Studying viral-like particles using fluorescence fluctuation spectroscopy

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Jolene Johnson Armstrong

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Adviser Dr. Joachim Mueller

October 2012

Acknowledgements

This dissertation would not have been possible without the support of my family, friends, and lab members. I would like to thank…

Joachim D. Müller, for your guildance and support. Your critical attitude and thorough approach to science has molded me into the scientist I am today.

Yan Chen for your support, scientific insights, and frequent conversations about science and everything else. You are a great role model and friend.

My labmates both past and present. Joseph Skinner, Bin Wu Lindsay, Hillesheim for all the help you provided in the lab and for being great senior lab memebers. Serkan Berk, Kwang Ho Hur, Issac Angert, Keir Fogarty, and Jinhui Li for the scientific conversations and putting up with me over the years. I would like to especially thank Patrick MacDonald, and Elizabeth Smith. I would not have made it without guys. Thank you for being not only scientific collaborators, but also supporting friends. I am grateful to all the members of the Mueller group for the great research environment and for making the Mueller lab a great place to spend 7 years.

Members of the Mansky lab for your insight and help with the HTLV-1 project. I would like to especially thank Iwen Grigsby for the scientific insight and interesting conversation.

My fellow physics graduate students who helped me get through those first years of classes, and for being great friends.

Gustavus and all the Gustavus physics majors of 2005.

My family including my parents, grandparents and siblings. Your love and support of my endless studies since I was a nerdy middle school student interested in science all the way through graduate school has made all the difference

Finally and most importantly, my husband Zach Armstrong. I literally would not have made it through without you. Your love and support provided meaning to my work and kept me going (that and all the food you cooked and laundry you did helped ☺)

This work has been supported by many sources of funding including, the Alworth Scholarship, the Pepin Fellowship, the Doctoral Dissertation Fellowship, the National Institutes of Health and the National Science Foundation.

Dedication

This dissertation is in memory of my grandmothers. I wish you could see this day, but I am grateful for everything you have done to get me here.

Table of Contents

List of Tables

List of Figures

Glossary of Abbreviations

1D-PCH	One-dimensional photon counting histogram
2D-PCH	Two-dimensional photon counting histogram
APD	Avalanche Photodiode
A3G	Apolipoprotein B mRNA-editing, enzyme-catalytic, polypeptide-like
CA	Capsid domain of Gag
cryo-TEM	Cryogenic Transmission Electron Microscopy
DMEM	Dulbecco's Modified Eagle Medium
EGFP	Enhanced Green Florescent Protein
EYFP	Enhanced Yellow Florescent Protein
FCA	Fluorescence Cumulant Analysis
FCS	Florescence Correlation Spectroscopy
FFS	Florescence Fluctuation Spectroscopy
FIDA	Fluorescence Intensity Distribution Analysis
FRET	Fluorescence Resonance Energy Transfer
Gag	Group specific antigen
HIV-1	Human Immunodeficiency Virus type 1
HTLV-1	Human T-lymphotropic Virus type 1
IFP	Intensity Fraction Plot
IPA	Isopropyl Alcohol
MA	Matrix domain of Gag
MuV	Murine leukemia virus
NA	Numerical Aperture
NC	Nucleocapsid domain of Gag
NFC	Nanofabrication Center
PBS	Phosphate Buffered Saline
PCH	Photon Counting Histogram
PDF	Probability Distribution Function
PEB	Post Exposure Bake
PSF	Point Spread Function
RSV	Rous Sarcoma Virus
SDS	Sodium Dodecycl Sulfate
TIFCA	Time-Integrated Fluorescence Cumulant Analysis
VLP	Viral Like Particle
VSV-G	Vesicular Stomatitis Virus G

Chapter 1: Introduction to Viral-like Particles

1.1 Introduction

Viruses are pathogens in every kingdom of life. They are neither truly living nor truly dead, as they rely on their host cell for replication and do little else besides infect and replicate. There are estimated to be 10^{31} viruses on earth (1). Many of these viruses are essential to human life and are integrated into our complex ecosystem, but some, such as the flu virus, are detrimental to human health. Because viruses rely on hosts for survival, they are orders of magnitude simpler than any other truly living system. The simplest virus, porcine circoviruses, contains only two genes: one for replicating its genome, and one for encoding a single protein that forms its shell (2). Some viruses are so simple that they can be reconstituted in buffer solution by combining purified copies of nucleic acids and a single protein molecule (3) . The simplicity of viruses allows us a glimpse into the complicated biological world. In the 1940's, Max Delbruck launched the pursuit of quantitative biology by demonstrating that studying viral replication leads to a better understanding of the basic molecular aspects of genetics and the synthesis of proteins by DNA (4). It is in his path that we follow. Viruses are the simplest model system we have in biology. It is our hope that studying this simple model promotes understanding of the underlying biological and physical principles that govern this simple system that straddles the line between life and non-life.

1.2 Fluorescence Fluctuation Spectroscopy (FFS)

Proteins are essential building blocks of biological systems and form the structure of viruses. Quantifying protein-protein interactions is notoriously difficult. Most approaches involve studying purified sample with biochemical assays. While these approaches have provided a host of valuable information, they provide mainly qualitative results. Fluorescence fluctuation spectroscopy (FFS), on the other hand, is a quantitative that determines the interactions of proteins and other biomolecules from

fluorescence signal fluctuations with single molecule sensitivity (5–7) FFS can directly quantify protein interactions inside a living cell (8), giving it a distinct advantage over previously used techniques. In FFS, proteins of interest are labeled with a fluorescent tag, often a fluorescent protein. Every time a labeled protein passes through a tiny optical observation volume (< 1 fL), it creates an observable fluctuation in fluorescence signal. Statistical analysis extracts information about the sample from these fluctuations. Through statistics, we can determine information about the transport properties, the concentration, and the oligomerization state of the labeled protein.

Our lab mainly uses FFS to study interactions of viral and other proteins directly inside cells. In this dissertation, I will step outside the cell and explore FFS as a tool to study the protein copy number of released viral particles. The protein copy number provides implicit information on the assembly of the virus as discussed in more detail below. The statistical analysis of FFS data and additional background information will be covered in Chapter 2.

1.3 Virus Model System

Recently, our group has extended FFS techniques to study retroviral assembly (9). Compared to other biological systems, viruses are relatively simple organisms consisting of only two or three main components; the genetic sequence (either RNA or DNA), a protein coat, and often a lipid membrane. While not all viruses have a lipid membrane, those that do are called enveloped. This lipid membrane is acquired from the host cell during assembly. The presence of a membrane and the inability to fabricate these viruses in a test tube presents significant challenges to study the composition and copy number of the proteins residing in the particle. In this dissertation, I will present a spectroscopic method of characterizing proteins within enveloped viruses that can be extended to study other bioparticles and vesicles.

Our lab first started studying the enveloped retrovirus HIV-1 (Human Immunodeficiency Virus Type 1) as our model system. The retrovirus family contains

many other viruses, most notably HTLV-1 (Human T-lymphotropic Virus Type 1), which will be discussed in Chapter 3 of this dissertation. Retroviruses store their genetic information as RNA and use reverse transcriptase to make a DNA copy of their genome. This DNA is incorporated into the host cell's DNA using integrase. Subsequently, the cellular machinery produces viral RNA and proteins, leading to the production of new particles (10).

All retroviruses contain Gag (group specific antigen). Gag plays a crucial role in the lifecycle of the virus, and is the main structural protein of the virus. It is produced in the cytoplasm of the host cell and is then targeted to the membrane. Gag assembles at the membrane, leading to the formation of a small bud (~140 nm diameter) with Gag attached to the membrane. At this step, the virus hijacks parts of the cellular machinery, such as the ESCRT complex, to drive particle budding and release of an immature virion (11). Protease, another viral protein also packaged into the virion, subsequently cleaves each Gag molecule into three new proteins: matrix (MA), capsid (CA) and nucleocapsid (NC). The matrix proteins remain attached to the membrane, while the capsid proteins self-assemble into a conical shaped core particle enclosing the NC and viral RNA. This maturation process results in fully infectious particles and is shown in Figure 1.1. Reviews on this topic are available (12–14).

Experiments have demonstrated that expressing Gag in the absence of other viral proteins in cells is sufficient to produce and release viral-like particles (VLPs) of the same size as authentic viral particles (15). Since VLPs are noninfectious, assemble by themselves, and have a much simpler composition than the original viruses, they are an ideal model system for studying the assembly of viruses. A simple model demonstrating the main differences between VLPs and authentic virions is shown in Figure 1.2. VLPs and immature viruses contain Gag and a lipid membrane acquired from the host cell. However, VLPs lack other viral proteins and the viral genome, which prevent VLPs from maturing and becoming infectious.

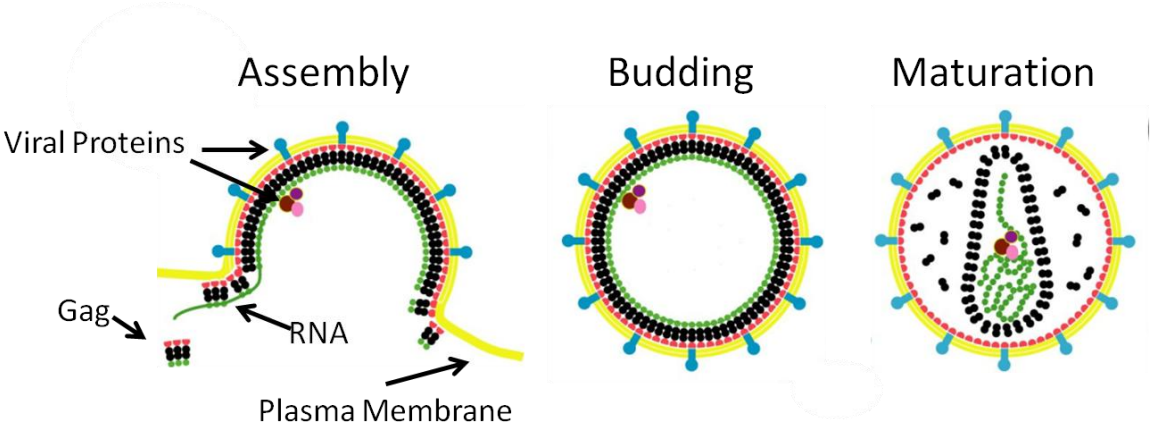


Figure 1.1 Simple model of HIV assembly and maturation process.
During budding, Gag assembles on the membrane along with RNA and other viral proteins. The virus buds off and is released as an immature virion. Finally, the virus undergoes a maturation process to become fully infectious.

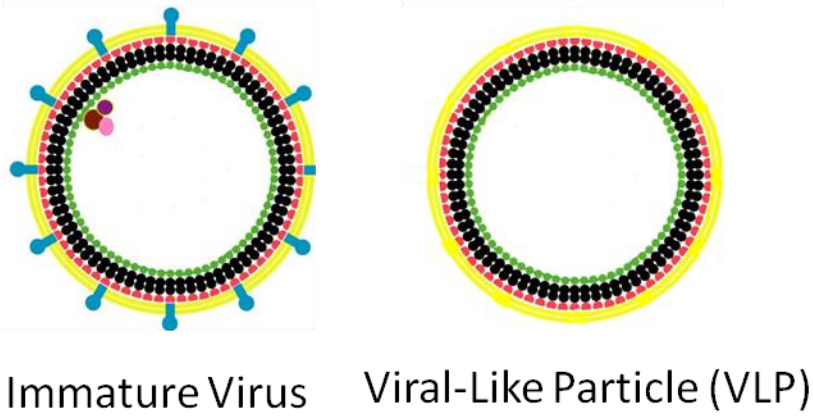


Figure 1.2 Model of immature virus compared to VLP
Both the immature virus and the VLP contain the Gag polyprotein and are enveloped by a lipid membrane acquired from the host cell. However, the VLP is lacking the other viral proteins that allow it to mature and become infectious.

A simple self-assembly model predicts an average stoichiometry of 5000 proteins per VLP (based on an average diameter of 140 nm). This estimation is based on the assumption that the Gag molecules are in molecular contact and form a crystalline lattice that covers the entire surface. It has been shown that Gag proteins engineered to be soluble form self-assembled particles of similar size (~140 nm diameter) as authentic particles (16). These particles can be quantified using cryo-TEM because they lack the enveloping membrane. This study resulted in a copy number of 5000 Gag proteins per VLP (15). In contrast, biochemical and TEM analysis of *in vivo* assembled VLPs and immature virions report numbers of Gag copies ranging from 1000 to 4000 (17). However, these studies are associated with large uncertainties and often require complicated purification steps.

Recent experiments have cast doubt on the self-assembly model of Gag in a closed-shell lattice. Our lab performed a fluorescence characterization of HIV-1 VLPs in which Gag was tagged with a fluorescent protein. FFS allowed us to determine the average copy number of Gag incorporated into the VLPs and the average hydrodynamic radius, which will be explained in more detail later. This study found a constant hydrodynamic radius of ~130 nm and a variable stoichiometry of ~750-2500 Gag proteins per VLP. This fluorescence study is corroborated by cryo-TEM tomography studies that reveal that most immature virions contain an incomplete, hexameric Gag shell, covering approximately 40-60% of the membrane surface (17, 18). The average number of Gag proteins per particle was estimated to be approximately 2,400 ± 700 per immature particle. Together, these results clearly demonstrate that models based on Gag self-assembly into a closed shell are not an accurate reflection of the assembly process. Current knowledge leads to a diagram of VLPs that contain patches of Gag on the inner leaflet of the lipid membrane (Figure 1.3).

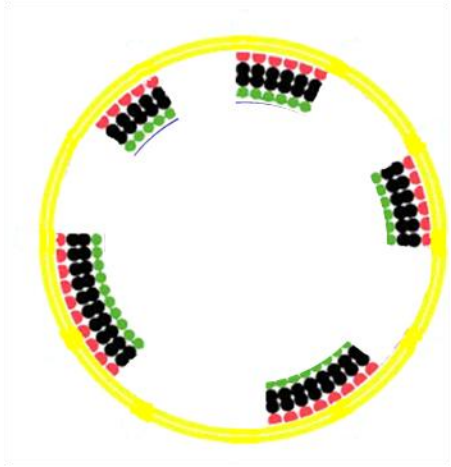


Figure 1.3 Model of VLP containing patches of Gag lattice

The observation of an incomplete Gag lattice, along with the ability to vary the average Gag copy number, demonstrates that the assembly process is complex and not yet fully understood. Other observations also give hints to this complexity. HIV-1 VLPs have been shown to have a bell shaped distribution in size, with VLPs observed from below 100 nm to above 200 nm (19). Current simple models do not explain how particles with such diverse sizes are released, and we do not know if the Gag copy number scales with surface area or if there are additional factors that drive Gag coverage. Additionally, we know that Gag assembly occurs at the membrane, but we also know that the membrane is not homogenous in composition. The membrane is made up mostly of a diverse group of lipids. Spontaneous partitioning of lipids into ordered and disordered phases has been observed in membrane model systems (20). The ordered phase is enriched with cholesterol and saturated lipids. HIV-1 is proposed to assemble at these lipid rafts based on its sensitivity to cellular cholesterol depletion (21). Specifically, HIV-1 assembly is proposed to occur at tetraspanin-enriched microdomains based on microscopy studies (22). It has been hypothesized that oligomers of membrane-bound Gag can stabilize and reorganize these microdomains, and this is an important step in the assembly process (23, 24). This observation along with the size variation observed serve as a reminder of the intricate complexity of the assembly process.

We would like to model viral assembly, but require additional information about the role of cellular factors, individual proteins and their interactions in the assembly process. FFS is a technique that can address some of these questions. In this dissertation, we will not study the assembly process directly, nor will we measure protein interactions in cells; instead, we will focus on studying the in vivo assembled and released VLPs. While studying the released particles does not allow direct viewing of the assembly process, it nevertheless allows to infer about the assembly process by studying how differences in cellular conditions, such a protein concentration, affect the VLPs.

1.4 Brief outline of dissertation

This dissertation focuses on developing and extending FFS techniques to study the protein copy number of viruses. Portions of this dissertation have been previously published. Part of Chapter 3, and most of chapter 5 have been published (25, 26) . The remainder of the work is not yet published.

Chapter 2 introduces FFS and covers the analysis methods used throughout this dissertation, focusing on autocorrelation and PCH. The extension of these techniques to dual-color FFS is also covered. Finally, we demonstrate the application of FFS to VLPs and point out some important details of the analysis.

Chapter 3 discusses the extension of FFS to study the human retrovirus HTLV-1. Studying HTLV-1 and comparing the results to our studies of HIV-1 will provide hints about which features are common among retroviruses and which features are virus specific. Our first results with HTLV-1 VLPs indicated that this retrovirus is significantly smaller than has been reported for other retroviruses. However, the VLPs still contained the same heterogeneous size distribution that was observed for other retroviruses. The copy number approximately scaled with the smaller surface area observed when we compare HTLV-1 VLPs to earlier results we had published with HIV-1 VLPs (9). In the course of our studies, we found that labeling HTLV-1 Gag alters the particle behavior, which was unexpected based on our experience with HIV-1 labeled Gag. We performed experiments that explore the origin of labeling interference and suggest a labeling method that has the potential to overcome the current impasse. While the current labeled HTLV-1 VLPs are not a good model system for studying biology, these particles did alert us to the adaptability of retroviral assembly. We took advantage of this model system to study the flexibility of viral assembly in response to changes in the structure and interactions of Gag proteins. The results of these experiments point to an assembly system capable of successful particle production, even if Gag-Gag and Gag-RNA interactions are significantly altered.

In previous FFS studies of VLPs, we focused on studying a single protein, namely Gag, as this is the only protein required to produce VLPs. However, complex viruses such as HIV-1 contain many other proteins including both viral and cellular proteins. Studying the relationship between the copy number of two proteins within VLPs can provide us with additional information about the assembly process. In Chapter 4, we expand our VLP FFS studies to the characterization of two proteins using dual color labeling. We explore the application of two-dimensional photon counting histogram (2D-PCH) analysis to determine the average copy number of two proteins in VLPs. Additionally, we present a new method called dual-color intensity fraction plots (IFP) that takes advantage of the low concentration of viral samples and allows for quick illumination of the qualitative distribution of protein copy number ratios. This information is useful in determining if 2D-PCH analysis is feasible. We apply these techniques to quantify the copy number of Apobec3G in HIV-1 VLPs, because this cellular protein has been implicated as a viral restriction factor in Vif-deficient HIV-1 viruses and packaging of Apobec3G into the virus is a critical step in its restriction activity.

VLP studies are time consuming due to the low particle concentrations typically encountered. Viral assembly is complicated and requires studying a variety of factors that may influence the assembly process. The low concentrations encountered limit the number of experiments, due to the exorbitant amount of time required to concentrate the sample and collect sufficient events in an FFS measurement for statistical analysis. In Chapter 5, we introduce flow-FFS as one method to overcome this limitation. Flow FFS is a powerful technique which allows us to increase the accessible concentration range of FFS experiments, decrease the measurement time per sample, and measure samples with very high background. To implement our flow studies, we designed and built microfluidic devices compatible with FFS and biological samples. We present an outline of the device fabrication process. Next, we demonstrate the first application combining hydrodynamic flow and PCH analysis to study the protein composition of viruses. These

experiments expand the range of feasible VLP studies and open up a lower concentration range than was previously inaccessible by FFS.

Chapter 2: Methods and FFS Introduction

This chapter will familiarize the reader with the basic background information required for the experiments and analysis techniques used throughout this dissertation. This chapter includes the basic concepts of fluorescence fluctuation spectroscopy (FFS), a description of the experimental setup, and a brief introduction to FFS analysis techniques. The analysis section focuses on autocorrelation and photon counting histogram (PCH), as these are the two analysis techniques used throughout this dissertation. PCH restricted to one color is used to characterize protein homo-complexes. The extension of PCH to study protein hetero-interactions by dual-color detection will be covered, along with model corrections necessary to account for the non-ideal detector affects of deadtime and afterpulsing. Finally, the application of FFS to study viruses will be discussed.

2.1 General overview of fluorescence fluctuation spectroscopy (FFS)

Fluorescence fluctuation spectroscopy is a powerful technique with the unique capability of quantifying protein interactions directly inside the living cell (8). In FFS, proteins of interest are labeled with a fluorescent tag. Every time a labeled protein passes through a tiny optical observation volume, it creates an observable fluctuation in the fluorescence signal. A depiction of this is shown in Figure 2.1.

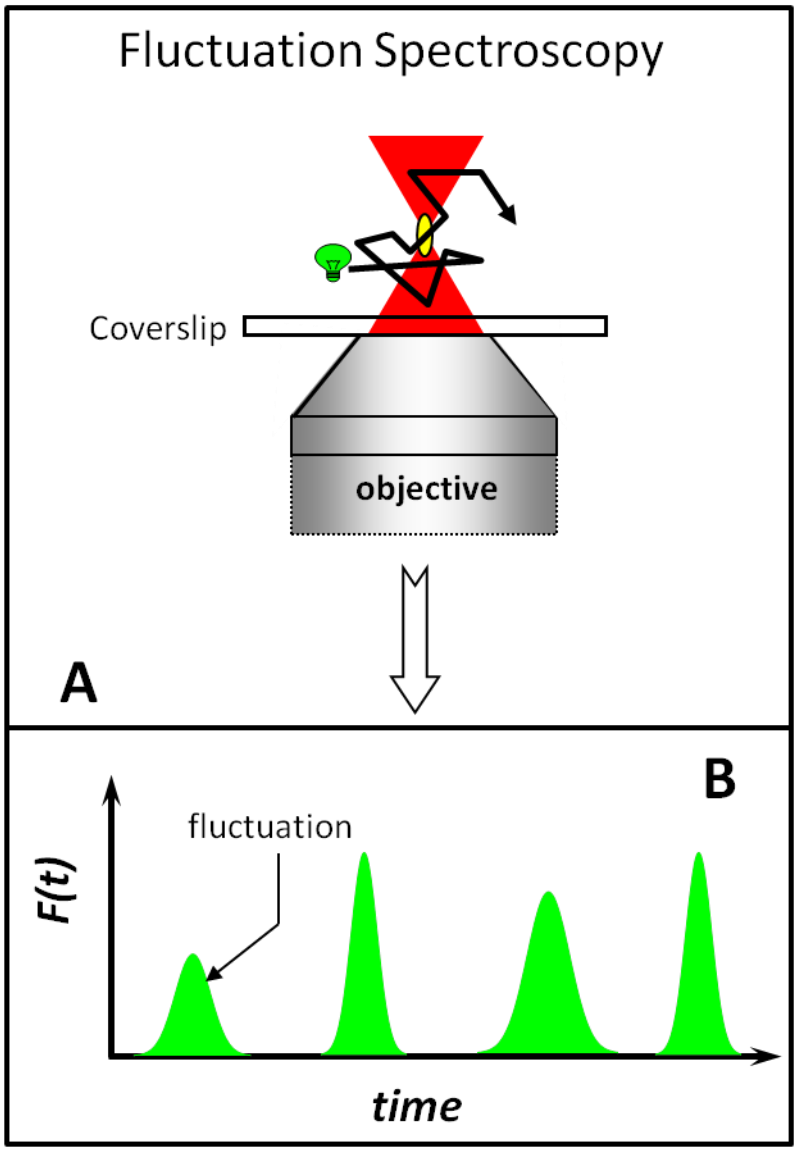


Figure 2.1 Basic concept of FFS
(A) In FFS the fluorophore, often attached to a protein of interest, passes through the observation volume by diffusion or other processes such as flow.
(B) This creates fluctuations in the fluorescence intensity as a function of time that can be analyzed to determine information about the sample like concentration and protein oligomerization state.

The fluctuations of the fluorescence signal contain information about the sample. The burst frequency directly reflects the protein concentration, while the average duration of a burst characterizes the typical time for a protein to cross the observation volume, which can be related to diffusion and the hydrodynamic size of the particle. Finally, the average burst amplitude describes how bright the fluorescence complex is, which measures the protein oligomerization state when proper calibrations are performed. Statistical analysis is required to extract the information contained in these fluctuations. Here we will briefly introduce the two types of analysis commonly used in this dissertation. Fluorescence correlation spectroscopy (FCS) (27–29), or autocorrelation analysis, characterizes the average residence time. If diffusion is the dominant process, the residence time is called the diffusion time, which can be related to the average hydrodynamic size. In this dissertation, we also will encounter FCS in the presence of hydrodynamic flow. In this case, flow is the dominant process and the residence time can be used to calculate the flow velocity (30, 31). A complimentary analysis approach is based on the photon counting histogram (PCH) and related techniques (32, 33). These analysis methods extract the concentration of the fluorescent complexes and their brightness from the distribution of photon counts.

2.2 Molecular brightness

Molecular brightness λ, defined as the detected photon count rate per fluorescent complex, provides a direct quantification of the average oligomeric state of a population of fluorescently tagged proteins (34). For example, EGFP is a monomeric fluorescent protein with a distinct brightness that depends on experimental parameters such as optics, excitation wavelength and detection efficiency. When these experimental parameters are kept constant, the brightness of a monomer EGFP molecule is a distinct repeatable value. A complex consisting of two EGFPs detected with the same setup will have twice the brightness of the monomer. Brightness thus scales with the number of fluorescent proteins in the complex (8). In this dissertation,

we will be studying particles containing hundreds to thousands of proteins, so it is useful to introduce the concept of normalized brightness. The normalized brightness b is the ratio of the complex brightness to the monomer brightness, and it returns the complex's stoichiometry or copy number.

$$b = \frac{\lambda_{complex}}{\lambda_{monomer}} \quad (2.1)$$

Another description of brightness often encountered in experiments is the counts per molecule per sampling time (ε). In the limit of short sampling times, it is related to λ by $\varepsilon = \lambda T$ where T is the sampling time. Brightness analysis has been shown to be a robust parameter, independent of sample concentration, for measuring oligomerization and stoichiometry in living cells (8, 35).

2.3 Two-photon microscopy

FFS requires a small observation volume, typically generated using confocal or two-photon microscopy (36, 37). In confocal microscopy, the small observation volume is created by placing a pinhole directly before the detector. In two-photon microscopy, a small excitation volume is created at the focal point of the objective due to the non-linear nature of two-photon absorption. The experiments in this dissertation are conducted using two-photon excitation. Two-photon excitation is the simultaneous absorption of two photons of half the energy (or twice the wavelength) as compared to the single photon typically absorbed by a fluorescent molecule (38). For example, if 450 nm light excited a fluorophore with one-photon excitation, 900 nm light is required for two-photon excitation. One advantage of two-photon excitation is the greater separation between excitation and emission wavelength. This wide separation ensures that the excitation light and the Raman scattering can be rejected without filtering out any of the fluorescence emission spectrum.

Another advantage of two-photon excitation is the inherent three-dimensional optical sectioning effect (39). Simultaneous absorption of two photons requires a high spatial and temporal photon density. The high spatial density is achieved by using

objectives with high numerical aperture (NA). A laser with approximately 100 fs pulses provides the temporal density. Efficient two-photon excitation occurs only at the focus of the laser beam, obviating the need for a pinhole that is required in confocal microscopy to achieve 3d sections. For n-photon excitation, the total number of photons absorbed per unit time n_{abs} is given by (40)

$$n_{abs}(t) = \int_V \sigma_n C I^n(\vec{r}, t) d\vec{r} \qquad (2.2)$$

where the n-photon cross section is σ_n, C is the fluorophore concentration, $I^n(\vec{r}, t)$is the incident intensity given by $I(\vec{r}, t) = I_0(t) psf_{1\gamma}(\vec{r})$. The temporal description of the incident light is given by $I_0(t)$ and is defined as the intensity at the focal point. The spatial distribution is given by $psf(\vec{r})$. For one-photon excitation the $psf_{1\gamma}(\vec{r})$ is modeled by a Gaussian profile in the radial direction and a Lorentzian in the axial direction. For two-photon excitation, this function is squared and thus we obtain a squared Lorentzian for the axial beam profile, referred to as $psf(\vec{r})$ for the remainder of this dissertation. This means the number of photons absorbed is at a maximum at the focal point of the objective and rapidly decays as a function of distance from the focal volume. Thus, the two-photon excitation volume is strictly limited to near the focal plane of the objective. Another advantage of two-photon excitation is that longer wavelengths scatter less and interact less with components in the cell, causing less damage and lower autofluorescence.

With two-photon excitation, a radial resolution of ~0.4 µm and an axial resolution of ~2.0 µm are achieved with the water objective used throughout this dissertation (41), giving an observation volume of ~0.1 fL. With this observation volume, one molecule in the excitation volume corresponds to a concentration of approximately 10 nM. In two-photon excitation, the observation volume and excitation volume are identical, unlike for one-photon excitation (42). This means photodamage is limited to the focal volume and less autofluorescence is generated, reducing background fluorescence.

2.4 Instrumentation

A schematic of the instrument used for two-photon fluorescence fluctuation experiments is shown in Figure 2.2. The setup is composed of a Zeiss Axiovert 200 microscope (Thornwood, NY) and a Mai Tai Deepsee (Spectra-Physics, Mountain View, CA) laser or a mode-locked Ti:sapphire laser (Tsunami, Spectra-Physics, Mountain View, CA) pumped by an intracavity doubled Nd:YVO4 laser. Both lasers' pulse widths are ~100 fs, and the wavelength is tunable from ~700-1000 nm for the Ti:sapphire laser and from ~650-1040 nm for the Mai Tai. The laser is guided into a beam expander to fill the back aperture of the objective. The expanded laser beam is reflected into the objective by a short pass dichroic mirror and focused into the sample. The fluorescence light is collected by the same objective and passes through the dichroic mirror toward the detectors. We chose a #740DCSPXR dichroic mirror (Chroma, Rockingham, VT), which transmits fluorescence from 400 nm – 750 nm to maximize the collection of red fluorescence while preventing the collection of any scattered laser light. A Zeiss 63x C-Apochromat water immersion objective (NA = 1.2) is used to focus the laser onto the sample. The collected fluorescence passes through two additional band-pass filters (FF01-750/SP-25, Semrock) to eliminate any residual scattered laser light. The detection pathway of the instrument includes an optional dichroic mirror to split the light into two detection channels. The light of each channel is detected by matching avalanche photodiodes (APD) (Perkin Elmer, Waltham Massachusetts). For single-channel FFS experiments, the dichroic mirror is removed and all fluorescence is detected by a single APD. The output of the APD was directly connected to a two-channel data acquisition card (FLEX02, Correlator.com, Bridgewater). The recorded photon counts are stored and later analyzed with programs written for IDL version 6.0 or later (Research System, Inc., Boulder, CO).

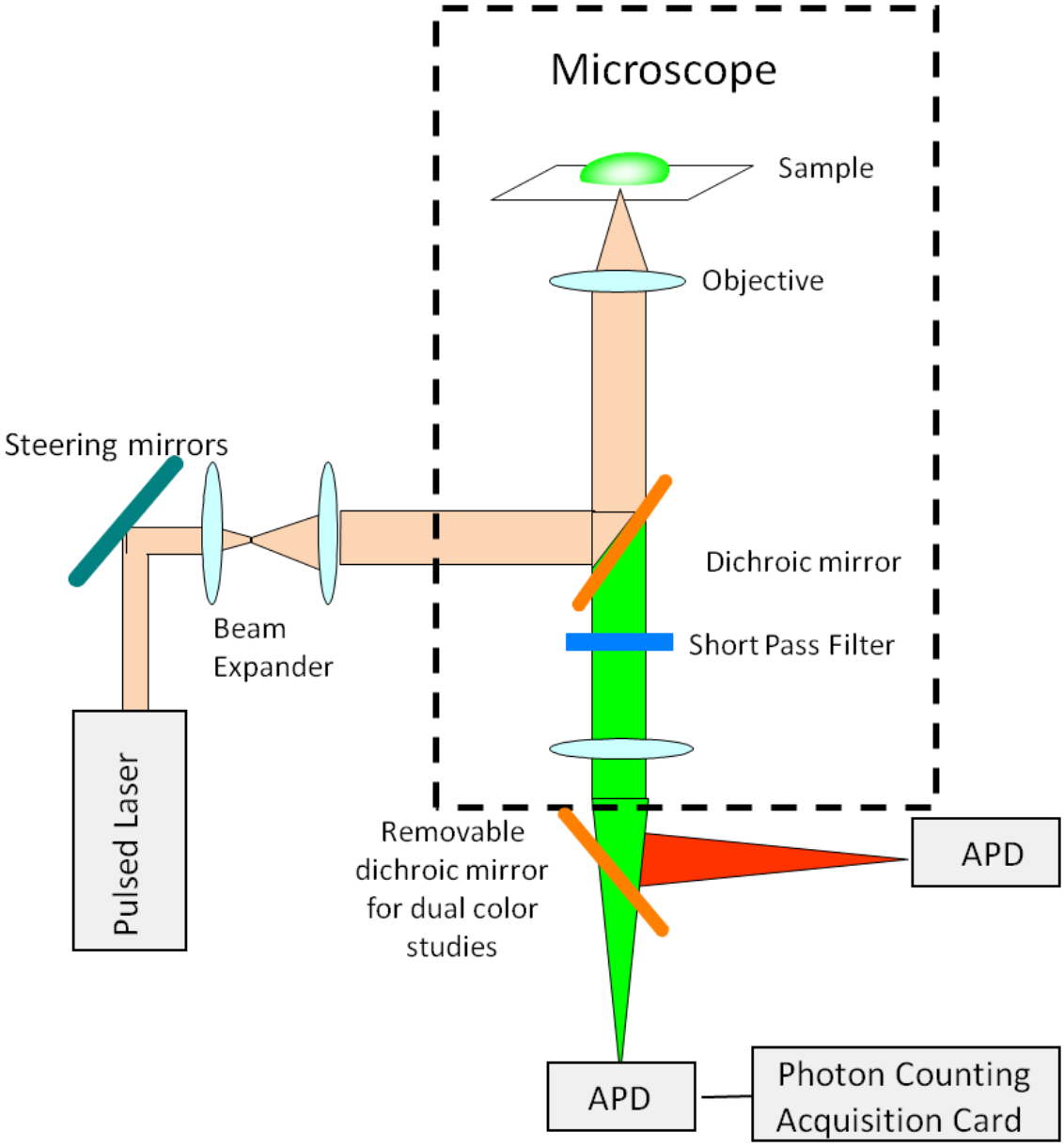


Figure 2.2 Schematic of two-photon instrument
A Ti-Sapphire or MaiTai Deepsee pulsed laser serves as a two-photon excitation source. Steering mirrors direct the laser into the instrument. A beam expander is inserted into the path just before the objective to ensure the back aperture of the objective is filled. The excitation light is reflected into a high NA objective by a dichroic mirror on the microscope. The same objective is used to collect the emitted fluorescence and direct it through a dichroic mirror and onto the APD for detection. For dual-color experiments, a second dichroic splits the emitted light into two colors, which are detected by different APDs. Photon counts are recorded to computer memory and stored for future analysis.

2.5 Analysis Methods

Many statistical tools exist to extract information from the recorded fluorescence fluctuations. FFS includes fluorescence correlation spectroscopy (FCS), moment analysis, photon counting histogram (PCH), and the related fluorescence cumulant analysis (FCA). FCS uses the autocorrelation function to analyze the temporal behavior of the fluctuations. PCH analyzes the amplitude distribution of these same fluctuations. Moment analysis is a computationally easier way to quickly analyze the fluctuation amplitude. In the next section, each method will be briefly introduced, focusing on FCS and PCH, as they are the most widely used methods in this dissertation.

2.5.1 Autocorrelation

The autocorrelation function measures the average duration of the fluorescence fluctuations or in other words, the similarity between a signal at time t and at a later lag time $t + \tau$. Because the theory is widely covered in literature (30), only a general introduction is given. The autocorrelation function is defined as

$$G(\tau) = \frac{\langle \delta F(t) \delta F(t+\tau) \rangle}{\langle F(t) \rangle^2} \tag{2.3}$$

where $F(t)$ is the measured fluorescence intensity at time *t,* and the fluorescent fluctuation $\delta F(t) = F(t) - \langle F(t) \rangle$ is the deviation of $F(t)$ from the mean $\langle F(t) \rangle$. Two properties are primarily determined from the autocorrelation function. The decay of the autocorrelation function characterizes kinetic information, and the fluctuation amplitude $g(0)$ represents the normalized variance of fluorescence intensity. The time zero value of the autocorrelation function is inversely proportional to the average number of molecules in the observation volume N

$$g(0) = \frac{\gamma_2}{N} \tag{2.4}$$

where γ_2 is a shape factor that depends on the geometry of the point spread function (psf). Throughout this dissertation, we used the shape factor (30) of a squared Gaussian-

Lorentzian point spread function, $\gamma_2 = 3/16$. N, the number of molecules inside the observation volume, is related to particle concentration c by

$$N = cV_{psf} \tag{2.5}$$

V_{psf} is the observation volume given by

$$V_{psf} = \int psf(\vec{r})d\vec{r} \tag{2.6}$$

The autocorrelation function can be used to distinguish multiple species if the diffusion time of each individual species is sufficiently different. For multiple non-interacting species, the autocorrelation function is a weighted superposition of all the individual correlation functions

$$G^{(n)}(\tau) = \sum_{i=0}^{n} f_i^2 G_i(\tau) \tag{2.7}$$

where n is the number of species, f_i is the fractional intensity of species i, and $G_i(\tau)$ is the autocorrelation function of species i. In practice, it is difficult to use the autocorrelation function to resolve individual species. The main reason for this is that the diffusion time is proportional to the cube root of molecular weight, assuming globular molecules, which means the diffusion time difference between a monomer and dimer is too small to be resolved by FCS (43). The autocorrelation function is good for distinguishing two species with a large size separation. Other analysis methods are needed for separating species with a small size difference.

For particles undergoing only diffusion (as is typical for solution measurements), the autocorrelation function provides information about the hydrodynamic size of the particles. Throughout this dissertation, a 2D Gaussian single species diffusion model sufficiently describes the data for single species solution measurements. The 2D Gaussian refers to the psf shape assumed in the theoretical model. The 2D Gaussian single species diffusion model is given by (44)

$$g_D(\tau) = g(0)\left(1 + \frac{\tau}{\tau_D}\right)^{-1} \tag{2.8}$$

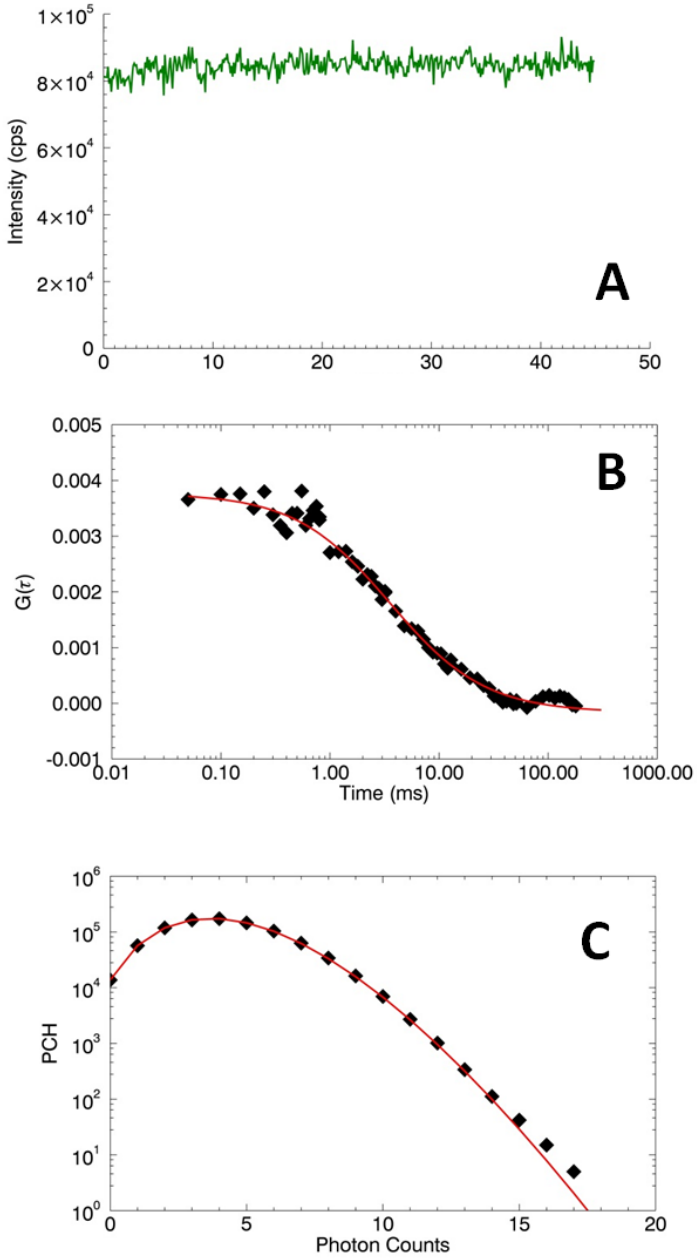


Figure 2.3 FFS Analysis of purified EGFP in solution
(A) The FFS intensity trace for 1 µM purified EGFP measured in a buffer of 50% glycerol and 50% magic buffer at 20 kHz for 45 sec. **(B)** Fit to single species 2D Gaussian diffusion model results in a diffusion time of 3.4 ms with a chi-squared of 0.8 **(C)** Fitting to a single species PCH, with deadtime and afterpulsing corrections, yields a brightness of 10,800 counts per second per molecule and an occupation number of 48 with a reduced chi-squared equal to 1.2.

A fit returns the fluctuation amplitude $g(0)$ and the diffusion time τ_D. Figure 2.3 shows the intensity trace for purified EGFP in solution along with the autocorrelation function fit to equation 2.8. For two-photon excitation the diffusion time is related to the diffusion coefficient D of the particle, and the beam waist $w_0$of the excitation light by (30)

$$\tau_d = \frac{w_0^2}{8D} \tag{2.9}$$

The hydrodynamic radius (R) can then be determined from the Stokes-Einstein relation,

$$D = \frac{k_B T}{6\pi\eta R} \tag{2.10}$$

where $k_B$is the Boltzmann constant, T is the temperature, and η is the viscosity.

A potential complication of FCS analysis is that the theory assumes a point particle approximation. Large particles such as VLPs are on the order of 100 nm. Using a point-particle approximation for particles of this size introduces an error of around 5% compared to finite size analysis (45). Because this error is within the statistical uncertainty of our measurement, we use point particle approximation when fitting VLP FCS data in this dissertation.

2.5.2 Moment analysis

Moment analysis is another way to extract brightness information from fluctuation data. The first moment ($\langle F \rangle$) and second moment ($\langle \Delta F^2 \rangle$) can be used to determine the average number of molecules within the sample volume (46, 47). Autocorrelation and moment analysis of brightness are of course related, because $g(0)$ is determined by the first two moments of the fluorescence intensity. The brightness can be calculated directly from moment analysis by,

$$\varepsilon = \frac{\langle \Delta F^2 \rangle - \langle F \rangle}{\gamma_2 \langle F \rangle} \tag{2.11}$$

Although moment analysis is not the technique we use to determine brightness or concentration in most experiments, we do take advantage of this simple formula to

quickly determine the quality of the data set. Using moment analysis across short sub-segments of data allows us to initially estimate the number of VLPs present and retake the data if any large particles become trapped in the laser beam.

2.5.3 Photon Counting Histogram (PCH) and Fluorescence Intensity Distribution Analysis (FIDA)

The photon counting histogram (PCH) and fluorescent-intensity distribution analysis (FIDA) are equivalent methods that were independently developed as a way to resolve protein multi-species by brightness (32, 33). Both methods analyze the distribution of photon counts but with slightly different mathematical approaches. The PCH is a histogram of photon counts that captures the amplitude distribution of fluctuations. PCH exploits the probability distribution of photon counts to distinguish species by difference in molecular brightness. Thus, it is a complementary method to the autocorrelation function, which describes the temporal behavior of fluctuations and can separate a mixture of species based on differences in diffusion.

Here a brief overview of the theory underlying PCH is presented. The low light levels typical of FCS experiments require the use of photon counting techniques. In the detection process, a charge separation is generated after the absorption of a photon by the detector (APD for the experiments in this dissertation). This step in the detection process involves the interaction of light from classical light sources. It is sufficient to treat the electromagnetic field classically and the atomic system quantum mechanically, resulting in the semiclassical description of the detection process (48). Mandel's formula describes the semiclassical case of photon counting statistics (49) by

$$p(k) = \int_0^\infty \frac{(\eta_W W)^k e^{-\eta_W W}}{k!} p(W) dW = \int_0^\infty Poi(k; \eta_W W) p(W) dW \qquad (2.12)$$

The probability $p(k)$ of observing k events during a sampling time T depends on the statistical properties of the light reaching the detector, and η_W the detection efficiency of the detector. For a given time-varying intensity $F(t)$ incident on a detector,

the integrated light intensity $W(t)$ reaching the detector during sampling time T is given by (50)

$$W(t) = \int_{t}^{t+T} F(t)dt \tag{2.13}$$

If we assume the sampling time T is short enough, then the fluctuations in W track the intensity fluctuations. The integrated intensity is then

$$W(t) = I(t)T \tag{2.14}$$

Additionally, the probability distribution of the integrated intensity is proportional to the probability distribution of the intensity, $p(W)dW = p(I)dI$. Equation 2.12 in the limit of short sampling times can then be written as

$$p(k) = \int_{0}^{\infty} Poi(k;\eta I)p(I)dI = \langle Poi(k;\eta I)\rangle \tag{2.15}$$

where $\eta = \eta_W T$. Looking at equation 2.15, the angular brackets represent the average of the Poissonian shot noise contribution over the intensity distribution. Light of constant intensity incident on a photodetector will yield a Poissonian photon count distribution. This shot noise is inherent in the photoelectric detection process (48). The second source of fluctuations is from fluctuations in the light intensity reaching the detector. Any fluctuations in the light intensity will cause the photon counting histogram $p(k)$ to broaden compared to a Poisson distribution. The photon counting histogram is then characterized as having a variance that is greater than its mean value, which is classified as super-Poissonian (32). In general terms then, PCH analysis is based on analyzing the super-Poissonian nature of FFS photon counting histograms.

