

DISSERTATION

HIV-1 GAG TRAFFICKING AND ASSEMBLY: MATHEMATICAL MODELS AND  
NUMERICAL SIMULATIONS

Submitted by

Roberto Munoz-Alicea

Department of Mathematics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2013

Doctoral Committee:

Advisor: Jianguo (James) Liu

Co-Advisor: Simon Tavener

Chaoping Chen

Jennifer Mueller

Patrick Shipman

Copyright by Roberto Munoz-Alicea 2013

All Rights Reserved

## ABSTRACT

### HIV-1 GAG TRAFFICKING AND ASSEMBLY: MATHEMATICAL MODELS AND NUMERICAL SIMULATIONS

AIDS (acquired immune deficiency syndrome) is an infectious disease that takes away many people's lives each year. Group-specific antigen (Gag) polyprotein precursor is the major structural component of HIV, the causing agent of AIDS. Gag is essential and sufficient for the formation of new HIV virus-like particles. The late stages of the HIV-1 life cycle include the transport of Gag proteins towards the cell membrane, the oligomerization of Gag near the cell membrane during the budding process, and core assembly during virion maturation. The mechanisms for Gag protein trafficking and assembly are not yet fully understood. In order to gain further insight into the mechanisms of HIV-1 replication, we develop and analyze mathematical models and numerical algorithms for intracellular Gag protein trafficking, Gag trimerization near the cell membrane, and HIV-1 core assembly.

Our preliminary results indicate that active transport plays an important role for Gag trafficking in the cytoplasm. This process can be mathematically modeled by convection-diffusion equations, which can be solved efficiently using characteristic finite element methods. We employ differential dynamical systems to model Gag trimerization and HIV-1 core assembly.

For the Gag trimerization model, we estimate relationships between the association and dissociation parameters as well as the Gag arrival and multimerization parameters. We also find expressions for the equilibrium concentrations of the monomer and trimer species, and show that the equilibrium is asymptotically stable.

For HIV-1 core assembly, we first consider a model developed by Zlonick and others, which regards assembly as a polymerization reaction. We utilize theoretical and numerical tools to confirm the stability of the equilibrium of CA intermediates. In addition, we propose a cascaded dynamical system model for HIV-1 core assembly. The model consists of two subsystems: one subsystem for nucleation and one for elongation. We perform simulations on the nucleation model, which suggests the existence of an equilibrium of the CA species.

**Keywords:** capsid, convection, diffusion, equilibrium, finite elements, Gag, HIV, micro-tubules, transport, trimerization, viral assembly

## ACKNOWLEDGEMENTS

I would like to thank my advisors,  
my committee members,  
and my parents.

## TABLE OF CONTENTS

ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	iv
Chapter 1. Biological Problems - HIV and Gag .....	1
1.1. AIDS and HIV .....	1
1.2. HIV and Gag .....	2
1.3. HIV Maturation .....	5
1.4. Roles of Mathematical Modeling for HIV and Gag .....	7
Chapter 2. Gag Trafficking Inside the Cytoplasm .....	11
2.1. Biological Background and Existing Work .....	11
2.2. A Quasi-2D Model for Gag Trafficking .....	13
2.3. Simulation Results .....	24
2.4. Parallelization of Numerical Simulations .....	26
2.5. Stochastic Aspects of Gag Trafficking .....	29
2.6. Remarks: Biological Implications .....	34
2.7. Remarks on Further Work .....	35
Chapter 3. HIV Gag Trimerization at the Plasma Membrane .....	37
3.1. Gag Trimerization and Multimerization .....	37
3.2. Kinetic and Equilibrium Constants .....	40
3.3. Mathematical Model for Two-species Gag Trimerization .....	42
3.4. Remarks: Biological Implications and Further Work .....	49

Chapter 4. HIV Core Assembly .....	52
4.1. Introduction .....	52
4.2. Existing Work .....	54
4.3. Dynamical System Model .....	60
4.4. Matlab Implementation .....	63
4.5. <i>In Silico</i> .....	63
4.6. Interpreting Results .....	66
4.7. Existence and Stability of Equilibrium .....	67
4.8. Cascaded Dynamical System Model .....	70
4.9. Remarks: Biological Implications and Further Work .....	74
Chapter 5. Summary .....	76
5.1. Gag Trafficking .....	76
5.2. Gag Trimerization .....	77
5.3. HIV Core Assembly .....	77
BIBLIOGRAPHY .....	79

## CHAPTER 1

# BIOLOGICAL PROBLEMS - HIV AND GAG

### 1.1. AIDS AND HIV

Human immunodeficiency virus type 1 (HIV-1), the causative agent of acquired immunodeficiency syndrome (AIDS), was discovered in early 1980's [1]. More than three million people die each year from AIDS, and roughly 40 million people are infected with HIV [1].

It is believed that HIV descended from simian immunodeficiency virus (SIV). From the evolutionary point of view, AIDS is a new disease and humans have not adapted to it yet. HIV attacks T lymphocytes and macrophages. In particular, HIV infects and kills CD4<sup>+</sup> T helper cells [1, 2], which coordinate the immune response against invading pathogens. Untreated infected individuals usually develop AIDS about 8-10 years after infection [1]. Most people with AIDS have very weak immune systems, and they eventually die due to the body's inability to fight off infections.

HIV is a retrovirus, which means that its genome is stored in RNA form instead of DNA. The replication process of a retrovirus is error-prone, and therefore retroviruses have a high mutation rate [1, 2]. On average, there is a one-point mutation in each replication cycle of HIV. This is due to the fact that the enzyme *reverse transcriptase*, which transcribes the virus's RNA genome into DNA form, "cannot correct nucleotide misincorporation errors" [2] (p. 1450). Only about one-third of HIV's coding sequences in its genome remains invariant. This presents a big dilemma. Not only is it very difficult for the immune system to fight off HIV, but also it is extremely challenging to develop effective vaccines against this pathogen [2].

Many different treatments for HIV have been developed, and there has been much progress in this regard. On the other hand, there is still much to be learned about the way the pathogen that causes AIDS replicates in the infected cell. We aim to better understand the HIV's life cycle in order to aid in determining more efficient ways to treat this disease.

There are two main different types of HIV: HIV-1 and HIV-2, which seem to have evolved separately from different strains of SIV. HIV-2 is less virulent and is not as widely spread [1, 2]. In this work, unless otherwise specified, HIV refers to HIV-1.

## 1.2. HIV AND GAG

In order to better comprehend HIV's replication cycle, we need to understand the virus's composition and structure. Several main proteins, such as Env, Pol, Gag, and others comprise HIV [2, 3, 4]. Our focus in this study is on the main structural component of HIV particles, the Gag protein [3, 5, 6].

**Gag.** Gag (group specific antigen) is essential in the formation of new HIV virions, each of which is enveloped, of spherical shape [2, 3], and consists of approximately 5,000 Gag molecules. *In vivo*, Gag is sufficient for the formation of new HIV virus-like particles (VLP's). Uncleaved Gag surrounds the HIV RNA genome and spreads radially in the immature virion [3, 5, 8, 9, 10].

The major domains of Gag are: matrix (MA), capsid (CA), nucleo-capsid (NC), and p6 (see Figure 1.1). Gag binds to the plasma membrane via its MA domain. The CA domain forms the outer shell of the mature HIV virion's core during the maturation stage. Gag multimerizes via its CA and NC domains. The NC domain binds to viral RNA. Gag utilizes its p6 domain and the host cell's plasma membrane during budding [3, 10, 12, 13].

**Retroviral Lifecycle.** HIV utilizes CD4 receptors to enter CD4 T cells and macrophages [2, 7]. The virus invades the host cell by fusing with the plasma membrane, and releases its genomic RNA and other proteins into the cytoplasm [2]. The enzyme reverse transcriptase converts the RNA into DNA, which is then integrated into the host cell's DNA by the enzyme integrase [2]. The host cell transcribes the viral DNA into viral mRNA, which then travels to the cytoplasm. From mRNA, new Gag proteins are synthesized in the cytoplasm on free polysomes (clusters of ribosomes) [14, 15]. Nermut et al. [15] found in their study that the Gag precursor pr55Gag first appeared in the cytoplasm of an infected cell at 28-30 hours post infection. (h.p.i.). The newly synthesized Gag proteins, associated to viral RNA, travel to the plasma membrane via diffusion and active transport [16]. Gag protein transport to the plasma membrane after synthesis takes about 5-10 minutes [7]. Transport mediated by motor proteins along microtubules has been proposed as a mechanism

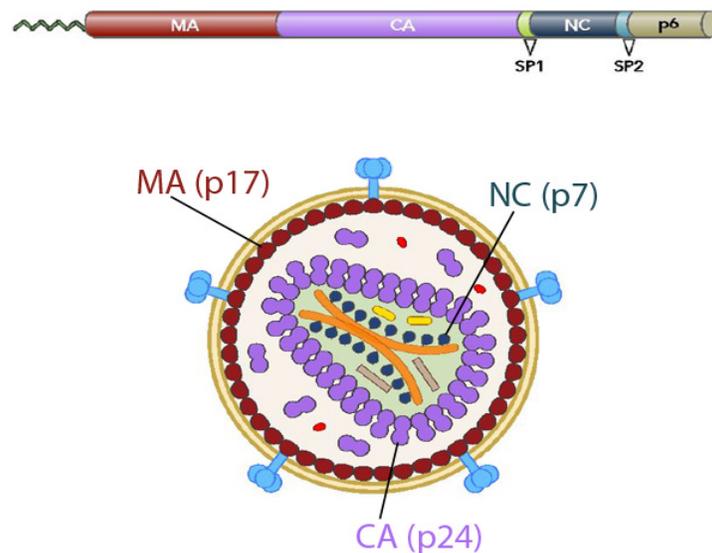


FIGURE 1.1. HIV and Gag. *The Gag domains MA, CA, NC, and p6 are the main structural components of the HIV particle. The MA domain binds to the interior side of the viral membrane. The CA domain forms the viral core shell. The NC domain is attached to the RNA inside the core. p6 is involved in the virion's budding from the host cell. Source: [7] (with slight modification).*

for intracellular active transport of Gag proteins [14, 17]. Multimerization of Gag molecules takes place at the plasma membrane, where new immature virions assemble right before or during budding. This process starts at around 6-8 hours after the Gag proteins are first synthesized in the cytoplasm [9, 15]. However, the speed at which the Gag proteins are produced and transported seems to vary depending on the type of cell. Nermut et al. [15] found in their experiments that the majority of Sf9 cells were in the budding stage at 40 to 44 h.p.i., while in T4 lymphocytes, budding took 2 or 3 days to begin. Assembly of the new virions at the plasma membrane takes about 5-6 minutes [9]. The immature virions leave the infected cell and begin the process of maturation. During this stage, each newly formed

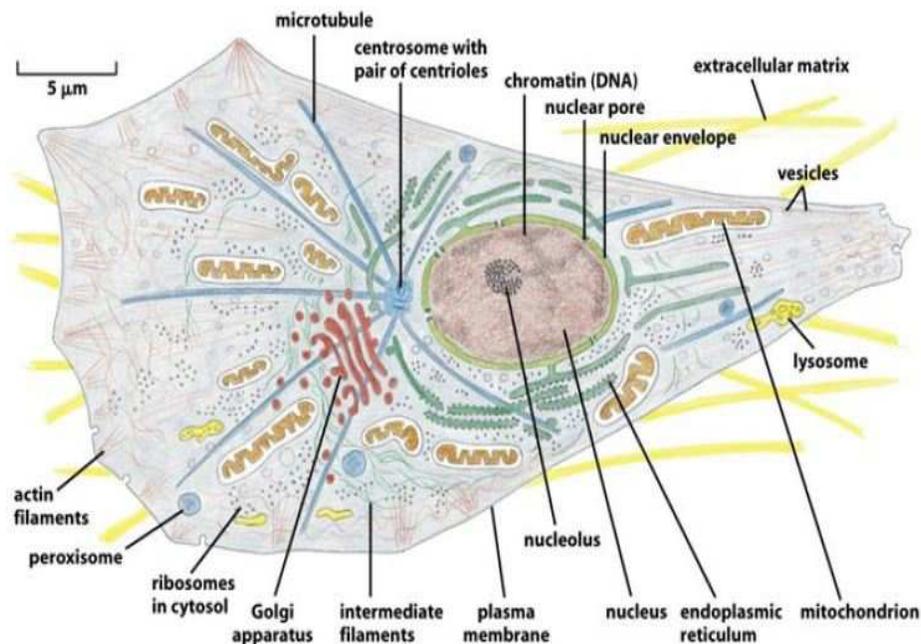


FIGURE 1.2. Cell Structure. *Human cells have various shapes and sizes, but most of them have similar components, some of which are: the nucleus, where the host's genome is packaged as DNA strands; a cytoplasm, which contains the cell's organelles; a nuclear envelope, that separates the nucleus from the cytoplasm; a plasma (cell) membrane, which gives the cell its shape and through which materials are exported and imported; microtubules, which serve as the cell's scaffold and are involved in cell movement and trafficking of particles in the cytoplasm. Source: [2].*

immature virion develops a core, which contains RNA and other proteins [3, 10] (see Figures 1.2, 1.3, and 1.4).

### 1.3. HIV MATURATION

New HIV virions that form near the plasma membrane of a host cell escape the cell during the process of budding. These immature virions are not yet infectious. They need to undergo a maturation process in order to be able to invade and infect other cells. During maturation, Gag proteins are cleaved by the enzyme protease. The MA domain remains

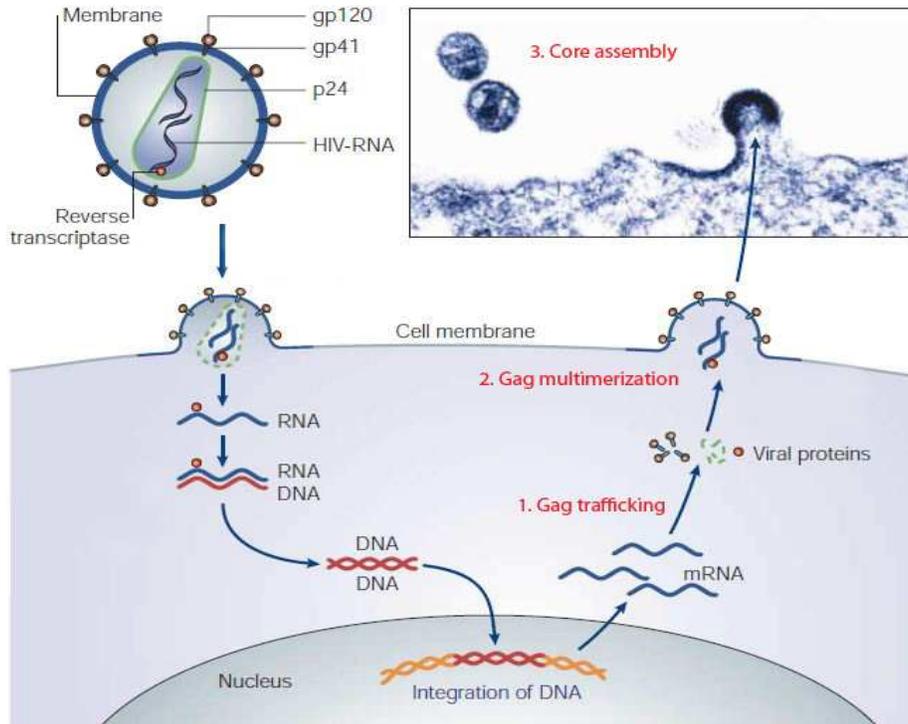


FIGURE 1.3. Retroviral Lifecycle. *HIV fuses with the host cell's plasma membrane and releases its RNA and enzymes in the cytoplasm. HIV's RNA is translated to DNA by the enzyme reverse transcriptase. The viral DNA is integrated into the host cell's DNA by the enzyme integrase. The host cell translates the viral DNA into viral mRNA. The mRNA produces Gag proteins in the cytoplasm, which travel to the cell membrane, the site of assembly of new immature virions. The virions leave the cell during the budding process, and develop a core and become infective during maturation. Source: [11] (with slight modification).*

attached to the virion's membrane, while about 1,500 of the CA proteins form the outer shell or capsid of the mature virion's core. NC is bound to two RNA strands, which are packaged inside the core, along with other viral proteins [3, 8, 10, 19] (see Figure 1.5).

**HIV Core Assembly.** The outer shell of the mature HIV core or capsid (see Figure 1.6) is composed of about 1,500 Gag capsid domain (CA) proteins [8, 20], bound together by weak inter-subunit interactions. The HIV core is a globally stable lattice. The stability of protein-protein interactions is required for assembly [3]. HIV core assembly consists of two main stages: nucleation and elongation.

**Nucleation and Elongation.** *Nucleation* is the formation of a nucleus (polygon) and is usually characterized by a lag phase [2, 21, 22]. For the HIV core, we assume the nucleus size to be  $n_{nuc} = 6$ , since CA hexamers are the main subunit in the process of HIV core assembly [18]. *Elongation* is the addition of subunits to the already formed nucleus,

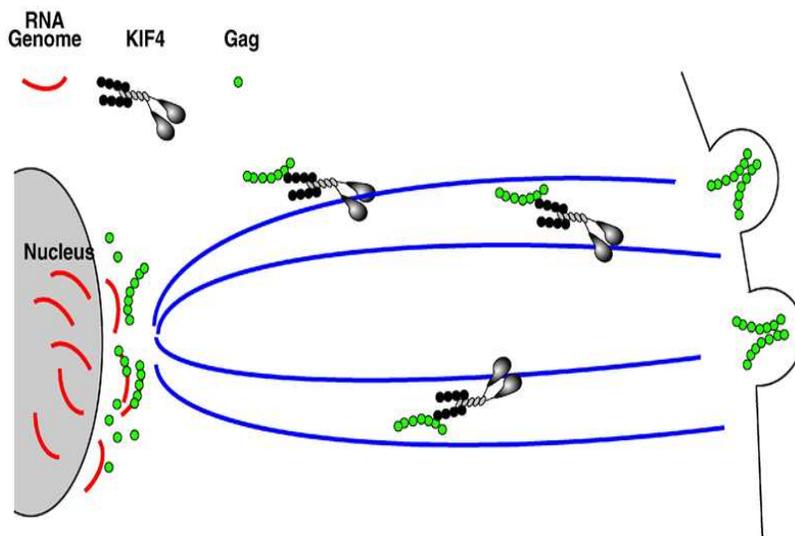


FIGURE 1.4. HIV Egress. *New mRNA forms in the host cell's nucleus and travels to the cytoplasm, where it synthesizes new Gag proteins. Studies suggest that these proteins can move through the cytoplasm via diffusion and active transport along microtubules, while attached to the motor protein KIF4 and carrying RNA, towards the cell membrane and form new HIV virions, which then bud off the cell.* **Source:** [13].

and it is usually characterized by a rapid growth phase and an equilibrium phase [2] . The process of elongation is typically fast compared to nucleation, in terms of the rate at which subunits are added to the growing structure [22].

#### 1.4. ROLES OF MATHEMATICAL MODELING FOR HIV AND GAG

Quantitative description of the aforementioned biological problems, in particular mathematical models, will help us understand the mechanisms of the HIV life cycle. Well-developed mathematical models shall provide quantitative characterization of detailed aspects of HIV reproduction. Here, we highlight the efforts and contributions of this thesis.

**Models for Gag Trafficking Inside Cytoplasm.** We first consider a mathematical model for Gag trafficking in the cytoplasm. We assume the cytoplasm is an annulus, where

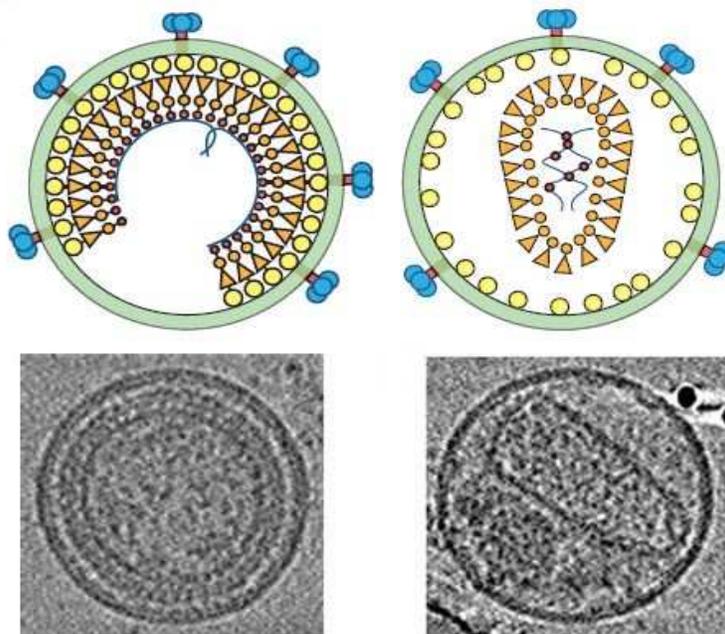


FIGURE 1.5. Immature and Mature HIV Virions. *About 5,000 Gag proteins are spread radially and uniformly inside the immature HIV virion (left). In the mature HIV virion (right), Gag's MA domain (yellow) is attached to the inner layer of the virion's membrane, while about 1,500 of the available 5,000 CA proteins form the capsid (outer shell) of the virion's core. The core packages two strands of viral RNA and other proteins. Source: [10].*

Gag proteins are produced uniformly at a constant rate. The model consists of a partial differential equation with a diffusion term and a convection term, accounting for the diffusion and active transport of Gag proteins. Active transport may take place, for instance, along microtubules via motor proteins. We also assume radial symmetry in the transport of Gag proteins. Our Robin boundary conditions reflect the assumptions that Gag proteins do not penetrate back into the nucleus and that no Gag proteins escape the cell through the plasma membrane until a Gag concentration threshold has been reached. The initial condition reflects the assumption that there are no Gag proteins in the cytoplasm at the start of the process. We perform numerical simulations for various combinations of parameter values. Our simulations agree principally with experimental results. We also comment on further work on the trafficking model by considering stochasticity of the parameters, in particular the active transport parameter.

