcgg tttcgctaaa tactggcagg

2461 cgtttcgtca gtatccccgt ttacagggcg gcttcgtctg ggactgggtg gatcagtcgc

2521 tgattaaata tgatgaaaac ggcaacccgt ggtacctcaa gtacatccac gttgctccat

2581 cctaaagaat ctattctcat ttcgataaaa cctatttact atctctcaat tgggagatat

2641 attttggcta aacccacgca attgatggca agtgttggca aacagagtca aatcaattgc

2701 aaactttggc taatagggaa tcatgcaata tggcttctga aatcgcaatc atcaaagtgc

2761 ctgcacctat cgttactctg caacaattcg cagagcttga gggtgtttct gaacgcaccg

2821 cctagcgctg gacaaccggc gacaaccctt gtgtaccaat cgaaccccgc acaatccgta

2881 aaggctgcaa gaaagcaggt ggcccgattc gcatttatta cgcacgctgg aaagaagagc

2941 agttgcgtaa ggcgttggga cattcccgtt ttcaactcgt catcggtgct taattcactt

3001 tatgtgaatt gtaaggatgc aacatgctcg agctgggtaa taagcgttgg caatttaacc

3061 gccagtcagg ctttctttca cagatgtgga ttggcgataa aaaacaactg ctgacgccgc

3121 tgcgcgatca gttcacccgt gcaccgctgg ataacgacat tggcgtaagt gaagcgaccc

3181 gcattgaccc taacgcctgg gtcgaacact ggaaggcggc gggccattac caggccgaag

3241 cagcgttgtt gcagtgcacg gcagatacac ttgctgatgc ggtgctgatt acgaccgctc

3301 acgcgtggca gcatcagggg aaaaccttat ttatcagccg gaaaacctac cggattgatg

3361 gtagtggtca aatggcgatt accgttgatg ttgaagtggc gagcgataca ccgcatccgg

3421 cgcggattgg cctgaactgc cagctggcgc aggtagcaga gcgggtaaac tggctcggat

3481 tagcggccgc aagaaaacta tcccgaccgc cttactgccg cctgttttga ccgctgggat

3541 ctgctgtaac agagcattag cgcaaggtga tttttgtctt cttgcgctaa ttttttccat

3601 tgtctagagt agcgatagcg gagtgtatac tggcttaact atgcggcatc agagcagatt

3661 gtactgagag tgcaccatat gcggtgtgaa ataccgcaca gatgcgtaag gagaaaatac

3721 cgcatcaggc gctcttccgc ttcctcgctc actgactcgc tgcgctcggt cgttcggctg

3781 cggcgagcgg tatcagctca ctcaaaggcg gtaatacggt tatccacaga atcaggggat

3841 aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc

3901 gcgttgctgg cgtttttcca taggctccgc ccccctgacg agcatcacaa aaatcgacgc

3961 tcaagtcaga ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga

4021 agctccctcg tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt

4081 ctcccttcgg gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcggtg

4141 taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc

4201 gccttatccg gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg

4261 gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc

4321 ttgaagtggt ggcctaacta cggctacact agaaggacag tatttggtat ctgcgctctg

4381 ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc

4441 gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct

4501 caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt

```
4561 taagggattt tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa
4621 aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa
4681 tgcttaatca gtgaggcacc tatctcagcg atctgtctat ttcgttcatc catagttgcc
4741 tgactccccg tcgtgtagat aactacgata cgggagggct taccatctgg ccccagtgct
4801 gcaatgatac cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca
4861 gccggaaggg ccgagcgcag aagtggtcct gcaactttat ccgcctccat ccagtctatt
4921 aattgttgcc gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttgtt
4981 gccattgctg caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc
5041 ggttcccaac gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcggttagc
5101 tccttcggtc ctccgatcgt tgtcagaagt aagttggccg cagtgttatc actcatggtt
5161 atggcagcac tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact
5221 ggtgagtact caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc
5281 ccggcgtcaa cacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt
5341 ggaaaacgtt cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg
5401 atgtaaccca ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct
5461 gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa
5521 tgttgaatac tcatactctt cctttttcaa tattattgaa gcatttatca gggttattgt
5581 ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc
5641 acatttcccc gaaaagtgcc acctgacgtc taagaaacca ttattatcat gacattaacc
5701 tataaaaata ggcgtatcac gaggcccttt cgtcttcaag aattctcatg tttgacagct
5761 tatcatcgat aagctttaat gcggtagtt
```

B. Primer Sequences

Primer	*Sequence*	*T_m*
S/pBR322/2211	agccatgacccagtcac	56°C
A/PBR322/3728	gtagcgatagcggagtgtat	55°C

C. MATLAB® Code

Code for Particle Analysis and Data Linker can be found in the GitHub repository of Gustavo Borjas: https://github.com/borjasgj/AFM-Analysis

Dr. Dan Kovari's code for the *DNATracer* program, containing Dr. Wang's algorithm, can be found in his GitHub repository: https://github.com/dkovari/DNAtracer

Dr. Haowei Wang's original code, and description of it, can also be found in his Emory dissertation Appendices: https://etd.library.emory.edu/concern/etds/707958495?locale=en