The probability distribution of the intensity depends on the psf mentioned in section 2.5.1. A typical model psf for two photon FFS experiments is the squared Gaussian-Lorentzian as mentioned earlier. If we assume the sampling time is short enough, then the photon counting probability for a single fluorescent particle diffusion in volume V can be calculated,

$$p^1(k,\varepsilon) = \frac{1}{V}\int_V Poi(k,\varepsilon psf(\overrightarrow{r}))\overrightarrow{dr} \tag{2.16}$$

ε is the molecular brightness measured in counts per sampling time described earlier. For N particles, the PCH is given by N repeated convolutions of the PCH for a single particle.

$$p^N(k;\varepsilon) = \underbrace{(p^1 \otimes .. \otimes p^1)}_{N-times}(k,\varepsilon) \tag{2.17}$$

In typical FFS systems, molecules can be exchanged between the observation volume and the reservoir. In this case, the probability to have N particles in the subvolume is also given by a Poisson distribution $p(N) = Poi(N, \overline{N})$. Finally, the PCH for N particles is given by the average of $p(k,\varepsilon)$, the individual probability functions weighted by the probability of observing N particles.

$$p(k;\varepsilon,\overline{N}) = \sum_{N=0}^{\infty} p^N(k;\varepsilon)p(N) \tag{2.18}$$

and the average photon counts are given by

$$\bar{k} = \varepsilon\overline{N} \tag{2.19}$$

The PCH of a single species is thus characterized by only two parameters, ε and $\overline{N}$, the brightness and average number of molecules. In Figure 2.3c, we perform PCH analysis on the same EGFP solution used for FCS analysis. Finally, if more than one species is present, the PCH of two species is the convolution of each individual species' PCH

$$p(k;\overline{N}_1,\overline{N}_2,\varepsilon_1,\varepsilon_2) = p(k;\overline{N}_1,\varepsilon_1) \otimes p(k;\overline{N}_2,\varepsilon_2) \tag{2.20}$$

For more than two independent species, each individual PCH must be convoluted successively.

2.6 Dual-color FFS experiments

In dual-color or dual-channel FFS, a dichroic mirror is inserted into the emission path to separate the emission of spectrally distinct fluorophores into two separate channels. Dual-color FFS has extended brightness analysis to allow the study of hetero-protein interactions (35). In a typical setup for studying two interacting proteins, one protein is labeled with EGFP, which has an emission spectrum with a maximum at 507

nm, and the other protein is tagged with mCherry, which has an emission spectrum centered around 610 nm (51). The emission dichroic then splits this 'green' and 'red' fluorescence into two separate channels.

Dual-channel FFS looks for coincident fluctuations in the two detectors to identify hetero-interactions. There are three main ways to analyze dual-color experiments: cross-correlation analysis (52); dual-color PCH (53)/two-dimensional FIDA (54); and hetero-species partition analysis (HSP) (55), which will not be discussed in this dissertation. Cross-correlation analysis was first developed to overcome autocorrelation analysis' shortcomings in resolving different species with similar diffusion coefficients (56). As was mentioned earlier, monomers and dimers have diffusion coefficients that differ by only a factor of 1.26, which is not sufficient to be resolved with traditional autocorrelation analysis (43). To overcome this problem the interacting proteins are labeled with two separate colors. If the proteins interact, then simultaneous fluctuations are visible in both channels.

Cross-correlation analysis correlates the signals in the two channels as a function of time. The theory and applications are discussed in the literature (57). Cross-correlation analysis has been used to study diffusion (56), enzyme kinetics (58), and protein-protein interactions (59). Ideally, if the proteins do not interact, the cross-correlation should be zero. However, in reality, spectral crosstalk between the channels exists. For instance, using the two-color setup described earlier, proteins carrying an mCherry label will have basically all of their signal passed through to the red channel (above 580 nm), but a protein carrying an EGFP label will have 92% of its signal reflected into the green detector with 8% of the signal passing through to the red detector. Without crosstalk, the time zero value of the cross-correlation function is inversely proportional to the number of molecules carrying both fluorophores. If there is crosstalk, a careful calibration is needed to accurately return the number of molecules carrying both labels (60).

Dual-color also increases the ability of PCH analysis to separate mixtures of interacting proteins (53). As in cross-correlation analysis, crosstalk limits the sensitivity of dual-color PCH analysis to resolve species. However, this dissertation deals with studying large, bright protein complexes with high background and no crosstalk from the red tag (mCherry). Dual-color PCH has high enough sensitivity to resolve brightness species and has the added advantage of a nice visual representation of the sample composition. The dual-color PCH is given by the probability of observing k_A counts in channel A and k_B counts in channel B. As an example, consider a dimer of protein XY with X carrying a red tag and Y carrying a green tag. If this dimer passes through the observation volume, the two-dimensional histogram has a higher probability of observing photons simultaneously in both channels compared to if the proteins did not associate. Figure 2.4a shows the theoretical 3D representation of a dimer with one mCherry-labeled subunit and one EGFP-labeled subunit. The effects of FRET are not included. For simplicity, dual color PCH is often graphed in a two-dimensional representation. This two-dimensional plot facilitates viewing the residuals to determine the quality of the fit. The two-dimensional representation of the same theoretical data shown in 3D is seen in Figure 2.4b.

The theory for dual-color PCH is an extension of the theory presented earlier in this chapter. It is fully developed in previous papers published by our lab (53). The instrumentation for dual-color FFS is nearly identical to the instrumentation previously shown for single color FFS. We simply insert a dichroic mirror in the emission path as shown in Figure 2.2. The choice of the exact properties of the dichroic mirror is very important as it determines the amount of crosstalk and hence the ability of each analysis technique to resolve species. The most common dichroic used in this dissertation is for proteins labeled with mCherry and EGFP. For this setup, the best dichroic is a 580 nm split with and additional band-pass filter (51). The emission spectra of EGFP and mCherry along with the transmission curve of the dichroic used is shown in Figure 2.5.

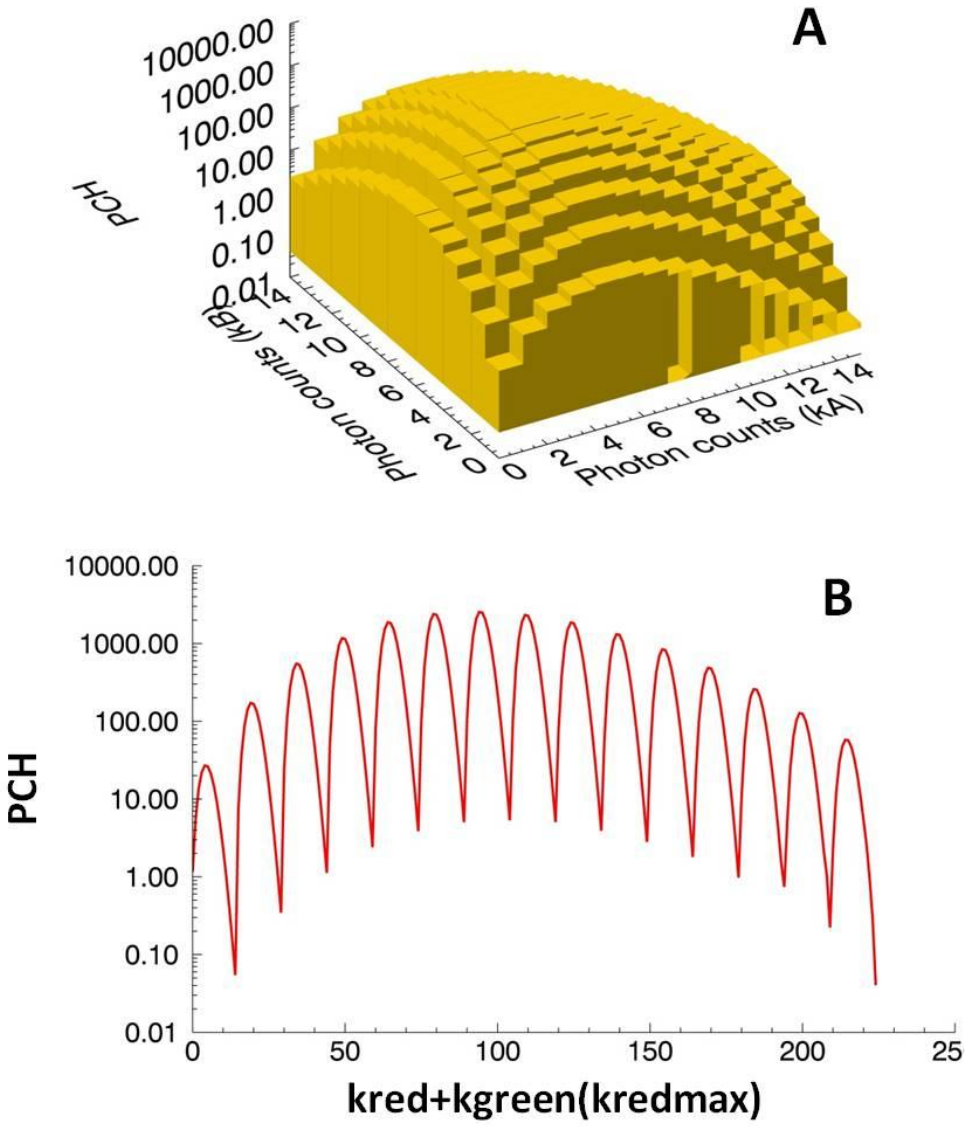


Figure 2.4 Theoretical 2D-PCH of a heterodimer of EGFP and mCherry
(A) Three dimensional representation of 2D-PCH for a dimer of EGFP and mCherry. The mCherry signal falls entirely into the red channel, while the EGFP signal falls mostly in the green channel with about 9% of the signal falling into the red channel. EGFP is approximately twice as bright as mCherry using the setup in Figure 2.2. Thus, the 2D-PCH of a dimer of mCherry and EGFP consists of a high probability of observing photons in both channels, with a slightly higher probability of observing more green counts. **(B)** Two-dimensional representation of data shown in panel (A). In the two-dimensional representation, each row of the 3D histogram is graphed sequentially, starting with $k_B = 0$. The two-dimensional representation is used to facilitate comparison of the model to its fit. kred is the counts in the red channel. kgreen is the counts in the green channel, and kredmax represents the maximum count of the red channel.

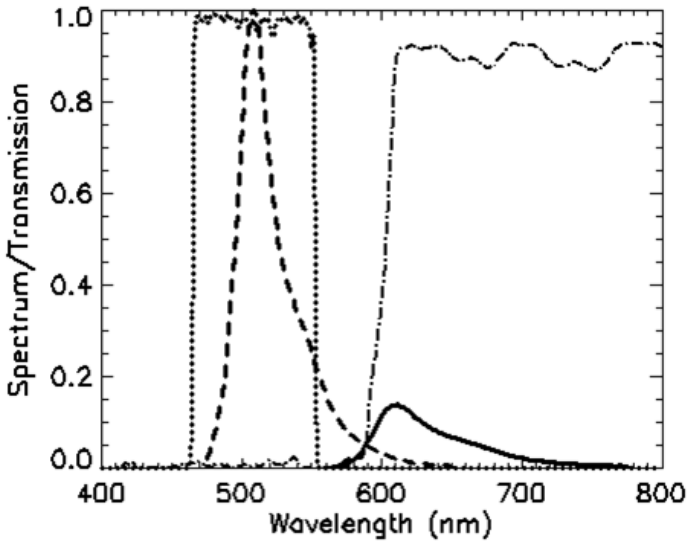


Figure 2.5 Emission spectrum of mCherry and EGFP and dichroic transmission curves used in dual-color FFS experiments

Fluorescence emission spectra of EGFP (dash) and mCherry (solid) are shown with the transmission curve of the dichroic mirror (dot-dash) used to separate the fluorescence into two channels overplotted. To eliminate mCherry fluorescence reflected by the dichroic mirror an additional 84 nm wide filter (dot) centered at 510 nm was used. The emission spectra are normalized such that their integrated areas are proportional to their brightness.

2.7 FRET in dual-color studies

FRET is the radiation-less transfer of energy through dipole-dipole interaction of an excited donor fluorophore with an acceptor fluorophore (61). In this interaction, the donor does not emit a photon (i.e. fluoresce). FRET only occurs if the emission spectrum of the donor overlaps with the absorption spectrum of the acceptor and its efficiency depends on the distance r between the donor and acceptor and scales as r^{-6}, which makes it an attractive 'ruler' for small distances. Literature reviews of FRET are available (62, 63). The full derivation of the effect of FRET on the donor and acceptor brightness can be found in (53). If FRET occurs, the donor fluorescence is quenched and the acceptor fluorescence is enhanced. The fractional reduction in the donor fluorescent intensity is defined as the FRET efficiency E

$$E = 1 - \frac{\langle I_d^F \rangle}{\langle I_d \rangle} \tag{2.21}$$

where $\langle I_d \rangle$ is the fluorescence intensity of the donor without FRET and $\langle I_d^F \rangle$ is the donor intensity in the presence of FRET. Brightness and intensity are proportional to each other (Equation 2.19), so the effect of FRET on the donor brightness is

$$\varepsilon_d^F = (1 - E)\varepsilon_d \tag{2.22}$$

where ε_d is the normal donor brightness. In the presence of FRET, the acceptor fluorescence intensity increases because it is excited both by the laser and by the energy transfer from the donor molecule. The increase in acceptor fluorescence involves the donor's cross-section because donor excitation is necessary to transfer energy to the acceptor, and the acceptors' cross-section. The acceptor brightness in the presence of FRET is

$$\varepsilon_a^F = \left(1 + \frac{\sigma_d}{\sigma_a} E\right) \varepsilon_a \tag{2.23}$$

The effect of FRET on a complex containing EGFP and mCherry will be an increase in the intensity in the red channel, an increase in brightness of mCherry, and a corresponding decrease in intensity in the green channel and decrease in brightness of

EGFP. If these effects are not measured and accounted for, quantitative studies cannot be done accurately. One way to quantify FRET is to measure the donor's lifetime in the presence and absence of FRET, and calculate the efficiency (62)

$$E = 1 - \frac{\langle \tau_d^F \rangle}{\langle \tau_d \rangle} \tag{2.24}$$

where $\langle \tau_d^F \rangle$ is the donor's average lifetime in the presence of FRET.

2.8 Deadtime and afterpulsing

Photodetectors used in photon counting experiments are never ideal. All brightness experiments that employ photodetectors, such as APDs, must contend with two primary artifacts, deadtime and afterpulsing (64, 65). Deadtime is a fixed period after the collection of a photon during which the detector cannot register any other events (Figure 2.6a). This temporary "blindness" leads to a decrease in the number of photons detected and is of particular concern in high-count rate experiments. Although VLP experiments typically have low overall count rates, each VLP event produces short periods of high count rates. Deadtime has the effect of narrowing the photon count distribution, which leads to an artificially low calculated brightness. At high-count rates, more events are lost since there is a greater probability of additional photons arriving while the detector is blind to new events. Our lab has APDs of deadtimes ranging from 25-50 ns. In high-count rate and VLP experiments, the effect of deadtime on brightness is not negligible and will result in lower brightness if there is no correction applied.

Afterpulses are spurious events that occur with a small probability after the detection of an actual photon (Figure 2.6b). This leads to an increase in the number of apparent photons detected. Afterpulses will have the effect of broadening the photon count distribution because the afterpulses artificially increase the counts in the higher channels, which leads to an artificially high brightness. For newer APD detectors the afterpulse probability can be sufficiently low such that its effect on many experiments is negligible. For instance, APDs used in our lab typically have an afterpulse probability of 0.2% to 0.5%.

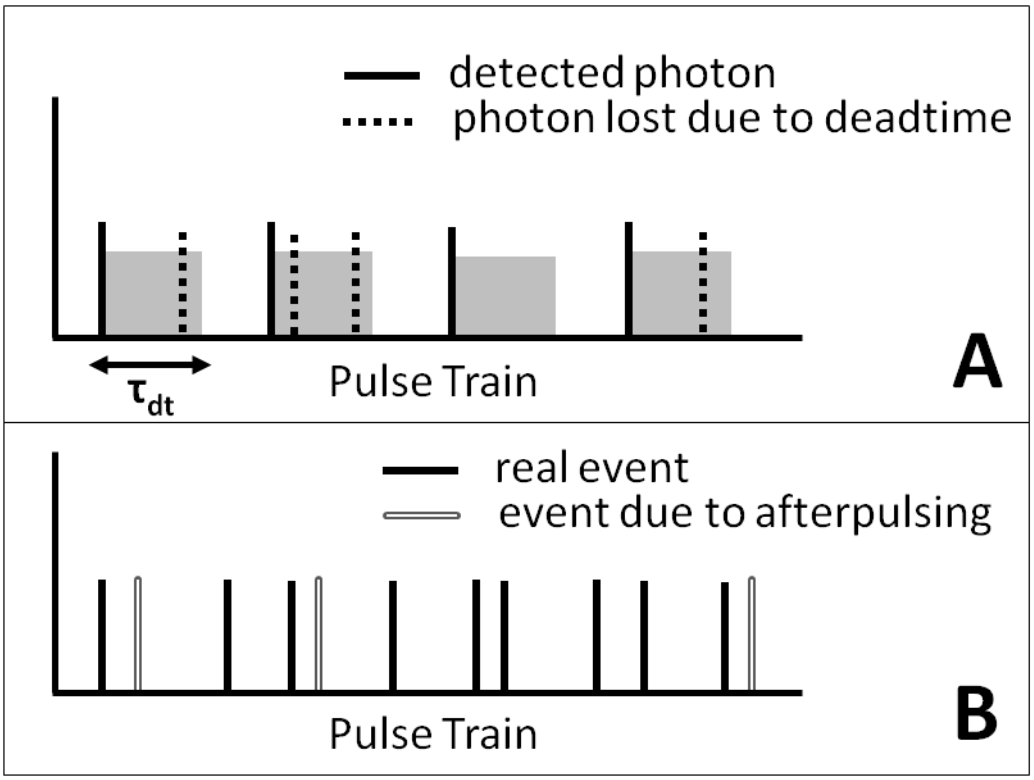


Figure 2.6 Conceptual picture of the non-ideal detector effects
(A) Deadtime. After a photon is counted by the detector, a deadtime of duration τ_{dt} occurs. No photons hitting the detector at this time are detected. **(B)** Afterpulses are false pulses generated with a probability q after the true detection event.

While detectors often include factory specifications for the deadtime and afterpulsing, for the most accurate experiments these numbers need to be independently verified. For characterization we use a very highly concentrated fluorescent dye solution to imitate a constant fluorescence intensity light source (64). It is useful to employ Mandel's Q-parameter (66) to characterize the width of the photon count distribution. For a constant intensity light source, the photon count distribution will be Poissonian, which results in $Q = 0$. However, the non-ideal detector effects modify the Q value of a constant intensity light source,

$$Q = -2\tau_{dt}I + 2q \tag{2.25}$$

where I is the intensity, τ_{dt} is the deadtime, and q is the afterpulsing probability (67). A graph of Q (obtained using moment analysis) versus collected photon intensity is produced by changing the intensity of the excitation light using a neutral density filter. This graph is then fit to the above equation, from which the deadtime and afterpulsing can be determined. Note that this equation is a first order approximation. The linear relationship between Q and I breaks down at sufficiently high intensities. Additionally, this equation measures the total deadtime in the system. If the deadtime is significantly higher than the manufacturer's specifications, this may be due to deadtime of the card, or software or hardware issues. This is another reason why it is important to measure the deadtime and afterpulsing for the experimental setup used. For reference, Table 2.1 lists the deadtime and afterpulsing parameters for the instruments used in this dissertation.

Methods to correct photon count distributions and moments have been developed and tested extensively (64, 68). If these corrections are not applied, the brightness of a fluorescent sample varies with the concentration of the fluorophore as the total intensity varies. To test these corrections, a dilution study of Alexa 488 in a 50% glycerol solution was performed. The data was analyzed using moment analysis without deadtime or afterpulsing corrections. The same data was also analyzed with PCH and TIFCA with the corrections. TIFCA (68, 69), an additional brightness analysis

	Instrument 1	Instrument 2
Afterpulse Probability	$q_A = 0.0026$ $q_B = 0.0043$	$q_A = 0.0054$ $q_B = 0.0027$
Deadtime	$\tau_{dtA} = 26$ ns $\tau_{dtB} = 26$ ns	$\tau_{dtA} = 50$ ns $\tau_{dtB} = 50$ ns

Table 2.1 Deadtime and afterpulse probabilities for our detectors
The subscript A refers to the APD detecting the through light. The subscript B refers to the APD detecting the reflected light.

method that is more robust than PCH, is shown here for completeness. As shown in Figure 2.7, with corrections the entire range of concentrations studied returns the same brightness. This is important for VLP experiments because it shows first that the deadtime and afterpulsing have been properly determined for this experimental setup, and second that the wide range of peak intensities detected during VLP experiments can be accurately fit using a deadtime and afterpulsing corrected PCH model.

2.9 Statistical analysis of rare bright particles such as VLPs

In typical FFS experiments, the fluctuations from individual proteins or protein complexes passing through the observation volume is not directly observed because the characteristic concentration ranges studied mean more than one particle is always present in the observation volume. An example of this type of intensity trace is shown in Figure 2.3a, which is data for purified EGFP in solution. When measuring a sample of VLPs or other rare bright particles the fluctuation from each VLP can be individually seen because of the low concentrations typically encountered. Figure 2.8a contains a characteristic intensity trace for a VLP FFS experiment. In this figure, you can see that normally no VLP is observed in the observation volume. This means that anywhere from 30% to 99% of the photons collected are from background. The analysis method chosen to analyze VLPs must be able to cleanly separate the overwhelming background from the signal of interest.

It is important to note that by looking at the intensity trace it is not possible to determine the brightness of individual VLPs. This is because of the inhomogeneous excitation profile. A low intensity spike may correspond to a dim VLP passing near the center of the observation volume, or a bright VLP passing near the edge of the observation volume. An FFS analysis method must be used so that the distribution of VLPs and the observation volume profile are both accounted for.

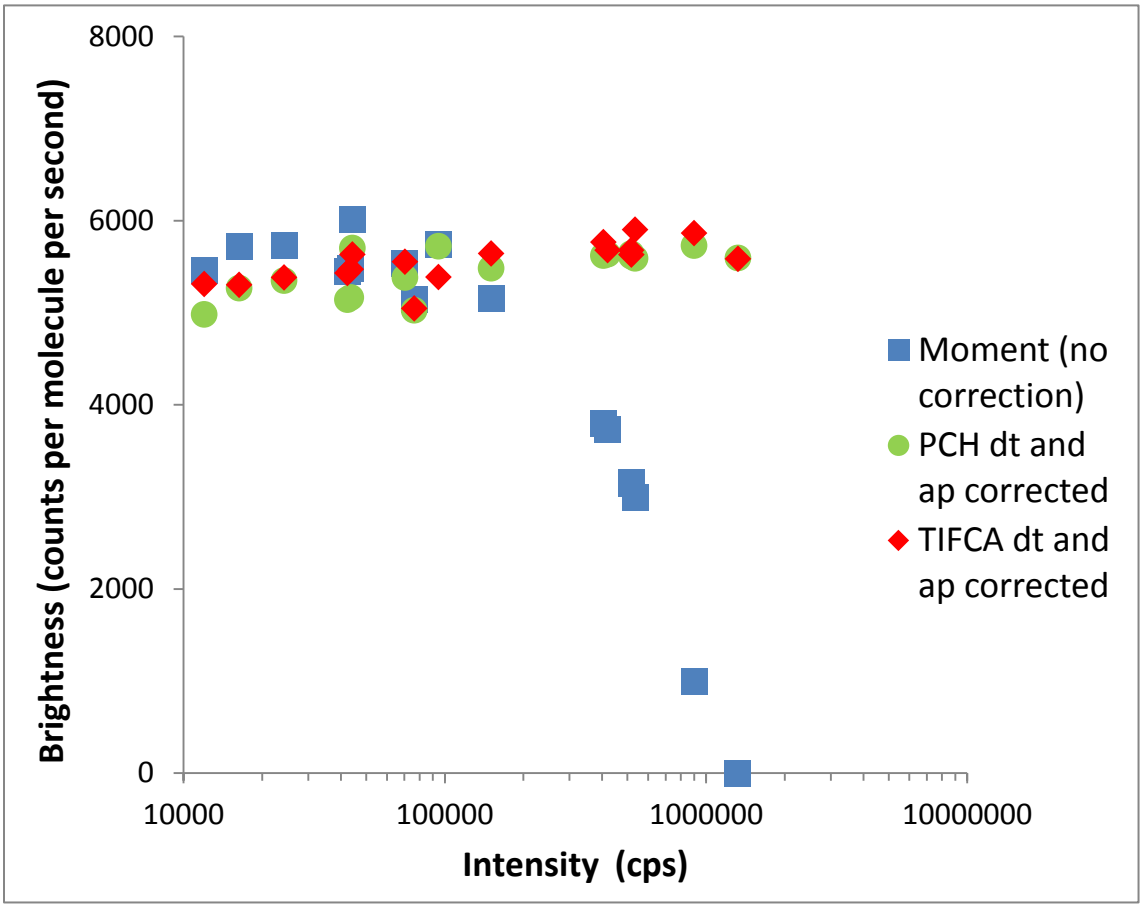


Figure 2.7 Effects of deadtime and afterpulsing in a dye dilution study
The fluorescent dye Alexa 488 was dissolved in a buffer solution of 50% glycerol and 50% HEPES buffer. The solution was successively diluted into the same buffer. The data was then analyzed using three different methods, simple moment analysis without deadtime or afterpulsing corrections, PCH with deadtime and afterpulsing corrections, and TIFCA with deadtime and afterpulsing corrections. At high-count rates, the uncorrected moment analysis brightness is reduced, while with the corrections in PCH and TIFCA, a constant brightness is achieved across the entire dilution experiment. The deadtime of the detector is 26 ns; the afterpulsing probability of the detector is 0.26%

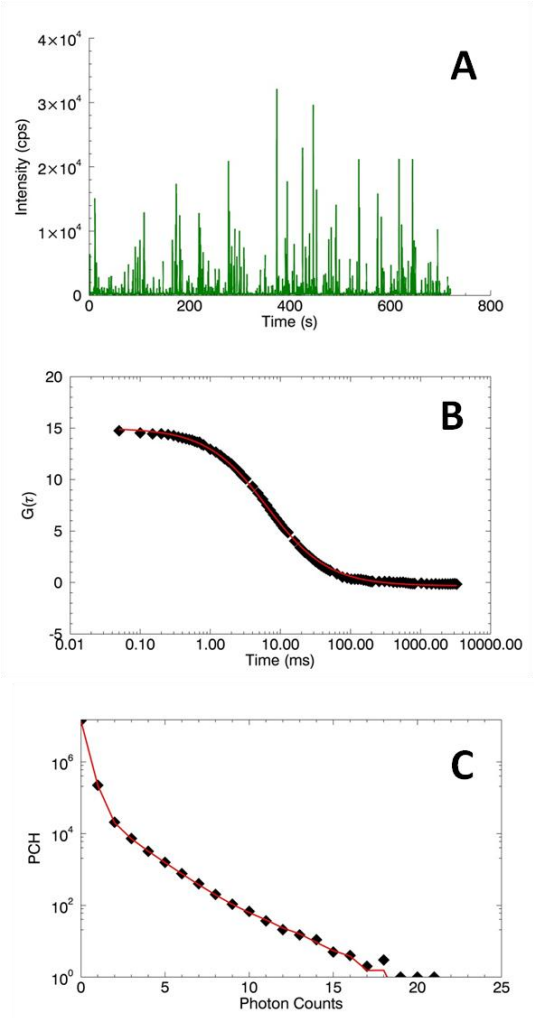


Figure 2.8 FFS analysis of VLP sample
(A) The FFS intensity trace for EYFP-labeled HIV-1 VLPs produced by COS cells. In the intensity trace, each spike represents the passing of a single VLP through the observation volume. For much of the acquisition time, there is no VLP in the observation volume so only background counts are collected. For this sample the concentration was approximately 50 pM, and 30% of the total counts were from background. **(B)** Fitting to a single species 2D Gaussian diffusion model results in a diffusion time of 6.4 ms with a chi-squared of 1.3. This corresponds to an average hydrodynamic radius of 125 nm. **(C)** Fit to a three species PCH with deadtime and afterpulsing corrections. The first species represents the dim high concentration background species. This is visible in the steep section of the PCH with counts less than 3. The second two species approximate the brightness distribution of the VLP sample. In this case species 1 has a concentration of 52 pM and a copy number of 670 Gag, while species 2 has a concentration of 3 pM and copy number of 1760 Gag. Combining the two results in an apparent brightness of 810 Gag per VLP. Note that the PCH fit did not require any additional knowledge or calibration of the background, but the background fit returned does match the background intensity observed.

Figure 2.8b shows the autocorrelation analysis of the VLP data shown in Figure 2.8a. Using the measured residence time and known beam waist (see Chapter 5 for explanation of beam waist calibration), this sample has an average VLP diameter of 125 nm. It is important to note that the autocorrelation function can almost always be fit assuming there is a single diffusing species for the VLP samples presented in this dissertation. This does not indicate that all VLPs in the sample have a single radius; instead, it indicates that this distribution can be approximated as a single radius species. In fact, cryo-TEM studies of HIV-1 particles reveal an approximately Gaussian size-distribution with an average diameter of ~140 nm (68). Ensemble averaging of FFS collapses this distribution to a single diffusion value as illustrated in Figure 2.9a.

We will rely on PCH to determine brightness of VLPs throughout this dissertation, as it is more reliable and applicable than the fluctuation amplitude of the autocorrelation function. PCH is particularly well suited for VLP analysis for a number of reasons. PCH has the unique ability to separate the background from the VLP signal as easily visualized by looking at the photon counting histogram (26, 51). A detailed description of this process is found in Chapter 5. A graph of the PCH provides a good visual representation of this signal separation and is shown in Figure 2.8c. The steep part of the curve at low counts (< 3 counts) corresponds to the background. VLPs are rarely detected, but when they are detected, they produce high photon count rates. The section of the PCH curve with the shallow slope at high counts captures the VLP species. A typical PCH fit of a VLP sample identifies the background species together with two VLP species that differ in their brightness (9). As was true for the autocorrelation function analysis, the two species PCH fit does not indicate that VLPs exist in only two distinct brightness species. This two species fit represents an approximation of the underlying brightness distribution. The exact shape of this distribution is not currently known. Figure 2.9 illustrates the concept with a hypothetical size and copy number distribution.

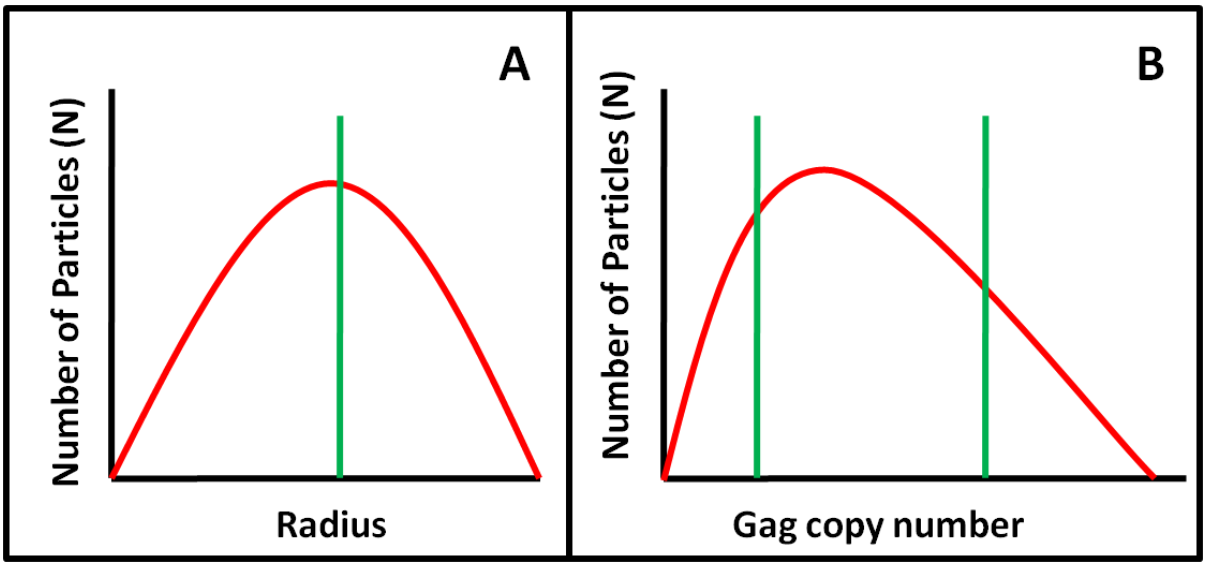


Figure 2.9 Hypothetical size and Gag copy number distribution for HIV-1 VLP
(A) The distribution of VLP sizes is represented by the red line. Autocorrelation fitting approximates this distribution as a single average radius represented by the green line. **(B)** The Gag copy number distribution of HIV-1 VLPs is unknown, but the red line represents a hypothetical distribution. PCH fitting approximates this distribution by returning two brightness species represented by the two green lines.

Chapter 3: FFS studies of HTLV-1 Assembly

3.1 Overview

This chapter presents experiments exploring HTLV-1 VLPs. Additional background will be provided in section 3.2. These studies were done in collaboration with the Mansky and Wei group here at the University of Minnesota. Dr. Iwen Grigsby was the lead virologist from the Mansky lab working on this project. She carried out many of the biological experiments and sample preparations including initial plasmid design, sample preparation for cryo-transmission electron microscopy (cryo-TEM), and biological control experiments. Jose Maldonado from the Mansky lab assisted her in sample preparation for cryo-TEM for the VLPs described in section 3.3. Dr. Wei Zhang did the cryo-TEM experiments described throughout this chapter. Other members of Dr. Mansky's and Dr. Mueller's labs helped design and implement experiments. I would especially like to thank Dr. Yan Chen for thoughtful discussions and plasmid design. While much of the work described in section 3.2 has been previously published (25), the rest of this chapter contains unpublished results.

This chapter is divided into three main sections. Section 3.2 describes our studies of VLPs produced from human codon-optimized HTLV-1 Gag-EYFP DNA. Section 3.3 chronicles experiments that uncovered differences in HTLV-1 VLPs produced with and without labels and briefly mentions additional experiments seeking the source of these differences. More detailed results of many of these experiments are found in Appendix B. Finally, in section 3.4 we briefly explore the robustness of VLP production to perturbations in the Gag molecule.

3.2 Viral-like particles from HTLV-1 Gag-EYFP

3.2.1 Introduction

HTLV-1 was discovered over three decades ago, making it the first human retrovirus discovered. There are approximately 15-20 million people infected with HTLV-

1 worldwide (70). HTLV-1 infection has been implicated in a number of disorders including adult T cell leukemia/lymphoma (ATLL) and HTLV-1 associated myelopathy/tropical paraparesis (HAM/TSP). Despite its impact on human health, HTLV-1 is one of the least understood retroviruses in terms of its replication, assembly, and particle release.

HTLV-1 has been difficult to study in the tissue culture system due in part to the fragility of the HTLV-1 proviral sequence and low levels of viral replication. To overcome these limitations, our collaborators developed a human codon-optimized HTLV-1 Gag gene fused to an EYFP reporter to create a robust level of HTLV Gag expression, which made FFS and cryo-TEM studies feasible. The human codon-optimized HTLV-1 Gag-EYFP expression construct used throughout this chapter is shown in Figure 3.1. This gene was tested to ensure expression, similar cellular localization as wild type HTLV-1, and VLP production (25).

Although HTLV-1 is a retrovirus with significant human health impacts, little is known about its assembly process. Despite the limited amino acid sequence homology between different retroviruses, the tertiary structure of their Gag domains exhibit high similarity (71–73). Much more is known about the retrovirus HIV-1, which has been extensively studied. Therefore, structural and assembly mechanisms of HIV-1 are generally used as a reference model for other retroviruses, including HTLV-1. Recently, our lab investigated the Gag copy number of HIV-1 VLP using FFS (9). We are particularly interested in determining if HTLV-1 VLPs exhibit similar copy number characteristics as HIV-1 VLPs.

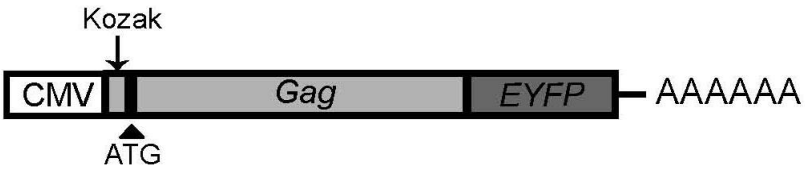


Figure 3.1 HTLV-1 Gag expression construct
The HTLV-1 Gag gene was codon-optimized to improve cellular expression and VLP production. A CMV promoter was used. A Kozak consensus sequence was inserted upstream of the ATG start codon. The EYFP label was inserted in-frame prior to the Gag gene stop codon.

3.2.2 Materials and Methods

3.2.2.1 Sample preparation

As previously discussed, the expression of Gag alone is sufficient to form VLPs. For measuring the copy number of Gag using FFS, 293T cells were cotransfected with constructs expressing pEYFP-N3 HTLV-1 and vesicular stomatitis virus G (VSV-G) protein at a 10:1 ratio, respectively. VSV-G is an envelope protein from the vesicular stomatitis virus that is commonly used in retroviral studies (74). Without VSV-G, VLPs of irregular shapes were observed, suggesting that this envelope protein helps stabilize the VLP membrane. Authentic HTLV-1 does include a similar envelope protein, but it is expressed less efficiently. Transfection was carried out using Genjet transfection reagent (SignaGen Labs, MD) with a ratio of 5 µg of DNA to 15 µL of reagent for a 25 cm^2 plate of cells. Genjet was found to produce more uniform and higher expression levels and resulted in higher virus production levels than TransFectin (BioRad, CA), which was previously used. Genjet transfection resulted in anywhere from a 5-20 fold increase in VLP concentration when directly compared with TransFectin. One hour before transfection, the cell medium was replaced with fresh phenol red free DMEM medium (Corning, VA) with 10% serum (JRScientific Inc, Woodland CA). Approximately 24 hours post-transfection, the medium containing the VLPs was removed from the cells. The medium was then spun at 14000 rpm for 2 min to remove large cell debris. Next, the VLPs were pelleted down by spinning the medium at 20000 rpm for 2 hours at 4˚C (Beckman model J2-MI, Brea, CA). The medium was removed and the VLPs were resuspended in 1x STE buffer (10 mM Tris-Cl, pH 7.4, 100 mM NaCl, 1 mM EDTA) at 4˚C overnight. Before FFS measurements, the sample was spun at 14000 rpm for 2 min to remove any large debris still present. Typically, the VLPs were concentrated by a factor of 10 for measurement. After each measurement, the samples were stored at 4˚C for future studies.

Other methods of VLP purification have previously been used. The most common alternative method is described here in brief. For a typical HIV-1 experiment,

32 hours after transfection the cell medium was collected and spun at 14000 rpm for 2 min to eliminate cell debris. A portion of non-concentrated VLP in cell supernatant was saved for future studies. The remaining solution was concentrated by a factor of 10 into Dulbecco's PBS (Biowhittaker, Walkersville, MD) using a Centricon filter (Millipore, Billerica, MA) at 14000 rpm. If the sample was not to be measured within 24 hours, it was stored at -80˚C and thawed at 4˚C when needed. This method worked quite well for HIV-1 with the exception of the occasional faulty filter. However, occasionally with HTLV-1, irregularly shaped HTLV-1 VLPs and fragments of VLPs were observed using cryo-TEM if this method was used. Additionally, a high percentage of HTLV-1 VLPs were lost during the concentration step. It is possible that HTLV-1 VLPs are more susceptible to bursting in PBS medium and deformations during the filtering process. Finally, although the filter method is quicker, it results in approximately 5 times the fluorescent background in the final sample compared with the pellet method. This is due to the relatively large amount of serum that remains in the final sample.

3.2.2.2 Experimental Setup

The experiments were carried out on a modified two-photon microscope as described in Chapter 2. Samples were measured with two-photon excitation at either 905 or 1000 nm. The VLPs were measured with a Zeiss 63x C-Apochromat water immersion objective (NA = 1.2) with excitation powers ranging from 0.1-0.3 mW as measured after the objective. FFS data were acquired at a sampling frequency of 20 kHz, stored, and subsequently analyzed with programs written for IDL version 6.4 (RSI, Boulder, CO). Measurements were performed by loading 200 µL of solution into an 8-well Nunc Lab-Tek Chamber Slide (Thermo Fisher Scientific, Pittsburgh PA) and focusing the two-photon spot 10 µm into the solution. Measurements ranged from 10-60 min depending on sample concentration.

3.2.2.3 Brightness Calibration

Determining the Gag copy number relies on an accurate calibration of the monomer brightness. A direct calibration measurement is difficult, because VLPs are

measured at low powers to avoid saturating the detector, to ensure that particles are not trapped by the beam, and to avoid bleaching these slow moving particles. At low powers, it is not possible to accurately measure the brightness of a monomeric fluorescent protein due to the poor signal-to-noise ratio. An alternative method of extrapolating the brightness of a monomer at these low powers was developed based on the relationship between power, intensity, brightness, and concentration. For each color label studied, a calibration of the concentration and brightness was performed at a sufficiently high power to measure the brightness, but well below the power at which bleaching and saturation were observed. A typical calibration power is 0.4 mW. Next, a power study was performed over a range of powers, including the power at which the VLPs are measured. The power range was small enough that changes in excitation volume are negligible. To extract the monomer brightness at low excitation powers, the linear relationship between brightness and intensity for constant concentrations was used,

$$I = \lambda N \tag{3.1}$$

where I is the fluorescent intensity excluding background. The monomer brightness at low powers was then calculated by

$$\lambda = \frac{(I_{total} - I_{back})}{N} \tag{3.2}$$

where I_{total} is the total intensity measured at the low power, I_{back} is the background intensity measured during a calibration (mostly stray room light and APD dark counts), and N is the occupation number determined by the calibration at a higher power. For completeness, this method was also compared to a method that uses the relationship between brightness and power squared using the following formula

$$\lambda_{low} = \frac{\lambda_{high} P_{low}^2}{P_{high}^2} \tag{3.3}$$

where λ_{low} is the brightness at the low power, λ_{high} is the brightness at the higher calibration power, P_{low} is the excitation light power at the low power setting and P_{high} is the higher calibration power. Both methods produced the same calculated brightness

within 5% uncertainty. As measurements of power at the sample can be difficult, Equation 3.2 was used to calculate the brightness for the remainder of this dissertation.

In addition to determining the monomer brightness at the desired VLP measurement power, the power study provides a quick way to ensure that no undesired effects such as bleaching, saturation, or high background are present. Data from a typical power study for all three fluorescent protein labels used is shown in Figure 3.2. To easily visualize any adverse effects, the calculated brightness (Equation 3.2) divided by the power squared is plotted as a function of power. Ideally, since brightness scales as power squared for two-photon excitation, this plot should result in a horizontal line in the absence of experimental artifacts. For each fluorescent protein, a constant value is observed between ~0.2 mW to 1.0 mW. Above 1.0 mW, the power normalized brightness starts to decrease, indicating that bleaching and/or saturation occurs. Note that the background has been properly accounted for, as there is no increase in brightness seen at lower powers. If background is not properly measured and subtracted as shown in Equation 3.2, the calculated brightness will be artificially high where background accounts for a significant percentage of the detected photon counts. The background was determined by measuring the intensity of the buffer under identical conditions. In general, this method allows for determining the monomer brightness across a wide range of powers, and provides a quick, qualitative check of the effect of artifacts in the data.

Lastly, to determine the range of powers free from bleaching, particle trapping, and other artifacts, power studies were also performed on the VLP samples. From 0.15 mW to 0.5 mW, the brightness scales with power squared and the diffusion time is constant, indicating that this is a safe range for measuring VLP samples. Above 0.5 mW, VLPs and cellular debris were often trapped, so powers > 0.5 mW were avoided.

