SINGLE MOLECULE FLUORESCENCE STUDIES OF BIOMOLECULAR INTERACTIONS

BY

REZA VAFABAKHSH

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Nigel Goldenfeld, Chair
Professor Taekjip Ha, Director of Research
Professor John D. Stack
Assistant Professor Yann R. Chemla

Abstract

Single molecule fluorescent techniques have become standard approaches to study protein-DNA interactions. However, these techniques have largely been confined by limitations in assays to studying the interaction between simple DNA substrates and a single protein. During my PhD I developed several novel assays to study a long-standing controversial biophysics question (flexibility of short dsDNA), genome packaging in viruses (Influenza and T4) and dynamics of challenging protein complexes (membrane proteins).

The classical view of DNA posits that DNA must be stiff below the persistence length (<150 base pair) but recent studies addressing this have yielded contradictory results. We developed a fluorescence-based, protein-free, assay for studying the cyclization of single DNA molecules in real time. The looping rate for short DNA molecules has remarkably weak length dependence between 67 and 106 bps, deviating significantly from the worm-like chain model. We propose that many biologically significant protein-DNA interactions that involve looping and bending of DNA below 100 bp likely use this intrinsic bendability of DNA.

One of the critical aspects of a virus life cycle is packaging of the viral genome. Different viruses have devised intelligent mechanisms to perform this task. I studied packaging mechanism in Bacteriophage T4 and Influenza. Influenza A virus possesses a segmented genome of eight, single-stranded RNAs. However, the exact copy number of each viral RNA segment per individual virus particles has been controversial for the past 50 years. To address this question we combined single molecule TIRF microscopy and multi-color fluorescent *in situ* hybridization (FISH) to study the composition of viral RNAs at single-virus particle resolution. Our results showed that a high percentage of virus particles package a single copy of each segment of viral RNAs. Our findings support a model that the packaging of influenza genome is a selective and robust process.

Finally we developed a single molecule fluorescence assay to study initiation and re-initiation of dsDNA packaging in the T4 bacteriophage. Using this assay we quantified the details of T4 “packasome” assembly. Also, we showed that the T4 packaging machine can package multiple DNA into the same head in burst-like fashion.

To my father and mother

Acknowledgements

I would like to thank my advisor, Taekjip Ha for his guidance and support throughout my graduate studies. His scientific insight, critical guidance, endless enthusiasm and vision for research were inspirational.

I would also like to thank Professors John Stack, Nigel Goldenfeld, and Yann Chemla for serving on my committee, and offering support and constructive comments.

I am grateful from Michelle Nahas, Chirlmin Joo, Rahul Roy and Hamza Balci who generously helped me with experiments and provided assistance when I joined the lab and started to work on my first project.

I am appreciative of the past and current Ha lab members. Specifically I give my heartfelt thanks to Ruobo Zhou, Kaushik Ragunathan, Sultan Doganay, Kyung suk Lee, Ankur Jain, Ibrahim Cisse, Gwangrog Lee, Jeehae Park, and Sinan Arslan who were wonderful labmates and constant sources of technical assistance and valuable discussions. Working with these people definitely improved my experience in Urbana and I will miss them!

During my graduate studies in the Ha lab I had the opportunity to collaborate with several great scientists from whom I learned a lot. I heartily thank Rob Phillips, Hernan Garcia, John Widom, Colin Nichols, Shizhen Wang, Katherine Henzler-Wildman, Supratik Dutta, Venigalla Rao, Li (Joyce) Dai, Vishal Kottadiel, Peter Palese and Yi-Ying Chou for their patience to work with me and generous assistance.

I will always be grateful to my M.Sc. thesis advisor Ramin Golestanian for introducing me to the experimental biophysics.

I would like to thank all of my friends in Urbana-Champaign with whom I have shared so many excellent times over the past six years. In particular, I would like to thank Colleen. Akbar, Soheil, Lyuda, Reza, Mohammad, Esi, Maro, Renee, Mustafa and Ehsan for all the good memories I will take with me.

I thank my father who encouraged me to study physics in college and was my first teacher. Finally, I am forever grateful to my parents for all the sacrifices they have made for me and supporting me to achieve my dreams.

Table of Contents

Chapter 1 Introduction 1
1.1 Fluorescence 1
1.2 Förster (Fluorescence) Resonance Energy Transfer (FRET) 2
1.3 Single Molecule Fluorescence 3
1.3.1 Total Internal Reflection Microscopy 3
1.3.2 Single Molecule Fluorescent Dyes 4
1.3.3 Imaging Buffer 5
1.4 Figures 7

Chapter 2 Mechanical Properties of DNA on Short Length Scales 11
2.1 Introduction 11
2.1.1 The Worm Like Chain Model (WLC) for DNA 12
2.1.2 Ensemble Methods for Studying DNA Mechanical Properties 13
2.1.3 Single Molecule Measurements of DNA Mechanical Properties 14
2.1.4 *j* factor 15
2.1.5 Previous Experiments to Quantify Short DNA Looping 17
2.2 Experimental Results 20
2.2.1 Single Molecule Cyclization Assay 20
2.2.2 Measuring *j* factor 23
2.2.3 Applications of the Single Molecule Cyclization Assay 24
2.3 Discussion 25
2.4 Experimental Procedures 26
2.5 Figures 34

Chapter 3 Studying the Genome Composition of Influenza Virus at the Single Virus Particle Level 45
3.1 Introduction 45
3.2 Experimental Results 47
3.2.1 Assay Design and Validation 47
3.2.2 Quantifying the Co-packaging of vRNAs 48
3.2.3 Quantifying the Copy Number of Each Packaged vRNA 49
3.3 Discussion 52
3.4 Conclusion 53
3.5 Experimental Procedures 53
3.6 Figures and Tables 57

Chapter 4 Analyzing DNA Packaging Initiation of Bacteriophage T4 by a Real-time Single Molecule Fluorescence Assay 67
4.1 Introduction 67
4.2 Experimental Results 70
4.2.1 Single Molecule Fluorescence Assay to Study DNA Packaging 70
4.2.2 Bacteriophage T4 Packaging Machine is Highly Promiscuous 71
4.2.3 Assembly of Bacteriophage T Packaging Machinery 72

4.2.4 DNA Can Directly Bind to the Viral Capsid Portal....73
4.2.5 Quantifying the DNA Packaging Initiation....75
4.2.6 Initiation of DNA Packaging Happens in Bursts....75
4.2.7 Quantifying the DNA Packaging Initiation Efficiency....76
4.2.8 Heat and Cold-Sensitive gp17 Mutants are Defective in Initiation....77
4.3 Discussion....78
4.4 Experimental Procedures....80
4.5 Figures....83

Appendix A PEGylation Procedure....97

References....100

Chapter 1
Introduction

1.1 Flourescence

The phenomenon of fluorescence occurs when a specimen absorbs a specific wavelength of light and consequently emits light as it return to its ground state from the excited state. British scientist Sir George G. Stokes first quantified this phenomena and coined the term “fluorescence”. Stokes wrote “It was certainly a curious sight to see the tube instantaneously light up when plunged into the invisible rays: it was literally darkness visible. Altogether the phenomenon had something of an unearthly appearance.” (*1*). Since then, fluorescence techniques have become an integral part of the modern biomedical research and it is impossible to imagine current developments without the use of fluorescent techniques.

The basic principle of fluorescence can be depicted in a Jablonski diagram (Figure 1.1A) (*2*). A molecule at room temperature primarily resides in its lowest energy state which is the singlet ground state (S_0). The molecule then can absorb a photon with certain wavelength and transit to the singlet excited state (S_1). Both the ground state and the excited state have multiple vibrational levels and depending on the excitations energy, a transition occurs between a vibrational level in S_0 and S_1, within 10^{-15} seconds. After excitation, the molecule is quickly relaxed to the lowest vibrational level of the excited electronic state. This process is called internal conversion (or vibrational relaxation) and occurs on the time scale of 10^{-15} -10^{-12} seconds. The molecule remains in the lowest vibrational level of the excited electronic state for a period on the order of 10^{-9} seconds, so called the fluorescence lifetime. Fluorescence emission occurs as the fluorophore decay from the singlet electronic excited states to an allowable vibrational level in the electronic ground state. The energy lost during the internal conversion results in the emitted photon has a lower energy (longer wavelength) than the excitation photon. This phenomenon is called Stokes shift. Alternatively, the excited molecule in S_1 can relax to the lowest excited triplet state (T_1) through a process called intersystem crossing. This transition is rare and takes 10^{-8}-10^{-5} seconds. Once in T_1 the molecule may relax to the

ground state via a slow radiative process (phosphorescence). Alternatively, the molecules in the triplet state can relax to the ground state by nonradiative processes where excitation energy is not converted into photons but is dissipated by thermal processes such as vibrational relaxation and collisional quenching.

1.2 Förster (Fluorescence) Resonance Energy Transfer (FRET)

Energy transfer is a non-radiative process through which a molecule in the excited state transfers energy without emission (Figure 1.1B). Förster (or Fluorescence) Resonance Energy Transfer (FRET) refers to the non-radiative transfer of an electronic excitation from a donor molecule (*D*) to an acceptor molecule (*A*) via dipole-dipole interaction (equation 1.1), first described by Förster in 1946 (*3*).

$$\begin{aligned} D + h\upsilon &\rightarrow D^* \\ D^* + A &\rightarrow D + A^* \\ A^* &\rightarrow A + h\upsilon' \end{aligned} \qquad (1.1)$$

The dipole-dipole interaction energy is of the form $V_{\text{int}} = A\frac{\kappa}{R^3}$ where A is a constant, R is the distance between the donor and acceptor dipoles and κ is a constant that depends on donor and acceptor dipole orientations. According to the Fermi's golden rule, the transition rate between the initial and final states is proportional to $\left|\langle f|V_{\text{int}}|i\rangle\right|^2$ where $|i\rangle = |D^*A\rangle$ and $|f\rangle = |DA^*\rangle$ are the initial and final states. Therefore the energy transfer rate would be $k_{ET} \propto \frac{1}{R^6}$. Exact calculation of the transfer rate yields:

$$k_{ET} = \frac{1}{\tau_D}\left(\frac{R_0}{r}\right)^6 \qquad (1.2)$$

where τ_D is the fluorescence lifetime of the donor in the absence of acceptor. R_0 which is the Förster distance the can be written as:

$$R_0{}^6 = (8.785\times10^{17})\,\phi_D\,\kappa^2 n^{-4} J(\upsilon)\ (\text{nm}^6) \qquad (1.3)$$

where ϕ_D is the donor quantum yield in the absence of the acceptor, n is the refractive index of the medium, κ^2 is the dipole orientation factor (ranging from 0 to 4 and it is 2/3 for a pair of freely rotating dipoles) (Figure 1.2A) and is calculated by (*4*):

$$\kappa^2 = (\cos\gamma - 3\cos\alpha\cos\beta)^2 \qquad (1.4)$$

J(ν) is the integral of the spectral overlap between the donor emission and the acceptor absorption (Figure 1.2B).

FRET efficiency, *E,* is a measure of what fraction of donor excitation is transferred to the acceptor and is given by $E = k_{ET}/(k_{ET} + k_r + k_{nr})$ where k_r and k_{nr} are radiative and non-radiative decay rates of the donor. Using these definitions, and eq. (1.2), the FRET efficiency can be written as:

$$E(R) = \frac{1}{1+\left(\frac{k_r + k_{nr}}{k_{ET}}\right)} = \frac{1}{1+\left(\frac{R}{R_0}\right)^6} \qquad (1.5)$$

The power 6 dependence of FRET efficiency on the distance between the donor and acceptor makes FRET a powerful tool to probe and measure the relative distance between donor and acceptor with very high sensitivity (Figure 1.3A). However it is important to remember the sensitivity of this method depends on the distance between donor and acceptor, with the highest sensitivity around R=R_0 (Figure 1.3B).

1.3 Single Molecule Fluorescence

Fluorescence-based single molecule techniques have become a staple approach in biomedical research in the past 20 years by providing previously unobtainable data on fundamental biological processes (*5*). Techniques based on single molecule FRET (*6*), fluorescence quenching (*7*), single molecule localization and tracking (*8*) and polarization spectroscopy (*9*) are being extensively used to study protein-nucleic interactions as well as conformational dynamics of proteins complexes and nucleic acids structures. In this section I will discuss practical aspects of single molecule fluorescence which I used in the following chapters.

1.3.1 Total Internal Reflection Microscopy

One of the major hurdles in detecting single molecules is the high background from diffusing fluorescent molecules in epi-fluorescence microscopy. A key advancement was achieved by using total internal reflection (TIR) microscopes and

probing surface tethered molecules (*10, 11*). In TIR microscopy an exponentioally decaying evanescent field is created on the surface (*12*). The "penetration depth" of the excitation field is 100 nm – 200 nm which greatly reduces the background fluorescence. Since TIR is a wide-field microscopy technique, it allows for probing hundreds of molecules at the same time which greatly enhances the yield of an experiment.

Figure 1.4A shows the schematic of the prism-type TIR. In brief, the excitation laser is guided towards the microscope sample holder using appropriate mirrors. The laser intensity is attenuated using the combination of a half-waveplate and a polarizing beam-splitting cube or with neutral density filters. In our setup, we used a solid-state laser at 532 nm and a HeNe laser at 633 nm. The two lasers were combined through the beam-splitting cube. Finally the beam is focused onto the objective with a TIR lens. Imaging is done through a 60X water immersion objective (UPLAPO60XW, 60x, 1.2NA, Olympus) (Figure 1.4B). In the detection path of the setup, a 550 nm long-pass filter and a 635 nm notch filter remove the scattered excitation light from the 532 nm and 633 nm lasers. The fluorescence emission from the excited fluorophores are then split using a 650 nm dichroic filter and imaged with an electron-multiplying charge-coupled device (EM-CCD) camera (*13*). An area of (75 µm × 75 µm) can be imaged onto the 8.2 mm × 8.2 mm CCD chip at one time.

Fluorescence signal is recorded in real time using home-written Visual C++ software either in 8 bits or 16 bits. The movies are analyzed using either a custom-written IDL or a C++ program. In brief, an average image is created by averaging few frames from the movie (5 to 12 frames). This averaging removes the noise and increases the detection accuracy. Then the fluorescent molecules that exhibit a Gaussian intensity profile are selected in the averaged image. These molecules are tracked over time and their fluorescent intensity is recorded to build the intensity time trace for each molecule. These intensity time traces were used for further analysis.

1.3.2 Single Molecule Fluorescent Dyes

The most popular dyes for single molecule studies are organic fluorophores such as Cyanine, Alexa and Atto fluorophores. These fluorophores are very bright (have high extinction coefficient and high quantum yield), water soluble, small in size and

photostable without large intensity fluctuations. In this thesis, Cy3 and Cy5 fluorophores were used in all the experiments presented (Figure 1.5). Also Cy3 and Cy5 are the most popular donor and acceptor pair for smFRET (Cy3 as the donor and Cy5 as the acceptor) because 1) their spectral separation is large (~100 nm) but they have considerable spectral overlap; 2) they are both photostable in oxygen-free environment and 3) their quantum yields (~0.25) are comparable. We used the 532 nm laser for Cy3 excitation and the 633 nm laser for Cy5 direct excitation. In the FRET experiment the FRET efficiency was calculated by:

$$E = \frac{I_A}{I_A + I_D} \qquad (1.4)$$

where I_A and I_D are the emission intensities of the acceptor and the donor respectively.

1.3.3 Imaging Buffer

One of the major limitations in performing single molecule fluorescence experiments is the photobleaching phenomena. Photobleaching occurs when a fluorophore permanently loses the ability to fluoresce due to photon-induced chemical damage and covalent modification. Another unfavorable photophysics related to fluorophores is “blinking” of fluorophores. Blinking happens when the fluorophore intensity repeatedly drops to zero and then comes back to normal (*14*). Molecular oxygen is primarily responsible for photobleaching, either through direct interaction with a fluorophore in the excited state or indirectly by producing free radicals in solution. Therefore, it is essential to remove molecular oxygen for achieving long observation times. However, O_2 is an efficient triplet state quencher and removing oxygen to prolong the photobleaching time of the fluorophore would increase the dark state lifetime to milliseconds (*15*). Therefore, to observe fluorophores for extended periods of time, oxygen must be removed from the solution but an alternative triplet-state quencher should be used.

In all the experiments in this thesis 3-4 mM Trolox was used as the triplet state quencher. Also, an oxygen scavenging system consisting of glucose oxidase (165 U/ml), catalase (2,170 U/ml) and β-D glucose (0.4% wt/wt) was used (*16*). Trolox is prepared by dissolving 8 – 10 mg Trolox powder in distilled water. Trolox is a weak acid and it does

not dissolve in water efficiently. We found that adding NaOH to the final concentration of 5 mM would completely dissolve the Trolox. After adding water and NaOH to trolox, the tube is covered with aluminum foil to protect from light and tumbled at room temperature for 2 hours. It is then kept in the fridge and could be used for up to a month. This imaging buffer allows us to probe single molecules for several minutes and collect more than a million photons before photobleaching.

It is worth mentioning that gluconic acid is an unfavorable side product of the aforementioned oxygen scavenging system. As a result, pH of the sample would drop after applying the imaging buffer (*17*). The rate of pH drop depends on the buffer strength of the imaging buffer however, it is important to supply fresh imaging buffer if the experiment is longer than 30 minutes.

1.4 Figures

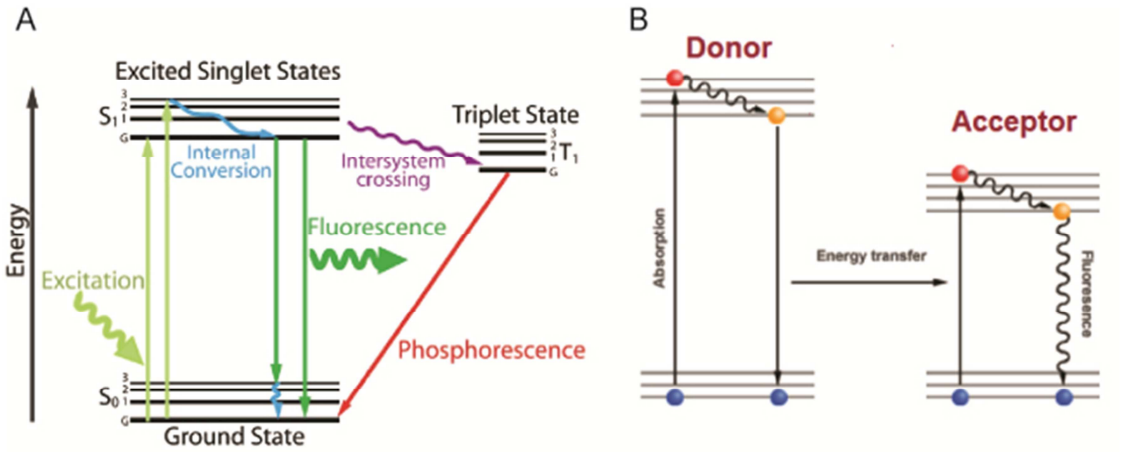

Figure 1.1 (A) Jablonski diagram of a typical fluorescent molecule. The ground state S_0, the singlet states S_1 and the triplet state T_1 are represented by thick black horizontal lines and the thinner horizontal black lines depict the vibrational states associated with each electronic state. Photophysical processes occur due to transitions between different states are represented by colored arrows. (B) The energy diagram for fluorescence resonance energy transfer (FRET) between donor and acceptor molecules.

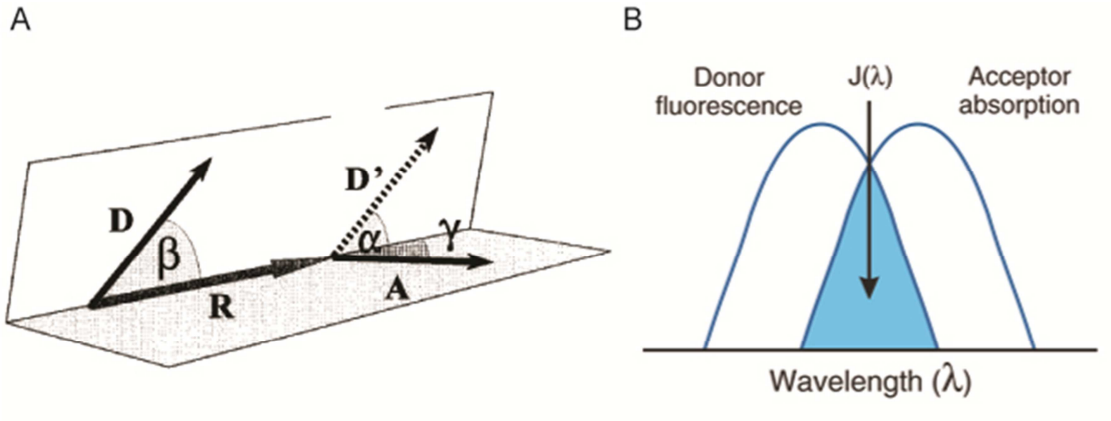

Figure 1.2 (A) The orientation parameter κ^2. The transition dipole vectors for the coupled donor and acceptor fluorophores are indicated by the arrows, labeled D and A. Vector D′ is generated by the in-plane translation of vector D to share its origin with vector A. The definition of κ^2 is given by Eq. (1.5) and is based on the angles shown. (B) Spectral overlap between the donor emission and the acceptor absorption used to calculate the Förster distance.

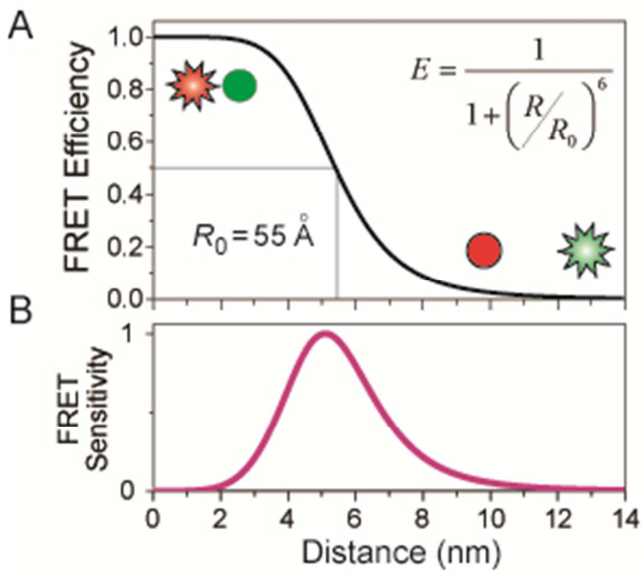

Figure 1.3 FRET Efficiency as a function of the distance between donor and acceptor. The Förster distance (R_0) for Cy3-Cy5 pair is about 55 Å. FRET efficiency at $R = R_0$ is 0.5. FRET sensitivity (defined as dE/dR) as a function of distance. As it is seen, the highest sensitivity is around $R = R_0$.

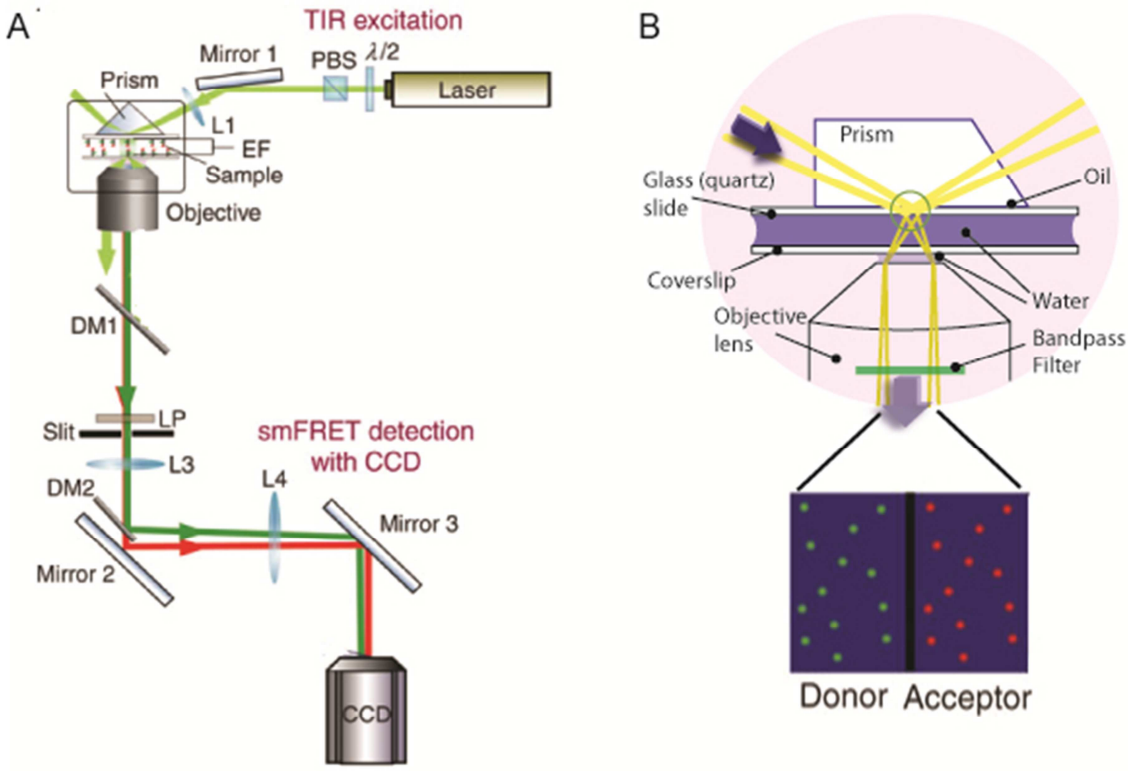

Figure 1.4 (A) Optical layout for prism-type total internal reflection (TIR) microscopy with typical dual-channel EM-CCD detection. (B) Prism is coupled to the slide with an immersion oil film. Excitation laser enters the sample through prism and excite the molecules that are immobilized on the slide surface. Fluorescent signal is then collected through a water immersion objective and imaged onto the EM-CCD.

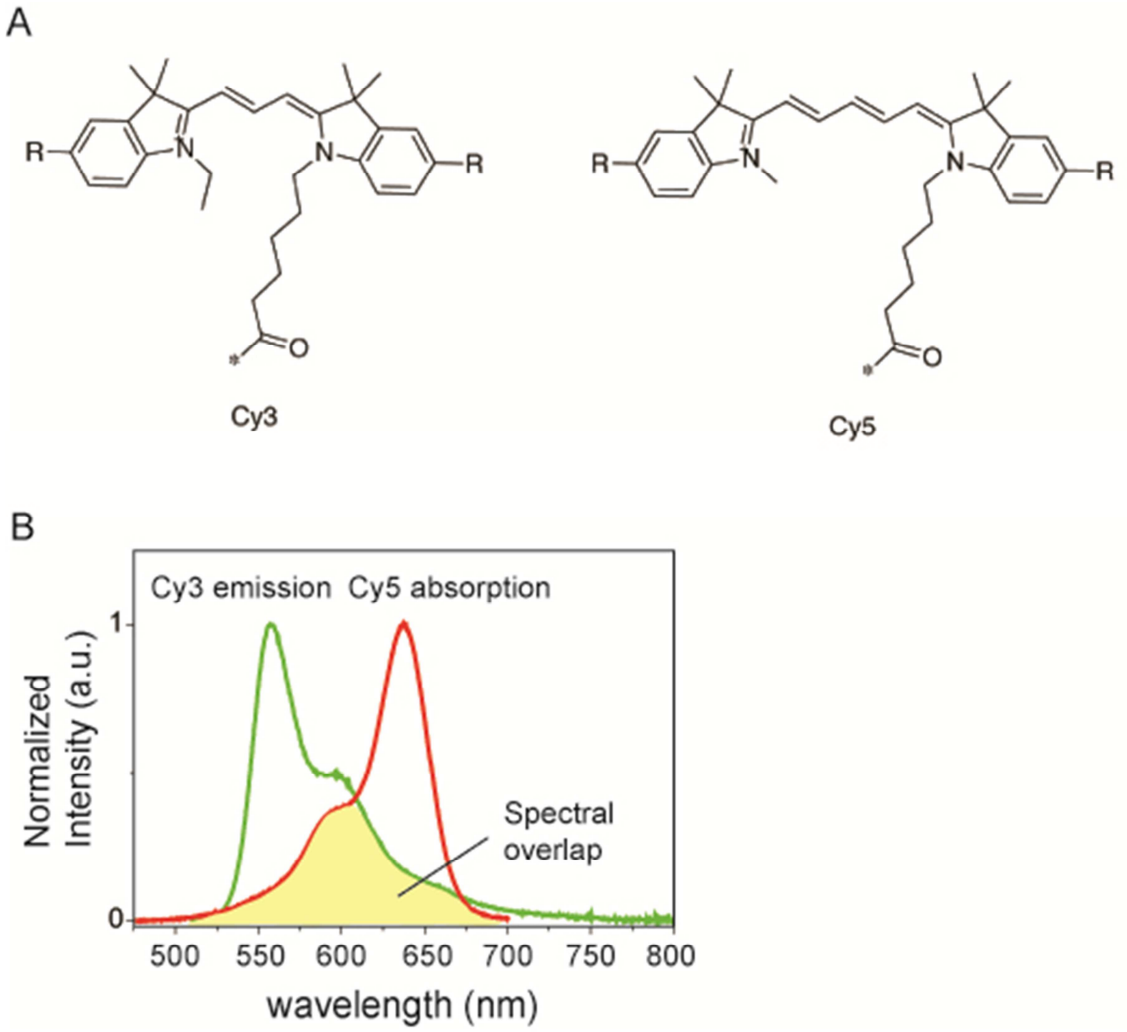

Figure 1.5 (A) Structure of Cy3 and Cy5 dyes. These dyes were used for all the experiments in this thesis as a FRET pair or individually. (B) Cy3 emission and Cy5 absorption have a considerable overlap which makes them an ideal FRET pair.

Chapter 2

Mechanical properties of DNA on Short Length Scales [*]

2.1 Introduction

DNA in the cell undergoes conformational and mechanical alteration over a wide range of length scales and time scales. Processes such as cell division, DNA duplication, and DNA packaging involve DNA management on the genome-wide scale whereas interactions such as epigenetic modifications of DNA, mismatch repair, and regulation of gene expression affects DNA locally. Because these interactions involve physical manipulation of DNA, the mechanical properties of DNA play a crucial role in regulating and managing these interactions. Therefore, to be able to understand and control such processes we need to have a clear picture of the DNA physical properties on the relevant length scales and timescales.

Bending and looping of segments of DNA below 100 base pair (bp) is ubiquitous in cellular processes such as regulated gene expression in bacteria and eukaryotes (*18, 19*), packaging of DNA in viral capsids and DNA storage complexes in eukaryotes (*20*). According to a widely used approximation, DNA duplex is modeled as an elastic rod and its mechanical properties are described by the worm-like chain (WLC) model. Persistence length (l_p) is a measure of the bending rigidity of DNA and for a DNA molecule that is several kilobasepair (kbp) or longer, l_p can be readily measured using bulk or single molecule manipulation tools and is about 50 nm or 150 basepair (*21*). In this framework, formation of DNA loops or sharp bends over length scales shorter than l_p incurs a large energetic cost which makes the probability of their spontaneous formation vanishingly

[*] This work in Chapter 2 has been published as the following paper and book chapter:

- Vafabakhsh, R. & Ha, T. "Extreme Bendability of DNA Less than 100 Base Pairs Long Revealed by Single-Molecule Cyclization", *Science* **337**, 1097-1101 (2012).
- R. Vafabakhsh, K. S. Lee, T. Ha "Recent Advances in studying mechanical properties of DNA", *Advances in Chemical Physics, Volume 150* (John Wiley & Sons, 2012).

small. Therefore, quantifying the intrinsic bendability of DNA at biologically important length scales is essential for understanding DNA-protein interactions.

In this chapter first I briefly review the theoretical and experimental approaches to study and model mechanical properties of double stranded DNA (dsDNA). Then I will introduce the single molecule cyclization assay and its application to examine theoretical models of DNA at biologically relevant length scales.

2.1.1 The Worm-Like Chain Model (WLC) for DNA

The simplest mechanical model for the DNA is the freely joined chain (FJC) model. This model treats a polymer as a chain of independent and uncorrelated segments with a fixed length (Kuhn length). A one dimensional freely joined chain polymer with N segments, each of length *b*, is mathematically equivalent to a one dimensional random walk consist of N steps each of length *b* (Figure 2.1A). The end-to-end distance probability distribution for such a polymer is (*22*):

$$P_N(R) = \frac{1}{\sqrt{2\pi N b^2}} e^{-R^2/2Nb^2} \qquad (2.1)$$

For this polymer, the variance of dsDNA length fluctuations (σ^2_{total}) depends on the nature of the correlation between individual basepairs' fluctuation (σ^2_{bp}). If fluctuations of single basepairs are completely uncorrelated then $\sigma^2_{total} = N\ \sigma^2_{bp}$, where *N* is the number of basepairs. On the other hand, if the thermal fluctuations of all basepairs are fully correlated, $\sigma^2_{total} = N^2\ \sigma^2_{bp}$ (*23*).

To model the bending of dsDNA for a chain we can assign a bending energy proportional to the cosine of the angle between tangent vectors of adjacent segments in the form of: $E = -g\left(\hat{t}_i \cdot \hat{t}_{i+1}\right)$ where g is the bending modulus. For such a chain, the thermal correlation between two segments *i* and *j* would be (*24*):

$$\left\langle \hat{t}_i \cdot \hat{t}_j \right\rangle = e^{|j-i|\mathrm{Ln}\left(\coth\left(\frac{g}{k_BT}\right)-\frac{k_BT}{g}\right)} \qquad (2.2)$$

For Eq. (2.2) the correlation of tangent vectors falls off exponentially with the characteristic length of $\frac{-1}{\mathrm{Ln}\left(\coth\left(\frac{g}{k_BT}\right)-\frac{k_BT}{g}\right)}$.