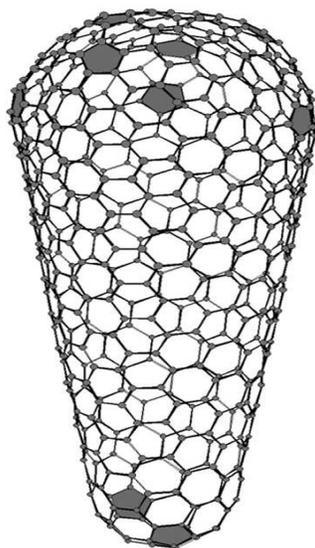


FIGURE 1.6. HIV Core. *The mature HIV core shell is composed of about 1,500 CA proteins, mostly in the form of hexamers. The core also contains 12 CA pentamers, 5 of them in the narrow end, and the other 7 in the wide end of the capsid. This arrangement gives the HIV core its conical shape. Source: [18].*

**Models for Gag Trimerization at Plasma Membrane.** We also develop and analyze a mathematical model for Gag trimerization at the plasma membrane. HIV Gag accumulates near the plasma membrane at certain “Gag hotspots”, where HIV virion assembly takes place. We assume that Gag proteins arrive at the site of assembly as monomers, at a constant rate. Three monomers can come together to form a trimer, and one trimer can decompose into three monomers. Trimers can come together to form higher order multimers, which bind to each other and to the plasma membrane as a new immature HIV virion assembles. Our trimerization model consists of a nonlinear dynamical system of two ordinary differential equations. The equations represent the rates of change in the concentrations of Gag monomers and trimers at a Gag hotspot. We show analytically and verify numerically the existence and stability of a unique equilibrium of the monomer and trimer species, regardless of parameter values, as long as all parameters are positive.

In addition, we estimate the ratios of Gag association and dissociation parameters and the ratio of Gag monomer arrival and Gag multimerization parameters based on values from literature and from experimental data. We also derive a condition on the model parameters that shift the Gag monomer-trimer equilibrium towards the trimer state. We verify the condition numerically. We then calculate a lower bound for the equilibrium association constant  $K_a$  for Gag monomers and trimers. Finally, we comment on further work on Gag trimerization modeling in terms of possible improvements to the model.

**Models for HIV Core Assembly.** We paid out attention to HIV core assembly also. We utilize nonlinear systems of ordinary differential equations for representing the rates of change of the concentrations of CA intermediates in the assembly of a mature HIV virion core.

The first model, based on a model proposed by Zlotnick and others for virus capsid assembly [23], considers hundreds of species of CA intermediates, from CA monomers up to a full capsid (mature HIV core shell). We ran numerical simulations for a total of  $N = 1,300$  and  $N = 5,000$  CA intermediates, assuming higher association rates and lower dissociation rates for larger CA multimer species. Our simulations provide verification of the existence of a stable equilibrium. We showed analytically the existence of a unique equilibrium, and also showed that the equilibrium is stable regardless of parameter values, as long as all parameters are positive, for the case when there are only three CA intermediates.

Next, we consider a cascaded dynamical system for HIV core assembly. There are two subsystems, the first one corresponding to the nucleation of a CA hexamer, and the second one corresponding to the elongation of the mature virion capsid. The first system consists of six nonlinear ordinary differential equations, representing the rates of change of the concentrations of CA intermediates 1 through 6, that is, monomers through hexamers. We consider various paths of assembly of these intermediates. We verify numerically the existence of a stable equilibrium for this six-species model. The elongation system of differential equations is similar to the model proposed by Zlotnick's group for capsid assembly, adapted to HIV core assembly. In particular, there are about  $N = 250$  differential equations (instead of  $N = 1,500$  equations), corresponding to the concentrations of CA 6-mers, 12-mers, ..., and capsid. That is, the basic subunit is now a CA hexamer instead of a monomer. We also comment on further work on HIV core assembly.

## CHAPTER 2

# GAG TRAFFICKING INSIDE THE CYTOPLASM

### 2.1. BIOLOGICAL BACKGROUND AND EXISTING WORK

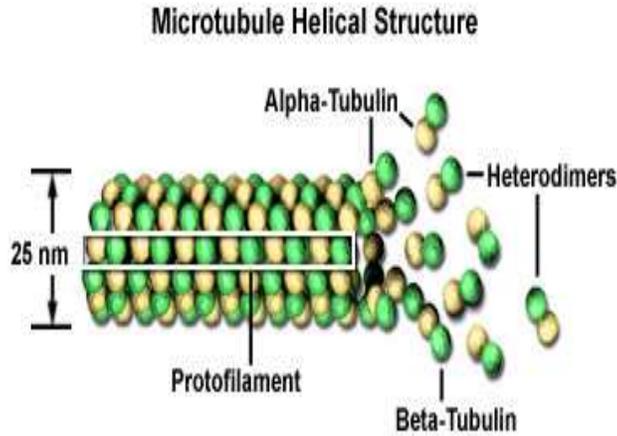


FIGURE 2.1. Microtubule Structure. *Microtubules are hollow cylinders, about 25nm in diameter, and are part of the cell's cytoskeleton.* **Source:** [24].

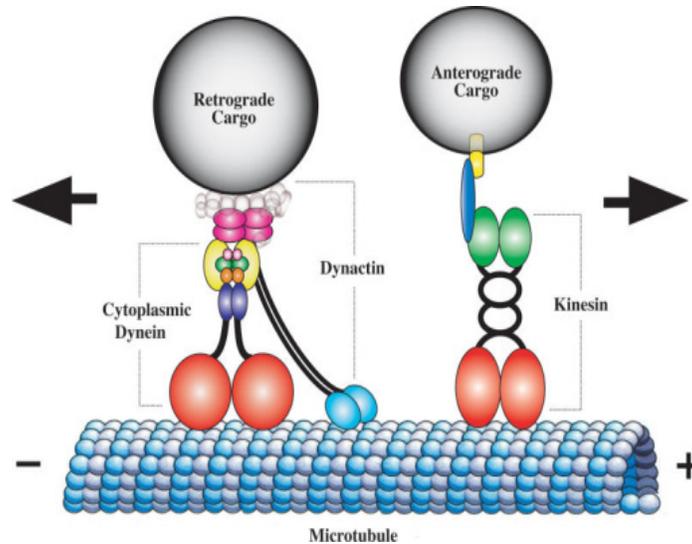


FIGURE 2.2. Motor Proteins. *Dyneins move in retrograde fashion, towards the minus-end of microtubules. Most kinesins exhibit anterograde motion, towards the plus-end of microtubules.* **Source:** [25].

Proteins and other particles may be transported inside the cell along microtubules (see Figure 2.1). These structures made of tubulin are hollow tubes, about 25 nm in diameter and

between 50 and 100  $\mu\text{m}$  long [28]. There are typically about 150 microtubules in a cell [28]. Microtubules radiate from the microtubule organization center (MOTC) near the nucleus towards the plasma membrane. They serve as tracks for motor proteins, which attach to cellular cargoes and move along microtubules either towards the minus-end (dyneins) or the plus-end (kinesins) of the microtubules (see Figures 2.2 and 2.3). Studies have found that motor proteins for movement towards the plus-end and the minus-end of microtubules are present simultaneously on cellular cargo [29].

The mechanisms of intracellular HIV Gag trafficking and virus assembly are poorly understood [7, 30]. Gag proteins are present in the cytosol mostly as monomers and low order oligomers [6, 9, 10]. In experiments, it has been found that kinesin KIF4 associates with retroviral Gag protein [31]. Moreover, KIF4 disruption slows down HIV Gag trafficking in the cytoplasm and inhibits production of new virus particles [14]. There is also evidence that disrupting microtubules with the drug nocodazole reduces HIV infection twofold [32]. In addition, cells contain a stable pool of nocodazole-resistant microtubules [14]. Martinez

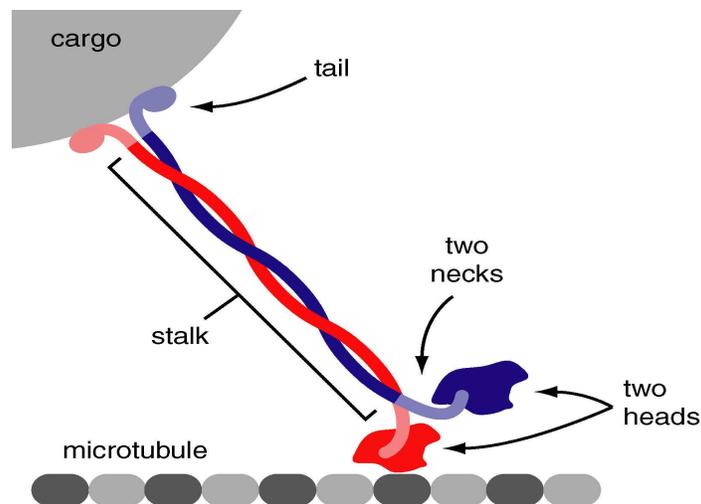


FIGURE 2.3. Structure of Kinesin. *Kinesin uses its two heads to walk along microtubules in 8nm steps [26]. Kinesin's tail is attached to cargo, which is transported along the microtubule towards the plasma membrane. Source: [27].*

and co-workers propose that HIV Gag trafficking might depend mostly on this stable pool of microtubules, whose selection could be enhanced by association with KIF4 [14].

In the following sections, we present and analyze a mathematical model for intracellular HIV Gag trafficking. Our model will consider Gag transport towards the plasma membrane. We aim to better understand the dynamics of Gag protein transport. Specifically, we look at the relationships between diffusion and active transport parameters and the assembly time of new virus particles at the plasma membrane. The numerical methods used in simulations are based on finite elements and characteristic tracking [16, 33, 34, 35, 36, 37, 38, 39].

## 2.2. A QUASI-2D MODEL FOR GAG TRAFFICKING

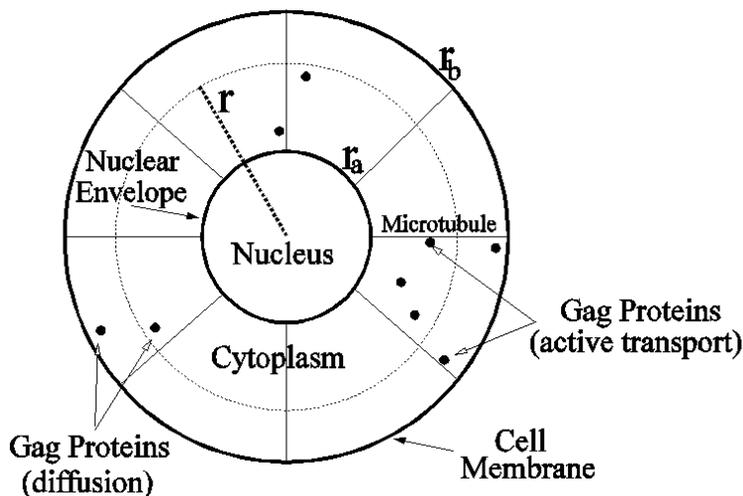


FIGURE 2.4. Annular Model of the Cytoplasm. *Our Gag trafficking model assumes that the host cell's cytoplasm has annular form and that Gag proteins are transported via diffusion and active transport. The transport is assumed to be symmetric with respect to the angle. Thus, the model considers motion in the radial direction. Here,  $r_a$  and  $r_b$  represent the radius of the nucleus and of the cell, respectively. Source: [16].*

2.2.1. ASSUMPTIONS. We list here the main assumptions of our mathematical model for Gag trafficking (see Figure 2.4):

- The cell's cytoplasm is an annulus.
- All Gag species are treated as quasi-monomers.
- There is no association or dissociation among Gag species during intracellular transport.
- Gag transport happens via diffusion and active transport along microtubules.
- Gag monomers are produced uniformly and constantly from mRNA in the cytoplasm.
- Gag particles do not penetrate back into the nucleus.
- Gag particles cannot escape through the plasma membrane before virions form.
- Initial monomeric Gag concentration in the cytoplasm is zero.

2.2.2. GAG ACTIVE TRANSPORT. **Variables and Parameters.** We define here the main quantities involved in our model for Gag trafficking:

- $s$  = speed of Gag on microtubules in the radial direction ( $\mu\text{m}/\text{sec}$ )
- $\mathbf{v} = (s \cos \theta, s \sin \theta)$  = velocity vector ( $\mu\text{m}/\text{sec}, \mu\text{m}/\text{sec}$ )
- $D$  = diffusion coefficient ( $\mu\text{m}^2/\text{sec}$ )
- $g$  = rate of production of new Gag protein ( $\mu\text{M}/\text{sec}$ )
- $P(r, \theta, t) = P(r, t)$  = concentration of Gag at distance  $r$  from the cell center at time  $t$  ( $\mu\text{M}$ )
- $T_v$  = time when new *virion* first appears on cell membrane, that is, the Gag concentration near plasma membrane reaches a certain threshold value (sec).

**Mass Conservation.** The mass conservation law asserts that temporal change in mass concentration is balanced by spatial change (flux) plus production (source, sink). Mathematically, this can be expressed as the following partial differential equation:

$$\frac{\partial P}{\partial t} = \nabla \cdot (D\nabla P - \mathbf{v}P) + g.$$

In polar coordinates, the total flux can be expressed as follows:

$$\nabla \cdot (\mathbf{v}P - D\nabla P) = \frac{1}{r} \frac{\partial}{\partial r} (srP - Dr \frac{\partial P}{\partial r}).$$

This leads to a mathematical model for active transport of quasi-monomers of Gag protein:

$$(1) \quad \left\{ \begin{array}{l} P_t = \frac{1}{r} \frac{\partial}{\partial r} (Dr \frac{\partial P}{\partial r} - srP) + g, r \in (r_a, r_b), t \in (0, T_v) \\ (Dr \frac{\partial P}{\partial r} - srP)|_{r_a} = (Dr \frac{\partial P}{\partial r} - srP)|_{r_b} = 0, t \in (0, T_v) \\ P(r, 0) = 0, r \in (r_a, r_b). \end{array} \right.$$

Robin boundary conditions with zero total flux at the two ends are proposed to reflect the assumption of no penetration back into the nucleus and no escape through the plasma membrane before new virions form. Numerical algorithms for the above initial boundary value problem can be developed based on characteristic finite element methods.

2.2.3. VARIATIONAL FORMULATION FOR THE QUASI-2D MODEL. We rewrite equation

(1) in an equivalent form as follows:

$$(2) \quad r \frac{\partial p}{\partial t} + \frac{\partial}{\partial r} \left( srp - Dr \frac{\partial p}{\partial r} \right) = gr, \quad r \in [r_a, r_b], \quad t \in [0, T_v].$$

The development of the numerical method is divided into two stages. A weak formulation based on a temporal discretization is first established. Then we apply a spatial finite element discretization to obtain a discrete algebraic system.

Let  $0 < t_1 < \dots < t_{n-1} < t_n < \dots < t_N = T_v$  be a temporal partition of  $[0, T_v]$  that is not necessarily uniform and  $\Delta t_n = t_n - t_{n-1}$  ( $n = 1, \dots, N$ ). We consider test functions defined on the space-time slab  $\Sigma = [r_a, r_b] \times [t_{n-1}, t_n]$ . Multiplying both sides of equation (2) by a typical test function  $\psi(r, t)$ , and integrating with respect to space and time, we obtain

$$(3) \quad \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} \left( r \frac{\partial p}{\partial t} + \frac{\partial}{\partial r} \left( srp - Dr \frac{\partial p}{\partial r} \right) \right) \psi dr dt = \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} gr \psi dr dt$$

For the first term on the left-hand side of equation (3), Fubini's theorem and integration by parts in time yield

$$(4) \quad \begin{aligned} \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} r \frac{\partial p}{\partial t} \psi dr dt &= \int_{r_a}^{r_b} \left( \int_{t_{n-1}}^{t_n} \frac{\partial p}{\partial t} \psi dt \right) r dr \\ &= \int_{r_a}^{r_b} p(r, t_n) \psi(r, t_n) r dr - \int_{r_a}^{r_b} p(r, t_{n-1}) \psi(r, t_{n-1}^+) r dr \\ &\quad - \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} p \frac{\partial \psi}{\partial t} r dr dt. \end{aligned}$$

For the second term on the left-hand side of equation (3), we apply integration by parts in space and the boundary condition in (1) to obtain

$$\begin{aligned}
& \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} \frac{\partial}{\partial r} \left( srp - Dr \frac{\partial p}{\partial r} \right) \psi dr dt \\
&= \int_{t_{n-1}}^{t_n} \left[ srp - Dr \frac{\partial p}{\partial r} \right]_{r_a}^{r_b} \psi dr dt \\
(5) \quad & - \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} \left( srp - Dr \frac{\partial p}{\partial r} \right) \frac{\partial \psi}{\partial r} dr dt \\
&= \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} \left( Dr \frac{\partial p}{\partial r} - srp \right) \frac{\partial \psi}{\partial r} dr dt.
\end{aligned}$$

A combination of the above two equations leads to

$$\begin{aligned}
& \int_{r_a}^{r_b} p(r, t_n) \psi(r, t_n) r dr - \int_{r_a}^{r_b} p(r, t_{n-1}) \psi(r, t_{n-1}^+) r dr \\
(6) \quad & - \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} p \frac{\partial \psi}{\partial t} r dr dt + \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} \left( Dr \frac{\partial p}{\partial r} - srp \right) \frac{\partial \psi}{\partial r} dr dt \\
&= \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} g \psi r dr dt,
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
& \int_{r_a}^{r_b} p(r, t_n) \psi(r, t_n) r dr + \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} D \frac{\partial p}{\partial r} \frac{\partial \psi}{\partial r} r dr dt \\
& - \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} p \left( \frac{\partial \psi}{\partial t} + s \frac{\partial \psi}{\partial r} \right) r dr dt \\
&= \int_{r_a}^{r_b} p(r, t_{n-1}) \psi(r, t_{n-1}^+) r dr + \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} g \psi r dr dt.
\end{aligned}$$

We require each space-time test function to satisfy the adjoint equation

$$(7) \quad \frac{\partial \psi}{\partial t} + s \frac{\partial \psi}{\partial r} = 0.$$

Clearly, this implies that any test function is a constant along each characteristic, i.e.,

$$(8) \quad \psi(r, t) = \psi(r^*, t^*).$$

We have now a weak formulation

$$(9) \quad \begin{aligned} & \int_{r_a}^{r_b} p(r, t_n) \psi(r, t_n) r dr + \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} D \frac{\partial p}{\partial r} \frac{\partial \psi}{\partial r} r dr dt \\ &= \int_{r_a}^{r_b} p(r, t_{n-1}) \psi(r, t_{n-1}^+) r dr + \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} g \psi r dr dt. \end{aligned}$$

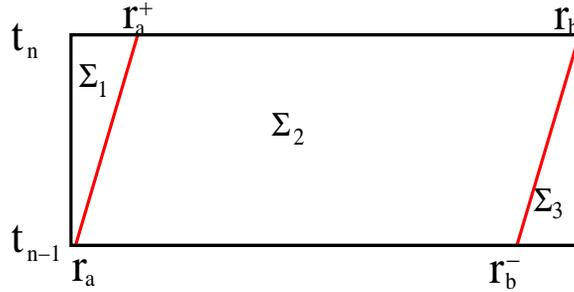


FIGURE 2.5. The space-time slab  $\Sigma = [r_a, r_b] \times [t_{n-1}, t_n]$  is divided into three regions  $\Sigma_1, \Sigma_2, \Sigma_3$  by two special characteristics: the one connecting  $(r_a, t_{n-1})$  and  $(r_a^+, t_n)$ , and the one connecting  $(r_b^-, t_{n-1})$  and  $(r_b, t_n)$ .

As shown in Figure 2.5, the space-time slab  $[r_a, r_b] \times [t_{n-1}, t_n]$  is divided into three regions by two special characteristics: the one connecting  $(r_a, t_{n-1})$  and  $(r_a^+, t_n)$ , and the one connecting  $(r_b^-, t_{n-1})$  and  $(r_b, t_n)$ , where

$$(10) \quad r_a^+ = r_a + s \Delta t_n, \quad r_b^- = r_b - s \Delta t_n.$$

2.2.4. FINITE ELEMENT DISCRETIZATION. Next we consider spatial discretization and approximate the diffusion term in the weak form (9). For time step  $t_n$ , let  $r_a = r_0 < r_1 < \dots < r_{j-1} < r_j < \dots < r_M = r_b$  with  $h_j = r_j - r_{j-1}$  ( $j = 1, \dots, M$ ) be a nonuniform spatial

mesh. We shall adopt local mesh refinements and Lagrangian  $\mathcal{P}_1$  spatial elements at each time step. Let  $\phi_j(r)(j = 0, \dots, M)$  be the nodal basis (hat) functions for a typical time step. Note that  $\phi_0$  has only the right half of the hat, whereas  $\phi_M$  has only the left half. It should be pointed out that the spatial partition, the number of spatial elements  $M$ , and the nodal basis functions are generally different for different time steps. We shall assume this difference is clear from context rather than introduce some awkward notations.