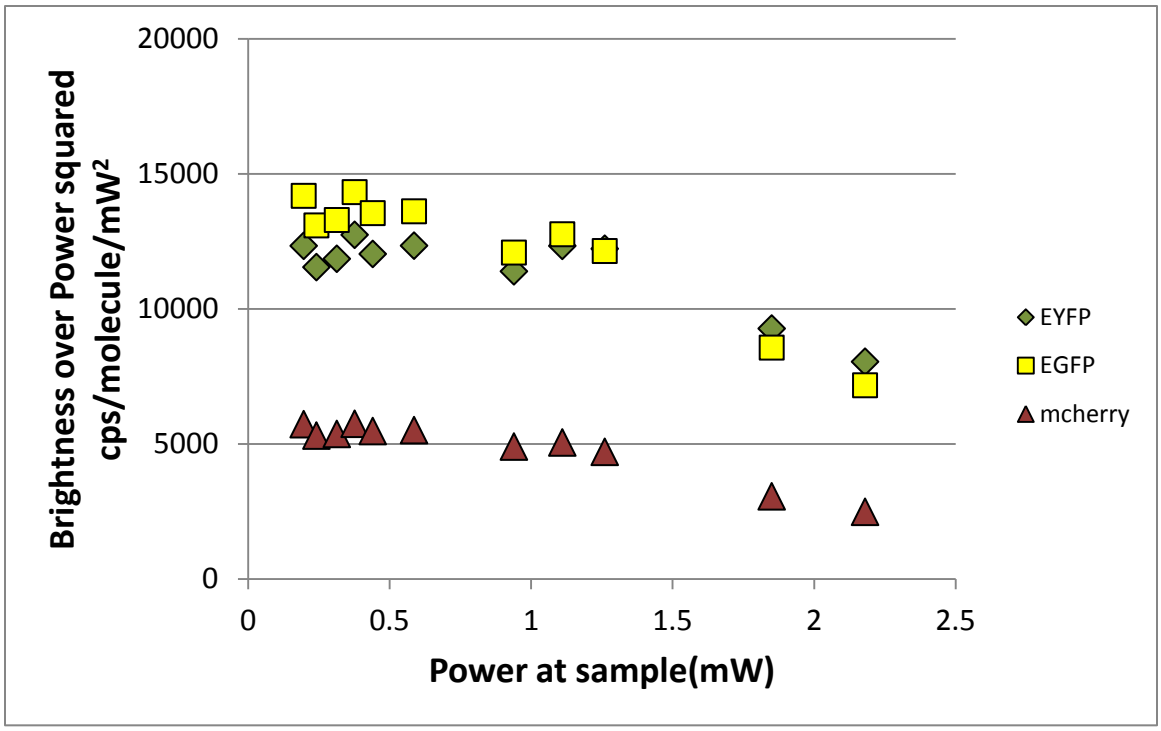


Figure 3.2 Power study of fluorescent proteins
Each fluorescent protein was purified from *E. coli* and measured in a solution of 50% glycerol, 50% HEPES buffer at 1000 nm over a range of powers. The intensity at each power was recorded and Equation 3.2 was used to calculate the brightness for each measurement. From approximately 0.1 mW to 1.0 mW, the brightness scales with power squared indicating the absence of artifacts.

3.2.3 Results of HTLV-1 Gag-EYFP studies

3.2.3.1 HTLV-1 Gag EYFP FFS Results

The raw intensity trace for a typical EYFP HTLV-1 VLP measurement is shown in Figure 3.3a. HTLV-1 VLP samples are relatively low in concentration (≤ 200 pM). Therefore, like HIV-1 VLP samples, each spike represents a single VLP passing through the observation volume. This raw data was analyzed using autocorrelation analysis (as described in Chapter 2) to determine the average hydrodynamic radius. A single species diffusion model accurately describes the correlation function and returns a diffusion time of 5.2 ms. Using 50 nm diameter fluorescent spheres as a calibration, this corresponds to a hydrodynamic diameter of 74 nm. This measurement is performed for 5 independently prepared samples and results in an average size of 74 ± 4 nm.

These measurements indicated that HTLV-1 VLPs are much smaller than HIV-1 VLPs. An earlier FFS study found HIV-1 VLPs produced in COS-1 cells have an average hydrodynamic diameter of 130 nm (9). For a more direct comparison and to avoid differences due to variations introduced by cell line differences, measurements are also performed on HIV-1 VLPs produced in 293T cells, harvested in parallel with HTLV-1 VLPs. From our observations, the average diameter for these HIV-1 VLPs was 115 ± 10 nm, meaning the average HTLV-1 VLP with Gag-EYFP was approximately 65% the diameter of the average HIV-1 VLP. This large size difference was not expected. From previous results using TEM with staining, the average diameter of HTLV-1 was reported between 95-110 nm (75). The origin of this size difference will be addressed later in this chapter.

Next, the data was analyzed with PCH to determine the Gag copy number and particle concentration. Since a model assuming a single VLP brightness species leads to poor fits of the experimental PCH data (reduced $\chi^2 \geq 10$), a second VLP brightness species was introduced into the fit model and was sufficient to reproduce the experimental data. A fit of the photon counting histogram to a 2-species model (reduced $\chi^2 = 2.1$) is shown in Figure 3.3b. The presence of two brightness species indicates brightness heterogeneity in the VLP sample.

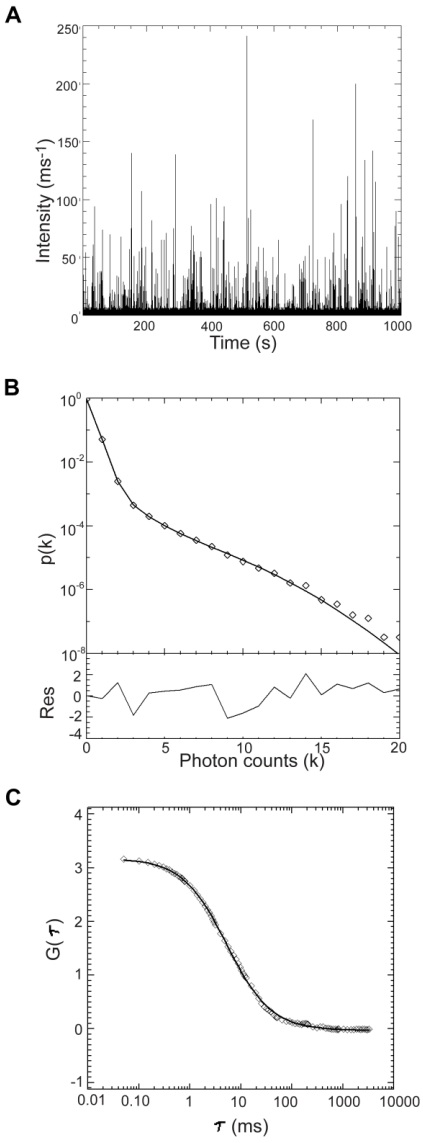


Figure 3.3 Fluorescence fluctuation spectroscopy analysis of HTLV-1 Gag-EYFP based VLPs **(A)** The fluorescence intensity trace shows discrete peaks corresponding to individual VLPs diffusing through the observation volume **(B)** Experimental photon counting histogram fit to a two species model with background identifies the concentration and Gag copy number of the VLPs. The presence of two species indicates the existence of heterogeneity in the Gag copy number of VLPs as previously observed with HIV-1 VLPs. The first VLP species has a copy number of 270 and a concentration of 20.5 pM. The second VLP species, which is brighter than the first, has a copy number of 800 and a concentration of 6.5 pM. A weighted average of the two species, calculated using equation 3.4, leads to an apparent Gag copy number of 530 Gag per VLP **(C)** A fit (solid line) of the autocorrelation function (diamonds) to a 2D Gaussian, single species diffusion only model identifies an average hydrodynamic diameter of 74 nm for the VLPs

The Gag copy number for each species was calculated using the monomer EYFP calibration. Species one had a concentration of 20.5 pM and a Gag copy number of 270, while species two had a concentration of 6.5 pM and a Gag copy number of 800. For a quick comparison between samples, it is illustrative to briefly ignore the brightness heterogeneity by calculating the average Gag copy number b_{avg} of the VLP sample according to (8),

$$b_{avg} = \frac{b_1^2 \cdot N_1 + b_2^2 \cdot N_2}{b_1 \cdot N_1 + b_2 \cdot N_2} \tag{3.4}$$

This results in an average Gag copy number for this sample of 520. We repeated this measurement multiple times for each sample and measured multiple independent samples. The average copy number was stable at 510 ± 50 copies per VLP. The copy number of each species was also relatively stable at 300 ± 60 and 880 ± 100 copies. Although the total concentration varied from sample to sample depending on cell concentration and conditions, the relative ratio of concentrations between the high and low copies was constant, with 19 ± 7% of the VLPs having the higher copy number. These results are summarized in Figure 3.4.

Recall that ~5000 Gag is required to completely fill the surface of a 140 nm HIV-1 VLP (15). Thus, a maximum Gag copy number of ~1300 is expected for the smaller (75 nm) HTLV-1 VLP, assuming that Gag proteins occupy a comparable surface area at each membrane. The observation of an average Gag copy number of 520 indicates that if a Gag lattice at the membrane of a HTLV-1 VLP is present, it covers less than half the available surface area on average. This means that the Gag coverage of HIV-1 and HTLV-1 VLPs is similar. This idea will be further explored later in this section.

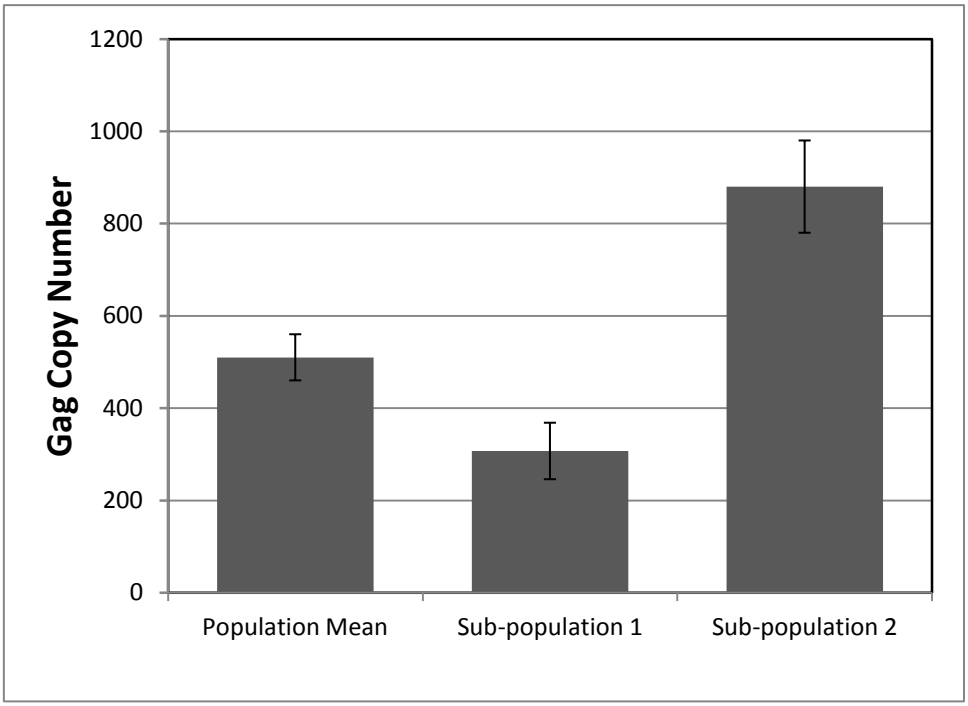


Figure 3.4 Gag copy number of HTLV-1 Gag-EYFP based VLPs
The Gag copy number was determined by FFS analysis of 5 independent VLP samples. The apparent copy number per VLP is shown together with the corresponding copy number of the two subpopulations identified by FFS analysis. The error bars represent the standard deviation based on multiple measurements.

3.2.4 Cryo-TEM HTLV-1 Gag EYFP

To independently confirm the small size of HTLV-1 VLPs reported by FFS measurements and to gain additional structural insights, the HTLV-1 Gag-EYFP VLP morphology was examined using cryo-TEM. Cryo-TEM is a powerful tool that provides insight into the structure of a broad range of samples in their native frozen-hydrated state. It allows the study of structures of specimens such as retroviruses that do not form well-diffracted crystals or that have molecular complexes that are too large for NMR spectroscopy. The cryo-TEM data presented in this chapter was collected by Dr. Wei Zhang with assistance from Dr. Iwen Grigsby. The samples for cryo-TEM measurements were prepared as described above with the addition of a purification step during which the VLPs were subjected to a 10-40% linear sucrose gradient. Figure 3.5a shows a representative VLP sample. With VSV-G present, the majority of the VLPs were observed to be spherical. The size of each individual particle was measured and a histogram of sizes was built as shown in Figure 3.5b. Similar to other retroviruses, a broad range of sizes was observed around a central peak. Particle sizes ranged from 30-237 nm, with over 25% of the total population between 70-80 nm. The average size of the particles was 71 ± 20 nm, in agreement with the FFS measurements.

The cryo-TEM data was then used to examine the VLP radial profile to gain hints of the structure. The majority of the VLPs did not have an inner structure that could be attributed to a Gag lattice or Gag structure. The occasional structure observed was less obvious than previously reported for HIV-1 VLPs (76). An average radial density profile for a number of particles with diameters between 70-80 nm was determined and is shown in Figure 3.6. From this profile, the MA domain is not distinguishable from the membrane. The density peaks observed between 300 Å and 400 Å are the outer and inner lipid membranes, and no additional density is observed near this region. Additionally, the density profile at the center of the particle is relatively flat, which suggests a homogenous inner density. These observations show that EYFP labeled HTLV-1 VLPs have a distinct structure compared to HIV-1 VLPs.

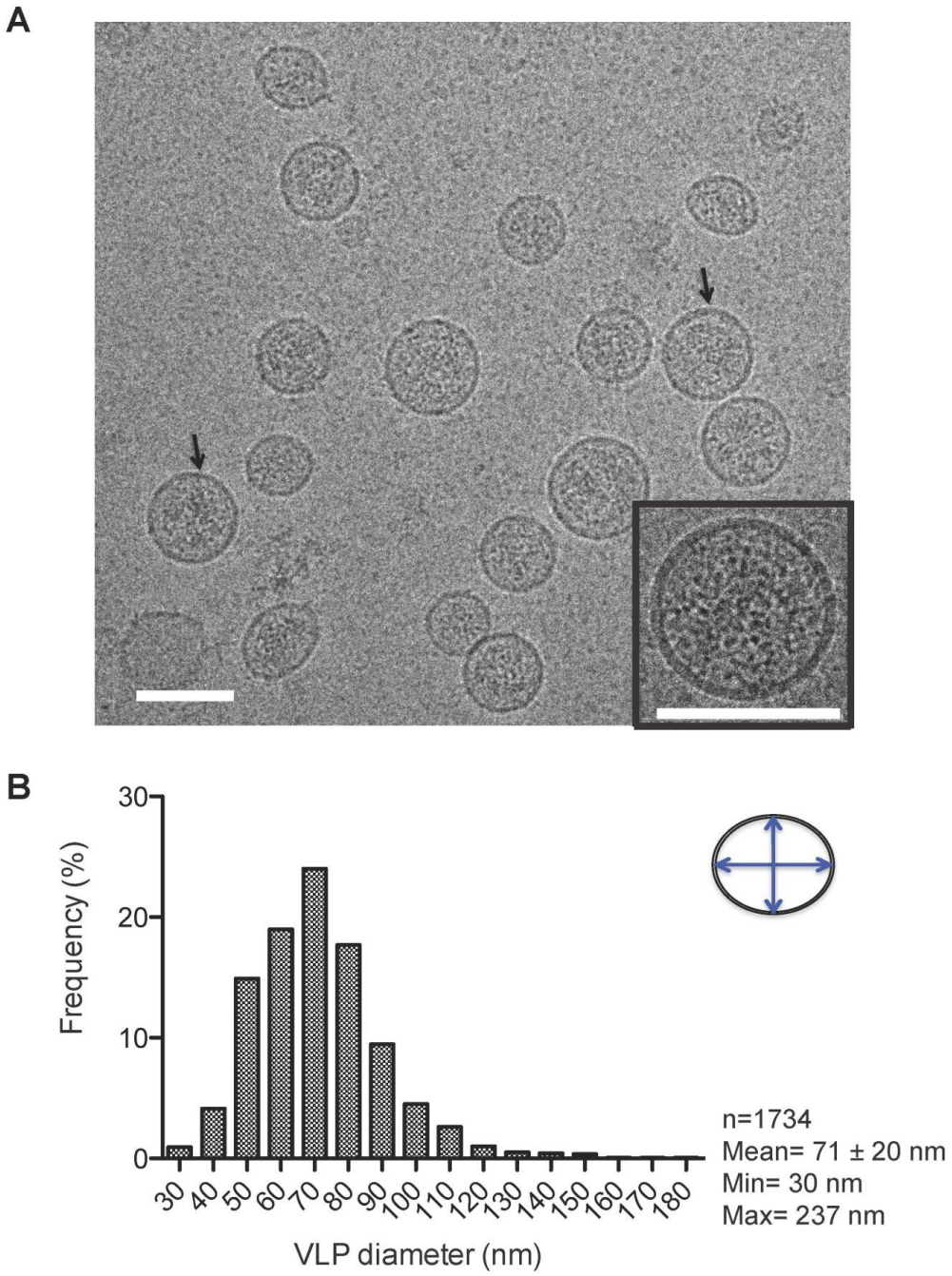


Figure 3.5 Cryo-TEM analysis of HTLV-1 Gag-based VLPs
(A) Cryo-TEM images of VLPs produced from 293T cells transfected with HTLV-1 Gag-EYFP. Examples of the partially occupied inner electron density are indicated by the arrows. The inset shows a magnified view of a representative VLP. Scale bar is 100 nm. **(B)** Distribution of VLP diameters. The particle diameter was determined by averaging the longest and shortest measurements as indicated in the diagram at the top right corner using the ImageJ software. 1735 VLPs were examined (mean = 71 ± 20 nm).

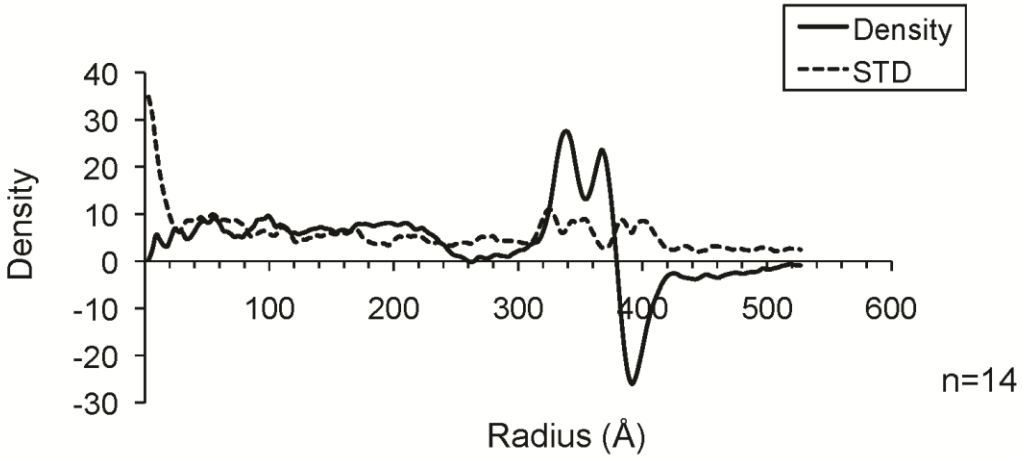


Figure 3.6 Radial density profile of the HTLV-1 EYFP VLPs
The solid line represents the average density measured. The dashed line indicates the standard deviation ($n = 14$). The two density peaks between 300-400 Å are the inner and outer lipid membranes.

3.2.5 Discussion of HTLV-1 Gag-EYFP results

HTLV-1 was the first human retrovirus to be discovered (77, 78), but little was known about its morphology due to the fragile nature of the proviral sequences and limited levels of viral gene expression. We developed a model system to study HTLV-1 Gag VLP production and morphology based on a human codon-optimized HTLV-1 Gag EYFP construct. This construct is stable, leads to high expression of HTLV-1 Gag and produces a sufficient number of VLPs to perform FFS and cryo-TEM measurements. Initial results from our system showed that the HTLV-1 Gag-EYFP VLPs produced had a diameter smaller than previously reported for HTLV-1 using TEM. The diameter from FFS studies was found to be 74 ± 4 nm, and 71 ± 20 nm with cryo-TEM. Cryo-TEM revealed the size is distributed over a wide range similar to results for other retroviruses (76).

The average Gag copy number of HTLV-1 VLPs was found to be approximately 500 using FFS. A low copy number species and a high copy number species were necessary to describe the PCH. Like previous HIV-1 experiments, this does not indicate the presence of two VLP species with distinct copy numbers; instead, it most likely points to a VLP sample with a range of copy numbers as one would expect based on the size variations observed by cryo-TEM. This concept was discussed in Chapter 2 and a model is shown in Figure 2.9. If we assume that HTLV-1 Gag is approximately the same size as HIV-1 Gag, we would expect a maximum copy number of ~1300 if the entire surface of the 75 nm diameter VLP was covered by Gag. Our studies indicate that on average less than half the surface of HTLV-1 VLPs is covered by Gag. It has been observed for HIV-1 VLPs that 40-60% of the surface is covered by a Gag shell (17, 18, 79). Our labs' previous HIV-1 FFS experiments also indicated less than half the surface of HIV-1 was covered with Gag (9). Thus, both HTLV-1 and HIV-1 seem to require an assembly model that does not rely on Gag self-assembly into a closed shell. Our data suggest that HIV-1 and HTLV-1 VLPs with EYFP labels have similar Gag coverage of their surface. The Gag copy number of HIV-1 and HTLV-1 VLPs was repeatedly tested in 293T

cells and COS-1 cells. The copy number ratio was found to be consistently between 3:1 and 4:1.

Of interest is the relative copy number between HIV-1 and HTLV-1, as they have closely related Gag structures, but the VLPs produced are very different in size. Our studies of HIV-1 VLPs produced in the same cell line under the same conditions resulted in an average copy number of ~1800. Our average HTLV-1 VLP copy number is ~500. This indicates that on average a single HIV-1 VLP has 3-4 times the amount of Gag compared to a HTLV-1 VLP. HIV-1 VLPs are nearly twice as big as HTLV-1 VLPs on average, and since surface area goes as radius squared, we would expect HIV-1 VLPs to contain approximately 4 times the amount of Gag if similar coverage was achieved. This approximately matches what we observed, so it seems HIV-1 and HTLV-1 with EYFP labels do have similar Gag coverage on their surface. The relative copy number between HIV-1 and HTLV-1 was repeatedly tested in 293T cells and COS-1 cells. It was found to be consistently a ratio between 3-4:1.

Interestingly, we did not observe any sections of organized structure near the membrane with our EYFP tagged HTLV-1 VLPs under cryo-TEM, although structure was observed for HIV-1 VLPs (data not shown) as has been reported by other groups (17). Initially we concluded that HTLV-1 Gag must be organized significantly differently than HIV-1 due to the relatively constant radial density profile across the VLP and lack of visible structure as shown in Figure 3.6. HIV-1 VLPs have been observed to contain patches of organization, which are due to an organized Gag lattice next to membrane. The lack of these patches in HTLV-1 indicates a significant difference in VLP organization. The explanation for this lack of structure in HTLV-1 is explored in the next section.

3.3 Evidence of EYFP label interference with HTLV-1 VLP assembly

Control experiments were conducted to assess the impact of the fluorescent tag on the behavior of the model system. Previous results with HTLV-1 Gag expression in

cells indicate that labeled and non-labeled Gag (which we refer to as ‘dark’ Gag) colocalize to the same areas of the cell including puncta on the cell membrane where it is presumed that VLP production occurs (25). Because EYFP does not interfere with HIV-1 Gag assembly into VLPs (9), we anticipated that the EYFP tag should also have no effect on HTLV-1 VLP assembly. Surprisingly, the initial studies of ‘dark’ HTLV-1 VLPs produced without the EYFP tag suggest that EYFP perturbs particle assembly. This section presents evidence of label interference and briefly introduces experiments that explore the origin of this interference.

3.3.1 Experimental evidence of label interference

3.3.1.1 Cryo-TEM of ‘dark’ Gag HTLV-1 VLPs

VLPs were produced using unlabeled or ‘dark’ Gag with the previously described methods. The ‘dark’ Gag uses the exact same HTLV-1 codon as described in Figure 3.1, except that EYFP has been removed. A representative cryo-TEM picture of these dark VLPs is shown in Figure 3.7a. Gag-only VLPs showed distinct morphology when compared to EYFP VLPs (Figure 3.5). The arrow in Figure 3.7a points to a portion of a VLP that appears to have an organized structure near the membrane. A survey of all the particles suggested that the vast majority of the dark particles do have patches of distinct structure near the membrane, similar to the patches observed in HIV-1 VLPs. These patches have been interpreted as sections of Gag lattice. Currently we are investigating further structural details about these organized structures using cryo-TEM and related methods.

In addition to the distinct structure, the dark particles appear larger than the EYFP labeled Gag particles. To quantify this observation, a large sample of dark HTLV-1 VLPs was measured and a size histogram was developed. The measurements were performed by Jose Maldonado, with results from a preliminary experiment shown in Figure 3.7b. The mean radius is 106 ± 34. Compared to the distribution for the EYFP labeled VLPs (Figure 3.5b), the size distribution is shifted toward larger values, but the broad distribution of sizes is still preserved. The size observed for the dark Gag only

HTLV-1 VLPs is more closely in line with the average size of other retroviruses including RSV, MLV, and HIV-1, which range from 100-140 nm in average diameter (76, 80, 81).

3.3.1.2 VLPs containing a mix of labeled and unlabeled Gag

Because FFS relies on a fluorescent tag, we are not able to study dark-only particles, but we can explore VLPs assembled from a mixture of dark and labeled Gag proteins. If the labeling of Gag with EYFP does not interfere with assembly, then Gag and Gag-EYFP will be incorporated into the particle without preference. The composition of the mixed Gag VLPs should reflect the relative concentration of these proteins with the cell. On the other hand, if EYFP does interfere with assembly, the dark Gag will be preferentially incorporated into the VLPs and the VLPs will be enriched with dark Gag compared to the cellular concentration. To test the incorporation of labeled and unlabeled Gag in HTLV VLPs, a DNA titration experiment was performed. As a control, we performed a side-by-side experiment of HIV-1 dark mixing, which has previously been shown to have a labeled Gag copy number that scales with the amount of labeled Gag (9). The results are shown in Figure 3.8.

As expected, the labeled Gag copy number is within the uncertainty proportional to the percentage of labeled Gag plasmid for HIV-1 VLPs. However, HTLV-1 VLPs show a copy number that seems independent of the percentage of dark Gag DNA. This observation indicates that either the dark Gag and labeled Gag do not mix, but form separate particles, or mixing occurs, but in a more complicated fashion than predicted.

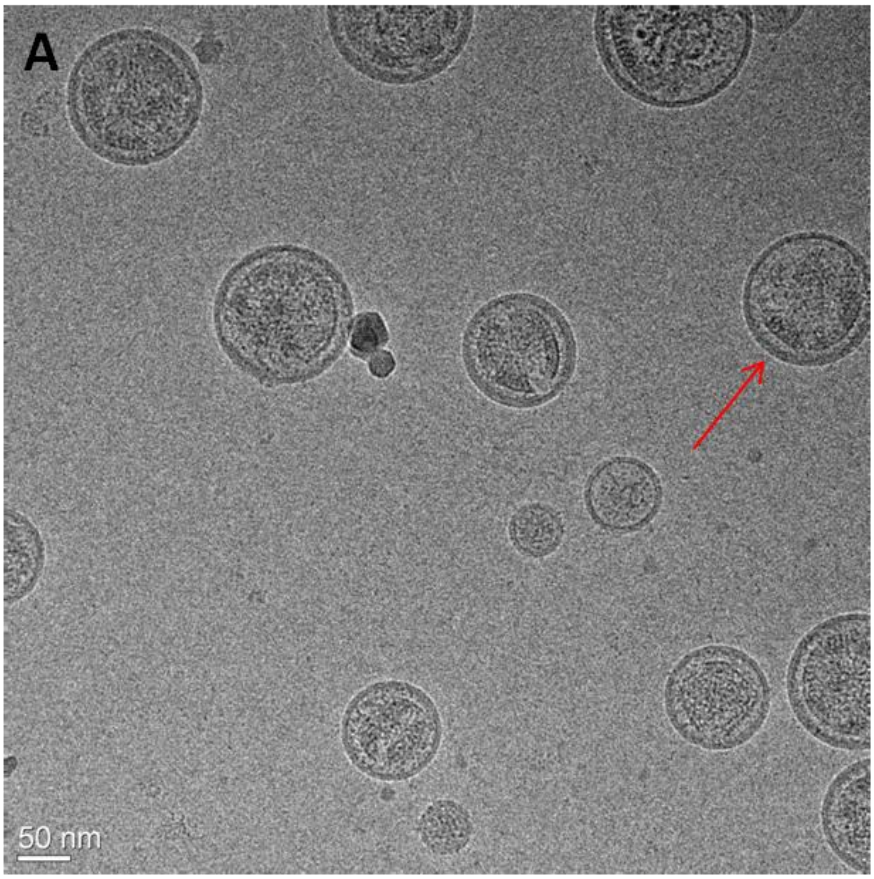
To further investigate this issue, a wider range of mixing ratios was explored. The main goal of this experiment was to identify a mixing ratio where labeled Gag is sufficiently diluted so that the perturbation from the label in the assembly process is negligible. If a safe mixing range were to be found, quantitative Gag labeling studies could be performed. HIV-1 VLPs were measured as a control and the results are as expected and previously reported (9). The extended HTLV-1 dark mixing results are shown in Appendix B, Figure B.1. No mixing ratio was found that allows for quantitative HTLV-1 studies. Surprisingly, the size of the VLPs was observed to grow with increasing

amounts of dark protein. At a mixing percent of 10% labeled Gag, the diameter of the particles was 104 ± 15 nm, which is in agreement with the average size for dark Gag only VLPs based on the preliminary cryo-TEM results shown in Figure 3.7.

VLPs produced from cells containing a 1:1 mixture of dark to labeled HTLV-1 Gag were also examined with cryo-TEM. The results are shown and further discussed in Appendix B.2. In general, the VLPs observed were peanut shaped with some sections resembling the structure of the dark only VLPs, and other sections displaying no distinct structure similar to the EYFP only particles. The cryo-TEM results provide further evidence that dark and labeled Gag do not correctly mix in HTLV-1 VLPs, although both types of Gag seem to be found on the same particle.

3.3.2 The search for a suitable fluorescently labeled Gag

The FFS data and cryo-TEM data all point to labeling of HTLV-1 Gag causing significant changes in assembly and presumably Gag-Gag interactions during particle formation. The size and particle morphology are significantly different between VLPs produced from labeled and unlabeled Gag. Additionally, the labeled and unlabeled Gag does not mix in an easily explainable fashion. FFS studies rely on a fluorescent label that does not change the properties of the proteins of interest in a substantial way. Discovering the source of the label interference along with possible alternative labeling methods for our HTLV-1 system is critical to performing biologically relevant experiments. We conducted a number of experiments designed to find the source of the label interference, which are described for reference in Appendix B. The following section summarizes the most exciting results, which were obtained for a construct carrying a linker between Gag and the label.



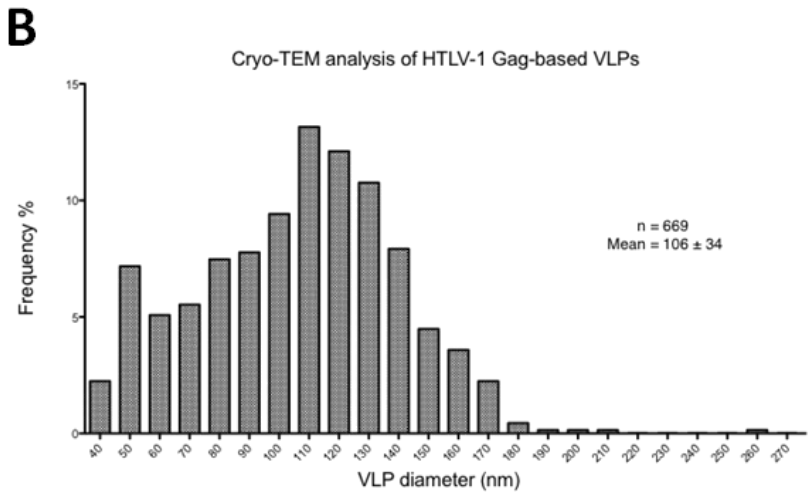


Figure 3.7 Cryo-TEM analysis of HTLV-1 'dark' Gag-based VLPs
(A) Cryo-TEM images of VLPs produced from transfecting 293T cells with the human codon-optimized HTLV-1 Gag without an EYFP tag. An example of the Gag lattice observed is indicated by the arrow. **(B**) Distribution of VLP diameters. The particle diameter was determined by averaging the longest and shortest measurements as before. 669 VLPs were examined (mean = 106 ± 34 nm).

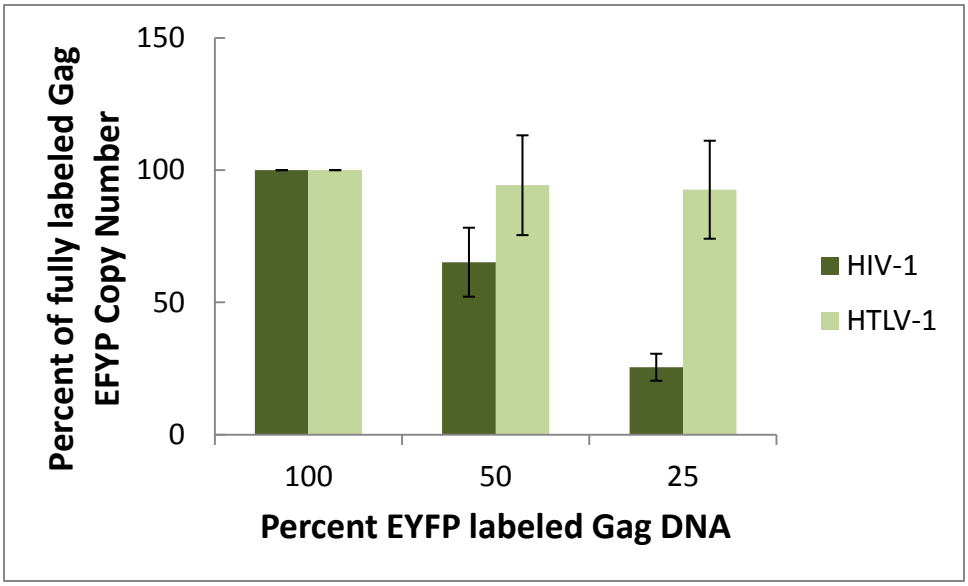


Figure 3.8 HIV-1 and HTLV-1 Gag stoichiometry in the presence of unlabeled Gag
VLPs produced from cells co-transfected with Gag-EYFP and Gag only. The Gag-EYFP copy number is shown as a percentage of copy number for fully labeled particles of each type.

3.3.2.1 HTLV-1 Gag-linker-EYFP studies

It has been shown that the addition of a linker or a sequence of flexible amino acids between the protein and its tag can diminish adverse labeling effects (82). If the EYFP label sterically interferes with conformational changes of Gag, introducing a suitably engineered linker may provide the necessary space and flexibility to overcome the steric restriction. Dr. Iwen Grigsby designed the helical linker inserted between the end of the HTLV Gag and the YFP label. This linker is 30 amino acids long and contains 5 repetitions of an amino acid sequence that forms alpha helical sections (83). Based on previous studies, this linker should add ~5 nm of separation between the NC domain of Gag and the EYFP. The gene construct was tested in 293T cells and found to express efficiently. The expressed proteins target the same areas of the cell as the previously described EYFP labeled Gag. Additionally, the linker construct produced particles of approximately the same concentration and copy number as the EYFP Gag only construct when measured with FFS. Next, we turned to FFS studies to quantitatively study the mixing of the linker construct with dark Gag. VLPs for a range of linker to dark Gag mixing ratios were produced. The diffusion time, apparent copy number, and concentration were determined for each sample, and results are shown in Figure 3.9.

If labeling Gag with linker-EYFP does not interfere with assembly, then Gag and Gag-linker-EYFP will be incorporated into the particle without preference, and the composition of the mixed Gag VLPs should reflect the relative concentration of these proteins within the cell. The results shown in Figure 3.9 indicate that even with the linker, labeling of HTLV-1 Gag interferes with assembly, although differently than in the absence of the linker. Compared to the results for the Gag-EYFP without a linker mixed with dark (shown in Figure B.1), we observe a distinct increase in both the copy number and size upon mixing labeled and unlabeled at a 1:2 ratio. The size of the VLPs is consistent with the size of the dark-only VLPs (Figure 3.7). Further dilution of the labeled Gag does not lead to further changes of size, but a gradual decrease in the copy number of labeled Gag is observed (Figure 3.9). The decrease in copy number with dilution for

constant VLP size is a promising observation for future quantitative FFS studies. However, ratios on the order of 1:10 or more seem required to exclude labeling effects, which approaches the limit of experimentally controlled mixing ratios possible due to the statistics of cotransfection (84).

The experimental uncertainty of these initial FFS experiments on mixed particles is quite high. Thus, additional experiments that repeat this study with more mixing ratios are required to confirm these results.

Cryo-TEM studies were also performed to corroborate the FFS results. HTLV-1 VLPs consisting of Gag-linker-EYFP only were observed to be closer in size to the dark Gag HTLV-1 VLPs than the Gag-EYFP HTLV-1 VLPs. Additionally, some structure was observed near the membrane, which could indicate the presence of Gag lattice as was observed for the dark only particles. More details and images can be found in Appendix B.2. Although the particle morphology and size distribution do not exactly match the dark Gag results, the addition of the linker before the EYFP does seem to reduce the effect of label interference based on these cryo-TEM results.

A cryo-TEM experiment looking at VLPs from cells expressing a 1:5 mixture of dark Gag and Gag-linker-EYFP was also performed. The experiment is described in detail in Appendix B.4 along with a sample image. The results for these mixed particles are promising. The histogram of size distributions is bell shaped with an average diameter of ~120 nm, and is qualitatively similar to the distribution of dark Gag VLP sizes (data not shown). In the majority of VLPs, there are sections of distinct structure near the membrane, similar to the patches of structure observed for the dark Gag VLPs. These results suggest that the addition of a linker decreases the label interference.

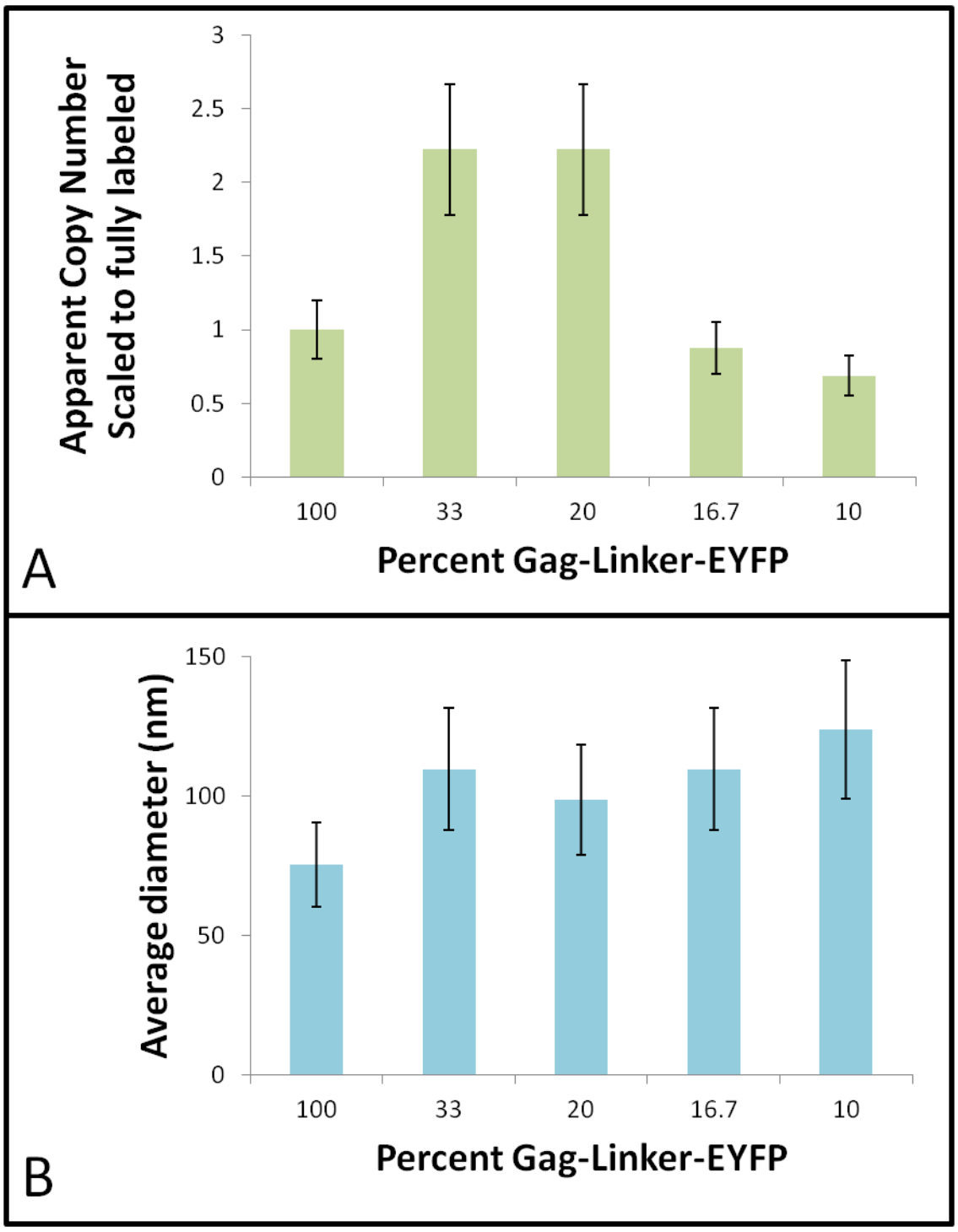


Figure 3.9 HTLV-1 VLP Gag-linker-EYFP 'dark' Gag DNA titration experiment
A. The apparent copy number for each VLP population was calculated and scaled to the Gag-linker-EYFP only copy number.
B. The average hydrodynamic radius was calculated using autocorrelation analysis.

3.3.3 Additional remarks on label interference studies

Our studies presented in Appendix B provide some insight into potential causes of label interference, but fail to overcome this problem. The most promising results of these experiments are the Gag-linker-EYFP mixed with dark Gag at a ratio of 1:3 or less. If the ratio is 1:3 or less, the size of the particles appears to be stable and similar to the dark Gag only VLP size. However, how the labeled copy number of these particles responds to changes in the ratio of DNA used during transfection is not completely understood. Cryo-TEM images of the Gag-linker-EYFP, dark Gag mixed particles are also promising. These particles appear similar in size to the dark only particles, and sections of distinct structure are visible as was the case for the dark Gag HTLV-1 VLPs.

The concept of viruses containing a small percentage of 'labeled' Gag is found in nature. Viral particle assembly naturally introduces a small percentage of 'labeled' Gags in the form of Gag-Pol. Retroviruses are thought to contain 5% Gag with Pol attached (10). This occurs because the Gag and Pol genes are next to each other in the viral RNA. Occasionally, the cellular machinery makes a 'mistake' and instead of finishing Gag completely it jumps to the Pol gene, leading to the fusing of these proteins. This naturally observed phenomena makes our dilution approach, introducing a small percentage of fluorescently labeled Gag, promising. Experiments are ongoing to identify an optimized fluorescently labeled Gag construct that is suitable for biologically relevant HTLV-1 VLPs assembly studies. The experiments presented in this section represent a first step in this goal.

3.4 Robustness of VLP assembly to perturbations of the Gag protein

It was pointed out by Crick and Watson that viruses should display symmetry due to the limited size of their genome (85).They argued that by using symmetry, viral particles can be built by one or a few types of protein, which are repeatedly used as building blocks for the capsid of the particle. This principle has been beautifully

confirmed for naked viruses, which have a capsid forming an icosahedral lattice or other lattices of high symmetry. The same principle has been assumed to apply to enveloped viruses that assemble at the plasma membrane of cells. For example, it is generally assumed that the formation of a Gag lattice is a necessary step in VLP assembly. Cryo-TEM studies of a range of retroviruses have consistently revealed at least partial areas of Gag lattice (17, 86). Originally, it was assumed this lattice structure was continuous, but new results have revealed that the immature HIV-1 virus contains an incomplete lattice. Some areas of the inner leaflet are covered with Gag lattice, while others are empty.

Although not part of our original HTLV-1 VLP study, our results provide some interesting observations that directly question common assumptions about retroviral assembly. The first remarkable observation is that the HTLV-1 Gag-EYFP construct produced VLPs, despite the substantial perturbation of the assembly process introduced by the label. Particles carrying the label were not only significantly smaller than regular particles, but also were entirely deficient of Gag lattice. The fact that these VLPs were even released suggests that the assembly process cannot be primarily driven by lattice assembly. While we currently do not know the forces responsible for the formation of budding sites at the membrane, it is clear that the process is more complex than originally anticipated. For example, experiments have shown that HIV-1 and HTLV-1 assembly involves lipid rafts, which are membrane microdomains that are enriched in cholesterol and saturated lipids (24). It is hypothesized that oligomerization of Gag serves as a mechanism to recruit microdomains at assembly sites, but the details of this mechanism are currently unknown (87). It is possible that the involvement of microdomains is relevant for our observation of particles without lattice structure. Additional studies are required to address this question.

We also conducted a series of experiments that investigate the robustness of VLP assembly to perturbations in Gag structure. This section contains the results of three such experiments. First, we studied a delta NC mutant of HTLV-1 Gag. NC is thought to play a critical role in RNA interactions, which in turn drives HTLV-1 VLP

assembly (88). The delta NC mutant contains EGFP after the CA domain, while the NC domain is completely eliminated. Finally, we studied mutants of HTLV-1 and HIV-1 Gag that have their CA domain swapped. Because the CA domain is thought to play a key role in Gag-Gag interactions during assembly (86), it may be that this domain determines key structural details of the VLP, such as size or morphology.

3.4.1 HTLV-1 Gag studies with Delta NC mutant

Previous studies of HIV-1 indicated that the NC domain plays a role in particle formation, most likely through NC-RNA interactions. In general, retroviral NC domains are nucleic acid binding proteins possessing potent nucleic acid condensing and chaperoning activities (88). The NC domain is involved in Gag-Gag multimerization through its association with the genomic RNA (89). Mutations in the NC domain result in viruses defective in the efficacy of Gag assembly, budding and particle morphogenesis for many retroviruses (HIV, SIV, MLV, RSV, BLV, MPMV) (88). In most cases, even with NC mutations, some Gag is targeted to plasma membrane assembly sites, but VLP production is at least partially affected or prevented. For HIV-1, the deletion of the Gag NC domain led to a delay of particle production when compared to the wild type Gag (90). To test the importance of the NC in the HTLV-1 system we developed a construct based on codon-optimized HTLV-1 Gag, but with a cut at amino acid 333 and an EGFP attached at the end. This deletion construct does not contain the NC domain. We transfected cells with this construct and examined the VLP properties using FFS. For simplicity, we collected particles after 28 hours and did not study the timing of particle release. Interestingly, this construct did result in VLP production with approximately the same concentration as the Gag-EYFP construct. The Gag copy number appeared to be slightly lower than for the regular Gag-EYFP construct, but the fit uncertainty is sufficiently high that this result needs to be confirmed by repeating the measurement. However, the result clearly demonstrates that NC-RNA interactions are not necessary for HTLV-1 VLP production.