Accordingly, in the continuum limit, the bending energy of the polymer can be written as

$$E = \frac{B}{2}\int_0^L ds \left|\frac{d\hat{t}}{ds}\right|^2 \qquad (2.3)$$

where B is the bending rigidity of the polymer (*25*). By extending the result for a discreet polymer to the continuous polymer, the tangent vector correlation can be written as:

$$\left\langle \hat{t}(s_1)\cdot\hat{t}(s_1)\right\rangle = e^{-(s_1-s_2)/P} \qquad (2.4)$$

where P is the persistence length of the polymer, related to its bending rigidity $P = B/k_B T$ (*26*). This is the basic worm like chain model (WLC) and has been the standard platform for modeling the mechanical properties of DNA.

2.1.2 Ensemble Methods for Studying DNA Mechanical Properties

Quantifying the mechanical properties of DNA have been an active subject of research since the discovery of DNA helical structure. Early studies used electron microscopy (*27*) and hydrodynamic and optical methods (*28*) to measure the persistence length of dsDNA. Hydrodynamic methods are based on measuring intrinsic sedimentation and viscosity coefficients of DNA molecules in solution as a function of molecular weight (*29*). By fitting the results of such experiments to a relevant theoretical model for a WLC polymer, a value of 600 ± 100 Å was estimated for the persistence length of DNA (*29, 30*).

Optical methods are based on light scattering. Light scattering measurements provide information about molecular weight and size of DNA coils which can be used to deduce the persistence length of DNA (*31*). Alternatively, angular dependence of light scattering from a polymer solution can be used to estimate the persistence length of DNA (*32*). Transient Electric Birefringence (TBE) has also been used to quantify DNA elasticity (*33*). In this method a short electric pulse (several microseconds) introduces optical anisotropy in the sample and gives rise to birefringence. After the field is switched off, the characteristic relaxation time of induced birefringence is measured. By fitting to an appropriate model, this relaxation time can be used to measure the rotational diffusion coefficient and persistence length of DNA (*34*). Using this approach a

persistence length of approximately 500 Å was determined for the double stranded DNA. Also it was shown that the persistence length of DNA has a weak dependence on monovalent ion concentrations above 1 mM concentration of ions.

The ensemble assays for measuring persistence length of DNA as a gauge of its elasticity suffer from fundamental limitations. In bulk assays the average behavior of many molecules is being probed and intermediate states or rare events cannot be resolved. In addition, the heterogeneity of molecules limit the accuracy and resolution of measurements (*35*). Moreover, all these bulk assays are indirect and based on inferring the persistence length from measuring other physical properties of DNA in solution. Therefore, although a great deal of mechanical information about DNA molecules was discovered using these bulk assays, these limitations hinder the applicability of bulk assays to probe structural and mechanical details of nucleic acids.

2.1.3 Single Molecule Measurements of DNA Mechanical Properties

Single molecule techniques such as optical tweezers have enabled us to probe the mechanical response of single DNA molecules under external force. To infer elastic properties of DNA from these experiments, one needs a proper model to fit to the experimental data and get those parameters. The average extension of a FJC polymer with the Kuhn length b and under an external force F can be written as:

$$\langle x \rangle = L\left(\coth\left(\frac{Fb}{k_{\mathrm{B}}T} \right) - \frac{k_{\mathrm{B}}T}{Fb} \right) \qquad (2.5)$$

This relationship has shown to describe the behavior of duplex DNA in the limit of small forces with a good accuracy (F < 0.08 pN) (*36*). In this regime DNA acts like a spring with an effective spring constant of $\frac{3k_{\mathrm{B}}T}{bL}$. Unfortunately, at the intermediate and large forces this simple model fails. However, WLC model provides a very good description of DNA in this force regime (F < 10 pN) (*37*) (Figure 2.1B). In this range, an approximate relationship which only accounts for the entropic energy contributions provide an excellent fit to the data (*38*):

$$F = \left(\frac{k_{\mathrm{B}}T}{P} \right)\left(\frac{1}{4(1-x/L_o)^2} + \frac{x}{L_o} - \frac{1}{4} \right) \qquad (2.6)$$

where P and L_o are persistence length and contour length of DNA. Since this model is purely entropic and describes an inextensible polymer with a fixed contour length, a deviation from this theory is observed at forces above 10 pN. This can be corrected by introducing a finite stretching modulus in the model which accounts for the enthalpic stretching and can fit the stretching data at forces up to 50 pN (*39*):

$$F = \left(\frac{k_B T}{P}\right)\left(\frac{1}{4(1-x/L_o+F/K_o)^2}+\frac{x}{L_o}-\frac{F}{K_o}-\frac{1}{4}\right) \qquad (2.7)$$

here K_o is the dsDNA stretching modulus (The force required to double the length of DNA) and it is estimated to be about 1000 pN (*39*). Figure 2.1B shows the force-extension data for a λ-DNA (contour length=16.3 μm) in 50 mM Na^+ along with fit to FJC, approximate Marko-Siggia (Equation 2.6) and modified Marko-Siggia (Equation 2.7). From such single molecule experiments the persistence length of DNA has been estimated to be around 45 nm to 50 nm depending on the buffer conditions (*21*, *39*).

2.1.4 *j* factor

For a simple WLC chain with the persistence length of P, the elastic energy cost to form a circular loop of length L based on Equation 2.3 is:

$$\frac{E_{circle}}{k_B T} = 2\pi^2 \frac{P}{L} \approx 3000 \frac{1}{N_{bp}} \qquad (2.8)$$

where N_{bp} is the number of basepairs in the duplex. For a 100 bp dsDNA this energy is about 30 $k_B T$ which is a large energy in the cellular level (1 ATP hydrolysis produces about 20 $k_B T$ energy). It is worth mentioning that the lowest energy of a dsDNA loop is not a circular shape but a tear-drop-shaped loop (*40*). In this case the elastic energy is $\frac{E_{teardrop}}{k_B T} = 14.055\frac{P}{L}$ which is about 71% energy of a circular loop.

The bending energy by itself is not enough to predict the probability of dsDNA looping; one must also sum over the bending fluctuations, thermally excited changes in shape. Cyclization experiments are often reported using a quantity known as the *j*-factor, which has the units of concentration. *j* factor can be interpreted as the effective concentration of one end of the dsDNA molecule in the vicinity of the other end. Exact derivation of *j* factor involves sophisticated analytic and numerical calculations however

I will outline a few approximate approaches to calculate this quantity.

The original Shimada-Yamakawa calculation was based on the saddle-point approximation of the Green's function (*41*). In brief, they discretized the continuous chain into n-segments, parameterized using Euler's angels, and calculated the minimum potential energy of such a chain. Then they expanded the energy around this minimum in terms of fluctuations in the chain configuration (i.e. twist and writhe). Finally they estimated the integral over all configurations and arrived at (Figure 2.2A):

$$j = \frac{1.66 \times 112.04}{P^3}\left(\frac{P}{L}\right)^5 e^{-E/k_B T} e^{0.246 L/P} \qquad (2.9)$$

This relationship is valid for L< 4*P* since for longer chains the entropic contributions would be considerable. In fact, for a chain with L >> *P*, we reach the Gaussian chain limit. In this limit, we may estimate the probability of finding the two ends of a long DNA molecule in the vicinity of each other, using the average end-to-end distance for such a chain which is $\sqrt{2PL}$. In this case the two ends of the chain are in a volume of $\approx (PL)^{3/2}$. Therefore the probability of finding the two ends together for a long chain (L >> *P*) scales as $j \approx \frac{1}{(PL)^{3/2}}$.

Alternatively, the numerical calculation for *j* factor can be done using the transfer matrix method (*42*). In this method, first the end-to-end vector distribution for a WLC polymer is represented through

$$\rho(\vec{R}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{R}} \frac{\int d^2\hat{t}_1 \cdots d^2\hat{t}_N \xi \exp\left[-\beta E + ib\vec{k}\cdot\sum_{j=1}^{N-1}\hat{t}_j\right]}{\int d^2\hat{t}_1 \cdots d^2\hat{t}_N e^{-\beta E}} \qquad (2.10)$$

where $\beta E = \sum_{i=1}^{N-1} \frac{a}{2}(\hat{t}_{i=1} - \hat{t}_i)^2$ is the total energy and *a* and *b* are the elastic constant and segment unit length of the polymer. ξ is a function that imposes a specific boundary conditionfor loop formation. Using the transfer matrix approach, Equation 2.10 can be written as $\rho(\vec{R}) = \frac{Z(\vec{R})}{Z}$ where $Z(\vec{R}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{R}} Z_k$ and

$Z_k = \sum_{l_1,m_1}\sum_{l_N,m_N}\int d^2\hat{t}_1 d^2\hat{t}_N \xi\langle\hat{t}_1|l_1,m_1\rangle\langle l_1,m_1|T_k{}^{N-1}|l_N,m_N\rangle\langle l_N,m_N|\hat{t}_N\rangle$. T_k is the k-dependent transfer matrix and its elements can be computed through

$$\langle l,m|T_k|l',m'\rangle = 4\pi(-1)^m\delta_{m,m'}\sum_j i^j(2j+1)\sqrt{(2l+1)(2l'+1)}\begin{pmatrix} j & l & l' \\ 0 & 0 & 0\end{pmatrix} \times\begin{pmatrix} j & l & l' \\ 0 & m & -m\end{pmatrix}e^{-a}i_{i'}(a)J_j(bk)$$

As it mentioned before, the boundary condition is handled by the function ξ. Figure 2.2B shows two important boundary conditions namely parallel and free boundary conditions. The parallel boundary condition requires the two ends of polymer come in contact parallel to each other for the cyclization to happen. In this case, $\xi = \delta^2(\hat{t}_1,\hat{t}_N)$. For the case of free boundary condition, the two ends can anneal with any angel they come into contact with each other. Having the $\rho(R)$ at hand, j could be easily calculated as $j = \frac{4\pi}{N_A}\rho(0)$ since j factor is essentially the end-to-end vector distribution at R=0. It is important to mention that the torsional stiffness of dsDNA is not considered here. Including this factor adds a modulation with the period of 10 bp to the j factor curve (Figure 2.2A). This is simply due to the fact that the two ends of DNA can anneal more readily when they are in-phase.

Calculating j factor for the condition that the two ends can anneal when they are within a distance r of each other (capture radius of r) is similar. In this case j can be calculate as: $j(r) = \frac{1}{\frac{4}{3}\pi r^3}\int_0^r dr'4\pi r'^2\rho(r')$.

2.1.5 Previous Experiments to Quantify Short DNA Looping

The most well-known experimental approach to quantify dsDNA cyclization is called the ligase assay (*43*). In this approach, dsDNA with sticky overhangs (usually 4 nucleotides) are incubated with the ligase protein. After certain incubation time, the ligation reaction is quenched. The ligation reaction yields different products including

linear monomers, circular monomers, linear dimers, circular dimers and so on (Figure 2.3A). Since the gel motility of these products is different, they can be separated with high accuracy on a PAGE gel (Figure 2.3B). As shown by Jacobson and Stockmayer the ring closure probability is $j = K_c / K_a$ where K_c is the cyclization constant and K_a is the biomolecular equilibrium constant for joining two molecules (*44*). Taylor and Hagerman showed that in the framework of the ligase assay, *j* can be measured in a single ligation reaction through:

$$j = 2M_0 \ \ Lim \ \left[C(t)/D(t)\right] \ t \to 0 \qquad (2.11)$$

where M_0 is the molar concentration of linear monomer at the start of the reaction and *C*(t) and *D*(t) are the time-dependent concentrations of monomer circles and linear dimers (*45*).

Cloutier and Widom were the first to apply the ligase assay to quantify cyclization of short DNA molecules (89 - 116 bps) (*46, 47*). Their results suggested that *j* factor for DNA molecules in this range, is up to 5 orders of magnitude higher than the WLC prediction (Figure 2.3C). In order to explain these unexpected findings, theoreticians proposed that the observed higher flexibility is due to thermally-induced formation of bubbles or kinks in the DNA (*48, 49*). Kinking of the DNA duplex would allow unusually large local bend angles which enhances DNA cyclization. In a similar way, bubbles would also create single stranded DNA (ssDNA) regions which are much more flexible than dsDNA and would increase the cyclization rate. All-atom molecular dynamics (MD) simulations were also applied to determine if kinking could be regarded as an explanation for the Cloutier and Widom result. Lankas and co-workers observed formation of such kinks in DNA circles (*50*). Local basepairing disruption was also observed in other simulations (*51, 52*). These studies suggested that formation of dsDNA circles smaller than 85 bp in length may require alteration of the WLC model to include sharp bending or kinking of DNA. Although the result of experiments using ligase assay and the theoretical and computer simulation work that ensued provided a solid evidence for deviation of dsDNA from WLC model at short length scales, follow up experiments challenged these results. Vologodskii and co-workers revisited cyclization experiments of

dsDNA fragments 105−130 bp in length and demonstrated that the WLC model remains adequate for predicting the cyclization probabilities of these fragments when assumptions of the assay are thoroughly met (*53*) (Figure 2.3D).

Since then other experimental approaches have been used to quantify dsDNA looping at length scales below 100 bp. AFM imaging of dsDNA molecules absorbed on a surface and analyzing their contours showed that the percentage of large bends in the DNA was larger than what you would expect based on the WLC model (Figure 2.4A) (*54*). These results were later challenged and attributed to the specific geometry of the experimental setup and interaction of dsDNA with the charged surface (*55, 56*). More recently, small angle x-ray scattering (SAXS) experiments examining the equilibrium end-to-end distance distributions of short DNA fragments with gold nanocrystal labels again challenged the WLC model at short DNA lengths (Figure 2.4B) (*23*). These experiments suggested a cooperative stretching behavior over two helical turns of the DNA double helix. This interpretation suggested new flexibility regimes for DNA beyond WLC predictions. However, later it was suggested that the fluctuation of the linker connecting the gold nanoparticle to the DNA can amplify the real stretching fluctuations through a linker leverage effect (*57*). This claim was later confirmed in fresh SAXS experiments and computer simulations that considered the effect of attachment linker (*58*).

Contradictory results from these measurements call for a more direct approach to study flexibility of DNA on short length scales. In addition, all existing bulk approaches suffer from inherent limitations such as limited range of physical conditions and formation of by-products other than monomer DNA circles (Figure 2.3A), which limit their applicability to other systems. Moreover, real-time looping events cannot be detected using these techniques and some of these approaches require extensive analysis which could complicate the results and conclusions. On the other hand, because of geometrical and technical limitations, single-molecule DNA-stretching approaches cannot be used to study the mechanics of very short DNA molecules. In the single molecule experiment, even for a moderate length of DNA with several hundred bp, many corrections are required to account for the finite length of the chain and the boundary conditions (*59*). In addition, because of the relatively long persistence length of dsDNA,

even 100 femto-Newton of force makes configurations with sharp local bending inaccessible. Therefore, DNA responses measured by mechanical stretching would not include any contribution from such sharply bent conformations even if they existed in the relaxed DNA.

2.2 Experimental Results

2.2.1 Single-Molecule Cyclization Assay

We developed a cyclization assay, based on single molecule fluorescence resonance energy transfer (smFRET) (*11, 60*), for directly monitoring the cyclization of single DNA molecules (Figure 2.5A). For the long-term observation while avoiding dimer formation, DNA molecules are immobilized on a polymer-coated surface through a biotin linker attached to a base at an internal DNA location. To avoid potential contributions from inherent DNA curvature, we avoided motifs such as A-tracts which are known to induce considerable intrinsic curvature (*61*). The DNA probe is a duplex with single-stranded extensions on both 5′ ends. Each DNA molecule is labeled with Cy3 (donor) and Cy5 (acceptor) fluorophores at the 5′ end of the strands. Single-stranded overhangs are complementary so that hybridization will trap the DNA molecules in the looped state. In the unlooped state, the donor and acceptor are distant from each other and the molecules show zero FRET. Looping brings the dyes close to each other, and the DNA molecules exhibit a high FRET signal. Therefore, the looped state can be clearly distinguished from the unlooped state based on the FRET value and the relative intensities of donor and acceptor (Figure 2.5B).

The experiment starts in a buffer without added ions in order to strongly favor the unlooped state. Introducing a buffer containing high concentration of Na^+ or Mg^{2+} can stabilize the looped state. Depending on the lengths of the DNA duplex and the single stranded overhangs, different behaviors were observed during the probing time window, typically ranging from less than 1 minute to up to 4 hours (Figure 2.5B). DNA molecules formed stable loops, showed dynamics between looped and unlooped conformations, or exhibited no looping events. For example, for a 91 bp initial dsDNA with 10 nt overhangs, looping was nearly irreversible and the looped high FRET accumulated to saturation within about 20 min (Figure 2.6). In this case, the looping rate *R* could be

determined by fitting the time evolution of the looped population with a single exponential function (Figure 2.6B).

We measured the looping rate *R* for a series of DNA molecules with a 10 nucleotide overhang on each end and with the circular size ranging from 67 to 106 bps (the circular size is the circumference of the DNA circle formed after looping, and is therefore the sum of the initial dsDNA length and the overhang length). The measured looping time, 1/*R*, varied from less than 10 minutes to more than 200 minutes (Figure 2.7A, black squares). This 20 fold change in the looping times is surprisingly small, because we expected that DNA significantly shorter than the persistent length would take dramatically longer to form a loop. However, the result is qualitatively consistent with the observations that short DNA loops induced by the lac repressor and AraC protein can form efficiently in vivo and in vitro (*19, 62-64*), suggesting that the protein-induced looping may use the intrinsic flexibility of the DNA. Also, our result is consistent with a recent in vivo study which predicted that bending energy for short DNA loops would be independent of the loop length (*54*). In addition to the length dependence, a variation in looping rate depending on the angular phase between the two cohesive ends would be expected. Indeed, our data displays oscillation in *R* with a period of about one helical turn (see data points between 93 bp and 106 bp).

We performed a variety of controls to confirm that the surprisingly weak length dependence of the looping rates was not an artifact of our experimental scheme. First, to rule out possible contributions from surface tethering and internal biotin labeling, we repeated the experiments for DNA molecules without any internal modification and confined in 200 nm diameter phospholipid vesicles that are permeable to ions (*65*). Because infusion of 1 M NaCl ruptured the vesicles, we instead used 10 mM Mg^{2+} to stabilize the looped state. This is the same ionic condition used in the standard ligase assay. For five DNA constructs ranging from 94 to 105 bp, the looping times were similar between the vesicle encapsulated measurements and the surface measurements (Figure 2.7A, red squares). Because DNA bending and torsional rigidity are sequence dependent (*66*), it remained possible that the high flexibility we observed was due to an extraordinarily flexible sequence. Therefore, we measured the looping times for a series of DNA constructs with identical loop length and overhang sequence but different

internal base composition (Figure 2.7B). Our standard sequence (R73) is not an outlier in terms of the looping rate. The DNA derived from a nucleosome positioning sequence (TA) (*67*) showed much faster (35 fold) looping than our standard sequence R73 (*47*). We also examined the looping behavior of sequences with potentially curvature inducing A tracks, $(A)_n$, where n=0, 10, 17, 26, 38 embedded in otherwise randomly chosen sequence and found that the looping rate varied from more than 30 fold higher (n=0) to 2 fold lower (n=26) than that of R73. The two orders of magnitude difference in the looping time for these sequences despite the fact that the final 12 bp of these duplexes on both ends are similar, rules out duplex end opening as a possible mechanism for the rapid looping observed in our assay.

Changing the length of overhang in our DNA constructs allowed us to observe real time looping-unlooping dynamics (Figure 2.8A) and evaluate the rates directly as a functional of salt. While the unlooping rate did not change between 0.5 M to 2 M Na^+, the looping rate increased 3 fold in this range (Figure 2.8B). However, since we observed the same increase in the bimolecular annealing rate (Figure 2.8C) we can attribute the acceleration in looping at higher salt solely to the annealing enhancement. Therefore, monovalent ion concentrations above 0.5 M do not have a detectable effect on dsDNA flexibility (*21*).

When a DNA loop forms, internal elastic energy stored in the loop would provide a shear force that promotes unlooping. Shorter DNA molecules would incur a stronger internal force, accelerating unlooping. Previously, FRET in combination with magnetic tweezers provided an estimate on the internal elastic force, f_{int}, in short DNA loops (*68, 69*). Unlooping would be opposed by the basepairing of single strand overhangs. AFM-based investigation of DNA duplex shearing indicated that the critical shear force, f_c, needed to rupture a short duplex increases linearly with increasing duplex length (*70, 71*). It was showed that 7 basepairs are required to form a stable duplex and estimated f_c per bp of 3-4 pN per each additional bp beyond 7 bp. The balance between f_{int}, determined primarily by the duplex length, and f_c which is controlled by the overhang length will determine the fraction of annealed state, $P_{annealed}$, in equilibrium. For the constructs with 10 nt overhang, 10 basepairs in the looped state can withstand forces up to 12 pN before they are sheared (f_c ~ 12 pN) (*70*) whereas even for DNA circles as short as 70 bp, f_{int} is

below 10 pN (*68*) which is smaller than f_c . This is likely to be the reason why we rarely observed unlooping with 10 nt overhang ($P_{annealed}$ ~1). Indeed, if we decrease the overhang length from 10 nt to 9 nt or 8 nt while keeping the 91 bp initial duplex length, $P_{annealed}$ becomes progressively smaller, approaching 0, likely because f_c is now smaller than f_{int} (Figure 2.9A). Consistent with this data, we observed that the unlooping rate increased by about two orders of magnitude while looping rate did not change in this range (Figure 2.9B). For the 4 nt overhangs typically used in the ligase assays, $P_{annealed}$ would be even smaller. Therefore, the main effect of longer overhangs in our assay compared to the 4 nt overhangs typical in the ligase assay is to increase the lifetime of the looped state. When a DNA loop forms, internal elastic energy stored in the loop is expected to provide a shear force that promotes unlooping. Indeed we found that 8 bp of duplex melts 20 times faster in a DNA circle than in a DNA dimer, likely due to internal tension in a circle which is absent in a dimer (Figure 2.9C).

2.2.2 Measuring *j* Factor

The looping rate, *R*, can be calculated as the product of the bimolecular association rate k_{on} between 10 nt long complementary strands and the effective concentration of one end of DNA in the vicinity of the other end, which we call the apparent *j* factor, j_{app} (*43*) (Figure 2.10A). Using a similar surface-based assay but for intermolecular annealing (Figure 2.10B), k_{on} was measured to be $0.78 \pm 0.07 \times 10^6 M^{-1}s^{-1}$ in 1 M NaCl and $0.26 \pm 0.04 \times 10^6 M^{-1}s^{-1}$ in 10 mM Mg^{2+}, both consistent with an earlier estimate for short oligonucleotide annealing (*72*). The corresponding apparent *j*-factors, calculated using $j_{app} = R/k_{on}$ are shown in Figure 2.10C. Our calculated apparent *j*-factors along with the prediction of the WLC model (*73, 74*) are plotted in Figure 2.10D. The solid line and the dashed line are the *j*-factor for a semiflexible polymer with parallel and free boundary conditions and the dotted line is the *j*-factor for a polymer with free boundary condition and 5 nm capture radius (the two ends anneal when they are closer than 5 nm) (*42, 75*). The measured j_{app} values matched the theoretical prediction for 201 bp DNA but deviated from the theoretical values by orders of magnitude for the shortest lengths examined even under the most liberal boundary condition.

Our observation that the looping rate does not drop precipitously with decreasing DNA length is in stark contrast to the steep drop in the *j*-factor predicted by the WLC model. In many biologically relevant protein-DNA interactions such as in some genetic switches, the DNA bendability plays an important role in determining the state of the switch by controlling the concentration of one protein binding site in the vicinity of the other binding site (*18, 76*). Our assay samples such equilibrium in which two DNA ends are in close proximity, but not annealed (dashed box in Figure 2.11A). In contrast, the ligase assay samples the equilibrium of the annealed state (dashed box in Figure 2.11B) (*43*). The equilibrium looped population, which is the substrate for ligase protein, is very sensitive to the unlooping rate (Figure 2.9). The looping rate is biologically more relevant because it reports on how quickly two regions of DNA are brought into close proximity whereas the unlooping rate is additionally influenced by the melting rate of the short duplex formed. Our assay could independently measure the looping rate without being affected by the loop instability caused by internal tension in the short DNA circles. Many DNA binding proteins may have evolved to use the extraordinary flexibility of the DNA to capture and further stabilize transiently bent or looped DNA conformations.

2.2.3 Applications of Single-Molecule Cyclization Assay

So far the most common assay to quantify the effect of DNA modifications, DNA sequence or protein or ligand binding, on the DNA cyclization has been the ligase assay (*77-80*). However since this assay relies on the activity of ligase, the range of buffer conditions that can be used (ions, pH, temperature …) is very limited. Moreover, external factors like other proteins or ligands could interfere with the ligation reaction. On the other hand, our assay is protein-free and is practically general in terms of buffer conditions.

We first set to test how DNA defects would affect looping of DNA. We prepared a series of DNA constructs with single backbone nick, double nicks or a single bp mismatch in the middle. We observed that the doubly nicked DNA could cyclize an order of magnitude faster than the intact DNA (Figure 2.12A). Also single mismatch in the middle (C:C) increased the looping rate of DNA by a factor of 5 (Figure 2.12B).

The nucleoid-associated protein HU is one of the most abundant proteins in

Escherichia coli and has been suggested to play an important role in bacterial nucleoid organization and regulation (*81*). Previously, using the ligase assay, it was suggested that binding of HU to the DNA facilitates DNA looping and bending (*82*). However, using AFM imaging and magnetic tweezers, it was recently observed that at high concentrations, HU makes the DNA more stiff, supposedly by forming filaments around dsDNA (*83*). We tested the effect of HU in our cyclization assay. Figure 2.13A shows the fraction of DNA loops after 5 minutes in the presence and absence of 100 nM HU. As it is seen addition of HU dramatically enhances the DNA looping (left panel). However in the presence of 2 µM HU we observed significant suppression of looping (Figure 2.13A, middle panel). This is consistent with previous observation that at such a high concentration, HU forms a filament on the DNA which makes the DNA stiffer. Interestingly, addition of 10 mM Mg^{2+} recovers the dramatic looping in the presence of 2 µM HU. This might be due to screening on the charges which could interfere with binding of HU to the DNA.

Finally we studied the effect of binding of DAPI on the DNA flexibility. DAPI is a dsDNA fluorescent stain which binds to the minor groove of DNA and it is used extensively in fluorescence microscopy to stain the cell nucleus. Our results showed that increasing concentration of DAPI decreases the looping rate. Consistent with this observation, the fractions of looped molecules at equilibrium decreased (Figure 2.13B) by increasing the DAPI concentration.

2.3 Discussion

As discussed in previous section, extended WLC models have previously been developed to explain the remarkable flexibility of short DNA by allowing for the formation of temporary bubbles or kinks, and molecular dynamics simulations observed emergence of kinks in small DNA minicircles. Others accommodate high bendability of short DNA by introducing non-harmonic elastic behavior. To gain insight into the mechanism of facile looping, we performed experiments on DNA constructs with a single backbone nick, double nicks or a single basepair mismatch in the middle and observed one to two orders of magnitude higher looping rate compared to our original DNA constructs (Figure 2.12). This significant enhancement in the looping rate confirms that

stable defects such as a single basepair mismatch or a nick can enhance global cyclizability of DNA, suggesting that similar but transient defects, if they are frequent enough, may explain the extreme bendability observed here. However, whether the high bendability of DNA at short length scales comes from transient kinks or bubbles or stems from anharmonic elasticity of DNA require improved computational methods and further studies.

We also used the assay to quantify the effect of DNA binding proteins or ligands on the global flexibility and cyclization of DNA molecules. For example we could observe the dual architectural role of HU in a single experiment and very efficiently. Our assay should become a fast and versatile tool to study the effect of DNA binding proteins, ligands or DNA modifications on the DNA flexibility.

2.4 Experimental Procedures

Synthesis and purification of oligonucleotides

All oligonucleotides were constructed by annealing, ligation and gel purification of 4 pieces of DNA (Integrated DNA Technologies). The sequences are randomly chosen with about 50% G-C content. For the constructs with 10 nucleotide overhang, Cy3 donor fluorophore is attached to the 5' end of a 35 nt DNA strand. Cy5 acceptor fluorophore is attached to the 5' end of another 35 nt strand which also has a biotin attached to a thymine base through a 5 carbon linker. The two Cy3 and Cy5 strands are common among all the constructs. We used two other unlabeled strands to make sequences with different lengths. Duplexes were annealed and ligated overnight at 16^{o} C using T4 ligase (NEB). Next, the final products were purified by polyacrylamide gels electrophoresis (Amersham Bioscience). Finally the DNA was recovered from the gel and the concentration was determined using NanoDrop (Thermo Scientific).

The constructs for studying the effect of DNA sequence on looping were made by annealing two single stranded DNA strands containing fluorophores and biotin. The annealed construct was then purified by gel electrophoresis and used.

Single molecule FRET measurements

Single molecule measurements were performed on a home-built, wide field prism-type total internal reflection microscope with a 532 nm laser (Coherent Inc.) for Cy3 excitation and a 633 nm laser (Melles Griot) for Cy5 excitation. Details of the setup have been described elsewhere (*16*). To eliminate surface interaction of immobilized DNA during experiments, the chamber surfaces were passivated with PEG polymer (Laysan Bio. Inc.). Briefly, quartz slides and glass cover slips were cleaned using Alconox (VWR International) acetone (Fisher Scientific) and 1 M potassium hydroxide (Fisher Scientific). After cleaning, the surfaces were amino-modified by treating with aminosilane (United Chemical Technologies). Finally the surfaces were coated with 98:2 (mol/mol) mPEG:biotinylated-PEG (Laysan Bio. Inc.). The measurements were done in chambers assembled by double sided tape sandwiched between a quartz slide and a coverslip.

After assembling the channel, NeutrAvidin (Thermo Scientific) was added (0.2 mg/mL) followed by incubation with the biotinilated DNA construct (50-100 pM) for 5 minutes. Finally, free DNA was flushed out with the appropriate buffer for imaging (see below).

Single molecule FRET measurements on surface tethered vesicles

Initially, lipid films were prepared by mixing 98:2 molar ratio of 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC) and 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(cap biotinyl) dissolved in chloroform (Avanti Polar Lipids). The mixture then dried by a stream of nitrogen and vacuumed for two hours. 100 nM of DNA molecules in 10 mM Tris and 50 mM NaCl was used to hydrate the film and form vesicles. Multiple cycles of freeze and thaw (in liquid nitrogen then water) followed by extrusion through a membrane with 200 nm pore size generates unilamellar and uniform vesicles, 200 nm in diameter (*23*). The lipid and DNA concentrations were carefully chosen so that most of the vesicles contain a single DNA molecule. Similar to the previous procedure, after immobilizing vesicles on the PEG surface through NeutrAvidin, the experiments performed in the appropriate imaging buffer.

Single DNA looping assay

After immobilizing DNA molecules on the surface, an imaging buffer which contains 50 mM Tris, 3mM trolox (Sigma-Aldrich) and oxygen scavenger system (0.8% dextrose, 0.1 mg/ml glucose oxidase and 0.02 mg/ml catalase) without cationic ions was flowed in. In this buffer, all DNA molecules on the surface were in the open state and no effective energy transfer detected within our time resolution. We start the experiment by flowing in an imaging buffer, same as above but containing high concentration of cations (0.5-1 M NaCl or 10-30 mM $MgCl_2$). At every time point, we analyzed about 4000 molecules (100 ms imaging time resolution) from different areas (70 um × 35 um) across the chamber (20 mm × 5 mm × 70 μm). The FRET efficiency was measured as $I_A/(I_A+I_D)$ where I_A and I_D are the respective acceptor and donor intensities for each molecule. Two Gaussian functions were used to fit the histogram of FRET efficiency from all molecules at each time point and the fraction of looped population was defined as $A_H/(A_H+A_L)$ where A_H and A_L are the area under the fitted Gaussian curve at high FRET and low FRET respectively. Fresh imaging buffer was reintroduced in the chamber every 30 minutes to avoid acidification of buffer due to activity of the oxygen scavenging system.

Calculating looping and unlooping rates from histograms

The fraction of looped molecules over time follows an exponential decay curve of the form $C\left(1-e^{-R\bullet t}\right)$ where C and R can be determined by fitting. On the other hand, if we assume a simple two state system for the looping process, $unlooping \underset{k_u}{\overset{k_l}{\rightleftarrows}} looping$ where k_l and k_u are looping and unlooping rates, the fraction of looped molecules after time t would be $\frac{A_H}{A_H+A_L}=\frac{k_l}{k_l+k_u}\left(1-e^{-(k_l+k_u)t}\right)$. From this simple model, k_l and k_u can be determined through fitting as $k_l=R\cdot C$ and $k_u=R\cdot(1-C)$ where $C=\frac{k_l}{k_l+k_u}$ and $R=k_l+k_u$.

Finally, for each DNA construct we measured the fraction of molecules with an active FRET pair. This fraction was about 70%-90% for different DNA constructs.

Therefore the fraction of looped molecules at saturation did not reach unity. We included this correction in our looping and unlooping rate measurements.

Calculating looping and unlooping rates from single molecule traces

We also calculated looping and unlooping rates from single molecule traces. We recorded long movies from single molecules (~15 min) with 600 ms time resolution. At the end of the recording we directly excited acceptors with a 633 nm laser to distinguish between unlooped and Cy5 photobleached molecules.