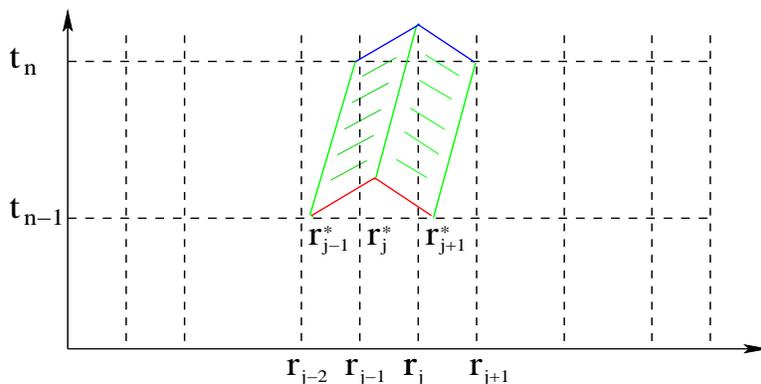


FIGURE 2.6. An illustration of the test functions (except the last one)

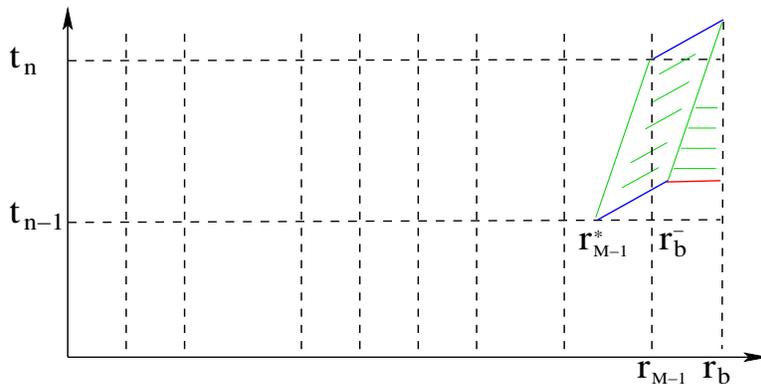


FIGURE 2.7. An illustration of the last test function  $\psi_M(r, t)$

Our finite element approximation is different from the traditional ones. The trial basis functions are spatial hat functions as described above, whereas the test basis functions are

space-time functions that satisfy the adjoint equation (7). To be precise, let us look at a typical time stepping from  $t_{n-1}$  to  $t_n$ . The unknown approximate solution  $P^{(n)}(r)$  at time step  $t_n$  is a linear combination of the spatial hat functions  $\phi_j(j = 0, \dots, M)$  with unknown coefficients  $\{P_j^{(n)}\}_{j=0}^M$ . There are  $M+1$  space-time test functions  $\psi(r, t)$  that are defined in the space-time slab  $\Sigma = [r_a, r_b] \times [t_{n-1}, t_n]$  and satisfy the adjoint equation. Any test function is a constant along each characteristic within its support. For  $0 \leq i \leq (M-1)$ ,  $\psi_i(r, t_n) = \phi_i(r)$ . So the support of  $\psi_0(r, t)$  is a region in  $\Sigma$  that is bounded by the boundary of  $\Sigma$  and the backtracking characteristic starting from  $(r_1, t_n)$ . The support of  $\psi_i(r, t)(i = 1, \dots, M-1)$  is a region in  $\Sigma$  that is bounded by the two backtracking characteristics starting from  $(r_{i-1}, t_n)$  and  $(r_{i+1}, t_n)$ , respectively. An illustration of these test functions is shown in Figure 2.6. However, the last test function  $\psi_M(r, t)$  is special. Its support consists of two parts. The first part is the region in  $\Sigma$  that is bounded by the two backtracking characteristics starting from  $(r_{M-1}, t_n)$  and  $(r_M, t_n)$ , respectively. Similarly,  $\psi_M(r, t_n) = \phi_M(r)$ . The second part of the support is  $\Sigma_3$  and  $\psi_M(r, t) \equiv 1$  on  $\Sigma_3$ .

For the diffusion term in the weak form (9), our treatment is similar to that in [38]. Because the diffusion is small, the test functions are constants along characteristics, and the solution changes the least along characteristics, we approximate the diffusion term as

$$(11) \quad \int_{t_{n-1}}^{t_n} \int_{r_a}^{r_b} D \frac{\partial P}{\partial r} \frac{\partial \psi}{\partial r} r dr dt \approx \int_{r_a}^{r_b} \Delta t(r, t_n) D \frac{\partial P^{(n)}(r)}{\partial r} \frac{\partial \psi}{\partial r}(r, t_n) r dr,$$

where  $\Delta t(r, t_n) = t_n - t^*$  and  $(r^*, t^*)$  is the foot of the characteristic starting from  $(r, t_n)$ . This approximation holds for all test basis functions. Clearly, if  $r_a \leq r < r_a^+$ , then  $t^* \in (t_{n-1}, t_n]$  and hence  $\Delta t(r, t_n) < \Delta t_n$ . But if  $r_a^+ \leq r \leq r_b$ , then  $t^* = t_{n-1}$  and hence  $\Delta t(r, t_n) = \Delta t_n$ .

Substituting (9) into (11), we obtain a finite element scheme for equation (1). At each time step  $t_n (1 \leq n \leq N)$ , seek

$$(12) \quad P^{(n)}(r) = \sum_{j=0}^M P_j^{(n)} \phi_j(r)$$

such that for any test function  $\psi(r, t)$ , the following holds

$$(13) \quad \begin{aligned} & \int_{r_a}^{r_b} P^{(n)}(r) \psi(r, t_n) r dr + \int_{r_a}^{r_b} \Delta t(r, t_n) D \frac{\partial P^{(n)}(r)}{\partial r} \frac{\partial \psi}{\partial r}(r, t_n) r dr \\ & = \int_{r_a}^{r_b} P^{(n-1)}(r) \psi(r, t_{n-1}^+) r dr + \int_{t_{n-1}^+}^{t_n} \int_{r_a}^{r_b} g \psi r dr dt. \end{aligned}$$

To start the time-stepping procedure,  $P^{(0)}(r)$  can be taken as the piecewise linear nodal interpolation of the initial condition  $p_0(r)$ .

Next we discuss how to evaluate the mass matrix, the stiffness matrix, and the right-hand side in the discrete linear system. The mass matrix

$$(14) \quad \mathbf{A} = [A_{i,j}]_{(M+1) \times (M+1)} = \left[ \int_{r_a}^{r_b} \phi_i(r) \phi_j(r) r dr \right]_{i,j=0,\dots,M}$$

is a symmetric tridiagonal matrix. In particular, we have

$$(15) \quad \begin{aligned} A_{0,0} &= \frac{1}{12}(r_1 - r_0)(r_1 + 3r_0), & A_{0,1} &= \frac{1}{12}(r_1^2 - r_0^2), \\ A_{i,i-1} &= \frac{1}{12}(r_i^2 - r_{i-1}^2), & A_{i,i+1} &= \frac{1}{12}(r_{i+1}^2 - r_i^2), \\ A_{i,i} &= \frac{1}{12}(r_i - r_{i-1})(3r_i + r_{i-1}) + \frac{1}{12}(r_{i+1} - r_i)(r_{i+1} + 3r_i), \\ A_{M,M-1} &= \frac{1}{12}(r_M^2 - r_{M-1}^2), & A_{M,M} &= \frac{1}{12}(r_M - r_{M-1})(3r_M + r_{M-1}), \end{aligned}$$

where  $1 \leq i \leq (M - 1)$ .

The stiffness matrix

$$(16) \quad \mathbf{B} = [B_{i,j}]_{(M+1) \times (M+1)} = \left[ \int_{r_a}^{r_b} \Delta t_n(r, t_n) D \frac{\partial \phi_i}{\partial r} \frac{\partial \phi_j}{\partial r} r dr \right]_{i,j=0,\dots,M}$$

is also a symmetric tridiagonal matrix, but its evaluation is a little bit more involved, since  $\Delta t_n(r, t_n)$  relies on characteristic tracking. Suppose  $r_a^+ \in (r_{I-1}, r_I]$  for some  $I \geq 1$  (but  $I \ll M$ ). It is known that  $\Delta t(r, t_n) = \Delta t_n$  for  $r \geq r_a^+$ . So explicit expressions are available for the following entries:

$$(17) \quad \begin{aligned} B_{i,i-1} &= -\Delta t_n D \frac{1}{2} \frac{r_i + r_{i-1}}{r_i - r_{i-1}}, & B_{i,i+1} &= -\Delta t_n D \frac{1}{2} \frac{r_{i+1} + r_i}{r_{i+1} - r_i}, \\ B_{i,i} &= \Delta t_n D \frac{1}{2} \left( \frac{r_i + r_{i-1}}{r_i - r_{i-1}} + \frac{r_{i+1} + r_i}{r_{i+1} - r_i} \right), & \text{for } (I+1) \leq i \leq (M-1), \\ -B_{M,M-1} &= B_{M,M} = \Delta t_n D \frac{1}{2} \frac{r_M + r_{M-1}}{r_M - r_{M-1}}. \end{aligned}$$

For the evaluation of  $B_{0,0}, B_{0,1} = B_{1,0}, B_{1,1}, B_{1,2}$  until  $B_{I,I-1}, B_{I,I}$ , Gaussian quadratures could be employed.

The evaluation of the right-hand side of (13) depends on the test functions and characteristic tracking. For the first few test basis functions, their supports do not intersect with  $[r_a, r_b] \times \{t_{n-1}\}$ , so  $\psi(r, t_{n-1}^+) = 0$  and hence the corresponding entries are zero. For a test function whose support intersects with  $[r_a, r_b] \times \{t_{n-1}\}$ , we apply change of variable to evaluate the integral. For simplicity of presentation, we assume that both  $\text{supp } \psi(r, t_n)$  and  $\text{supp } \psi(r, t_{n-1}^+)$  are in  $[r_a, r_b]$ . We replace the dummy integral variable  $r$  by  $r^*$  and rewrite the integral as

$$\int_{r_a}^{r_b} P^{(n-1)}(r^*) \psi(r^*, t_{n-1}^+) r^* dr^*.$$

Note that there is a one-to-one mapping between  $\text{supp } \psi(r, t_n)$  and  $\text{supp } \psi(r, t_{n-1}^+)$ , so we have [35]

$$\int_{r_a}^{r_b} P^{(n-1)}(r^*) \psi(r^*, t_{n-1}^+) r^* dr^* = e^{-2s\Delta t_n} \int_{r_a}^{r_b} P^{(n-1)}(r^*) \psi(r, t_n) r dr.$$

The above right-hand side could be evaluated by a Gaussian quadrature on  $\text{supp } \psi(r, t_n)$ , therein  $P^{(n-1)}(r^*)$  is evaluated through linear interpolation after  $r^*$  is located.

Once again the evaluation of the right-hand side involving the last test basis function  $\psi_M(r, t)$  is divided into two parts due to the two-part structure of its support. The calculation on the left part is similar to what is just described above. The right part is simply

$$\int_{r_b^-}^{r_b} P^{(n-1)}(r) r dr,$$

because  $\psi_M(r, t) \equiv 1$  on  $\Sigma_3$ . An explicit formula is available for the above integral, since  $P^{(n-1)}(r)$  is piecewise linear on  $[r_b^-, r_b]$ .

Finally, we obtain a tridiagonal linear system with unknowns  $\{P_j^{(n)}\}_{j=0}^M$  and the coefficient matrix  $(\mathbf{A} + \mathbf{B})$ . The linear system can be solved directly with spatial and temporal complexities both  $\mathcal{O}(M)$ .

**Mass Conservation.** By construction,  $\sum_{i=0}^M \psi_i(r, t) \equiv 1$  on the space-time slab  $[r_a, r_b] \times [t_{n-1}, t_n]$ . Combined with (13), this implies that

$$(18) \quad \int_{r_a}^{r_b} P^{(n)}(r) r dr = \int_{r_a}^{r_b} P^{(n-1)}(r) r dr.$$

Therefore, our numerical scheme is mass-conservative by design. Since piecewise linear approximations to the unknown concentration are adopted in our numerical scheme, an explicit formula for computing the total mass at a typical time step  $t_n$  is available as follows

$$(19) \quad \int_{r_a}^{r_b} P^{(n)}(r) r dr = \sum_{j=1}^M \left[ \frac{1}{2} (r_{j-1} + r_j) (r_j P_{j-1}^{(n)} - r_{j-1} P_j^{(n)}) + \frac{1}{3} (r_{j-1}^2 + r_{j-1} r_j + r_j^2) (P_j^{(n)} - P_{j-1}^{(n)}) \right].$$

However, the evaluations of the coefficient matrix entries and the right-hand side of the linear system involve quadrature errors and round-off errors, the actual total mass might fluctuate slightly from time to time. The above formula can be used for checking mass conservation in numerical experiments.

### 2.3. SIMULATION RESULTS

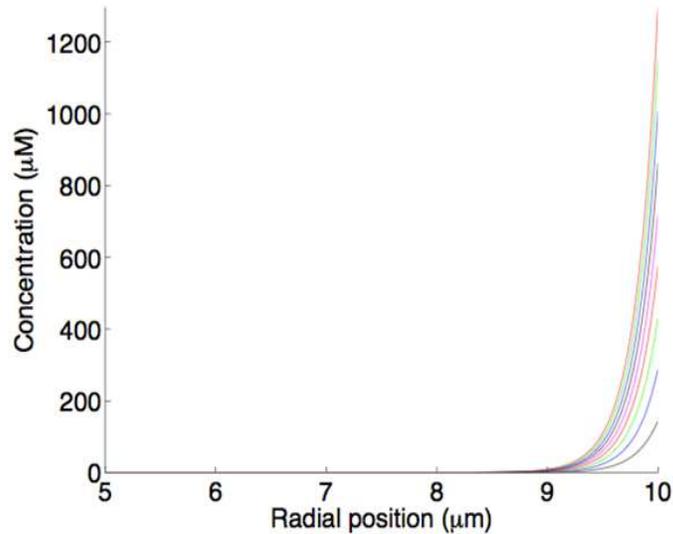


FIGURE 2.8. Simulated Radial Distribution of Gag Concentration. *Speed*  $s = 2$ , *Diffusion*  $D = 0.04$ ,  $r_a = 5$ ,  $r_b = 10$  [40, 41].

We estimate the threshold concentration  $P_v$ , that is, the Gag concentration at the plasma membrane needed for a new HIV virion to form, as follows. We assume it takes 4,500 Gag

TABLE 2.1. Simulated Time of Virion Assembly. **Source:** [16]

Speed $s$ ( $\mu\text{m}/\text{sec}$ )	Time $T_v$
2.00	09h 46m 41s
1.00	10h 14m 35s
0.40	10h 32m 57s
0.20	13h 31m 29s
0.10	21h 09m 08s
0.05	34h 20m 07s

proteins to form a new virion of radius  $0.07 \mu\text{m}$ . These values are within the ranges reported in literature for HIV virions [3, 5, 8, 9, 10, 42, 43]. We then convert 4,500 to *moles* by

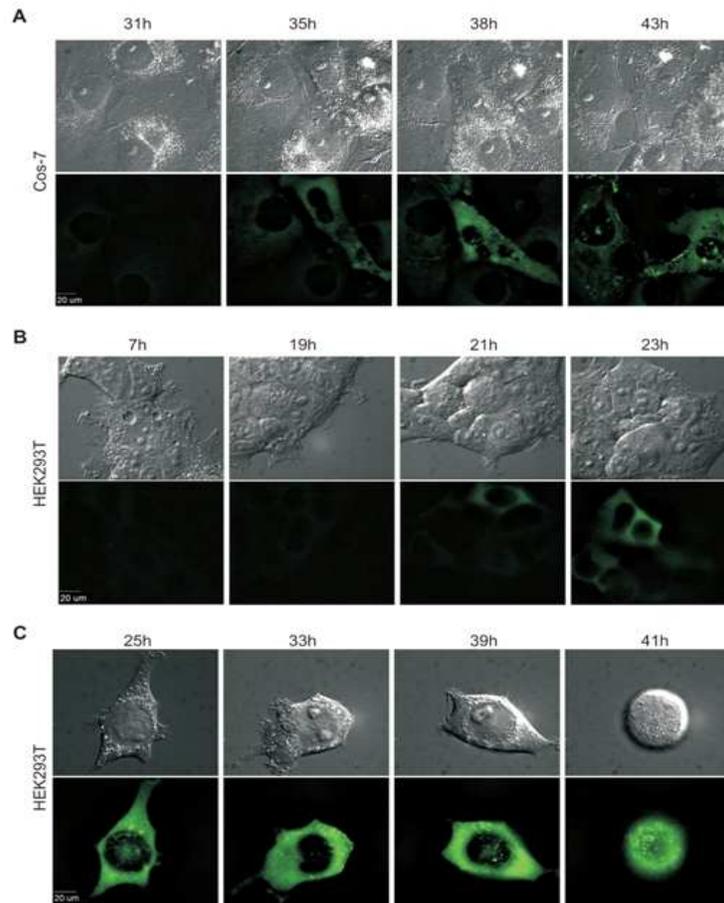


FIGURE 2.9. Experimental Results. *Gag* was tagged with green-fluorescent protein (GFP). A weak GFP signal begins to appear at about 35 h.p.i. on *Cos-7* cells, and at around 20 h.p.i. on *HEK293T* cells. Clearly visible puncta (green dots), which may represent new HIV virions, appear at around 40 h.p.i. on *Cos-7* cells, and at around 25 h.p.i. on *HEK293T* cells. **Source:** [16].

multiplying by  $\frac{1}{6.022 \times 10^{23}}$ . To convert the resulting value into  $\mu\text{M}$ , we divide by the volume of a virion, calculated as the volume of a sphere of radius  $0.07 \mu\text{m}$ . The resulting formula, after simplifying, is  $P_v = \frac{\frac{6.022}{45}}{\frac{4}{3}\pi(0.07)^3} \approx 5201 \mu\text{M}$ .

Similarly, we estimate the value of the parameter  $g$ , the rate of production of new Gag proteins in the cytoplasm, as follows. We assume that 1,500 Gag molecules are produced inside the cytoplasm of a cell each second [40]. We also assume that the cytoplasm is an annulus as depicted in Figure 2.4, with  $r_a = 5\mu\text{m}$  and  $r_b = 10\mu\text{m}$ . To convert to volume, we assume that the cell is a cylinder, with height equal to  $5\mu\text{m}$  [40]. We then convert this value to  $\mu\text{M/s}$  by using the formula  $g = \frac{\frac{15}{6.022}}{\pi(10^2 - 5^2)(5)} \approx 0.002114 \mu\text{M/s}$ .

In addition, we set the value of the diffusion coefficient  $D = 0.04 \mu\text{m}^2/\text{s}$  [16, 28]. The value of the active transport coefficient  $s$  is varied within an acceptable range for transport of cargo via motor proteins along microtubules [2, 32, 41, 44, 45].

Our simulation results agree principally with experimental results (see Figures 2.8 and 2.9, and Table 2.1). If HIV entry into the host cell takes roughly the same time as the HIV egress, then the time post infection for the first virions to appear is about twice the time of egress [16]. In experiments, Nermut et al. found that some cells produced VLPs after 36 h.p.i. and the majority reached VLP budding and release at about 40-44 h.p.i. [15].

## 2.4. PARALLELIZATION OF NUMERICAL SIMULATIONS

**Motivation.** In order to better understand distribution of parameter values, numerical simulations are needed for a very large collection of parameter values  $s, D$ . This is a case of Single Instruction Multiple Data (SIMD) [16].

**Overview.** There are 4 major parameters:  $s, D, \Delta t, g$ . There are also many for-loops in C++ code. We use MPI (message passing interface) for *coarse-grain* parallelism (runs with different sets of parameters) and OpenMP for *fine-grain* parallelism (for-loops inside each run).

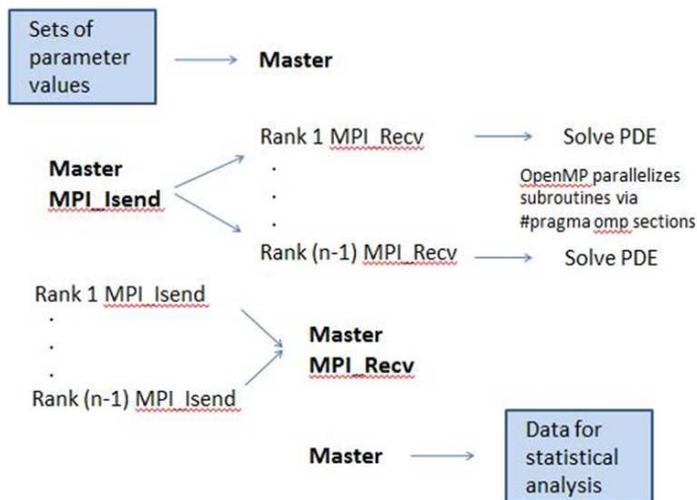


FIGURE 2.10. Parallelization: *Master-Worker*. **Source:** [16].

**MPI and OpenMP.** For MPI parallelization, we use a *master-worker* approach. An input file contains a collection of sets of parameter values. The master rank reads the file and distributes the jobs among the worker ranks. Each worker sends its output, i.e., the calculated virion time  $T_v$ , to the master rank, which then writes the output onto another file for later analysis (see Figure 2.10).