Generally, it is assumed that RNA interactions are crucial for viral assembly (91). This assumption is supported by *in vitro* results. In vitro assembly of Gag proteins into VLPs requires the addition of RNA (16). Infectivity of authentic viruses requires that their genome be packaged into the virus; however, these results suggest that RNA interactions are not prerequisite for forming VLPs *in vivo*. The ability to form particles with limited RNA interactions again hints at a more complex assembly process then the current models suggest.

3.4.2 HTLV-1 and HIV-1 VLPs with swapped CA domain

The CA domain is thought to play a key role in Gag-Gag interactions during assembly (86). CA-CA interactions are important for the assembly of immature VLPs and the maturation process, where CA assembles into a conical shape. Previously, it has been reported that for HIV-1 and RSV the CA domain dictates virion size and morphology (86). To investigate if this is also true for HTLV-1, we developed two chimeras of HTLV and HIV with the CA domain swapped. HTLV/HIV-CA is the HTLV Gag with the CA domain replaced by the HIV CA domain. Similarly, HIV/HTLV-CA is the HIV Gag with the CA domain replaced by the HTLV CA domain. As there is a large size and Gag copy number difference between HIV-1 VLPs and HTLV-1 VLPs, the constructs could illuminate the importance of the CA domain in driving this difference. 293T cells were transfected with HIV-1, HTLV-1, HTLV/HIV-CA or HIV/HTLV-CA Gag. The chimeras with the CA domain switched produced fewer VLPs in both cases (about 3-5x lower). The VLPs from the HTLV/HIV-CA chimera were nearly identical to the HTLV-1 only particles in terms of copy numbers. The VLPs from the HIV/HTLV-CA had a copy number closer to that for the HIV-1 only particles, although always slightly lower. The results are shown in Figure 3.10. It was difficult to determine the size of these particles using FFS due to the low concentration of the VLPs. However, the similarity of Gag copy number for Gag and the chimeric Gag with the CA-domain swapped already suggests that the CA domain is not the driving force behind the differences observed in HTLV-1 particle morphology.

Thus, domains other than the CA domain seem to play an important role in HTLV-1 VLP assembly.

3.4.3 Concluding remarks on the adaptability of HTLV-1 VLP assembly

The surprising lack of Gag lattice observed in the cryo-TEM experiments studying HTLV-1 Gag-EYFP and the ability to alter the average size of HTLV-1 particles, suggested to us that the retroviral assembly process is more flexible than previously realized. While these VLP particles are not biologically relevant, the fact that so many different changes in Gag structure result in successful production of VLPs is an intriguing observation. Neither RNA interactions nor a Gag lattice seems to be required to produce particles. This suggests that our protein constructs could serve as interesting model systems for elucidating the fundamental processes required to form VLPs. Such studies need to be performed at the plasma membrane in order to directly investigate the assembly process. Our group is currently developing and testing spectroscopic techniques suitable for quantitative studies of the assembly process.

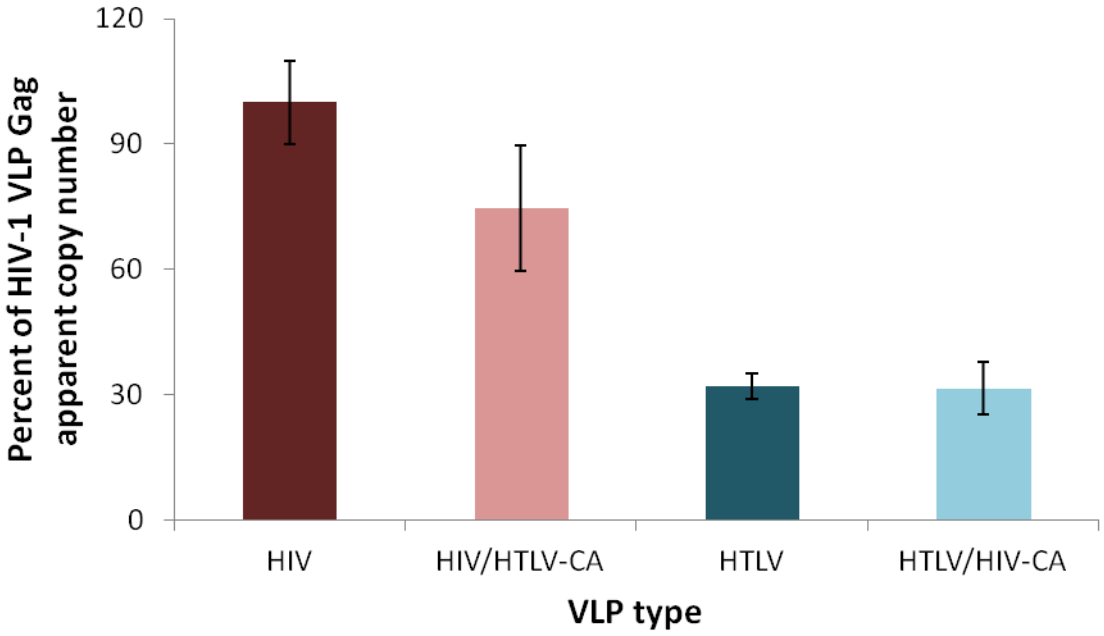


Figure 3.10 Copy number of VLPs with CA domain swap
VLPs were created from HIV-1, HTLV-1; HIV-1 with its CA domain replaced with HTLV-1 CA, or HTLV-1 with its CA domain replaced with HIV-1 CA. The apparent copy number of each species was determined using PCH analysis, and then was scaled to percent of the HIV-1 Gag copy number. Switching the CA domain does not significantly change the Gag copy number.

Chapter 4: Dual-Color FFS on VLPs: characterizing the copy number of APOBEC3G in HIV-1 VLPs

4.1 Introduction

While the Gag polyprotein is the sole viral protein necessary to form VLPs, viruses contain a variety of proteins. In retroviruses such as HIV-1, some of these proteins are viral proteins like protease and integrase, while others are cellular proteins such as clatherin, APOBEC3G, and actin (92, 93). The physical interactions that lead to the incorporation of this wide range of proteins are currently not well understood. We sought to clarify the mechanisms and copy number of proteins selectively incorporated into VLPs. Dual-color FFS applied in cells to study hetero-protein interactions is an instrumental tool for identifying the molecular mechanisms important for the packaging of proteins into virions. In fact, our lab is currently undertaking cellular studies in this area. In this chapter, I will take a complementary approach and focus on the use of dual-color FFS to determine the copy number of two different proteins located inside VLPs.

Biochemical methods can detect the presence of various other viral and cellular proteins within VLPs. However, there are a number of difficulties with these methods. First, due to the density gradient method used during most VLP purifications, the sample contains not only VLPs but also some cellular vesicles (94, 95), which can result in false positives. Methods do exist to try to distinguish viruses from vesicles, but some of these methods require additional vesicle digestion steps that can lead to the digestion of proteins on the surface of viruses as well (92). Additionally, biochemical assays provide mainly qualitative results. Quantitative interpretation of these experiments requires extensive controls and assumptions about the nature of the sample. These techniques only offer an indirect measure of protein incorporation and can be very time consuming. FFS studies of protein incorporation overcome many of these difficulties by providing quick and quantitative results, with little sample required.

One cellular protein that is known to be packaged into Vif-deficient HIV-1 viruses is APOBEC3G (apolipoprotein B mRNA-editing, enzyme-catalytic, polypeptide-like 3G) (94–96). Humans have seven APOBEC (A3) proteins; four of them (A3D, A3F, A3G and A3H) have been demonstrated to be potent Vif-deficient HIV-1 restriction factors (99). We focused on the first member demonstrated to be a retroviral restriction factor, A3G. Its restrictive activity functions by deaminating deoxycytidines in single-strand DNA to deoxyuridines, resulting in massive G-to-A hypermutation, which interferes with viral replication. Studies have demonstrated that in addition to its DNA editing abilities, packaging of A3G into virions is a necessary step in its function as a restriction factor (100).

The HIV-1 virus overcomes this cellular defense by encoding a protein called viral infectivity factor (Vif) that triggers the degradation of the retroviral restriction factor A3G (101). In Vif-deficient viruses, the degradation of A3G is blocked, and the infectivity of the virus is restricted by A3G. The biological function of A3G has been widely studied with the hope that these studies may lead to a treatment that overcomes the inactivation by Vif to restore its restrictive function. The selective packaging of A3G into virions is thought to be dependent on its interaction and formation of complexes with both the NC domain of HIV-1 and RNA (102). These complexes are then packaged into the budding viral particles and enter the new host cell (103, 104). While packaging of A3G is well studied, quantitative information on the amount of A3G packaged and/or required for function as a restriction factor is sparse. Using biochemical methods, one recent study estimated that HIV-1 Vif-deficient VLPs contain 7 ± 4 A3G proteins (105). We would like to develop a spectroscopic method that determines the copy number of A3G quickly and quantitatively. Using Dual-color FFS measurements on VLPs carrying both labeled A3G and labeled Gag appears to be a very promising approach. Figure 4.1 shows a model of this double-labeled system based on the current knowledge of A3G incorporation into HIV-1 VLPs.

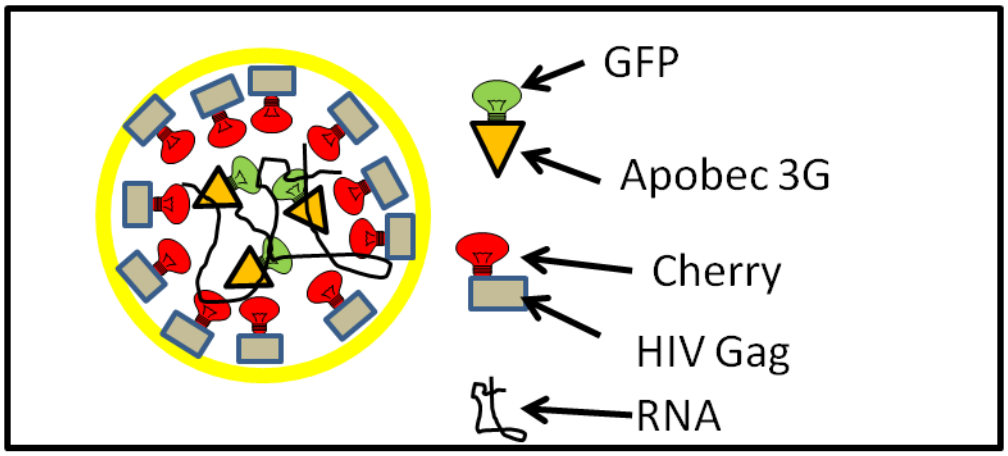


Figure 4.1 Model of A3G in HIV-1 VLP
Gag lines the inside of the membrane, with the mCherry tag on the inside attached to the NC domain. A3G labeled with EGFP is located inside the VLP, possibly interacting with RNA or Gag. The number of proteins shown is not to scale. Based on current knowledge we expect approximately 1000-2000 copies of Gag, forming an incomplete lattice around the inner boundary of the VLP, and a much smaller number of A3Gs somewhere within the VLP.

Dual-color FFS studies allow for the investigation of two interacting protein species by labeling each species with a different color and studying the coincident detection of both tags. While this concept has previously been limited to small protein complexes (106, 107), studying the presence of two proteins within a VLP is a natural extension. However, because VLPs are bright but rare particles, we first have to demonstrate the feasibility of the technique using well-defined model systems. The challenges of dual-color FFS on VLPs prompted us to introduce a new analysis method called intensity fraction plots (IFP), which allows for a quick, qualitative inspection of the dual copy number heterogeneity of the system. Using FFS and IFP, we characterized the average number of A3Gs per VLP and the effect of relative DNA concentrations on incorporation.

4.2 Theory

The normalized brightness b_n of an *n*-mer where each monomer is labeled with a fluorescent protein (such as EGFP and EYFP) is simply calculated by dividing the *n*-mer brightness by the monomer brightness

$$b_n = \frac{\lambda_n}{\lambda_1} \tag{4.1}$$

This normalized brightness gives the oligomerization state of the protein, $b_n = n$. One of the most common tag combinations used in dual-color FFS is EGFP and mCherry. Their spectral overlap is relatively small, making them a sensitive pair for resolving interacting protein species. mCherry has been extensively studied and has been shown to be suitable for quantitative dual-color studies with EGFP (51). One complication of working with mCherry is the presence of two distinct long-lived states with different brightness values, unlike EGFP and EYFP, which have a single, uniform brightness state. Consequently, the normalized brightness of an *n*-mer will be less than *n*, and cannot be described using equation 4.1. Fortunately, if the monomer and dimer brightness values of mCherry are known, the normalized brightness can still be correctly calculated by

$$b_n = (n-1)b_2 - (n-2). \tag{4.2}$$

where b_2 is the normalized dimer brightness. For a full derivation, see Wu et al. (51). Similarly, the true number of molecules for an *n*-mer needs to be calculated by

$$\frac{N_n}{N} = \frac{n}{b_n}(b_2 - 1). \quad (4.3)$$

These corrections will be applied when reporting the normalized brightness of protein complexes carrying mCherry labels.

4.3 Materials and methods

4.3.1 FFS measurements

The instrumentation for the dual-color FFS experiments is described in Chapter 2. A 580 nm dichroic was used to split the emission of mCherry and EGFP into two separate detection channels. All of the mCherry fluorescence was transmitted to the first "red" channel, while ~92% of the EGFP fluorescence was reflected to the second "green" channel. The remaining 8% of the EGFP signal was detected by the red channel. As previously described, the monomer brightness of each protein was separately calibrated in solution, along with the exact intensity split between the channels. Measurements were taken at 1000 nm to excite both proteins, at 905 nm to selectively excite EGFP-labeled proteins, and at 1040 nm to selectively excite mCherry-labeled proteins. The sampling frequency was 20 kHz for all measurements; at this frequency, no undersampling occurs. The excitation power for VLP samples was between 0.1-0.4 mW to avoid unwanted effects, such as photobleaching and optical trapping. The calibration and VLP data were fit to the theoretical PCH functions and non-ideal detector effects were included in the fit as described in Chapter 2.

4.3.2 Sample preparation

To validate and characterize the dual-color approach, we used 100 nm-diameter, internally dyed fluorescent polymer microspheres (Fluospheres Invitrogen, Grand Island, NY) as a model system. The microspheres were diluted and suspended into a buffer containing 0.5% w/v sodium dodecycl sulfate (SDS) and 14.7% w/v sucrose. Green

spheres with an emission peak at 515 nm and red spheres with an emission peak at 605 nm were used to approximate bright VLPs carrying either EGFP or mCherry. The SDS prevents aggregation, while the sucrose is added to match the density of the buffer to the spheres (1.06 g/mL), thus avoiding sedimentation. The sample was vortexed and sonicated for 30 min immediately before measurements to ensure monodispersion. 200 µL of the sample was loaded into an eight-well chamber and the sample was measured for 1-5 min at a frequency of 20 kHz with excitation powers between 0.05 and 0.2 mW. These conditions avoid undersampling, photobleaching, saturation and particle trapping.

COS-1 cells and 293T cells were cultured and transfected as described in Chapter 3. Cells were transfected with plasmids expressing either tagged HIV-1 Gag, or tagged A3G, or both proteins. In all cases, the total amount of Genjet was kept constant at 5 µg of DNA and 15 µL of Genjet per 25 cm^2 plate. VLPs were collected and purified as reported in Chapter 3. *In vivo* monomer and dimer brightness calibrations were performed in both 293T and COS-1 cells (8).

4.4 Model system results

4.4.1 Quantitative copy number with EGFP, EYFP and mCherry labels

As a first step in dual-color FFS experiments, it is necessary to confirm that all tags are interchangeable and produce results matching those of the corresponding single-color experiments. For our dual-color FFS studies of VLPs, we needed to measure the brightness of VLPs carrying EGFP, EYFP and mCherry labels. These brightnesses will then be converted into copy numbers to confirm that copy number is independent of label choice.

Accurately measuring the VLP copy number requires us to revisit the brightness doubling of a dimeric protein complex. Measurement of the dimeric brightness is fundamental to accurately convert VLP brightness to copy number.

Previously, we reported that the brightness of EGFP and EYFP doubles for a dimeric complex within an accuracy of 10% (108). Thus, EGFP and EYFP are both well described by a single brightness state (8, 9), and the normalized brightness of an *n*-mer complex of either of these proteins is given by equation 4.1. mCherry, as previously discussed, fails to show brightness doubling for the dimer, which indicates the presence of more than one brightness state, and equation 4.2 must be used to determine the normalized brightness (51). This conversion requires a precise experimental value of the dimer brightness b_2. Therefore, while the brightness of dimers of EGFP, EYFP and mCherry have been measured *in vivo*, we reexamined the dimer brightness in the COS-1 and 293T cells that were used in these experiments. For EYFP and EGFP we measured dimer constructs in which each monomer is directly fused to the other. For mCherry, our dimer consisted of two monomers separated by the ligand binding domain of the RAR nuclear receptor. The addition of this protein is simply to act as a spacer, which reduces hetero-FRET between the different brightness states of mCherry (51). Table 4.1 summarizes previously reported and current results for 293T cells and COS-1 cells. All dimer calibrations agree with earlier results within uncertainty. We will use a measured doubling factor of 1.65 for mCherry copy number corrections throughout the rest of the chapter. It should be noted that the uncertainty in normalized dimer brightness affects the uncertainty in reported copy numbers. In comparison to the other two labels, the normalized dimer brightness of mCherry had higher variability in day-to-day measurements. The data reported in Table 4.1 represents the average of two different experiments for each cell line, with approximately 15 cells expressing monomers and 15 cells expressing dimers measured per experiment.

Normalized Dimer Brightness (b_2)	**$EYFP_2$**	**$EGFP_2$**	**$mCherry_2$**
293T cells	1.95 ± .06	1.90 ± .07	1.62 ± .12
COS-1 cells	1.92 ± .07	1.95 ± .05	1.65 ± .10
Previously reported	2.02 ± .05 (108)	1.97 ± .10 (8)	1.70 ± .05 (51)

Table 4.1 Normalized dimer brightness of EYFP, EGFP and mCherry
The measured normalized dimer brightness for each cell line used in this chapter is shown along with previously reported results. The measured results are based on two experiments in which approximately 15 cells expressing monomers and 15 cells expressing dimers were measured for each color. All measured results agree with previously reported results within uncertainty.

Next, we created VLPs from HIV-1 Gag tagged with each of the different labels. A total of three species (two brightness species and a background species) are necessary to describe the VLP population. The copy number of each species is calculated using equations 4.1 and 4.2. For simplicity, we only reported the apparent copy number of the VLP species (Figure 4.2). The apparent copy number for all three labels agrees with previously reported results for HIV-1 VLPs, under similar transfection conditions. The concentration and individual species copy numbers also agree within uncertainty for all colors (data not shown).

Quantitative brightness analysis requires robust labels with stable brightnesses. Interactions between fluorophores can result in energy transfer, interactions with the environment can result in quenching, and some fluorophores are sensitive to changes in pH. All of these processes produce brightness changes. Our experience has shown that EGFP is a very robust label, which is most likely attributable to the fluorophore being located inside the protein, protected from the environment (thus no quenching). FRET occurs when labels with overlapping emission and excitation spectra are brought into close proximity. If a label exists in only a single brightness state (EYFP, EGFP), this does not change the brightness and is referred to as homo-FRET. The Gag copy number for VLPs labeled with EYFP and EGFP match well because both labels have a single brightness state and the fluorophore is protected from the environment by the protein matrix.

For labels with more than one brightness state, such as mCherry, the possibility of hetero-FRET between adjacent molecules in different brightness states complicates the interpretation of the brightness of large complexes, such as VLPs. Although the copy number of mCherry agrees with the copy number for the other tags, we noticed that it was slightly but consistently lower than the other labels. This could be attributed to the difficulty in precisely calibrating the dimer brightness of mCherry as required for copy number calculations, or alternatively as the result of FRET between closely packed mCherry proteins within a VLP. In HIV-1 VLPs, much of the Gag is in patches of tight

lattice structure with a spacing of 8 nm between the CA domains of neighboring Gag molecules (17). At this tight spacing, the brightness may be affected by hetero-FRET between different brightness states of mCherry. When two mCherry proteins were fused directly together and measured in cells, the normalized dimer brightness was 1.50 ± 0.10. We hypothesized that the equivalent normalized dimer brightness will likely reside between 1.5 and 1.65 for the spacing in VLPs. To determine if a small amount of FRET between brightness states is responsible for the slight decrease in apparent copy number of mCherry VLPs, lifetime studies on VLPs need to be performed. The change in fluorescence lifetime would quantify the amount of FRET occurring. However, the data in Figure 4.2 demonstrate that the effect of FRET would be small if present at all. Thus, the results shown in Figure 4.2 indicate that all three colors can be used for quantitative FFS studies of VLPs.

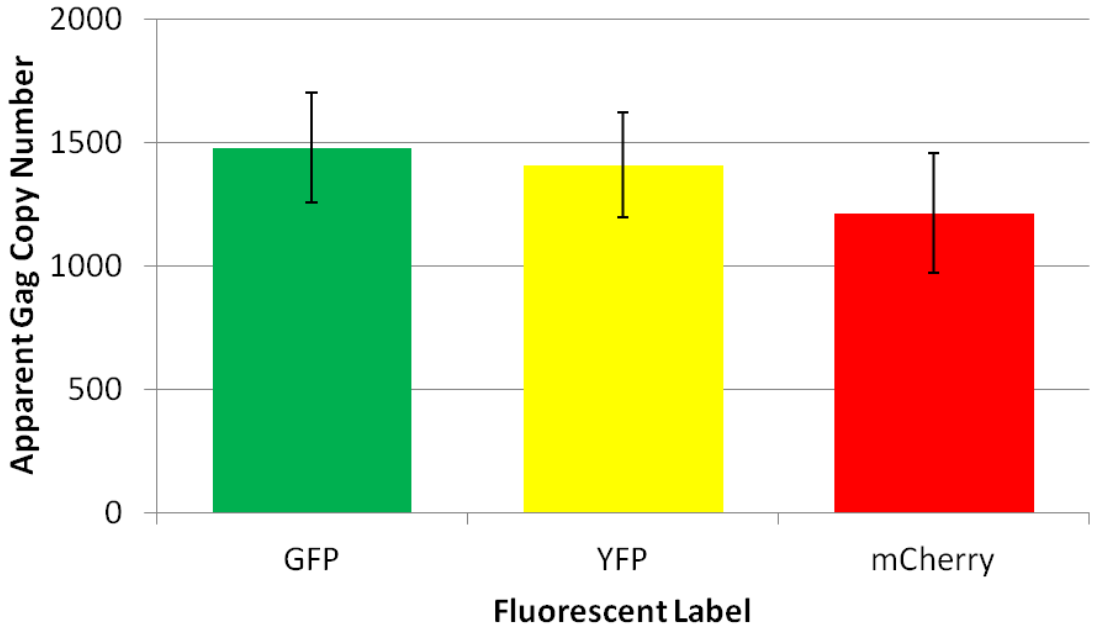


Figure 4.2 Apparent copy number of HIV-1 VLPs match for all tags
Single color HIV-1 VLPs were produced by transfecting 293T cells with EYFP, EGFP or mCherry labels attached to the NC domain of Gag. EYFP and EGFP dimers have double the brightness of monomers in cells, so the standard normalized brightness equation was used to determine the apparent copy number. mCherry dimers have a normalized brightness of 1.65 ± 0.1 in cells due to the presence of two brightness states in mCherry monomers. Equation 4.2 was used to calculate the apparent copy number for mCherry. With the mCherry dimer correction all colors produce the same copy number VLPs within uncertainty, allowing the use of any tag in dual-color experiments.

4.4.2 Protein copy number of VLPs with 1D- and 2D-PCH

Although 2D-PCH has been used extensively for studying small numbers of interacting proteins (106, 107), it has not yet been used to study complexes containing hundreds to thousands of proteins such as VLPs. While we expected the results for 1D-PCH and 2D-PCH to match, this prediction needs to be verified experimentally. For example, the background is not evenly split between the two detection channels, which could give rise to complications. Background from the VLP buffer contributes almost entirely to the red channel signal. This asymmetry may introduce a bias into the analysis if the fit does not properly identify the background contribution from each channel.

Thus, we compared 1D-PCH and 2D-PCH analysis for the same VLP sample. A HIV-1 VLP sample with EGFP-labeled Gag was measured. First, the data from the green channel was analyzed by 1D-PCH. Next, the FFS data from both channels of the same measurement were analyzed by 2D-PCH. We expected both methods to return two VLP species with matching copy numbers and concentrations. We also expected the brightness ratio between the red and green channels for the VLP species to match the intensity ratio measured in our EGFP calibration. As shown in Figure 4.3, a three species model (two VLP species, one background species) is sufficient to describe both the single- and dual-channel measurements at the original frequency of 20 kHz. Additionally, the fit parameters (Table 4.2) demonstrate that 1D- and 2D-PCH return the same copy number and concentration for each species. The brightness split between the two channels was as expected. The calibration identified that 8.2% of the total EGFP signal was detected in the red channel. The 2D-PCH fit of the EGFP VLPs resulted in the same brightness split. We concluded that 2D-PCH could be used to quantitatively study VLP samples.

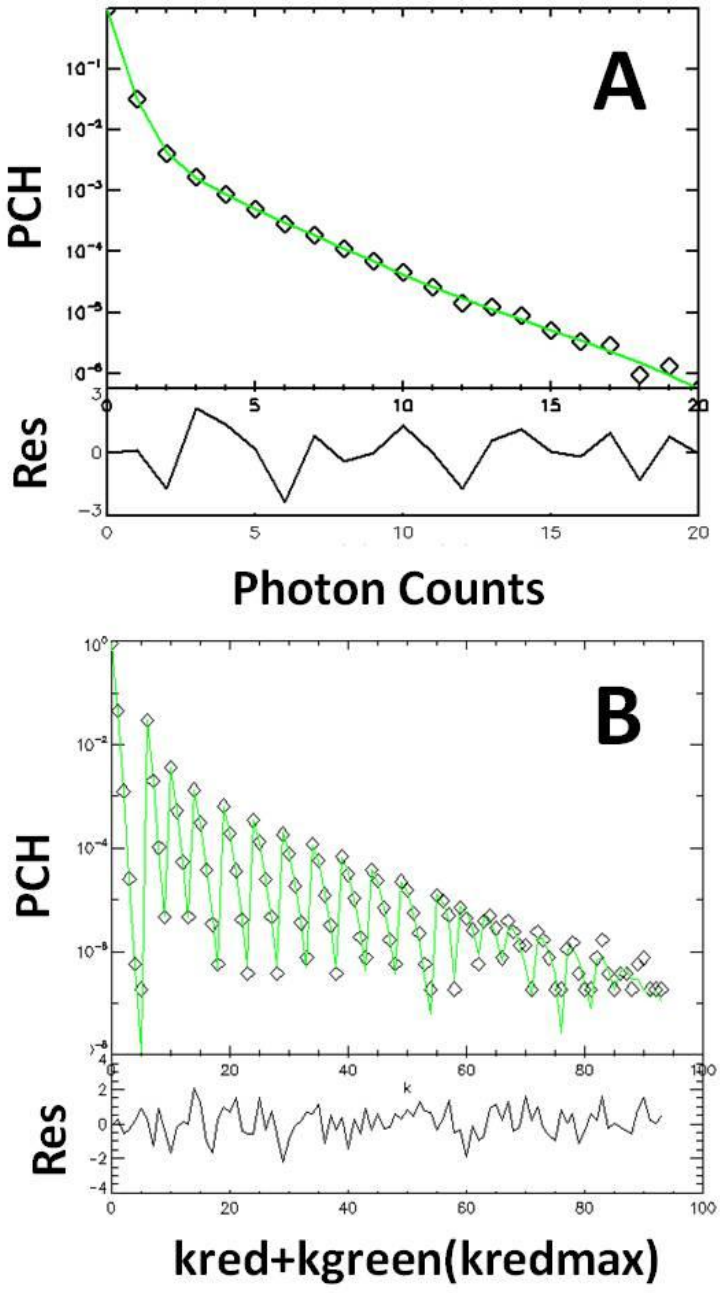


Figure 4.3 1D-PCH and 2D-PCH return the same copy number and concentration for single color VLPs

HIV-1 VLPs with Gag-EGFP were produced in 293T cells and then measured using a two-channel setup. 91.8% of the signal is reflected into the green channel. **(A)** The green channel was analyzed using 1D-PCH with deadtime and afterpulsing corrections. **(B)** 2D-PCH with deadtime and afterpulsing corrections was used to analyze both channels. Both analysis methods were performed at the original sampling frequency of 20 kHz. The results are shown in Table 4.2. Both analysis methods lead to fit parameters that match within uncertainty.

Analysis Method	Copy number species A	Copy number species B	Apparent Copy Number	Concentration species 1 (pM)	Concentration Species 2 (pM)	Reduced Chi-squared
1D-PCH	543 ± 20	1150 ± 50	660 ± 30	100 ± 10	11 ± 2	1.66
2D-PCH	540 ± 20	1110 ± 50	700 ± 30	120 ± 10	15 ± 3	0.81

Table 4.2 Copy number and concentration from 1D- and 2D-PCH fit of VLPs
Description of experiment found in Figure 4.3 caption

4.4.3 2D-PCH of mixed red and green spheres

The power of 2D-PCH lies in its ability to resolve species based on the color of their tags. As an initial test of 2D-PCH on bright particles, we measured a mixture of single color microspheres. Microspheres have a similar size to VLPs, are very bright, and are typically measured at low concentrations. These spheres represent an ideal test system because previous PCH analysis has indicated that a single brightness species is sufficient to describe the sample (26). This means that the sample contains little heterogeneity in dye amount, or brightness per sphere. Additionally, they have high photostability and, with proper preparation, do not interact. First, we measured a sample containing a single color of microspheres and analyzed the results with 2D-PCH. Next, red and green microspheres were mixed at a 1:1 ratio and the FFS data was again analyzed with 2D-PCH (Figure 4.4a). All measurements were taken with the 580 nm dichroic in place so brightness comparisons could be made. After aggregates were removed from the intensity trace (see Appendix A for peak removal information), the single color microsphere sample was fit with one bright species and one background species. The mixture of red and green spheres required a fit with one species for each color of microspheres and one background species. A comparison of the fit parameters for the single color and mixed color spheres is compiled in Figures 4.4b and 4.4c. Brightness and concentration of the mixture match the results obtained for the single color sphere samples. These results demonstrate the feasibility of using 2D-PCH to identify mixtures of single-colored rare and bright particles.

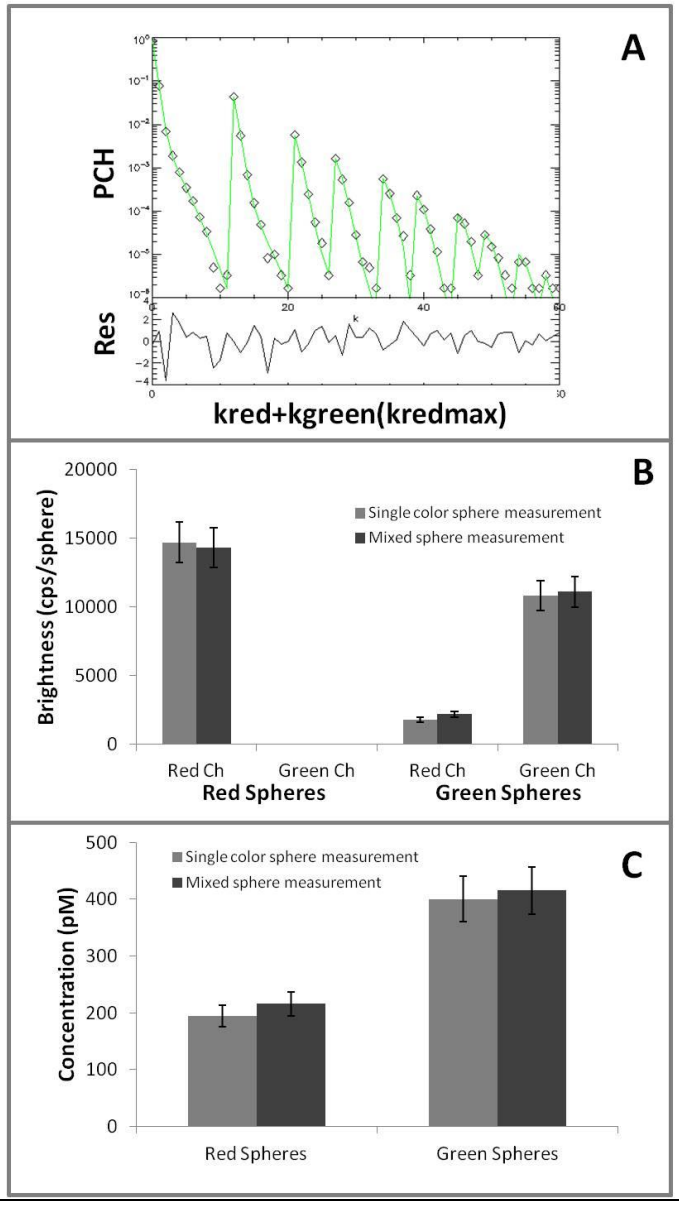


Figure 4.4 Dual color PCH analysis of microspheres
Three samples of 100 nm fluorescently-labeled spheres were prepared in a buffer of 14.7% sucrose and 0.5% SDS. The samples were red spheres only, green spheres only and a 1:1 mixture of red and green spheres, all created to have the same final concentration. Each sample was measured and 2D-PCH analysis was performed on the original data sampled at 20 kHz. **(A)** 2D-PCH fit of the photon counting histogram from the 1:1 mixture of red and green spheres to a three species model (reduced chi-squared of 1.4) **(B)** Comparison of brightness returned from single color sphere measurements and mixed sphere measurements. **(C)** Comparison of concentration returned from single color sphere measurements and mixed sphere measurements. The concentrations from each measurement match. The red spheres have a lower concentration in both cases because the stock solution concentrations were different, but the same dilution factor was used.

4.5 Two color VLP experiments

4.5.1 2D-PCH applied to mixed single color VLP sample

Microspheres were chosen as our model system because they are more homogenous in brightness than VLP samples and can be fit with a single brightness species. PCH analysis indicates that most single color VLP samples contain two brightness species but, as discussed in Chapter 2, this most likely approximates an underlying continuous brightness distribution. We investigated the effect of this heterogeneity on our ability to resolve multi-label species using 2D-PCH. The simplest mixed color VLP system consists of VLPs containing either only the red label or only the green label. To model this system we mixed HIV-1 VLPs harvested from cells expressing only EGFP labeled Gag, and HIV-1 VLPs harvested from cells expressing only mCherry labeled Gag. As each single color VLP sample can be approximated by two species in a 1D-PCH and 2D-PCH fitting, we expected a mix of these samples to contain 4 resolvable species (2 of each color) when analyzed with 2D-PCH. Our hypothetical model of the number of resolvable species for single color VLPs is shown in Figure 4.5a. This hypothesis needed to be tested because it is possible that due to the higher sensitivity of 2D-PCH, more than 4 species would be required to fully describe the 2D-PCH of the mixed sample.

The results of 2D-PCH fitting of mixed single color VLPs is summarized in Figure 4.6a and 4.6b, and the fit parameters are shown in Table 4.3. For comparison to previous results, we also measured each color VLP sample separately and analyzed these results using 2D-PCH. The single color VLP measurement results are reported in the same table and figure. All data was rebinned to 5 kHz before fitting. Rebinning increased the signal-to-noise ratio and it enabled us to resolve 4 distinct VLP species in the mixed color sample. 2D-PCH with 5 total species (2 red VLPs, 2 green VLPs, and background) returned a chi-squared of 2.7 for the mixed color VLP sample. Despite the relatively high chi-squared, remarkably, the results from measuring the mixed single

color VLP sample match the results from measuring each color independently. These results show that our hypothesis about the number of species required to fit a complex mixture of single color VLPs is correct, and that 2D-PCH can be used to fit VLPs containing a mixture of single color VLPs.

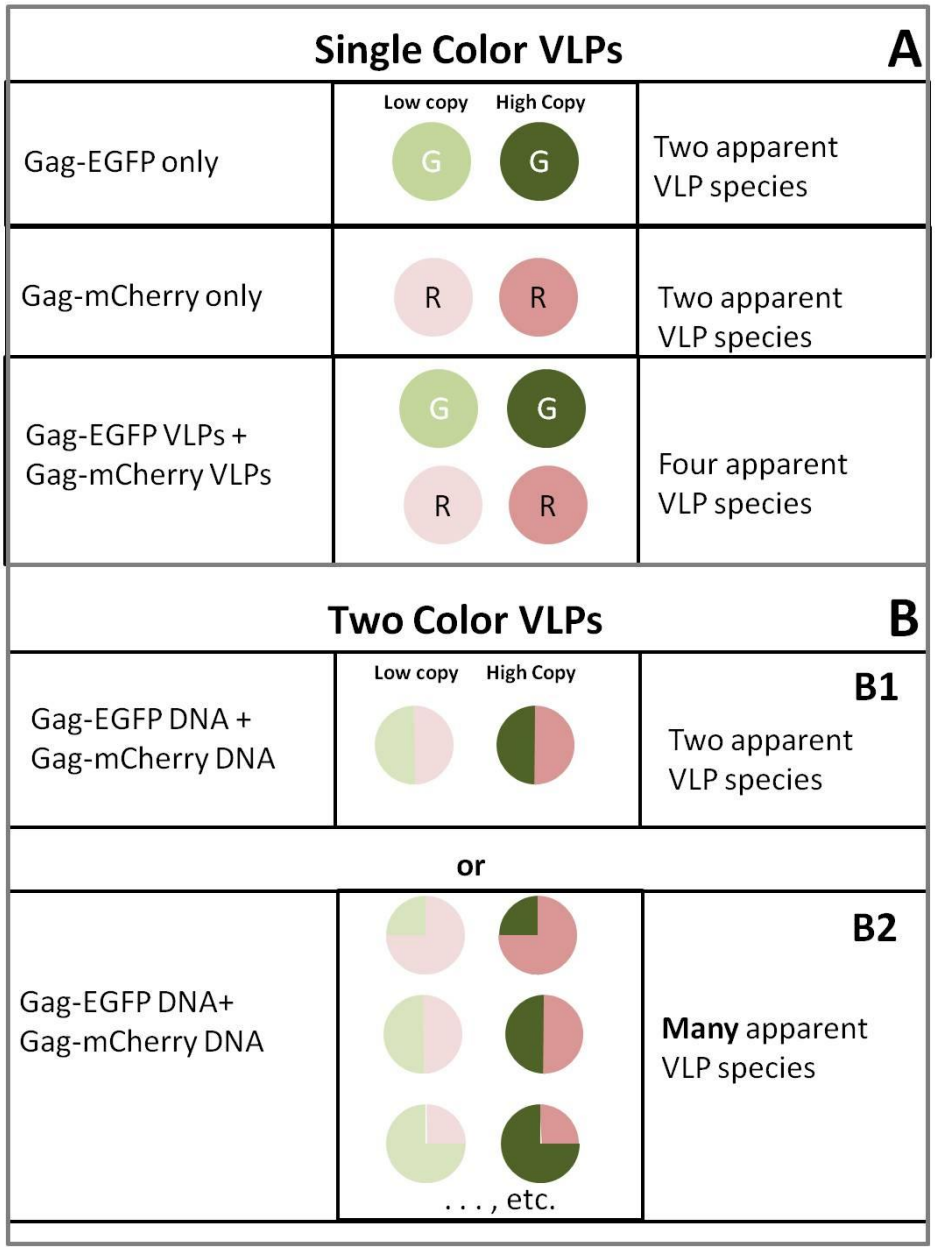


Figure 4.5 VLP apparent copy number species with mixed label models
(A) PCH resolves two VLP brightness species in samples containing VLPs with either all green or all red-labeled Gag. These two brightness species correspond to low and high Gag copy number VLPs. When these single color VLPs are mixed together, we expect to resolve 4 VLP species corresponding to a low and high copy number species of each color. **(B)** A sample containing both Gag-mCherry and Gag-EGFP on individual VLPs could contain a narrow distribution of protein labels, which would be approximated by 2D-PCH as two brightness species containing the same color ratio, but still a low and high total copy number. Alternatively these dual-color VLPs could contain a wide distribution of protein color labels that 2D-PCH may approximate as more than 2 species with different brightness ratios between channels for each species.

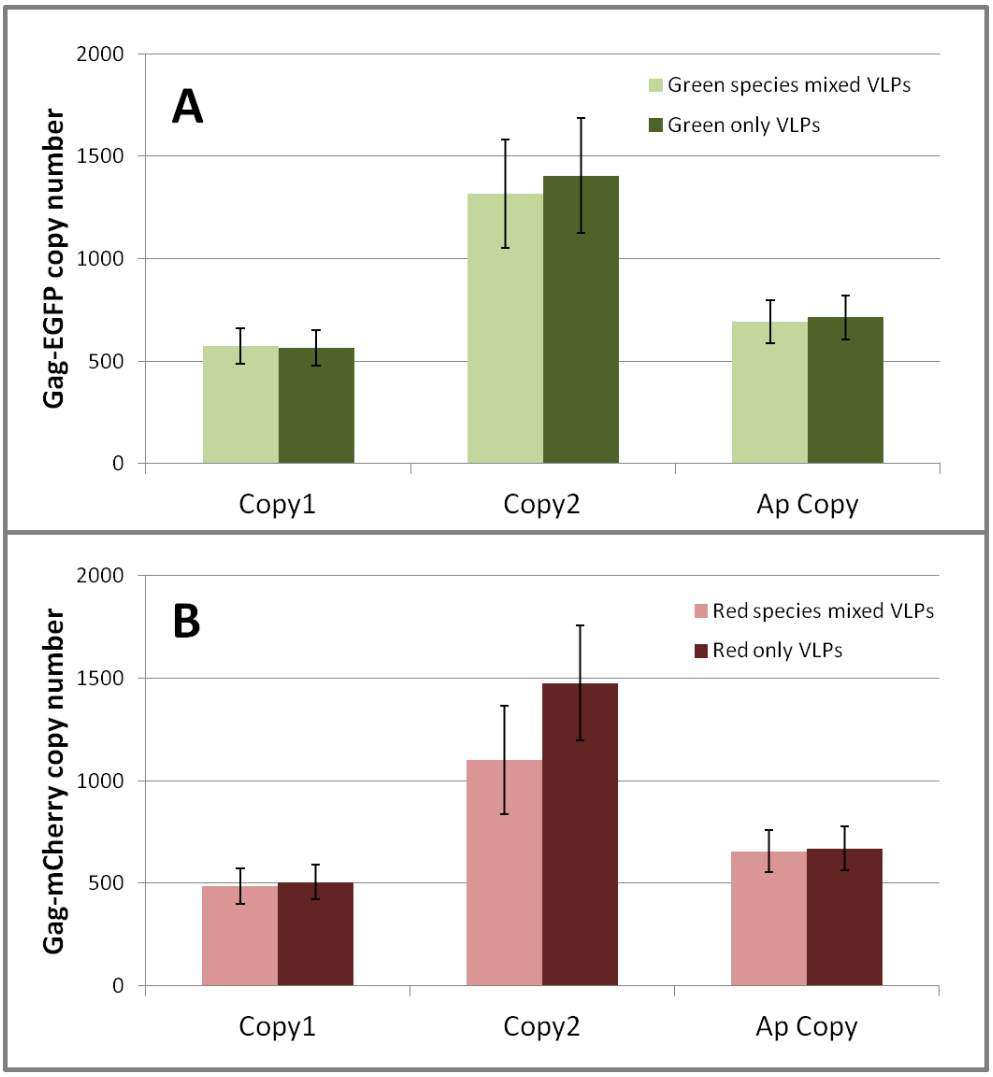


Figure 4.6 Comparison of copy number from 2D-PCH fitting of single color VLP samples with mixed single color VLP sample

2D-PCH analysis of a sample containing a mixture of red VLPs and green VLPs returned 4 VLP species. Two of these species are green only VLPs, and two of these species are red only VLPs based on the brightness ratio between channels for each species. For comparison, the single color VLP samples were also measured individually **(A)** The light color bars show the green species copy numbers calculated from 2D-PCH fitting of the mixed color VLP sample. The darker bars show the copy number calculated from 2D-PCH fitting of the green VLP only sample. The copy numbers of each species match between measurements. **(B)** Similarly, the light color bars represent the calculated red species copy numbers from 2D-PCH fitting of the mixed color VLP sample. The darker bars are the copy numbers calculated from the 2D-PCH fitting of the red only VLP sample. The copy numbers of each species agree. The uncertainty was calculated from the standard deviation of multiple measurements. The exact fit parameters are shown in Table 4.3.