By analyzing several hundred single molecule traces we measured looping rate from $N_{L\rightarrow H}/T_{low}$ where $N_{L\rightarrow H}$ is the total number of transitions from low to high FRET state and T_{low} is the total time spent in low FRET state. Similarly, the unlooping rate was calculated as $N_{H\rightarrow L}/T_{high}$ where $N_{H\rightarrow L}$ is the total number of transitions from high to low FRET state and T_{high} is the total time spent in the high FRET state.

DNA sequences

Bold, underlined T denotes the position of the biotin, attached internally to a thymine base. For vesicle experiments, we used the same DNA sequence, without an internal biotin.

67 bp circle size (10 nt overhang)
Cy3-cagaatccgtgctagtacctcaatatagactccctttgacctgactatcctcacctccaccgtttca
cgatcatggagttatatctgagggaaactggactga**<u>T</u>**aggagtggaggtggcaaagtgtcttaggca-Cy5

69 bp circle size (10 nt overhang)
Cy3-cagaatccgtgctagtacctcaatatagactccctttgacccatgactatcctcacctccaccgtttca
cgatcatggagttatatctgagggaaactgggtactga**<u>T</u>**aggagtggaggtggcaaagtgtcttaggca-Cy5

71 bp circle size (10 nt overhang)
Cy3-cagaatccgtgctagtacctcaatatagactcccttctaattgactgactatcctcacctccaccgtttca
cgatcatggagttatatctgagggaagattaactgactga**<u>T</u>**aggagtggaggtggcaaagtgtcttaggca-Cy5

73 bp circle size (10 nt overhang)
Cy3-cagaatccgtgctagtacctcaatatagactcccttctaattgaccatgactatcctcacctccaccgtttca
cgatcatggagttatatctgagggaagattaactggtactga**T**aggagtggaggtggcaaagtgtcttaggca-Cy5

74 bp circle size (10 nt overhang)
Cy3-cagaatccgtgctagtacctcaatatagactcccttaagttgacccatgactatcctcacctccaccgtttca
cgatcatggagttatatctgagggaaattcaactgggtactga**T**aggagtggaggtggcaaagtgtcttaggca-Cy5

76 bp circle size (10 nt overhang)
Cy3-cagaatccgtgctagtacctcaatatagactcccttctaagttgacctcatgactatcctcacctccaccgtttca
cgatcatggagttatatctgagggaagattcaactggagtactga**T**aggagtggaggtggcaaagtgtcttaggca-Cy5

93 bp circle size (10 nt overhang)
Cy3-
cagaatccgtgctagtacctcaatatagactcccttaatacttctcctatgacttctaattgacccatgactatcctcacctccaccgtt
tca
cgatcatggagttatatctgagggaattatgaagaggatactgaagattaactgggtactga**T**aggagtggaggtggcaaagtg
tcttaggca-Cy5

94 bp circle size (10 nt overhang)
Cy3-
cagaatccgtgctagtacctcaatatagactcccttaatacttctcctatgacttctaattgacccatgactatcctcacctccaccgt
ttca
cgatcatggagttatatctgagggaaattatgaagaggatactgaagattaactgggtactga**T**aggagtggaggtggcaaagt
gtcttaggca-Cy5

97 bp circle size (10 nt overhang)
Cy3-
cagaatccgtgctagtacctcaatatagactcccctatgttaatacttctcctatgacttctaattgacccatgactatcctcacctccac
cgtttca

cgatcatggagttatatctgagggatacaattatgaagaggatactgaagattaactgggtactga**T**aggagtggaggtggcaa
agtgtcttaggca-Cy5

99 bp circle size (10 nt overhang)
Cy3-
cagaatccgtgctagtacctcaatatagactccctagatgttaatacttctcctatgacttctaattgacccatgactatcctcacctcc
accgtttca
cgatcatggagttatatctgagggatctacaattatgaagaggatactgaagattaactgggtactga**T**aggagtggaggtggca
aagtgtcttaggca-Cy5

101 bp circle size (10 nt overhang)
Cy3-
cagaatccgtgctagtacctcaatatagactccctgtagatgttaatacttctcctatgacttctaattgacccatgactatcctcacct
ccaccgtttca
cgatcatggagttatatctgagggacatctacaattatgaagaggatactgaagattaactgggtactga**T**aggagtggaggtgg
caaagtgtcttaggca-Cy5

103 bp circle size (10 nt overhang)
Cy3-
cagaatccgtgctagtacctcaatatagactccctacgtagatgttaatacttctcctatgacttctaattgacccatgactatcctcac
ctccaccgtttca
cgatcatggagttatatctgagggatgcatctacaattatgaagaggatactgaagattaactgggtactga**T**aggagtggaggt
ggcaaagtgtcttaggca-Cy5

105 bp circle size (10 nt overhang)
Cy3-
cagaatccgtgctagtacctcaatatagactccctgaacgtagatgttaatacttctcctatgacttctaattgacccatgactatcctc
acctccaccgtttca
cgatcatggagttatatctgagggacttgcatctacaattatgaagaggatactgaagattaactgggtactga**T**aggagtggag
gtggcaaagtgtcttaggca-Cy5

106 bp circle size (10 nt overhang)

Cy3-

cagaatccgtgctagtacctcaatatagactccctagaacgtagatgttaatacttctcctatgacttctaattgacccatgactatcct
cacctccaccgtttca

cgatcatggagttatatctgagggatcttgcatctacaattatgaagaggatactgaagattaactgggtactga**T**aggagtggag
gtggcaaagtgtcttaggca-Cy5

201 bp circle size (10 nt overhang)

Cy3cagaatccgtgctagtacctcaatatagactccctatcagtacgaagctgggctataccgttcttattgtccttaataccaccg
acgagttgtacgccctctcatccgaagacgacacgtacctgggaaaagaacgtagatgttaatacttctcctatgacttctaattga
cccatgactatcctcacctccaccgtttca

cgatcatggagttatatctgagggatagtcatgcttcgacccgatatggcaagaataacaggaattatggtggctgctcaacatgc
gggagagtaggcttctgctgtgcatggacccttttcttgcatctacaattatgaagaggatactgaagattaactgggtactga**T**a
ggagtggaggtggcaaagtgtcttaggca-Cy5

69 bp circle size with a single C:C mismatch (10 nt overhang)

Cy3-cagaatccgtgctagtacctcaatatagactccctttgacccatgactatcctcacctccaccgtttca

cgatcatggagttatatctgagggaaactcggtactga**T**aggagtggaggtggcaaagtgtcttaggca-Cy5

97 bp circle size with double nicks (10 nt overhang)

Cy3-

cagaatccgtgctagtacctcaatatagactccct|atgttaatacttctcctatgacttctaattgacccatgactatcctcacctcca
ccgtttca

cgatcatggagttatatctgagggatacaattatgaagaggatactgaagattaactgggta|ctga**T**aggagtggaggtggcaa
agtgtcttaggca-Cy5

73bp- TA (10 nt overhang)

Cy3-cagaatccgtagctctagcaccgcttaaacgcacgtacgcgc**T**gtctaccgcgttttaaccgccaataggatt
tcgagatcgtggcgaatttgcgtgcatgcgcgacagatggcgcaaaattggcggttatcctaagtcttaggca-Cy5

73bp- E817 (10 nt overhang)

Cy3-cagaatccgtttttatttatcgcctccacggtgctttttttt**T**ttttttttggccgtgttatctcgagttagt

aaaataaatagcggaggtgccacgaaaaaaaaaaaaaaaaccggcacaatagagctcaatcagtcttaggca-Cy5

73bp- E838 (10 nt overhang)

Cy3-cagaatccgtttttatttatcgttttttttttttttttttt**T**ttttttttttttttttttatctcgagttagt
aaaataaatagcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaatagagctcaatcagtcttaggca-Cy5

73bp-E8A10 (10 nt overhang)

Cy3-cagaatccgtttttatttatcgcctccacggtgctgtttttt**T**tttctgttggccgtgttatctcgagttagt
aaaataaatagcggaggtgccacgacaaaaaaaaaagacaaccggcacaatagagctcaatcagtcttaggca-Cy5

73bp-E8A26 (10 nt overhang)

Cy3-cagaatccgtttttatttatcgcctccttttttttttttttt**T**ttttttttttccgtgttatctcgagttagt
aaaataaatagcggaggaaaaaaaaaaaaaaaaaaaaaaaaaaaaggcacaatagagctcaatcagtcttaggca-Cy5

2.5 Figures

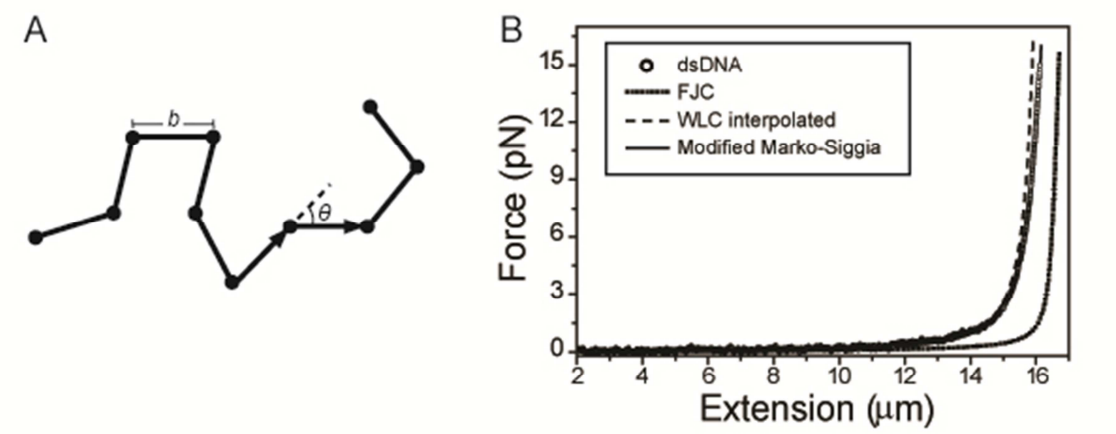

Figure 2.1 (A) Schematic of a freely-jointed chain. The chain consists of non-interacting segments of length "b". Elastic energy of such a chain depends on the angle between adjacent segments. (B) Force-extension data (black circles) for a double-stranded λ-DNA pulled with optical tweezers in 50mM Na^+. The data is fitted to the freely joined chain model (dotted line), WLC-based model by Marko–Siggia (dashed line) and the modified Marko–Siggia model (solid line). The fit to the modified Marko–Siggia model provides the persistence length of P = 50 nm and stretch modulus of 1000 pN for this molecule.

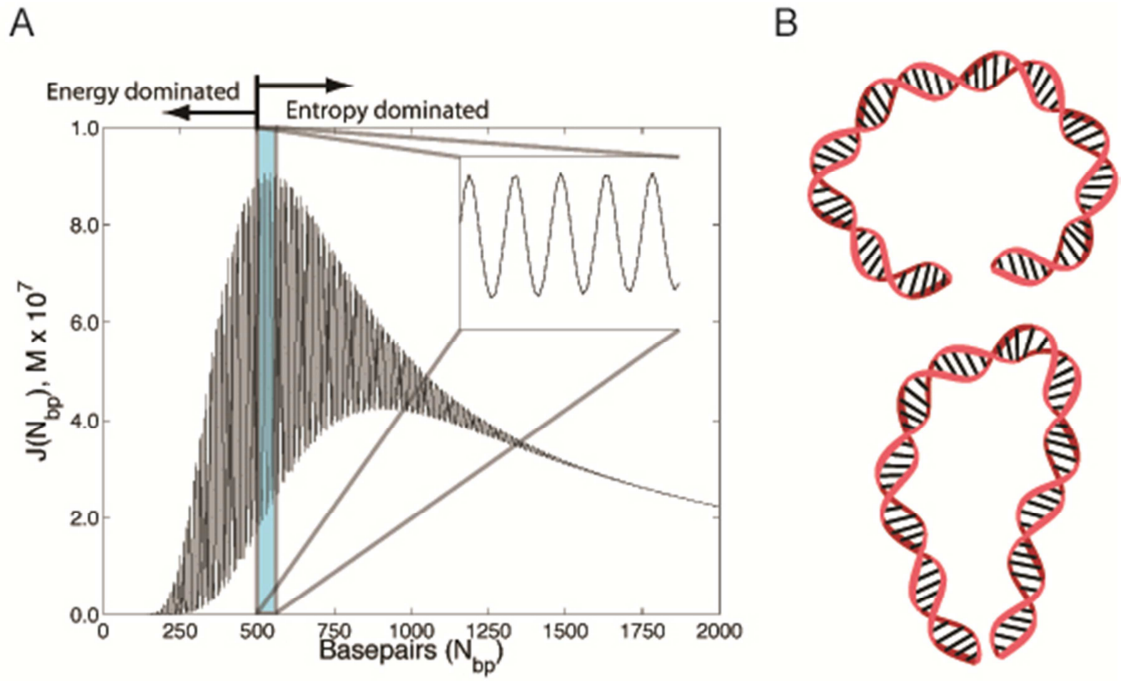

Figure 2.2 (A) The ring-closure probability or *j* factor (units *M*) versus the chain length in number of basepairs (N_{bp}) for the WLC with persistence length $l_p = 53\ nm$, torsional persistence length $l_t = 110\ nm$. The peak of *j* is around 500 bp. The inset shows a close-up of the data in the range $500\ bp < N_{bp} < 550\ bp$. (B) Two important boundary conditions in calculating the *j* factor. Parallel boundary condition (top) relevant to the ligase assay and free boundary condition (bottom)

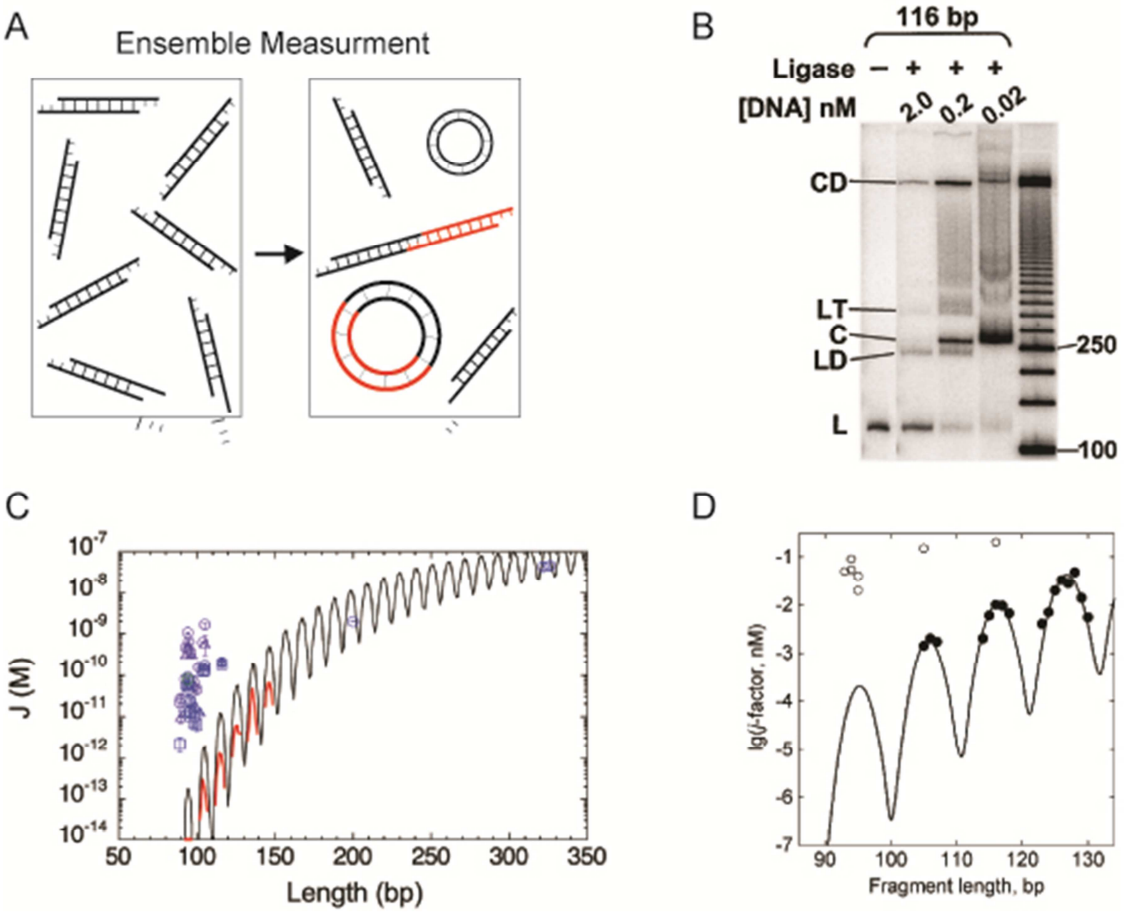

Figure 2.3 (A) Schematic of the ligase assay. Linear DNA molecules with 4 nucleotides overhang are incubated with the ligase. Ligase traps the molecules in whatever conformation they are (eg. circular monomer, dimer, circular dimer ….) (B) The ligation reaction is quenched and the products are separated on a gel. Further analysis of the gel determines the *j*-factor. (C) Measured *j*-factors by Cloutier and Widom (*47*) compared to the predictions of the WLC model (black curve) for DNA cyclization. Although the theory fits experimental data for longer lengths, it deviates up to 6 folds at shorter lengths. (D) *j*-factor of short DNA fragments as a function of their length. Experimental data from Cloutier and Widom (*47*) (open circles) are shown together with the results obtained by Du *et al* (*53*) (filled circles) which fitted well to the WLC model. The discrepancy between the two experiments was assigned to thw wrong ligase concentration used by Cloutier and Widom.

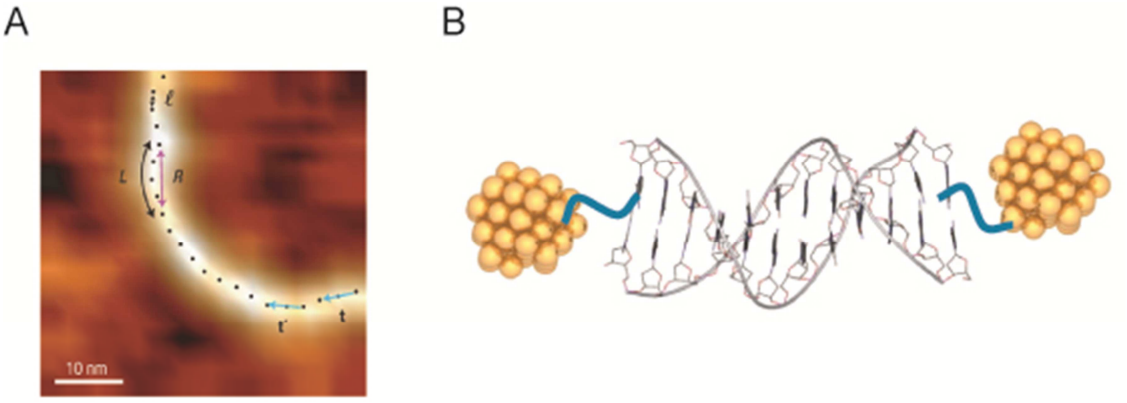

Figure 2.4 (A) AFM image of dsDNA deposited on mica (*54*). In this study, the probability distribution of angles between tangents (short blue arrows) was measured and compared to the WLC prediction. (B) Model of a dsDNA molecule labeled with gold nanocrystal at the ends, used in SAXS experiments in (*23*).

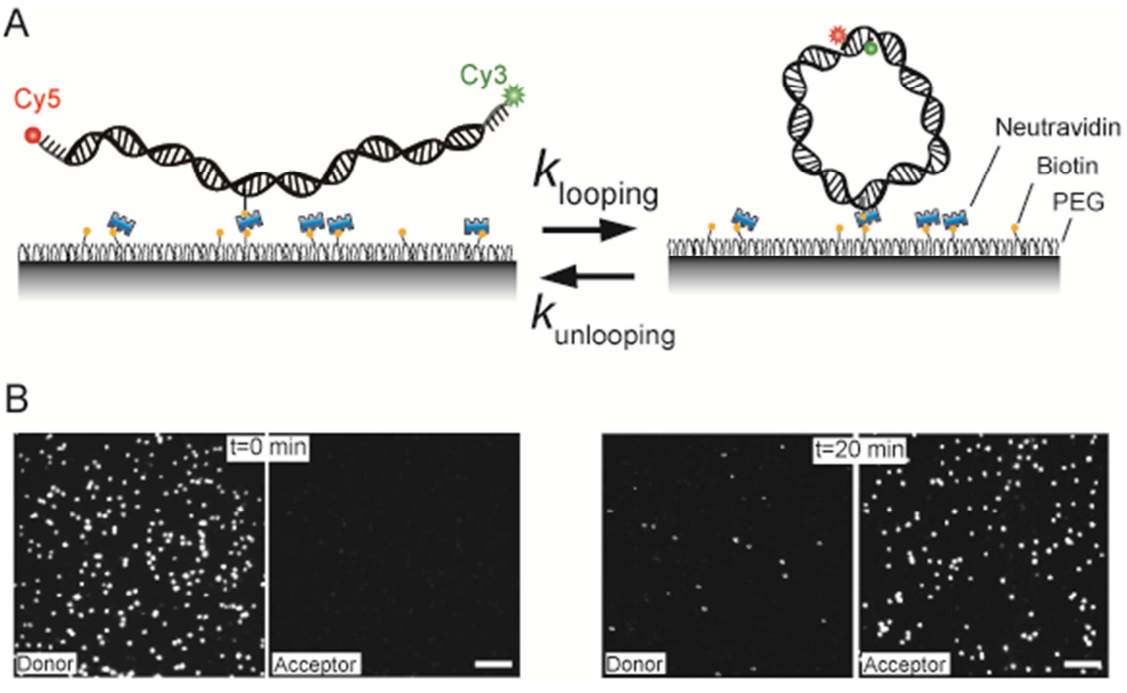

Figure 2.5 (A) Schematic of the single molecule cyclization assay. Donor (Cy3) and acceptor (Cy5) labeled DNA molecules were immobilized on the surface via biotin-

neutravidin interaction. (B) Fluorescence images of single 91-bp DNA molecules in corresponding donor and acceptor channels are shown before (left panels) and 20 min after adding high salt (1 M NaCl) buffer (right panels). Scale bar, 5 μm.

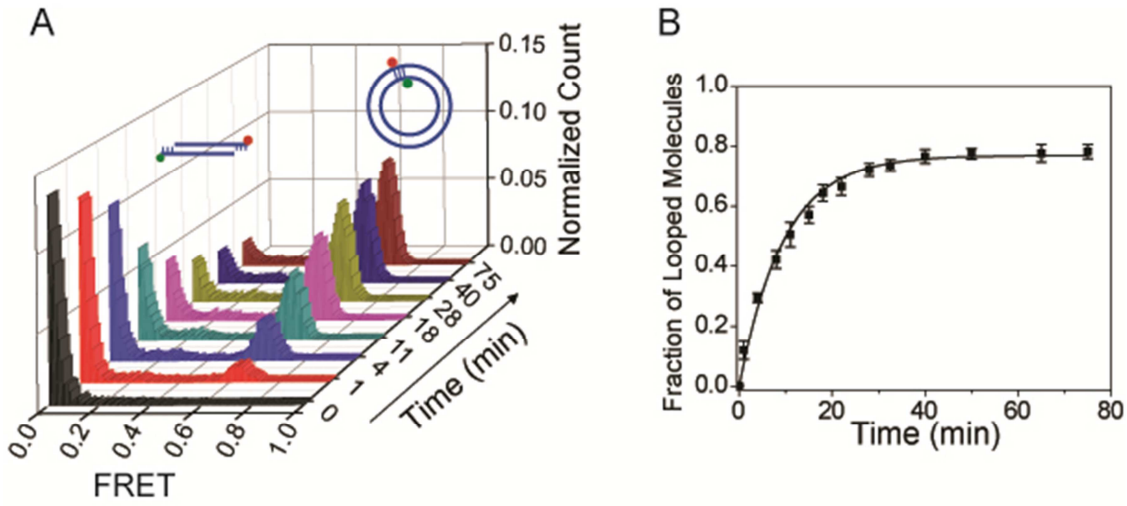

Figure 2.6 (A) Histograms of FRET efficiency as a function of time (t = 0 is when high salt was introduced) show the evolution of looped (high FRET) and unlooped (low FRET) populations. (B) Fraction of looped DNA (high FRET population) as a function of time, measured from the histograms in A. An exponential fit to this curve gives R. Error bars indicate ± SEM; n = 5.

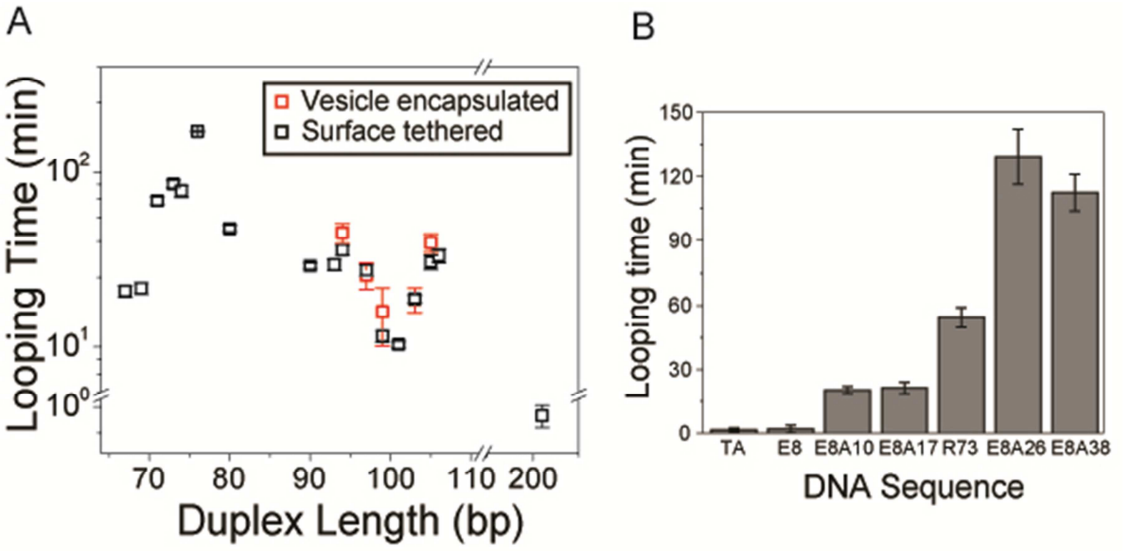

Figure 2.7 (A) Looping time as a function of DNA circular length for surface-tethered DNA (black squares) and vesicle-encapsulated DNA molecules (red squares). (B) Looping time for 7 DNA sequences with 63-bp duplex length and 10-nt overhang. R73 is the standard sequence used in (A). Poly-A constructs were constructed by inserting n = 10, 17, 26, and 38 consecutive A bases in the middle of a random sequence (E8). Error bars indicate ± SEM; $n \geq 3$.

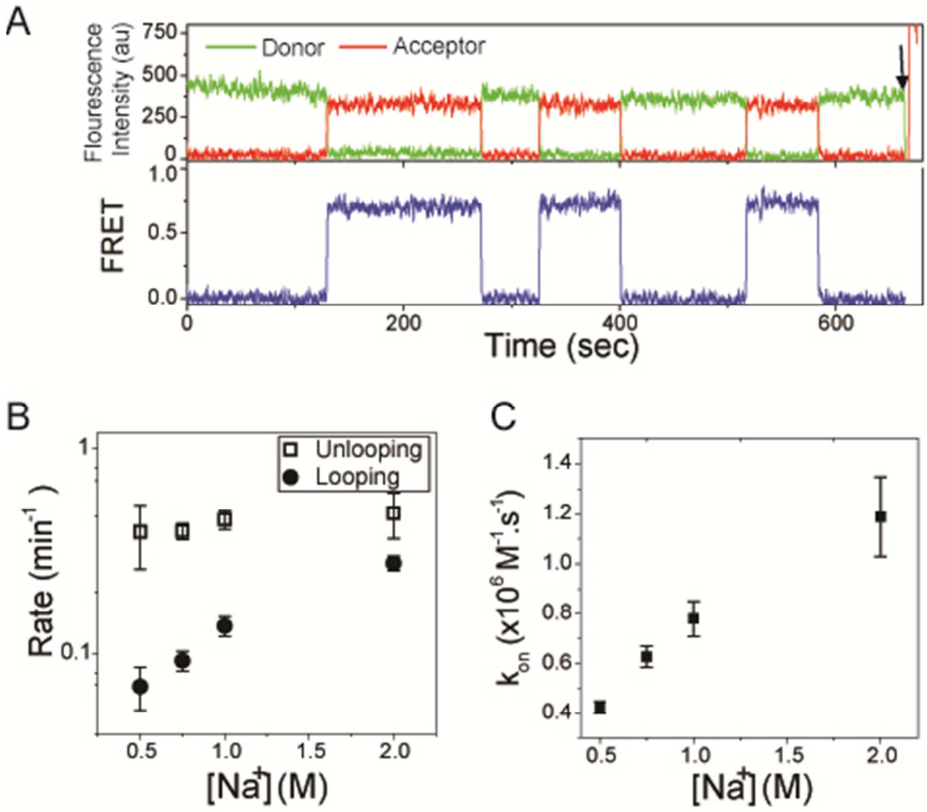

Figure 2.8 (A) Representative fluorescence intensities (top, green for donor and red for acceptor) and corresponding FRET efficiency (bottom, blue) time traces measured from a single DNA molecule in 750 mM NaCl. The DNA has 91-bp initial dsDNA with 8-nt single-stranded overhangs. The arrow indicates a direct acceptor excitation to verify that the acceptor has not photobleached. (B) Looping and unlooping rates as a function of [NaCl]. The DNA has 91-bp initial dsDNA with 8-nt single-stranded overhangs. (C) k_{on} measured as shown in fig. S2 shows the same 3-fold increase as the looping rate with increasing [NaCl]. Data are means ± SEM ($n \geq 300$ molecules).

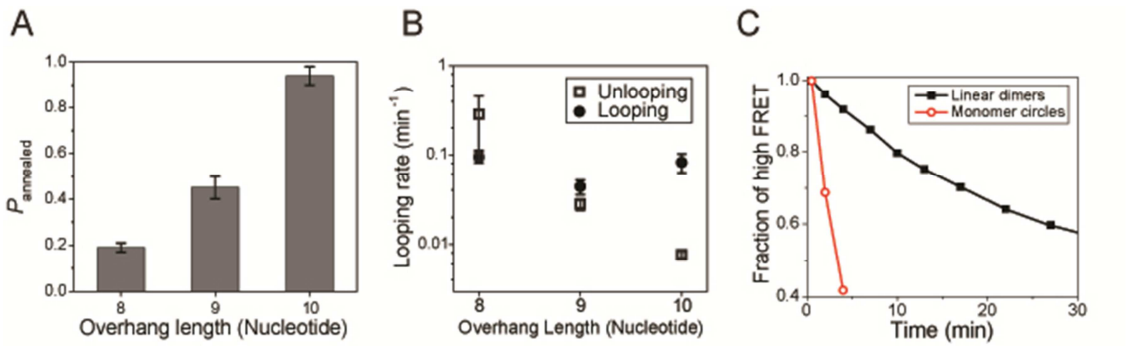

Figure 2.9 (A) Equilibrium fraction of DNA circles, $P_{annealed}$, as a function of single stranded overhang length. All constructs have the same 91 bp initial dsDNA but with different overhang lengths (8, 9 or 10 nt). (B) Looping and unlooping rates measured as a function of ss-overhang length. All constructs have the same 91 bp initial dsDNA but with different overhang lengths (8, 9 or 10 nt). Data are means ± SEM ($N \geq 300$ molecules). (C) Comparing melting of DNA dimers formed from DNA monomers with 8 nt single stranded overhang and monomer DNA circles of the same size and overhang. The melting rate of dimers, determined by an exponential decay fit is about 0.013 min^{-1}, which is more than 20 times smaller than the melting rate of DNA loops formed using the same 8 nt overhangs on 91 bp initial dsDNA (0.29 min^{-1}).