We use OpenMP for parallelization within each worker rank. We first run gprof [47] for parallelability analysis. gprof creates a code profile that helps us determine the parts of the code (for-loops) that take longest to run. Then we use *parallel sections* to parallelize such loops. Our parallel code was implemented and run on the Colorado State University (CSU) Cray supercomputer (see Figure 2.11).

**Parallelization Speedup and Efficiency.** Basic definitions:

- Speedup  $S_i = T_1/T_i$
- Efficiency  $E_i = T_1/(iT_i)$
- $i$  = number of allocated cores
- $T_i$  = time that it takes  $i$  cores to run a program

**Parallel Simulations.** Parameter values:

- $s \in \{1, 1.2, 1.4, \dots, 3.0\} \mu\text{m/s}$
- $D \in \{0.02, 0.04, 0.06\} \mu\text{m}^2/\text{s}$
- $\Delta t \in \{0.125, 0.25, 0.5, 1.0\} \text{ s}$

We fix  $g = 0.002114 \mu\text{M/s}$ . The simulations take longer to finish for smaller values of  $s, D$ , and  $\Delta t$ .

OpenMP parallelization had the most impact when fewer ranks were used. Execution time was reduced by about 44% when 1 worker and 5 threads were used. We also observe that there is no need to run the code with 133 ranks (132 workers). The parallel efficiency was high with 5 threads and between 20 and 40 ranks (see Tables 2.2 and 2.3).

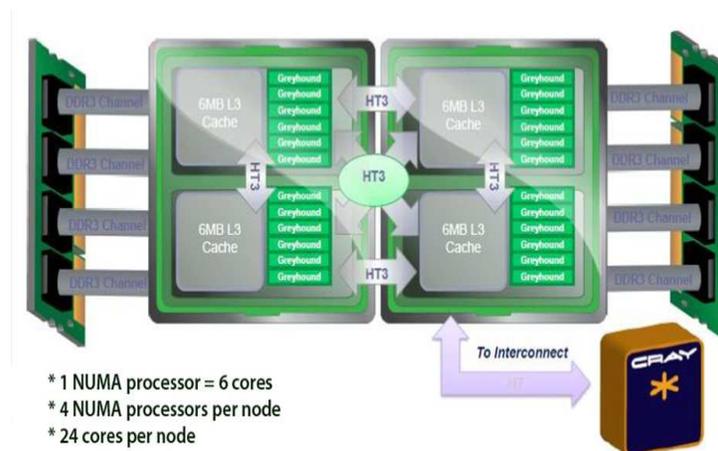


FIGURE 2.11. Cray Architecture. **Source:** [46].

TABLE 2.2. Effects of the number of threads (OpenMP). **Source:** [16].

$n$ (# of ranks)	$d$ (# of threads)	Runtime (seconds)	Speedup
2	1	929.67	
2	5	521.33	1.78
133	1	21.00	
133	2	17.33	1.23

TABLE 2.3. Effects of the number of ranks. **Source:** [16].

$n$ (# of ranks)	$d$ (# of threads)	Runtime (seconds)	Efficiency
2	5	521.33	
20	5	31.33	0.81
30	5	23.33	0.71
40	5	19.00	0.65
50	4	22.00	0.45
133	2	17.33	0.21

TABLE 2.4. Job Unevenness. **Source:** [16].

$n$ (# of ranks)	$d$ (# of threads)	Runtime (seconds)	Efficiency
100	2	21.33	0.23
110	2	24.00	0.18
120	2	22.33	0.18
130	2	17.33	0.21
140	2	17.67	0.20

Executing the code with 110 and 120 workers took longer than with 100 ranks! (See Table 2.4). Possible reasons are:

- There is additional overhead costs for additional allocated cores.
- There is unevenness of jobs: some cores finish their jobs quickly while others take longer.

## 2.5. STOCHASTIC ASPECTS OF GAG TRAFFICKING

Intracellular HIV Gag trafficking is dependent on microtubules and motor proteins, such as kinesin. This protein exhibits predominantly anterograde (plus-ended) movement along

microtubules. Kinesin moves either one or two steps at a time in a rather erratic manner. The protein obtains energy by hydrolyzing ATP in order to move along a microtubule. For each ATP molecule hydrolyzed, conventional kinesin takes one 8nm step forward [44]. At times, the protein stops and even detaches from the microtubule. The kinesin may re-attach and continue its movement toward the cell periphery. Conventional kinesin may travel for hundreds of ATPase cycles before detaching from a microtubule [2, 44, 48].

Cargo attaches to kinesin based on protein binding affinity. The cargo may be membrane-enclosed or free. In the absence of cargo, kinesin proteins tend to exhibit decreased binding to microtubules and inhibited motility [49]. Conformational changes in kinesin might explain this difference in binding and motility. Friedman and Vale [49] observe that cargo-free kinesin is folded in such a way that the head and tail are in close proximity. Removal of the kinesin tail promotes attachment to microtubules and increases microtubule-stimulated ATPase activity. Thus, they conclude that the kinesin tail represses ATPase and the motility of this protein along microtubules [49].

It has been observed that organelles move bidirectionally inside cells. Evidence suggests that both types of motor proteins, kinesins and dyneins, attach to cargo simultaneously, and directionality of movement is determined by a complex mechanism which ensures that when one type of motor is active, the other one is inactive [29, 41]. In addition, the cell type, the pH value, the type of cargo, and other factors can affect intracellular transport speed.

2.5.1. STOCHASTIC TRANSPORT OF GAG ON MICROTUBULES. The speed at which cargoes move along microtubules is still an active area of research. According to Sodeik and others [32], GFP-tagged HIV particles have a peak velocity of 1  $\mu\text{m/s}$  as they travel through the cytoplasm.

Kinesins move along microtubules at a wide range of velocities [45]. Kinesin velocity can be regarded as a random variable with a probability distribution. In experiments, Setz and Surrey [45] estimated the average kinesin velocity to be about  $0.56 \mu\text{m/s}$  (see figure 2.13). In contrast, as observed by Gazzola and co-workers [41] in *in vitro* experiments, intracellular

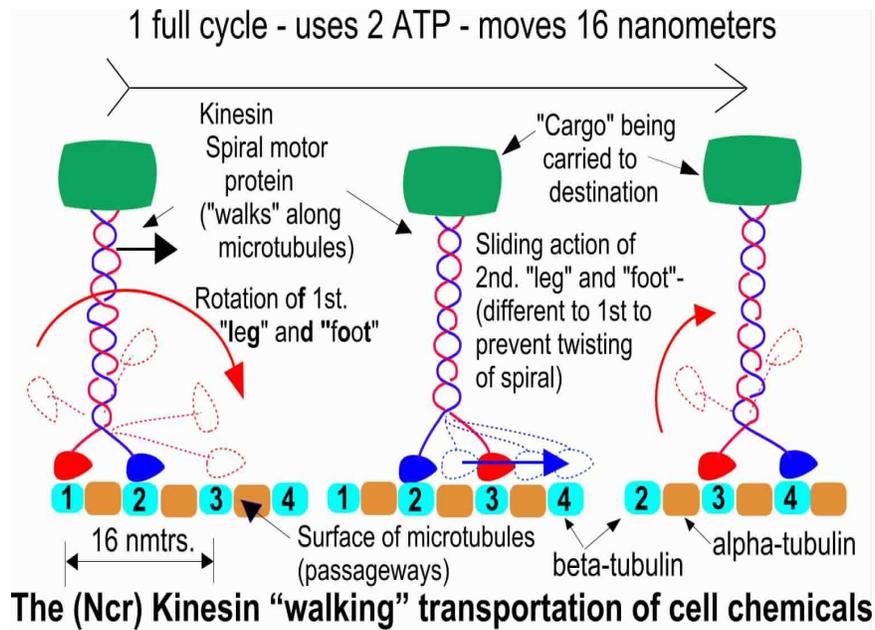


FIGURE 2.12. Kinesin Walking and ATP Cycles. **Source:** [50].

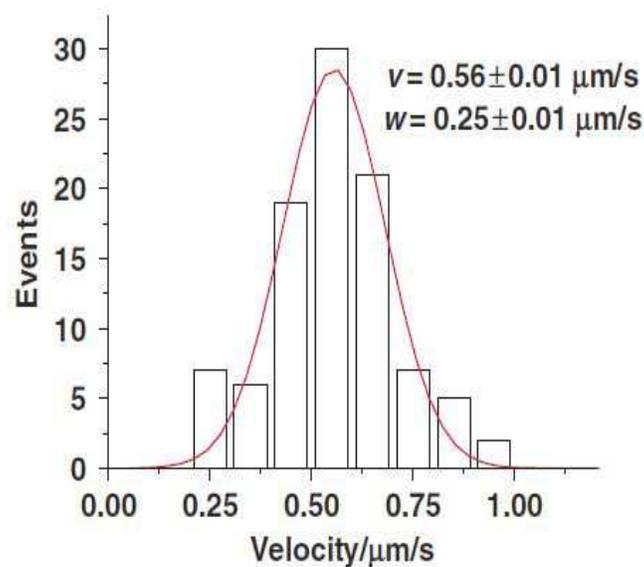


FIGURE 2.13. Distribution of Kinesin Speeds.  $\mathbf{v}$  = mean,  $\mathbf{w}$  = standard deviation. **Source:** [45].

cargo, such as vesicles and intracellular viruses, have exhibited greater velocities than the maximum speeds measured for motor proteins without load. Such maximum velocities listed in [41] are:

- $3 \mu\text{m/s}$  for dynein
- $0.4 \mu\text{m/s}$  for kinesin-1
- $3 \mu\text{m/s}$  for kinesin-1
- $0.8 \mu\text{m/s}$  for kinesin-1
- $0.5 \mu\text{m/s}$  kinesin-2.

Similar motor protein velocities have been reported by Alberts, Coy, Seitz, and others [2, 44, 45].

2.5.2. STOCHASTIC ASPECT OF GAG TRANSPORT AT THE CELL LEVEL. For stochastic analysis of Gag trafficking, one can focus on transport at the cell level instead of considering transport at the molecular level. In other words, the emphasis can be on the overall variability of transport velocities, Gag production rates, and nuclear and cell radii among cells, without taking into consideration the specific mechanisms involved in the intracellular transport of proteins.

We then can consider equation (1), where we have, over the space of all cells:

- Random input: speed ( $s$ ), diffusion ( $D$ ), production rate ( $g$ ), nuclear radius ( $r_a$ ), cell radius ( $r_b$ )
- Random output: the time needed for puncta appearance ( $T_v$ )
- Parameter space: tensor grids vs sparse grids

There are several facts that should be taken into consideration for research on stochastic aspects of HIV Gag trafficking, particularly relating to the speed at which particles are being

transported and the amount of time it takes for Gag particles to assemble at the plasma membrane. Here we list some of the most important data obtained from literature about HIV Gag trafficking.

Some known facts regarding Gag transport and assembly time:

- The first Gag proteins appear in the cytoplasm at 28-30 h.p.i. [15].
- Budding starts 6-8 hours later, that is, 34-38 h.p.i. [15].
- Some cells produce VLPs after 36 h.p.i., while the majority of cells reach the VLP budding and release state at about 40-44 h.p.i. [15].
- A subpopulation of Gag protein reaches the plasma membrane within 5-10 min (after Gag synthesis begins) [7, 15].

Some known facts about the shape and size of VLPs:

- HIV Gag protein assembles into 100 to 120 nm diameter particles in mammalian cells [51].
- Recombinant HIV Gag assembles *in vitro* into particles only 25-30 nm in diameter if inositol phosphates not added [51].
- VLPs are about 100-150 nm in diameter, and contain several thousand Gag molecules [42, 43].
- Gag proteins in immature HIV virions are highly extended rods: length about 20 nm, width of only 2 to 3 nm [43].

Some known facts about diffusion of particles in the cytoplasm are:

- The cell's cytoskeleton (actin microfilaments, microtubules, and others) contribute 3000  $\mu\text{m}^2$  of surface area in a typical mammalian tissue culture cell [28]
- Some typical diffusion coefficients for various cargoes in the cytoplasm are [28]:

- $2.6 \times 10^{-3} \mu\text{m}^2/\text{s}$  for beads in SW3T3 fibroblasts
- $4.0 \times 10^{-3} \mu\text{m}^2/\text{s}$  for CV1 fibroblasts
- $2.5 \times 10^{-2} \mu\text{m}^2/\text{s}$  for endogenous vesicles
- $3.0 \times 10^{-3} \mu\text{m}^2/\text{s}$  for fluorescently stained chromatin granules
- $3.9 \times 10^{-4} \mu\text{m}^2/\text{s}$  to  $7.4 \times 10^{-3} \mu\text{m}^2/\text{s}$  for secretory vesicles tagged with GFP.

## 2.6. REMARKS: BIOLOGICAL IMPLICATIONS

As noted above, our numerical results agree principally with experimental results. Therefore, for appropriate combinations of parameter values it is possible to obtain virion assembly times  $T_v$  that make sense according to the literature and experiments performed. This indicates that the assumptions regarding diffusion and active transport are valid in the case of HIV Gag trafficking:

- Diffusion alone is not enough to form a new HIV virion in a realistic time-frame.
- Gag particles move through the cytoplasm via diffusion and active transport.
- In addition, our simplifying assumptions, such as the annular model of the cytoplasm, linear transport in the radial direction and angular symmetry of transport do not seem to have significant effect on the final results and might be feasible assumptions for most cell types.

The importance of active transport could motivate some new clues for HIV. This means that we could look for ways to suppress or disrupt active transport inside the cytoplasm, for example, disrupting KIF4 or disrupting microtubules.

## 2.7. REMARKS ON FURTHER WORK

Modeling Gag trafficking in the cytoplasm is a new research area in mathematical biology. Presented in this dissertation is some preliminary work. Further work can be done in this new area by developing further mathematical models that have assumptions better reflecting the biological process and/or considering further complexity in the process.

Here we list some ideas for further work:

- (1) Geometric domain other than the annulus could be used to better reflect the shape of cells.
- (2) Solving the stochastic transport equation to better characterize the randomness in the trafficking process. The stochastic system can be formulated as follows:

$$(20) \quad \left\{ \begin{array}{l} P_t = \frac{1}{r} \frac{\partial}{\partial r} (D(\omega) r \frac{\partial P}{\partial r} - s(\omega) r P) + g(\omega), \\ \\ r \in (r_a(\omega), r_b(\omega)), t \in (0, T_v), \omega \in \Omega \\ \\ (D(\omega) r \frac{\partial P}{\partial r} - s(\omega) r P)|_{r_a(\omega)} = (D(\omega) r \frac{\partial P}{\partial r} - s(\omega) r P)|_{r_b(\omega)} = 0, \\ \\ t \in (0, T_v), \omega \in \Omega \\ \\ P(r, 0) = 0, \quad r \in (r_a(\omega), r_b(\omega)), \omega \in \Omega. \end{array} \right.$$

Here, the set  $\Omega$  is an abstract space whose elements are all possible outcomes  $\omega$ . The input parameters  $D, s, g, r_a$  and  $r_b$  are independent random variables. The solution  $P$  is now a random quantity [52].

- (3) In addition, instead of employing characteristic finite element methods to numerically solve the model, one could use Crank-Nicolson or other numerical methods and evaluate the advantages and disadvantages of using these methods.

## CHAPTER 3

# HIV GAG TRIMERIZATION AT THE PLASMA MEMBRANE

### 3.1. GAG TRIMERIZATION AND MULTIMERIZATION

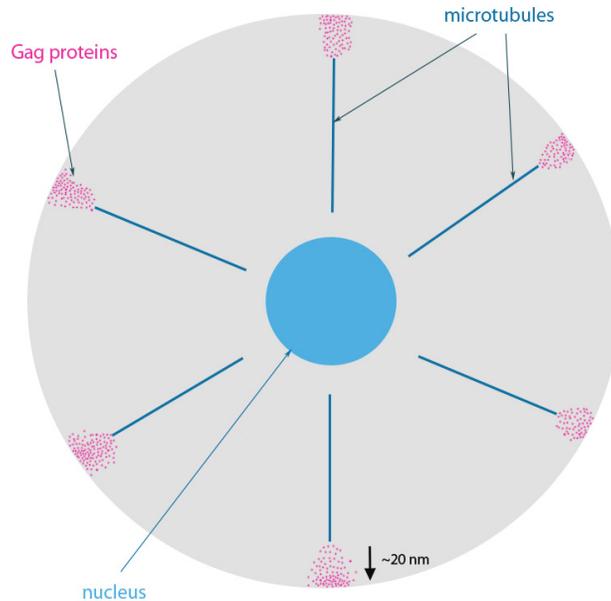


FIGURE 3.1. Gag Hotspots. *The sites of Gag multimerization and HIV assembly comprise about 10% of the plasma membrane [40].*

Gag proteins are transported inside the cell via diffusion and active transport towards the plasma membrane (see Chapter 2). These proteins exist mostly as monomers and low order multimers in the cytoplasm [6, 10, 12, 53]. In particular, Gag monomers and trimers have significant relevance in the process of Gag multimerization and HIV assembly. Gag trimers could be regarded as the basic building blocks or subunits for the immature HIV virion capsid. HIV virus-like particles (VLPs) do not seem to form significantly inside the host cell [9, 10]. Thus, trimers and higher order multimers form mainly at “Gag hotspots” near or on the plasma membrane (see Figure 3.1). They attach to the membrane during the

early stages of HIV assembly and budding. Gag proteins are recruited to the site of assembly either directly from the cytosol [10] or via lipid rafts [6, 54]. As the local Gag concentration at the plasma membrane increases, Gag multimerization is enhanced, which speeds up virion assembly [6, 9, 30, 55, 56]. The stable binding of Gag proteins to the plasma membrane is then crucial for successful virion formation, and it is achieved through a myristoyl switch [6, 30, 55, 56].

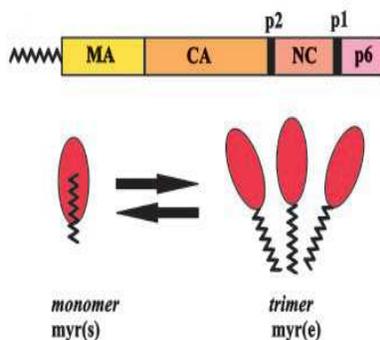


FIGURE 3.2. HIV-1 Gag Domains and Myristoyl Switch. *The main HIV-1 Gag domains are: matrix (MA), capsid (CA), nucleocapsid (NC), and p6. Myristate is bound to the MA domain. It can exist in either sequestered state (myr(s)), in Gag monomers, or exposed state (myr(e)), in Gag trimers.* **Source:** [6].

3.1.1. MYRISTOYL SWITCH. The main HIV-1 Gag precursor Pr55gag domains are: matrix (MA), capsid (CA), nucleocapsid (NC), and p6 (see Section 1.2 and Figure 3.2). In addition, myristate, a 14-carbon saturated fatty acid, covalently modifies the membrane-binding (M) domain of MA at its N-terminal region. This process is called myristoylation. A “myristoyl switch”, which promotes tight anchoring of Gag proteins to the plasma membrane, consists of myristate plus a cluster of basic residues within the M domain that synergizes with myristate [6, 56]. This myristoyl switch is, thus, essential in the assembly of new HIV virions.

3.1.2. GAG STATUS NEAR THE PLASMA MEMBRANE. The myristate moiety in HIV-1 MA exists in two different states: monomeric sequestered ( $\text{myr}(s)$ ) and trimeric exposed ( $\text{myr}(e)$ ) (see Figures 3.2 and 3.3). These two states of myristate in myr-MA exist in equilibrium in the cytosol.

The exposure of myristate is coupled with trimerization [6, 54, 55, 56]. In particular, binding of the MA domain to  $PI(4,5)P_2$  helps localize Gag proteins on virion assembly sites at the plasma membrane, which promotes Gag trimerization and triggers myristate

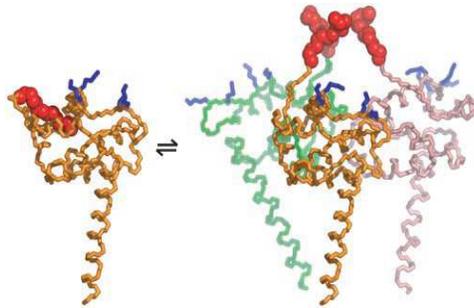


FIGURE 3.3. Molecular Model for  $\text{myr}(e)$ -MA Trimer. *This model was proposed by Tang et al. for  $\text{myr}(s)$ -MA monomer and  $\text{myr}(e)$ -MA trimer at equilibrium, based on observations made in experiments that examined myristoylated MA proteins. Binding of the  $\text{myr}(e)$ -MA trimer to the plasma membrane may disrupt myr-myrist interactions. The myristate is shown here in red. Source: [56].*

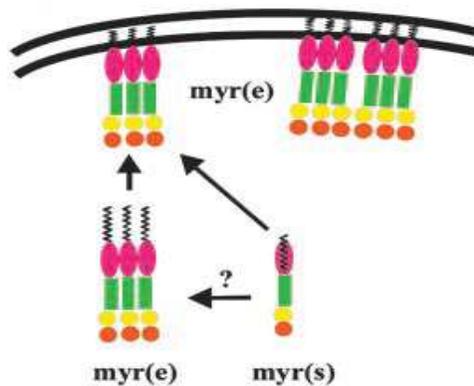


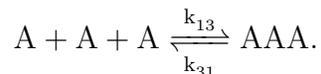
FIGURE 3.4. Trimerization and Membrane Binding of Gag.  *$\text{myr}(e)$  Gag trimers either bind or are bound to the plasma membrane, and they form higher-order multimers as the new HIV virion assembles. Source: [6].*

exposure [30, 54] (see Figure 3.4). According to Jouvenet and coworkers, “Gag concentration influences both the interaction between Gag molecules and the interaction between Gag and the plasma membrane and is very probably a key determinant of assembly kinetics” [9]. Thus, the increase in Gag concentration at the plasma membrane enhances multimerization and myristate exposure, which in turn ensure tight anchoring of Gag to the plasma membrane. In addition, myristate exposure is also enhanced by the CA and NC subdomains and the binding of Gag to RNA [6, 56], and is modulated by pH [55]. Interestingly, the myristate group remains sequestered in HIV-2 MA in the presence of  $PI(4,5)P_2$  [54, 55]. Thus, the assembly mechanisms for two closely related viruses may be quite different.