Sample	**Green species**					$\varepsilon_{red}/\varepsilon_{total}$ **as %**	
	Copy1	**Copy2**	**Ap Copy**	**N1**	**N2**	**Species 1**	**Species 2**
Mixed	576 ± 40	1320 ± 120	694 ± 30	58 ± 8	4.8 ± 2	7.8 ± 0.5	10.1 ± 1.0
EGFP VLP	565 ± 40	1400 ± 100	715 ± 30	78 ± 10	6.8 ± 2	8.3 ± 0.5	8.8 ± 0.5
	Red species						
Mixed	486 ± 30	1100 ± 110	656 ± 30	91 ± 10	6.4 ± 2	100.0 ± 0.1	100.0 ± 0.1
mCh VLP	505 ± 30	1480 ± 100	670 ± 30	93 ± 10	6.3 ± 5	100.0 ± 0.1	100.0 ± 0.1

Table 4.3 Summary of 2D-PCH fit parameters for single color VLP experiments

4.5.2 Heterogeneity of VLPs with two color labels on a single VLP

Most biological samples of interest will contain VLPs containing red and green labels on a single VLP instead of a mixture of two distinctly colored VLPs, as was discussed above. For instance, if we label Gag with an mCherry and A3G with an EGFP label, we expect that A3G packaging results in VLPs carrying both red and green labels.

We first started with VLPs carrying Gag labeled with two different colors as a model system. Cells transfected with Gag-mCherry and Gag-EGFP DNA are expected to produce VLPs that contain both labels simultaneously. The number of VLP species required to fit this sample is difficult to predict. If the VLPs assemble with a fixed proportion of green to red labels, we expect that two brightness species with identical brightness color ratios are sufficient to model the sample (Figure 4.5b1).

However, it has been shown that not all cells express exactly the same ratio of proteins as the mixing ratio of plasmid DNA used in co-transfection experiments. A distribution of protein expression levels is generally observed. This distribution depends on the types of proteins co-expressed, the transfection method used, and other factors (84). Because we expected that labeled Gag is incorporated into VLPs without preference, the heterogeneity of the copy number ratios observed in VLPs should mirror the heterogeneity of the expression ratio of the differently labeled Gag observed in cells. Therefore, a strong variation in expression ratios across the cell population would give rise to a very complex mixture of VLPs. For example, some particles could carry only green or red labels, while other particles could contain both labels with varying ratios. In this case, the number of species required by 2D-PCH to approximate the sample is not clear. Identifying a suitable 2D-PCH model for a VLP sample with significant heterogeneity in the copy number ratio of the green and red labeled Gag (illustrated in Figure 4.5b2) would be difficult if not impossible.

To test the feasibility of 2D-PCH analysis, we collected and measured VLPs produced from cells that were transfected with a 1:1 ratio of Gag-mCherry to Gag-EGFP using TransFectin as the transfection reagent. This data was very difficult to fit with 2D-

PCH. A 3 VLP species 2D-PCH fit of the data at the original 20 kHz sampling frequency resulted in a reduced chi-squared of 12 and residuals that indicated a poor fit (Figure 4.7a). Increasing the number of species did not result in acceptable fits. Additionally, rebinning the data to 5 kHz as was necessary to fit the mixed single-color VLPs, because rebinning typically increases the signal-to-noise ratio and thus is beneficial for data fitting. In this case, however, the 5 kHz data could not be fit with 2D-PCH models of 2-8 brightness species. The raw intensity rebinned to 100 ms bins is shown in Figure 4.7b. This intensity trace hints that the sample may contain a wide range of color ratios, as there appear to be only red spikes and only green spikes. Our hypothesis is that this transfection method results in a VLP sample that more closely resembles the model shown in Figure 4.5b2. However, the current analysis methods are not sufficient to definitively identify the makeup of this sample.

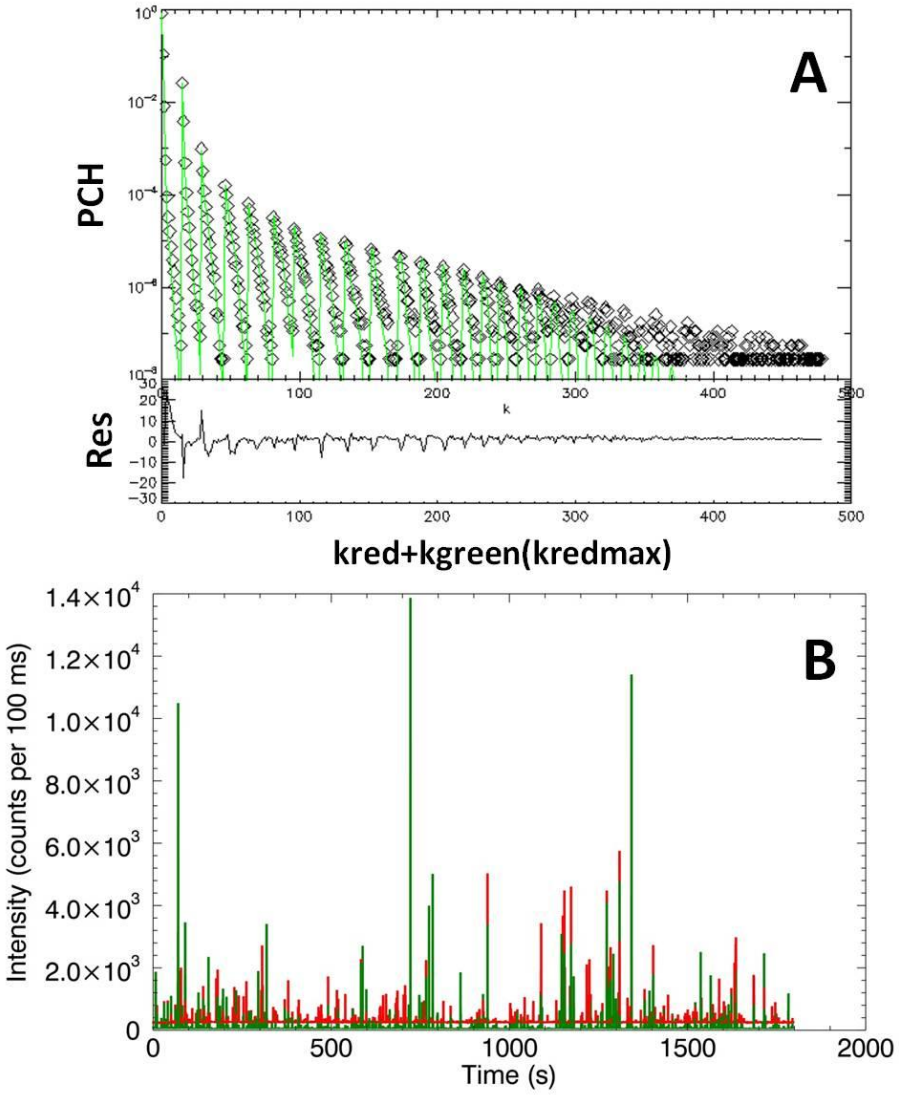


Figure 4.7 2D-PCH models do not fit mixed color Gag VLPs with TransFectin
VLPs collected from cells transfected with a mixture of Gag-mCherry and Gag-EGFP using TransFectin were measured and analyzed using 2D-PCH. No 2D-PCH models were found that produced a reduced chi-squared under 12 when data was analyzed at the original frequency of 20 kHz. **(A)** 2D-PCH of mixed color TransFectin VLPs with a 4 species fit (three VLP and one background species) results in a reduced chi-squared of 12. The brightness ratio between the two channels was different for all three VLP species. **(B)** Graph of raw intensity data of red and green rebinned to 100 ms and overlaid. The raw intensity data hints that the difficulty fitting 2D-PCH may be due to heterogeneity in the color ratios of VLPs, but when the data is displayed in this form it is difficult to clearly see this.

4.5.3 Introduction to Dual-color Intensity Fraction Plots

We introduced the dual-color intensity fraction plot (IFP) as a new analysis method that allows for the qualitative assessment of the distribution of ratios of two distinctly-colored proteins residing in bright but rare particles such as VLPs. The concept of the intensity fraction plot is shown in Figure 4.8. This method relies on two special properties of VLP measurements that have already been mentioned. VLP experiments typically have very low concentrations, but a high number of photons collected per VLP (which is related to brightness). At concentrations less than ~100 pM, the time between VLP events is much greater than the average time a single VLP spends in the observation volume. Additionally, the probability of more than one VLP being present in the observation volume simultaneously is very low. Because of these two facts, the raw intensity data can be rebinned in such a way that one bin represents either the integrated intensity of a single VLP in both channels, or the background in both channels. The ratio of the integrated intensity of the red channel divided by the total integrated intensity is called the fractional intensity f_I and is defined as

$$f_I = \frac{I_r}{I_r + I_g} \tag{4.4}$$

A graph of the intensity fraction for the total population of VLPs in an FFS experiment can then be plotted to determine the range of intensity fractions of individual VLPs.

Figure 4.8a shows a 3 s segment of FFS data from VLPs labeled only with EYFP. With a 580 nm emission dichroic, the intensity from EYFP is split roughly 81% into the green channel and 19% into the red channel. The data has been rebinned from its original frequency to 1 ms per bin to improve the differentiation of VLP events from the background. In order to generate the IFP, this data is further rebinned to 100 ms per bin. The choice of bin time will be discussed in the next section. Two example bins are shown. The first bracket is a bin containing only background counts, which is representative of the majority of bins. The intensity fraction for this bin merely returns

the intensity fraction of the background. As the background has a low integrated total intensity compared to VLP events, this bin is located in the bottom left of the dual-color intensity fraction plot. The second bracket shown contains a VLP event. The intensity fraction for this bin is the intensity fraction for the single VLP and represents one IFP data point.

Figure 4.8b shows the intensity fractions of all bins for the EYFP-labeled VLP experiment. There is a large cluster of points with red intensities at less than 100 counts per 100 ms due to the high percentage of bins containing only background signals. Above this background threshold, each peak represents a single VLP event. The range of total intensities per VLP is due to both the copy number variations of individual VLPs and to the inhomogeneous excitation profile. The inhomogeneity of the excitation profile prevents a direct calculation of the copy number from the intensity plot and is the reason 2D-PCH analysis is necessary to determine brightness. The slope of the IFP is the average dual-color fractional intensity of the VLP population. This sample contains only EYFP-labeled VLPs, so a narrow distribution centered on the EYFP intensity fraction, as determined during calibration, is expected. A linear fit of the IFP returns an intensity fraction of 0.195 or 19.5% of the signal in the red channel, matching both the 2D-PCH brightness fraction of both species and the fractional intensity observed from a calibration with a monomeric EYFP sample. The fit also includes a y-intercept, which returns the average red intensity per bin due to background, as this is the dominant non-VLP contribution to background. This plot represents the IFP for a simple sample consisting of VLPs carrying the same fluorescent label. The intensity ratio is constant for each VLP event, because the green fluorescence is split into the two detection channels by a fixed proportion that depends only on the optics and the spectrum of the dye, but is independent of the diffusion path through the observation volume. For samples with VLPs carrying more than one color, the spread of intensity ratios reveals information about the heterogeneity of the copy number ratio for each of the distinctly colored proteins. If there is a single slope with a narrow distribution of individual VLP intensity

fractions, this indicates a fixed copy number ratio across the entire population of VLPs. This scenario would match Figure 4.5b1. A wide distribution of individual VLP intensity fractions is evidence that the copy number ratio varies widely across the VLP population, as depicted in the second diagram of Figure 4.5b2. Using IFP plots to qualitatively distinguish between these scenarios will be addressed in section 4.5.4.

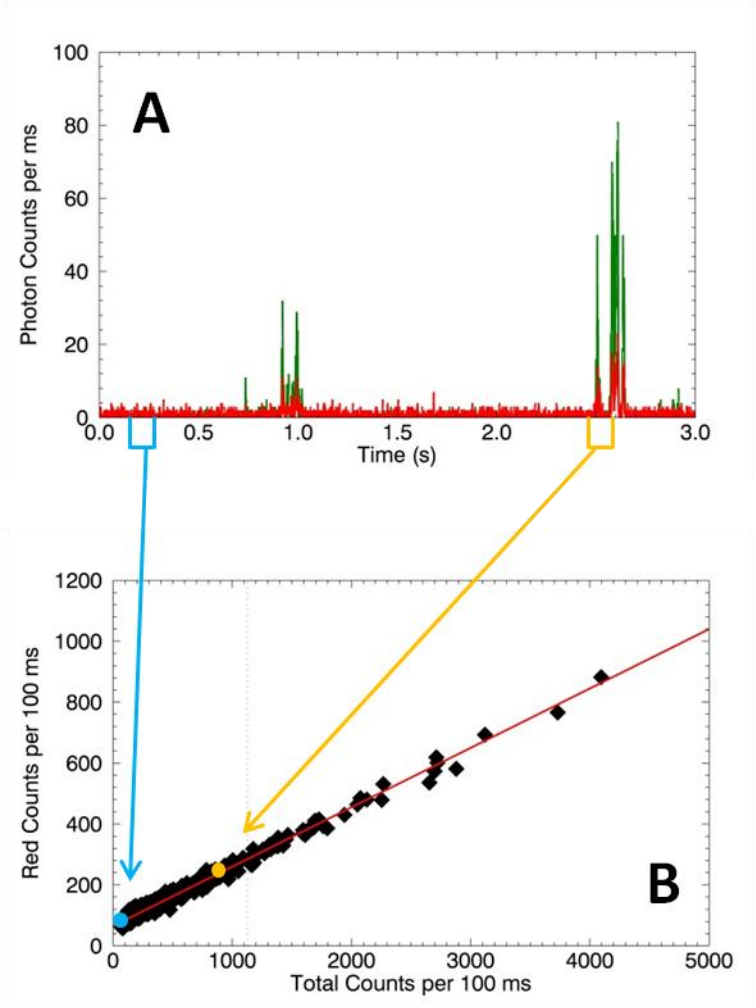


Figure 4.8 Concept of Intensity Fraction Plot (IFP) analysis
(A) A 3 s segment of raw data rebinned to 1ms per bin from HIV-1 VLPs labeled with EYFP. **(B)** IFP for data rebinned to 100 ms per bin is shown. Two example bins are highlighted. The first bin contains only background and results in a point in the bottom left hand corner of the IFP. Points clustered around this area are the dominant bin type and simply represents the intensity fraction of the background. The second example bin contains the signal from a single VLP event. The corresponding data point is in the middle of the IFP and represents the intensity fraction for a single VLP event. The IFP is built from graphing the intensity fractions for all VLP events. This graph is fit to a straight line, which represents the average intensity fraction of the VLP population.

4.5.4 Relationship between IFP and 2D-PCH

2D-PCH analysis returns the brightness in both channels for each species, which can be used to calculate the brightness fraction f_b for each species

$$f_b = \frac{\varepsilon_r}{\varepsilon_r + \varepsilon_g} \quad (4.5)$$

The VLP brightness is linearly related to integrated intensity for individual particles, so the brightness fraction for the VLP species should match the average intensity fraction ($f_b = f_I$), as long as background is negligable. This allows us to compare the brightness fraction for each VLP species to the average intensity fraction determined by the IFP. First, we consider single-color VLPs, such as the Gag-EYFP HIV-1 VLPs. The brightness fraction determined by 2D-PCH for both species is identical as only one color label is present. Thus, the IFP should also contain a single slope that matches the brightness fraction. The data for Gag-EYFP HIV-1 VLPs shown in Figure 4.8 demonstrates this is the case. The slope determined by fitting the IFP to a linear equation matches the brightness ratio of both species determined by 2D-PCH.

Next, we turn to the sample containing a 1:1 mixture of red VLPs and green VLPs (Figure 4.5a). The 2D-PCH analysis of this data identified 4 VLP brightness species, with two species having a brightness fraction corresponding to red VLPs, and two species having a brightness fraction corresponding to green VLPs. We expected the IFP to contain only two distinct slopes, with one slope characterizing the signature of green VLPs containing EGFP and the other, steeper slope characterizing the signature of red VLPs containing mCherry. This prediction is confirmed by the IFP for the mixture (Figure 4.9). The fact that almost all data points fall on either the red or green VLP slope means that nearly every IFP data point captures no more than one VLP event, as discussed in the next section. The result demonstrates the power of IFP to quickly and directly visualize the heterogeneity in the brightness fraction for the entire VLP population without complicated analysis.

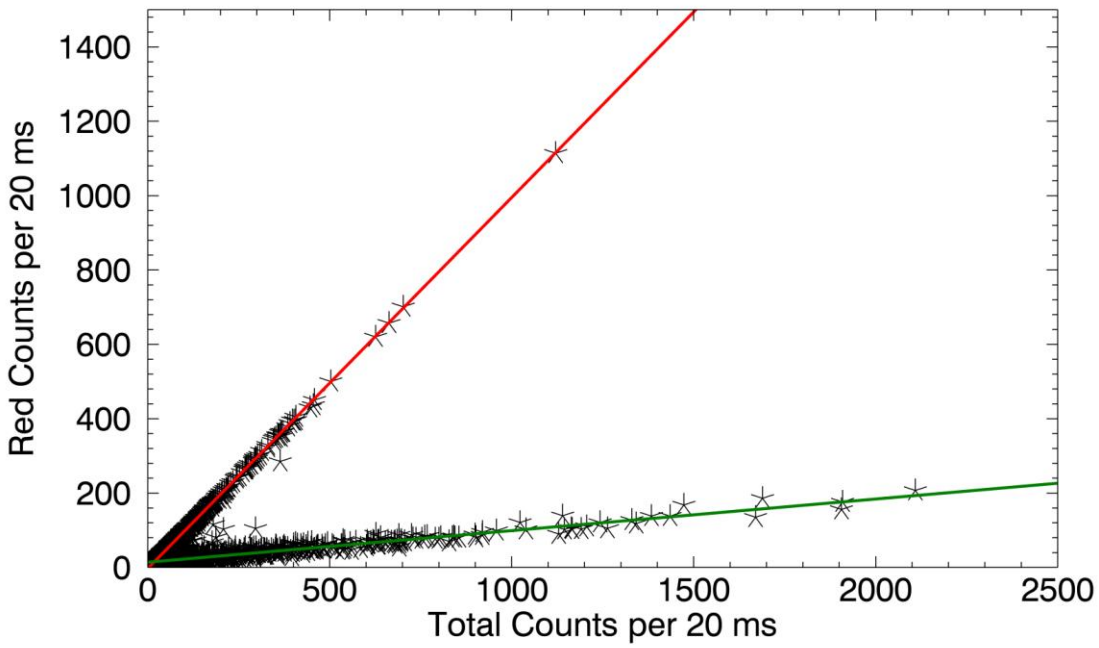


Figure 4.9 The Intensity fraction from IFP matches the brightness fraction from 2D-PCH fitting The IFP for a mix of red only VLPs and green only VLPs with a 20 ms bin time. Nearly all points fall on two distinct slopes, with the shallow slope matching the brightness fraction ratio for the green VLPs, and the steeper slope matching the brightness fraction for the red VLPs. The y-axis intercept of these lines represents the background intensity in the red channel.

4.5.5 The effect of bin size on intensity fraction plots

The choice of bin size for IFP data is important. Ideally, the bin size will be big enough so that each VLP is fully contained within a single bin and a good signal-to-noise ratio is obtained. However, the bin size must also be small enough so that the probability of a bin containing more than one VLP is very low. It is difficult to give general rules for determining the best bin size for a range of samples due to variations in concentration and background from sample to sample. However, Figure 4.10 outlines the general concept for determining the best bin size using the 1:1 mixed green and red VLP sample ($c \approx 100$ pM) shown in Figure 4.7. In Figure 4.10a, a bin size of 1 ms was chosen. Because the diffusion time for VLPs is on the order of 10 ms, a 1 ms bin time splits each VLP's fluorescence signal into multiple bins, as shown in the diagram. This creates two problems in interpreting the IFP. First, the signal-to-noise ratio is low and the spread of intensity fractions for individual bins increases, as observed for the green-labeled VLPs (Figure 4.9a). Second, VLPs that spend longer times in the observation volume are overrepresented on the intensity fraction plot. For instance, a VLP that spends 20 ms diffusing through the observation volume will have 20 points on the intensity fraction plot, while a VLP that just grazes the edge of the observation volume may only be represented by a few points. For these reasons, bin times shorter than 10 ms are generally not a good choice for typical VLP samples with concentrations in the range of 10-100 pM.

Figure 4.10b shows the best bin time for the mixed single color VLP experiment. Choosing 20 ms bins ensures that most VLP events fall into a single bin, as depicted in the diagram. Thus, the vast majority of bins contain the signature of only red or only green VLPs. Some VLPs are represented by more than one data point on the IFP due to the wide distribution of diffusion times and the possibility that a VLP event is split across two neighboring data bins. However, the spread of the data is greatly reduced due to the increase in signal through rebinning.

Figure 4.10c demonstrates what happens if too long a binning time is chosen (500 ms). The probability that more than one VLP visits the observation volume within 500 ms is non-negligible for a VLP concentration of 100 pM. The diagram shows both a red and green VLP event falling into the same bin (which extends past the edge of the axis) even though the two events do not overlap in time. When this happens, the data point on the IFP represents the average intensity fraction of the two VLPs. In our case, with a VLP population consisting of only red and green particles, a bin containing red and green VLPs will result in a data point falling between the slopes of the red and the green VLP populations. This confounds the interpretation of the data and could give rise to the false impression of VLPs aggregating. Thus, in order to avoid skewing the results of the VLP intensity fraction distribution, large bin times need to be avoided.

4.5.6 IFP of dual-color intensity ratio for VLPs produced using TransFectin

After introducing the intensity fraction plot and its relationship to 2D-PCH analysis, we returned to the original problem of determining if heterogeneity in copy number ratios is responsible for the failure of 2D-PCH to describe mixed color VLPs produced using TransFectin. Figure 4.11 shows the IFP of the VLP data from Figure 4.7b. The IFP clearly shows that the intensity fraction is heterogeneous. The red line represents VLPs that contain only red labels, while the green line represents VLPs that contain only green labels. This IFP shows VLPs that fall on these lines as well as a wide distribution between them, quickly and clearly demonstrating that sample heterogeneity is likely the reason that a 2D-PCH fit was not successful.

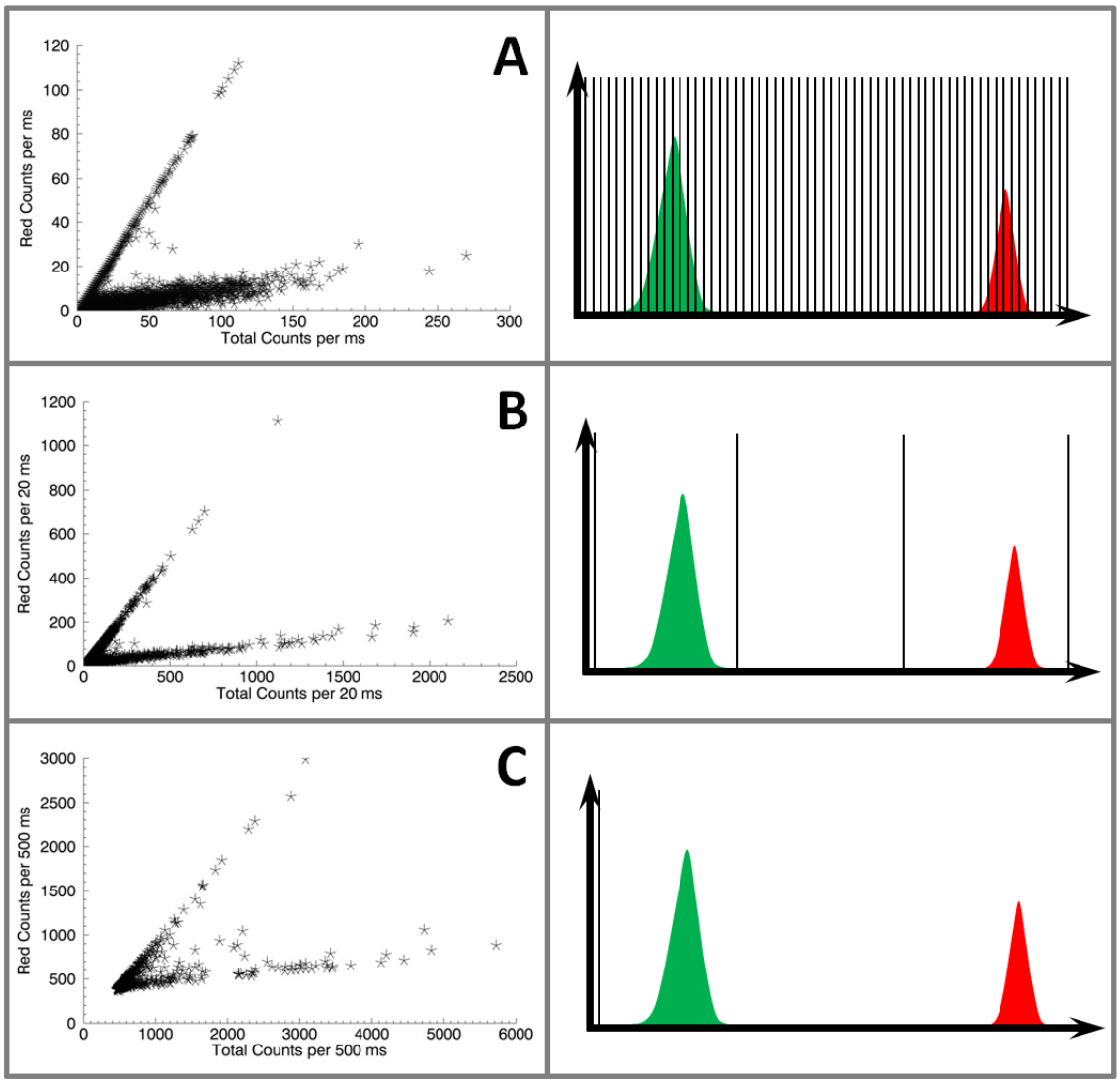


Figure 4.10 Effect of binning size on IFP
This figure explores the effect of the bin size on the IFP for the mix of single color red and green VLPs previously discussed. **(A)** VLPs diffuse through the observation volume in approximately 10 ms. When the bin size is chosen to be 1 ms, each VLP event is divided into multiple neighboring bins. This leads to multiple data points on the IFP for each VLP event. **(B)** When the bin size is chosen to be 20 ms, most bins contain the entire VLP event. The low concentration ensures that the probability of putting two VLPs in the same bin remains low. **(C)** When a large bin size is chosen, there is a non-negligible probability of combining two VLP events in the same bin. In the diagram, the edge of the bin is past the edge of the graph.

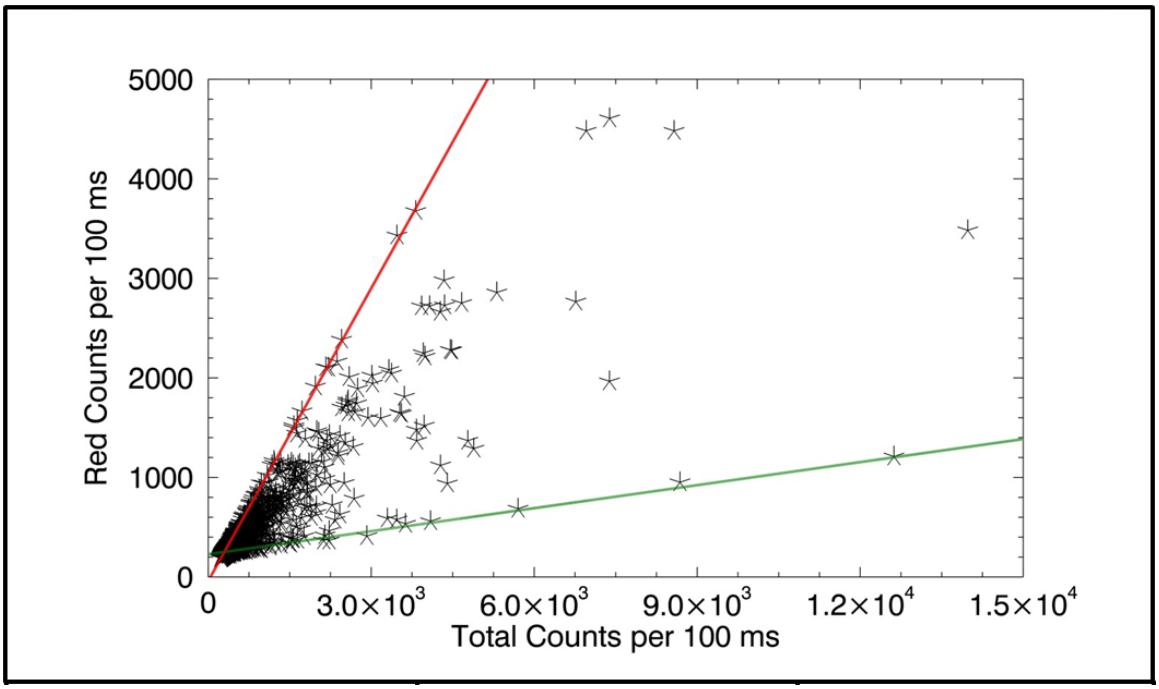


Figure 4.11 IFP for VLPs transfected with a mixture of Gag-mCherry and Gag-EGFP using TransFectin transfection reagent
The IFP illuminates the presence of a heterogeneous intensity fraction across the VLP population. The red line represents the intensity fraction for VLPs containing only mCherry labels. The green line represents the intensity fraction for VLPs containing only EGFP labels. Multiple VLPs fall on each line, plus a broad distribution in between these two limiting slopes.

4.5.7 IFP for VLPs produced using Genjet

2D-PCH analysis is powerful because it quantitatively determines protein copy numbers. However, our current transfection procedure leads to copy number ratios of VLPs that are too broad for meaningful and successful 2D-PCH fitting. The broad distribution of VLP copy number ratios is most likely due to the broad distribution of protein expression levels that is observed in cells transfected with TransFectin reagent (84). Our lab has observed that Genjet, a different transfection reagent, results in more homogenous cellular expression ratios. We collected VLPs transfected with a 1:1 ratio of Gag-mCherry to Gag-EGFP using the Genjet reagent and measured the VLPs as previously described. Figure 4.12 shows that IFP for these VLPs. Surprisingly, simply changing the transfection method resulted in a much narrower distribution of intensity fraction ratios. The IFP shown in Figure 4.12 suggests that these VLPs could potentially be analyzed by 2D-PCH fitting.

4.5.8 2D-PCH analysis of VLPs produced using Genjet

The IFP for the VLPs produced using Genjet transfection indicates that fruitful 2D-PCH analysis may be possible. Based on the observation of a single slope, we tried fitting the 2D-PCH to only 2 VLP species with the same brightness ratio. This simple model describes the data with a reduced chi-squared of 1.8. The data and 2D-PCH fit are shown in Figure 4.13. 2D-PCH identified two brightness species with identical brightness fractions. The existence of a single brightness fraction gives rise to an IFP with a single, sloped line, representing the average ratio of the protein labels. As further confirmation of the relationship between IFP and 2D-PCH, the single brightness fraction identified by 2D-PCH is graphed as a solid line in the IFP of Figure 4.12, and provides a good approximation of the average slope of the IFP data.

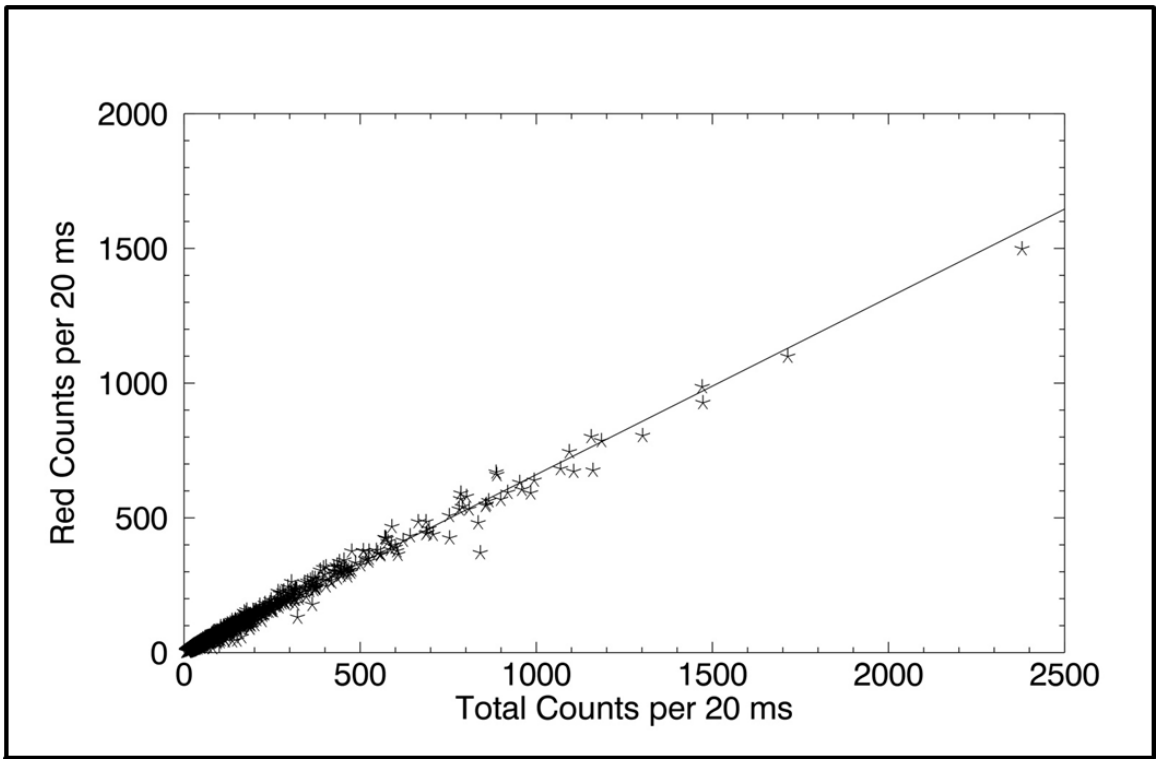


Figure 4.12 IFP for VLPs transfected with a mixture of Gag-mCherry and Gag-EGFP using Genjet transfection reagent

The IFP illuminates the relatively narrow distribution of intensity fractions in VLPs co-transfected with red and green labeled Gag using Genjet transfection reagent

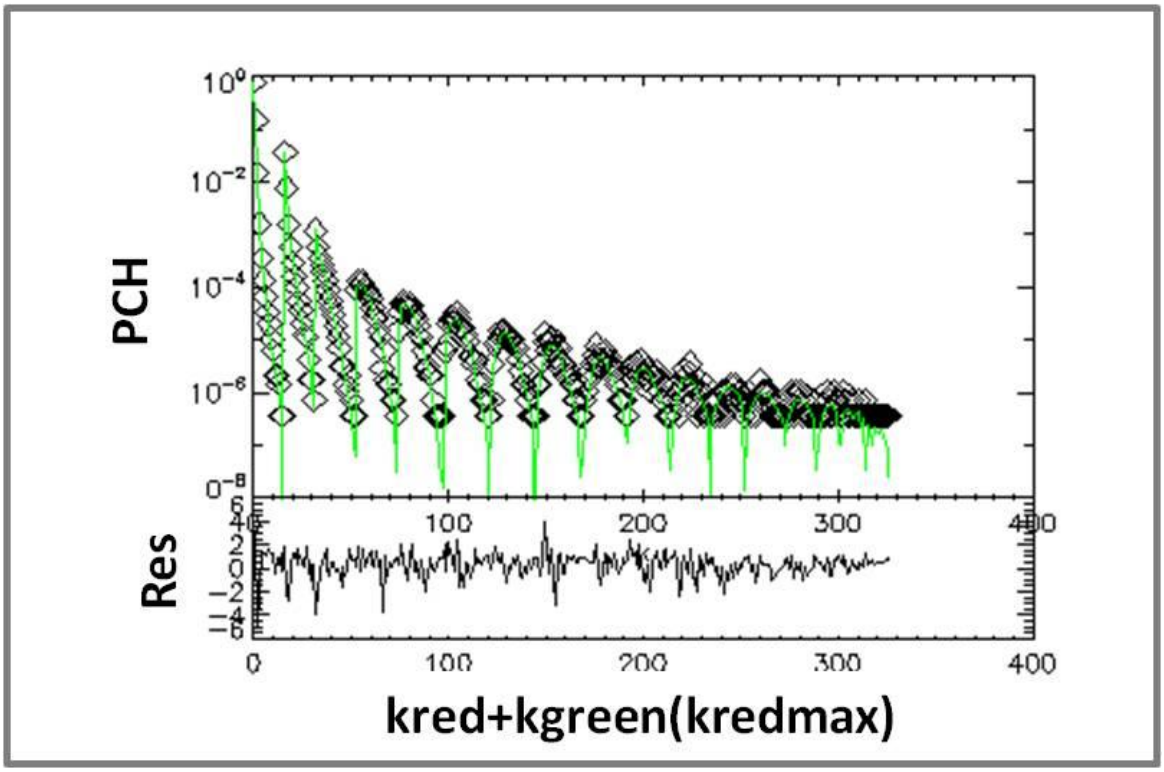


Figure 4.13 2D-PCH fit of VLPs transfected with EGFP and mCherry labeled Gag using Genjet reagent

The data was rebinned to 5 kHz to increase signal to noise. The fit returned a total of three brightness species (one background species and two VLP species with matching brightness ratios). The 2D-PCH data (symbols) and fit (green line) are shown above. The large undershoot observed between the data and its fit is due to the fit returning fractional counts of less than one.

In the analysis of VLPs containing red and green labels, we did not take the effects of FRET into account. With the close spacing of Gag in the HIV-1 Gag lattice, we expected a non-negligible amount of FRET. Indeed, when the sample was measured at 1040 nm, the Gag-mCherry copy number was lower than the copy number measured at 1000 nm. This indicates that some amount of FRET is occurring, because the brightness of the acceptor (mCherry) is higher in the presence of FRET. Thus, quantitative analysis of the IFP requires that FRET be taken into account. This can be accomplished by an additional fluorescence lifetime measurement of the donor. However, if we expect little to no FRET, the above analysis will be sufficient. We will return to this point at the end of the chapter.

4.6 A3G results

We have now developed the tools needed to tackle our biological model system: VLPs containing mCherry labeled Gag and EGFP labeled A3G. As discussed in the introduction, packaging of A3G into HIV-1 virions is necessary for viral restriction. VLPs containing A3G-EGFP and Gag-mCherry were created by transiently transfecting 293T cells with a 1:1 ratio of A3G-EGFP DNA to Gag-mCherry DNA using Genjet to minimize the range of cellular co-expression ratios. The supernatant from the cells was collected, purified, and measured as previously described. After the removal of statistically rare intensity peaks as described in Appendix A, we explored the heterogeneity of the A3G HIV-1 VLPs by studying the IFP shown in Figure 4.14a. The IFP reveals a narrow distribution of intensity ratios. The average intensity ratio is represented by the black line. The red line represents the intensity ratio for VLPs that contain only Gag-mCherry. Very few of the data points fall on the red line, indicating that most VLPs contain at least some A3G.

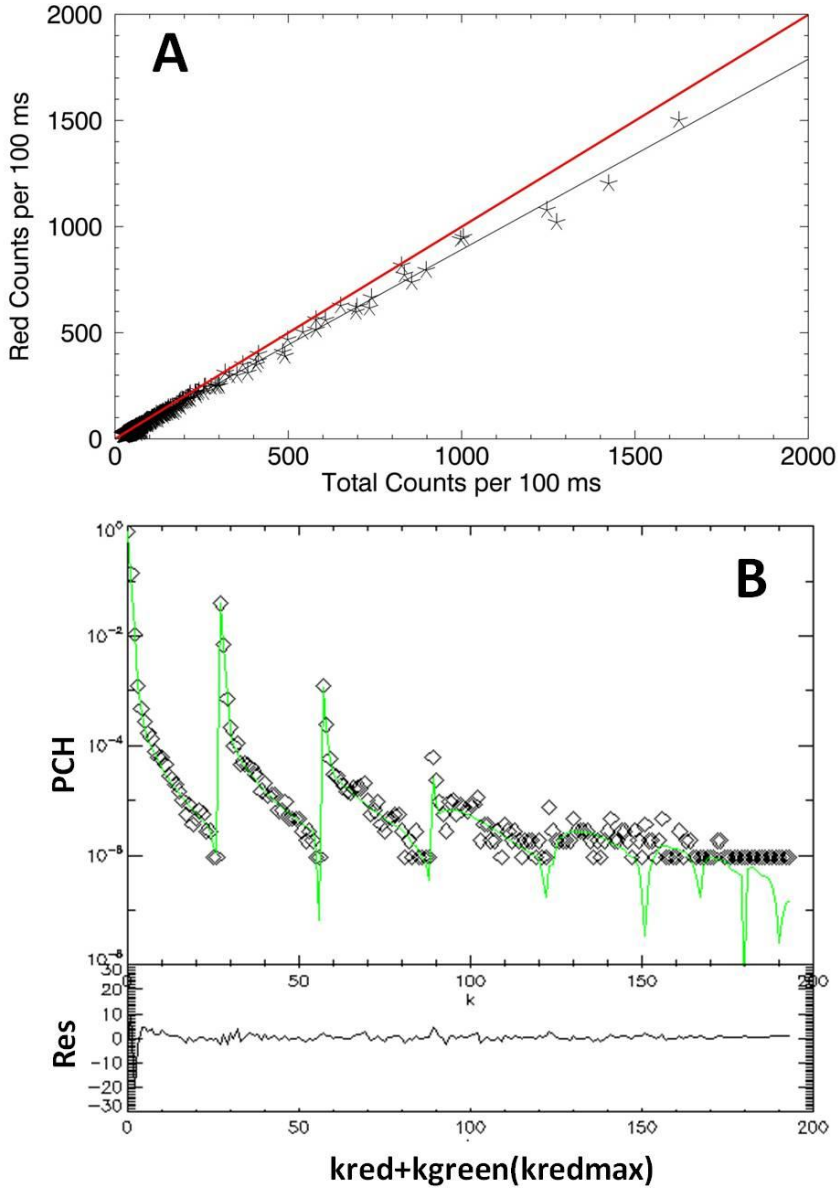


Figure 4.14 IFP and 2D-PCH analysis of VLPs containing Gag-mCherry and A3G-EGFP
(A) The IFP for VLPs produced from cells transfected with a 1:1 ratio of A3G-EGFP DNA to HIV-1 Gag-mCherry with Genjet. The black line is the average intensity fraction that matches the results from both species' 2D-PCH brightness fractions. The red line represents the intensity fraction expected for VLPs containing only Gag-mCherry. Most VLPs do not fall directly on the red line and instead contain at least some A3G-EGFP. **(B)** The 2D-PCH fit of the same data returns two VLP species with the same brightness fractions. The average A3G copy number is 89, and the reduced chi-squared is 4. The high chi-squared is due to high residuals when fitting data representing the background in the red only channels. This could be due to a population of red only background fluorescence that is difficult to account for during fitting.

Next, we turned to 2D-PCH analysis to determine the apparent A3G copy number per VLP. After the removal of statistically rare peaks as described in Appendix A, the raw data was fit using 2D-PCH. Due to the low signal-to-noise ratio in the green channel, it was necessary to rebin the raw data to 1 ms per bin to increase the signal. The data is fit to a two species model plus background with a chi-squared of 4 (Figure 4.14b). Fitting more species does not further reduce the chi-squared or result in a change of the apparent copy number. The residuals show large variations in the first channels, which mainly represent the background in the red channel. The exact cause of this misfit is still being investigated, but one possible explanation of the misfit is the presence of a red only fluorescent background species. Regardless of the high chi-squared, the analysis returns an apparent A3G copy number that is consistent across multiple samples. Both VLP species contain a small green brightness and a much higher red brightness. The brightness ratio varies slightly between the species, but species with lower green brightness also have lower red brightness. Potential FRET effects are discussed later in the chapter. The apparent green, or A3G, copy number is easily calculated using equation 4.1. For the 1:1 ratio of A3G to Gag DNA the copy number is 90 ± 20 based on 2D-PCH fit results. The red channel brightness is more difficult to calculate due to crosstalk of EGFP into the red channel. Using the measured intensity fraction from the monomer EGFP calibration and the copy number determined for A3G-EGFP, the apparent Gag copy number can be calculated to ~1000 ± 300 copies, which is within the range of Gag copy numbers previously reported. The average ratio of A3G to Gag is thus approximately 1:10.

4.6.1 A3G copy number depends on the A3G to Gag plasmid ratio

After demonstrating the incorporation of A3G into VLPs, we sought to investigate whether the incorporation of A3G depends on its cellular concentration. We decreased the ratio of A3G to Gag DNA plasmid used in transfection, which we predicted would result in lower cellular A3G concentration. This assumption needed to be confirmed by experiments measuring the cellular A3G concentration by FFS. The VLPs

for each DNA transfection ratio were measured and fit using 2D-PCH as described above. The apparent copy number for each ratio is shown in Figure 4.15. These results demonstrate that the amount of A3G incorporated depends on the DNA transfection ratio and likely reflects the accompanying changes in cellular A3G concentration.

The apparent Gag copy number did not vary much for the different ratios of DNA, while the A3G copy number varied by nearly a factor of 3. The A3G copy number went from 90 to 35 copies, while the Gag copy number went from 1000 to 1200, which is identical within experimental uncertainties. We speculate that using Genjet instead of Trasfectin as a transfection reagent is responsible for keeping the Gag copy number constant even while the amount of Gag DNA was varied during transfection. Although the mechanism by which Genjet achieves a constant Gag copy number is currently not understood, we utilized this transfection method because it reduced the heterogeneity of the copy number ratio. Since the amount of Gag was approximately constant, the ratio of A3G to Gag per VLP varied from approximately 1:10 to 1:40. The uncertainties are high, but it does appear that the average A3G:Gag ratio in VLPs can be changed by varying the DNA ratios used during transfection.