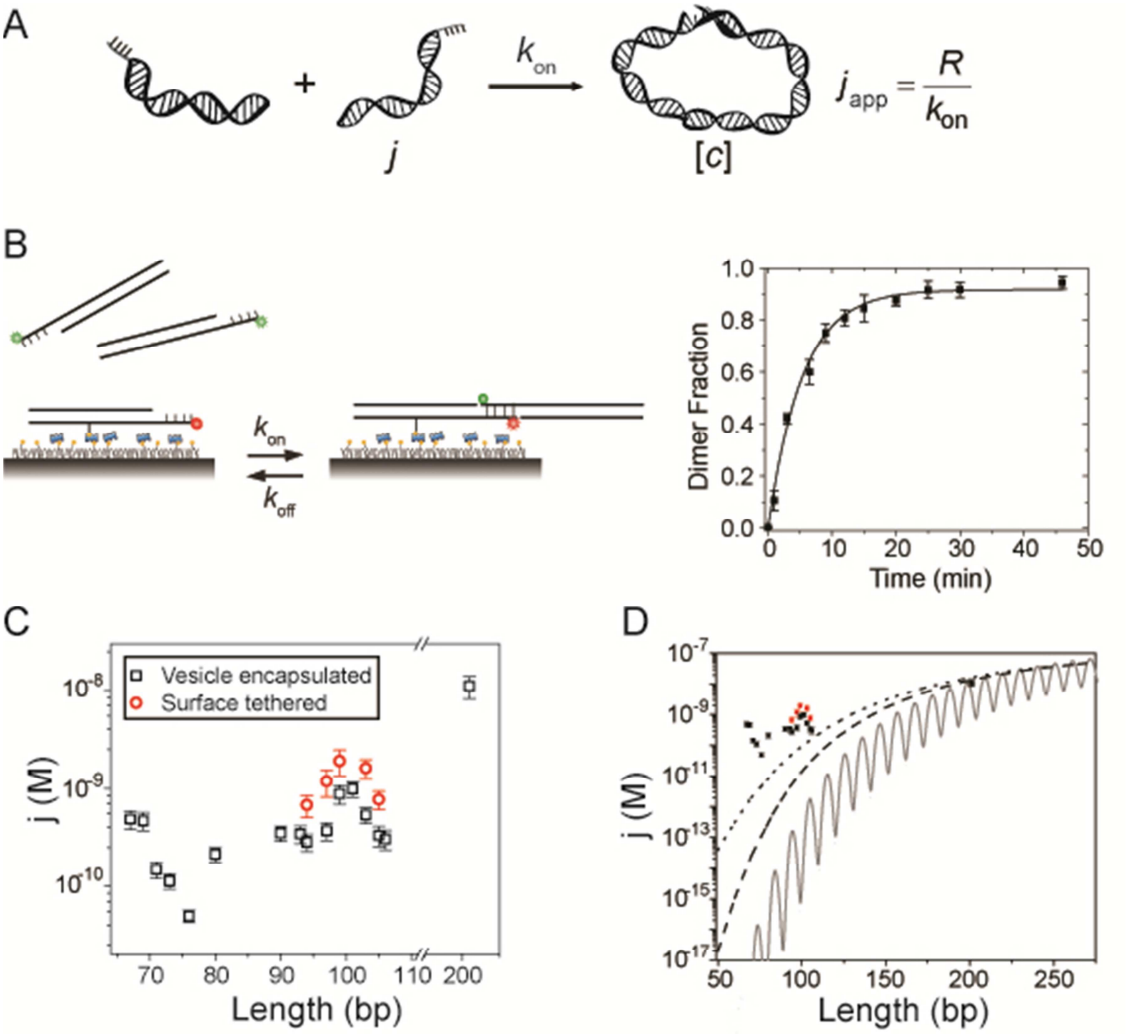

Figure 2.10 (A) The model to relate *R*, k_{on}, and apparent *j* factor. (B) Schematic of single molecule FRET assay to measure bimolecular association rate k_{on}. By measuring the fraction of high FRET DNA molecules, as a function of time and by fitting the data to an exponential curve, one can determine the apparent rate of dimer formation. Dividing the apparent binding rate by the donor-labeled DNA concentration in solution gives k_{on}. Data are means ± SEM ($N \geq 5$). (C) *j* factor for surface-tethered DNA (black squares) and vesicle-encapsulated DNA (red circles). (E) Measured *j* factor for surface-tethered DNA (black squares) and vesicle-encapsulated DNA (red squares). Solid black curve is the Shimada-Yamakawa prediction for DNA cyclization. Dashed line and dotted line are the WLC predictions for the *j* factor of DNA circles with free boundary condition and for DNA molecules with 5-nm capture radius, respectively. Error bars indicate ± SEMn ≥ 3.

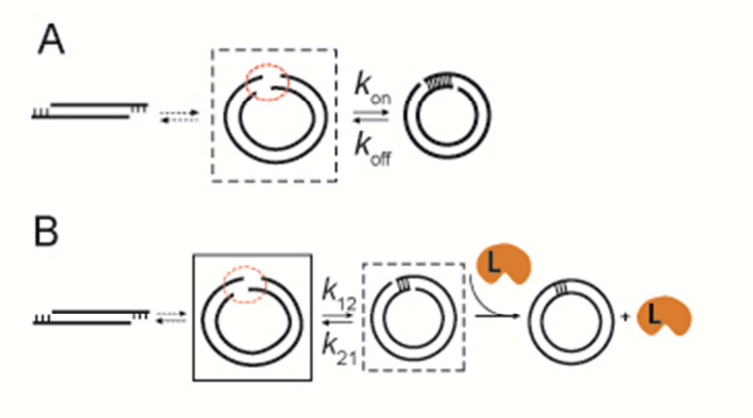

Figure 2.11 (A) Our assay reports on the equilibrium population of the intermediate state (dashed box), with the two DNA ends in close proximity. (B) Schematic representation of DNA cyclization reaction steps in the ligase assay. The intermediate state with the two DNA ends in close proximity (solid box) does not get sampled in this assay. Instead, the ligase samples the equilibrium population of the annealed state (dashed box). Ligase protein is labeled L.

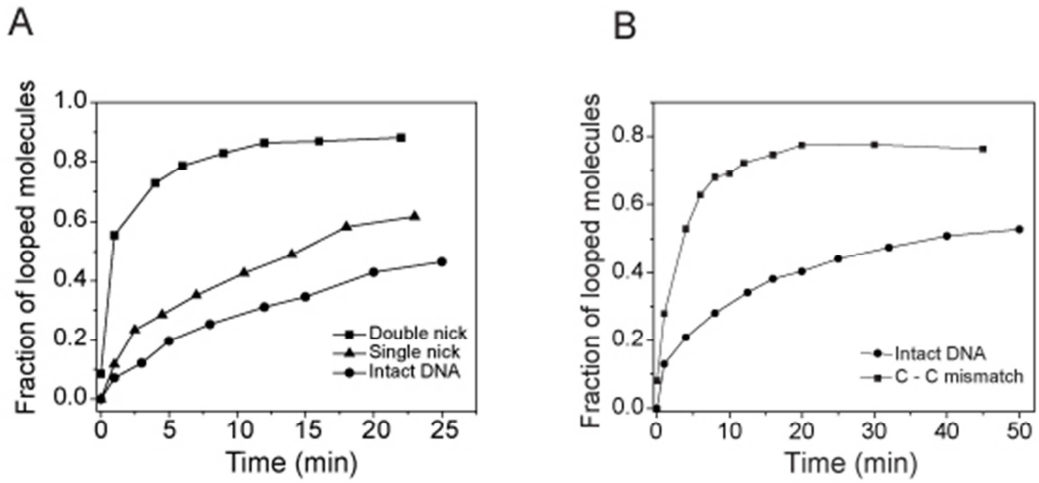

Figure 2.12 (A) Effect of nick on DNA looping for DNA constructs with one or two backbone nicks. The DNA has 87 bp initial dsDNA with 10 nt overhangs. (B) Effect of single base mismatch (C:C) on DNA looping. The mismatched DNA circularize 5 times faster than the intact DNA. The DNA has 59 bp initial dsDNA with 10 nt overhang.

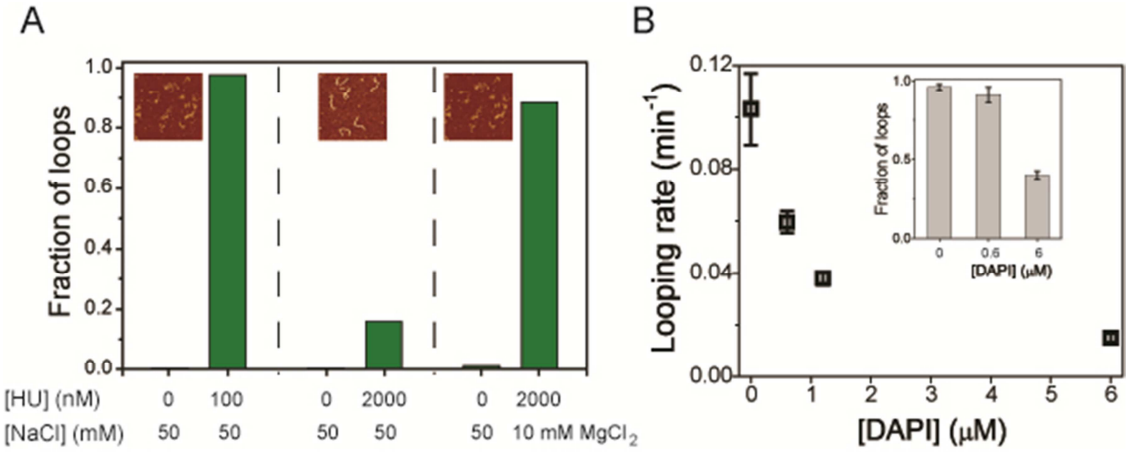

Figure 2.13 (A) Effect of HU protein on the cyclization of dsDNA. Dual architectural role of HU protein can be observed very easily through our cyclization assay. While at low concentration (100 nM) HU dramatically enhances DNA looping, at high concentration (2 μM), it greatly suppresses DNA looping. However, high Mg^{2+} can recover efficient looping even at high HU concentration. (inset shows AFM image of HU bound to dsDNA from (*83*)) (B) DAPI decreases the looping rate, observed through our cyclization assay.

Chapter 3
Studying the Genome Composition of Influenza Virus at the Single Virus Particle Level[†]

3.1 Introduction

Influenza A virus is among the most prominent health threats around the world. In the United States alone, influenza causes more than 40,000 deaths and about 1.5 million hospitalizations each year (*84*). Due to its impact on the global health, it is important to have a comprehensive understanding of the Influenza virus biology and its life cycle. Influenza A virus is a negative-sense single stranded RNA virus and a member of *Orthomyxoviridae* viruses (from the Greek word *orthos*, meaning "standard, correct," and *myxa*, meaning "mucus") (*85*). The genome of influenza A virus is composed of eight single-stranded RNA segments that encode for 11 viral genes: hemagglutinin (HA), neuraminidase (NA), matrix 1 (M1), matrix 2 (M2), nucleoprotein (NP), non-structural protein 1 (NSP1), non-structural protein 2 (NS2), polymerase basic protein 1 – F2 (PB1-F2) and three subunits of the influenza RNA polymerase (PB1, PB2 and PA) (*86*). The size of viral RNA (vRNA) segments and the genes they encode are listed in table 3.1 (*87*). The vRNA segments share a common architecture; a long central coding region flanked by relatively short untranslated regions (UTRs) and partially complementary sequences at the termini (Figure 3.1A). Nucleoprotein (NP) wraps around individual vRNAs in a nonspecific manner to form ribonucleoprotein (RNP) structures (Figure 3.1B). One copy of the three influenza polymerase proteins (PB1, PB2 and PA) is also bound to the end of each RNP structure (Figure 3.1C).

[†] This work has been published as a paper:

- Chou, Y.Y., R. Vafabakhsh, S. Doğanay, Q. Gao, T. Ha and P. Palese, "One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis", *PNAS* **109**, 9101-9106 (2012).
- Gao, Q., Y. Y. Chou, S. Doganay, R. Vafabakhsh, T. Ha and P. Palese, "The influenza A virus PB2, PA, NP and M segments play a pivotal role during genome packaging", *J. Virology* **86**, 7043-7051 (2012).

The segmented nature of influenza genome provides the virus with an evolutionary advantage. If a single host cell gets infected by two different strains of influenza A virus, the new virus particles can have RNA segments that originated from either of the infecting viruses. This process of mixing of the genetic materials from multiple virus strains to form a viral strain with novel gene combination is called reassortment. Genetic reassortment plays a vital role in the evolution of the influenza virus and has historically been linked to the emergence of pandemic strains. For example, the 2009 H1N1 pandemic strain is a reassortant of avian, human, and swine influenza viruses (*88*) (Figure 3.2).

Genome segmentation of influenza acts as a double-edged sword. Although, as it was mentioned above, it enables the virus to produce novel genetic combinations through reassortment, it provokes a challenge in the virus assembly step. In order for a single virion to be infectious, it needs to package at least one copy of each vRNA. There are two models for the packaging of influenza vRNAs: (i) the random packaging model and (ii) the specific packaging model.

In the random packaging model, there is no mechanism to distinguish between different vRNA segments and a random number of vRNA segments get packaged in each virion (Figure 3.3A) (*89*, *90*). According to this scheme, a fully infectious virion would acquire a complete genome purely through chance, with the probability of success being increased by packaging more segments than eight. If the eight influenza viral RNA segments were randomly packaged into new particles, we would expect to observe one infectious particle for every 400 particles assembled ($8!/8^8$). This value is incompatible with the approximately 1/10–1/100 proportion of infectious influenza virions (*91*). However, if each virion could package more than eight vRNA, this ratio would increase further. In fact, this ratio could fall within the range of infectious to noninfectious particles that occur in virus stocks if each virus could non-specifically package 10-11 vRNA segments (*89*). The observation of virus particles with nine or more vRNA segments supports the random incorporation model (*89*).

The selective packaging model on the other hand suggests that each vRNA contains unique features that distinguish it from the other vRNA segments (Figure 3.3B). Therefore, one copy of each vRNA segment is specifically packaged during the viral

assembly (*92-94*). The identification of segment-specific packaging sequences located at the terminal ends of all eight vRNAs supports the specific packaging mechanism (*94, 95*). Further support for the specific packaging model came recently with the electron microscopy analysis of influenza virus particles that showed eight RNP complexes inside the virus arranged in a "7+1" architecture (*96, 97*) (Figure 3.3C).

Even though there has been growing evidence supporting the selective packaging mechanism, conclusive proof that each virus particle contains eight unique vRNA segments has been missing, mainly because of the low sensitivity and low yield of current bulk methods. In order to study the genome packaging of influenza viruses with high sensitivity, a method that allow stoichiometric analysis of vRNAs within single virus particles and with individual vRNA sensitivity is required.

In this chapter I describe an experimental scheme which allows the detection and quantification of the viral RNAs within influenza virions at single-virus resolution and single vRNA sensitivity. To achieve the individual v RNA specificity we adapted the single-molecule fluorescent in situ hybridization technique (smFISH) (*98*). Moreover to accomplish single particle resolution, we used the total internal reflection fluorescence microscopy (TIRF) on the surface immobilized virus particles (*60*). Using this combined approach we showed that most influenza virus particles package eight distinct viral RNAs. These data provide direct evidence that the packaging mechanism in the influenza virus is very robust and supports a specific mechanism for the influenza genome packaging.

3.2 Experimental Results

3.2.1 Assay Design and Validation

We used prism-type total internal reflection fluorescence microscopy (TIRF) (*60*) to probe individual virus particles immobilized on the chamber surface. The experimental chamber was assembled as described previously. In brief, a double-sided tape is sandwiched between a microscope slide and coverslip. The typical dimension of the experimental chamber is about 16 mm × 5 mm × 100 µm. Furthermore, to eliminate the nonspecific sticking of virus particles, the slide surface was passivated with PEG polymer (polyethylene glycol) sparsely supplemented with biotinylated PEG, as described in

appendix A. Influenza A/Puerto Rico/8/34 (PR8) virus particles were anchored onto this passivated surface through biotinylated anti-HA antibody (Figure 3.4A). The surface density of virus particles was such that individual virus particles could be resolved as well-separated diffraction limited spots (~250 particles in 2,500-μm^2 imaging area). After immobilizing the virus particles on the surface, fluorescently labeled FISH probes were used for probing a specific target vRNA.

We first validated the assay by performing smFISH using a set of 48 Cy3-labeled probes against the NA vRNA or a set of 48 Cy5-labeled probes against the PB2 vRNA segments. In each case, more than 50 fluorescent spots were detected (Figure 3.4B and 3.4C). On the other hand, control experiments using FISH probes against a non-influenza RNA (RIG-I) or performing the assay in chambers without the surface immobilized virus particles showed only few dim spots (Figure 3.4B and 3.4C). The concentrations of the antibody, virus particles, FISH probes and incubation times were the same between the control and the positive experiment. Therefore, we believe that the reproducible large difference in the number of spots (more than an order of magnitude) observed between control experiments and experiments probing for influenza vRNAs demonstrated the high specificity of our assay in detecting influenza vRNA segments.

3.2.2 Quantifying the Co-packaging of vRNAs

After establishing the smFISH assay to probe for specific vRNA, we set to quantify the co-packaging of different vRNAs in the individual virus particles. We performed hybridization with two sets of probes targeting different viral RNAs, one probe set labeled with the Cy3 fluorophore and the other set labeled with Cy5 fluorophore (Figure 3.5B).

First, in order to evaluate the specificity and accuracy of our colocalization assay, we performed the hybridization using two sets of probes with different fluorophores (Cy3 and Cy5) targeting two different regions of the same vRNA (Figure 3.5A). The two sets of probes against different regions of the NA segments showed over 90% (93.1%±2.2%) colocalization between Cy3 and Cy5 spots and 84.8%±2.0% colocalization for probes against the M segment. The percentages of spots showing only Cy3 or Cy5 signals were comparable and were all below 10% in both cases, indicating that the two differently

labeled probe sets share similar sensitivity and detect comparable number of viral RNAs (Figure 3.5C). While we observed similar intensity distributions between colocalized spots and spots that were not colocalized, intrinsic defects of some virus particles or the limitation of our detection and analysis could be the reason why we did not see 100% co-localization between the Cy3 and Cy5 labeled probes against the same vRNA. It is worth mentioning that we also consistently observed a slightly higher number of spots detected in the Cy3 channel. This is due to the fact that the fluorescent emission of "junk" particles on the surface, introduced with antibodies or the virus particles, fall mainly in the Cy3 channel.

To determine if influenza virus efficiently packages vRNA of different identities, we first quantified the co-packaging efficiency PB2 segment, detected using Cy3 labeled FISH probes, with the other seven vRNA segments, detected using Cy5 labeled probes. Colocalizataion efficiency was calculated as the number of colocalized Cy3 and Cy5 spots over the total number of Cy3 fluorescent spots corresponding to the PB2 segment. The PB2 segment showed high percentages of colocalization with all the other viral RNA tested (Figure 3.6A). These results demonstrated that the PB2 segment was efficiently co-packaged with other seven vRNAs into virus particles and over 50% of virus particles that packaged the PB2 segments also packaged the other seven segments ($0.911\times0.944\times0.922\times0.9\times0.889\times0.9\times0.9$). Colocalization analysis of the PB1 vRNA with the NA vRNA or the M vRNA also showed over 90% colocalization (97.0%±1.86% for PB1 and NA, 95.4%±2.93% for PB1 and M) (Figure 3.6B). In all these experiments, Cy5 labeled probes targeting the paired vRNAs detect comparable number of spots to the Cy3 probes targeting the PB2 segment. This also results in high percentages of a particular vRNA that co-packaged with the PB2 segment suggesting that other seven gene segments are efficiently co-packaged with the PB2 segment as well (Figure 3.6C). These data firmly suggest that influenza virus proficiently packages heterogeneous viral RNAs into virons and majority of virus particles the eight different vRNA segments.

3.2.3 Quantifying the Copy Number of Each Packaged vRNA

In order to investigate if the heterogeneous viral RNAs packaged into each virus particle were unique, we have to determine the copy number of packaged vRNA

segments. To do so we hybridized Cy5 labeled FISH probes to a specific vRNA. We then photobleached the fluorescent spots and counted the number of photobleaching steps. The number of photobleaching steps for each spot corresponds to the number of FISH probes that were hybridized to the specific vRNA inside that virus particle. If a vRNA segment is unique in the virus particle, only one copy of that RNA should be detected by the single-virus particle FISH analysis. In other words, the distribution of the number of photobleaching steps among a population of virus particles should exhibit a singly peaked profile.

We first performed single-virus photobleaching analysis on the HA vRNA segment in the PR8 virus. In this experiment, fifteen HA targeting FISH probes singly labeled with Cy5 were used for hybridization (Figure 3.7B). Movies of several imaging areas were obtained and the time traces of fluorescence intensity for the spots were analyzed. As the fluorophores photobleache over time the fluorescence intensity drops, yielding a step-wise profile. The number of steps corresponds to the number of probes that were hybridized to HA target (Figure 3.7A). We then counted the number of photobleaching steps for over 800 fluorescent spots corresponding to the HA vRNA. Figure 3.7D (black data points) shows the distribution of the number of photobleaching steps for FISH probes against the HA vRNA in the PR8 virus particles. A single sharp peak around seven photobleaching steps was observed, demonstrating that on average seven probes are likely to hybridize with an HA viral RNA in a single virus particle using our experimental design. In addition, particles displaying more than 15 photobleaching steps were observed and because a maximum of 15 probes could be hybridized to one viral RNA segment, these particles likely represent aggregates of virus particles or virus particles containing multiple targeted viral RNAs.

In order to demonstrate that PR8 virus particles package only one copy of the HA RNA instead of two or more copies, we performed the photobleaching analysis with a virus strain that was reverse engineered to carry two copies of the HA gene (*99*). In this virus, PB2 packaging signals were used to package additional copy of the HA gene. PB2 gene itself was packaged using the NA packaging signals. The engineered 2HA-PR8 virus carries the mutated PB2 segment and two HA segments. To determine if both HA segments were incorporated into the virus particles, viral RNA extracted from purified

2HA-PR8 virus was resolved using RNA electrophoresis followed by silver staining. The PB2-HAmut-PB2 segment migrating between the PA segment and the HA segment was observed, showing that the additional HA segment was incorporated into virus particles (Figure 3.6C).

To test if the photobleaching analysis has the resolution to resolve the number of target segments packaged into virus particles, we performed FISH and photobleaching analysis with the 2HA-PR8 virus under the same experimental conditions as the wild type. The histogram of photobleaching steps for the probes against the HA segment in the 2HA-PR8 virus exhibited two peaks (Figure 3.7D, red data points). The first peak appeared around seven photobleaching steps which was the same peak position as for the PR8 virus. The second peak appeared around 13 steps, which is approximately double the position of the first peak. These results showed that in the 2HA-PR8 virus population, there was a populations of virions that contained the same number of HA segment as the wild type PR8 virus and another population of virus particles that packaged as twice as many. To further confirm that the appearance of the second peak in the 2HA-PR8 virus is not due to intrinsic disorganization of viral RNA packaging, we performed FISH and photobleaching analysis using fifteen Cy5 probes against the NP segment in the wild type PR8 and 2HA-PR8 virus. The photobleaching step histograms of NP segment in both viruses displayed single peak distributions, demonstrating that 2HA-PR8 virus was not defective in packaging unmodified viral RNAs and the stoichiometry of the packaged NP segment was the same as that of the wild type virus (Figure 3.7E). Since it was unlikely that a 2HA-PR8 virus packaged 4 copies or more HA segments, these data strongly suggested that the double peak distribution seen was representative of two viral HA segments being incorporated into a single virion. Therefore, the single peak distribution of the photobleaching step histogram seen for the HA segment in PR8 virus arguably represents only one copy of the HA segment packaged into virus particles. Our data also strongly suggests that a single copy of NP is incorporated into a single virion.

We next wanted to investigate whether only a single copy of the other viral RNA segments were package into PR8 virus particles. We performed FISH and photobleaching step analysis for the other six segments: PB2, PB1, PA, NA, M and NS. In each case fifteen Cy5 probes were used for hybridization and photobleaching steps for more than

500 fluorescence spots were counted. The histograms of the photobleaching steps for the segments examined all exhibited singly peaked profiles (Figure 3.8). Gaussian distribution fits to the photobleaching profiles in Figure 3.8 showed that more than 90% of the virus particles package a single copy of a specific gene (Figure 3.9 A and B).

3.3 Discussion

Using two-color single-particle FISH analysis we achieved direct quantification of co-packaging of different vRNAs into the same virion. The PB2 segment was shown to be co-packaged with each of the other vRNA segments at efficiencies around 90%. This indicated that a large proportion of virions that packaged the PB2 segment also package the other seven segments. It is of note that colocalization between Cy3 and Cy5 FISH probes against different segments of the same vRNA was only approximately 90%, possibly due to detection and analysis limitations or defects in some vRNA (Figure 3.5C). Therefore we are underestimating the percentage of virions that incorporate all the eight vRNA segments due to the limitations of our detection or uneven hybridization probability for different FISH probes. Nevertheless, these results quantitatively show that heterogeneous viral RNA segments are selectively packaged within an influenza virion with high efficiency.

We applied photobleaching analysis to resolve the copy number of a specific viral RNA segment that is packaged into individual virions. Results from the analysis of the number of HA vRNA segments in the wild type PR8 and 2HA-PR8 viruses showed that only one copy of the HA vRNA was incorporated per PR8 virus particle. Analysis of other viral RNA segments also suggested that a single copy of each vRNA segment was packaged into individual virus particles. Although we estimated about 90% of virus particles contain single copy of a specific gene, we believe that this fraction is an under-estimation of the true percentage of virus particles that package single copy of each vRNA. It has been shown that about 10% to 20% of virus particles would self-aggregate after purification (*100*). Moreover the immobilization of two virus particle within a diffraction limited area (1% to 4% of virus particles at the particle density that we do our experiments) or uneven hybridization probability among probes targeting the same vRNA

(due to the fact that vRNA segments within the virus envelope are non-specifically wrapped around with proteins in the form of ribonucleoprotein structures) can generate a shoulder towards higher number of photobleaching steps. Considering all these points, we believe that the majority of the virus particles contain a single copy of a specific gene. Thus, we believe the percentage of virus particles containing single copy of each vRNA is close to 100% and the percentage of virus particles that package eight single copy of each of the eight viral RNA is much higher than the estimated 40% (0.9^8) (Figure 3.6). Normalized Cy3 and Cy5 intensities of individual fluorescent spots, in the case of hybridization with Cy3 and Cy5 FISH probes against different regions of the NA vRNA, showed very high intensity correlation between these channels. This confirms that the spots with high intensity most probably are coming from virus aggregates (Figure 3.9C and D).

3.4 Conclusion

In conclusion, we have established a new experimental scheme with single virus particle sensitivity to interrogate the influenza vRNA for its identity and composition. This technique allowed us to understand the stoichiometry and composition of the segmented vRNAs in the influenza virus particles. Using this technique, we showed that one influenza virus particle packages eight unique viral RNA segments, providing quantitative evidence that the genome packaging of influenza virus is a highly selective process. This novel strategy can be further applied to study the packaging mechanisms of other viruses with a segmented genome.

3. 5 Experimental Procedures

Virus Purification

Wild type PR8 virus and the 2-HA PR8 virus were grown in 10-day embryonated chicken eggs (Charles River Laboratories, SPAFAS) at 37°C for 48 hours and 60 hours, respectively. The embryos were then killed and the allantoic fluid was harvested. The allantoic fluid was clarified by centrifugation at 4000 rpm at 4°C for 30 minutes. The clarified allantoic fluid was then layered on a 20% sucrose cushion and centrifuged at

4°C, 25000 rpm for 2 hours using a Beckman SW28 rotor. The pelleted virus was resuspended in 1XNTE buffer (100 mM NaCl, 10mM Tris-HCl pH 7.4, 1mM EDTA) and layered onto a 10% - 30% iodixanol density gradient (Sigmal-Aldrich, MO) for centrifugation at 25000 rpm for 3 hours at 4°C. The virus was extracted from the gradient using a syringe and pelleted by centrifugation at 25000 rpm at 4°C, for 1.5 hours. The purified virus was resuspended in 1X NTE buffer and stored at -80°C before use.

Sample Preparation

Flow chambers were prepared on polyethylene glycol (PEG) passivated slides doped with biotinylated PEG which were then bound with NeutrAvidin (Thermo). To immobilize the influenza virus particles, biotinylated mouse-monoclonal anti-hemagglutinin antibodies were captured on the surface by incubating the antibody at a 15 nM concentration for 20 min at room temperature. The virus was diluted in T50 buffer (10mM Tris-HCl, pH 8.0, 50mM NaCl) and incubated for 30 min at room temperature in the chamber with immobilized antibody. The PR8 virus was diluted 1:150 and the 2-HA PR8 virus was diluted 1:30 from the stock to obtain well-isolated spots on the surface (~250 particles per 2500 μm^2 imaging area). After each step, the unbound antibodies and virus particles were washed away with T50 buffer. The antibodies and virus were then fixed with 4% paraformaldehyde in T50 buffer for 10 min. After one wash with T50 buffer, the virus particles were permeablized by 10 min incubation with 0.25% Triton X-100 to expose the viral RNPs. The flow chambers were then washed twice with T50 supplemented with 2mM RNase inhibitor vanadyl ribonucleoside complexes (VRC) (New England BioLabs) before hybridization.

Fluorescence *in situ* hybridization (FISH) was performed following protocols published earlier (*98, 101*). For each influenza viral RNA, 30-48 probes were designed and synthesized (Biosearch technologies, Novato, CA). The probes were labeled with Cy3 or Cy5 fluorophores and HPLC purified according to published protocol (*102*). The hybridization reactions were carried out by incubating the permeabilized virus with the hybridization solution containing each probe at a 4 nM concentration at 37°C for 3 hours. Different numbers of probes were mixed for the hybridization reactions according to experimental purposes. After hybridization, the chambers were washed with the wash

buffer (2X SSC, 10% formamide and 2mM VRC) at 37°C for 30 min and then incubated in 2X SSC before imaging.

Single-Molecule Data Acquisition

Single-molecule imaging was performed using a prism type total internal reflection fluorescence (TIRF) microscope as described earlier (*60*). Briefly, a 532 nm Nd:YAG laser and a 632 nm Helium:Neon laser was guided through the prism to generate an evanescent field of illumination for Cy3 and Cyu5 excitation. A water-immersion objective (60X, numerical aperture 1.2, Olympus) was used to collect the signal and the Cy3 back-scattered light was removed using a 550 nm long-pass filter. A 635 nm notch filter was used to allow for the simultaneous imaging of Cy3 and Cy5. The fluorescence signal was sent to a high-speed EMCCD camera (iXon DV 887-BI, Andor Technology). Time resolution of 0.1 sec was used for the data acquisition. The imaging buffer consisting of 4 mM Trolox, 2X SSC buffer, 2mM VRC and with an oxygen scavenging system (1 mg/ml glucose oxidase, 0.4% (w/v) D-glucose and 0.04 mg/ml catalase) was injected into the flow chamber before data acquisition. The measurements were performed at room temperature (22 ± 1°C).

Photobleaching Analysis

The surface containing virus particles with hybridized FISH probes against a specific vRNA were imaged until all of the fluorescent spots were photobleached. Single-molecule fluorescence time traces for each spot were acquired and analyzed for the number of photobleaching steps by a semi-automated algorithm. As distinct bleaching steps could be identified, the average step size of a single bleaching step was calculated for each trace and the total number of photobleaching steps was determined based on the starting fluorescence intensity of that trace. All the fluorescence traces were examined manually, traces with no clear photobleaching steps or derived from fluorescent spots exhibiting oval shape (virus particles aggregates) were discarded. In each case, at least 300 particles were analyzed.

Colocalization Analysis

Colocalization between Cy3 and Cy5 spots was analyzed as previously described (*103, 104*). In brief, coordinate of fluorescent spots upon Cy3 and Cy5 excitation were recorded. Then by using a mapping file, the two channels were mapped onto each other. If the center of two spots were closer than 1.5 pixels we considered the two spots as colocalized. Our control experiments showed that at the surface density that we performed our experiments this criteria would produce a maximum of 4% false colocalization. The colocalization efficiency was calculated as the percentage of the number of colocalized spots over the total number of Cy3 spots.

Fitting Analysis

To determine the percentage of particles containing single copy of each vRNA the photobleaching histograms were fitted with double Gaussian distributions. Initially we determined the position of the first peak by fitting a single Gaussian function to the first half of the data. Then a double Gaussian fit was done to the data by fixing position of the first peak and bounding the position and width of the second peak. All fits were done using Origin Lab software. The accuracy of the fits was also verified by using a fitting MATLAB (Mathworks) code.

3.6 Figures and Tables

Segment	Segment length (nucleotides)	Encoded protein(s)	Protein (amino acids)	Protein function
1	2341	PB2	759	Polymerase subunit; mRNA cap recognition
2	2341	PB1	757	Polymerase subunit; RNA elongation, endonuclease activity
		PB1-F2	87	Pro-apoptotic activity
3	2233	PA	716	Polymerase subunit; protease activity
4	1778	HA	550	Surface glycoprotein; major antigen, receptor binding and fusion activities
5	1565	NP	498	RNA binding protein; nuclear import regulation
6	1413	NA	454	Surface glycoprotein; sialidase activity, virus release
7	1027	M1	252	Matrix protein; vRNP interaction, RNA nuclear export regulation, viral budding
		M2	97	Ion channel; virus uncoating and assembly
8	890	NS1	230	Interferon antagonist protein; regulation of host gene expression
		NEP/NS2	121	Nuclear export of RNA

Table 3.1 The genomic segments of influenza A/Puerto Rico/8/1934 (H1N1) virus and their encoded proteins. The PB2, PA, HA, NP and NA proteins are each encoded on a separate RNA segment. The M2 and NEP are both expressed from spliced mRNAs, while PB1-F2 is encoded in a +1 alternate reading frame. Adapted from the paper by Bouvier *et al* (*87*) with permission from Elsevier.

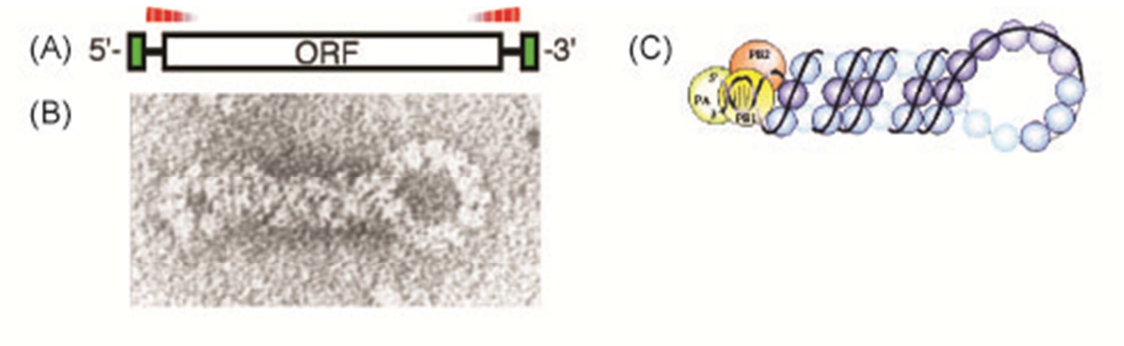

Figure 3.1 General structure of influenza A viral RNA and RNPs. (A) Schematic diagram of a typical vRNA segment. It consists of a large open reading frame (ORF) (open box) flanked by short untranslated regions (UTRs) (lines) and terminal promoter sequences (green boxes) that form the polymerase binding site and are essentially identical in all segments. The proposed packaging signals are shown as red wedges and overlap the UTRs and terminal coding regions. (B) EM image of an RNP negatively stained with uranyl acetate. Approximate magnification is ×300,000. Adapted from a paper by Jennings *et al* (*105*). (C) Cartoon model of Influenza RNP organization. Blue spheres represent NP monomers (1 monomer per 20 nucleotides) with associated vRNA molecule (black line). The single-stranded vRNA is coiled into a hairpin structure) which forms the binding site for the heterotrimeric RNA-dependent RNA polymerase (*106*).