Based on the above findings from previous studies, we develop and analyze a mathematical model for HIV-1 Gag trimerization near the plasma membrane. The model considers two main species: Gag monomers, corresponding to the myr(s) state, and Gag trimers, corresponding to the myr(e) state. We aim to better understand the relationships among the variables and parameters involved in the dynamics of Gag trimerization and multimerization at the site of virion assembly. We also analyze the model to show the existence of a stable equilibrium between the monomeric myr(s) and trimeric myr(e) states of Gag protein.

### 3.2. KINETIC AND EQUILIBRIUM CONSTANTS

Table 3.1 defines the general kinetic association and dissociation constants in a chemical reaction. The chemical reaction can be represented with the equation



In the context of our model of Gag trimerization,  $k_a = k_{13}$  and  $k_d = k_{31}$  correspond to the rates of Gag trimerization and monomerization, respectively. The reaction towards the right corresponds to trimerization, with  $k_{13}$  measured in  $1\text{mol}^{-2}\text{s}^{-1}$ , and the reaction towards the left corresponds to monomerization, with  $k_{31}$  measured in  $1\text{s}^{-1}$ .

TABLE 3.1. Relevant Kinetic Constants. **Source:** [57].

	Association rate constant, $k_a$	Dissociation rate constant, $k_d$
Definition	$A + B \rightarrow AB$	$AB \rightarrow A + B$
Description	Reaction rate of $AB$ formation: number of $AB$ complexes formed per unit time at unit concentration of $A$ and $B$	Dissociation rate of $AB$ : number of $AB$ complexes dissociating per unit time
Units	$1\text{mol}^{-1}\text{s}^{-1}$	$1\text{s}^{-1}$

Table 3.2 defines the equilibrium association constant  $K_a$ . In the context of our mathematical model,  $K_a$  is measured in  $1\text{mol}^{-2}$  and is defined as [40]

$$(21) \quad K_a = \frac{\text{concentration of Gag trimers}}{(\text{concentration of Gag monomers})^3} = \frac{[AAA]}{[A]^3}.$$

TABLE 3.2. Relevant Equilibrium Constant. **Source:** [57].

	Equilibrium association constant $K_a$
Definition	$\frac{[AB]}{[A][B]} = \frac{k_a}{k_d}$
Description	Affinity to association: high $K_a$ means high affinity to associate
Units	$1\text{mol}^{-1}$

### 3.3. MATHEMATICAL MODEL FOR TWO-SPECIES GAG TRIMERIZATION

3.3.1. ASSUMPTIONS, VARIABLES, AND PARAMETERS. Listed below are the main assumptions for our model of HIV-1 Gag trimerization near the plasma membrane:

- Gag molecules arrive in monomeric form at the plasma membrane at a constant rate.
- Trimerization is enhanced by increased Gag monomer concentration.
- Multimerization is enhanced by increased Gag trimer concentration.
- Molecules do not leave the area near the plasma membrane after they arrive.
- Once trimers attach and multimerize at the plasma membrane, they do not break down into smaller molecules.

The variables and parameters utilized in our model are described below:

- $P_1$  = concentration of Gag monomers
- $P_3$  = concentration of Gag trimers
- $g$  = arrival rate of new Gag monomers to the assembly site at the plasma membrane
- $k_{13}, k_{31}$  = Gag association (trimerization) and dissociation (monomerization) rates, respectively
- $\gamma$  = multimerization rate of trimers at the plasma membrane

3.3.2. TWO-SPECIES MODEL: EQUATIONS. We assume that Gag monomers arrive at the site of Gag multimerization and virion assembly (“Gag hotspot”) at a constant rate. Three Gag monomers can associate and give rise to a Gag trimer. A trimer can also dissociate into three monomers. Gag trimers that multimerize and are tightly bound to the plasma membrane are “harvested out” and do not break down into smaller molecules.

The resulting model is:

$$(22) \quad \begin{cases} \frac{dP_1}{dt} = g - 3k_{13}(P_1)^3 + 3k_{31}P_3, \\ \frac{dP_3}{dt} = k_{13}(P_1)^3 - k_{31}P_3 - \gamma P_3. \end{cases}$$

3.3.3. MATLAB IMPLEMENTATION OF THE TWO-SPECIES MODEL. We describe here our Matlab code, consisting of three files:

- `setMonTriParams.m`: This function sets values for the model parameters  $g$ ,  $k_{13}$ ,  $k_{31}$ , and  $\gamma$ . We later calculate estimates for  $\frac{k_{13}}{k_{31}}$  and  $\frac{g}{\gamma}$ , which can be used to set the parameters values.
- `MonTri.m`: This function implements the model equations for the rates of change of the concentrations of Gag monomers and trimers. `MonTri` first calls the function `setMonTriParams` to get model parameters values, and then it uses these values to implement the equations of the model. `MonTri` returns a vector with the differential equations.
- `simuMonTri.m`: This script solves the differential equations numerically. It first sets the initial and final times and initial values. Next, Matlab built-in solver `ode45` is called to solve the differential equation model implemented with the `MonTri` function. Then, the script produces a plot of the concentrations of Gag monomers and trimers vs time.

3.3.4. TWO-SPECIES MODEL: EQUILIBRIUM AND ANALYSIS. The equilibrium expressions for the concentrations of myr(s) monomers and myr(e) trimers are:

$$(23) \quad \begin{cases} P_1^* = \left[ \frac{g}{3k_{13}} \left( \frac{k_{31}}{\gamma} + 1 \right) \right]^{1/3}, \\ P_3^* = \frac{g}{3\gamma}. \end{cases}$$

Notice that the equilibrium concentration for myr(e)-Gag trimers does not depend on the association and dissociation parameters  $k_{13}$  and  $k_{31}$ , but it only depends on the Gag monomer arrival and Gag trimer harvesting parameters. In particular, an increase in the monomer arrival rate or a decrease in the trimer harvesting rate will increase the equilibrium concentration of myr(e)-Gag trimers. Additionally, the equilibrium concentrations are positive, assuming that all parameter values and either  $P_1(0)$  or  $P_3(0)$  are positive. If  $P_1^* = 0$  or  $P_3^* = 0$ , then  $g = 0$ , which implies that  $P_1^* = 0 = P_3^*$ .

The Jacobian matrix for system (22) is:

$$J = \begin{bmatrix} -9k_{13}P_1^2 & 3k_{31} \\ 3k_{13}P_1^2 & -k_{31} - \gamma \end{bmatrix}.$$

The eigenvalues of matrix  $J$  are:

$$\lambda_{\pm} = \frac{-(9k_{13}P_1^2 + k_{31} + \gamma) \pm \sqrt{(9k_{13}P_1^2 + k_{31} + \gamma)^2 - 36k_{13}\gamma P_1^2}}{2}.$$

Let  $B = 9k_{13}P_1^2 + k_{31} + \gamma$ . Then

$$\lambda_{\pm} = \frac{-B \pm \sqrt{B^2 - 36k_{13}\gamma P_1^2}}{2}.$$

If  $B^2 - 36k_{13}\gamma P_1^2 \leq 0$ , then  $Re(\lambda_{\pm}) < 0$ . Otherwise,  $\lambda_- < 0$  and

$$\begin{aligned} & -36k_{13}\gamma P_1^2 < 0 \\ \Rightarrow & B^2 - 36k_{13}\gamma P_1^2 < B^2 \\ \Rightarrow & \sqrt{(B^2 - 36k_{13}\gamma P_1^2)} < B \\ \Rightarrow & -B + \sqrt{(B^2 - 36k_{13}\gamma P_1^2)} < 0. \end{aligned}$$

Hence,  $\lambda_+ < 0$ .

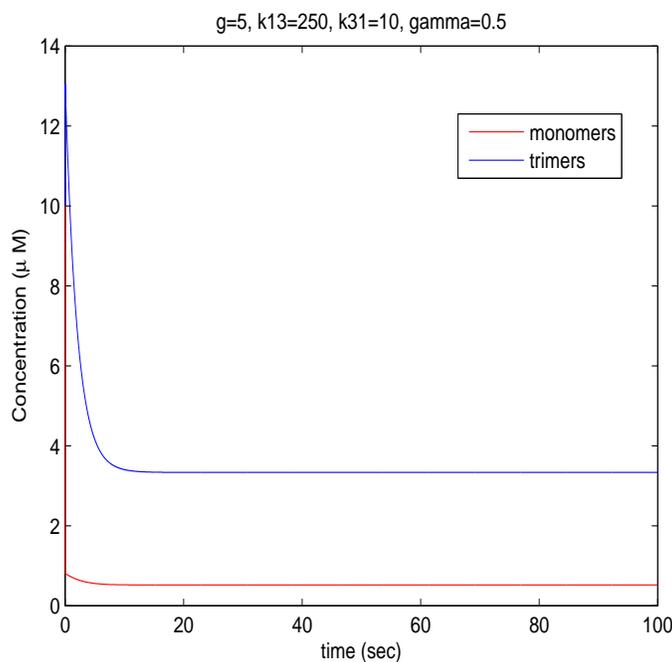


FIGURE 3.5. Gag Trimerization Model Simulations. *Matlab simulations confirm our theoretical results. There is a stable equilibrium.*

Therefore, the equilibrium is (locally) asymptotically stable, regardless of parameter values. Numerical simulations, utilizing Matlab's `ode45` subroutine and `pplane8` [58], confirm our result (see Figures 3.5 and 3.6). This result agrees with the literature, since it is known that myr-Gag monomers and trimers approach an equilibrium state [6, 56]. Notice that the two nullclines are the same when  $g = 0 = \gamma$  (see Figure 3.6). In that case, when there is no

monomer arrival to the Gag hotspots and no higher-order multimerization, all points on the curve  $P_3 = \frac{k_{13}(P_1)^3}{k_{31}}$  are equilibrium points.

3.3.5. ESTIMATES FOR  $\frac{k_{13}}{k_{31}}$  AND  $\frac{g}{\gamma}$ . For this trimerization process, the ranges of biological quantities are hard to measure, so in this subsection we provide estimates for  $\frac{k_{13}}{k_{31}}$  and

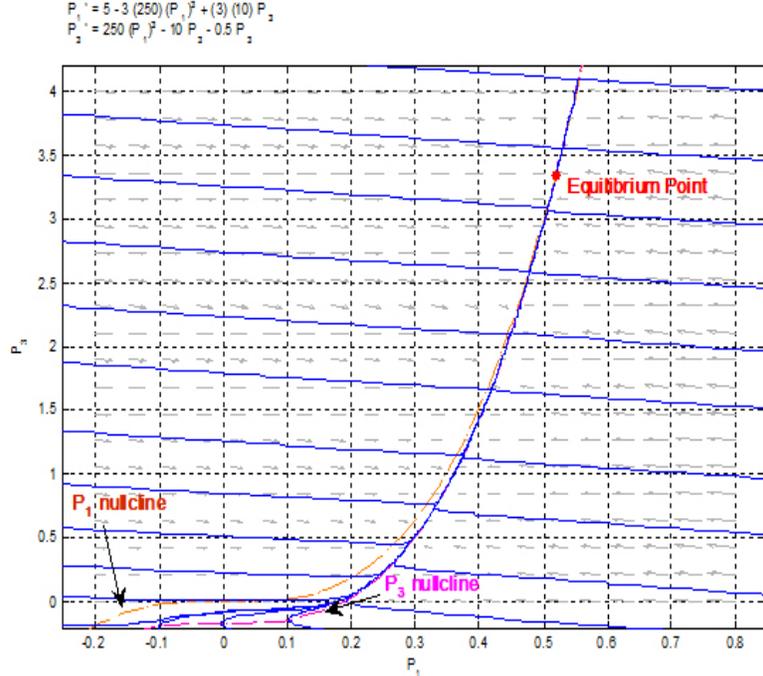


FIGURE 3.6. Gag Trimerization Phase Portrait. The equilibrium point is located at about  $(0.52, 3.3)$ . The  $P_1$  nullcline is  $P_3 = \frac{k_{13}(P_1)^3 - g/3}{k_{31}}$ . The  $P_3$  nullcline is  $P_3 = \frac{k_{13}(P_1)^3}{k_{31} + \gamma}$ .

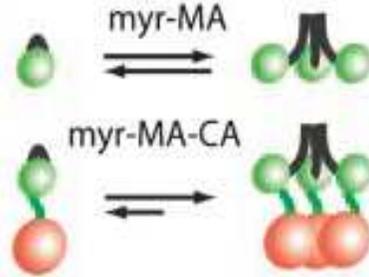


FIGURE 3.7. myr-MA and myr-MA-CA Equilibria.  $K_a = (2.5 \pm 0.6) \times 10^8 M^{-2}$  for myr-MA, and  $K_a = (4.6 \pm 1.1) \times 10^9 M^{-2}$  for myr-MA-CA at  $20^\circ C$ . Source: [56].

$\frac{g}{\gamma}$ . From equation (21) and equations (23), it is known that

$$(24) \quad K_a = \frac{P_3^*}{P_1^{*3}} = \frac{k_{13}}{k_{31} + \gamma}.$$

Then,  $K_a = \frac{k_{13}}{k_{31}}$  is a good approximation for small values of  $\gamma$ , that is, when there is little multimerization.

Tang and co-workers [56] estimated the equilibrium association constants for myr-MA and myr-MA-CA at 20°C (see Figure 3.7). In particular, the estimated equilibrium association constant for myr-MA-CA is

$$K_a = (4.6 \pm 1.1) \times 10^9 M^{-2},$$

which is about 20 times larger than the  $K_a$  value for myr-MA. We expect the corresponding value of the equilibrium association constant  $K_a$  for myr-Gag to be larger than that for myr-MA-CA, since the inclusion of the NC domain and association to RNA promote Gag-Gag interactions and, thus, trimerization [6, 56].

From experimental data, it is known that at the site of virion assembly near the cell membrane, the total Gag protein concentration at equilibrium

$$(25) \quad c = P_1^* + 3P_3^*$$

is close to 10  $\mu\text{M}$ , but could be as large as 1,000  $\mu\text{M}$  [40], taking into account that the threshold concentration of Gag for the formation of a new HIV virion with 1,500 Gag proteins is about 1,409  $\mu\text{M}$ . Based on the estimate from Tang et al. [56], we let  $K_a = 4.6 \times 10^9 M^{-2}$  and use Mathematica to solve equations (24) and (25). We calculate that the concentration

of Gag trimers  $P_3^*$  is somewhere between 1.20218  $\mu\text{M}$  and 319.63  $\mu\text{M}$ . Since  $P_3^* = \frac{g}{3\gamma}$ , we conclude that  $\frac{g}{\gamma}$  is between 3.60653  $\mu\text{M}$  and 958.89  $\mu\text{M}$ .

3.3.6. CONDITION FOR EQUILIBRIUM SHIFT TO GAG TRIMERS. Here we determine conditions on  $K_a$  and the model parameters that will ensure an equilibrium shift towards the trimeric myr(e)-Gag state. When 50% of the Gag protein concentration at equilibrium is monomeric, we have  $P_1^* = 3P_3^*$ . Hence, by equation (21),

$$K_a = \frac{P_3^*}{(3P_3^*)^3} = \frac{1}{27(P_3^*)^2}.$$

Thus, the equilibrium is shifted towards the trimeric state when

$$(26) \quad K_a > \frac{1}{27(P_3^*)^2}.$$

If we use equations (23) and (24) to write condition (26) in terms of the model parameters, we obtain the equivalent condition

$$(27) \quad \frac{g}{\gamma} > \sqrt{\frac{k_{31} + \gamma}{3k_{13}}}.$$

Notice that for the example illustrated in Figure 3.5, with the concentration of trimers higher than the concentration of monomers at equilibrium, condition (27) is satisfied, since we have

$$\frac{g}{\gamma} = 10 > 0.12 \approx \sqrt{\frac{k_{31} + \gamma}{3k_{13}}}.$$

According to Tang et al. [56], at equilibrium, the myr-MA concentration is 50% monomeric when the total concentration is  $73 \mu\text{M}$ , and the myr-MA-CA concentration is 50% monomeric when the total concentration is  $17 \mu\text{M}$ <sup>1</sup> at  $20^\circ\text{C}$ . Now, suppose that the myr-Gag equilibrium concentration is 50% monomeric when the total concentration is  $10 \mu\text{M}$  [40]. Then, the concentration of trimers  $P_3^* \approx 1.67 \mu\text{M}$ , and condition (26) implies that

$$K_a > 1.33 \times 10^{10} M^{-2}.$$

The lower bound for  $K_a$  found here is about 3 times larger than the estimate for myr-MA-CA  $K_a$  by Tang et al. [56]. Comparing the  $K_a$  values given by Tang et al. for myr-CA and myr-CA-MA [56] (see Figure 3.7), this estimate seems reasonable for myr-Gag.

### 3.4. REMARKS: BIOLOGICAL IMPLICATIONS AND FURTHER WORK

Our results show that there is a stable equilibrium between the myr-Gag monomer and trimer species. Even though our model could potentially be enhanced, we can see that our assumptions here are sufficient for capturing this important feature of the dynamics of Gag trimerization at the plasma membrane, which has been observed in experiments on myr-MA and myr-MA-CA. Thus, our assumptions of the dynamics of Gag trimerization and multimerization at the plasma membrane, such as:

- Gag proteins arrive to the Gag hotspots as monomers at a constant rate;
- Gag multimers, which are formed from associations of trimers and attach tightly to the plasma membrane, do not break down into smaller structures;

---

<sup>1</sup>This is a recent correction on Tang et al., PNAS (2004) [59].

- the relationships between the variables and the monomerization, trimerization, and multimerization parameters, as given by the model equations

seem to be reasonable.

Following, we comment on possible further work avenues on Gag trimerization modeling.

- (1) Literature research and simulations may be performed to approximate the values of the parameters for the rate of Gag monomer arrival  $g$  and/or Gag multimerization  $\gamma$ . Only one of the two is necessary, if the approximation for the ratio of these parameters calculated here is used. Similarly, more research may allow one to obtain approximations for the association and dissociation rates  $k_{13}$  and  $k_{31}$ . These parameter estimates can also help to determine a more accurate estimate for the value of myr-Gag  $K_a$ , the equilibrium association constant.
- (2) There are two timescales in the Gag trimerization model that we proposed and analyzed here (see 22). In the first (fast) phase, near time  $t = 0$ , the concentrations of monomers and trimers change very rapidly. In the second (slow) phase, the concentrations of monomers and trimers approach equilibrium. Multiple timescale analysis via perturbation methods may be performed on the trimerization model in order to further investigate the behavior of the solution near  $t = 0$ .
- (3) The mathematical model for Gag trimerization proposed here could be further improved to better reflect the dynamics of Gag localization and HIV virion assembly at the plasma membrane. For instance, instead of assuming that the arrival rate of Gag monomers  $g$  is constant, one could consider  $g = g(c)$ . That is,  $g$  might depend on the total concentration  $c$  of Gag protein at the plasma membrane. As a Gag hotspot becomes saturated, fewer Gag proteins may be transported to the area. A

more accurate model might include a carrying capacity term for the arrival rate of Gag monomers at the plasma membrane, thus, requiring  $g$  to depend on the value of the concentration  $c$ .

Moreover, we should point out that one might consider a three-species model for Gag trimerization. Such a model could include the concentrations of Gag monomers in myr(s) state, Gag monomers in myr(e) state, and Gag trimers. On the other hand, according to Tang and co-workers, myristate exposure and trimerization are coupled together [6, 56], which then eliminates the need for the intermediate species of Gag monomers in myr(e) state.

## CHAPTER 4

# HIV CORE ASSEMBLY

### 4.1. INTRODUCTION

When enough Gag proteins have accumulated near the cell membrane, a new immature HIV virion forms and buds off the host cell. In order for this new virion to become infective, the virion needs to go through the maturation process. During maturation, an enzyme called protease (PR) cleaves the Gag proteins inside the virion into their main domains: matrix (MA), capsid (CA), and nucleo-capsid (NC). Approximately 1,200-5,000 MA molecules remain attached to the inner layer surrounding the virion [9, 10, 19], while roughly 1,200-1,500 CA molecules assemble to form the outer shell or *capsid* of the mature HIV virion's core [10, 18, 19] (see Table 4.1). The interior of the core contains two single strands of RNA associated to thousands of NC proteins, as well as a few hundred copies of reverse transcriptase and integrase enzymes [18, 19].