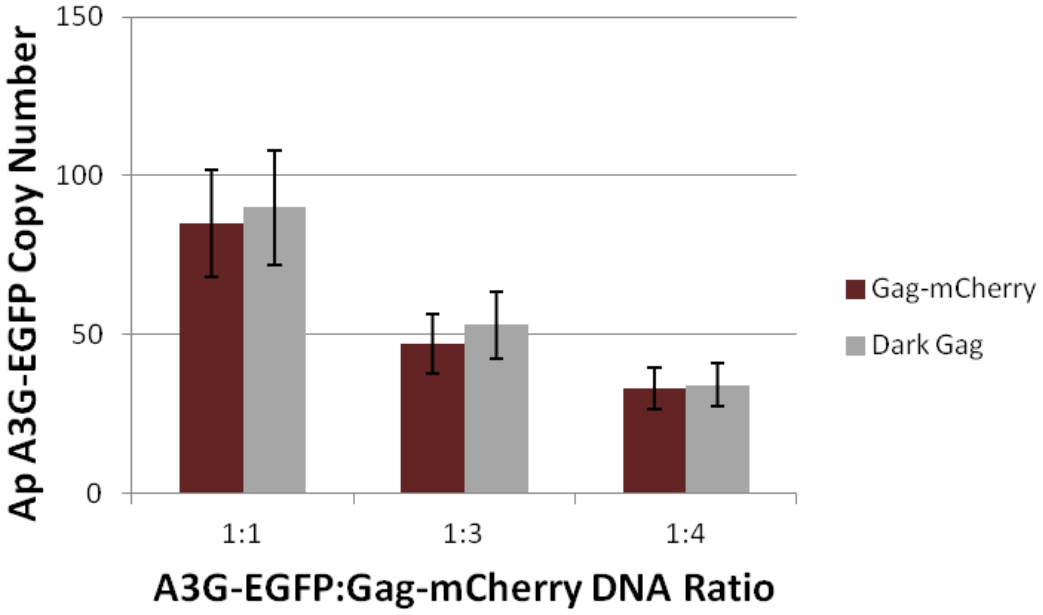


Figure 4.15 A3G copy number of VLPs
The ratio of A3G-EGFP to HIV-1 Gag-mCherry plasmid was varied from 1:1 to 1:4. The apparent copy number of A3G was determined from 2D-PCH analysis, and the uncertainty is based on the standard deviation of two independent experiments. The experiment was also conducted with unlabeled HIV-1 Gag to study the effects of FRET on determining the A3G copy number. The copy numbers are unchanged in the presence of dark Gag at every DNA ratio tested. This result suggests that FRET effects are negligible.

4.6.2 Dark Gag labeled A3G experiments

So far, we have ignored FRET effects in our analysis. To investigate the effects of FRET we performed side-by-side experiments with labeled and unlabeled Gag, and compared the A3G copy numbers. We prepared VLPs with A3G-EGFP and dark Gag with different DNA ratios as previously described. Next, we performed 1D-PCH analysis to determine the A3G copy number. 1D-PCH analysis required rebinning to 2 kHz, and 2 VLP species to describe the data. The grey bars in Figure 4.15 represent the average A3G copy numbers determined for VLPs containing unlabeled Gag. The A3G copy numbers for VLPs containing labeled and unlabeled Gag agree. This indicates that FRET effects are negligible. If FRET effects were not negligible, we would have expected an apparent reduction in the A3G copy number due to the energy transfer from the EGFP donor to the mCherry acceptor.

4.7 Discussion/Conclusion

Dual-color studies of VLPs rely on two differently colored protein labels with large differences in emission fluorescence and a predictable brightness for a large protein complex. EGFP and mCherry are currently the most promising pair, as their emission spectra have little overlap and they both have stable monomer brightnesses. We confirmed that the brightnesses of EGFP, EYFP and mCherry dimers behave as previously reported. Additionally, we demonstrated that all three labels produce VLPs with approximately the same Gag copy number. However, we observed that mCherry VLPs consistently have an apparent slightly lower copy number than EGFP or EYFP VLPs, and occasionally this difference was statistically significant. We hypothesize that hetero-FRET between the brightness species of adjacent mCherry labels within the VLP is responsible for this effect. This hypothesis needs to be confirmed with lifetime studies, but if proven true, this observation is further evidence of the need for a red fluorescent protein with a single brightness state. This would also eliminate the issues with the

precise determination of the dimer brightness when the protein configuration within the complex is unknown.

Next, we studied whether 2D-PCH is a suitable tool for characterizing VLP samples. Our analysis shows that 2D-PCH and 1D-PCH return the same brightness species and concentration for single color samples. We also successfully characterized a mixture of green VLPs and red VLPs by 2D-PCH with four brightness species, demonstrating the potential of 2D-PCH to study complex dual-color VLP samples. We then created a model system of single VLPs containing two color labels by mixing red- and green-labeled Gag DNA during transfection. This data could not be accurately fit with any 2D-PCH model, which we hypothesized was due to heterogeneity of the copy number ratio in the VLP sample. Because it was difficult to identify heterogeneity with the currently available analysis techniques, we developed the IFP.

IFP takes advantage of the unique nature of VLP samples. VLP samples contain rare but high intensity events that typically do not overlap in time. We discussed the proper choice of binning time that results in the best signal-to-noise ratio, while ensuring that most data points do not contain more than one complete VLP event. The IFP then allowed us to look at the intensity ratio of single VLPs, which can be related to the copy number ratio with 2D-PCH analysis and calibrations. The IFP for the VLPs produced using TransFectin reagent confirmed our hypothesis that these VLPs were very heterogeneous in their proportion of colored labels. This heterogeneity most likely mirrors the underlying heterogeneity observed in protein expression across the cell population. A previous study with TransFectin demonstrated that a small percentage of cells express only one protein or the other, which would explain the small number of single color VLPs in our sample (84). While the IFP did illuminate the heterogeneity of protein copy numbers, currently this method does not allow us to quantitatively determine protein copy numbers. Ideally, we would like to find a system with less heterogeneity so that we can use 2D-PCH to quantitatively determine copy numbers. We were able to achieve such a system by simply switching to a different transfection

reagent that based on our observations results in more homogenous cellular expression ratios. The IFP for VLPs produced using Genjet showed a very narrow distribution of intensity fractions. This experiment demonstrates the utility of IFP in quickly assessing the heterogeneity of the copy number ratio, which informs what data analysis methods are suitable and provides helpful feedback in selecting the best sample preparation methods.

We applied 2D-PCH to study the incorporation of A3G into HIV-1 VLPs. The results of these experiments show that the ratio of A3G to Gag depends on the DNA ratio used during transfection. The most likely explanation of this result is that varying the amount of A3G protein concentration to Gag protein concentration within a cell results in different amounts of A3G incorporation. However, our experiments need to be followed up with studies on how the DNA ratio affects cellular protein concentration for these two proteins. Additionally studying the interactions of these two proteins inside of cells will provide further hints for the mechanism of A3G incorporation into VLPs.

Initially we ignored the effects of FRET in quantifying the A3G copy number. Studies with unlabeled Gag resulted in A3G copy numbers that overlapped with the A3G copy number measured in the presence of labeled Gag. This result shows that FRET effects are negligible for our system. This observation has implications for the location and interaction of A3G within VLPs. FRET will occur if the EGFP and mCherry label are within approximately 5 nm of each other. If the model of individual A3G proteins interacting directly with the NC domain of Gag is correct (102), there should theoretically be a significant amount of FRET. The lack of observable FRET suggests at least two possible models for the location of A3G within HIV-1 VLPs. In the first model (Figure 4.16a), we assume that the A3G molecules are free and uniformly distributed throughout the available space within the VLP. The membrane is approximately 5 nm thick, the Gag lattice structure is approximately 30 nm thick (79), and the A3G must be more than 5 nm away from the mCherry label to avoid FRET (109). Additionally, the

diagram shows that half the membrane does not have a Gag lattice and would be accessible to the A3G proteins. With these assumptions, we expect that only ~5% of the A3G molecules are close enough to the Gag-mCherry to FRET, which is consistent with the results shown in Figure 4.15. Another possible model is shown in Figure 4.16b. In this model, the A3Gs are attached to pieces of RNA, much like beads on a string, and the RNA is not tied entirely to the Gag-NC. Here, we would also expect very little FRET, because only the first A3G on the RNA string would be close enough to mCherry for FRET. Our data cannot distinguish between these two models. However, future fluorescence polarization studies should allow us to distinguish between both scenarios.

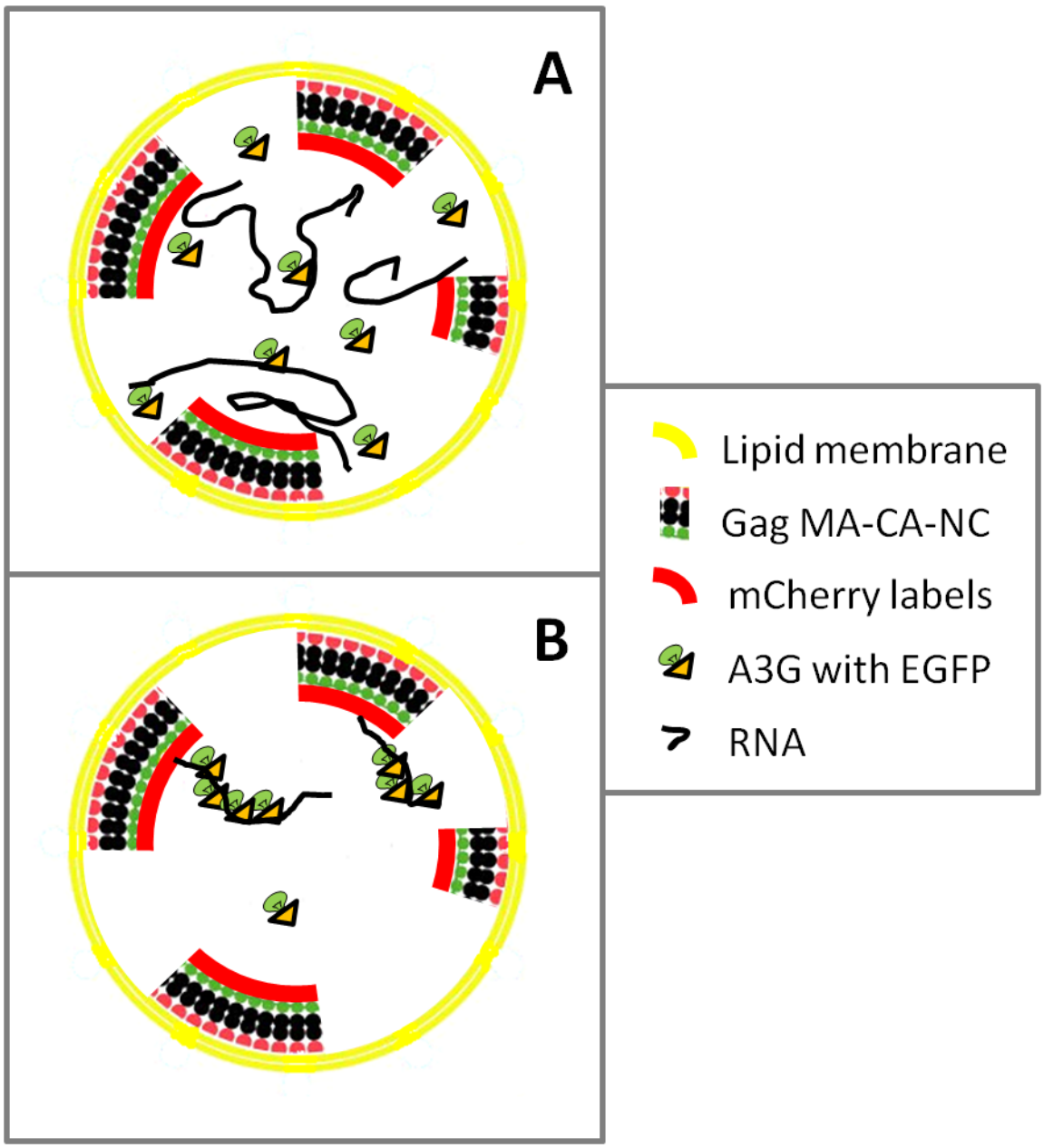


Figure 4.16 Model of A3G location in HIV-1 VLPs
The observation of negligible FRET effects suggests two possible models of the location of A3G within HIV-1 VLPs. The majority of the A3G must be more than 5 nm away from the mCherry attached to the NC domain of Gag. Two possible models could explain this observation. **(A)** The A3G could be randomly distributed throughout the entire VLP including sections near the membrane without Gag structure (**B)** The A3G could be attached to RNA like beads on a string. Only the A3G closest to the NC domain would experience FRET with the mCherry acceptor.

Chapter 5: Characterization of Brightness and Stoichiometry of Bright Particles by Flow-Fluorescence Fluctuation Spectroscopy

5.1 Introduction

Biological particles, such as viruses and vesicles, are relatively large and contain tens to thousands of proteins. These particles can be hard to study using FCS because they often exist at very low concentrations and diffuse slowly due to their large size. Thus particles are infrequently detected, which necessitates very long data acquisition times to build up the required statistics for FFS analysis. Often such experiments cannot be performed because the measurement time is exorbitant. Flowing or moving the sample during data acquisition is a simple method to significantly increase the event rate of particles passing through the observation volume (110). This chapter focuses on FFS measurements of flowing samples, which we refer to as flow-FFS. We further refer to FFS measurements on a resting sample as stationary-FFS to contrast the two methods. While the effect of flow on the autocorrelation function is well known (31), the influence of flow on PCH and brightness has not yet been investigated. Here, we characterize flow-FFS and demonstrate that brightness and concentration determined by PCH analysis are unaffected by flow speed as long as undersampling is avoided. We further provide guidance on optimizing flow-FFS through the proper selection of flow speed and sampling time.

Even in the presence of flow, FCS experiments at low concentrations remain challenging. The fluorescence background of samples at subnanomolar concentrations often overwhelms the contributions from the particles and consequently reduces the fluctuation amplitude of the autocorrelation function. Differentiating between the particle signal and the background based on the autocorrelation function of a flowing sample is not feasible. Remarkably, PCH, unlike autocorrelation analysis, has the

capability to separate the signal of sparse but bright particles from the background. This property of PCH is crucial for the quantitative characterization of the brightness and concentration of particles at very low concentrations. We characterize flow-FFS on fluorescent microspheres flowing through a microfluidic channel in concentrations ranging from 1 nM to 10 fM. Next, flow-FFS and stationary-FFS are applied to determine the stoichiometry of the HIV-1 Gag polyprotein within VLPs. We demonstrate that flow-FFS determines the stoichiometry from VLPs at lower concentrations and at significantly shorter data acquisition times than for stationary-FFS. This study establishes the utility of flow-FFS as a sensitive tool to gain quantitative information regarding the composition of complex macromolecular particles. Much of the material presented in this chapter has appeared in published form (26). This is first publication of the device fabrication section.

5.2 Material and methods

5.2.1 Experimental setup

The experiments were carried out on a modified two-photon microscope as previously described (Chapter 2) using two-photon excitation at a wavelength of 905 nm. The fluorescent microspheres and VLPs were measured with a Zeiss 63x C-Apochromat water immersion objective (NA = 1.2) with excitation powers ranging from 0.1-1.0 mW as measured after the objective. FFS data are acquired at sampling frequencies from 20-200 kHz, stored, and subsequently analyzed with programs written for IDL version 6.4 (RSI, Boulder, CO).

Stationary measurements were performed by loading 200 µL of solution into an 8-well Nunc Lab-Tek Chamber Slide (Thermo Fisher Scientific, Pittsburgh PA). Flow measurements were performed at the center of a 20 µm tall microfluidic channels with widths ranging from 20-200 µm. Constant pressure driven flow was achieved using a syringe pump (Kent Scientific Corporation, Torrington, CT). Flow velocities ranging from

2 – 44 mm/s were measured at the center of the channel. Fabrication of the microchannel devices is described below.

Internally labeled green fluorescent spheres (Duke Scientific, Fremont, CA) were suspended in a buffer as previously described (Chapter 4). VLPs were obtained from Cos-1 cells (ATCC, Manassas, VA) transfected with Gag-EYFP vector maintained in 10% fetal bovine serum and DMEM (without phenol red). Transfection was carried out using TransFectin according to the manufacturer's protocol (BioRad, Hercules, CA). Cells were maintained at 80% confluency on the day of transfection. For a standard VLP experiment, 2.1 μg of vector and 1.25 μg of TransFectin were mixed and added to a cell culture plate with a growth area of 25 cm^2. 32 hours after transfection the cell medium was collected and spun at 14000 rpm for 2 min to eliminate cell debris. A portion of non-concentrated VLP in cell supernatant was saved for future studies. The remaining solution was concentrated by a factor of 10 into Dulbecco's PBS (Biowhittaker, Walkersville, MD) using a Centricon filter (Millipore, Billerica, MA) at 14000 rpm.

5.2.2 Device fabrication

There are five basic steps in the device fabrication process; mask design, mold fabrication using soft photolithography, imprinting of channels in PDMS, oxidation and sealing of PDMS and glass, and finally tubing connection. The mask design was realized using ICED CAD software. The design file was then turned into a physical mask at the Nanofabrication Center (NFC) by NFC staff members.

The molds are fabricated on 4-inch silicon wafers that are precleaned in a piranha bath for 30 min at 120˚C (111). The wafers then undergo a dehydration bake at 200˚C for 10 min. The dehydration bake is crucial to remove all liquid and to prepare the wafer for processing. SU-8 (MicroChem), a negative, epoxy-type, near-UV photoresist (365 nm), was then spun on. The thickness of the SU-8 layer can be altered from 1-200 μm depending on the type of SU-8 used and the spin speeds. For the experiments described in this chapter, we use final spin speeds varying from 1000-3500 rpm and three formulations of SU-8: 10, 50, and 2010. The processing guidelines provided by

MicroChem were used as a starting point for determining the processing procedure. There are 4 basic steps in SU-8 processing. First, after the photoresist is spun on, the wafer is soft baked at 65˚C for 2 min and then 95˚C for 3 min. This step removes the solvent from the photoresist. In order to achieve good adhesion and crosslinking, it is crucial that most of the solvent is removed during the soft bake. As adhesion between the SU-8 and the silicon has been a problem, the soft bake time was generally increased by about 20% from the recommended times. The second step is exposure using UV light to link the exposed areas to the substrate. This is done on a Karl Suss MA6 Mask aligner. The exposure times range from 7-30 s depending on the thickness of SU-8 and the length of the soft bake. Thicker layers require longer exposure times, as do wafers that undergo a longer soft bake. The exposure time relative to the soft bake time is the most crucial factor affecting the final adhesion quality. The third step is the post exposure bake (PEB). In this step, the illuminated SU-8 is polymerized through a cationic photoamplification mechanism (112). Internal stress and cracking of the SU-8 has been an issue, so a longer (10 min), lower temperature (65˚C) PEB is used. The final step is development, which strips away the unexposed areas of SU-8. The development time is greatly dependent on solution agitation level and type of structures, and can range from 30 s to 5 min. In general, a long development time with gentle agitation produced the cleanest wafer with no SU-8 loss. After development is complete, the wafer is rinsed with isopropyl alcohol (IPA), then with deionized water, and finally dried using a stream of nitrogen gas. The mold is then complete and ready for use. Typically, a well made mold can be reused anywhere from 20-100 times before the adhesion between the SU-8 and substrate fails. Figure 5.1 outlines the photolithography process and shows a picture of a completed mask and mold.

After the mold has been fabricated, the next step is the imprint of the channels into PDMS. PDMS has two components, a base and a curing agent. When mixed, the silicon hydride groups present in the curing agent react with the vinyl groups present in

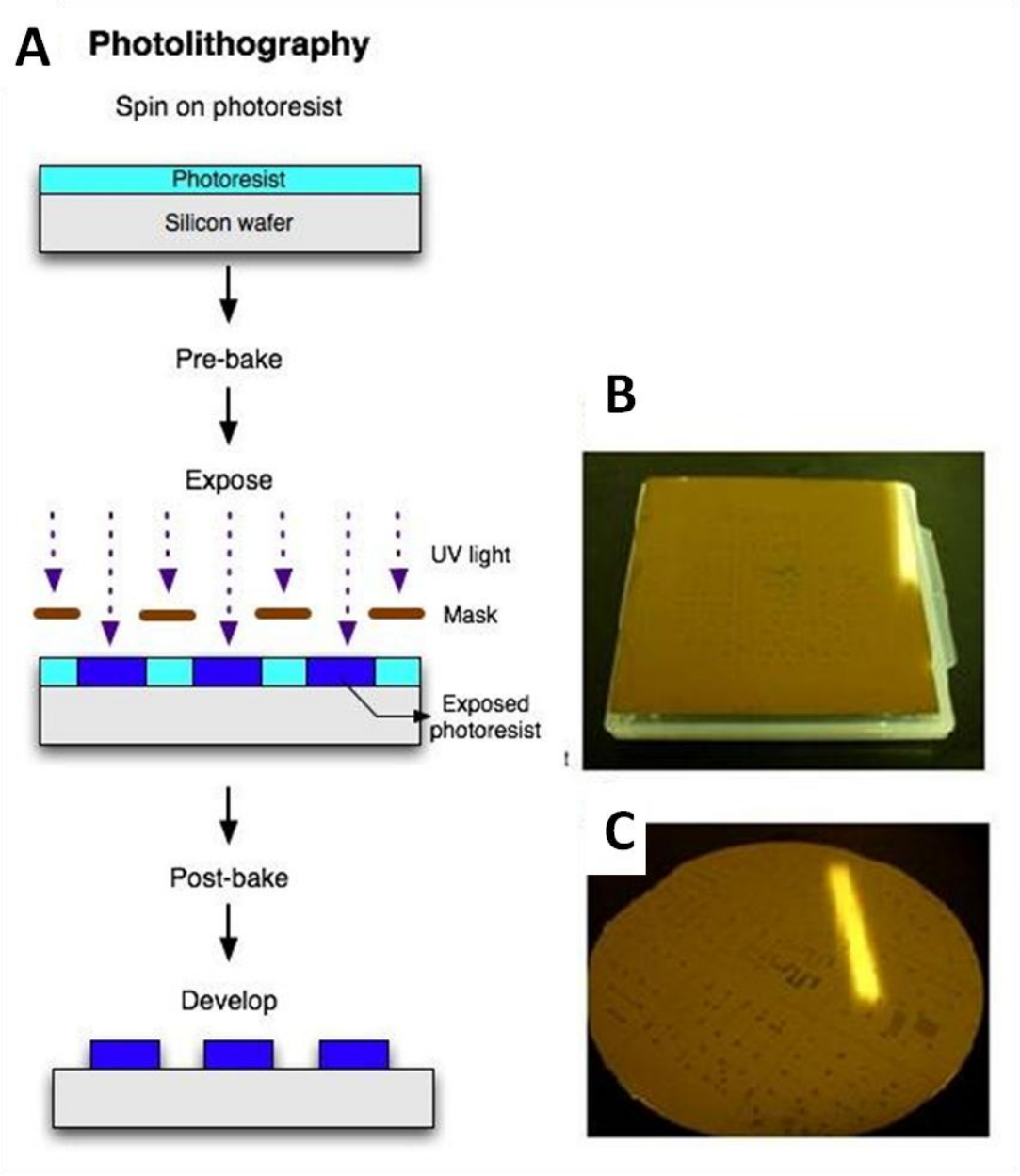


Figure 5.1 Overview of photolithography for microfluidic device fabrication
(A) Outline of photolithography process for creating molds used in microfluidic device fabrication.
(B) Photograph of mask used (on top of white box for contrast).
(C) Photograph of completed mold.

the base to form a cross-linked elastomeric solid (113). Typically, the recommended ratio of 10:1 base:curing agent was followed when mixing. Sometimes a 5:1 ratio was used to decrease the flexibility of the PDMS and encourage better bonding to the glass substrate, especially as the PDMS stock solutions aged. Increasing the curing agent also decreases non-specific protein absorption, which can be critical in applications studying 'dirty' biological samples (114). After mixing, the liquid polymer is poured over the mold, degassed and cured. We have found that no pretreatment of the mold is necessary to prevent sticking as the low surface free energy and elasticity of the PDMS allows for an easy release. PDMS can be cured at a wide range of temperatures, ranging from 48 hours at room temperature to 3 min at 150˚C. The temperature and length of curing time does have some effect on the elastic properties of the final polymer. We have found that curing overnight (8-12 hours) at 50˚C provides the best final product. This combination of time and temperature reduces mold damage, ensures complete curing before bonding, and provides the best elasticity of the PDMS for glass bonding. After curing, the PDMS is carefully peeled off the mold and cut into small sections.

Before bonding the glass and PDMS, both pieces are thoroughly cleaned by rinsing with methanol, IPA, acetone, and water, and then individually dried with nitrogen gas. This step is critical as even the smallest amount of surface contamination can cause catastrophic adhesion failure. We use Fisher Microscope Cover Glass 45 mm by 50 mm with thickness of 0.13-0.17 mm as the glass for bonding. Both the glass and PDMS pieces are placed in an ozone machine for 30 s to oxidize the surface by adding silanol (SiOH) groups to the surface. The ozone treatment is critical for creating permanent bonds. This step must be done in small batches because the glass and PDMS must be bonded within 30 sec of the ozone treatment to create a permanent bond. After the ozone treatment, the PDMS and glass are quickly brought into contact and carefully pressed together. This step must be completed within 30 s of ozone treatment to ensure long-term bonding. The device is then baked on a hotplate for 10 min at 75˚C

to make the bond semi-permanent. We have found that this bond will deteriorate with time and after 6-12 months is no longer reliable.

The final step in device fabrication is tubing connection and assembly. The basic procedure is to punch holes through the PDMS to the marked wells using a flat tipped 21-gauge needle that has been sharpened using sandpaper. If this procedure is done correctly, the PDMS core comes out with the needle and a small opening is left to insert the tubing. Tygon microbore tubing with an inner diameter of 0.25 mm is used. Tygon tubing is ideal for our applications because it is nearly chemically inert, flexible but still stiff enough to reduce kinking, and compatible with most sterilization techniques. Once the Tygon tubing is inserted into the PDMS holes, the PDMS conforms around it to form a fairly strong seal. The seal can be improved by pouring a small amount of uncured PDMS around the hole, or by gently heating the device. Once completed, the devices are stored protected from dust and remain viable for up to one year. A completed device is shown in Figure 5.2. It is possible to reuse channels for multiple experiments, but care must be taken to soak and rinse channels between uses. Additionally, the channels should be checked for fluorescent particle contamination using wide field fluorescent light before reuse.

5.3 Theory and data analysis

Data analysis was performed as previously described (Chapter 2 and 3). Here we summarize the analysis steps and highlight new theory. The autocorrelation function of stationary-FFS measurements was fit to a single species diffusion model for a 2D Gaussian geometry (Equation 2.8). This equation provided a sufficient approximation to determine the fluctuation amplitude $g(0)$ and the diffusion time. In the presence of uniform flow with velocity v_f, the autocorrelation function decays more quickly. FFS measurements of flowing samples were conducted at flow speeds where diffusion effects on the autocorrelation function are negligible, which we refer to as flow-dominant FFS experiments. In this limit, the autocorrelation function is approximated by

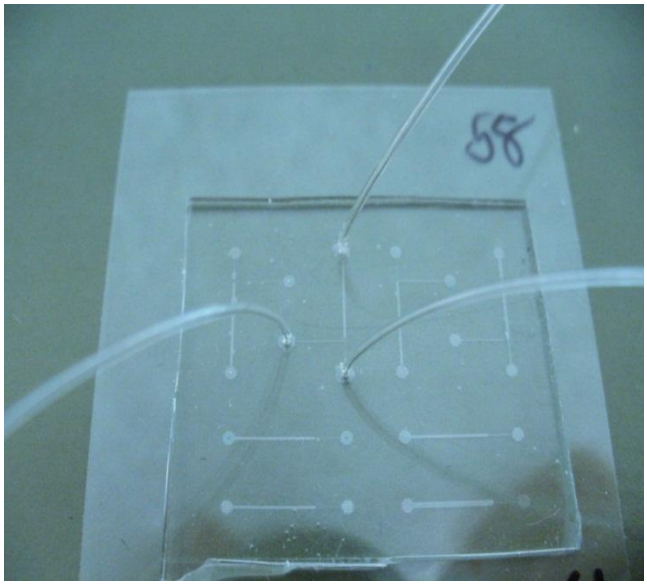


Figure 5.2 Completed microfluidic device
Picture of a completed microfluidic device with 2 input channels and a single output channel. Channel width is 100 µm, and channel height is 30 µm.

$$G_F(\tau) = g(0)exp\left[-\frac{(\tau/\tau_F)^2}{1+\tau/\tau_F}\right] = g(0)f_F(\tau) \quad (5.1)$$

The autocorrelation function was fit with Equation 5.1 to determine the flow time given by

$$\tau_F = w_0/(\sqrt{n}v_F) \quad (5.2)$$

and the flow speed v_F using $n = 2$ for two photon excitation. The flow time characterizes the time it takes to cross the observation volume by flow alone. Flow-dominated conditions were established by choosing flow times that are at least 50 times shorter than the diffusion time, $\tau_F/\tau_D \leq 0.02$.

The PCH function $p_k(k)$ is calculated by histogramming the photon counts *k*. The experimental PCH of flowing and stationary-FFS data is fit to an n-species PCH model with each species characterized by its brightness λ and occupation number N. Each PCH fit was corrected for deadtime and afterpulsing of the detector as previously described (Chapter 2). Error analysis was carried out as previously described (115). The concentration is calculated by

$$c = N/(N_A \cdot V_O) \quad (5.3)$$

using Avogadro's number N_A and the observation V_O determined by calibration measurements. The beam waist was calibrated from an FCS measurement of 26 nm diameter, internally labeled fluorescent spheres (Duke Scientific, Fremont, CA) in aqueous solution. The diffusion coefficient calculated from the Stokes-Einstein relation (Equation 2.10) was inserted into Equation 2.9 to convert the diffusion time into a beam waist of 0.38 ± 0.02 µm. The optical observation volume V_O was calibrated by PCH analysis of a dye solution with known concentration $c = 80$ nM using Equation 5.3. The calibration returned a sample volume of 0.08 – 0.15 fL depending on instrument parameters and the index of refraction of the measured sample. All channel measurements were performed at the center of the channel unless otherwise noted because this area has the smallest velocity gradient as a function of height. In order to determine the Gag copy number of VLPs tagged with yellow fluorescent protein (EYFP),

the brightness of monomeric EYFP, λ_{EYFP}, was determined as previously described (Chapter 3). The brightness λ of the VLP sample determined by PCH analysis was then converted into the normalized brightness.

The introduction of flow changes the rate of the intensity fluctuations, but the probability distribution and moments of the photon counts remain unchanged as long as undersampling is avoided. In this chapter, we ensure the absence of undersampling by choosing a sampling time T that is faster than the characteristic time scale of fluorescence intensity fluctuations of the sample (50). Thus, the shape of the autocorrelation function depends on flow, but the photon counting histogram (PCH) of the sample is independent of flow. Similarly, the fluctuation amplitude $g(0)$ of the autocorrelation function is independent of flow. A derivation of these theoretical results is found below.

5.3.1 The probability distribution function of fluorescence for diffusion and flow

An FFS experiment detects the fluorescence $F(t)$ of particles present within a small optical observation volume. The observation volume, or psf $O(\vec{r})$ is typically approximated by a 3-dimensional Gaussian (3DG) or by a squared Gaussian-Lorentzian (GL) function, which for convenience is normalized to one at the origin, $O(0) = 1$. The observation volume is defined by $V_O = \int O(\vec{r})d\Omega$. The instantaneous fluorescence $F(t)$ is time-dependent because the particles are free to move, which gives rise to fluctuations in the signal as they pass through the observation volume. Statistical analysis of FFS data requires that the process which gives rise to the fluctuations is stationary. By definition, a stationary process is a stochastic process whose probability distribution function (pdf) does not change when shifted in time or space. In other words, in order to apply FFS, the pdf $p(F)$ of the fluorescence has to be a constant function with respect to time and space. In addition, the autocorrelation function of a stationary process only depends on the time difference τ, not on the absolute time.

Diffusion is a stationary process with a time and space independent probability distribution function, $p_{\#}(n)$, of finding n particles at a small volume element located at $\vec{r}$. The pdf $p(F)$ of the fluorescence is

$$p(F) = \int_{\Omega} \delta\big(F - \lambda \cdot O(\vec{r})\big) \cdot p_{\#}(n) d\Omega \tag{5.4}$$

where λ is the brightness of a single particle in units of counts per second (cps), δ is Dirac's-delta function, and integration is carried out over the sample volume Ω. Note that flow corresponds to a translation of the sample in space. Because $p_{\#}(n)$ is independent of time and space, a translation in space does not affect the functional form of $p_{\#}(n)$. Consequently, the presence or absence of flow does not affect the pdf $p(F)$ of the fluorescence. Because $p(F)$ uniquely determines the moments of F, the variance and mean of the fluorescence are identical whether flow is present or not. Hence, the fluctuation amplitude $g(0) \equiv \langle \Delta F^2 \rangle / \langle F \rangle^2$ and Mandel's Q-parameter $Q \equiv \langle \Delta F^2 \rangle / \langle F \rangle$ are independent of flow.

5.3.2 Sampling time and flow

We previously pointed out that $p(F)$ is independent of the presence of flow. However, the detector records the fluorescence integrated over a finite sampling time T (50),

$$W(t) = \int_{t-T/2}^{t+T/2} F(t')dt' \tag{5.5}$$

If the sampling time T is faster than the time scale of fluorescence intensity fluctuations, then the integrated intensity W faithfully tracks the intensity F and Equation 5.5 is approximated by

$$W(t) = F(t)T \tag{5.6}$$

In this chapter, we choose experimental conditions where this approximation is valid. Because the experiment tracks the instantaneous fluorescence

$$F(t) = W(t)/T \tag{5.7}$$

the conclusion that the probability distribution and moments of *F* and W are independent of flow remains valid (32).

However, the rate of intensity fluctuations increases with the flow speed. Once the flow speed is sufficiently high, the approximation (Equation 5.6) breaks down. We refer to these conditions, where the sampling time T is insufficient to track $F(t)$, as undersampling. Under these circumstances, rapid fluctuations are integrated out and the statistical properties, such as the pdf and the moments of W, now explicitly depend on the flow speed. Thus, the flow speed v_F has to be kept below a critical value, which depends on the sampling time T and beam waist, in order to ensure the validity of Equation 5.6. We determine the critical flow speed by observing the onset of flow speed dependent changes in the pdf or moments of W.

5.3.3 Photon Counting Histogram (PCH) with flow

Until now, we ignored the fact that a photon counting detector records the number of photon counts k in the sampling time T instead of the integrated intensity W. Mandel's formula relates the pdf $p(W)$ of the integrated light intensity W absorbed by the detector to the pdf $p(k)$ of the photon counts k (49). We refer to $p(k)$ as the photon counting histogram. In the absence of undersampling, Mandel's formula determines PCH by (32)

$$p(k) = \int_0^\infty Poi(k, F \cdot T) p(F) dF \qquad (5.8)$$

where $Poi(k, x)$ is the Poisson distribution with an average photon count of k. Because $p(F)$ is independent of flow speed, the PCH $p(k)$ is also independent of flow speed as long as undersampling is avoided. Because the PCH determines the brightness λ and concentration N of each fluorescent species in the sample, these parameters are independent of flow speed in the absence of undersampling.

5.3.4 Event sampling of FFS

FFS relies on the statistics generated by particles crossing the observation volume. A sufficient number of particle events must be detected to ensure that the statistics of the sample are representative of the parent population. While the minimum number of events required for FFS analysis depends on a variety of experimental conditions, we have found empirically that a minimum of ~1000 particle events are needed for the analysis of HIV-1 VLPs (9). Fluorescent microspheres, on the other hand, only require ~100 events for FFS analysis as explained later. We now estimate the event rate dN/dt for a stationary sample ($v_F = 0$) by modeling the process as a diffusion limited reaction (116),

$$\frac{dN_D}{dt} \approx 4\pi \cdot R_O \cdot D \cdot c \cdot N_A \tag{5.9}$$

for particles at concentration *c* with diffusion coefficient *D*. The radius R_O is the characteristic radius of the observation volume. In the limit where flow is fast enough that diffusion effects can be neglected, we may derive the event rate by calculating the number of particles pushed through the cross section of the observation volume per time. For a flow speed v_F the event rate is

$$\frac{dN_F}{dt} \approx A_O v_F c N_A = \pi R_O^2 v_F c N_A \tag{5.10}$$

with $A_O = \pi R_O^2$ representing the cross-sectional area of the observation volume.

To illustrate the effect of flow on the event rate, consider HIV-1 VLPs with an average diameter of 130 nm. This size corresponds to a diffusion coefficient of 3.3 μm^2/s based on the Stokes-Einstein relation for an aqueous solution. The concentrated VLP samples with *c* ≈ 20 pM in our experiment lead to a stationary event rate of $dN/dt = 0.5$ events/sec as obtained by Equation 5.9. A flow rate of 2 mm/s increases the event rate to ~75 events/sec. Here the ratio dN_F/dN_D of the flow to stationary event rate equals 150. Because flow achieves a much higher event rate, it significantly decreases the minimum data acquisition time to collect a sufficient number of events.

For example, stationary VLP measurements require a measurement time of ~30 minutes to collect 1000 events, while flow of 2 mm/s reduces the time down to 10 s.

5.4 Results

5.4.1 Flow- versus stationary-FFS

To explore the influence of flow on FFS, we performed experiments on a 1.5 nM sample of fluorescent microspheres with a diameter of 100 nm. The sample was loaded into a microfluidic channel of width 50 μm and height 20 μm. FFS data were collected with the focus at the center of the microfluidic channel for 20 s at a sampling frequency of 200 kHz . The first FFS experiment was conducted with a flow speed set to 9 mm/s at the channel's center. At this speed, flow dominates diffusion and a flow-only model (Equation 5.1) is used to fit the autocorrelation function (Figure 5.3a). The flow time τ_F of 31 μs determined by the fit returns a flow velocity that agrees with the experimental value of 9 mm/s. Note that the microspheres are sufficiently small to neglect their finite size in FFS analysis (45). Next, we collected a second data set on the same sample inside the channel, but this time without flow. The autocorrelation function for the stationary sample is shown in Figure 5.3a together with a fit to a single species diffusion model. Using the calibrated beam waist the diffusion coefficient of the stationary spheres is 2.8 ± 0.2 $\mu m^2/s$, which agrees with the value of 2.8 $\mu m^2/s$ predicted by the Stokes-Einstein relation.

The PCH $p_k(k)$ of the flowing and stationary samples are identical within experimental uncertainty (Figure 5.3b). Thus, the brightness and concentration determined by PCH analysis are independent of flow. This result verifies that the probability distribution of the photon counts and its moments are independent of flow as discussed in the theory section. Consequently, the fluctuation amplitudes, $g(0)$, for the flowing and stationary sample also have to be identical as experimentally observed (Figure 5.3a).

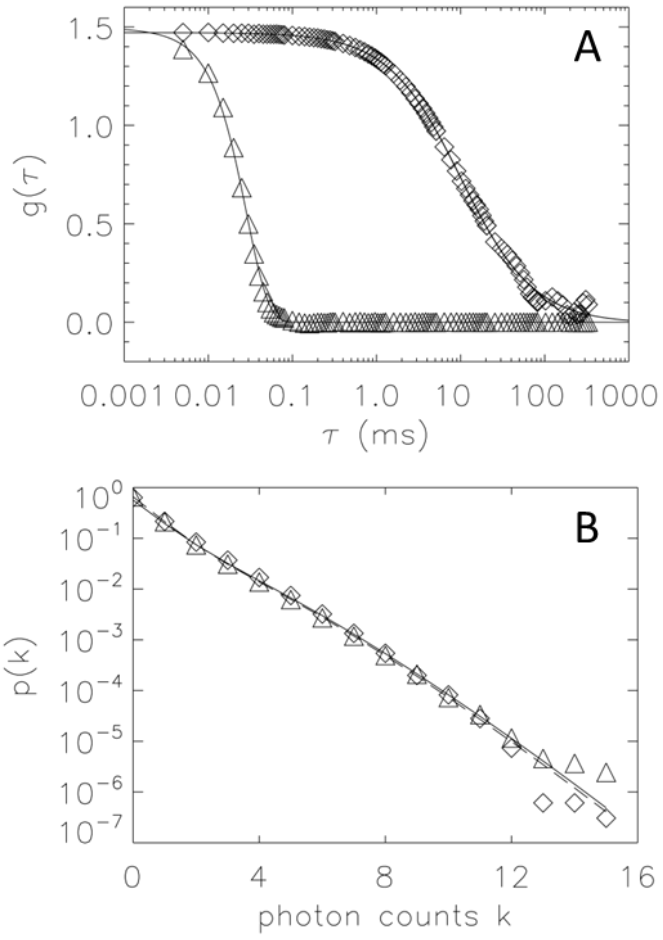


Figure 5.3 Comparison between flow- and stationary-FFS
A sample of 100 nm diameter spheres at a concentration of 1.5 nM was measured with and without flow at a sampling frequency of 200 kHz for 20 s. **(A)** The autocorrelation functions for stationary (diamonds) and flowing (triangles) samples are fit as explained in the text. A diffusion coefficient of 2.8 μm^2/s was determined for the spheres of the stationary sample. The flowing sample yielded a flow speed of 9 mm/s. The fluctuation amplitudes, $g(0)$, of both samples were identical and correspond to a concentration of 1.5 nM. A brightness of 9.2x10^5 cps was determined from $g(0)$ and the average intensity. **(B)** The PCHs of the stationary (diamonds) and flowing (triangles) sample fall on top of each other. Each PCH was fit to a single species model (χ^2 of 0.9 and 1.3). The brightness and concentration of the stationary and flowing samples agree and correspond to $c = 1.5 \pm 0.1$ nM and a brightness of (9.2 ± 0.6)x10^5 cps per sphere.

5.4.2 Event sampling for flow- and stationary-FFS

Fluorescence fluctuation experiments were carried out on a stationary sample with fluorescent microspheres (diameter = 100 nm) at low concentrations (110 fM). Data was collected for 160 s at 200 kHz. The intensity trace (Figure 5.4a) exhibits one spike indicating the passage of one sphere during the experiment. This result is consistent with an event rate of ~1/1800 s^{-1} as determined by Equation 5.9 with $D = 2.8$ μm^2/s, $c = 110$ fM, and $R_O \approx 1$ μm. Clearly, the number of events is insufficient for FFS analysis. It would take ~2 days of measurement time to acquire the minimum sampling population for FFS analysis (~100 events). By applying flow to the sample, the number of events greatly increases as seen by the intensity trace (Figure 5.4b). Data were collected for 160 s with a flow speed of 8.7 mm/s. The event rate for the flowing sample (Equation 5.10) is ~2 s^{-1}, which is more than a 1000-fold increase compared to the stationary sample. The data collected for 160 s contains more than 100 events, which is sufficient for FFS analysis. The autocorrelation and PCH curve are shown in Figure 5.5 and will be discussed further below. These experimental curves are reproducible when repeating the measurement (data not shown), which indicates that adequate sampling is achieved by the experiment.

5.4.3 The effect of background on FFS analysis

The autocorrelation function (Figure 5.5a) is fit to a flow-only model (solid line) with a flow time of 30 μs, which agrees with a flow speed of 8.7 mm/s. However, the value of the fluctuation amplitude, $g(0) \approx 0.7$, corresponds to a concentration of 3 nM, which is inconsistent with the expected value of 110 fM. We will address this discrepancy after discussing the PCH function. The PCH function (Figure 5.5b) displays two distinct slopes pointing towards the presence of two brightness species. One slope is steep and occurs at photon counts $k < 3$, which likely corresponds to a dim fluorescent background from the sample.

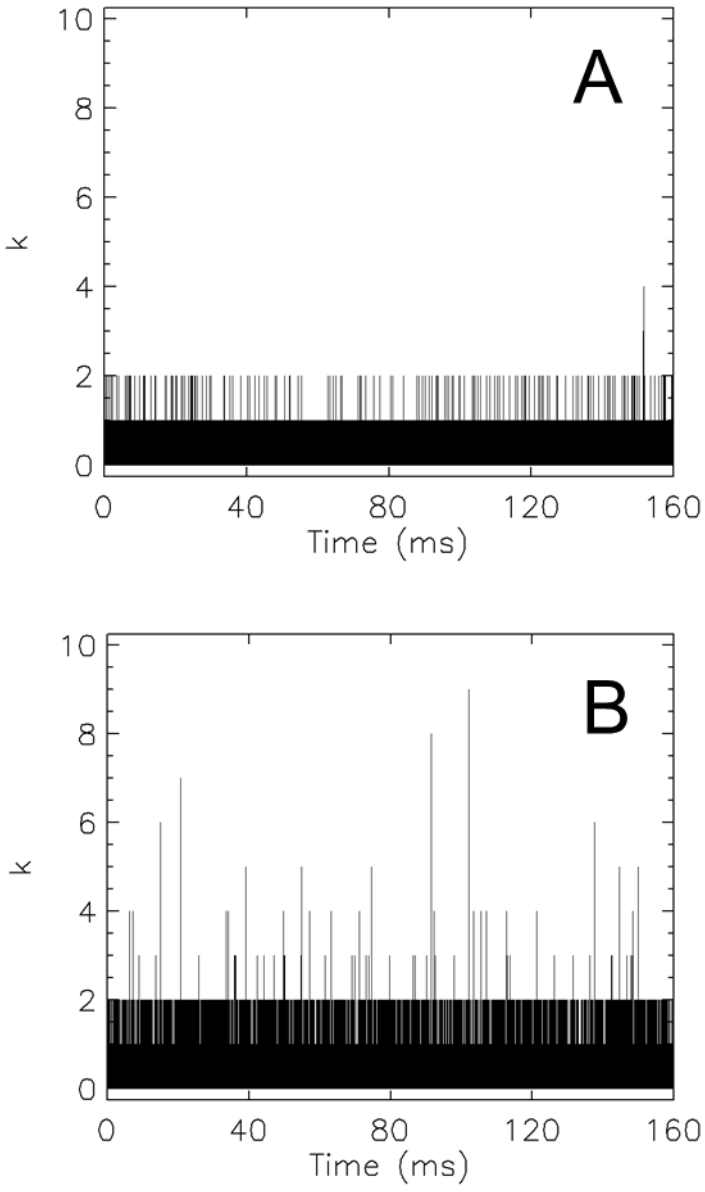


Figure 5.4 Intensity traces of a low concentration sample with and without flow
100 nm fluorescent spheres at a concentration of 110 fM were measured with and without flow for 160 s at a sampling frequency of 200 kHz.
(A) The intensity trace of a stationary sample showing a single event.
(B) The intensity trace of the same sample flowing at 8.7 mm/s. Flow provides a sufficient increase in the event rate to make FFS analysis feasible.