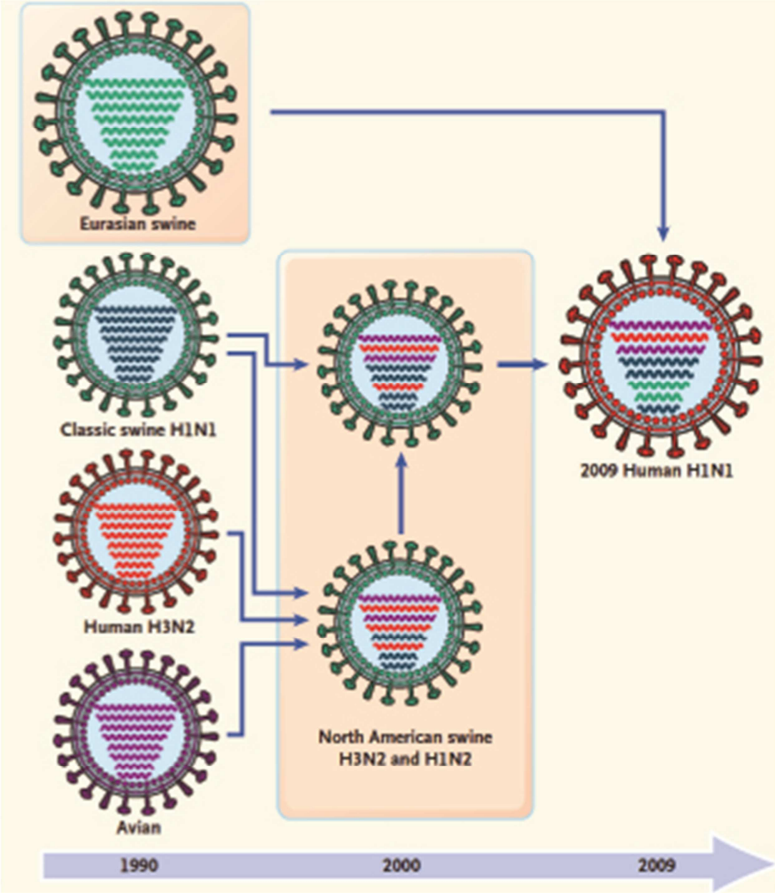

Figure 3.2 History of reassortment events in the evolution of the 2009 influenza A (H1N1) virus. The segments of the human 2009 influenza A (H1N1) virus have coexisted in swine influenza A virus strains for more than 10 years. The ancestors of neuraminidase have not been observed for almost 20 years. The mixing vessel for the current reassortment is likely to be a swine host but remains unknown. Figure reproduced with permission from Trifonov *et al* (*88*). Copyright Massachusetts Medical Society.

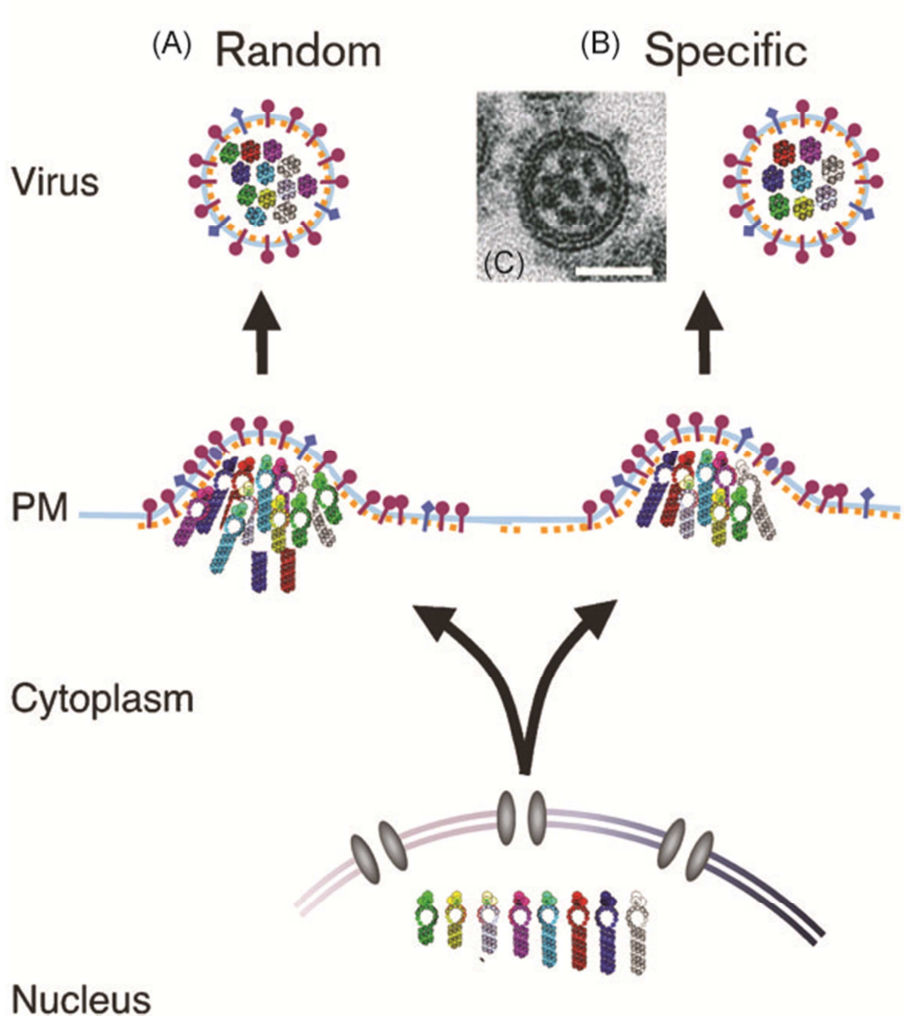

Figure 3.3 Cartoon models for two possible influenza genome packaging mechanisms. Replication of the viral genome takes place in the nucleus of the infected cell. Then the viral RNPs are transported to the plasma membrane where they get packaged into the budding virions. (A) The random packaging model proposes that an arbitrary number of RNPs are packaged so that a reasonable proportion of virions contain at least one copy of each segment. (B) The specific model proposes that each vRNA contains unique features that allow it to be distinguished from other vRNA and cellular RNA. According to this model, each virus selectively packages eight unique viral RNPs. (C) A negatively stained EM section through an H1N1 influenza virion showing the distinctive "7+1" array of RNPs. Adapted from the paper by Noda *et al*. (*96*), with permission from the Nature Publishing Group.

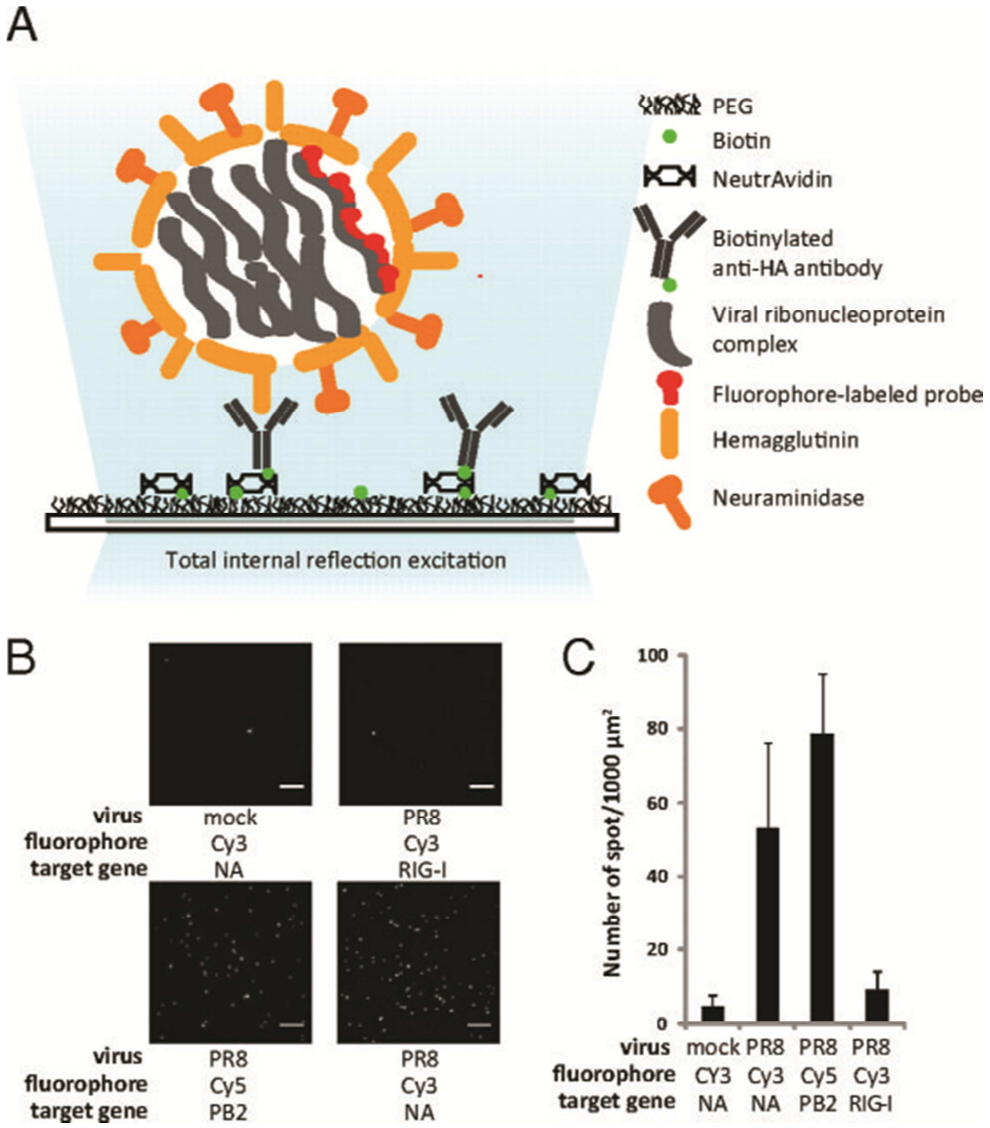

Figure 3.4 (A) Schematic for FISH analysis on single-virus particle. The virus particles were immobilized through interaction with biotinylated antibodies against the HA protein and neutravidine. After fixation and permeabilization of virus particles, viral RNA hybridized with FISH probes were visualized using TIRF microscopy. (B) Specificity of our single virus FISH assay. TIRF images of the slide surface in the presence or absence of virus particles and hybridized with FISH probes are shown. In each case, a mixture of 48 singly labeled DNA FISH probes targeting a specific vRNA was used. The error bar is 5 μm. (C) Average number of fluorescent spots in B, per 1,000 μm^2 is shown (number of fields, n > 10).

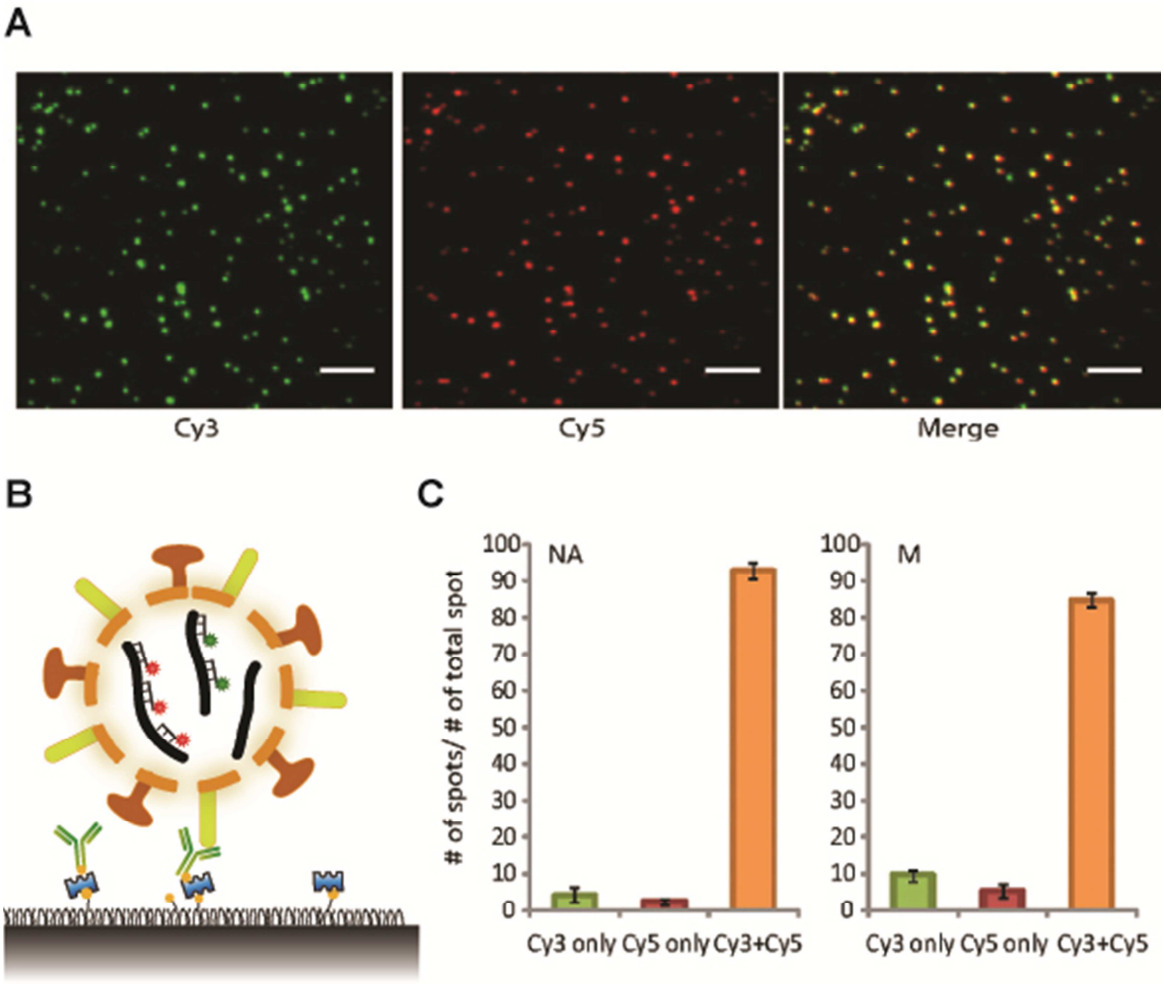

Figure 3.5 (A) Representative TIRF images for colocalization analysis is shown here. Immobilized virus particles were hybridized with 23 Cy3-labeled probes and 23 Cy5-labeled probes against different parts of the NA viral RNA segment. TIRF images of particles labeled with Cy3 probes (Left) and Cy5 probes (Center) are shown. The overlay image of the two images is shown (Right). (Scale bar: 5 μm). (B) Schematic representation of co-localization analysis. (C) Colocalization efficiency of two fluorescence probe sets targeting the same viral RNA. FISH analyses were carried out on PR8 virus immobilized imaging surfaces using 23 Cy3 labeled probes and 23 Cy5-labeled probes against the NA viral segments or 16 Cy3-labeled probes and 16 Cy5-labeled probes against the M viral segments. The percentages of the number of the Cy3-only spots, the Cy5-only spots and the dual-labeled spots over the total number of spots are plotted.

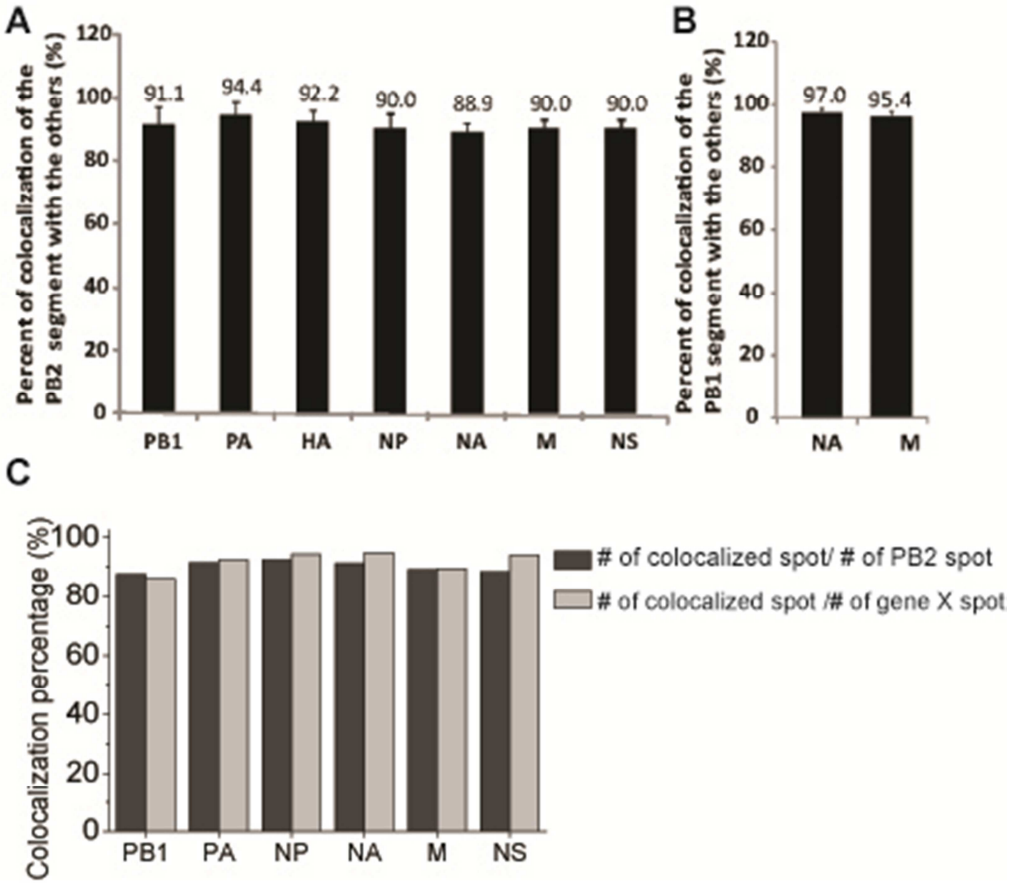

Figure 3.6 Co-packaging of different vRNA segments into individual virus particles. (A) Colocalization efficiency of the PB2 segment with the PB1, PA, HA, NP, NA, M and NS segment is shown. In each case, 48 Cy3-labeled probes targeting the PB2 segment and 15 Cy5-labeled probes targeting the paired segment were used. (B) Colocalization efficiency of the PB1 segment with the NA or M segment. FISH analysis was performed by mixing 48 Cy3-labeled probes targeting the PB1 segment with 15 Cy5-labeled probes targeting the NA or M segments. The colocalization efficiency is calculated as the number of Cy3 and Cy5 colocalized spots over the total number of Cy3 spots. The data shown are normalized by setting the colocalization efficiency of Cy3 and Cy5 probe sets against the NA segment as 100% (see Figure ?.5B). Error bars denotes SD (n > 10). (C) Co-packaging efficiency of the PB2 segment with the PB1, PA, NP, NA, M, or NS segment (# of colocalized spots/ # of PB2 spots) and the co-packaging efficiency of PB1, PA, NP, NA, M or NS with the PB2 segment (# of colocalized spots/ # of vRNA X spots) are shown.

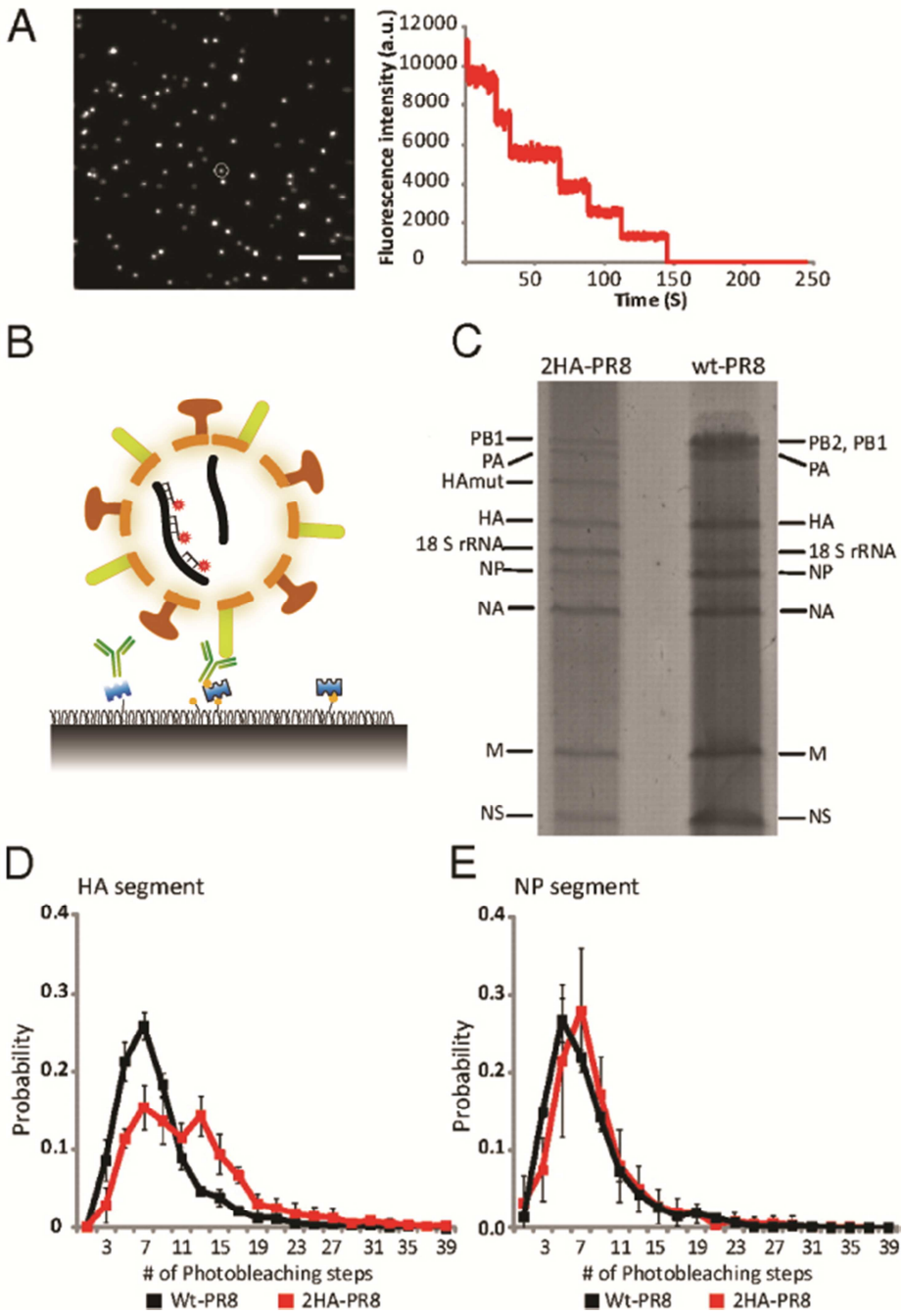

Figure 3.7 PR8 virus particle packages one copy of the HA segment. (A) Photobleaching analysis of the HA vRNA in PR8 viruses hybridized with 15 Cy5-labeled probes. Representative image of the slide surface under the TIRF illumination. The graph shows the plot of total fluorescence intensity versus time for the circled Cy5 spot (left) exhibiting seven photobleaching steps. (B) Schematic of the photobleaching assay. (C)

Bulk analysis of viral RNA packaging in the 2HA-PR8 virus. Purified viral RNAs from the 2HA-PR8 virus and PR8 virus were resolved using a 2.8% acrylamide gel, followed by silver staining. The identities of the bands were labeled based on their sizes and previous findings. (D) Histogram of the photobleaching steps analyzed for the FISH probes targeting the HA vRNA in PR8 (black data points) and 2HA-PR8 (red data points) viruses. Error bars denote SDs (n ≥ 5). (E) Histogram of the photobleaching steps analyzed for the FISH probes targeting NP vRNA in PR8 (black data points) virus and 2HA-PR8 (red data points) virus. Error bars denote SDs (n ≥ 3).

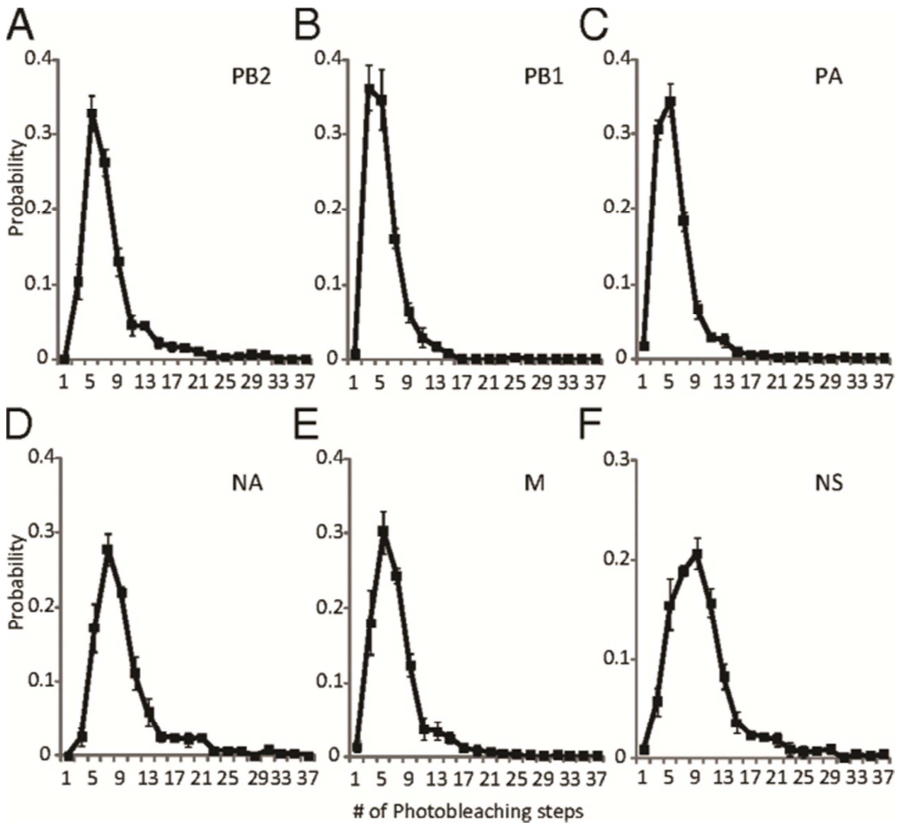

Figure 3.8 PR8 virus packages single copy of each viral RNA segment. Histograms of the photobleaching analysis are shown for the PB2 (A), PB1 (B), PA(C), NA (D), M (E), and NS (F) segments in the PR8 virus. Error bars denote SDs. n ≥ 3.

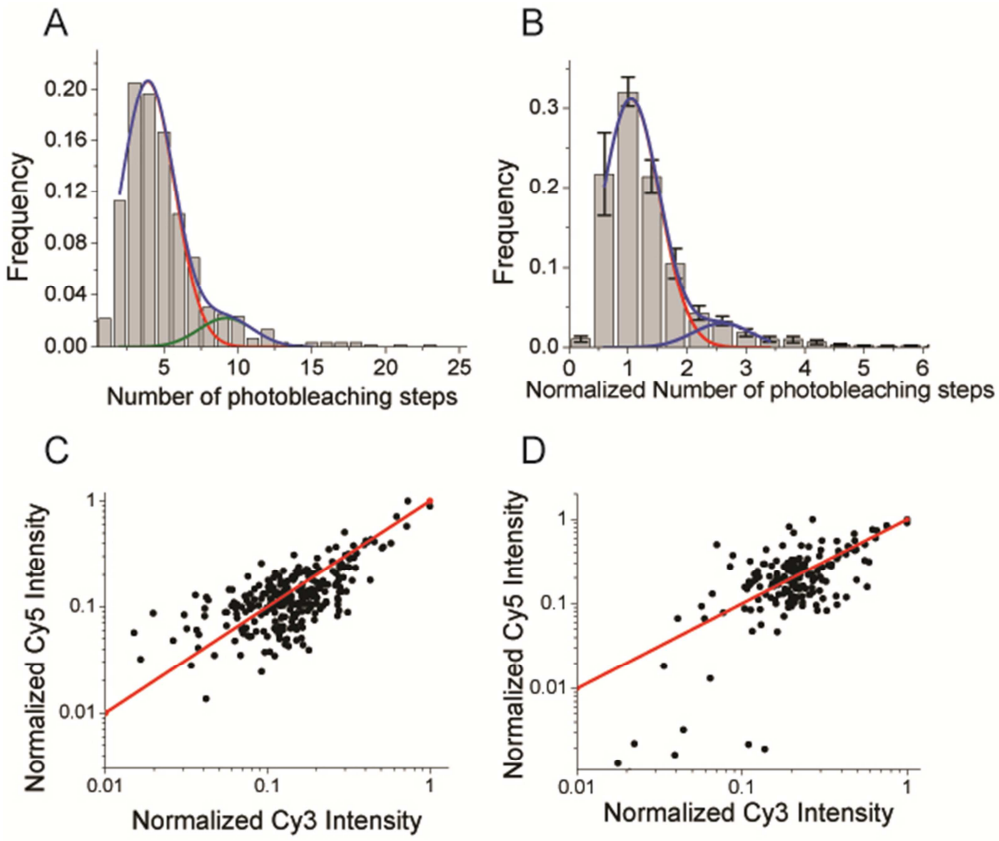

Figure 3.9 Double Gaussian fit to the histograms of the number of photobleaching steps. (A) Double Gaussian fit to the histogram of the photobleaching steps analyzed for the FISH probes hybridized with the PA viral RNA. The red curve represents Gaussian fit to the main peak and shows 90% of the population contains a single copy of the viral RNA. (B) Double Gaussian fit to the average histogram. This histogram was obtained by averaging the PB2, PB1, M, NP and PA photobleaching histograms normalized to their main peak positions. The red curve represents Gaussian fit to the main peak and shows 91% of the population contains one copy of a specific viral RNA. The blue curve is Gaussian fits to the second peaks and the black curve is the overall fit overlaid on the histograms. (C) Normalized Cy3 and Cy5 intensity of individual spots for virus particles hybridized with Cy3 and Cy5 labeled FISH probes against different regions of NA vRNA. (D) Normalized Cy3 and Cy5 intensity of individual spots for virus particles hybridized with Cy3 and Cy5 labeled FISH probes against PB2 and PA vRNA.

Chapter 4
Analyzing DNA Packaging Initiation of Bacteriophage T4 by a Real-time Single Molecule Fluorescence Assay [3]

4.1 Introduction

Bacteriophages (commonly known as phages) are viruses that infect bacteria and use the host cellular machinery to replicate. T4 bacteriophage is a lytic phage that infects the *E-coli* bacteria. Structurally, T4 is a tailed double stranded DNA (dsDNA) virus that packages its 168,903-bp long genome, coding for about 300 genes, inside an elongated icosahedral head which is 120 nm long and 86 nm wide (Figure 4.1) (*107*).

The T4 life cycle involves consecutive steps of adsorption, penetration, replication, assembly, and host lysis and escape. After the life cycle is complete, the host cell bursts open and releases 100-200 new virions. The lytic life cycle of T4 from initial infection to the lysis of the bacterial host cell takes about 25 minutes (Figure 4.2) (*108*).

A critical and fascinating step in the assembly of T4 virus involves packaging of its dsDNA genome into an empty preformed procapsid via the unique portal vertex. The molecular motor that "stuffs" the dsDNA into the phage procapsids usually consists of three proteins: the portal protein and two DNA-packaging proteins (large and small terminase subunits) (*109*). The portal protein of the tailed bacteriophages that have been examined, including T4, is a dodecametric ring arranged radially and with a central channel for DNA entry and ejection (*110, 111*) (Figure 4.3A). No ATP-binding or ATP cleavage site has been found in portal proteins (*109*). In the case of T4, a dodecameric ring of protein gp20 (61 kDa) forms the portal ring. Phage terminases are DNA packaging enzymes that contain the ATPase and endonuclease activity. These enzymes

[3] The work in this chapter is published and in preparation as following papers:

- Zhang, Z., V. I. Kottadiel, R. Vafabakhsh, L. Dai, Y. R. Chemla, T. Ha and V. B. Rao, "A Promiscuous DNA Packaging Machine from Bacteriophage T4", PLoS Biology 9 (2), e1000592 (2011).
- R. Vafabakhsh, K. R. Kondabagil, L. Dai, V. B. Rao, T. Ha "Analyzing DNA Packaging Initiation of Bacteriophage T4 by a Real-time Single Molecule Fluorescence Assay", *to be submitted*.

utilize the chemical energy of ATP hydrolysis to translocate the DNA into the capsid. Terminase proteins also have to recognize the viral DNA and bring it to the capsid for packaging.

In the T4 packaging system, gp17 (70 kDa) and gp16 (18 kDa) constitute the large and small terminase complexes (*112*). The precise role of the small terminase subunit gp16 as part of the packaging motor is still unclear. Although gp16 deletion mutants produce empty proheads (*113*), gp16 is not required *in vitro* packaging assays (*114, 115*). It has been proposed that gp16 is involved in the initial DNA recognition and further acts as the regulator of ATPase and nuclease activities of gp17 (*116*). Unlike the case of gp16, gp17 is necessary and sufficient for T4 DNA packaging *in vitro* (*117*). The N-terminal domain of gp17 (amino acids 1–360) has ATPase activity while the C-terminal domain (amino acids 361–610) has nuclease activity (*118*). The crystal structure of the large terminase gp17 has been solved to 1.8Å resolution and confirms the ATPase activity of WalkerA/B motif in the N-terminal domain of gp17 (*119*) (Figure 4.3B).