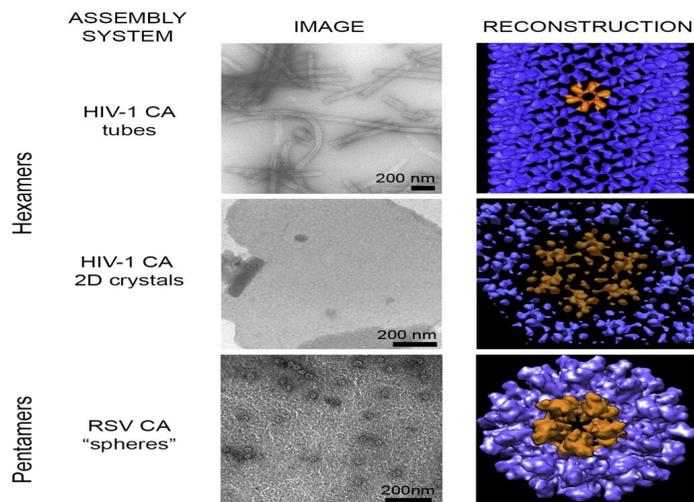


FIGURE 4.1. *In Vitro* and Reconstructed Images of HIV-1 CA Helical Tubes, 2D Hexagonal Crystals, and RSV Icosahedral Particles. **Source:** [60].

We investigate mathematical models for the assembly of the CA proteins into the mature virion's capsid. In particular, the model, proposed by Zlonick's group and studied also by Hagan's group (see, for instance, [22, 23]), consists of a system of ordinary differential equations for the rates of change of the concentrations of CA subunit, intermediates, and capsid. We apply this model to HIV by looking at two stages of the assembly process: nucleation and elongation. Since a mature HIV capsid contains only 12 CA pentamers,

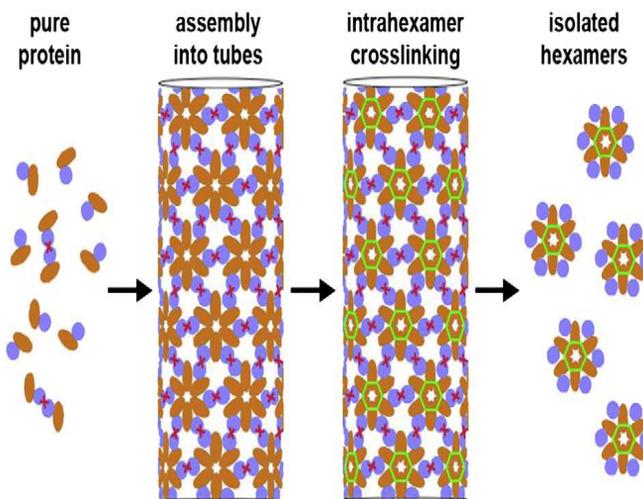


FIGURE 4.2. Disulfide Crosslinking Strategy Used to Obtain Soluble HIV-1 CA Hexamers for 3D Crystalization. **Source:** [60].

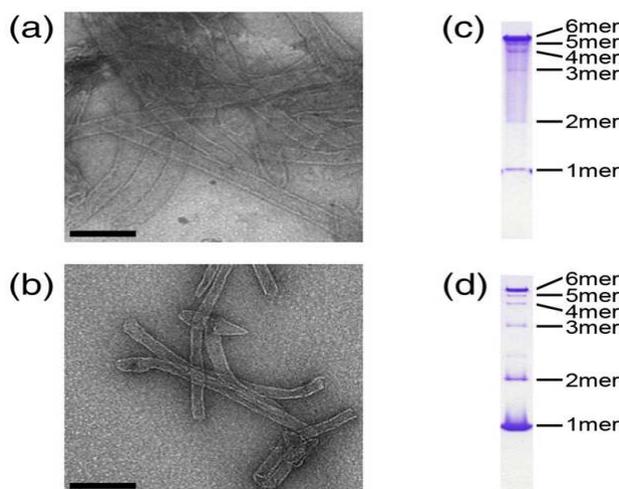


FIGURE 4.3. (a)(b) EM images of CA tubes for two different constructs; (c)(d) profiles of cross-linked soluble hexamers, and expected positions of cross-linked  $n$ -mers. The scale band represents 500nm. **Source:** [60].

we assume that a nucleus is a hexagon of CA molecules (see Figures 1.6, 4.1, 4.2, 4.3, and 4.4). Hence, nucleation for a single mature HIV capsid ends when the hexagon of CA proteins is formed. Also, the forward rates for the reaction are smaller compared to elongation. We perform numerical simulations of the model by using the Matlab subroutine `ode45` and plotting the concentration profiles for several intermediates, subunit, and capsid. Our simulation results imply that the concentrations of all species approach an equilibrium. We verify our results analytically. In particular, we show the existence of a unique and stable equilibrium for the case  $N = 3$ .

#### 4.2. EXISTING WORK

Compared to mathematical models for Gag trafficking and multimerization, there is more work on modeling HIV core assembly or broadly viral assembly [61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77].

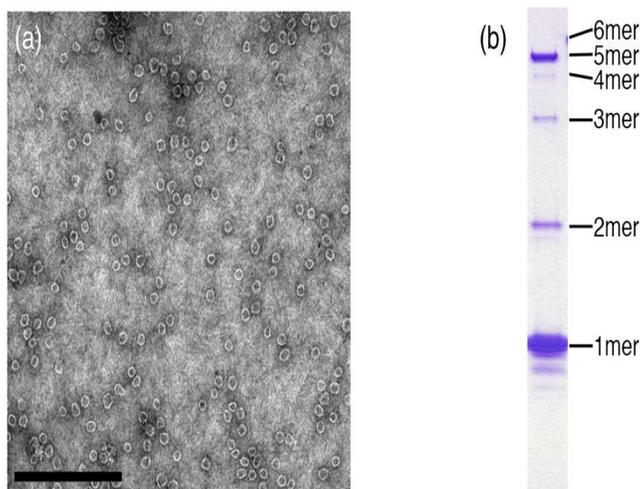


FIGURE 4.4. (a) Spherical particles detected via EM, indicating the insertion of pentamers. (b) Profile of cross-linked soluble pentamers with expected positions of cross-linked  $n$ -mers. **Source:** [60].

4.2.1. GANSER'S GROUP. Ganser and co-workers studied the structure of the core shell for HIV virus-like particles (VLPs) via electron cryotomography [19]. They observed a consistent tendency of the CA proteins to arrange in fullerene-like conical shapes. Cylindrical and irregularly-shaped capsids were also observed (see Figure 4.5).

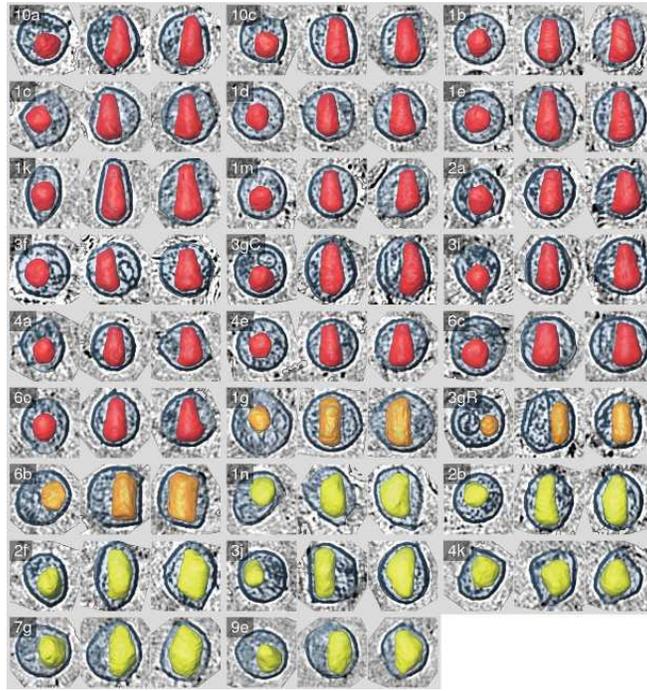


FIGURE 4.5. HIV-1 Virus-Like Particles (VLPs). *Conical (red), cylindrical (orange), and irregularly-shaped (yellow) capsids were observed via electron cryotomography of a cell culture. Source: [19].*

4.2.2. ZLOTNICK'S GROUP. Reference [23] is one of Zlotnick's earliest works on virus capsid assembly. A system of differential equations for the rates of change of the concentrations of capsid subunits and intermediates is developed. These equations model capsid assembly as a polymerization reaction. It is assumed that only one subunit (which could be an oligomer) at a time associates to or dissociates from the growing structure. Association of intermediates is not considered, as their concentration, in general, is low. For the most part, in their simulations, only one forward rate constant  $k_f$  is used, and the backward rate

TABLE 4.1. HIV-1 VLPs Statistics. **Source:** [19].

VLP	Calculated numbers of molecules	
	MA	CA
10a*	4700	1600
10c	3600	1100
1b	3500	1000
1c	4000	1000
1d	3700	1000
1e	4200	1200
1g*	5100	3600
1k	4300	1300
1m	3400	900
2a	3900	1100
3f	4300	1000
3g*	4200	2000
3i	3900	1000
4a	2800	900
4e	3400	900
6c	4500	1200
6e	3400	1000
6b	4900	1400
1n	4400	1200
2b	3700	1200
2f	4000	1300
3j	4000	1200
4k	3600	1000
7g	4200	1400
9e	3300	1100
Mean	4000	1300
Stddev	530	540

Asterisks (\*) mark VLPs with multiple capsids.

constants  $k_b n$  are calculated from the forward rate constant  $k_f$  and an association constant  $K_a n$  with the equation:  $k_b n = k_f / K_a n$ . The equations used are

$$(28) \quad \frac{d[\text{polymer}_n]}{dt} = k_f([\text{polymer}_{n-1}] - [\text{polymer}_n])[\text{freesubunit}] - k_b[\text{polymer}_n].$$

In simulations, these equations were used to model the formation of a dodecahedron, where each subunit is a pentagon. Numerical solutions of these equations were found using a fourth-order Runge-Kutta method with STELLA. Two types of assembly models were considered:

*equilibrium assembly (EQ)*, where intermediates form and break apart in a way that there are enough subunits available for the reaction to continue, and *kinetically limited (KL)*, in which the early reaction is slow in order to avoid kinetic trapping. Both models feature sigmoidal growth of capsid, with an initial lag phase followed by rapid growth. The EQ model is more susceptible to kinetic traps but does not require nucleation, as opposed to the KL model. The EQ model is more appropriate for assembly processes where the association energies are weak. The rate of formation of a nucleus can be approximated by the equation

$$(29) \quad \frac{d[\text{nucleus}]}{dt} \approx k_{fn}[\text{subunit}]^n$$

where  $n$  is the number of subunits in the nucleus and  $k_{fn}$  represents the net rate for nucleus assembly.

In [78], rate equations are used to model assembly of dodecahedral (formed by twelve pentamers) and icosahedral capsids (formed by 30 tetramers). Expressions are provided for association constants for subunits and capsids as well as the equilibrium concentrations of all species in terms of subunit. A typical equation from the system of rate equations A2 in [78] has the form

$$(30) \quad \frac{d[m]}{dt} = f_m s_m [u][m-1] - f_{m+1} s_{m+1} [u][m] + b_{m+1}[m+1] - b_m[m]$$

where  $s_m$  is a degeneracy statistical factor, and  $f_m$  and  $b_m$  are forward and backward rates, respectively, for the  $m$ th species. The backward rates are calculated from the forward rate  $f_m$ .

In [20], Katen and Zlotnick reiterate that virus capsid subunits interact through weak contact energies, which leads to a globally stable, and yet dynamic, structure. Assembly

is modeled, as before, as polymerization. First, there is a nucleation event. Then the structure elongates “through a series of faster and/or more stable additions of subunit, until an equilibrium between the polymer and the free subunits is reached” [20]. These steps correspond to the lag phase, which ends when the nucleus is formed, the rapid growth phase, and the saturation phase. On the other hand, *in vitro*, many capsids are assembled simultaneously. The lag phase, in this case, ends with the formation of an “assembly line” of intermediates and the first capsids begin to accumulate. During the elongation phase, concurrent nucleation and elongation events happen, giving rise to more capsids. The system used to describe this reaction is similar to the equations (30):

$$\begin{aligned}
 \frac{d[nuc + n]}{dt} = & k_{elong,n-1}[nuc + n - 1][subunit] + \\
 (31) \quad & k_{dissoc,n+1}[nuc + n + 1][subunit] \\
 & - k_{elong,n}[nuc + n][subunit] + (other\ terms).
 \end{aligned}$$

Katen and Zlotnick also discuss conditions for kinetic trapping: 1) if the association energies (and, hence, forward rates) are too high, which causes the formation of intermediates too rapidly, using up all of the subunits without forming complete capsids; 2) if nucleation happens too quickly, compared to elongation, which produces many metastable intermediates; 3) if there is “off-path assembly”, giving rise to metastable intermediates, as before.

4.2.3. HAGAN’S GROUP. Hagan and Elrad consider in [22] a model of rate equations developed by Zlotnick’s group (see eqn. (33) below). The model disregards malformed capsids. It also assumes that only individual subunits associate and dissociate, and that there is only one forward and one backward coefficient for each intermediate. Two types of

simulations are considered: canonical (NVT) simulations, corresponding to *in vitro* empty-capsid experiments; and steady-state ensemble simulations, where the concentration of free subunits is independent of time, in order to measure the mean elongation times of growing capsids. They also consider the nucleation-growth (NG) model and a classical nucleation (CNT) model. As opposed to Zlotnick and co-workers, who use two forward rates,  $f_{nuc}$  and  $f_{elong}$ , for the NG model, Hagan and Elrad distinguish between nucleation and elongation by using different intersubunit association free energies. Sigmoidal growth of capsids, as previously described, was observed. Their results show good agreement with experimental median assembly times. The main conclusions of this study are the following: 1) Lag times correspond to the initial mean elongation times. 2) Mean elongation times vary inversely with free subunit concentration. 3) If there is a critical nucleus size, it can be identified from the concentration dependence of the median assembly time.

In [79], Kivenson and Hagan study capsid assembly around a polymer, such as RNA. They model capsid as a cubic lattice and use Monte Carlo (MC) simulations. They observe that, under optimal conditions, capsid assembly and incorporation of the polymer proceed simultaneously and, in fact, the polymer helps stabilize the addition of subunits to the growing capsids. The process is similar to capsid assembly described before. In general, there is a nucleation stage, where a small partial capsid is formed attached to a polymer. Then, the capsid grows by adding subunits to the partial capsid. A final phase follows, where the capsid is completed. This last stage can be delayed until the polymer is completely encased inside the capsid. In addition, there can be polymer-related kinetic traps, for instance, if the polymer is too long and does not fit inside the capsid. Thus, successful capsid assembly around a polymer is dependent on the polymer length. In particular, there is an optimal

polymer length for maximum efficiency of capsid assembly. Another important observation in Kivenson and Hagan's work [79] is that the rates of nucleation and growth increase with polymer length and with polymer-subunit interaction strength.

### 4.3. DYNAMICAL SYSTEM MODEL

Our approach is to model HIV core assembly using dynamical systems. Capsid assembly is a complicated process and its mechanism has only begun to unravel. Several computational and mathematical models for capsid assembly of viruses have been developed and analyzed in recent years. The assembly model that we consider here, based on Zlotnick's group's model [23], is a differential dynamical system, similar to a population model for interacting species. Each species in this case represents a particular capsid intermediate or n-mer. Each equation in the model describes the rate of change of the concentration of a particular n-mer with respect to time. Several simplifications of the assembly process are used in building this model.

**Assumptions.** Listed below are important assumptions for our models:

- Malformed capsids are not considered. All capsids have the same number of CA subunits, and there are no lattices larger than a capsid.
- Only single subunits bind or unbind. Thus, partial capsids do not bind with each other.
- There is only one association and dissociation coefficient for each intermediate size. Both coefficients are, then, average rates of association and dissociation for a particular n-mer.

The above assumptions are based on those described in [22]. We adopt the model to describe the dynamics of HIV capsid assembly by choosing parameter values appropriately.

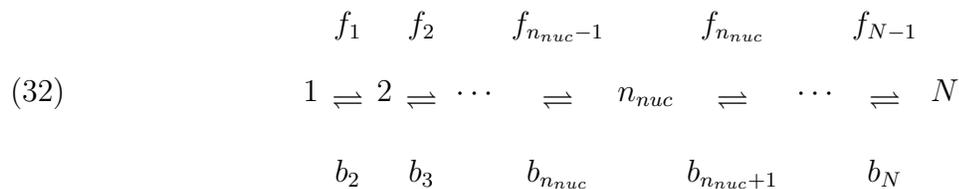
**Variables and Parameters.** Shown below is a list of the variables and parameters

in the dynamical system model and a brief description of each one:

- $c_n$  = concentration of the  $n$ th capsid intermediate (n-mer),
- $f_n$  = forward (association) rate,
- $b_n$  = backward (dissociation) rate,
- $N$  = number of subunits that form the capsid.

In the case of the mature HIV capsid (core shell),  $N \approx 1200 - 1500$  [10, 18, 19]. This implies that there are about 1200-1500 n-mers in the pathway between subunit and capsid. Hence, we have between 1200 and 1500 species in the model. Each species has its own rate equation, with its own forward and backward rates. Assuming that the capsid is the most stable of the species (based on the work by Zlotnick, Hagan, and others as described previously), we assume that the values of the forward coefficients  $f_n$  tend to increase with  $n$ . Similarly, we assume that the backward coefficients  $b_n$  tend to decrease with  $n$ .

**Intermediates: Association and Dissociation.** The diagram below describes the assembly reaction in two phases: nucleation and elongation.



Based on Equation (1) in [22]

The path to a capsid starts with one CA subunit (a monomer). One CA protein at a time is added to the growing capsid. The process is slow at first, until the first nucleus with  $n_{nuc}$  subunits, which we assume is a hexagon ( $n_{nuc} = 6$ ) in the case of HIV, forms. The process then continues at a faster pace as more and more CA units are added to the growing lattice until the capsid, composed of  $N$  CA proteins, is complete.

**Full Equations.** The assembly model that we consider is a dynamical system of first-order, autonomous, nonlinear ordinary differential equations. We emphasize the fact that the model is based on concentrations of subunits, intermediates, and capsids. Each equation describes the rate of change of the concentration of a particular n-mer with respect to time. The size of the system, i.e., the number of equations and variables, is determined by the number of subunits in a capsid,  $N$ . The equations for the capsid assembly model are (see [20, 22, 23])

$$\begin{aligned}
 \frac{dc_1}{dt} &= -2f_1c_1^2 + b_2c_2 + \sum_{n=2}^N(-f_nc_nc_1 + b_nc_n) \\
 \frac{dc_n}{dt} &= f_{n-1}c_1c_{n-1} - f_nc_1c_n - b_nc_n + b_{n+1}c_{n+1} \\
 (33) \qquad & n = 2, \dots, N \\
 & f_N = b_{N+1} = 0.
 \end{aligned}$$

The equation  $f_N = 0$  represents the fact that the last n-mer, the capsid, does not associate with any subunits. The largest n-mer possible is the capsid. For this reason, we also have  $b_{N+1} = 0$ . Since there is no  $(N + 1)st$  n-mer, there is no dissociation rate for such n-mer. Additionally, notice that the equations do not include a dissociation rate  $b_1$ , as it is impossible for subunits to dissociate.

#### 4.4. MATLAB IMPLEMENTATION

Our Matlab code consists of three files, which we describe below:

- `setAssemParams.m`: This function sets the forward and backward parameter values  $f_n$  and  $b_n$ . The parameters are chosen so that  $f_n$  increases monotonically and  $b_n$  decreases monotonically with respect to  $n$ . The function receives the number of subunits in a capsid (capsid size) and returns the vectors of  $f_n$  and  $b_n$  values.
- `CapAssem.m`: This function implements the model equations (concentrations of n-mers). The function receives the number of subunits in a capsid and vector variables for time and concentrations. It first calls `setAssemParams` to get the forward and backward parameters values, and then it uses these values to implement the equations of the model. `CapAssem` returns a vector with the differential equations.
- `simuCapAssem.m`: This script solves the differential equations numerically. It first sets the initial and final times, the capsid size, and initial values. Next, it calls `ode45` to solve the differential equation model implemented with the `CapAssem` function. Then, the script produces a plot of concentration profiles for some of the capsid intermediates, as well as a plot of the subunit and the capsid concentrations versus time.

#### 4.5. *In Silico*

In this section, we describe our analysis of the model through numerical simulations performed in Matlab.

**Parameters.** As mentioned in the previous section, we chose arbitrary functions for the forward and backward coefficients  $f_n$  and  $b_n$ . More specifically,  $f_n$  increases and  $b_n$  decreases as  $n$  increases. This condition guarantees that the most stable of the species is the capsid, while all other species, aside from the subunit, will tend to exist in very low concentrations.

**Initial Values.** The initial concentrations of n-mers and capsids were set to 0. We ran the code for two different concentrations (numbers) of subunits (CA): 1300 and 5000.

*In silico* for  $c_1(0) = 1300$ : See Figures 4.6 and 4.7.

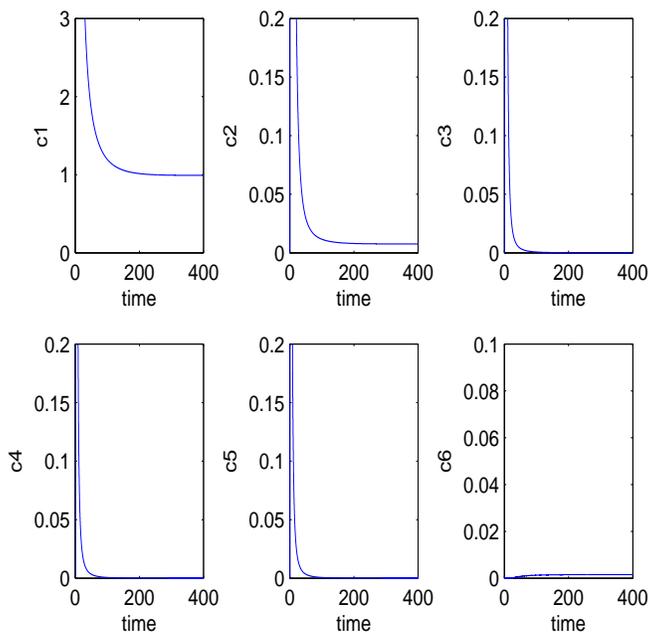


FIGURE 4.6. Concentration Profiles for Intermediates.