The second slope is less steep and extends to large photon count numbers as expected for bright particles. A single species fit fails to describe the PCH function, but a two-species PCH fit (reduced $\chi^2 = 0.9$) reproduces the experimental curve (Figure 5.5b). The first species has $\lambda_S = 1.03\text{x}10^6$ cps and $N_S = 9.28\text{x}10^{-6}$, which corresponds to a concentration of 107 fM, in good agreement with the expected value of 110 fM. The second species has a much lower brightness ($\lambda_B = 157$ cps) and a higher concentration ($N_B = 11.44$) than the first species and therefore describes the background.

Note that according to PCH analysis the intensity of the background ($F_B = \lambda_B N_B$) is ~1800 cps, while the fluorescent spheres only contribute an intensity ($F_S = \lambda_S N_S$) of ~10 cps to the fluorescence of the sample. We will use subscript B and S for the background and sample, respectively. Let us evaluate briefly the influence of background on FFS data. The autocorrelation function $g_F(\tau)$ of the sphere sample in the presence of background is (30, 115)

$$g_F(\tau) = \left(\frac{F_S}{F_S + F_B}\right)^2 g_S(0) f_F(\tau) + \left(\frac{F_B}{F_S + F_B}\right)^2 g_B(0) f_F(\tau) \qquad (5.11)$$

Without background, the fluctuation amplitude of the sample would be $g_F(\tau) = g_S(0)$.

The presence of a dim background leads to a suppression of the fluctuation amplitude, $g_F(\tau) < g_S(0)$. In fact, if we insert the PCH fit parameters into Equation 5.11 with $g_i(0) = \gamma_2 / N_i$, we get 0.6, which is in close agreement with the experimentally measured fluctuation amplitude of 0.7 (Figure 5.5a). In other words, background fluorescence is responsible for the reduction of the fluctuation amplitude, which leads to incorrect brightness and concentration values. Thus PCH analysis can be used to correct the influence of background on the $g(0)$ value.

Note that it is impossible to separate the background contribution from the sample by autocorrelation analysis, because temporal correlations induced by background and sample are identical for flow-dominated conditions. This point will be discussed in more detail later. While the temporal correlations are identical, the brightness of the background is much less than for the sample spheres.

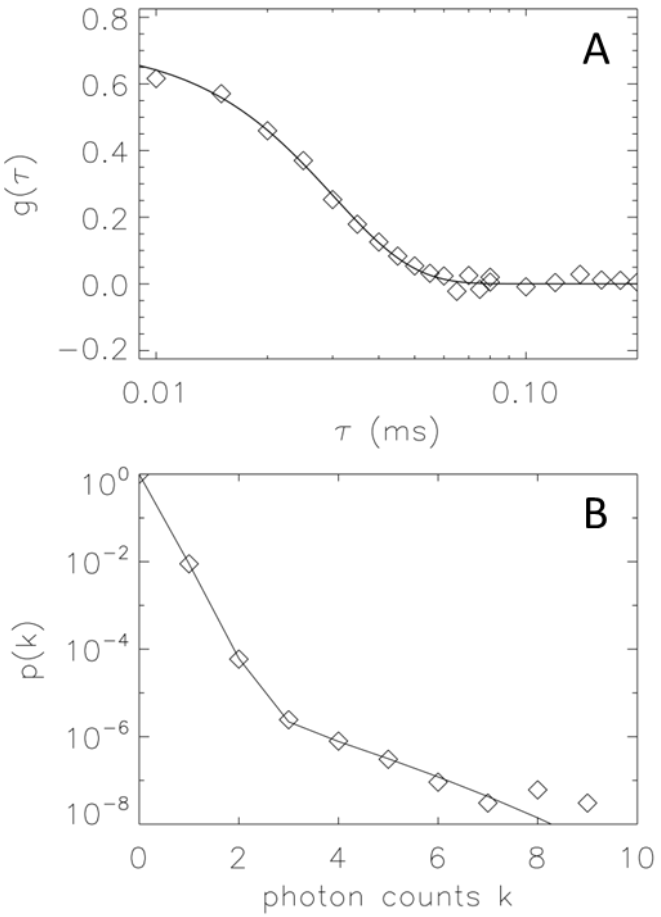


Figure 5.5 Flow-FFS of a low concentration sample
100 nm fluorescent spheres at a concentration of 110 fM were measured for 160 s at a flow speed of ~9 mm/s and a sampling frequency of 200 kHz. **(A)** The autocorrelation function of flowing spheres is shown together with the fit to the theoretical model. The flow time corresponds to a velocity of 8.7 mm/s which is in agreement with the expected value. However, the fluctuation amplitude $g(0)$ differs from the expected value by more than a factor of 1000, because background dominates. **(B)** PCH analysis of a flowing sample is plotted with a 2 species fit. The first species identifies the background, which has a very low brightness, contributes 99.3% of the counts, and is the main source for the PCH with $k \leq 3$. The second species models the rare but bright spheres, and is represented by the tail section of the PCH curve. The fit returns a brightness of ~10^6 cps per sphere and a concentration of 107 fM, which agrees with the expected value based on the dilution factor.

PCH exploits the brightness contrast to separate the background signal from the sample. Because background becomes a significant factor for ultralow sample concentrations, we will rely for the rest of the chapter on the remarkable ability of PCH to distinguish sample from background. We have found that the brightness of the background of all samples studied in this paper is sufficiently low, so that the PCH of the background is indistinguishable from a Poisson distribution. Thus, the experimental PCH is well approximated by convoluting the PCH from the particles with a Poisson distribution for the background.

5.4.4 The effect of flow velocity on brightness and concentration

As mentioned in the theory section the PCH, and therefore brightness and concentration, are not affected by flow speed as long as undersampling is avoided. Here we verify this prediction experimentally by varying the flow speed of a sample containing 100 nm fluorescent microspheres at 1.5 nM concentration. All measurements are performed at a sampling frequency of 200 kHz ($T = 5$ µs). The speed was varied by changing the flow rate of the syringe pump, and the flow velocity was measured using autocorrelation analysis. We also measured a stationary sample ($v_F = 0$). The brightness and concentration for each run was determined from a single species PCH fit, as the background contribution was negligible. Figure 5.6 demonstrates that brightness λ and occupation number N are constant for flow velocities ranging from 0 to 25 mm/s. However, undersampling leads to a reduction in the brightness and an increase in the concentration at a flow velocity of 44 mm/s. Thus, the critical flow speed demarcating the onset of undersampling effects is between 25 and 44 mm/s, which corresponds to a flow time of ~10 µs at a sampling time of T = 5 µs. Since undersampling depends on the flow and sampling time, their ratio provides a useful indicator for identifying undersampling conditions. Our experimental result shows that undersampling effects occur for $\tau_F/T < 2$. We verified that this relationship remains valid for sampling times other than $T = 5$ µs (data not shown). In other words, PCH accurately determines the brightness and concentration provided data is sampled at

least twice as fast as the flow time. All experiments presented in this chapter are conducted with $\tau_F/T \geq 2$.

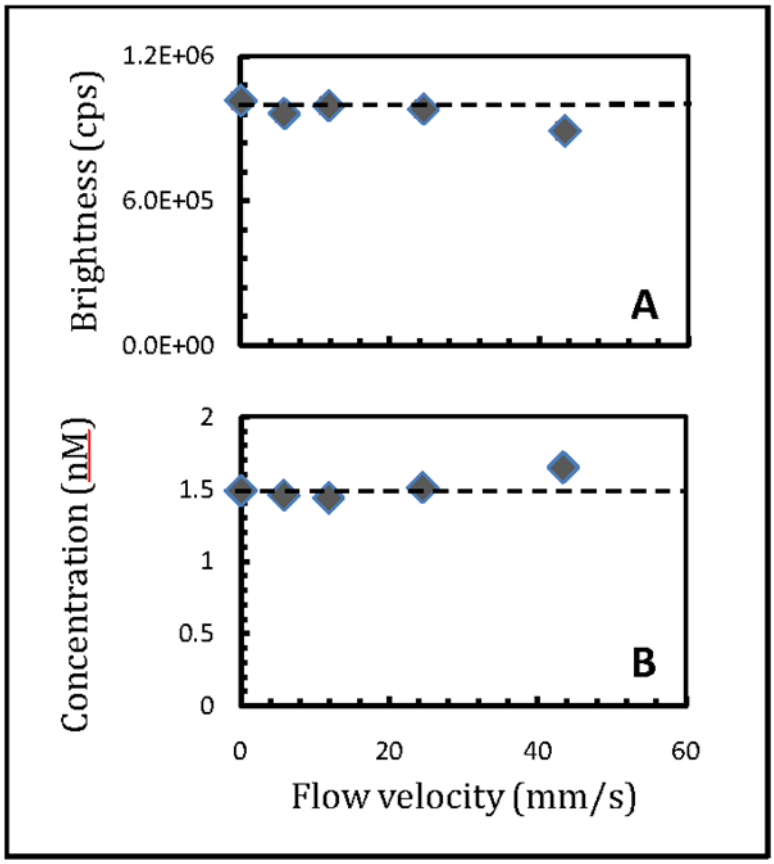


Figure 5.6 Brightness and concentration versus flow velocity
The velocity of a sample with 100-nm diameter fluorescent spheres at 1.5 nM was varied from 0 to 44 mm/s. Both the brightness **(A)** and concentration **(B)** determined by PCH analysis are independent of flow speed in the range from 0 to 25 mm/s. Undersampling is observed at a speed of 44 mm/s, which leads to an apparent increase in the particle concentration and a decrease in the brightness. The error was determined from the standard deviation of multiple measurements ($n = 4$) at each velocity. The symbol size is larger than the error.

5.4.5 Brightness versus concentrations of flowing particles

A dilution study was conducted on a flowing sample to evaluate the concentration ranges over which brightness and concentration are recovered by PCH analysis. A concentration of 100 nm spheres, which was initially 1.1 nM, was successively diluted by 10x between measurements. Each sample from 1.1 nM to 1.1 pM was measured at the same flow speed for 1 min. The 110 fM sample was measured for 3 min, and the 11 fM sample was measured for 10 min. The increased measurement time was necessary to collect a sufficient number of events for statistical accuracy. When necessary, the PCH was analyzed with a two-species model to separate the sample brightness and concentration from the background fluorescence. Figure 5.7a shows that brightness is independent of sphere concentration as expected. In addition, the fitted occupation number N_S is linear with concentration (Figure 5.7b). Experiments cover the concentration range from 1.1 nM down to 11 fM. While it is possible to measure at concentrations lower than 11 fM, the associated reduction in the event rate requires either longer measurement times or higher flow rates to ensure adequate sampling for statistical analysis of the data.

5.4.6 Brightness and Gag copy number of VLPs

In the past, our lab has prepared HIV-1 VLPs containing EYFP-tagged Gag and performed stationary-FFS measurements to determine the stoichiometry of the EYFP-labeled Gag within the VLPs (9, 25). These measurements present a significant challenge because of the low event rate due to low concentrations and slow diffusion. The measurement time required to collect a sufficient number of events ranged from 15-60 min. Applying flow-FFS should significantly shorten the measurement time and simultaneously increase the number of events.

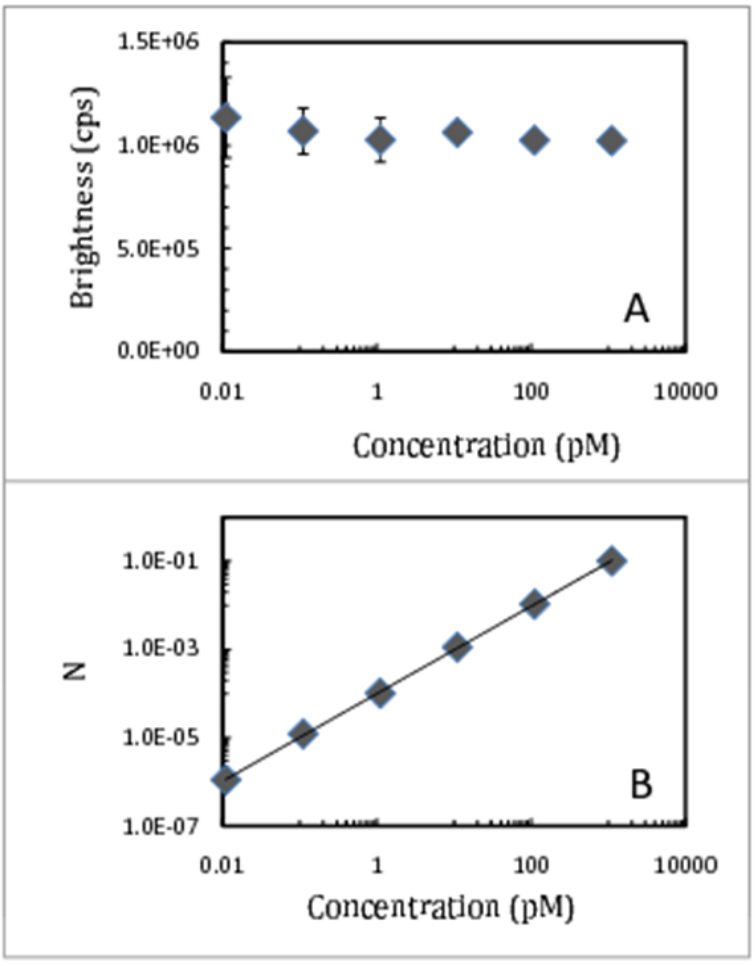


Figure 5.7 Serial dilution study of fluorescent spheres by flow-FFS
A sample of 100 nm fluorescent spheres at an initial concentration of 1.1 nM was successively diluted by factors of 10 and measured at a sampling frequency of 200 kHz and a flow velocity of ~5 mm/s. The measurement time was increased with decreasing concentration. Concentrations down to 1.1 pM were measured for 1 min. At 110 fM and 11 fM the sample was measured for 3 min and 10 min, respectively. PCH analysis determined the brightness and occupation number of the spheres.
(A) The brightness of the spheres remains constant.
(B) The occupation number *N* scales linearly with concentration as expected. The error was determined from the standard deviation of multiple measurements at each concentration ($n = 6$, except for 11 fM where $n = 3$). Error bars are not visible if the symbol size is larger than the error.

Flow experiments were performed on a 10x concentrated VLP sample. The brightness of monomeric EYFP was measured for calibration at the center of the channel as previously described (Chapter 3). The channel was then rinsed with PBS and the concentrated VLP solution was loaded. Pressure driven flow was used to obtain a flow rate of approximately 2 mm/s. This flow rate results in ~150 times increase in the number of events compared to stationary-FFS. Measurements were taken at 20 kHz for 80 s. Autocorrelation analysis was performed to determine the flow velocity in order to ensure that undersampling is avoided. The PCH of the data was fit to a single brightness species model with a Poissonian background. However, this model was insufficient to describe the data (reduced $\chi^2 > 20$), which indicated the presence of brightness heterogeneity of the VLP sample as was previously observed for stationary VLP samples. We added a second brightness species to the model, which resulted in a good description (reduced $\chi^2 < 2$) of the data (Figure 5.8). The majority of photon counts with $k < 3$ account for the background, which contributes ~90% of the signal, while the long tail at low $p(k)$ represents the VLP signal. Our result agrees with the published stationary-FFS study, which also identified two brightness species. The two brightness species were interpreted as an approximation of a brightness distribution of the VLP sample by the PCH fit. Thus, the two brightness species point towards a distribution of Gag copy numbers of the VLP sample. The previous study also observed that changing the amount of plasmid used in the transfection of cells resulted in brightness changes of the VLPs. A higher amount of plasmid resulted in increased brightness values for the two species, which reflects a change in the Gag copy number distribution. We confirmed this dependence of the two brightness species on the amount of transfected plasmid by flow-FFS (data not shown).

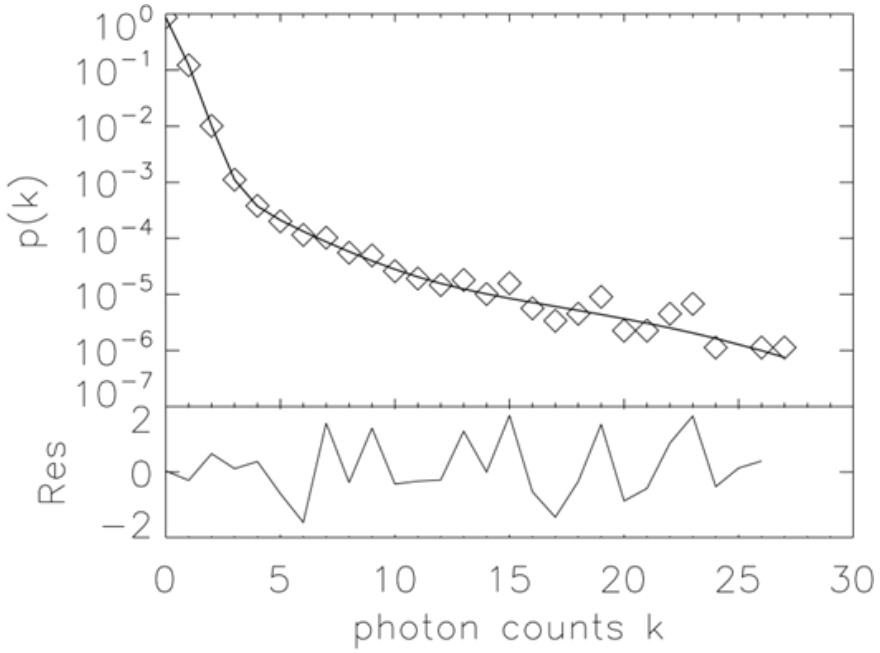


Figure 5.8 PCH of flowing VLPs
HIV-1 VLPs collected from cos-1 cells transfected with 2.1 µg of DNA and concentrated in PBS were measured at a flow velocity of 2.2 mm/s for 50 s at a sampling frequency of 20 kHz. The PCH (diamonds) and fit (line) to a 2-species model with Poissonian background are shown ($\chi^2 = 1.1$). The normalized residuals of the fit are shown in the lower panel. The background contributes 88% of the photon counts and dominates the PCH for $k \leq 3$. The VLP events are captured by the long tail of the PCH curve. The first brightness species corresponds to a Gag copy number of 1300 with a concentration of 20 pM. The second brightness species corresponds to a copy number of 3900 and a concentration of 3 pM.

Because the properties of the VLPs depend on the sample preparation conditions, a direct comparison between the current and the former study is difficult. We therefore performed stationary VLP measurements in addition to flow-FFS on the same sample in order to directly compare both methods. A 10x concentrated VLP sample was measured by flow-FFS with a flow speed of 2 mm/s for 80 s at 20 kHz. A two-species fit with Poissonian background describes the PCH data within experimental uncertainty. The uncertainty of the fit parameters was established by repeated measurements ($n = 5$) of the same sample. Gag copy numbers of the two fitted species are determined from the normalized brightness as described in the Chapter 3. Next, the same 10x concentrated sample was measured in the absence of flow for 30 min at 20 kHz. Autocorrelation analysis returned a diffusion time that corresponds to a VLP diameter of 130 ± 10 nm in agreement with previously reported numbers (9). PCH analysis required two brightness species with a Poissonian background to describe the experiment. The Gag copy numbers and concentration of the stationary and flow-FFS measurements agree with one another (Table 5.1).

The cell supernatant contains VLPs that are subsequently concentrated 10x into PBS. To test whether it is possible to avoid the time-consuming concentration step we measured the cell supernatant by flow-FFS. The cell medium of the supernatant presents a challenge as the background can be greatly increased from the concentrated sample. For the medium tested (DMEM without phenol red) the background intensity increased by approximately a factor of 3 from the concentrated sample. The VLPs in cell medium were measured under flow ($v_F = 2$ mm/s) at 20 kHz with a collection time of 160 s. Autocorrelation and PCH analysis were performed as described above. Again, a two-species model with Poissonian background was required to fit the PCH data. The copy numbers agreed with the values obtained for the 10x concentrated sample (Table 5.1). The concentrations of the original and concentrated samples differed by approximately a factor of 10, while retaining the same concentration relationship between the two VLP species. Measurements of non-concentrated VLPs in cell medium

were also performed for different types of cell medium including DMEM with phenol red. Although phenol red increases the background counts by a factor of 10, the copy numbers and concentration returned were consistent with measurements of samples concentrated into PBS. Taken together, our results indicate that characterization of VLPs by flow-FFS not only is feasible, but also offers significant advantages over stationary-FFS. VLPs can be measured directly in cell supernatant without concentrating the sample using much shorter data acquisition times than required for stationary-FFS.

Sample preparation	Acquisition Time (s)	Gag Copy #: Species 1	Gag Copy #: Species 2	Concentration Species 1 (pM)	Concentration Species 2 (pM)
10x concentrated, PBS, stationary	1800	1370 ± 400	4000 ± 1000	16 ± 4	3.0 ± 1.5
10x concentrated, PBS, flowing	50	1320 ± 100	4200 ± 400	18 ± 3	3.0 ± 0.5
1x concentrated, cell medium, flowing	160	1350 ± 200	4000 ± 600	2.0 ± 0.4	0.4 ± 0.2

Table 5.1 Gag copy number and concentration of VLPs
HIV-1 VLPs were collected from the supernatant of Cos-1 cells. The 10x concentrated sample was measured multiple times ($n = 5$) with and without flow. In addition, FFS measurements ($n = 5$) were also performed directly on the unconcentrated cell supernatant. FFS analysis determined the concentration and Gag copy numbers together with their standard deviations. The table also lists the data acquisition time. Flow greatly reduces the measurement time, while returning FFS results in agreement with stationary measurements.

5.5 Discussion

The experimental results establish that flow-FFS offers significant advantages over conventional FFS of bright biomolecular particles at low concentrations. Here, we briefly discuss the main factors that need to be considered in flow-FFS. Flow increases the event rate of particle detection, which is widely exploited for the measurement of samples at low concentrations. However, even in the presence of flow, FCS experiments of bright particles at low concentrations remain challenging, because of background fluorescence. Because flow-dominated FCS cannot distinguish particles from background, an independent FCS measurement of the background could potentially serve as a calibration, which may be used to correct the sample from the background effect. Unfortunately, even if it is possible to prepare a calibration sample for the background, this experimental strategy is from a practical point of view impossible. To illustrate the problem consider a sample with 1050 cps. If the background from the calibration sample is 1000 cps, then 50 cps are from the bright particles, which leads to a correction factor for $g(0)$ of 440 based on Equation 5.11. However, let us assume that the actual background of the sample is 1040 cps, which just differs from the calibration sample by 4%. The correction factor for $g(0)$ is now 11000. Thus, 4% uncertainty in the background signal leads to a 25-fold difference in estimating$g(0)$. Because of the inherent variability of biological sample preparations and the slight variations from measurement to measurement, it is clear that the required accuracy of the calibration approach cannot be achieved.

Because it is currently impossible to incorporate background into the analysis of sparse but bright particles, an alternative analysis was introduced that is based on fluorescent peak detection and uses a threshold to distinguish signal from background (110, 117, 118). While it has been shown that the particle concentration is proportional to the area under the peaks, quantitative modeling of the peaks is still under development and brightness analysis is not yet feasible.

This work introduces the first method capable of quantifying sparse but bright particles in the presence of background. PCH distinguishes background from particles by the shape of the photon count distribution. Because the theory of PCH is known, the concentration and brightness of the particles is determined from an analysis of the photon count distribution. It is further possible to detect heterogeneity in the brightness of the fluorescent particles using PCH as demonstrated for the VLP sample. Because PCH determines the background from a measurement of the actual sample, any uncertainty associated with a calibration measurement is avoided. PCH solves the problem associated with fluorescence background and thereby provides quantitative interpretation of $g(0)$, brightness, and concentration of bright particles at low concentrations.

Because we perform quantitative PCH, a minimum number of events are required for a statistically meaningful analysis. While it is difficult to pinpoint the exact number of events necessary, the following observations might provide a useful starting point. We have found that a minimum of ~100 events is suitable if the particles are of uniform brightness. However, in the presence of brightness heterogeneity, such as encountered for the VLPs, ~1000 events are needed to identify two brightness species. The data acquisition time is determined by dividing the number of events by the event rate, which is proportional to the particle concentration and the flow speed. Thus, in our case it takes about 5 min to acquire 100 events for a concentration of 10 fM at a flow speed of 20 mm/s.

It is tempting to increase the flow speed because it reduces the data acquisition time. However, this approach is not suitable in the presence of a significant amount of background counts. To illustrate the problem consider particles at a concentration of 1 pM ($N = 8\times10^{-5}$). Each particle carries 500 copies of EYFP, and each EYFP has a brightness of 300 cps. While the brightness of the particle is 150000 cps, the intensity of the particles is only ~10 cps. If the background is 1000 cps, the background constitutes 99% of the measured signal ($f_B = 0.99$). PCH needs to separate the signal of the

particles from the background. For this to occur the presence of the fluorescent particles has to add a component to the PCH that clearly distinguishes it from Poisson-background. Here, the sampling time becomes important, because deviation from the Poisson distribution depends on ε, which in the absence of undersampling is given by $\varepsilon = \lambda T$(50). The longest sampling time that avoids undersampling needs to be selected to maximize ε, which for flow is $T = \tau_F/2$. Thus, the maximum photon counts per particle depend on the flow speed ($\varepsilon = \lambda\tau_F/2$), which leads to a tradeoff between maximizing ε and minimizing the data acquisition time.

The significance of ε becomes apparent by continuing to examine our earlier example of particles with brightness of 150000 cps at a concentration of 1 pM with 1000 cps from background. A flow speed of 2 mm/s leads to an optimal sampling frequency of 20 kHz, which implies a mean of $\langle k \rangle = 0.05$ for the Poissonian background and ε = 7.5 for the particles. We modeled the PCH (squares) of this sample (Figure 5.9). The deviation from the Poisson of the background (dashed line) is evident, ensuring a clean identification of the particle properties by PCH analysis. Now consider a 10-fold increase in the flow speed. The sampling frequency increases to 200 kHz, which results in $\langle k \rangle = 0.005$ for the background and $\varepsilon = 0.75$ for the particles. The modeled PCH of the sample (diamonds) is now almost indistinguishable from the Poissonian background (dashed line). This example demonstrates the importance of the photon counts per sampling time, ε, for separating the particle signal from the background. The parameters chosen in this example are close to the actual parameters found in the VLP experiments, which explains why a flow speed of 2 mm/s was adopted for our measurements. The minimum value of ε required for characterizing particles depends on the background level, the particle concentration, and the data acquisition time. The appropriate value for ε is best discovered by performing the same type of modeling as discussed above (see Figure 5.9) using estimated parameters for the brightness and background counts of the system in question. As a rule of thumb, $\varepsilon > 1$ is typically required to distinguish the sample from the Poissonian background.

Although the average intensity of the flowing sample is very low, the passage of a particle results in a spike of the intensity. These transient intensities are high enough that deadtime and afterpulsing artifacts of the detector need to be taken into account in PCH analysis (64). In addition, the peak intensities may be high enough to saturate the detector, which in our case occurs at ~10^7 cps. In fact, experimental conditions where the VLPs and microspheres lead to saturation are easy to achieve, and we have encountered such cases in our work. Under saturating conditions, the PCH curve $p(k)$ exhibits a cutoff in k, which corresponds to the saturating intensity. In addition, the measured PCH deviates significantly from the model function. We avoid saturation by reducing the excitation power at the sample. In this chapter we choose conditions where the maximum photon count, $k_{Max} = F_{Limit}T$, observed in the PCH function corresponds to intensities of no more than $F_{Limit} =$ 2x10^6 cps. Because the PCH correction algorithms for deadtime and afterpulsing have previously been tested up to intensities of 2x10^6 cps (119), the PCH analysis of flowing particles is expected to be free of systematic artifacts. As an additional test, we measured a flowing sample as a function of excitation power and calculated its brightness and concentration by PCH analysis. For peak intensities up to 2x10^6 cps the brightness scaled with the square of the excitation power, while the concentration remained constant (data not shown). This result confirms that under our experimental conditions PCH analysis is well behaved.

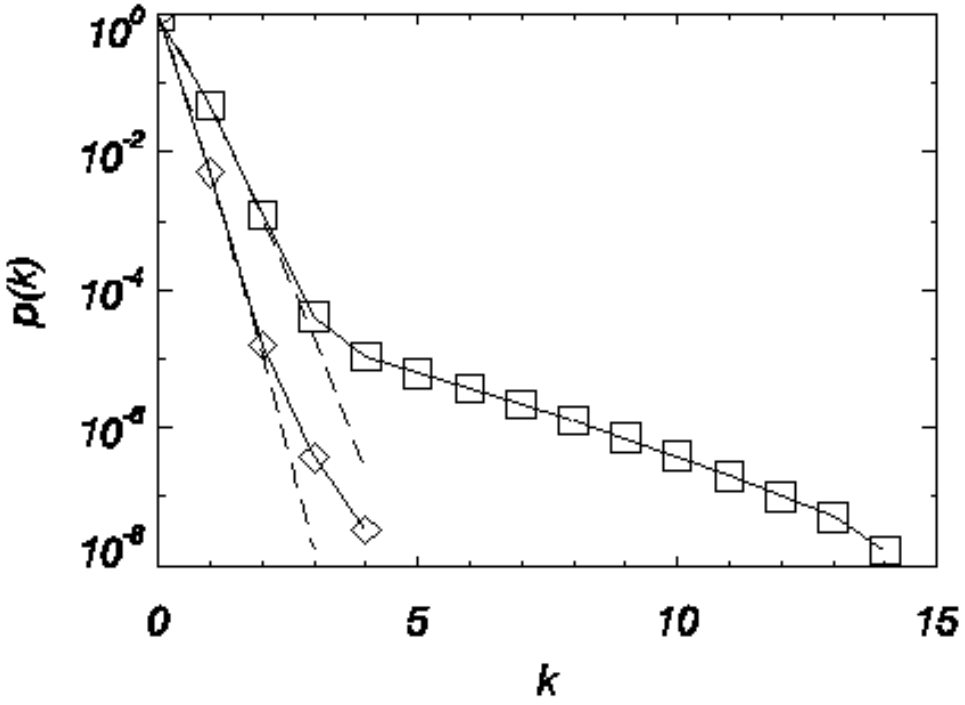


Figure 5.9 Modeled PCH curve at two sampling frequencies

The PCH for particles with a brightness of 1.5x10^{5} cps at 1 pM concentration ($N = 8\text{x}10^{-5}$) in the presence of a Poissonian background of 1000 cps were modeled for two different sampling frequencies. The PCH (squares) for a sampling frequency of 20 kHz provides a clear separation of the sample from the Poissonian background (dashed line). However, a sampling frequency of 200 kHz results in a PCH (diamonds) that is almost indistinguishable from the Poissonian background (dashed line). Thus, successful separation between sample and background by PCH requires careful consideration of the sampling frequency **(**for details, see text).

As mentioned earlier, flow together with fluorescence peak analysis has been used to detect bright particles at low concentrations and estimate their concentration. A distinct advantage of flow-FFS is the ability not only to measure the concentration, but also to provide information about the protein copy number of rare but bright particles. Our earlier study of HIV-1 VLPs by stationary-FFS established that the Gag copy number of VLPs is heterogeneous and depends on the sample preparation conditions. The reason for this complex behavior of Gag stoichiometry is not currently understood and requires further investigation. In the past, every VLP sample had to be concentrated 10-fold in order to reduce the minimum acquisition time required for PCH analysis to 30 minutes. The time-consuming procedure of preparing and measuring VLPs severely limits our ability to systematically investigate the factors that influence Gag stoichiometry. We have shown that flow-FFS provides the same information regarding Gag stoichiometry as stationary-FFS, but does so in less than 5 minutes of measurement time and without concentrating the sample. Thus, flow-FFS presents an attractive alternative for the future characterization of VLP stoichiometry. The only disadvantage of flow-FFS compared to stationary-FFS is the lack of information regarding diffusion, which is useful for determining the size of VLPs.

The highest Gag copy number achieved in the previous study of HIV-1 VLPs is ~2500, while a self assembly model predicts a maximum of 4000–5000 (15). In this work, we not only increased the statistical accuracy of the results, but also succeeded in measuring copy numbers that approach the theoretical limit. While this chapter focused on viral particles, flow-FFS is suitable for the characterization of any bright fluorescent particle, provided it contains a sufficient number of fluorescent labels. With the setup described in the chapter, ~50 EYFP-labeled proteins per particle are needed to separate signal from background at picomolar concentrations. Further technical improvements, such as decreasing the background using different sample purification methods, splitting the red fluorescent background into another channel, and experimenting with binning time have decreased the required amount of EYFP labels to approximately 20. Particle

size is another factor that needs to be considered. The VLPs are small enough that their finite size can be ignored in FFS analysis. However, particles that approach the size of the observation volume require analysis models that take size into account (45).

We have shown that flow-FFS accurately determines the brightness and concentration of particles at low concentrations. This study has successfully employed flow-FFS on concentrations as low as 10 fM. Furthermore, we have shown that measurement of rare but bright particles requires flow, which increases the event rate of particle detection, and PCH analysis, which separates the particle signal from background. This chapter contains detailed information on selecting experimental parameters and provides guidelines for conducting flow-FFS experiments. We emphasize that brightness is a unique parameter because it identifies the copy number of labeled proteins within a particle. Specifically, we determine the HIV-1 Gag copy number of VLPs at picomolar concentration. This work demonstrates that flow-FFS is a promising method for extracting quantitative information about the composition of large supramolecular complexes.

Chapter 6: Concluding Remarks

Understanding the complex process of retroviral assembly is a daunting task. Although many proteins are involved, the expression of Gag alone is sufficient to initiate the whole process. There are two main approaches for studying how Gag accomplishes this feat. One approach is based on cell studies that observe the relevant protein interactions in real time. Many labs including ours have taken this approach using a variety of methods. FFS is especially powerful, because it provides a quantitative characterization of protein interactions in cells. However, FFS in cells remains a challenging technique, because the cell environment leads to poor signal-to-noise and poor signal-to-background issues. An alternative approach to assess viral assembly is to study the assembled object and deduce information about the process. This second method was the focus of this work. In this dissertation, I developed and applied FFS methods to investigate VLPs, which provide a simplified model of the assembled immature virus.

We started by extending our previous FFS characterization of HIV-1 VLPs to the retrovirus HTLV-1. The initial goal was to find out if the results of our HIV-1 VLP study are applicable to other retroviruses. The first results from these experiments indicated that HTLV-1 VLPs were significantly different in size and Gag organization than observed for HIV-1 VLPs. Upon further investigation, it became clear that the differences were at least partially due to the fluorescent label interfering with viral assembly. Because biologically relevant FFS studies require a label that does not interfere with the assembly process, we performed experiments that ultimately identified a promising labeling strategy. Continuation with the HTLV-1 characterization will require additional work to optimize the linker between Gag and the fluorescent label.

Some of the results presented in Chapter 3 not only demonstrated label interference, but also led us to question a few model assumptions about the retroviral assembly process. Our results suggest that the production of VLPs is surprisingly robust to changes in the HTLV-1 Gag structure, does not depend on the formation of a Gag

lattice, and does not require RNA interactions. These observations imply that none of the current models properly identifies the driving force behind VLP assembly.

Complex biological particles such as viruses, ribosomes, and transport vesicles incorporate a mixture of biological molecules such as proteins, RNA, and lipids. Studying both the exact composition and relative ratios of the different biomolecules within these particles can provide us with hints about how each particle assembles. Specifically, studying the copy number ratio between multiple proteins could yield new insights. For instance, if two proteins form tight complexes with a high affinity, this will lead to a fixed copy number ratio between these proteins. Alternatively, the absence of correlation between the copy numbers suggests that the two proteins are not directly interacting. To prepare for measurements of two protein copy numbers we extended dual-color PCH to bright particles that contain two differently colored fluorescent proteins. While 2D-PCH successfully determined the average copy numbers in some cases, we encountered situations where 2D-PCH completely failed to fit the data. We developed IFP, a new method that takes advantage of the presence of bright but rare particles, to display the range of VLP intensity fractions. IFP identified that samples with a wide distribution of intensity fractions lead to failure of 2D-PCH analysis. Only samples with a narrow distribution are suitable for 2D-PCH analysis. In this particular study, we were able to narrow the range of intensity fractions by altering the transfection method. However, this solution is not a viable option for all studies, and future development of IFP into a rigorous and quantitative technique may offer an alternative method of quantifying the distribution of copy number ratios.

The potential for FRET in dual-color measurements present an additional complication that must be taken into account. Unfortunately, there exists no general theory to model FRET for complexes consisting of many proteins, because the close proximity of donors and acceptors leads to complex energy transfer pathways. We were able to show that for our A3G / Gag system FRET effects are minimal. However, future

theoretical and experimental work is required to address the effect of FRET in VLP particles to fully take advantage of dual-color FFS studies.

Because we observed almost no FRET, the majority of A3G and Gag molecules cannot be in close proximity. Additionally, our data suggest that the A3G copy number per VLP scales with the cellular A3G concentration, suggesting a weak association between Gag and A3G. Our group is currently performing FFS studies in the cytoplasm of cells to directly quantify this observation by measuring the association between Gag and A3G.

The dual-color FFS techniques described in this dissertation can be applied to other biological particles that are both bright and rare, such as vesicles. In fact, the work presented here inspired our lab to expand our research into cellular vesicles. Vesicles are similar in size to viruses and also contain a variety of proteins that coat the vesicle surface. Studying the recruitment and exchange of proteins associated with vesicles sheds light on vesicle recycling, trafficking, and other vesicle-related pathways.

VLP samples are typically low in concentration. Although FFS is a technique renowned for its sensitivity down to the single molecule level, the initial concentration of many VLP samples is even too low for FFS. Studying samples with similar characteristics to VLPs, such as authentic viral particles, exosomes, and cellular vesicles, also suffers from the same problem. Thus, we were forced to increase the concentration of the sample using the techniques discussed. While this concentration step did allow us to perform FFS studies of VLPs, there were significant drawbacks. The concentration step can be quite time consuming, and developing an appropriate technique for each sample is not trivial. Additionally, the concentration procedure can result in both sample loss and an increase in the fluorescence background. This fluorescence background is typically sufficiently severe that most techniques, including FCS, fail to characterize the sample.

We realized that the combination of flow and brightness analysis would both improve the event rate of a low-concentration sample and separate the signal from the

background. Because VLPs are so bright, there is a large contrast in brightness between any particles that contribute to background fluorescence and the VLP signal. PCH takes advantage of this contrast to cleanly separate the background from the signal. Note that this capability is a unique feature of PCH. The result of combining flow with FFS is a reduction of measurement time by an order of magnitude, and the elimination of the time consuming and non-trivial concentration of the sample. This advance is significant because the complexity of viral assembly requires experiments with a significant number of samples. Our results suggest that the combination of flow and PCH analysis is attractive for a wide range of bright but rare particles, such as vesicles and other viruses.

References

1. Breitbart, M., and F. Rohwer. 2005. Here a virus, there a virus, everywhere the same virus? Trends in Microbiology. 13: 278–284.

2. Allan, G.M., and J.A. Ellis. 2000. Porcine Circoviruses: A Review. J VET Diagn Invest. 12: 3–14.

3. Fraenkel-Conrat, H., and R.C. Williams. 1955. Reconstitution of Active Tobacco Mosaic Virus from Its Inactive Protein and Nucleic Acid Components. PNAS. 41: 690–698.

4. Cairns, J. 1992. Phage and the Origins of Molecular Biology. CSHL Press.

5. Berland, K.M., P.T. So, and E. Gratton. 1995. Two-photon fluorescence correlation spectroscopy: method and application to the intracellular environment. Biophys J. 68: 694–701.

6. Qian, H., and E.L. Elson. 1991. Analysis of confocal laser-microscope optics for 3-D fluorescence correlation spectroscopy. Appl. Opt. 30: 1185–1195.

7. Webb, W.W. 2001. Fluorescence Correlation Spectroscopy: Inception, Biophysical Experimentations, and Prospectus. Appl. Opt. 40: 3969–3983.

8. Chen, Y., L.-N. Wei, and J.D. Müller. 2003. Probing protein oligomerization in living cells with fluorescence fluctuation spectroscopy. Proc Natl Acad Sci U S A. 100: 15492–15497.

9. Chen, Y., B. Wu, K. Musier-Forsyth, L.M. Mansky, and J.D. Mueller. 2009. Fluorescence Fluctuation Spectroscopy on Viral-Like Particles Reveals Variable Gag Stoichiometry. Biophysical Journal. 96: 1961–1969.

10. Göttlinger, H.G. 2001. The HIV-1 assembly machine. AIDS. 15 Suppl 5: S13–20.

11. Morita, E. 2012. Differential requirements of mammalian ESCRTs in multivesicular body formation, virus budding and cell division. FEBS J. 279: 1399–1406.

12. Balasubramaniam, M., and E.O. Freed. 2011. New Insights into HIV Assembly and Trafficking. Physiology. 26: 236–251.

13. Briggs, J.A.G., and H.-G. Kräusslich. 2011. The Molecular Architecture of HIV. Journal of Molecular Biology. 410: 491–500.

14. Sundquist, W.I., and H.-G. Kräusslich. 2012. HIV-1 Assembly, Budding, and Maturation. Cold Spring Harb Perspect Med. .

15. Briggs, J.A.G., M.N. Simon, I. Gross, H.-G. Kräusslich, S.D. Fuller, et al. 2004. The stoichiometry of Gag protein in HIV-1. Nature Structural & Molecular Biology. 11: 672–675.

16. Campbell, S., and V.M. Vogt. 1995. Self-assembly in vitro of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. J. Virol. 69: 6487–6497.

17. Briggs, J. a. G., J.D. Riches, B. Glass, V. Bartonova, G. Zanetti, et al. 2009. Structure and assembly of immature HIV. PNAS. 106: 11090–11095.

18. Carlson, L.-A., J.A.G. Briggs, B. Glass, J.D. Riches, M.N. Simon, et al. 2008. Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. Cell Host Microbe. 4: 592–599.

19. Fischlechner, M., and E. Donath. 2007. Viruses as Building Blocks for Materials and Devices. Angewandte Chemie International Edition. 46: 3184–3193.

20. Lingwood, D., and K. Simons. 2010. Lipid rafts as a membrane-organizing principle. Science. 327: 46–50.

21. Hogue, I.B., G.N. Llewellyn, and A. Ono. 2012. Dynamic Association between HIV-1 Gag and Membrane Domains. Mol Biol Int. 2012.

22. Nydegger, S., S. Khurana, D.N. Krementsov, M. Foti, and M. Thali. 2006. Mapping of tetraspanin-enriched microdomains that can function as gateways for HIV-1. J Cell Biol. 173: 795–807.

23. Ono, A., and E.O. Freed. 2001. Plasma membrane rafts play a critical role in HIV-1 assembly and release. Proc. Natl. Acad. Sci. U.S.A. 98: 13925–13930.

24. Ono, A., and E.O. Freed. 2005. Role of lipid rafts in virus replication. Adv. Virus Res. 64: 311–358.

25. Grigsby, I.F., W. Zhang, J.L. Johnson, K.H. Fogarty, Y. Chen, et al. 2010. Biophysical analysis of HTLV-1 particles reveals novel insights into particle morphology and Gag stoichiometry. Retrovirology. 7: 75.

26. Johnson, J., Y. Chen, and J.D. Mueller. 2010. Characterization of Brightness and Stoichiometry of Bright Particles by Flow-Fluorescence Fluctuation Spectroscopy. Biophysical Journal. 99: 3084–3092.

27. Elson, E.L., and D. Magde. 1974. Fluorescence correlation spectroscopy. I. Conceptual basis and theory. Biopolymers. 13: 1–27.

28. Thompson, N.L., A.M. Lieto, and N.W. Allen. 2002. Recent advances in fluorescence correlation spectroscopy. Current Opinion in Structural Biology. 12: 634–641.

29. Van Orden, A., K. Fogarty, and J. Jung. 2004. Fluorescence Fluctuation Spectroscopy: A Coming of Age Story. Appl. Spectrosc. 58: 122A–137A.

30. Thompson, N. 2002. Fluorescence Correlation Spectroscopy. In: Lakowicz J, CD Geddes, JR Lakowicz, editors. Topics in Fluorescence Spectroscopy. Springer US. pp. 337–378.

31. Magde, D., W.W. Webb, and E.L. Elson. 1978. Fluorescence correlation spectroscopy. III. Uniform translation and laminar flow. Biopolymers. 17: 361–376.

32. Chen, Y., J.D. Müller, P.T.C. So, and E. Gratton. 1999. The Photon Counting Histogram in Fluorescence Fluctuation Spectroscopy. Biophysical Journal. 77: 553–567.

33. Kask, P., K. Palo, D. Ullmann, and K. Gall. 1999. Fluorescence-intensity distribution analysis and its application in biomolecular detection technology. Proc.Natl.Acad.Sci.U.S.A. 96: 13756–13761.

34. Chen, Y., J.D. Müller, Q. Ruan, and E. Gratton. 2002. Molecular brightness characterization of EGFP in vivo by fluorescence fluctuation spectroscopy. Biophys J. 82: 133–144.