Five molecules of gp17 assemble on the gp20 portal, forming a pentameric motor with a central DNA translocation channel that is about 2 nm wide (Figure 4.3C). Structural and biochemical studies have suggested that conformational change in gp17 between relaxed and tensed states, induced by ATP hydrolysis and stabilized by electrostatic interactions, drives packaging of DNA (*119*). The T4 genome exists in the form of a branched multimeric head-to-tail polymer of viral DNA, or so called concatemer (*120*). The terminase complex initiates the packaging by first generating double stranded ends from the concatemeric DNA. After a “headful” length of DNA (171 kb or 102% of the genome) being packaged, presumably a “headful” signal is transmitted to the terminase (*121*). Consequently, the terminase complex makes the second cut and terminates the DNA packaging. It is suggested that the initial cleavage of the concatemer and the final cut after the headful packaging, occur without sequence specificity. It takes about 3-5 minutes for T4 to package its full 170 kbp genome *in vivo*. DNA within the head is nearly as compact as crystalline DNA (~500 mg/ml) (*122*) which results to an internal pressure of up to 6 MPa (equivalent to more than ten times the pressure inside a bottle of champagne).

In the past 10 years single molecule techniques have provided us with a more in-depth picture of the mechano-chemical details of packaging machines. High-resolution optical tweezers experiments on the bacteriophage Φ29 system have demonstrated that packaging happens in 10 bp increments with substeps of 2.5 bp per ATP hydrolyzed (*123*). In the case of T4 bacteriophage, using single molecule optical tweezers, (*124*) it has been shown that the packaging motor can generate forces as high as 60 pN and can package DNA at a rate of up to 2,000 bp/s (*125*). These findings put the T4 packaging system among the most powerful molecular motors reported to date. More recently, single molecule fluorescent and FRET assays have been used to study subunit stoichiometry of phage packasome and kinetic details of DNA packaging. Single molecule imaging and photobleaching analysis have demonstrated the existence of 6 pRNA subunits within the bacteriophage Φ29 packaging machinery (*126*). Fluorescence correlation spectroscopy (FCS) have been adapted to monitor the kinetics of DNA packaging in near real-time and showed that both ends of the packaged DNA are held in proximity of the portal entry and exit channel of the capsid (*127, 128*). Finally, single molecule FRET studies have suggested that DNA compresses up to 25% as it translocates through the portal channel (*129*).

Although biochemical, structural and single molecule approaches in the past 50 years have yielded great progress in understanding different aspects of viral genome packaging, due to transient and complex nature of interactions at the initiation stage, not much is known about the nature and dynamical aspects of these interactions.

Here we report a real time single molecule fluorescence assay that demonstrates consecutive packaging of dsDNA molecules by the phage T4 packaging motor, one molecule at a time. The reconstituted complexes were specifically immobilized on a passivated slide through antibody interactions. ATP and fluorescently labeled dsDNA were added and individual packaging machines, each carrying out successive DNA translocation, were imaged in real-time by total internal reflection microscopy (TIRF). Using this assay we quantified the time required for assembly of the packaging complex and initiation packaging. Also we demonstrated that DNA can directly interact with the capsid portal which can highlight the active role of portal in DNA packaging. Subtle changes in the Walker A P-loop, such as increase in chain length by a single carbon, can

lead to severe defects in packaging initiation. Our results suggest a model in which packaging initiation, analogous to the firing of an automobile engine, requires rapid firing of ATPase subunits in succession to lock the DNA in the packaging machinery in translocation mode. Our single molecule approach provides a quick and accurate way to quantify viral assembly and initiation in different conditions and to resolve the effect of different mutations or other proteins on the assembly or initiation steps of packaging.

4.2 Experimental Results

4.2.1 Single Molecule Fluorescence Assay to Study DNA Packaging

We used prism-type total internal reflection fluorescence microscopy (TIRF) (*60*) to probe individual virus particles immobilized on the chamber surface. The experimental chamber was assembled as described previously. In brief, strips of double-sided tape were sandwiched between a microscope slide and a coverslip to make 5 chambers per slide. The typical dimension of the experimental chamber was about 16 mm × 4 mm × 100 µm. For the long-term observation of single virus particles, purified T4 bacteriophage heads, pre-assembled with the packaging motor (in the presence of ATPγS and a 120 bp unlabeled priming DNA) (*125*), were immobilized on a PEG polymer coated surface (Figure 3.4). The specific immobilization was done through successive incubation and wash of neutravidine, biotinylated secondary antibody and primary polyclonal antibody against capsid proteins.

After immobilizing the virus complexes and before initiating the packaging, less than 15 dim fluorescent spots could be detected on the surface per imaging area (70 µm × 35 µm) (Figure 3.5A); possibly due to the surface impurities. However, after initiation of packaging by introducing ATP and fluorescently labeled DNA, fluorescent spots appeared on the surface and grow in intensity as the heads packaged more DNA fragments (Figure 3.5A). In a typical experiment, the number of fluorescent spots per imaging area after injection of labeled DNA and ATP was 10-20 folds more than the controls (Figure 3.5B). Furthermore, DNase I treatment of the chamber after packaging, only slightly reduced the number spots on the surface. Therefore, the detected fluorescent spots on the surface came from the labeled DNA packaged inside the virus capsid (Figure

3.5C). Based on these data we believe that the reproducible large difference in the number of fluorescent spots demonstrates the high sensitivity of our assay.

4.2.2 Bacteriophage T4 Packaging Machine is Highly Promiscuous

The formation of viral particles is thought to be a sequential and irreversible process in which the assembly of one component generates a new site or conformational state that allows for the assembly of the next component, and so on (*130*). During the packaging step, once a full length genome has been packaged, a conformational change in the portal causes the terminase to detach. It has been assumed that the motor cannot reattach, after a headfull packaging. However, recent complementary results from bulk packaging assays, single molecule optical tweezers and single molecule total internal reflection fluorescence assays (smTIRF) have challenged this picture (*131*) (Figure 4.6). Initially, bulk packaging assays showed that both partial and full mature heads efficiently packaged short DNA fragments upon incubation with DNA, ATP and gp17. Single molecule optical tweezers experiments confirmed that packaging motor can assemble onto partial and full mature heads and package the DNA. The packaging rate of these assembled packaging machines was shown to be similar to that of machines assembled on packaging-naïve proheads (~800–1,100 bp/s). Finally, single molecule fluorescence measurements showed that the mature phage heads undergo repeated packaging initiations, packaging multiple DNA molecules into the same head. Short 39-bp Cy5-end-labeled and 83-bp Cy3-end-labeled DNAs were packaged into proheads, partial heads, and full heads in bulk. The heads were then immobilized on the chamber surface as described earlier and imaged in a TIRF microscope. Consistent with results from the bulk assays, the average number of fluorescent spots corresponding to partial heads that had packaged the labeled DNA was about 5-fold greater than for the proheads, and about 10-fold greater than for the full heads (Figure 4.7A, 4.7B). Control experiments, which omitted gp17, had 0–2 bright spots, suggesting that nonspecific fluorescence of any surface-bound material is negligible. Photobleaching analysis showed that the partial heads contained on average five to six DNA molecules per head, whereas the proheads and full heads contained four DNA molecules per head (Figure 4.7C). Thus, the mature phage heads, like the procapsids, can undergo multiple packaging initiations. The single

molecule data also suggested that the large difference in packaging efficiency between the partial head, the full head, and the prohead arose from the inability to initiate packaging in a large fraction of full heads and proheads. For heads that are capable of initiating DNA packaging, the number of molecules packaged is only slightly different between the three species.

These data suggest that the gp17 motor binds to both immature and mature capsids and continues to insert multiple DNA molecules into the capsid. Such promiscuous activity for the motor protein may help explain the fundamental relationship between genome size and capsid volume that has been observed in double-stranded DNA bacteriophages and other viruses.

4.2.3 Assembly of Bacteriophage T4 Packaging Machinery

As mentioned in 4.2.1, the packaging complexes were preassembled by mixing and incubating the heads with gp17, ATPγS and a short unlabeled priming DNA prior to immobilizing them on the surface. In order to inspect the functional details of gp17 assembly onto the capsid, we preassembled separate packaging complexes using the same concentration of heads and gp17 but omitting ATPγS or priming DNA. After immobilization of packaging complexes onto separate channels as described in the section 4.2.1, we initiated the packaging by injecting ATP and 35bp Cy5-labeled dsDNA. We then quantified the fraction of active complexes by determining the average number of fluorescent spots per imaging area (70 μm × 35 μm) from at least 30 different imaging areas for each sample. Figure 4.8 shows the fraction of active complexes showing multiple initiation events and after DNase I treatment. As it is seen, in the absence of ATPγS in the assembly step, only about 3% of complexes showed successful initiation compared to the complexes formed in the presence of ATPγS. ATPγS is the nonhydrolyzable ATP analog and should mimic the effect of ATP binding (and not hydrolysis). The fact that active packaging complexes form very inefficiently in the absence of ATPγS highlights the role of ATP binding in the assembly step and its direct role in stabilizing the interaction between gp17 and the portal protein (gp20). In most multimeric protein complexes and filaments, ATP or other cofactors are positioned at the interface between neighboring subunits mediating assembly of the structure (*132*).

However, in gp17 the ATP binding site is on the top surface of the N terminal subdomain II where gp17 is interacting with the portal protein (*119*). The higher stability of complexes in the presence of bound ATP hint to the role of ATP binding in "gluing" individual motor subunits to the viral capsid. Also the weak interaction between neighboring gp17 subunits enables the motor to achieve high packaging rates. On the other hand DNA in the channel can provide electrostatic support and therefore stabilizes the motor complex. This is why in the absence of the priming DNA the efficiency of active complex formation drops to 15% (Figure 4.8).

4.2.4 DNA Can Directly Bind to the Capsid Portal

The experiments in the previous section were done after immobilizing the pre-assembled complexes on the surface. We decided to test whether assembly step also could be done in the chamber itself. To test this, we directly immobilized partial heads on the surface without prior assembly with DNA or gp17. Surprisingly when we flowed in 2 nM Cy5-labeled DNA we observed many fluorescent spots on the surface (Figure 4.9A). When we increased the concentration of the immobilized heads on the surface, we observed more fluorescent spots on the surface, indicating that the DNA was binding to the capsid. Also to confirm that DNA is binding to the heads and not to the slide surface, we repeated the experiment, using the same DNA concentration, in a chamber without the virus heads. In this case, we observed more than 30 times less number of spots (19 spots vs. 634 spots), firmly indicating that the observed spots in the previous experiment were not due to non-specific binding of DNA onto the surface (Figure 4.9B). Most of the detected spots (>80%) showed single step photobleaching, suggesting that only one DNA molecule was bound to the capsid. The capsid surface area is very large and if the DNA was non-specifically "sticking" on the capsid surface, we would expect to see a wide range of number of photobleaching steps. We think the spots that show more than one photobleaching step are from virus head aggregates or multiple virus particles being immobilized within diffraction limited spot. To further attest that the DNA was bound to the head portal and not the capsid surface, we repeated the experiment in a chamber containing the preassembled packaging complex. In this case we expect that the portal be occupied by the gp17 protein. After injecting DNA into this chamber we observed an

order of magnitude fewer number of spots than when we did the experiment with the capsid only (75 spots vs. 634 spots) (Figure 4.9C, 4.9D).

We observed two distinct behaviors for the DNA molecules that interacted with the virus head. While most of the molecules showed fast binding and unbinding of DNA, a small fraction of them showed tight binding to the capsid (Figure 4.10A). For the fluctuating traces, increasing the DNA concentration increased the DNA binding rate while it did not have a significant effect on the unbinding rate (Figure 4.10B). In order to further prove that the DNA signal that we see is actually coming from a DNA binding to the capsid portal, we did the following experiment. After immobilizing the capsids on the surface and introducing Cy5-labeled DNA, we flowed in gp17 and ATP without additional DNA. After a short incubation, we flowed in excess amount of DNase I and incubated for 20 minutes. If the DNA was originally bound to the portal, packaging motor could assemble and package this DNA inside the capsid. The packaged DNA would be protected from the DNase I digestion and hence we should see the signal. On the other hand, if the DNA was stuck to the capsid surface or the slide surface, we do not expect to see signal after the DNase I treatment. As a control, we repeated the same experiment but without the packaging step (without the gp17+ATP step). Figure 4.10C shows the result of counting the number of spots on the surface in the two cases. As it is seen, when we had the packaging step, we observed 10 fold more number of spots on the surface. Furthermore, using a two-color approach, we observed that about 30% of capsids that packaged the initial Cy5 DNA, could later package multiple Cy3 DNA by providing gp17, Cy3 DNA and ATP after the DNase I step. All these data prove that DNA can bind to the portal directly and then upon access to gp17 and ATP, the same DNA can be packaged inside the capsid. I will discuss some dynamical aspects of this interaction in the next section. It is worth mentioning that both of the interactions shown in the Figure 4.10A can package DNA upon availability of ATP and gp17. In order to show this, virus heads were immobilized on the slide surface and 2nM DNA was flushed into the chamber. After 140 seconds, 1mM ATP, 2nM DNA and 1μM gp17 were introduced. We observed active packaging for complexes that showed both stable DNA binding or transient DNA binding prior to addition of motor and ATP (Figure 4.10D and 4.10E).

4.2.5 Quantifying the DNA Packaging Initiation

The pre-assembled packaging complexes that are immobilized on the surface can be image in real-time while they are packaging DNA molecules, one after the another (Figure 4.4). Analyzing individual packaging heads shows stepwise increase in the fluorescent intensity after addition of ATP and DNA (Figure 4.11A). Each step in these traces corresponds to encapsidation of a new Cy5-labeled DNA fragment inside the diffraction limited volume of a virus head. Depending on the DNA and ATP concentrations, it takes about 10 to 100 seconds for the motor to package a DNA. On the other hand, it has been shown that the average DNA translocation rate for the T4 packaging motor is more than 700 bp/s (*125*). Therefore, the 45 bp DNA used in our assay should get packaged in about 50 ms. Therefore, the time delay we observe between successive packaging events is the "initiation time" required for the motor to bind a new DNA, load with ATP and re-initiate the packaging.

In order to get insights into motor assembly dynamics, we can flow in the motor along with the labeled DNA and ATP into the chamber which has the heads immobilized on the surface. In this case, motor needs to assemble before packaging initiation happens. By comparing the first packaging time in this configuration and the time required for the preassembled complex to initiate packaging, we can measure the motor assembly timescales (Figure 4.11B). For example in the case shown in Figure 4.11B we measured an assembly time of at least 44 seconds even in the presence of saturating ATP (1 mM) and gp17 (1 μM) concentrations. Surprisingly, this is a relatively large time compared with the time required for the packaging of the whole T4 genome which is 3 to 5 minutes.

4.2.6 Initiation of DNA Packaging Happens in Bursts

Having established a robust assay to probe packaging in the page T4 system and to further quantify the packaging reinitation we followed individual virus heads as they packaged multiple short Cy5-labeled dsDNA molecules over long periods of time (Figure 4.12A). Each head showed stepwise increase in the fluorescence intensity over time with each step corresponding to the packaging of a new Cy5-labeled DNA. The discreet intensity drops in the intensity traces correspond to photobleaching of individual Cy5 fluorophores. Figure 4.12B shows the initiation times for pre-assembled capsids at 2 nM

DNA and 1mM ATP, in the absence of free gp17. As it is highlighted in the graph, the distribution of initiation times exhibits a double exponential behavior. This behavior is very reproducible and was observed at a wide range of DNA and ATP concentrations. Figure 4.12C show the short (τ_s) and long (τ_l) initiation time constants measured as a function of the DNA concentration and in the presence of saturating ATP concentration (1 mM). As it is seen, the short initiation time depends on the DNA concentration and therefore diffusion limited. However, the long initiation time does not have DNA concentration dependence. On the other hand, both of these timescales exhibit ATP dependence at constant DNA concentration (2 nM DNA) (Figure 4.12D) We assign the long initiation time constant (τ_l) to be pausing time between activity periods during which the packaging motor needs to perhaps recoordinate to initiate further packaging .

Based on this data we propose the following model. The packaging complex packages DNA with the rate k_1. This rate depends on DNA and ATP concentration. However the motor can transit into an inactive state with a rate k_2 which depends on ATP concentration. Finally the motor recovers with the rate k_3 and resumes packaging initiation (Figure 4.12E)

4.2.7 Quantifying the DNA Packaging Initiation Efficiency

In order to quantify initiation efficiency of the heads we terminated packaging at different times by quickly substituting the packaging buffer with a buffer lacking DNA and ATP. We then photobleached individual virus heads and the number of photobleaching steps were counted. Figure 4.13A shows a representative trace exhibiting 11 photobleaching steps. Figure 4.13B shows the normalized histogram of the number of packaged DNA molecules in each virus head at different times after the initial flow of DNA and ATP, as quantified by the photobleaching analysis. The observed heterogeneity in the number of packaged DNA molecules per head is consistent with the bursting model observed for mRNA transcription in bacteria (*133*). Here, both the pause time and the packaging time are exponentially distributed and the observed broad distributions are the result of the combined effect of these two sources of randomness.

The average number of DNA molecules per capsid can also be measured after long incubation with ATP and labeled DNA. As Figure 4.12C shows, the average number of DNA molecules packaged per capsid after 30 min incubation does not vary much over an order of magnitude of DNA concentration and in the presence of saturating ATP concentration. On the other hand, this number drops dramatically by decreasing ATP concentration and in the presence of 2 nM DNA (Figure 4.13D). These two different behaviors highlight the critical effect of ATP hydrolysis in the initiation efficiency.

4.2.8 Heat and Cold Sensitive gp17 Mutants are Defective in Initiation

In order to further investigate the effect of ATP hydrolysis on the packaging initiation we studied motors with mutation in the ATPase domain. Walker A P-loop of the AAA ATPase superfamily, one of the ancient and highly conserved motifs, has the consensus sequence G/PXXXGKS/T. In phage packaging ATPases, the loop is shortened by one amino acid and the first amino acid is often a serine instead of a glycine (Figure 4.14A). There is also another threonine after the conserved GKS/T sequence. Relatively conservative mutations at these residues produce temperature sensitive mutants (cold sensitive mutant; *cs* S161T and heat sensitive mutant; *hs* T168Q) (*134*). The *cs*161T and *hs*T168Q mutant proteins were purified and bulk functional analyses showed that these mutants exhibited about 10-fold lower DNA-packaging activity and ATP hydrolysis (Figure 4.14B and 4.14C).

Single molecule analyses with complexes formed with S161T mutant motor showed about 10 to 20 fold fewer heads capable of initiating and packaging multiple DNA molecules than the wild type gp17. Figure 4.15A shows typical time traces for packaging complexes with the S161T mutant motor. As it is seen, the packaging time is much longer for these heads. The average number of DNA molecules packaged in the heads with mutant motor is less than 2 at saturating ATP concentration and 4 nM DNA and after 30 min incubation. This is only comparable to wildtype motor in limiting ATP concentrations (Figure 4.13D). The number of fluorescent spots per imaging area (70 µm×35 µm) when using pre-assembled complexes is 20±6 while without prior motor preassembly we observe an order of magnitude more fluorescent spots on the surface (219±23) (Figure 4.15B). If the mutant motor could not assemble onto the capsid, then

the capsid portal would be accessible to DNA and DNA molecules could bind directly to the capsid as discussed in section 4.2.4. In that case we should have seen many more spots as in Figures 4.9A and 4.9D. Therefore, the S161T mutant motors can assemble onto the capsid efficiently. To investigate the reason for low packaging efficiency of the mutant motors, we did the experiment without pre-assembly by immobilizing the heads on the surface and then flowing in Cy5 DNA together with ATP and S161T gp17. After 30 min incubation we still observed more than 200 spots per imaging area. However, more than 75% of these spots exhibit only single photobleaching step. Since this measurement is done after DNase I treatment, the DNA molecules are either inside the capsid it stuck in the motor channel. This means that although the mutant motor can assemble, it is not capable of initiation and packaging the DNA. Quantification of packaging time showed that both the short and long packaging initiation times for these mutants were longer than the times measured with wildtype motors (Figure 4.15C). The short packaging time (τ_s) for this mutant is more than 71 s which is about 5 fold slower than the wildtype motor at the same DNA and ATP concentrations.

4.3 Discussion

DNA packaging into the viral capsid is a complex process consisting of initiation, elongation and termination. It also involves orchestrated coordination and sequential action of multiple proteins. Recent single molecule experiments using optical tweezers have shed light on the mechanochemical aspects of the elongation phase which involves translocation of viral DNA into the viral capsid by multimeric ATPases (*124*). However due to the transient and dynamic nature of the interactions, it has been challenging and cumbersome to study and quantify packaging initiation and viral assembly using bulk or single molecule techniques. In our assay we used fluorescence from encapsidated DNA molecules to study DNA packaging in real-time and at single virus particle level. Moreover, we used the assay to analyze the packaging motor assembly steps and to quantify the effect of mutations on the packaging initiation.

Our results demonstrated that the phage T4 carries can carry successive packaging initiations with inactive phases separating packaging bursts. Also, our assay uncovers the heterogeneity in the initiation among individual virus particles, a property that is not

evident in bulk assays. Optical tweezers experiments studying continuous packaging of a single, long dsDNA in the T4 system have shown a large heterogeneity in the T4 packaging rates. Bacteriophage T4 can package DNA with the average rate of 690 bp/s and standard deviation of 340 bp/s, reaching up to 2000 bp/s (static dispersion) (*125*). Additionally, individual heads show variable packaging rates as they package the DNA (dynamic dispersion). In fact, it was shown that for more than 50% of T4 complexes measured in an optical tweezers experiment, the standard deviation of individual head's packaging velocity was more than two fold larger than the expected variation from a Poissonian stepper (*125*). The reason for this high variability is not known. However our observed initiation pausing and the reported static and dynamic dispersion in the T4 packaging rates might have a similar underlying molecular mechanism. Based on the crystal structure of gp17 it is proposed that the motor has several different conformational states (*119*) and the observed variability might be due to the protein getting stuck in certain conformational state or loss of coordination among the subunits.

Our data combined with our previous results show that the T4 packaging machine is promiscuous, translocating into any head intermediate, any DNA, any number of times. The multiple translocations may represent an evolutionary relic of the headful packaging mechanism. An ancient packaging machine probably encapsidated many pieces of random DNA, which upon recombination and selection led to the evolution of viral genomes that fit the capsid shell.

In an infected bacterium, initiation involves formation of a holo-terminase complex between the small terminase protein gp16, large terminase protein gp17 and DNA, followed by docking of the terminase-DNA complex on the portal and insertion of the end into the translocation channel. The current bulk and single molecule methods are not capable of measuring the timescales involve in assembly and initiation. However, using our approach it is now straightforward to directly quantify the effect of other proteins on the overall assembly and initiation efficiency of the packaging motor. In our *in vitro* system, the wild-type packaging machine takes about a minute to assemble and initiate packaging even though the DNA is already cut and DNA, ATP and gp17 are present in excess.

Binding of DNA to the motor must be accompanied by rapid firing of motor subunits in succession resulting in translocation of several base pairs of DNA into the portal channel. Otherwise, the bound DNA would dissociate as the end is free to bind and release rather freely. Thus packaging initiation, analogous to firing of an automobile engine, requires rapid ATP firing of motor subunits so the DNA has translocated deep enough inside the capsid that it cannot diffuses out. The very inefficient packaging initiation efficiency for the *cs*S161T and *hs*T168Q mutants is consistent with the above hypothesis. Our single molecule analysis showed that although these motors can assemble onto the capsid, their initiation time is several folds longer than the wildtype motors. Bulk biochemical assays also show that these mutants, with a slight change in the ATP binding pocket, are defective and exhibiting 10-20 fold reduced rate of ATP hydrolysis in bulk.

4.4 Experimental Procedures

Bulk in vitro DNA Packaging

In vitro DNA packaging assays were performed by the procedure described earlier (*114*). The reaction mixture contained purified proheads, partial heads, or full heads ($0.5–1\times10^{10}$ particles), purified full-length gp17 (1.5 mM), and DNA (300 ng of 50- to 766-bp ladder DNA, 100 ng of Cy3 83-bp DNA, 50 ng of Cy5 39-bp DNA, or 600 ng of 48.5-kb phage λ DNA). The λ DNA was packaged using a buffer containing 30 mM Tris-HCl (pH 7.5), 100 mM NaCl, 3 mM MgCl2, and 1 mM ATP. The Cy3 and Cy5 DNAs were packaged using the 5% PEG buffer as described earlier (*114*). The packaging reactions were terminated by the addition of DNase I, and the encapsidated DNase I–resistant DNA was released by treatment with proteinase-K and analyzed by agarose gel electrophoresis. Each experiment included one to several negative controls that lacked one of the essential packaging components: head, gp17, ATP, or DNA. Packaging efficiency is defined as the number of DNA molecules packaged per the number of head particles used in the packaging reaction.

Single Molecule Fluorescent Analysis of DNA Packaging

Single molecule fluorescence experiments to quantify packaging efficiency of different heads were performed on a wide field prism type total internal reflection microscope with a 532 laser (Coherent) for Cy3 excitation or a 630 laser (Melles Griot) for Cy5 excitation. Immobilized capsids were imaged by a CCD camera (iXon DV 887-BI; Andor Technology) at 100-ms time resolution. A homemade C++ program was used to record and analyze the images (*60*). To minimize nonspecific surface binding, clean quartz slides and glass cover slips were surface-passivated with PEG and 3% biotinylated PEG (Laysan Bio) (Appendix A). After assembling the channel, NeutrAvidin (Thermo Scientific) was added (0.2 mg/ml), followed by incubation with biotinylated protein-G (Rockland Immunochemicals) (25 nM) for 30 min at room temperature. Subsequently, polyclonal anti-T4 antibody (15 nM) was added and incubated for 1 h. The packaged heads with 83-bp Cy3 and 39- bp Cy5 DNAs were applied to separate channels and incubated for 20 min. Prior to immobilization, the packaging reaction mixtures were treated with DNase I (10 mg/ml) at room temperature for about 20 h to digest any unpackaged or nonspecifically bound Cy3 and Cy5 DNAs. The unbound packaged heads were washed off, and immobilized capsids were imaged in iX PEG buffer (50 mM Tris-Cl buffer (pH 8), 5% PEG, 5 mM MgCl2, 1 mM spermidine, 1 mM putrescene, 60 mM NaCl), and the oxygen scavenger system (0.8% dextrose, 0.1 mg/ml glucose oxidase, 0.02 mg/ml catalase, and 3 mM Trolox).

To monitor the real-time packaging, assembled packaging complexes were first prepared by mixing 0.5×10^{10} partial heads (per slide channel), 1 µM gp17, 1 mM ATPγS, and 200 nM priming DNA (120 bp dsDNA) in 1X PEG buffer and incubating the mixture for 15 minutes at room temperature (5µL per slide chamber). The pre-assembled complex (5µL) was then mixed with 15µL of T50-BSA buffer (10 mM Tris (pH 7.5), 50 mM NaCl and 1 mg/mL BSA) and flushed into the antibody coated slide as described above. The complex was incubated for 30 min and free complexes were washed away with T50-BSA buffer containing 1 mM ATPγS. The heads were imaged while packaging the DNA in the imaging buffer; consisted of 3mM trolox, 1X PEG buffer (above text) and the oxygen scavenger system (0.8% dextrose, 0.1 mg/ml glucose oxidase, 0.02 mg/ml catalase) plus DNA and ATP at the relevant concentrations.

Data Analysis

Single molecule movies were acquired and analyzed to build the single molecule traces, using home-built software. All the subsequent analyses were done in custom MATLAB codes. To calculate the packaging times, individual traces were examined by eye, one by one. The packaging times (corresponding to the step-wise fluorescence increase which lasted for more than 2 seconds) were manually selected and used for subsequent analysis.

To compare packaging efficiencies in different conditions, the same number of heads was immobilized in different channels with the same incubation time. The number of complexes used for immobilization in each case was many orders of magnitude less than available binding sites on the slide. Packaging efficiency was quantified by determining the average number of fluorescent spots per area (70 µm × 35 µm) from at least 30 different imaging areas for each sample.

4.5 Figures

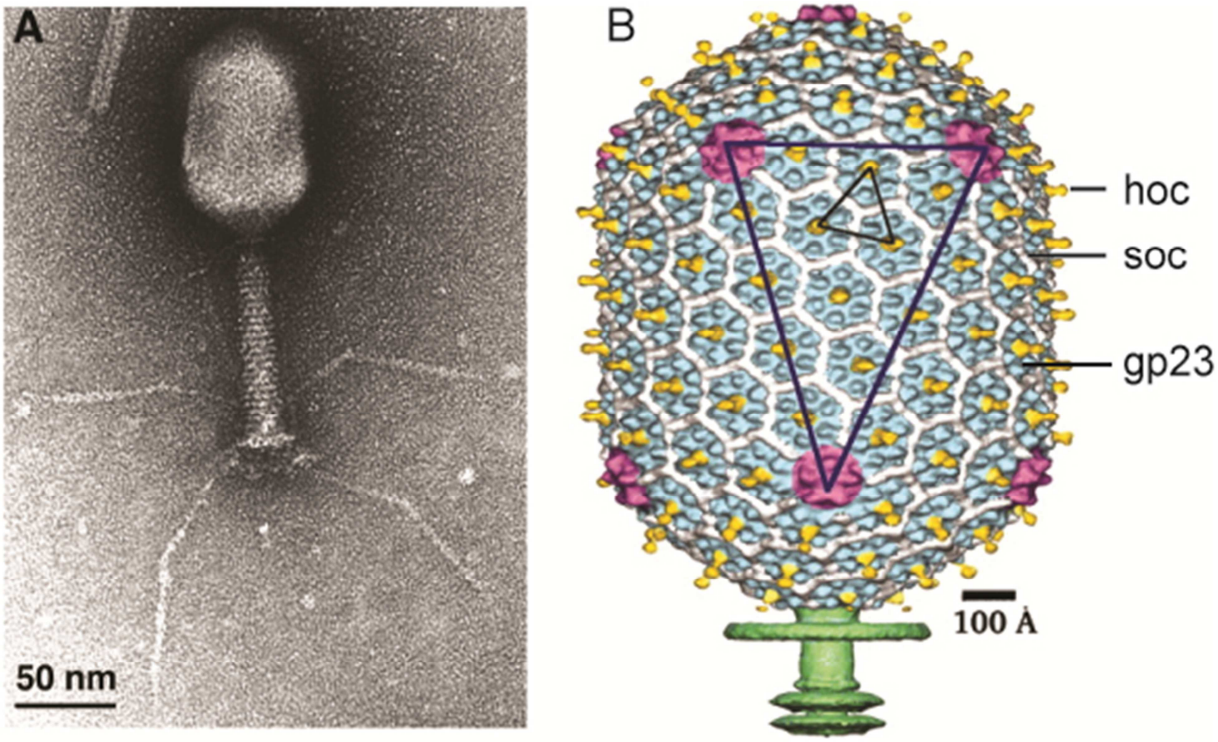

Figure 4.1 (A) Electron microscopy of bacteriophage T4 (*107*) showing the phage head that encapsidates its 171 kb genome, neck, tail and tail fibers. (B) Structure of the bacteriophage T4 head. gp23* is shown in blue, gp24* is in magenta, soc is in white, hoc is in yellow, and the tail is in green (*135*).

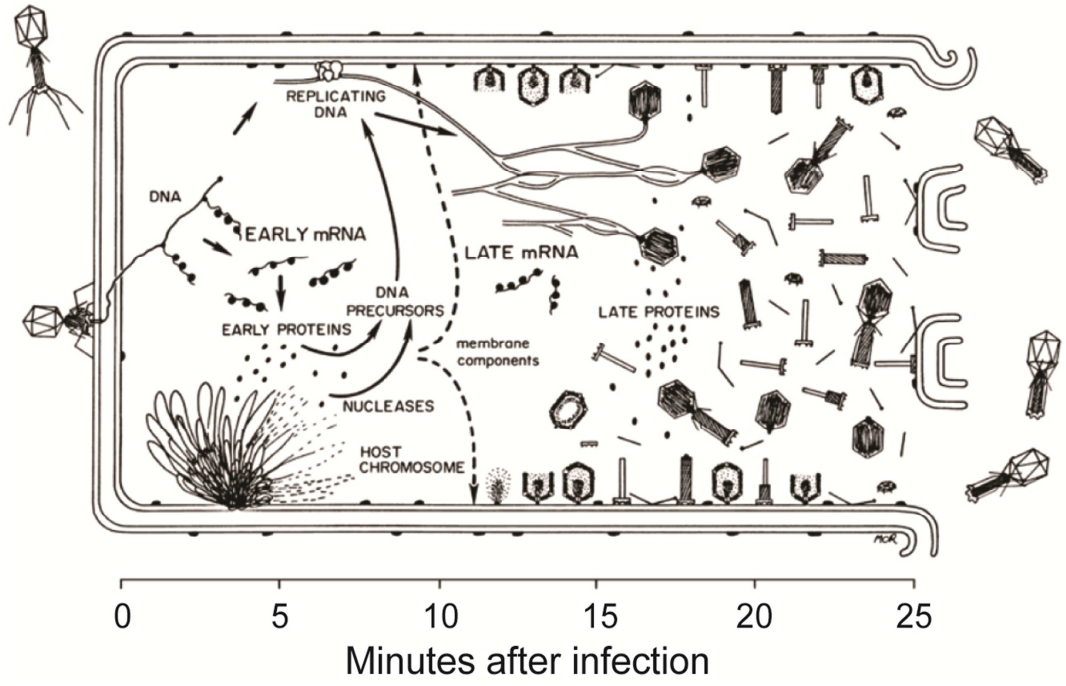

Figure 4.2 Life cycle of Bacteriophage T4 (*108*). Bacteriophage T4 completes its life cycle from infection to release within a ~25 minute time frame. After attaching and injecting its DNA into the *E. coli*, T4 hijacks the host cellular machinery to replicate its own DNA. Simultaneously it makes copies of its structural proteins, which then get assembled together and make mature phages. About 150 new T4 phages are released by lysing the infected bacteria.