*In silico* for  $c_1(0) = 5000$ : See Figures 4.8 and 4.9.

**Observations.** The concentrations of subunit and intermediates decrease rapidly and reach equilibrium at values close to zero. Capsid concentration increases in sigmoidal fashion and reaches equilibrium at values that correspond closely to the total number of subunits divided by 250 (see Figures 4.6, 4.7, 4.8, and 4.9).

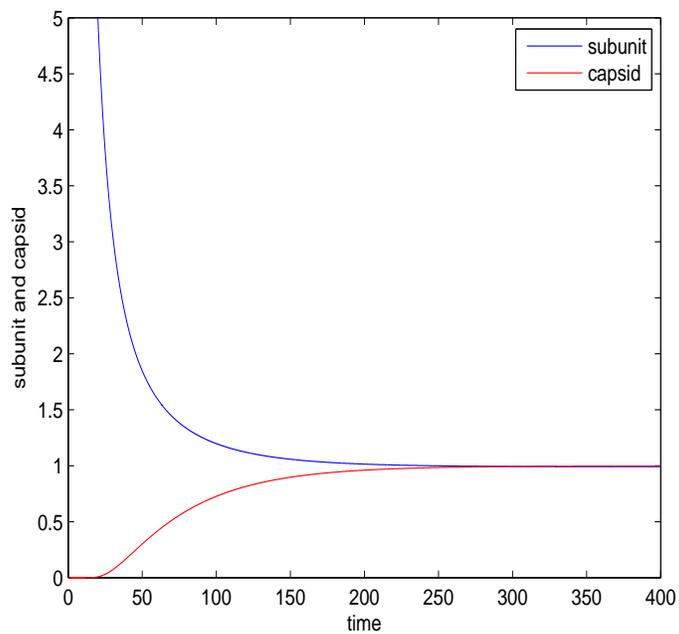


FIGURE 4.7. Concentration for Subunit and Capsid.

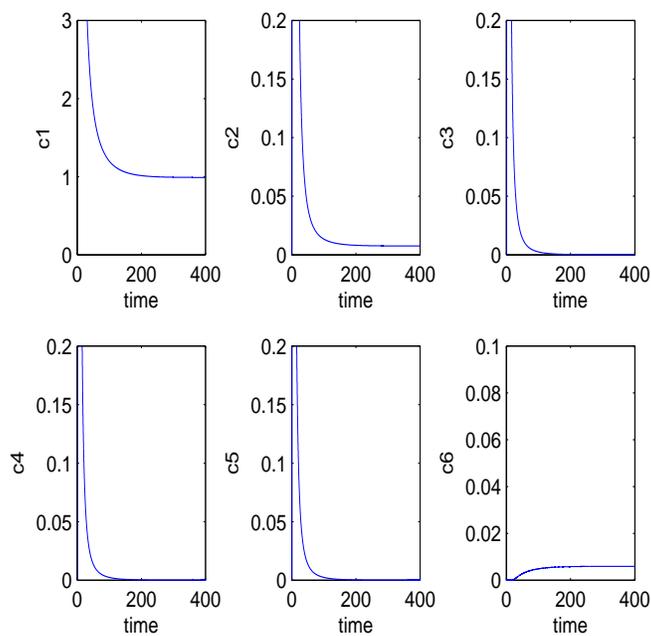


FIGURE 4.8. Concentration Profiles for Intermediates.

#### 4.6. INTERPRETING RESULTS

*In vivo* and *in vitro*, we expect the formation of a single core, although multiple cores may form [18, 19].

*In silico*, capsids are more stable than intermediates [16, 22].

The results agree with the literature, as it is known that the CA intermediates reach an equilibrium [22]. It is interesting to notice that, under our assumptions for the forward and backward parameters, the most important intermediates in the reaction seem to be the subunits and the full capsid. At equilibrium, all other intermediates exist at very low concentrations.

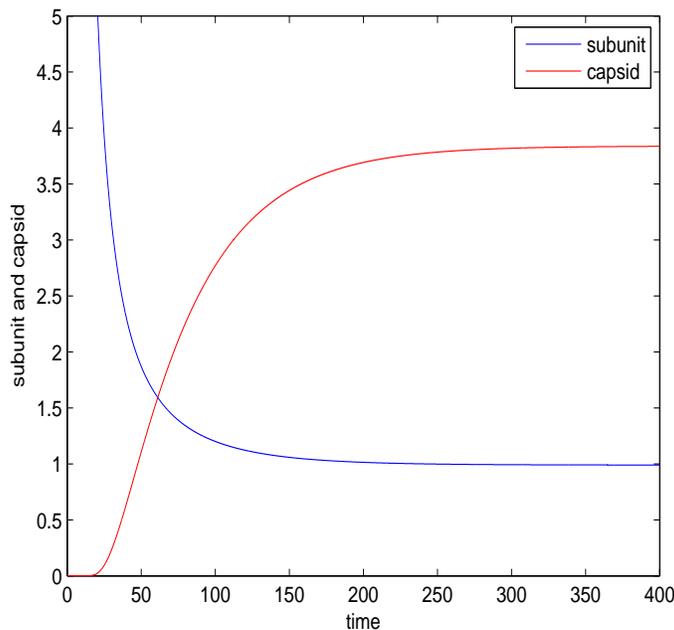


FIGURE 4.9. Concentration Profiles for Subunit and Capsid.

#### 4.7. EXISTENCE AND STABILITY OF EQUILIBRIUM

**Dynamical System for N=3.** For simplicity, we show that there exists a unique, stable equilibrium for  $N = 3$ . The equations for the system are

$$(34) \quad \begin{cases} \frac{dc_1}{dt} = -2f_1c_1^2 + 2b_2c_2 + b_3c_3 - f_2c_1c_2 \\ \frac{dc_2}{dt} = f_1c_1^2 - f_2c_1c_2 - b_2c_2 + b_3c_3 \\ \frac{dc_3}{dt} = f_2c_1c_2 - b_3c_3 \end{cases}$$

The Jacobian determinant is

$$(35) \quad \begin{vmatrix} -4f_1c_1 - f_2c_2 & 2b_2 - f_2c_1 & b_3 \\ 2c_1f_1 & -b_2 - f_2c_1 & b_3 \\ f_2c_2 & f_2c_1 & -b_3 \end{vmatrix} = 0.$$

This implies that our equations are not independent. We reduce the system by imposing the mass conservation condition that the total concentration of subunit is constant (see A4 in [78]):

$$c_1(t) + 2c_2(t) + 3c_3(t) = c_1(0) \text{ for all } t \Rightarrow c_3(t) = \frac{c_1(0) - c_1(t) - 2c_2(t)}{3}.$$

The reduced system is then:

$$(36) \quad \begin{cases} \frac{dc_1}{dt} = -2f_1c_1^2 + 2b_2c_2 + b_3 \frac{c_1(0) - c_1 - 2c_2}{3} - f_2c_2c_1 \\ \frac{dc_2}{dt} = f_1c_1^2 - f_2c_2c_1 - b_2c_2 + b_3 \frac{c_1(0) - c_1 - 2c_2}{3} \end{cases}$$

The Jacobian determinant for the reduced system is

$$(37) \quad \begin{vmatrix} \frac{-b_3}{3} - 4f_1c_1 - f_2c_2 & 2b_2 - 2\frac{b_3}{3} - f_2c_1 \\ \frac{-b_3}{3} - 2f_1c_1 - f_2c_2 & b_2 - 2\frac{b_3}{3} - f_2c_1 \end{vmatrix} \\ = b_2b_3 + 4b_3f_1c_1 + 3b_2f_2c_2 + 6f_2f_1c_1^2 \neq 0.$$

By the Implicit Function Theorem, we conclude that a unique equilibrium exists.

**Equilibrium for a System of Size N.** We can find an expression for the equilibrium of the full system (33) by setting the left hand side of the equations to zero and solving for all variables in terms of  $c_1^*$ , the concentration of subunits at equilibrium [78].

The equilibrium concentration for an intermediate of size  $n$  is

$$(38) \quad c_n^* = \frac{f_{n-1}f_{n-2} \cdots f_2f_1}{b_nb_{n-1} \cdots b_3b_2} c_1^{*n}, \quad n = 2, 3, \dots, N.$$

Assuming that all of the forward and backward coefficients are positive, this equilibrium is well defined. Moreover, equation (38) implies that, at equilibrium, all of the intermediates have positive concentration values. If  $c_m^* = 0$  for some  $m$ , then  $c_1^* = 0$ , which implies that  $c_n^* = 0$  for all  $n$ . We know that this is not possible, since we assume that the initial concentration of subunits is nonzero.

**Stability Analysis for N=3.** Let

$$B = b_2 + b_3 + 4f_1c_1 + f_2(c_1 + c_2),$$

$$C = 2f_1c_1(2b_3 + 3f_2c_1) + b_2(b_3 + 3f_2c_2),$$

$$D = B^2 - 4C.$$

The eigenvalues of the Jacobian matrix are

$$\lambda_{\pm} = \frac{1}{2}(-B \pm \sqrt{D}).$$

Assume that the equilibrium concentrations  $c_1$  and  $c_2$  are both nonnegative and that the forward and backward parameters are all positive. With these assumptions, we have  $B > 0$ , which implies that the equilibrium is asymptotically stable if  $D \leq 0$ . Suppose that  $D > 0$ . Then, the equilibrium is unstable if  $D > B^2$ . But  $D = B^2 - 4C < B^2$ , since  $C > 0$  by the assumptions on the equilibrium and parameters. Therefore, the equilibrium is asymptotically stable.

**Kinetic Trapping.** Kinetic trapping can happen, for instance, if association rates for nucleation are too large compared to the association rates for elongation [22]. In this case, nuclei would form too quickly, depleting the subunit pool and halting the reaction before capsids form.

#### 4.8. CASCADED DYNAMICAL SYSTEM MODEL

The capsid assembly process can be considered in two phases: nucleation and elongation. While these two stages are not completely separate from one another, we can reasonably assume that most of the significant nucleation occurs during the initial stage of the reaction, that is, the lag phase. After the lag phase, most of the formed nuclei elongate (grow) quickly to form larger intermediates and capsids. We can assume that this second phase, characterized by rapid growth and accumulation, is dominated by elongation events.

We consider two separate systems of equations, one for each phase of capsid assembly.

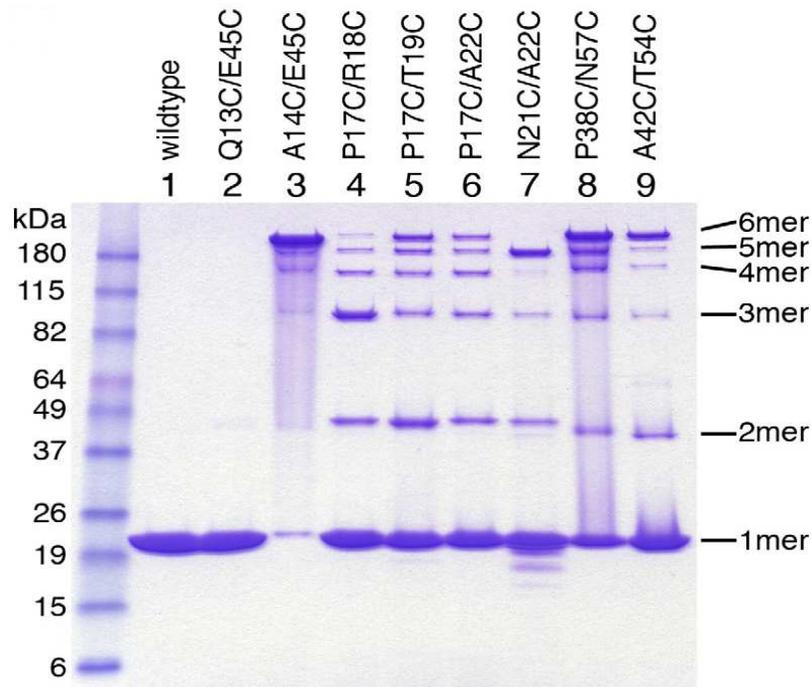


FIGURE 4.10. HIV-CA N-mer Profiles During Nucleation. *On most experiments depicted here, CA subunits and 6-mers were the most prevalent. Source: [60].*

**Nucleation Model.** The first system models nucleation. For HIV, we assume that a nucleus is comprised of six CA subunits, a hexagon or *hexamer*. As evidenced in [60] (see Figure 4.10), most intermediates during nucleation are present in the form of subunits and hexamers. Our model for this phase of capsid assembly consists of six differential equations, one for each of the first six species  $c_n$ , where  $n = 1, 2, 3, 4, 5, 6$ . As before,  $c_1$  corresponds to CA subunits, and  $c_6$  corresponds to hexamers (nuclei).

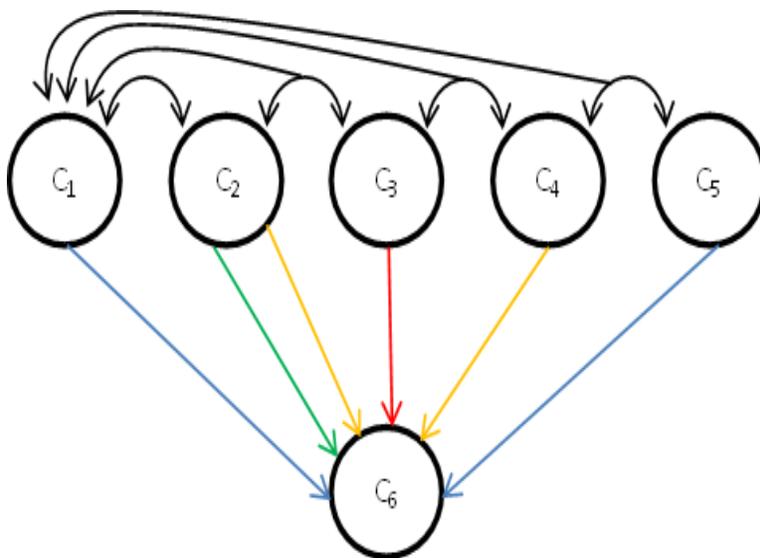


FIGURE 4.11. Assembly Paths for Nucleation Phase. *CA hexamers can form through various combinations of CA species: one monomer and one pentamer (blue), three dimers (green), a dimer and a tetramer (yellow), two trimers (red). Pentamers and smaller CA intermediates dissociate by the removal of one CA monomer at a time.*

Figure 4.11 describes the different pathways considered in the nucleation model. Two subunits come together to form a dimer, a subunit and a dimer form a trimer, and so forth. Also, a pentamer dissociates into a subunit and a tetramer, a tetramer dissociates into a subunit and a trimer, and so on. The largest intermediate in this case is the nucleus or hexamer. We consider four different pathways for the formation of hexamers. A hexamer

forms from the association of: a subunit and a pentamer (blue), three dimers (green), a dimer and a tetramer (yellow), and two trimers (red). We assume that hexamers are much more stable structures than the dimers, trimers, tetramers, and pentamers, and therefore we neglect their dissociation into smaller intermediates.

The equations for the nucleation phase are the following:

$$\begin{aligned}
\frac{dc_1}{dt} &= -2f_{11}c_1^2 - f_{12}c_1c_2 - f_{13}c_1c_3 - f_{14}c_1c_4 - f_{15}c_1c_5 + b_2c_2 + \sum_{i=2}^{i=5} b_i c_i \\
\frac{dc_2}{dt} &= f_{11}c_1^2 + b_3c_3 - 3f_{22}c_2^3 - f_{24}c_2c_4 - f_{12}c_1c_2 - b_2c_2 \\
\frac{dc_3}{dt} &= f_{12}c_1c_2 + b_4c_4 - f_{13}c_1c_3 - 2f_{33}c_3^2 - b_3c_3 \\
(39) \quad \frac{dc_4}{dt} &= f_{13}c_1c_3 + b_5c_5 - f_{24}c_2c_4 - f_{14}c_1c_4 - b_4c_4 \\
\frac{dc_5}{dt} &= f_{14}c_1c_4 - f_{15}c_1c_5 - b_5c_5 \\
\frac{dc_6}{dt} &= f_{15}c_1c_5 + f_{22}c_2^3 + f_{24}c_2c_4 + f_{33}c_3^2
\end{aligned}$$

where  $f_{ij}$  is the coefficient for the association of an  $i$ -mer and a  $j$ -mer, and  $b_i$  is the dissociation coefficient for an  $i$ -mer.

We implemented these equations in Matlab, in a similar manner to model (33) (see Figures 4.12 and 4.13). The concentration of hexamers increases and approaches the initial concentration of CA subunits, while the concentrations of all other intermediates approach zero.

**Elongation Model.** The elongation model equations are similar to equations (33). In this case, each subunit is now a hexamer.  $N$  is the number of hexamers in a HIV-1 capsid. The model can be written as

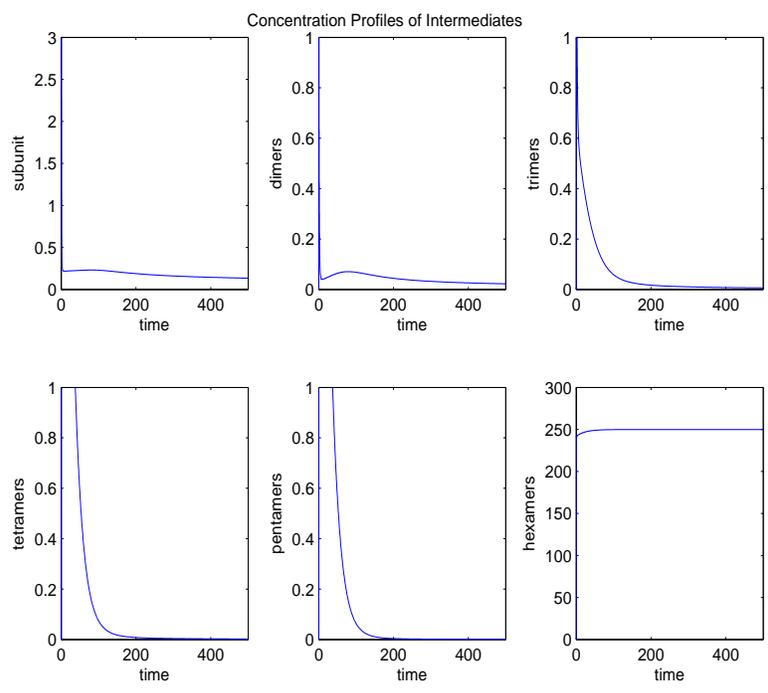


FIGURE 4.12. Concentration Profiles for CA Intermediates of Sizes 1 through 6: Nucleation.

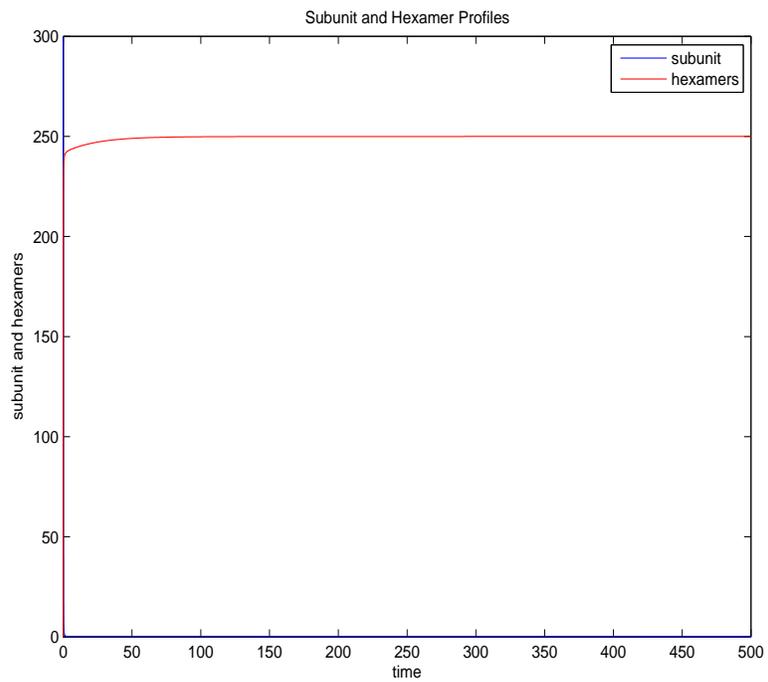


FIGURE 4.13. Concentration Profiles: CA Subunits and Hexamers.

$$\begin{aligned}
\frac{dh_1}{dt} &= -2f_1h_1^2 + b_2h_2 + \sum_{n=2}^N(-f_nh_nh_1 + b_nh_n) \\
\frac{dh_n}{dt} &= f_{n-1}h_1h_{n-1} - f_nh_1h_n - b_nh_n + b_{n+1}h_{n+1} \\
(40) \qquad \qquad n &= 2, \dots, N \\
f_N = b_{N+1} &= 0.
\end{aligned}$$

Here,  $h_n$  represents the concentration of an intermediate formed by  $n$  hexamers.