35. Chen, Y., and J.D. Müller. 2007. Determining the stoichiometry of protein heterocomplexes in living cells with fluorescence fluctuation spectroscopy. Proc Natl Acad Sci U S A. 104: 3147–3152.

36. Diaspro, A., G. Chirico, and M. Collini. 2005. Two-photon fluorescence excitation and related techniques in biological microscopy. Q. Rev. Biophys. 38: 97–166.

37. Murray, J.M. 2007. Practical Aspects of Quantitative Confocal Microscopy. In: Digital Microscopy, 3rd Edition. Academic Press. pp. 467–478.

38. Göppert-Mayer, M. 1931. Über Elementarakte mit zwei Quantensprüngen. Annalen der Physik. 401: 273–294.

39. Denk, W., J.H. Strickler, and W.W. Webb. 1990. Two-photon laser scanning fluorescence microscopy. Science. 248: 73–76.

40. Xu, C., and W. Webb. 2002. Multiphoton Excitation of Molecular Fluorophores and Nonlinear Laser Microscopy. In: Lakowicz J, CD Geddes, JR Lakowicz, editors. Topics in Fluorescence Spectroscopy. Springer US. pp. 471–540.

41. Macdonald, P.J., Y. Chen, X. Wang, Y. Chen, and J.D. Mueller. 2010. Brightness Analysis by Z-Scan Fluorescence Fluctuation Spectroscopy for the Study of Protein Interactions within Living Cells. Biophysical Journal. 99: 979–988.

42. So, P.T.C., C.Y. Dong, B.R. Masters, and K.M. Berland. 2000. Two-Photon Excitation Fluorescence Microscopy. Annual Review of Biomedical Engineering. 2: 399–429.

43. Meseth, U., T. Wohland, R. Rigler, and H. Vogel. 1999. Resolution of fluorescence correlation measurements. Biophys J. 76: 1619–1631.

44. Magde, D., E. Elson, and W.W. Webb. 1972. Thermodynamic Fluctuations in a Reacting System—Measurement by Fluorescence Correlation Spectroscopy. Phys. Rev. Lett. 29: 705–708.

45. Wu, B., Y. Chen, and J.D. Müller. 2008. Fluorescence Correlation Spectroscopy of Finite-Sized Particles. Biophysical Journal. 94: 2800–2808.

46. Qian, H., and E.L. Elson. 1990. On the analysis of high order moments of fluorescence fluctuations. Biophys. J. 57: 375–380.

47. Qian, H., and E.L. Elson. 1990. Distribution of molecular aggregation by analysis of fluctuation moments. Proc. Natl. Acad. Sci. U.S.A. 87: 5479–5483.

48. Saleh, B. 1978. Photoelectron statistics with applications to spectroscopy and optical communication. Springer Series in Optical Sciences, Berlin: Springer, 1978. -1.

49. Mandel, L. 1958. Fluctuations of Photon Beams and their Correlations. Proceedings of the Physical Society. 72: 1037–1048.

50. Müller, J.D. 2004. Cumulant Analysis in Fluorescence Fluctuation Spectroscopy. Biophysical Journal. 86: 3981–3992.

51. Wu, B., Y. Chen, and J.D. Müller. 2009. Fluorescence Fluctuation Spectroscopy of mCherry in Living Cells. Biophys J. 96: 2391–2404.

52. Heinze, K.G., A. Koltermann, and P. Schwille. 2000. Simultaneous two-photon excitation of distinct labels for dual-color fluorescence crosscorrelation analysis. PNAS. 97: 10377–10382.

53. Chen, Y., M. Tekmen, L. Hillesheim, J. Skinner, B. Wu, et al. 2005. Dual-Color Photon-Counting Histogram. Biophysical Journal. 88: 2177–2192.

54. Palo, K., Ü. Mets, S. Jäger, P. Kask, and K. Gall. 2000. Fluorescence Intensity Multiple Distributions Analysis: Concurrent Determination of Diffusion Times and Molecular Brightness. Biophysical Journal. 79: 2858–2866.

55. Wu, B., Y. Chen, and J.D. Müller. 2010. Heterospecies partition analysis reveals binding curve and stoichiometry of protein interactions in living cells. PNAS. 107: 4117–4122.

56. Schwille, P., F.J. Meyer-Almes, and R. Rigler. 1997. Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution. Biophys J. 72: 1878–1886.

57. Bacia, K., and P. Schwille. 2003. A dynamic view of cellular processes by in vivo fluorescence auto- and cross-correlation spectroscopy. Methods. 29: 74–85.

58. Kettling, U., A. Koltermann, P. Schwille, and M. Eigen. 1998. Real-time enzyme kinetics monitored by dual-color fluorescence cross-correlation spectroscopy. PNAS. 95: 1416–1420.

59. Kim, S.A., K.G. Heinze, M.N. Waxham, and P. Schwille. 2004. Intracellular calmodulin availability accessed with two-photon cross-correlation. PNAS. 101: 105–110.

60. Bacia, K., Z. Petrášek, and P. Schwille. 2012. Correcting for spectral cross-talk in dual-color fluorescence cross-correlation spectroscopy. Chemphyschem. 13: 1221–1231.

61. Förster, T. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. Annalen der Physik. 437: 55–75.

62. Lakowicz, J. 2006. Principles of Fluorescence Spectroscopy. Springer.

63. J. Thor, J., and K. J. Hellingwerf. 2002. 8 Fluorescence Resonance Energy Transfer (FRET) Applications Using Green Fluorescent Protein. In: Hicks BW, editor. Humana Press. pp. 101–119.

64. Hillesheim, L.N., and J.D. Müller. 2003. The photon counting histogram in fluorescence fluctuation spectroscopy with non-ideal photodetectors. Biophys.J. 85: 1948–1958.

65. Hillesheim, L.N., and J.D. Müller. 2005. The Dual-Color Photon Counting Histogram with Non-Ideal Photodetectors. Biophysical Journal. 89: 3491–3507.

66. Mandel, L. 1979. Sub-Poissonian photon statistics in resonance fluorescence. Opt.Lett. 4: 205–207.

67. Finn, M.A., G.W. Greenless, T.W. Hodapp, and D.A. Lewis. 1988. Real-time elimination of dead-time and afterpulsing in counting systems. Rev.Sci.Instrum. 59: 2457–2459.

68. Wu, B., and J.D. Mueller. 2005. Time-integrated fluorescence cumulant analysis in fluorescence fluctuation spectroscopy. Biophys. J. 89: 2721–2735.

69. Wu, B., Y. Chen, and J.D. Müller. 2006. Dual-Color Time-Integrated Fluorescence Cumulant Analysis. Biophys J. 91: 2687–2698.

70. Proietti, F.A., A.B.F. Carneiro-Proietti, B.C. Catalan-Soares, and E.L. Murphy. 2005. Global epidemiology of HTLV-I infection and associated diseases. Oncogene. 24: 6058–6068.

71. Bertola, F., C. Manigand, P. Picard, M. Belghazi, and G. Precigoux. 2000. Human T-lymphotrophic virus type I nucleocapsid protein NCp15: structural study and stability of the N-terminal zinc-finger. Biochem J. 352: 293–300.

72. Christensen, A.M., M.A. Massiah, B.G. Turner, W.I. Sundquist, and M.F. Summers. 1996. Three-dimensional structure of the HTLV-II matrix protein and comparative analysis of matrix proteins from the different classes of pathogenic human retroviruses. J. Mol. Biol. 264: 1117–1131.

73. Cornilescu, C.C., F. Bouamr, X. Yao, C. Carter, and N. Tjandra. 2001. Structural analysis of the N-terminal domain of the human T-cell leukemia virus capsid protein. J. Mol. Biol. 306: 783–797.

74. Burns, J.C., T. Friedmann, W. Driever, M. Burrascano, and J.K. Yee. 1993. Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. PNAS. 90: 8033–8037.

75. Doultree, J.C., R.E. Kiernan, J.Y. Lee, D.S. Bowden, D.A. McPhee, et al. 1992. A new electron microscope positive staining method for viruses in suspension. J. Virol. Methods. 37: 321–335.

76. Fuller, S.D., T. Wilk, B.E. Gowen, H.-G. Kräusslich, and V.M. Vogt. 1997. Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle. Current Biology. 7: 729–738.

77. Poiesz, B.J., F.W. Ruscetti, A.F. Gazdar, P.A. Bunn, J.D. Minna, et al. 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. PNAS. 77: 7415–7419.

78. Yoshida, M., I. Miyoshi, and Y. Hinuma. 1982. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proc. Natl. Acad. Sci. U.S.A. 79: 2031–2035.

79. Wright, E.R., J.B. Schooler, H.J. Ding, C. Kieffer, C. Fillmore, et al. 2007. Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. EMBO J. 26: 2218–2226.

80. Houzet, L., B. Gay, Z. Morichaud, L. Briant, and M. Mougel. 2006. Intracellular assembly and budding of the Murine Leukemia Virus in infected cells. Retrovirology. 3: 12.

81. Yu, F., S.M. Joshi, Y.M. Ma, R.L. Kingston, M.N. Simon, et al. 2001. Characterization of Rous Sarcoma Virus Gag Particles Assembled In Vitro. J. Virol. 75: 2753–2764.

82. Zhang, J., R.E. Campbell, A.Y. Ting, and R.Y. Tsien. 2002. Creating new fluorescent probes for cell biology. Nature Reviews Molecular Cell Biology. 3: 906–918.

83. Arai, R., H. Ueda, A. Kitayama, N. Kamiya, and T. Nagamune. 2001. Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng. 14: 529–532.

84. Smith, E., and J. Mueller. 2012. The statistics of protein expression ratios for cellular fluorescence studies. European Biophysics Journal. 41: 341–352.

85. Crick, F.H.C., and J.D. Watson. 1956. Structure of Small Viruses. , Published online: 10 March 1956; | doi:10.1038/177473a0. 177: 473–475.

86. Ako-Adjei, D., M.C. Johnson, and V.M. Vogt. 2005. The Retroviral Capsid Domain Dictates Virion Size, Morphology, and Coassembly of Gag into Virus-Like Particles. J Virol. 79: 13463–13472.

87. Ono, A. 2010. Relationships between plasma membrane microdomains and HIV-1 assembly. Biol Cell. 102: 335–350.

88. Muriaux, D., and J.-L. Darlix. 2010. Properties and functions of the nucleocapsid protein in virus assembly. RNA Biol. 7: 744–753.

89. Cimarelli, A., and J.L. Darlix. 2002. Assembling the human immunodeficiency virus type 1. Cell. Mol. Life Sci. 59: 1166–1184.

90. Ott, D.E., L.V. Coren, and T. Shatzer. 2009. The Nucleocapsid Region of Human Immunodeficiency Virus Type 1 Gag Assists in the Coordination of Assembly and Gag Processing: Role for RNA-Gag Binding in the Early Stages of Assembly. J Virol. 83: 7718–7727.

91. Rein, A., S.A.K. Datta, C.P. Jones, and K. Musier-Forsyth. 2011. Diverse interactions of retroviral Gag proteins with RNAs. Trends in Biochemical Sciences. 36: 373–380.

92. Ott, D.E. 2008. Cellular proteins detected in HIV-1. Reviews in Medical Virology. 18: 159–175.

93. Swanson, C.M., and M.H. Malim. 2008. SnapShot: HIV-1 Proteins. Cell. 133: 742–742.e1.

94. Bess, J.W., Jr, R.J. Gorelick, W.J. Bosche, L.E. Henderson, and L.O. Arthur. 1997. Microvesicles are a source of contaminating cellular proteins found in purified HIV-1 preparations. Virology. 230: 134–144.

95. Gluschankof P., Mondor I., Gelderblom H.R., and Sattentau Q.J. 1997. Cell Membrane Vesicles Are a Major Contaminant of Gradient-Enriched Human Immunodeficiency Virus Type-1 Preparations. Virology. 230: 125–133.

96. Sheehy, A.M., N.C. Gaddis, J.D. Choi, and M.H. Malim. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature. 418: 646–650.

97. Harris, R.S., K.N. Bishop, A.M. Sheehy, H.M. Craig, S.K. Petersen-Mahrt, et al. 2003. DNA Deamination Mediates Innate Immunity to Retroviral Infection. Cell. 113: 803–809.

98. Mangeat, B., P. Turelli, G. Caron, M. Friedli, L. Perrin, et al. 2003. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature. 424: 99–103.

99. Hultquist, J.F., J.A. Lengyel, E.W. Refsland, R.S. LaRue, L. Lackey, et al. 2011. Human and Rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H Demonstrate a Conserved Capacity To Restrict Vif-Deficient HIV-1. J. Virol. 85: 11220–11234.

100. Malim, M.H. 2009. APOBEC proteins and intrinsic resistance to HIV-1 infection. Philos Trans R Soc Lond B Biol Sci. 364: 675–687.

101. Marin, M., K.M. Rose, S.L. Kozak, and D. Kabat. 2003. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nature Medicine. 9: 1398–1403.

102. Bogerd, H.P., and B.R. Cullen. 2008. Single-stranded RNA facilitates nucleocapsid: APOBEC3G complex formation. RNA. 14: 1228–1236.

103. Alce, T.M., and W. Popik. 2004. APOBEC3G Is Incorporated into Virus-like Particles by a Direct Interaction with HIV-1 Gag Nucleocapsid Protein. J. Biol. Chem. 279: 34083–34086.

104. Schäfer, A., H.P. Bogerd, and B.R. Cullen. 2004. Specific packaging of APOBEC3G into HIV-1 virions is mediated by the nucleocapsid domain of the gag polyprotein precursor. Virology. 328: 163–168.

105. Xu, H., E. Chertova, J. Chen, D.E. Ott, J.D. Roser, et al. 2007. Stoichiometry of the antiviral protein APOBEC3G in HIV-1 virions. Virology. 360: 247–256.

106. Slaughter, B.D., J.M. Huff, W. Wiegraebe, J.W. Schwartz, and R. Li. 2008. SAM domain-based protein oligomerization observed by live-cell fluorescence fluctuation spectroscopy. PLoS ONE. 3: e1931.

107. Hillesheim, L.N., Y. Chen, and J.D. Müller. 2006. Dual-Color Photon Counting Histogram Analysis of mRFP1 and EGFP in Living Cells. Biophys J. 91: 4273–4284.

108. Chen, Y., J. Johnson, P. Macdonald, B. Wu, and J.D. Mueller. 2010. Chapter 16 - Observing Protein Interactions and Their Stoichiometry in Living Cells by Brightness Analysis of Fluorescence Fluctuation Experiments. In: Single Molecule Tools: Fluorescence Based Approaches, Part A. Academic Press. pp. 345–363.

109. Akrap, N., T. Seidel, and B.G. Barisas. 2010. Förster distances for FRET between mCherry and other Visible Fluorescent Proteins. Anal Biochem. 402: 105–106.

110. Bieschke, J., A. Giese, W. Schulz-Schaeffer, I. Zerr, S. Poser, et al. 2000. Ultrasensitive detection of pathological prion protein aggregates by dual-color scanning for intensely fluorescent targets. PNAS. 97: 5468–5473.

111. Banks, D. 2006. Microengineering, MEMS, and Interfacing: A Practical Guide. CRC Press.

112. Ito, H. 1997. Chemical amplification resists: History and development within IBM. IBM Journal of Research and Development. 41: 119 –130.

113. McDonald, J.C., and G.M. Whitesides. 2002. Poly(dimethylsiloxane) as a Material for Fabricating Microfluidic Devices. Acc. Chem. Res. 35: 491–499.

114. Mata, A., A.J. Fleischman, and S. Roy. 2005. Characterization of polydimethylsiloxane (PDMS) properties for biomedical micro/nanosystems. Biomed Microdevices. 7: 281–293.

115. Müller, J.D., Y. Chen, and E. Gratton. 2000. Resolving Heterogeneity on the Single Molecular Level with the Photon-Counting Histogram. Biophysical Journal. 78: 474–486.

116. Waite, T.R. 1957. Theoretical Treatment of the Kinetics of Diffusion-Limited Reactions. Phys. Rev. 107: 463–470.

117. Perevoshchikova, I.V., D.B. Zorov, and Y.N. Antonenko. 2008. Peak intensity analysis as a method for estimation of fluorescent probe binding to artificial and natural nanoparticles: tetramethylrhodamine uptake by isolated mitochondria. Biochim. Biophys. Acta. 1778: 2182–2190.

118. Van Craenenbroeck, E., G. Matthys, J. Beirlant, and Y. Engelborghs. 1999. A Statistical Analysis of Fluorescence Correlation Data. Journal of Fluorescence. 9: 325–331.

119. Sanchez-Andres, A., Y. Chen, and J.D. Müller. 2005. Molecular Brightness Determined from a Generalized Form of Mandel's Q-Parameter. Biophysical Journal. 89: 3531–3547.

120. Shaner, N.C., R.E. Campbell, P.A. Steinbach, B.N.G. Giepmans, A.E. Palmer, et al. 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nature Biotechnology. 22: 1567–1572.

121. Griffin, B.A., S.R. Adams, and R.Y. Tsien. 1998. Specific Covalent Labeling of Recombinant Protein Molecules Inside Live Cells. Science. 281: 269–272.

122. Hübner, W., P. Chen, A.D. Portillo, Y. Liu, R.E. Gordon, et al. 2007. Sequence of Human Immunodeficiency Virus Type 1 (HIV-1) Gag Localization and Oligomerization Monitored with Live Confocal Imaging of a Replication-Competent, Fluorescently Tagged HIV-1. J Virol. 81: 12596–12607.

Appendix A: Removal of Rare Events from FFS Data

In this appendix, we comment on the removal of high intensity, low frequency events, which is necessary before the raw data can be analyzed using FFS methods. This data filtering is required to remove low statistics events and to ensure proper fitting of the remaining data.

A.1 Comments on the removal of high intensity peaks

Occasionally, an unusually high intensity spike with an amplitude 3-10 times that of any other peaks was observed during an FFS experiment. The number of these events varied with sample type, sample preparation, and measurement time. Not all samples contained these events, but of the ones that do, 1-10 events per 30 min measurement were observed. The low probability of these events does not allow for accurate statistical analysis. These events results in a characteristic tail observed in PCH. This tail is long and nearly flat, indicating the extreme intensity and low probability respectively. This histogram cannot be fit to any current model. For these reasons, removal of these rare peaks is necessary for accurate data analysis.

These spikes most likely represent the passage of a large fluorescent aggregate through the observation volume, such as a piece of cell membrane with fluorescent proteins attached. These events could also represent an aggregate of VLPs, or possibly even very rare, high copy number VLPs, but the low event rate does not provide sufficient statistics for accurate analysis. We are aware that the sample sometimes contains fluorescent and non-fluorescent debris, because occasionally this debris becomes trapped by the laser, resulting in the need to retake data. If practical, data sets with spikes were immediately retaken. However, due to time limitations or the difficulty of taking data that was free of spikes, sometimes the data was kept and the spike was removed by software from the raw data before analysis. Here again, we stress that peaks were only removed if they had a distinctly different amplitude then the vast

majority of peaks. In situations where these peaks were not clearly identifiable as significantly different, they were not removed.

A.2 Dual color peak removal

We observed that when measuring the supernatant of cells transfected with A3G-EGFP alone, or A3G-EGFP and Gag-mCherry a small number of events with a peak appearing only in the green detection channel are observed. We refer to these vents as “green only”. Typically anywhere from 1-6 green only peaks are observed during a typical measurement. It is necessary to remove these statistically rare events before further analysis. If the peaks are not removed, 2D-PCH fitting results in high reduced chi-squareduared values.

We presume these spikes represent aggregates or cell debris containing A3G, although further characterization is necessary to confirm this hypothesis. The scarcity of events makes it impossible to characterize their copy number by FFS methods, but leaving them in the raw data results in difficulties when fitting with 2D-PCH. Fortunately, they can be easily identified (using IFP to find green only events) and are rare enough that individual events can be removed before analyzing dual color experiments with Gag-mCherry and A3G-EGFP. Figure A.1 shows the IFP of VLPs from a 1:1 transfection mixture of A3G-EGFP to Gag-mCherry. For this data set, a single green only event is observed.

In addition to these green only peaks, we occasionally observe single events with an intensity fraction that deviates significantly from the average intensity fraction, but does not represent a green only particle. The source of these rare events is currently not known, and they are removed before further analysis. The IFP enables us to identify and specifically remove these rare events, which allows for subsequent 2D-PCH analysis.

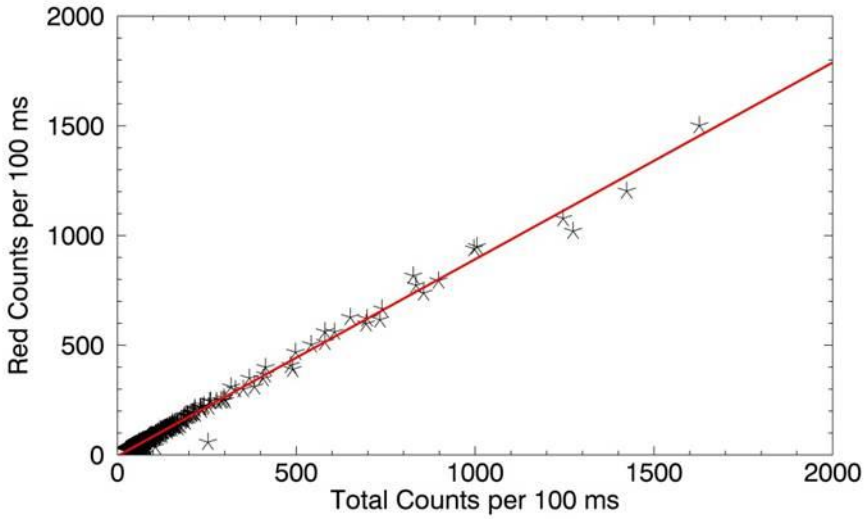


Figure A.1 IFP showing a single green only peak that will be removed before further analysis

The supernatant of cells transfected with a 1:1 mixture of A3G-EGFP to Gag-mCherry was collected, purified and measured. At approximately 250 total counts, a single green only event was identified and removed before further analysis.

Appendix B: Extended Results of HTLV-1 Studies

This appendix describes additional HTLV-1 VLP experiments and supporting data referred to in Chapter 3.

B.1 FFS of VLPs carrying HTLV-1 Gag-EYFP and dark Gag

To further investigate the issues with HTLV-1 mixing, we explored a wide range of mixing ratios. The goal of these studies was to find a mixing ratio that would allow quantitative labeling of a small percentage of the Gag while maintaining authentic VLP morphology. The results for HIV-1 VLPs were as expected and previously reported (9). HIV-1 Gag-EYFP copy number scales with the percentage of labeled Gag used during transfection, even with only 10% of the Gag labeled. The results for HTLV-1 VLP experiments are shown in Figure B.1.

The increase in size with increasing percentage of dark Gag DNA suggests that the labeled and unlabeled Gag may be incorporated into the same particle, although the results are not conclusive because of the high uncertainty in the reported FFS parameters. The uncertainty was calculated as the standard deviation of two independently measured samples. It was not feasible to test mixing ratios in excess of 1:9 labeled:dark protein due to limitations imposed by transfection statistics (84). Thus, while the increase in VLP diameter is encouraging, the minimal variation in Gag copy number with varying amounts of dark Gag prohibit quantitative experiments with the construct.

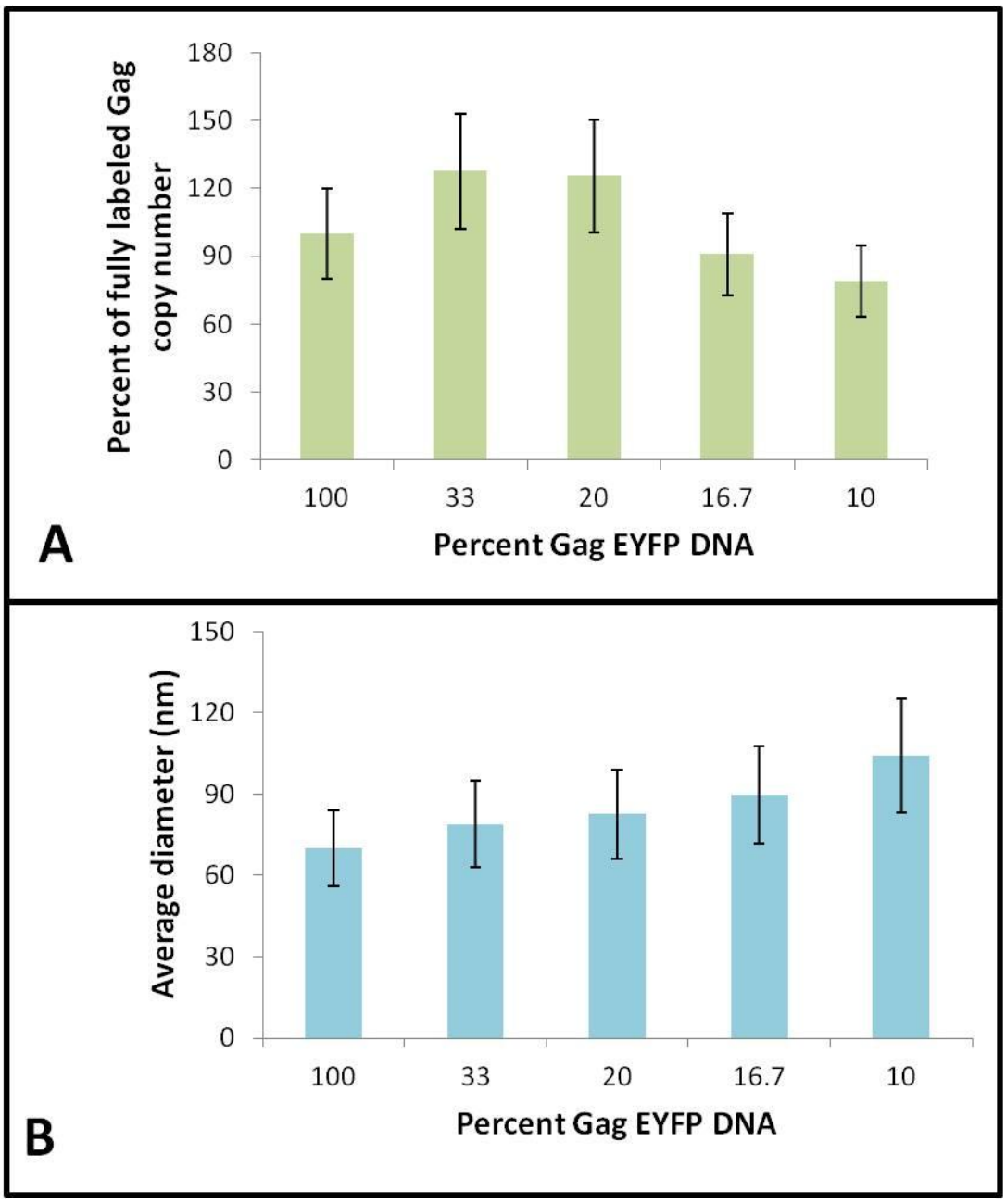


Figure B.1 HTLV-1 VLP extended 'dark' Gag DNA titration experiment
(A) The apparent copy number for each VLP population was calculated and scaled to the Gag-EYFP only copy number.
(B) The average hydrodynamic diameter was calculated using autocorrelation analysis.

B.2 Cryo-TEM of VLPs from cells transfected with a 1:1 ratio of dark Gag to Gag-EYFP

To study the particle morphology of VLPs containing dark and labeled Gag, Jose Maldonado collected HLTV-1 VLPs from cells transfected with a 1:1 ratio of EYFP labeled Gag and dark Gag and Dr. Wei Zhang studied the particles using cryo-TEM. Cryo-TEM does not directly verify the mixing of Gag particles because the limited spatial resolution is insufficient to determine if individual Gag molecules contain an EYFP. However, investigating the particle morphology and structure can give insights into whether mixing occurs or is absent. For instance, because of the distinct differences in structure between labeled and unlabeled VLPs, a split into populations with two distinct types of structure near the membrane would indicate a lack of mixing at the VLP level. We would expect dark only particles to contain sections of structure near the membrane, and EYFP only particles to have a constant radial density profile. A representative VLP observed is shown in Figure B.2c. The most striking observation about these particles is their "peanut" shape. These particles almost look like a fusion of sections of the dark Gag only structure (the larger oval) with the less structured EYFP labeled Gag structure (smaller oval in upper right). For comparison, a representative Gag-EYFP VLP and 'dark' Gag VLP is included in Figure B.2a and B.2b respectively. The "peanut" shape of these particles hints at a certain level of segregation of the two Gag types. From intracellular immunofluorescent studies, we know that the labeled and unlabeled Gag colocalize to the same budding site (unpublished data, Mansky lab). This matches our previous observation that both the labeled and unlabeled Gag are targeted to the same area of the cell (25). Therefore, while labeled and unlabeled Gag assembles into the same particle and colocalize in the cell, mixing at the molecular level appears absent. Note that this type of peanut shaped VLP had not been observed when labeled and unlabeled Gag was mixed for HIV-1. From the cryo-TEM picture, it appears that labeled and

unlabeled Gag form VLPs with distinct curvature and morphology, even though they must be budding from the same area of the cell to form these peanut structures.

B.3 FFS studies looking for the source of labeling interference

B.3.1 Label-label interactions experiments

Label-label interactions potentially influence viral assembly. These interactions occur when a fluorescent label interacts with a nearby label. This interaction may change the Gag-Gag interactions we are interested in studying. While this was not observed with HIV-1 Gag, it has been reported that at high enough concentrations EGFP engages in label-label interactions (82). The concentrations where these self-interactions were observed are not realized in the cytoplasm, but, near the cell membrane where Gag assembles, the Gag concentration is significantly higher. To test this hypothesis, we created constructs of HTLV-1 Gag labeled with three different colored tags, EYFP, EGFP, and mCherry. EYFP and EGFP have nearly identical structure on the protein surface as they are derived from the same source. mCherry is derived from a different organism, and therefore has a distinct surface structure compared to EGFP and EYFP (120). First, we transfected cells to produce VLPs carrying a single color tag and measured the size and copy number using FFS. All color tags were observed to result in VLPs of the same size as the EYFP tagged VLPs (75 nm). Additionally, the copy number of the VLPs was tag independent and matched that previously reported for EYFP tagged Gag VLPs. The corrections necessary for using mCherry as a tag are discussed in Chapter 4. The lack of a difference in size or copy number between VLPs containing EGFP, EYFP or mCherry suggests the absence of label specific interactions.

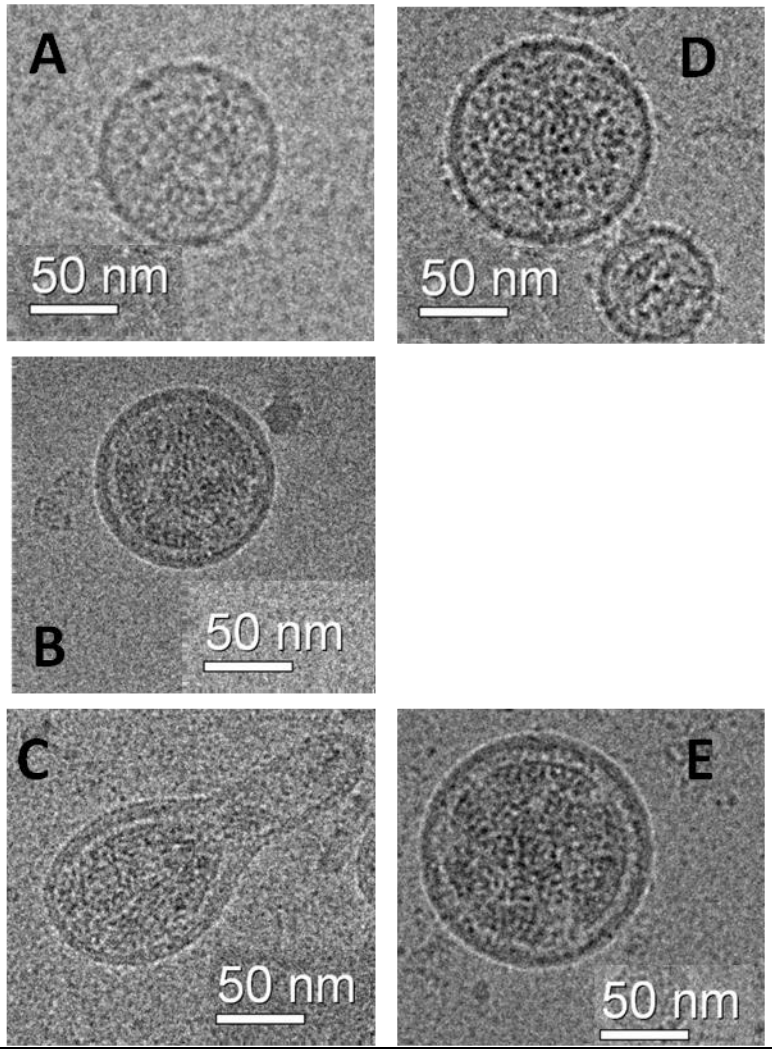


Figure B.2 Example Cryo-TEM images of HTLV-1 VLP morphologies.
(A) Representative Cryo-TEM of VLP produced from 293T cell transfected with HTLV-1 Gag-EYFP construct.
(B) Representative Cryo-TEM of VLP produced from 293T cell transfected with 'dark' HTLV-1 Gag.
(C) Representative Cryo-TEM of VLP produced from 293T cell transfected with 1:1 mixture of 'dark' Gag and Gag-EYFP.
(D) Representative Cryo-TEM of VLP produced from 293T cell transfected with only Gag-helical linker-EYFP construct.
(E) Representative Cryo-TEM of VLP produced from 293T cell transfected with 1:5 ratio of HTLV-1 Gag-EYFP construct to 'dark' Gag.
Each of these cryo-TEM experiments looked at hundreds to thousands of individual VLPs. These VLPs were chosen because they demonstrate the most common particle morphology observed for each sample.

Cells were also transfected with a mixture of Gag-EGFP and Gag-mCherry. If label-label interactions occur, we expect preferential incorporation of one label over the other label. If the interaction is strong enough we would expect to observe VLPs containing only mCherry or only EGFP. The signal was split into red and green detection channels and analyzed for the presence of each color tag on the VLPs. We found that all VLPs produced had a mixture of red- and green-tagged Gag, and the average diameter matched that for the VLPs labeled with a single color. Without accounting for FRET, it is difficult to quantify the mixing, but we did not observe any VLPs with a majority or near majority of red or green labels. These results along with the consistent size provide hints that label-label interactions are not the main source of the label interference problem.

B.3.2 Label size experiments

Next, we tested whether the size of the tag affects mixing and particle morphology. We performed a set of experiments to investigate the effect of tag size on Gag mixing. First, we made a number of different 'dark-labeled' versions of the HTLV-1 Gag. These 'dark-labeled' Gags contained varying amounts of EYFP attached to the end of the NC domain, although the section encoding the fluorophore was always removed. If the tag size were the determining factor in mixing between unlabeled and labeled particles, we would expect that as the size of the tag on the dark-labeled Gag decreases, the amount of mixing with the regular Gag-EYFP should also decrease. The three dark-labeled Gag-EGFP cuts tested were 180, 120, and 30/30. These numbers represent the number of amino acids of EGFP remaining after the cut. The 30/30 construct is so named because it contains the first 30 amino acids of EGFP and the last 30. This construct was used instead of 60 because it was already made for a previous experiment. Note, these cuts were chosen for their even spacing and ability to quickly characterize a range of tag sizes. We did not carefully examine the structure of the EGFP left on each construct to ensure proper folding of the peptide sequence. Again, if mixing

occurs as expected we should see the EYFP labeled copy number decrease as the percentage of dark-labeled Gag DNA increases.

The dark-labeled Gag with the longest dark labels (180 and 120) mix with the EYFP labeled Gag approximately as expected. The dark-labeled Gag copy number approximately scales with the percentage of fully labeled Gag. The shortest dark-labeled Gag (30/30) does not mix as expected. Thus, the 30/30 construct behaves similar as dark Gag. These results (Figure B.3) point to tag size as one potential source of label interference.

There are not many commercially available fluorescent protein tags that are smaller in size than standard fluorescent proteins such as EYFP. One system available is the ReAsH tag (121). ReAsH uses a 6 amino acid tag, the smallest expression tag available. Given this, we tested how well Gag labeled with this tag would interact with Gag-EYFP. In this system, the tag itself is not fluorescent. Rather, a fluorescent dye is introduced to the cells that have the tag expressed, and the dye binds to the tag with a high affinity. In order to keep the tag dark, we did not introduce the ReAsH dye into the system in these experiments. The results indicated that ReAsH labeled Gag behaves similarly to dark Gag when mixed with Gag-EYFP (data not shown). This experiment suggests that using a smaller tag might be one solution to the label interaction problem. However, the ReAsH system is not compatible with quantitative FFS studies due to dye-induced background and the difficulty in quantifying the exact percentage of tagged proteins that are labeled with the dye. For these reasons, we continued searching for a system that is compatible with fluorescent protein labeling.

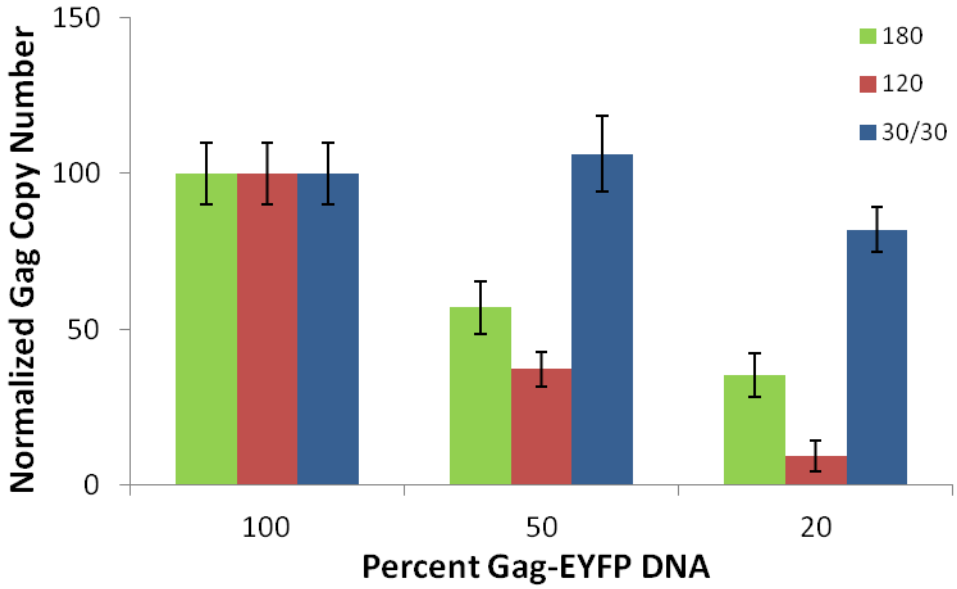


Figure B.3 FFS studies of label size on HTLV-1 Gag mixing
HTLV-1 Gag-EYFP was mixed with three different dark-labeled Gags that contained varying lengths of non-fluorescent amino acid sequences attached to the NC domain. These sequences consisted of varying amounts of EGFP. 180 means the first 180 amino acids of EGFP were attached to NC; 120 means 120 EGFP amino acids were attached; and 30/30 means the first 30 and last 30 amino acids of EGFP were attached. The larger dark-labeled Gag resulted in Gag-EYFP copy numbers that scaled with the amount of Gag-EYFP DNA used in transfection. The shorter 30/30 dark label did not indicate the presence of mixing. The result of these experiments suggests that label size affects the Gag-Gag interactions in HTLV-1 Gag assembly.

B.3.3 Tag location experiments

Another possible explanation of the observed label interference is that the location of the tag at the end of HTLV-1 Gag affects protein targeting, degradation, or other cellular differences that lead to the observed assembly differences. This hypothesis was tested by changing the position of the fluorescent tag within the Gag molecule. We examined the VLPs produced by each new construct and studied their mixing behavior with dark Gag. Here, we tested inserting the EYFP tag between the MA and CA domain, which we refer to as Gag-iYFP. The same approach has been successfully used in the study of labeled HIV-1 virions (122). Inserting a GFP between the MA and CA domain of HIV-1 Gag construct produces particles of similar size and morphology as the authentic virus.

The HTLV-1 Gag-iYFP construct produced VLPs, but the concentration was much lower (factor of ~10) than for the Gag-EYFP construct. This made quantitative FFS studies difficult, and we were unable to collect enough VLPs for cryo-TEM studies. Based on FFS studies, Gag-iYFP produces VLPs with the same apparent copy number as the standard EYFP tag after the NC domain as seen in Figure B.4.

Additionally, the Gag-iYFP construct appears to behave as the regular Gag-EYFP construct when mixed with dark Gag. The copy number seems to be approximately constant as dark Gag is mixed in (Figure B.4). Combined, these results indicate that while it is possible to insert the fluorescent tag between the MA and CA domain of the Gag protein, the resultant VLPs suffer from similar problems as the VLPs produced with EYFP tagged at the end of Gag.

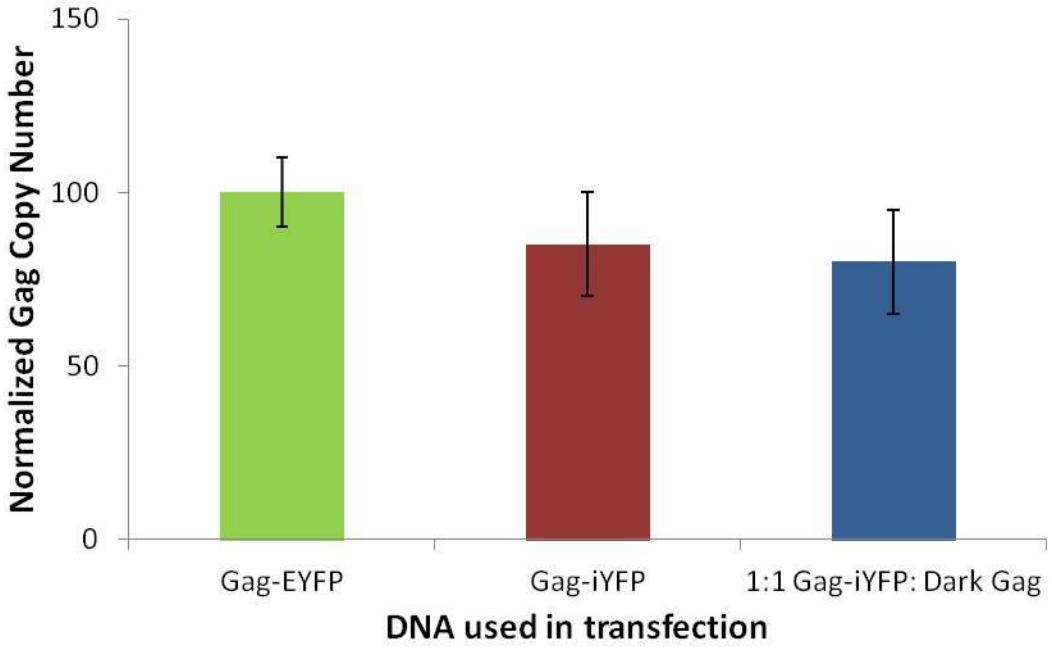


Figure B.4 Gag-iYFP HTLV-1 copy number results
Gag-iYFP, which contains an EYFP label after the MA domain instead of at the end of the NC domain, produces VLPs with a similar copy number as HTLV-1 Gag-EYFP VLPs. Additionally the labeled Gag copy number does not decrease when Gag-iYFP is coexpressed with dark Gag. These results suggest that changing the location of the label does not improve the mixing of dark and labeled Gag in HTLV-1 VLPs.

B.4 HTLV-1 Gag-linker-EYFP Cryo-TEM studies

B.4.1 HTLV-1 Gag-linker-EYFP only VLPs Cryo-TEM results

The particle morphology and size of HTLV-1 Gag-linker-EYFP was examined using cryo-TEM by Dr. Wei Zhang and Dr. Iwen Grigsby. A typical VLP produced is shown in Figure B.2d. These VLPs were closer in size to the VLPs with the EYFP label without a linker than the dark Gag VLPs. Although a size histogram was not measured, it appears that on average the VLPs formed with the Gag helical linker EYFP construct are slightly larger than the Gag-EYFP VLPs. Studies are continuing to determine if the size difference between the Gag-EYFP VLPs and the Gag-linker-EYFP VLPs is significant. Importantly, there is some distinct structure observable near the membrane as was seen for the dark Gag VLPs. This structure is less well defined in general than what was observed for the dark Gag VLPs. However, since structure was very rarely observed for the Gag-EYFP VLPs, the observation of any structure is promising. These cryo-TEM results point to the addition of a linker as a promising candidate for quantitative labeling experiments.

B.4.2 Cryo-TEM results of HTLV-1 Gag-linker-EYFP mixed with HTLV-1 dark Gag VLPs

Cryo-TEM experiments looking at mixtures of dark Gag and Gag-linker-EYFP will help provide a better idea of the particle morphology and size distribution. Jose Maldonado collected VLPs produced from cells transfected with a 1:5 mixture of EYFP linker-Gag DNA and dark Gag DNA. Dr. Wei Zhang then examined these particles using cryo-TEM. This ratio was chosen based FFS experiments indicated that at this ratio the size of these VLPs is similar to dark Gag VLPs and the amount of Gag-linker-EYFP incorporated into VLPs starts to decline with decreasing plasmid ratio (see Figure 3.9). Figure B.2e shows the cryo-TEM image of a typical VLP. The cryo-TEM results for the mixed particles are promising. The average diameter of the particles was roughly 120 nm, which is close to the value observed for dark HTLV-1 VLPs (114 ± 29 nm), and

qualitatively the structure of the majority of the particles closely resembled the sections of Gag lattice seen in the dark only particles.