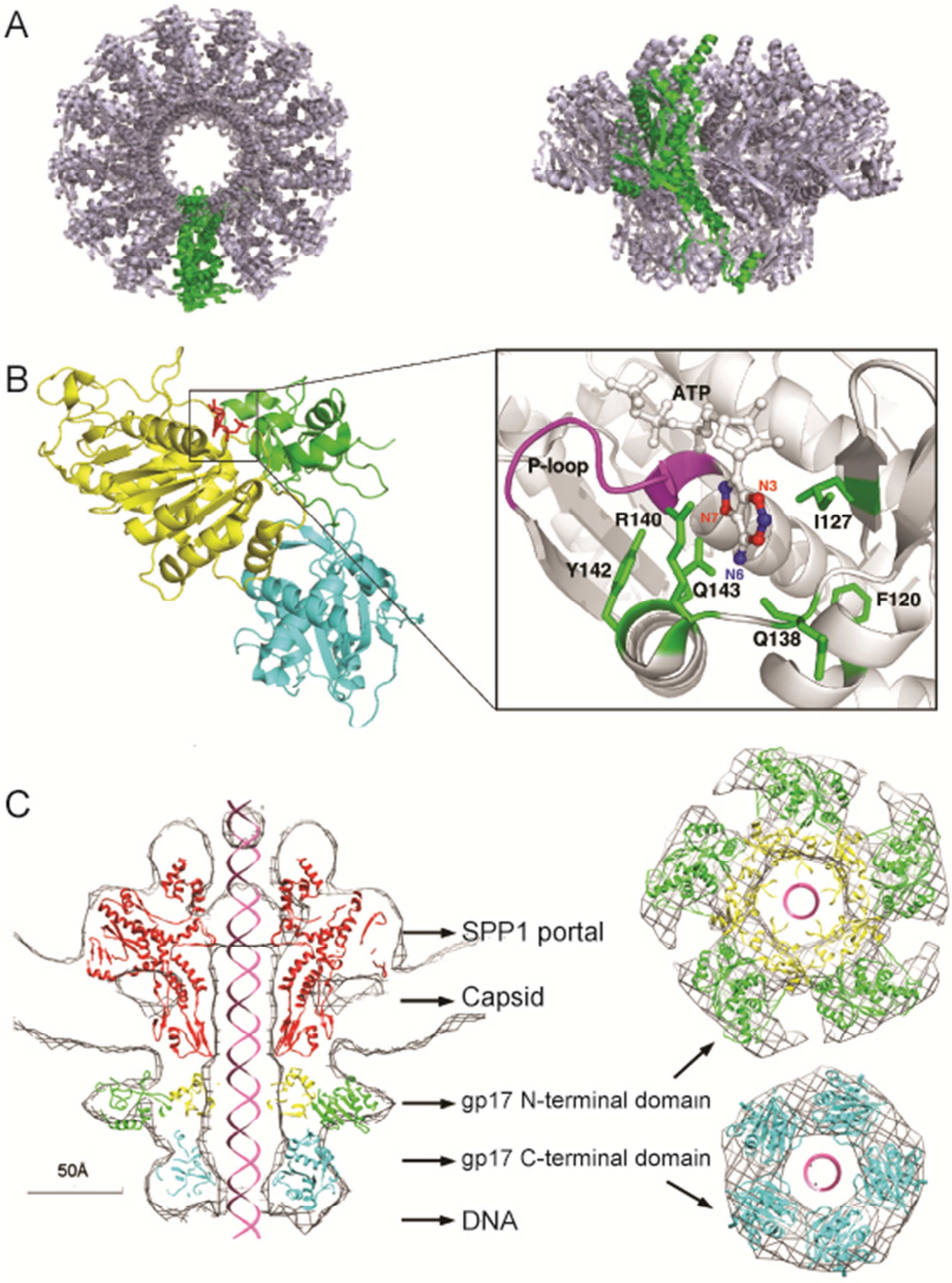

Figure 4.3 (A) Top and front views of Φ29 12-fold portal structure. A single subunit of the portal ring is colored in magenta (*120*). (C) The gp17 crystal structure. N-Subdomain I (amino acids 59 – 313; yellow), N-subdomain II (amino acids 1 – 58 and 314 – 360; green), and C-domain (amino acids 361 – 564; cyan). The ATP binding pocket is

enlarged to show the Walker A P-loop (magenta), bound ATP and adenine binding residues (green) (*120*). (B) Structure of the dodecameric SPP1 portal (red) (based on the crystal structure of the (13-mer) and crystal structures of the N-terminal subdomain I (green), subdomain II (yellow), and C-terminal domain (cyan) of T4 gp17, fitted into the cryo-EM density (*119*).

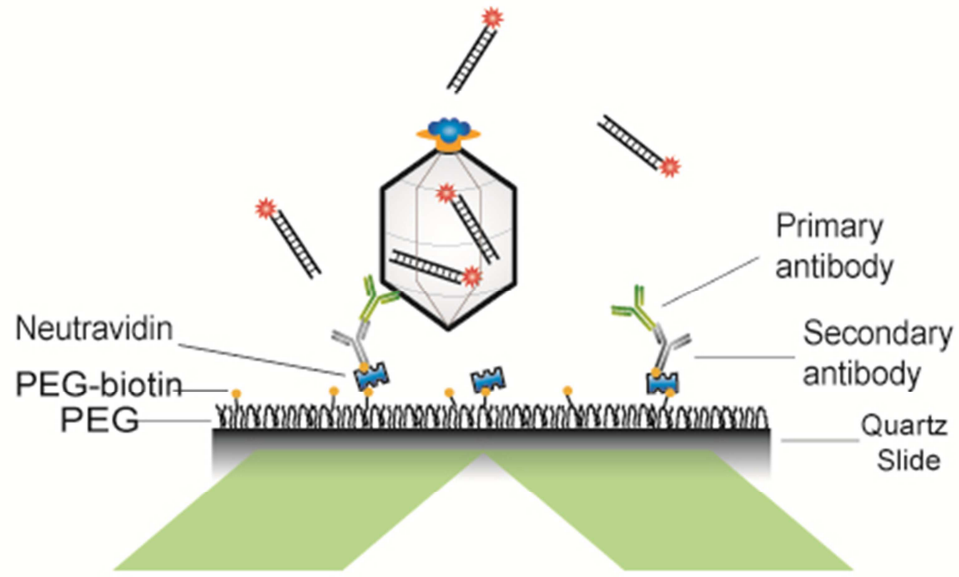

Figure 4.4 Schematic design of the single molecule fluorescence experiments. Packaging complexes were immobilized on a polymer coated surface through specific interaction with neutravidine, biotinylated secondary antibody and primary antibody against T4 capsid. The capsids were imaged in a TIRF microscope. Free fluorescently labeled DNA molecules which diffuse freely only contribute to the background. However if they get packaged inside a virus head, a well-defined step-wise increase in the fluorescent intensity of that spot is observed.

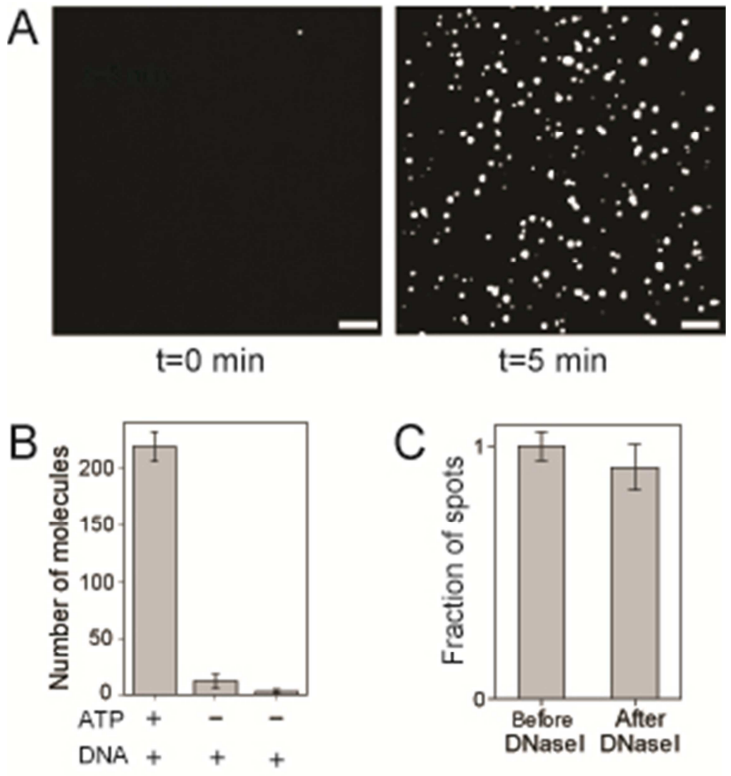

Figure 4.5 Specificity of our packaging assay. (A) (left) Pre-assembled packaging complexes immobilized on the slide surface before introducing fluorescently labeled DNA and ATP. Only very few dim spots are visible, possibly due to the surface impurities. (right) 5 minutes after introducing 4 nM Cy5-lebeled DNA and 1mM ATP, bright fluorescent spots are visible on the surface. Each spot corresponds to multiple DNA molecules being packaged inside individual virus heads. Scale bar is 5 µm. (B) Large difference in the number of molecules per imaging area (70µm × 35µm) denoted the specificity of our assay. Error bars are standard deviation of the number of spots from at least 30 different imaging areas. (C) DNase I treatment of the surface only slightly reduces the number of observed spots on the surface, proving that the fluorescent spots are indeed coming from encapsidated DNA molecules.

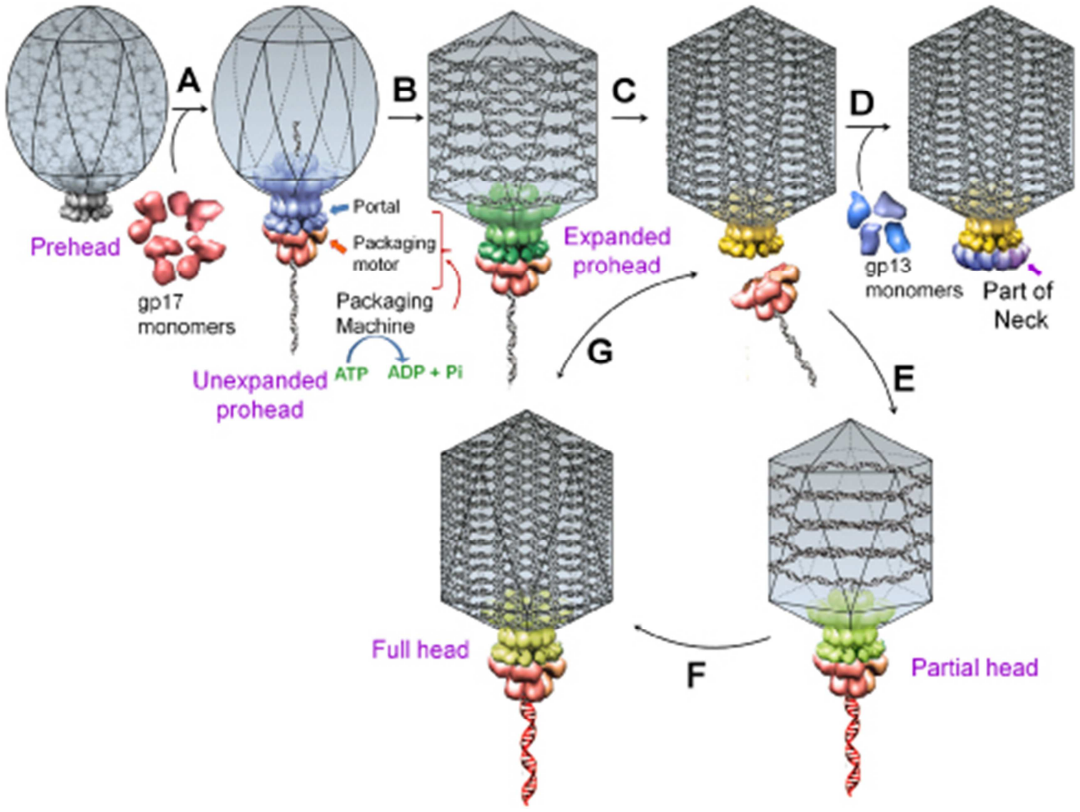

Figure 4.6 A schematic of DNA packaging by sequential assembly and promiscuous assembly. The major capsid protein assembles around a scaffolding core into a prehead. The core is removed by proteolysis to produce an empty unexpanded prohead (A). The unexpanded prohead normally has a round shape, but in phage T4 it has angular geometry. The packaging motor–DNA complex docks on portal and initiates packaging. The prohead expands after about 10%–25% of the DNA is packaged (B). After headful packaging, the motor cuts the concatemeric DNA and dissociates from the DNA-full head (C). The neck proteins (gp13, gp14, and gp15) assemble on portal to seal off the DNA-full head and provide a platform for tail assembly (D). The various colors of portal represent different conformational states. In promiscuous assembly, the packaging motor assembles on a partial head produced by ejection of packaged DNA (E) or a full head (G), and refills the head with new fragments of DNA ([F] and [G]; new DNA fragments shown in red) (*131*).

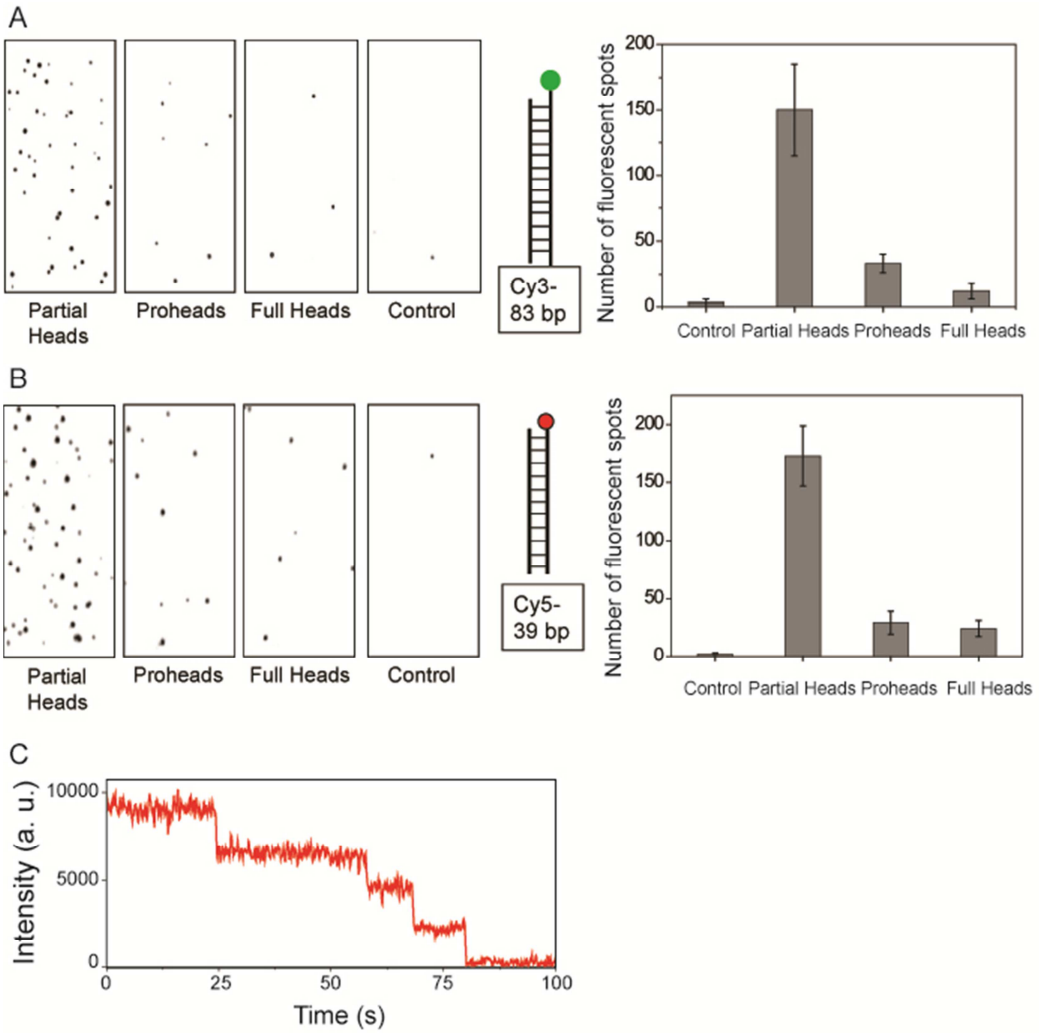

Figure 4.7 Single molecule fluorescence measurements of refilled heads. Quantification of packaging by single molecule fluorescence assay. (A) Fluorescence images of immobilized T4 heads packaged with Cy3 (83-bp) DNA. The average number of fluorescent spots per imaging area for each preparation method is shown. (B) Fluorescence images of immobilized T4 heads packaged with Cy5 (39-bp) DNA along with quantification. One-fourth of the 70 µm×35 µm imaging area is shown in each case. The number of fluorescence spots in more than 30 imaging areas was averaged in each case. Error bars denote the standard deviation of the mean. (C) Typical photobleaching profile from a single immobilized head, packaged with four Cy5-labeled DNA.

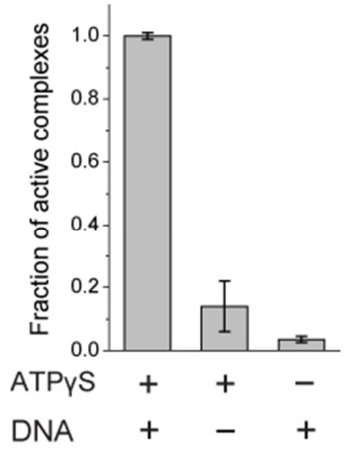

Figure 4.8 Quantifying the effect of cofactor (ATPγS) or priming DNA on the successful assembly of packaging complex.

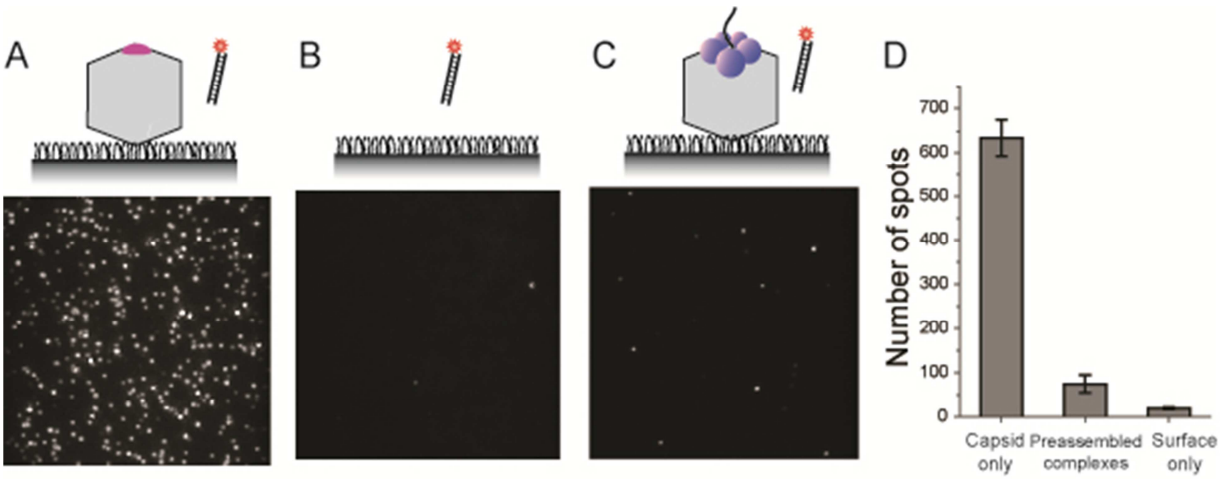

Figure 4.9 Direct demonstration of interaction of DNA with the head portal. (A) Virus heads immobilized on the surface bind Cy5-labeled DNA molecules directly. Our controls showed that the DNA binds to the head portal (magenta). (B) Very few background fluorescent spots are visible before immobilizing the heads. (C) Pre-assembled virus complexes do not interact with the DNA. The pre-assembly was done by mixing virus head, gp17 (purple) and an unlabeled priming DNA (black). In each case 2 nM of Cy5-labeled DNA was used. The representative images are 35µm×35µm. (D) Average number of spots on the surface for the cases shown in A, B and C. The number of molecules in at least 30 different imaging areas (70µm×35µm) was averaged in each case. Error bars are the standard deviation of the mean.

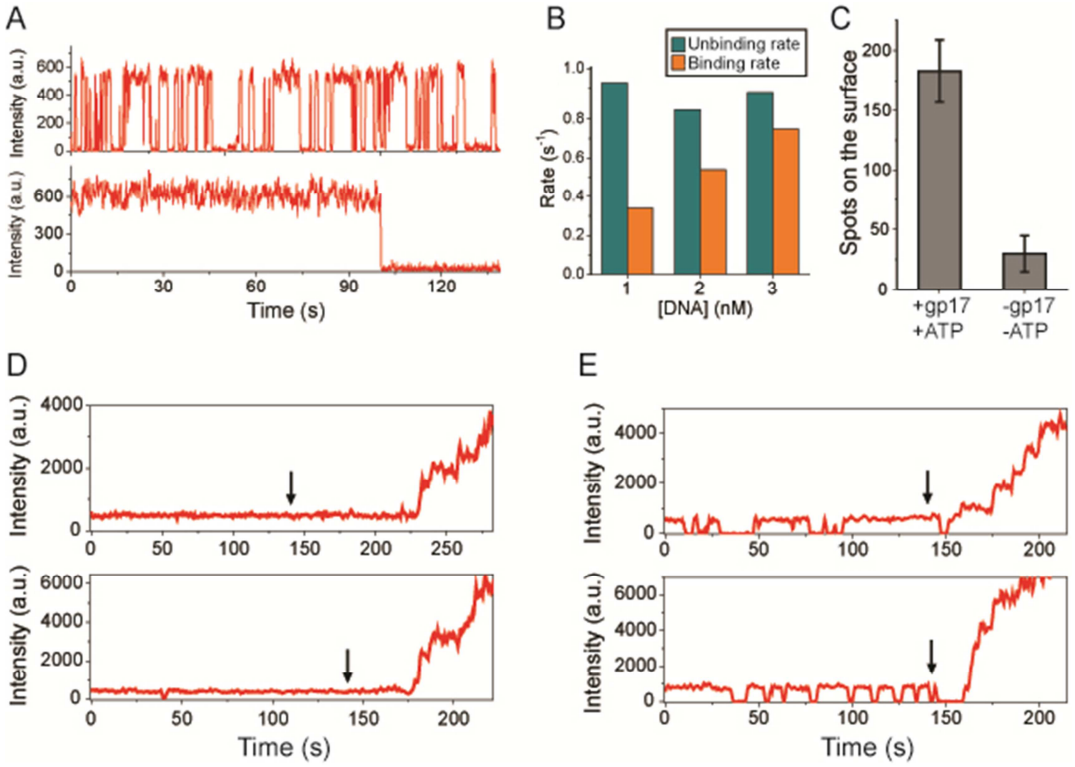

Figure 4.10 Quantifying the DNA-capsid interaction. (A) Most of the DNA molecules exhibit rapid binding, unbinding of DNA to the capsid (top), while some of them show stable binding to the capsid (bottom). (B) Increasing DNA concentration does not change the unbinding rate of DNA from capsid while it increases the binding rate in a linear way. (C) After providing gp17 and ATP, the DNA that was bound to the capsid can get packaged and resist DNase I digestion suggesting that the DNA was bound to the portal. (D) and (E) both kind of capsid-DNA interactions shown in (A) would package DNA efficiently upon addition of ATP, DNA and gp17. Surface -immobilized virus heads were incubated with 2nM Cy5-labeled DNA. While imaging the surface, at the time shown by the arrow, 1mM ATP, 2nM Cy5 DNA and 1µM gp17 were flowed into the chamber and real-time packaging was recorded.

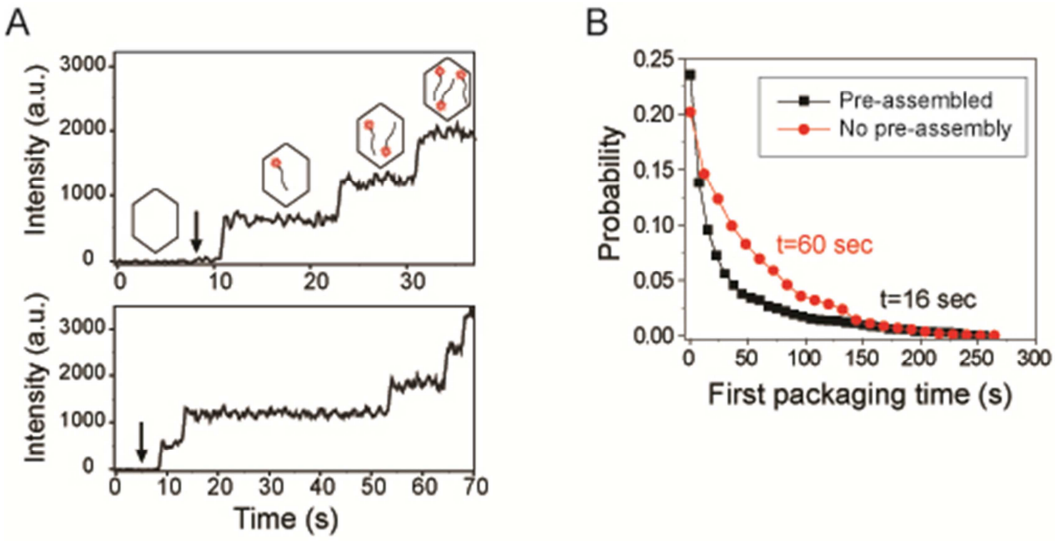

Figure 4.11 Monitoring packaging initiation in real-time. (A) Typical fluorescence intensity time traces of virus heads packaging DNA molecules one after another. The arrow denotes when ATP and labeled DNA was flushed into the chamber. (B) The first packaging time (time from when 4 nM DNA and 1 mM ATP were flowed in until the first DNA got packaged) for the pre-assembled complexes (black) and without prior pre-assembly (1 µM gp17 was also added along with DNA and ATP) (red). Exponential decay fit to these curves gives the packaging time of 16 sec for the pre-assembled complexes and 60 sec for the complexes without pre-assembly.

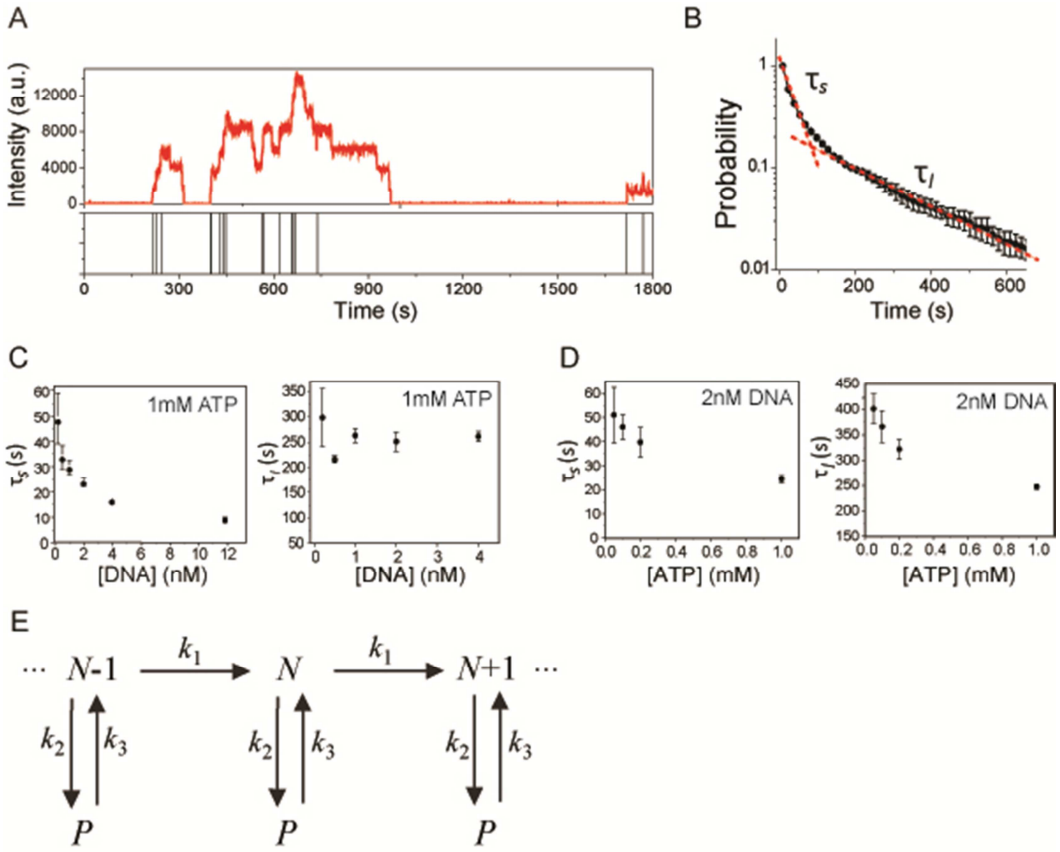

Figure 4.12 Quantifying the packaging initiation time as a function of ATP and DNA concentrations. (A) Fluorescence intensity time trace for a head that is packaging Cy5 labeled DNA molecules. Step-wise increase in the intensity corresponds to a new DNA molecule being packaged while the intensity drops correspond to photobleaching of individual DNA molecules. Each bar in the lower panel denotes the moment a new DNA molecule is packaged. (B) Packaging initiation times for pre-assembled complexes in 2 nM DNA and 1 mM ATP. Initiation times exhibit a double exponential distribution. (C) Short initiation time (τ_s) and long initiation time (τ_l) as a function of DNA concentration and in the presence of saturating ATP (1 mM). (D) Short initiation time (τ_s) and long initiation time (τ_l) as a function of ATP concentration and in the presence of 2nM DNA. (E) Proposed model for packaging initiation. Packaging motor packages DNA with the rate k_1. With a certain rate (k_2), the motor can transit into a paused state from which it recovers with the rate k_3.

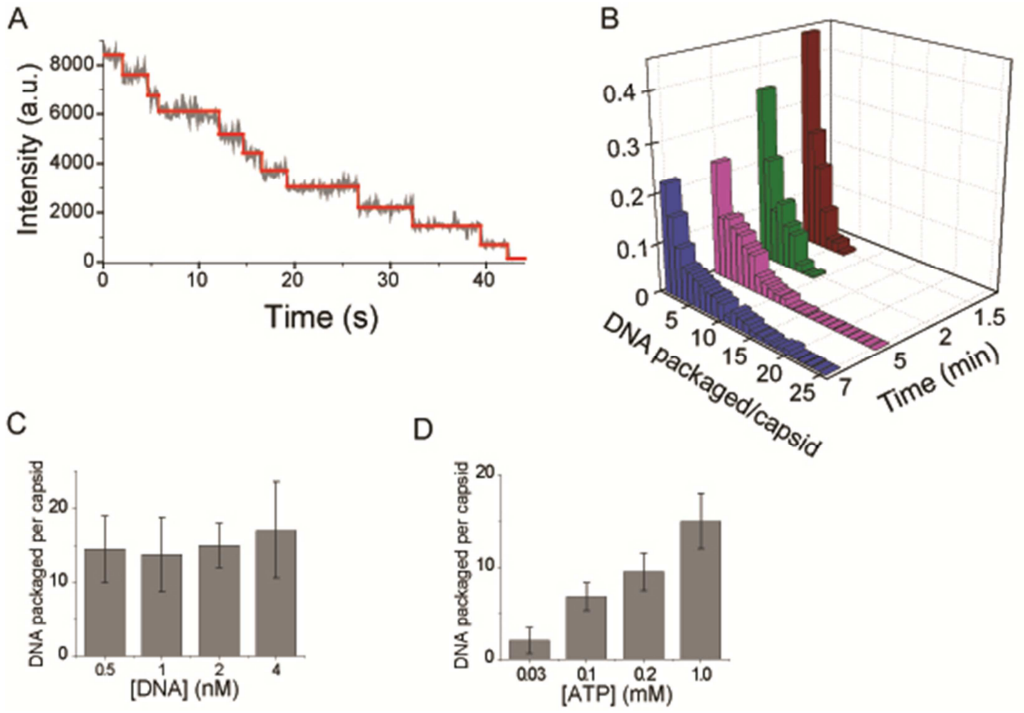

Figure 4.13 Quantifying the packaging efficiency using photobleaching analysis. (A) The photobleaching profile of a single head shows 11 distinct photobleaching steps. (B) Normalized distribution of the number of packaged DNA molecules in heads as a function of time. 2nM DNA and 1mM ATP was injected into a chamber containing surface immobilized pre-assembled packaging complexes. (C) Average number of DNA molecules per capsid after 30 min incubation with different DNA concentrations and 1mM ATP. (D) Average number of DNA molecules per head after 30 min incubation with 2nM DNA and varying ATP concentrations. In each case more than 150 heads were analyzed. Error bar is the standard deviation of the mean.