#### 4.9. REMARKS: BIOLOGICAL IMPLICATIONS AND FURTHER WORK

Our analytical and numerical results show the existence of a stable equilibrium of CA intermediates, which agrees with experimental results. Even though the dynamics of HIV core assembly are not well understood, the assumptions of the models described here, such as:

- modeling capsid assembly as a polymerization event
- the forward coefficient values increase and backward coefficient values decrease with the size of CA intermediates

seem reasonable enough to capture this feature.

Possible avenues for further research on HIV core assembly are identified as follows:

- (1) Further investigate the relationships between the forward and backward coefficients, and the concentrations of intermediates via numerical simulations.
- (2) Additional numerical simulations of the nucleation model for several combinations of parameter values would produce various values for the concentration of hexamers.

These values can be used as the initial concentrations of subunit for the elongation model, which would then lead to different equilibrium concentrations of capsid and intermediates. This would aid in understanding the effects of the various assembly pathways considered in the nucleation process in terms of the cascaded dynamical system model.

- (3) The models may be modified by considering other assembly pathways in the nucleation and elongation stages, to gain further insight into the significance of multiple pathways in HIV core assembly.

## CHAPTER 5

### SUMMARY

In this chapter we summarize the major contributions of this dissertation.

#### 5.1. GAG TRAFFICKING

We proposed a partial differential equation model for Gag trafficking that includes both diffusion and active transport. We developed a characteristic finite element scheme to solve the model numerically. In addition, we implemented the numerical scheme in parallel, utilizing a master-worker approach that combines MPI and OpenMP. We ran simulations for various parameter values and calculated the time  $T_v$  of virion assembly for each parameter combination. In particular, we focused on the active transport parameter  $s$ . We varied this parameter within an acceptable range of values, according to motor protein speeds reported in literature. Our numerical simulations agree principally with biological experiments. The times  $T_v$  obtained are within a reasonable range for Gag transport and virion assembly. Our results indicate that diffusion is not enough for Gag trafficking. We deduce that at least one active transport mechanism, e.g., via motor proteins along microtubules, is involved.

We also identified randomness in Gag trafficking. At the molecular level, motor proteins such as the kinesin KIF4, which have been shown to play an important role in HIV Gag trafficking, move erratically along microtubules. Cellular cargoes have also been observed to be attached to more than one type of motor protein at a time. At the cell level, we observed that kinesin speeds may vary widely within a cell. Also, diffusion and Gag production rates vary from one type of cell to another. In order to study the randomness in Gag trafficking at the cell level, one could study the relationships between the distributions of the diffusion, active transport, and Gag production parameters and the time of virion assembly.

## 5.2. GAG TRIMERIZATION

We developed and studied a two-species model for Gag trimerization near the plasma membrane. The model is a system of two ordinary differential equations, corresponding to the rates of change of concentrations of Gag monomers and trimers. Parameters considered in the model are the rates of: Gag arrival, trimerization, monomerization, and multimerization. We found the equilibrium solution, and showed that the equilibrium is stable regardless of parameter values. We verified our results via numerical simulations. Our conclusions agree with the literature, as an equilibrium between monomeric and trimeric particles in myr-MA and myr-MA-CA experiments has been observed. In addition, we estimated the ratios of model parameters based on information obtained from experiments and from literature. We also determined a condition for equilibrium shift toward the trimeric Gag state and used this condition to estimate a lower bound for the equilibrium association constant  $K_a$  for myr-Gag.

Additionally, we identified avenues for further research on Gag trimerization modeling, such as: utilizing the parameter ratios and the  $K_a$  estimates found here to find estimates for all model parameters, refining the model by considering carrying capacity based on Gag saturation at the Gag hotspot near the plasma membrane, and considering a three-species model that includes Gag monomers in both myr(s) and myr(e) states.

## 5.3. HIV CORE ASSEMBLY

We considered models for HIV core assembly. We first adapted a dynamical system model for capsid assembly, proposed by Zlotnick and others, to HIV core assembly. The system consists of ordinary differential equations that represent the rates of change of the concentrations of CA intermediates. As part of our analysis, we showed the existence of a

unique, stable equilibrium of the concentration of CA intermediates for the case when there are three species. We also found an expression for the equilibrium concentrations in terms of the equilibrium value of the CA subunit for model with  $N$  CA species. We examined the theoretical analysis against numerical simulations, including the existence and stability of the equilibrium and the sigmoidal growth of capsid concentration.

Next, we proposed a cascaded dynamical system model for HIV core assembly. This new model consists of two subsystems, representing nucleation and elongation. The nucleation model consists of six CA species, subunit through nucleus (hexamers), and considers various assembly pathways that lead to the formation of the more stable species, i.e., the hexamers. The model disregards dissociation of hexamers. Numerical simulations suggest that an equilibrium exists, with all CA species approaching zero concentration, except for the hexamers. The elongation model is similar to Zlotnick's model, with each subunit representing now a CA hexamer.

Suggestions for further work on HIV core assembly models include: investigating the relationships between the forward (association) and backward (dissociation) coefficients, and studying the significance of various assembly pathways in the nucleation and elongation processes.

## BIBLIOGRAPHY

- [1] P. H. Raven, G. B. Johnson, J. B. Losos, K. A. Mason, and S. R. Singer, *Biology, 8th ed.* New York, NY: McGraw Hill, 2008.
- [2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell.* New York, NY: Garland Science, 2008.
- [3] L. Briant, B. Gay, C. Devaux, and N. Chazal, “Hiv-1 assembly, release and maturation,” *W. J. of AIDS*, vol. 1, pp. 111–130, 2011.
- [4] A. C. van der Kuyl and B. Berkhout, “The biased nucleotide composition of the hiv genome: a constant factor in a highly variable virus,” *Retrovirology*, vol. 9, no. 92, pp. 1–14, 2012.
- [5] J. A. G. Briggs, M. N. Simon, I. Gross, H. G. Kräusslich, S. D. Fuller, V. M. Vogt, and M. C. Johnson, “The stoichiometry of gag protein in hiv-1,” *Nat. Struct. Mol. Biol.*, vol. 106, no. 27, pp. 11090–11095, 2009.
- [6] M. D. Resh, “A myristoyl switch regulates membrane binding of hiv-1 gag,” *PNAS*, vol. 101, no. 2, pp. 417–418, 2004.
- [7] M. Balasubramaniam and E. O. Freed, “New insights into hiv assembly and trafficking,” *Physiology*, vol. 26, pp. 236–251, 2011.
- [8] B. K. Ganser-Pornillos, M. Yeager, and O. Pornillos, “Assembly and architecture of hiv,” *Adv. Exp. Med. Biol.*, vol. 726, pp. 441–465, 2012.
- [9] N. Jouvenet, P. D. Bieniasz, and S. M. Simon, “Imaging the biogenesis of individual hiv-1 virions in live cells,” *Nature*, vol. 454, pp. 236–240, 2008.

- [10] W. I. Sundquist and H. G. Kräusslich, “Hiv-1 assembly, budding, and maturation,” *Cold Spring Harb. Perspect. Med.*, pp. 1–24, 2012. doi: 10.1101/cshperspect.a006924.
- [11] V. Simon and D. D. Ho, “Hiv-1 dynamics *In Vivo*: Implications for therapy,” *Nat. Rev. Microbiol.*, vol. 1, no. 3, pp. 181–190, 2003.
- [12] S. B. Kutluay and P. D. Bieniasz, “Analysis of the initiating events in hiv-1 particle assembly and genome packaging,” *PLoS Pathog.*, vol. 6, no. 11, pp. 1–13, 2010.
- [13] B. M. Ward, “The taking of the cytoskeleton one two three: How viruses utilize the cytoskeleton during egress,” *Virol.*, vol. 411, pp. 244–250, 2011.
- [14] N. W. Martinez, X. Xue, R. G. Berro, G. Kreitzer, and M. D. Resh, “Kinesin kif4 regulates intracellular trafficking and stability of the human immunodeficiency virus type 1 gag polyprotein,” *J. of Virology*, vol. 82, pp. 9973–9950, 2008.
- [15] M. V. Nermut, W. H. Zhang, G. Francis, F. Ciampor, Y. Morikawa, and I. M. Jones, “Time course of gag protein assembly in hiv-1-infected cells: A study by immunoelectron microscopy,” *Virology*, vol. 305, pp. 219–227, 2003.
- [16] J. Liu, R. Munoz-Alicea, T. Huang, S. Tavener, and C. Chen, “A mathematical model for intracellular hiv-1 gag protein transport and its parallel numerical simulations,” *Procedia. Comput. Sci.*, vol. 9, pp. 679–688, 2012.
- [17] A. Ploubidou and M. Way, “Viral transport and the cytoskeleton,” *Cur. Op. Cell Biol.*, vol. 13, pp. 97–105, 2001.
- [18] B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, and W. I. Sundquist, “Assembly and analysis of conical models for the hiv-1 core,” *Science*, vol. 283, pp. 80–83, 1999.
- [19] J. Benjamin, B. K. Ganser-Pornillos, W. F. Tivol, W. I. Sundquist, and G. J. Jensen, “Three-dimensional structure of hiv-1 virus-like particles by electron cryotomography,”

- J. Mol. Biol.*, vol. 346, pp. 577–588, 2005.
- [20] S. Katen and A. Zlotnick, “The thermodynamics of virus capsid assembly,” *Methods Enzymol.*, vol. 455, pp. 395–417, 2009.
- [21] M. F. Hagan and D. Chandler, “Dynamic pathways for viral capsid assembly,” *Biophys. J.*, vol. 91, pp. 42–54, 2006.
- [22] M. F. Hagan and O. M. Elrad, “Understanding the concentration dependence of viral capsid assembly kinetics - the origin of the lag time and identifying the critical nucleus size,” *Biophys. J.*, vol. 98, pp. 1065–1074, 2010.
- [23] A. Zlotnick, J. M. Johnson, P. W. Wingfield, S. J. Stahl, and D. Endres, “A theoretical model successfully identifies features of hepatitis b virus capsid assembly,” *Biochemistry*, vol. 38, no. 44, pp. 14644–14652, 1999.
- [24] <http://micro.magnet.fsu.edu>.
- [25] J. E. Duncan and L. S. B. Goldstein, “The genetics of axonal transport and axonal transport disorders,” *PLoS Genet.*, vol. 2, no. 9, 2006. e124. doi: 10.1731/journal.pgen.0020124.
- [26] C. L. Asbury, A. N. Fehr, and S. M. Block, “Kinesin moves by an asymmetric hand-over-hand mechanism,” *Science*, vol. 302, pp. 2130–2134, 2003.
- [27] <http://news.stanford.edu>.
- [28] K. Luby-Phelps, “Cytoarchitecture and physical properties of cytoplasm: Volume, viscosity, diffusion, intracellular surface area,” *Int. Rev. of Cytol.*, vol. 192, pp. 189–221, 2000.
- [29] M. E. Welte, “Bidirectional transport along microtubules,” *Current Biology*, vol. 14, pp. R525–527, 2004.

- [30] R. H. Ghanam, A. B. Samal, T. F. Fernandez, and J. S. Saad, “Role of the hiv-1 matrix protein in gag intracellular trafficking and targeting to the plasma membrane for virus assembly,” *Front. in Microbiol.*, vol. 3, no. 55, pp. 1–15, 2012.
- [31] Y. Tang, U. Winkler, E. O. Freed, T. A. Torrey, W. Kim, H. Li, S. P. Goff, and H. C. M. III, “Cellular motor protein kif-4 associates with retroviral gag,” *J. Of Virol.*, vol. 73, no. 12, pp. 10508–10513, 1999.
- [32] B. Sodeik, “Unchain my heart, baby let me go: the entry and intracellular transport of hiv,” *T. J. of Cell Biol.*, vol. 159, no. 3, pp. 393–395, 2002.
- [33] T. Arbogast and C. S. Huang, “A fully mass and volume conserving implementation of a characteristic method for transport problems,” *SIAM J. Sci. Comput.*, vol. 28, pp. 2001–2022, 2006.
- [34] J. D. Jr. and T. F. Russell, “Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures,” *SIAM J. Numer. Ana.*, vol. 19, pp. 871–885, 1982.
- [35] J. Liu and C. Chen, “A characteristic finite element method with local mesh refinements for the lamm equation in analytical ultracentrifugation,” *Numer. Meth. PDEs*, vol. 25, pp. 292–310, 2009.
- [36] J. Liu and R. E. Ewing, “An operator splitting method for nonlinear reactive transport equations and its implementation based on dll and com,” *Proceedings of Current Trends in High Performance Computing and its Applications*, pp. 93–102, 2005.
- [37] J. Liu and S. Tavener, “Semi-implicit spectral collocation methods for reaction-diffusion equations on annuli,” *Numer. Meth. PDEs*, vol. 27, pp. 1113–1129, 2011.

- [38] H. Wang, H. K. Dahle, R. E. Ewing, M. S. Espedal, R. C. Sharpley, and S. Man, “An ellam scheme for advection-diffusion equations in two dimensions,” *SIAM J. Sci. Comput.*, vol. 20, pp. 2160–2194, 1999.
- [39] S. Zhao, J. Ovadia, X. Liu, Y. T. Zhang, and Q. Nie, “Operator splitting implicit integration factor methods for stiff reaction-diffusion-advection systems,” *J. Comput. Phy.*, vol. 230, pp. 5996–6009, 2011.
- [40] C. Chen. Personal communication, 2013.
- [41] M. Gazzola, C. J. Burckhardt, B. Bayati, M. Engelke, U. F. Greber, and P. Koumoutsakos, “A stochastic model for microtubule motors describes the in vivo cytoplasmic transport of human adenovirus,” *PLOS Comp. Biol.*, vol. 5, no. 12, pp. 1–6, 2009. e1000623. doi:10.1371/journal.pcbi.1000623.
- [42] S. A. K. Datta, Z. Zhao, P. K. Clark, S. Tarasov, J. N. Alexandratos, S. J. Campbell, M. Kvaratskhelia, J. Lebowitz, and A. Rein, “Interactions between hiv-1 gag molecules in solution: An inositol phosphate-mediated switch,” *J. Mol. Biol.*, vol. 365, pp. 799–811, 2007.
- [43] S. A. K. Datta, X. Zuo, P. K. Clark, S. J. Campbell, Y. X. Wang, and A. Rein, “Solution properties of murine leukemia virus gag protein: Differences from hiv-1 gag,” *J. Virol.*, vol. 85, no. 23, pp. 12733–12741, 2011.
- [44] D. L. Coy, M. Wagenbach, and J. Howard, “Kinesin takes one 8-nm step for each atp that it hydrolyzes,” *J. Biol. Chem.*, vol. 274, pp. 3667–3671, 1999.
- [45] A. Seitz and T. Surrey, “Processive movement of single kinesins on crowded microtubules visualized using quantum dots,” *T. EMBO J.*, vol. 25, pp. 267–277, 2006.

- [46] R. Casey, “Istec cray: High performance computing system.” [http://istec.colostate.edu/pdf/cray/tutorial\\_3\\_30\\_11.pdf](http://istec.colostate.edu/pdf/cray/tutorial_3_30_11.pdf), 2011.
- [47] [http://people.sc.fsu.edu/~jburkard/cpp\\_src/gprof/gprof.html](http://people.sc.fsu.edu/~jburkard/cpp_src/gprof/gprof.html).
- [48] S. Lakamper and E. Meyhofer, “The e-hook of tubulin interacts with kinesin head to increase processivity and speed,” *Biophys. J.*, vol. 89, pp. 3223–3234, 2005.
- [49] D. S. Friedman and R. D. Vale, “Single-molecule analysis of kinesin motility reveals regulation by the cargobinding tail domain,” *Nat. Cell Bio.*, vol. 1, pp. 293–297, 1999.
- [50] <http://mbowden.surf3.net>.
- [51] S. Campbell, R. J. Fisher, E. M. Towler, S. Fox, H. J. Issaq, T. Wolfe, L. R. Phillips, and A. Rein, “Modulation of hiv-like particle assembly in vitro by inositol phosphates,” *PNAS*, vol. 98, no. 19, pp. 10875–10879, 2001.
- [52] D. Xiu, *Numerical Methods for Stochastic Computations: A Spectral Methods Approach*. Princeton, New Jersey: PUP, 2010.
- [53] N. Jouvenet, S. M. Simon, and P. D. Bieniasz, “Imaging the interaction of hiv-1 genomes and gag during assembly of individual viral particles,” *PNAS*, vol. 106, no. 45, pp. 19114–19119, 2009.
- [54] E. Hamard-Peron and D. Muriaux, “Retroviral matrix and lipids, the intimate interaction,” *Retrovirology*, vol. 8, no. 15, pp. 1–13, 2011.
- [55] E. L. Fledderman, K. Fujii, R. H. Ghanam, K. Waki, P. E. Prevelige, E. O. Freed, and J. S. Saad, “Myristate exposure in the hiv-1 matrix protein is modulated by ph,” *Biochemistry*, vol. 49, no. 44, pp. 9551–9562, 2010.
- [56] C. Tang, E. Loeliger, P. Luncsford, I. Kinde, D. Beckett, and M. F. Summers, “Entropic switch regulates myristate exposure in the hiv-1 matrix protein,” *PNAS*, vol. 101, no. 2,

- pp. 517–522, 2004.
- [57] R. A. Frazier and R. B. M. Schasfoort, *Handbook of Surface Plasmon Resonance*. Cambridge, UK: RSC, 2008.
- [58] J. C. Polking. <http://math.rice.edu/~dfield/>, 2009.
- [59] C. Tang. Personal communication, 2013.
- [60] M. Yeager, “Design of in vitro symmetric complexes and analysis by hybrid methods reveal mechanisms of hiv capsid assembly,” *J. Mol. Biol.*, vol. 410, pp. 534–552, 2011.
- [61] J. A. G. Briggs, K. Grünewald, B. Glass, F. Förster, H. G. Kräusslich, and S. D. Fuller, “The mechanism of hiv-1 core assembly: insights from three dimensional reconstructions of authentic virions,” *Structure*, vol. 14, pp. 15–20, 2006.
- [62] J. A. G. Briggs and H. G. Kräusslich, “The molecular architecture of hiv,” *J. Mol. Biol.*, vol. 410, pp. 491–500, 2011.
- [63] J. A. G. Briggs, J. D. Riches, B. Glass, V. Bartonova, G. Zanetti, and H. G. Kräusslich, “Structure and assembly of immature hiv,” *PNAS*, vol. 106, no. 27, pp. 11090–11095, 2009.
- [64] O. Elrad and M. F. Hagan, “Mechanisms of size control and polymorphism in viral capsid assembly,” *Nano Lett.*, vol. 8, no. 11, pp. 3850–3857, 2008.
- [65] O. M. Elrad and M. F. Hagan, “Encapsulation of a polymer by an icosahedral virus,” *Phys. Biol.*, vol. 7, no. 045003, pp. 1–17, 2010.
- [66] D. Endres, M. Miyahara, P. Moisant, and A. Zlotnick, “A reaction landscape identifies the intermediates critical for self-assembly of virus capsids and other polyhedral structures,” *Protein Sci.*, vol. 14, pp. 1518–1525, 2005.

- [67] M. F. Hagan, “Controlling viral capsid assembly with templating,” *Phys. Rev. E*, vol. 77, no. 5, pp. 1–10, 2008.
- [68] M. F. Hagan, “A theory for viral capsid assembly around electrostatic cores,” *J. Chem. Phys.*, vol. 130, no. 114902, pp. 1–17, 2009.
- [69] M. F. Hagan, O. M. Elrad, and R. L. Jack, “Mechanisms of kinetic trapping in self-assembly and phase transformation,” *J. Chem. Phys.*, vol. 135, no. 104115, pp. 1–13, 2011.
- [70] P. Moisant, H. Neeman, and A. Zlotnick, “Exploring the paths of (virus) assembly,” *Biophys. J.*, vol. 99, pp. 1350–1357, 2010.
- [71] O. Pornillos, B. K. Ganser-Pornillos, and M. Yeager, “Atomic level modeling of the hiv capsid,” *Nature*, vol. 469, no. 7330, pp. 424–427, 2011.
- [72] S. Singh and A. Zlotnick, “Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly,” *J. of Biol. Chem.*, vol. 278, no. 20, pp. 18249–18255, 2003.
- [73] S. Whitelam, E. H. Feng, M. F. Hagan, and P. L. Geissler, “The role of collective motion in examples of coarsening and self-assembly,” *Soft Matter*, vol. 5, pp. 1251–1262, 2009.
- [74] A. Zlotnick, “Are weak protein–protein interactions the general rule in capsid assembly?,” *Virology*, vol. 315, pp. 269–274, 2003.
- [75] A. Zlotnick, “Distinguishing reversible from irreversible virus capsid assembly,” *J. Mol. Biol.*, vol. 366, pp. 14–18, 2007.
- [76] A. Zlotnick and S. Mukhopadhyay, “Virus assembly, allostery and antivirals,” *Trends in Microbiology*, vol. 19, no. 1, pp. 14–23, 2011.
- [77] A. Zlotnick and S. J. Stray, “How does your virus grow? understanding and interfering with virus assembly,” *Trends in Biotechnology*, vol. 21, no. 12, pp. 536–542, 2003.

- [78] D. Endres and A. Zlotnick, “Model-based analysis of assembly kinetics for virus capsids or other spherical polymers,” *Biophys. J.*, vol. 83, pp. 1217–1230, 2002.
- [79] A. Kivenson and M. F. Hagan, “Mechanisms of capsid assembly around a polymer,” *Biophys. J.*, vol. 99, pp. 619–628, 2010.