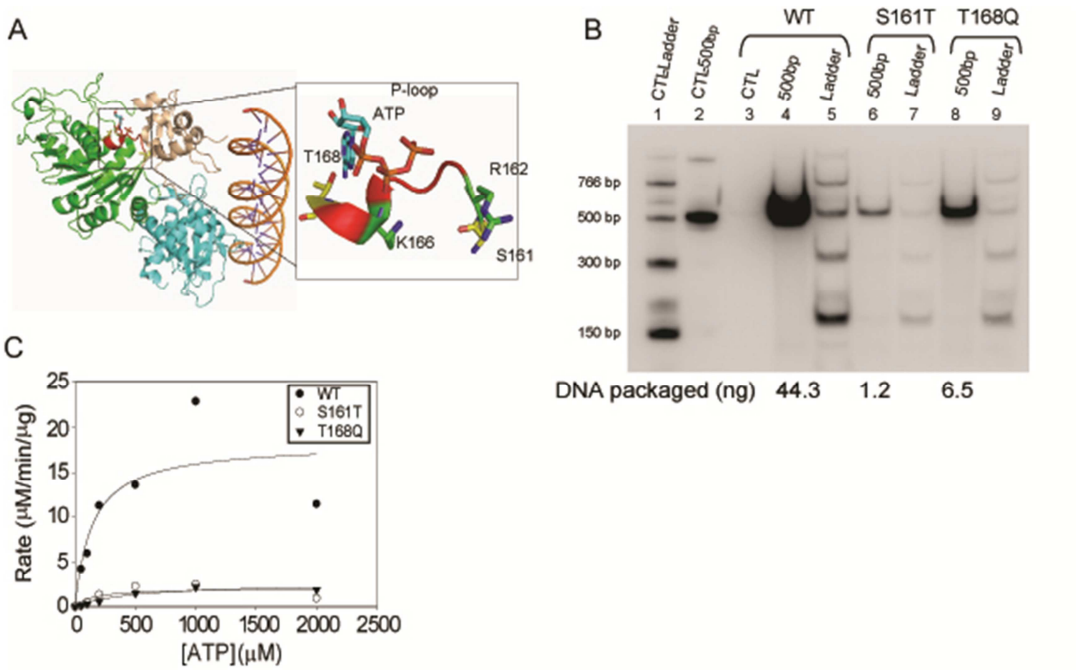

Figure 4.14 Bulk quantification of temperature sensitive mutants. (A) ATPase domain of gp17. Single mutations at S161T or T168Q results in cold sensitive (*cs*) or heat sensitive (*hs*) mutants. (B) Bulk packaging assay shows that temperature sensitive mutants are up to 30 fold less efficient in packaging DNA. (C) ATPase assay shows the temperature sensitive mutants have 10 fold less ATPase activity compared to wildtype gp17.

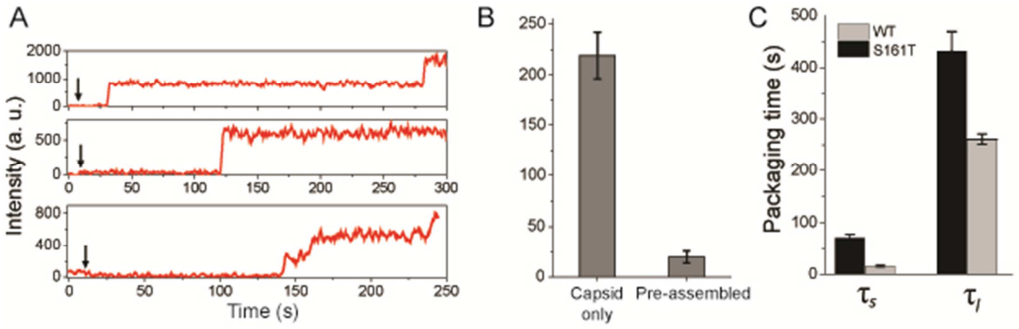

Figure 4.15 Single molecule quantification of temperature sensitive mutants. (A) representative fluorescence intensity time traces of virus heads, assembled with S161T gp17, packaging DNA molecules in real-time. The arrow denotes when 1mM ATP and 4 nM Cy5-labeled DNA were flushed into the chamber. (B) Number of DNase I-resistant fluorescent spots with different preparation methods. (C) Both short and long packaging times are much longer for the S161T mutant compared to the wildtype gp17.

Appendix A
PEGylation Procedure

One of the key requirements to do single molecule fluorescence experiments on surface tethered samples is having a good passivated surface. Biomolecules are usually charged and can easily stick to the slide surface. General passivation methods such as using bovine serum albumin (BSA) or lipid bilayers are quick and straightforward to use. However, these methods are not very reproducible and usually result in patchy coverage with areas of no passivation. The most efficient and robust passivation method is to covalently link polymer chains at high density to the surface. Polyethyleneglycol (PEG) is the most common polymer used for surface passivation. A good passivated surface, when incubated with 4 nM labeled dsDNA binds less than 20 molecules per 2,500 μm^2 of surface area. Proteins are usually more "sticky" and incubation of 5 nM labeled Rep helicase could result in up to 100 fluorescent spots per 2,500 μm^2, on a well passivated surface. It is important to note that a non passivated surface would be completely covered with fluorescent spots at that concentration of labeled rep protein.

There protocol for preparing PEGylated surface has three steps: 1- Cleaning of the slides and coverslip, 2- Aminosilanization of the surfaces and 3- PEGylating the surfaces. Here I will discuss each step and

Slide and Coverslips Cleaning

Impurities and dirt on the surface would inhibit PEGylation in that area and reduces the quality of surface passivation. The purpose of this step is to have a surface as clean as possible. We start with the slides that have holes already drilled in them. Since preparing the slides is a lengthy and cumbersome process, it is better to make as many channels as possible on each slide. Each slide can accommodate up to six channels. For the recycled slides we have to remove the tape and coverslip before the cleaning process. Keeping the used slides in water for a few days would significantly help with that.

1. Required: Glass beaker, razor blade, jars for holding the slides, Milli-Q water, methanol, acetone, 10% Alconox, 3M KOH, quartz slides, 40 mm coverslips (No. 1.5), flask.
2. Remove the tape from recycled slides. Use razor to shave the edges of the slide and remove epoxy.
3. Scrub the slides with acetone or methanol to remove the tape residues. Use a kimwipe to scrub resistant tape residues. The rule of the thumb is to scrub the slides so that no more tape is visible by eye. Finally scrub the slides with Alconox.
4. Rinse the slides with Milli-Q water and microwave them for 15 minutes.
5. Rinse the slides with Milli-Q and place them in glass jar for sonication. Place the coverslips in a separate jar. Start with a few extra coverslips since there is a good chance some of them break during the preparation.
6. Sonicate the slides and coverslips in acetone for 30 minutes.
7. In the meantime clean a glass flask for preparing the silanization solution in the next step. Sonicate the flask, filled with 1 M KOH, for 30 minutes.
8. Thaw the aminosilane in dark.
9. Dump the acetone out in a proper waste container and sonicate the slides and coverslip in 3 M KOH for 20 minutes.
10. Rinse the flask with Milli-Q, fill it with methanol and sonicate it for 30 minutes.
11. Rinse slides and coverslips with Milli-Q.

Aminosilanization

The purpose of this step is to link aminosilane to the clean slide and coverslip surface. m-PEG would covalently bind to this surface in the next step.

Required: Clean quartz slides and coverslips from the previous step, Propane torch, glass tips, acetic acid, Aminosilane, clean flask, PEG prep boxes.

12. Fill the jars containing the slides and coverslips with methanol and dry the methanol by gently blowing with nitrogen gas. The purpose of this step is to remove water since it can interfere with aminosilanization step.
13. Burn the coverslips in the propane flame. The coverslips easily deform or break when exposed to heat so each surface of the coverslips should not be burned more than 2 seconds. Return the coverslips to the jar after burning.
14. Burn the slides. Quartz slides can easily withstand the propane flame. Spend about 4 minutes to thoroughly burn the slide surfaces. Then cool the slides with a gentle nitrogen flow and return them to the jar.
15. Prepare Aminosilanization mixture in the clean flas: 150 mL methanol + 7.5 mL acetic acid + 1.5 mL aminosilane. Use glass pipette for acetic acid and aminosilane. Mix the solution well and pour in the slide and coverslip jars.

16. Incubate for 10 min then sonicate the jars for 1 min and incubate for another 10 min.
17. In the meantime, prepare pegylation buffer: 84 mg $NaHCO_3$ in 10 mL Milli-Q.
18. Put the aminosilane bottle in the desiccator to remove moisture before returning it to -20C freezer.

PEGylation

Required: Biotin-PEG, m-PEG, Sodium bicarbonate buffer, PEG assembly boxes.

19. In the meantime, prepare pegylation buffer: 84 mg $NaHCO_3$ in 10 mL Milli-Q. While the slides are incubating with aminosilane, prepare pegylation buffer: 84 mg $NaHCO_3$ in 10 mL Milli-Q.
20. Also, weight PEG powder and set it aside. For 5 slides use 80 mg mPEG and 1-2 mg biotin-PEG. It is important not to add too much biotin-PEG since it can increase the background spots.
21. Dump out the aminosilane solution in the proper waste container.
22. Rinse the jarsd 4 times with methanol
23. Dry the jars with gentle nitrogen blow and place them in the assembly boxes.
24. Add 50 µL of pegylation buffer to the PEG and mix well quickly by pipetting up and down.
25. Centrifuge at 10,000 rpm for 1 min.
26. After adding the buffer, PEG starts to hydrolyze quickly. It is important to do the mixing and centrifuge steps quickly.
27. Apply 60 µL of PEG solution to each slide and then place the coverslip on top. Note not to form bubbles between the slide and coverslip. Try to remove bubbles if they form. It is a good practice to PEGylate the same surface of the slides for the recycled slides
28. Store the slides in a dark, flat drawer. Slides can be kept overnight however the best quality is achieved in 4 to 6 hours.
29. Disassemble and rinse the slides and coverslips thoroughly with Milli-Q.
30. Put each slide and coverslip pair in a clean 50 mL tube, with the PEGylated surfaces facing away from each other. Store in -20C.

References:

1. G. G. Stokes, On the change of refrangibility of light. *Philos. Trans. R. Soc. Lond.* **142**, 463 (1852).
2. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*. (Springer, ed. 3rd, 2006).
3. T. Forster, Experimental and Theoretical Investigation of the Intermolecular Transfer of Electronic Excitation Energy. *Zeitschrift Naturforsch A* **4**, 321 (1949).
4. A. Iqbal *et al.*, Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proceedings of the National Academy of Sciences* **105**, 11176 (August 12, 2008, 2008).
5. C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, T. Ha, in *Annu. Rev. Biochem.* (Annual Reviews, Palo Alto, 2008), vol. 77, pp. 51-76.
6. T. Ha, Single-Molecule Fluorescence Resonance Energy Transfer. *Methods* **25**, 78 (2001).
7. X. Zhuang *et al.*, Fluorescence quenching: A tool for single-molecule protein-folding study. *Proceedings of the National Academy of Sciences* **97**, 14241 (December 19, 2000, 2000).
8. A. Yildiz *et al.*, Myosin V Walks Hand-Over-Hand: Single Fluorophore Imaging with 1.5-nm Localization. *Science* **300**, 2061 (June 27, 2003, 2003).
9. K. Adachi *et al.*, Stepping rotation of F1-ATPase visualized through angle-resolved single-fluorophore imaging. *Proceedings of the National Academy of Sciences* **97**, 7243 (June 20, 2000, 2000).
10. T. Funatsu, Y. Harada, M. Tokunaga, K. Saito, T. Yanagida, Imaging of Single Fluorescent Molecules and Individual Atp Turnovers by Single Myosin Molecules in Aqueous-Solution. *Nature* **374**, 555 (Apr 6, 1995).
11. T. Ha *et al.*, Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proceedings of the National Academy of Sciences* **93**, 6264 (June 25, 1996, 1996).
12. D. Axelrod, Total internal reflection fluorescence microscopy in cell biology. *Method Enzymol* **361**, 1 (2003).
13. W. E. Moerner, D. P. Fromm, Methods of single-molecule fluorescence spectroscopy and microscopy. *Rev. Sci. Instrum.* **74**, 3597 (Aug, 2003).
14. T. Ha, P. Tinnefeld, in *Annual Review of Physical Chemistry, Vol 63*, M. A. Johnson, T. J. Martinez, Eds. (Annual Reviews, Palo Alto, 2012), vol. 63, pp. 595-617.
15. R. Zondervan, F. Kulzer, S. B. Orlinskii, M. Orrit, Photoblinking of Rhodamine 6G in Poly(vinyl alcohol): Radical Dark State Formed through the Triplet. *The Journal of Physical Chemistry A* **107**, 6770 (2003/09/01, 2003).
16. I. Rasnik, S. A. McKinney, T. Ha, Nonblinking and longlasting single-molecule fluorescence imaging. *Nat. Methods* **3**, 891 (Nov, 2006).
17. X. Shi, J. Lim, T. Ha, Acidification of the Oxygen Scavenging System in Single-Molecule Fluorescence Studies: In Situ Sensing with a Ratiometric Dual-Emission Probe. *Analytical Chemistry* **82**, 6132 (2010/07/15, 2010).
18. E. M. Blackwood, J. T. Kadonaga, Going the distance: A current view of enhancer action. *Science* **281**, 60 (Jul, 1998).

19. O. K. Wong, M. Guthold, D. A. Erie, J. Gelles, Interconvertible Lac Repressor–DNA Loops Revealed by Single-Molecule Experiments. *PLoS Biol* **6**, e232 (2008).
20. T. J. Richmond, C. A. Davey, The structure of DNA in the nucleosome core. *Nature* **423**, 145 (May 8, 2003).
21. C. G. Baumann, S. B. Smith, V. A. Bloomfield, C. Bustamante, Ionic effects on the elasticity of single DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 6185 (Jun 10, 1997).
22. R. Phillips, J. Kondev, J. Theriot, *Physical Biology of the Cell*. (Garland Science, 2008).
23. R. S. Mathew-Fenn, R. Das, P. A. B. Harbury, Remeasuring the Double Helix. *Science* **322**, 446 (October 17, 2008, 2008).
24. J. F. Marko, in *Les Houches Summer School Proceedings*, D. Chatenay, R. Monasson, D. Thieffry, J. Dalibard, Eds. (Elsevier, 2005), pp. 211–270.
25. Schellma.Ja, Flexibility of DNA. *Biopolymers* **13**, 217 (1974).
26. L. Landau, E. Lifshitz, *Theory of Elasticity*. (Pergamon, New York, 1986).
27. C. Frontali *et al.*, Absolute Method for the Determination of the Persistence Length of Native DNA from Electron-Micrographs. *Biopolymers* **18**, 1353 (1979).
28. C. W. Schmid, F. P. Rinehart, J. E. Hearst, Statistical Length of DNA from Light Scattering. *Biopolymers* **10**, 883 (1971).
29. J. B. Hays, M. E. Magar, B. H. Zimm, Persistence Length of DNA. *Biopolymers* **8**, 531 (1969).
30. R. Ullman, Intrinsic Viscosity of Wormlike Polymer Chains. *J. Chem. Phys.* **49**, 5486 (1968).
31. D. Jolly, H. Eisenberg, Photon Correlation Spectroscopy, Total Intensity Light-Scattering with Laser Radiation, and Hydrodynamic Studies of a Well Fractionated DNA Sample. *Biopolymers* **15**, 61 (1976).
32. P. Sharp, Bloomfie.Va, Light Scattering from Wormlike Chains with Excluded Volume Effects. *Biopolymers* **6**, 1201 (1968).
33. P. J. Hagerman, Investigation of the Flexibility of DNA Using Transient Electric Birefringence. *Biopolymers* **20**, 1503 (1981).
34. P. J. Hagerman, B. H. Zimm, Monte-Carlo Approach to the Analysis of the Rotational Diffusion of Wormlike Chains. *Biopolymers* **20**, 1481 (1981).
35. S. Weiss, Fluorescence Spectroscopy of Single Biomolecules. *Science* **283**, 1676 (March 12, 1999, 1999).
36. S. B. Smith, L. Finzi, C. Bustamante, Direct Mechanical Measurements of the Elasticity of Single DNA-Molecules by Using Magnetic Beads. *Science* **258**, 1122 (Nov 13, 1992).
37. C. Bustamante, J. F. Marko, E. D. Siggia, S. Smith, Entropic Elasticity of Lambda-Phage DNA. *Science* **265**, 1599 (Sep 9, 1994).
38. J. F. Marko, E. D. Siggia, Stretching DNA. *Macromolecules* **28**, 8759 (Dec 18, 1995).
39. M. D. Wang, H. Yin, R. Landick, J. Gelles, S. M. Block, Stretching DNA with optical tweezers. *Biophysical journal* **72**, 1335 (Mar, 1997).
40. H. Yamakawa, Stockmay.Wh, Statistical Mechanics of Wormlike Chains .2. Excluded Volume Effects. *J. Chem. Phys.* **57**, 2843 (1972).

41. J. Shimada, H. Yamakawa, Ring-closure probabilities for twisted wormlike chains. Application to DNA. *Macromolecules* **17**, 689 (1984/04/01, 1984).
42. J. Yan, R. Kawamura, J. F. Marko, Statistics of loop formation along double helix DNAs. *Phys Rev E* **71**, (Jun, 2005).
43. D. Shore, J. Langowski, R. L. Baldwin, DNA flexibility studied by covalent closure of short fragments into circles. *Proceedings of the National Academy of Sciences* **78**, 4833 (August 1, 1981, 1981).
44. H. Jacobson, W. H. Stockmayer, Intramolecular Reaction in Polycondensations. I. The Theory of Linear Systems. *The Journal of Chemical Physics* **18**, 1600 (1950).
45. W. H. Taylor, P. J. Hagerman, Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure: II. NaCl-dependence of DNA flexibility and helical repeat. *Journal of Molecular Biology* **212**, 363 (1990).
46. T. E. Cloutier, J. Widom, Spontaneous sharp bending of double-stranded DNA. *Mol Cell* **14**, 355 (May 7, 2004).
47. T. E. Cloutier, J. Widom, DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 3645 (Mar 8, 2005).
48. J. Yan, J. F. Marko, Localized Single-Stranded Bubble Mechanism for Cyclization of Short Double Helix DNA. *Physical Review Letters* **93**, 108108 (2004).
49. P. A. Wiggins, R. Phillips, P. C. Nelson, Exact theory of kinkable elastic polymers. *Phys Rev E* **71**, 021909 (2005).
50. F. Lankas, R. Lavery, J. H. Maddocks, Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure* **14**, 1527 (Oct, 2006).
51. G. L. Randall, L. Zechiedrich, B. M. Pettitt, In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form. *Nucleic Acids Res* **37**, 5568 (Sep, 2009).
52. T. A. Lionberger *et al.*, Cooperative kinking at distant sites in mechanically stressed DNA. *Nucleic Acids Res* **39**, 9820 (December 1, 2011, 2011).
53. Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, A. Vologodskii, Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5397 (April 12, 2005, 2005).
54. P. A. Wiggins *et al.*, High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat Nanotechnol* **1**, 137 (Nov, 2006).
55. G. Witz, K. Rechendorff, J. Adamcik, G. Dietler, Conformation of Circular DNA in Two Dimensions. *Physical Review Letters* **101**, 148103 (2008).
56. F. G. A. Faas, B. Rieger, L. J. van Vliet, D. I. Cherny, DNA Deformations near Charged Surfaces: Electron and Atomic Force Microscopy Views. *Biophysical journal* **97**, 1148 (2009).
57. N. B. Becker, R. Everaers, Comment on “Remeasuring the Double Helix”. *Science* **325**, 538 (July 31, 2009, 2009).
58. A. J. Mastroianni, D. A. Sivak, P. L. Geissler, A. P. Alivisatos, Probing the Conformational Distributions of Subpersistence Length DNA. *Biophysical journal* **97**, 1408 (2009).

59. Y. Seol, J. Y. Li, P. C. Nelson, T. T. Perkins, M. D. Betterton, Elasticity of short DNA molecules: Theory and experiment for contour lengths of 0.6-7 mu m. *Biophysical journal* **93**, 4360 (Dec, 2007).
60. R. Roy, S. Hohng, T. Ha, A practical guide to single-molecule FRET. *Nat. Methods* **5**, 507 (Jun, 2008).
61. A. M. Burkhoff, T. D. Tullius, Structural details of an adenine tract that does not cause DNA to bend. *Nature* **331**, 455 (Feb, 1988).
62. L. Han *et al.*, Concentration and Length Dependence of DNA Looping in Transcriptional Regulation. *Plos One* **4**, e5621 (2009).
63. N. A. Becker, J. D. Kahn, L. James Maher Iii, Bacterial Repression Loops Require Enhanced DNA Flexibility. *Journal of Molecular Biology* **349**, 716 (2005).
64. D. H. Lee, R. F. Schleif, In vivo DNA loops in araCBAD: size limits and helical repeat. *Proceedings of the National Academy of Sciences* **86**, 476 (January 1, 1989, 1989).
65. I. Cisse, B. Okumus, C. Joo, T. Ha, Fueling protein–DNA interactions inside porous nanocontainers. *Proceedings of the National Academy of Sciences* **104**, 12646 (July 31, 2007, 2007).
66. P. J. Hagerman, Flexibility of DNA. *Annual Review of Biophysics and Biophysical Chemistry* **17**, 265 (1988).
67. P. T. Lowary, J. Widom, Nucleosome packaging and nucleosome positioning of genomic DNA. *Proceedings of the National Academy of Sciences* **94**, 1183 (February 18, 1997, 1997).
68. H. Shroff *et al.*, Optical measurement of mechanical forces inside short DNA loops. *Biophysical journal* **94**, 2179 (Mar 15, 2008).
69. H. Shroff *et al.*, Biocompatible force sensor with optical readout and dimensions of 6 nm(3). *Nano Lett* **5**, 1509 (Jul, 2005).
70. K. Hatch, C. Danilowicz, V. Coljee, M. Prentiss, Demonstration that the shear force required to separate short double-stranded DNA does not increase significantly with sequence length for sequences longer than 25 base pairs. *Phys Rev E* **78**, 011920 (2008).
71. P.-G. de Gennes, Maximum pull out force on DNA hybrids. *Comptes Rendus de l'Académie des Sciences - Series IV - Physics* **2**, 1505 (2001).
72. D. Porschke, M. Eigen, Co-operative non-enzymic base recognition. III. Kinetics of the helix-coil transition of the oligoribouridylic-oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *Journal of Molecular Biology* **62**, 361 (1971).
73. J. Shimada, H. Yamakawa, Ring-closure probabilities for twisted wormlike chains. *Macromolecules* **17**, 689 (1984).
74. K. B. Towles, J. F. Beausang, H. G. Garcia, R. Phillips, P. C. Nelson, First-principles calculation of DNA looping in tethered particle experiments. *Physical Biology* **6**, (Jun, 2009).
75. N. Douarche, S. Cocco, Protein-mediated DNA loops: Effects of protein bridge size and kinks. *Phys Rev E* **72**, (Dec, 2005).
76. R. Schleif, DNA Looping. *Annu. Rev. Biochem.* **61**, 199 (1992).

77. T. T. Paull, M. J. Haykinson, R. C. Johnson, The nonspecific DNA-binding and -bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures. *Genes & Development* **7**, 1521 (August 1, 1993, 1993).
78. D. Shore, R. L. Baldwin, Energetics of DNA twisting: I. Relation between twist and cyclization probability. *Journal of Molecular Biology* **170**, 957 (1983).
79. D. Kotlarz, A. Fritsch, H. Buc, Variations of Intramolecular Ligation Rates Allow the Detection of Protein-Induced Bends in DNA. *Embo J.* **5**, 799 (Apr, 1986).
80. P. M. Pil, C. S. Chow, S. J. Lippard, HIGH-MOBILITY-GROUP-1 PROTEIN MEDIATES DNA BENDING AS DETERMINED BY RING CLOSURES. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 9465 (Oct, 1993).
81. T. A. Azam, A. Iwata, A. Nishimura, S. Ueda, A. Ishihama, Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J Bacteriol* **181**, 6361 (Oct, 1999).
82. Y. Hodgesgarcia, P. J. Hagerman, D. E. Pettijohn, DNA Ring-Closure Mediated by Protein Hu. *Journal of Biological Chemistry* **264**, 14621 (Sep 5, 1989).
83. J. van Noort, S. Verbrugge, N. Goosen, C. Dekker, R. T. Dame, Dual architectural roles of HU: Formation of flexible hinges and rigid filaments. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6969 (May 4, 2004, 2004).
84. W. W. Thompson *et al.*, Influenza-associated hospitalizations in the United States. *JAMA-J. Am. Med. Assoc.* **292**, 1333 (Sep, 2004).
85. T. K. W. Cheung, L. L. M. Poon, Biology of Influenza A Virus. *Annals of the New York Academy of Sciences* **1102**, 1 (2007).
86. B. N. Fields, D. M. Knipe, P. M. Howley, *Field's Virology*. (Lippincott Williams & Wilkins, 2007).
87. N. M. Bouvier, P. Palese, The biology of influenza viruses. *Vaccine* **26, Supplement 4**, D49 (2008).
88. V. Trifonov, H. Khiabanian, R. Rabadan, Geographic Dependence, Surveillance, and Origins of the 2009 Influenza A (H1N1) Virus. *New England Journal of Medicine* **361**, 115 (2009).
89. M. Enami, G. Sharma, C. Benham, P. Palese, An influenza virus containing nine different RNA segments. *Virology* **185**, 291 (1991).
90. C. T. Bancroft, T. G. Parslow, Evidence for Segment-Nonspecific Packaging of the Influenza A Virus Genome. *Journal of Virology* **76**, 7133 (2002).
91. H. B. Donald, A. Isaacs, Counts of Influenza Virus Particles. *Journal of General Microbiology* **10**, 457 (June 1, 1954, 1954).
92. S. D. Duhaut, J. W. McCauley, Defective RNAs Inhibit the Assembly of Influenza Virus Genome Segments in a Segment-Specific Manner. *Virology* **216**, 326 (1996).
93. E. C. Hutchinson, J. C. von Kirchbach, J. R. Gog, P. Digard, Genome packaging in influenza A virus. *Journal of General Virology* **91**, 313 (February 1, 2010, 2010).
94. G. A. Marsh, R. Rabadán, A. J. Levine, P. Palese, Highly Conserved Regions of Influenza A Virus Polymerase Gene Segments Are Critical for Efficient Viral RNA Packaging. *Journal of Virology* **82**, 2295 (March 1, 2008, 2008).

95. G. A. Marsh, R. Hatami, P. Palese, Specific Residues of the Influenza A Virus Hemagglutinin Viral RNA Are Important for Efficient Packaging into Budding Virions. *Journal of Virology* **81**, 9727 (September 15, 2007, 2007).
96. T. Noda *et al.*, Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature* **439**, 490 (Jan, 2006).
97. T. Noda *et al.*, Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nat. Commun.* **3**, (Jan, 2012).
98. A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877 (Oct, 2008).
99. Q. Gao, A. Lowen, T. Wang, P. Palese, A Nine-Segment Influenza A Virus Carrying Subtype H1 and H3 Hemagglutinins. *Journal of Virology* **84**, 8062 (2010).
100. Z. Wei *et al.*, Biophysical characterization of influenza virus subpopulations using field flow fractionation and multiangle light scattering: Correlation of particle counts, size distribution and infectivity. *J. Virol. Methods* **144**, 122 (Sep, 2007).
101. A. M. Femino, F. S. Fay, K. Fogarty, R. H. Singer, Visualization of Single RNA Transcripts in Situ. *Science* **280**, 585 (April 24, 1998, 1998).
102. A. Raj, S. Tyagi, in *Methods in Enzymology*, G. W. Nils, Ed. (Academic Press, 2010), vol. Volume 472, pp. 365-386.
103. A. Jain *et al.*, Probing cellular protein complexes using single-molecule pull-down. *Nature* **473**, 484 (May, 2011).
104. M. H. Ulbrich, E. Y. Isacoff, Subunit counting in membrane-bound proteins. *Nat. Methods* **4**, 319 (Apr, 2007).
105. P. A. Jennings, J. T. Finch, G. Winter, J. S. Robertson, Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? *Cell* **34**, 619 (1983).
106. A. Portela, P. Digard, The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication. *Journal of General Virology* **83**, 723 (2002).
107. E. S. Miller *et al.*, Bacteriophage T4 Genome. *Microbiology and Molecular Biology Reviews* **67**, 86 (2003).
108. R. Calendar, Ed., *The Bacteriophages*, (Oxford University Press, USA, 2005).
109. S. R. Casjens, The DNA-packaging nanomotor of tailed bacteriophages. *Nat. Rev. Microbiol.* **9**, 647 (Sep, 2011).
110. R. A. Driedonks, A. Engel, B. tenHeggeler, R. van Driel, Gene 20 product of bacteriophage T4 its purification and structure. *Journal of Molecular Biology* **152**, 641 (1981).
111. A. A. Simpson *et al.*, Structure of the bacteriophage phi 29 DNA packaging motor. *Nature* **408**, 745 (Dec, 2000).
112. V. B. Rao, L. W. Black, in *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*, C. E. Catalano, Ed. (Springer, 2005), pp. 40-58.
113. R. B. Luftig, C. Ganz, Bacteriophage T4 Head Morphogenesis IV. Comparison of Gene 16-, 17-, and 49-Defective Head Structures2. *Journal of Virology* **10**, 545 (September 1, 1972, 1972).

114. K. R. Kondabagil, Z. Zhang, V. B. Rao, The DNA Translocating ATPase of Bacteriophage T4 Packaging Motor. *Journal of Molecular Biology* **363**, 786 (2006).
115. L. W. Black, G. Peng, Mechanistic Coupling of Bacteriophage T4 DNA Packaging to Components of the Replication-dependent Late Transcription Machinery. *Journal of Biological Chemistry* **281**, 25635 (September 1, 2006, 2006).
116. A. S. Al-Zahrani *et al.*, The Small Terminase, gp16, of Bacteriophage T4 Is a Regulator of the DNA Packaging Motor. *Journal of Biological Chemistry* **284**, 24490 (September 4, 2009, 2009).
117. V. B. Rao, L. W. Black, Cloning, overexpression and purification of the terminase proteins gp16 and gp17 of bacteriophage T4: Construction of a defined in-vitro DNA packaging system using purified terminase proteins. *Journal of Molecular Biology* **200**, 475 (1988).
118. S. Kanamaru, K. Kondabagil, M. G. Rossmann, V. B. Rao, The Functional Domains of Bacteriophage T4 Terminase. *Journal of Biological Chemistry* **279**, 40795 (September 24, 2004, 2004).
119. S. Sun *et al.*, The Structure of the Phage T4 DNA Packaging Motor Suggests a Mechanism Dependent on Electrostatic Forces. *Cell* **135**, 1251 (2008).
120. V. B. Rao, M. Feiss, in *Annual Review of Genetics*. (Annual Reviews, Palo Alto, 2008), vol. 42, pp. 647-681.
121. Y. Xiang *et al.*, Structural changes of bacteriophage phi 29 upon DNA packaging and release. *Embo J.* **25**, 5229 (Nov, 2006).
122. W. C. Earnshaw, S. R. Casjens, DNA packaging by the double-stranded DNA bacteriophages. *Cell* **21**, 319 (1980).
123. J. R. Moffitt *et al.*, Intersubunit coordination in a homomeric ring ATPase. *Nature* **457**, 446 (Jan, 2009).
124. C. Bustamante, W. Cheng, Y. X. Mejia, Revisiting the Central Dogma One Molecule at a Time. *Cell* **144**, 480 (2011).
125. D. N. Fuller, D. M. Raymer, V. I. Kottadiel, V. B. Rao, D. E. Smith, Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. *Proceedings of the National Academy of Sciences* **104**, 16868 (October 23, 2007, 2007).
126. D. Shu, H. Zhang, J. S. Jin, P. X. Guo, Counting of six pRNAs of phi29 DNA-packaging motor with customized single-molecule dual-view system. *Embo J.* **26**, 527 (Jan, 2007).
127. C. R. Sabanayagam, M. Oram, J. R. Lakowicz, L. W. Black, Viral DNA Packaging Studied by Fluorescence Correlation Spectroscopy. *Biophysical journal* **93**, L17 (2007).
128. K. Ray, J. Ma, M. Oram, J. R. Lakowicz, L. W. Black, Single-Molecule and FRET Fluorescence Correlation Spectroscopy Analyses of Phage DNA Packaging: Colocalization of Packaged Phage T4 DNA Ends within the Capsid. *Journal of Molecular Biology* **395**, 1102 (2010).
129. K. Ray, C. R. Sabanayagam, J. R. Lakowicz, L. W. Black, DNA crunching by a viral packaging motor: Compression of a procapsid-portal stalled Y-DNA substrate. *Virology* **398**, 224 (2010).

130. W. B. Wood, Genetic-control of Organelle Assembly at the Molecular-level.3. Bacteriophage T4 Morphogenesis as a Model for Assembly of Subcellular Structure *Q. Rev. Biol.* **55**, 353 (1980).
131. Z. Zhang *et al.*, A Promiscuous DNA Packaging Machine from Bacteriophage T4. *PLoS Biol* **9**, e1000592 (2011).
132. S. S. Patel, K. M. Picha, Structure and function of hexameric helicases. *Annu. Rev. Biochem.* **69**, 651 (2000).
133. I. Golding, J. Paulsson, S. M. Zawilski, E. C. Cox, Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* **123**, 1025 (2005).
134. V. B. Rao, M. S. Mitchell, The N-terminal ATPase site in the large terminase protein gp17 is critically required for DNA packaging in bacteriophage T4. *Journal of Molecular Biology* **314**, 401 (2001).
135. A. Fokine *et al.*, Molecular architecture of the prolate head of bacteriophage T4. *PNAS* **101**, 6003 (2004).