DISSERTATION

MATHEMATICAL MODELS FOR HIV-1 VIRAL CAPSID STRUCTURE AND ASSEMBLY

Submitted by

Farrah Sadre-Marandi

Department of Mathematics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2015

Doctoral Committee:

    Advisor: Jiangguo (James) Liu
    Co-Advisor: Simon Tavener

    Chaoping Chen
    Alexander Hulpke
    Yongcheng Zhou

Abstract

Mathematical Models for HIV-1 Viral Capsid Structure and Assembly

HIV-1 (human immunodeficiency virus type 1) is a retrovirus that causes the acquired immunodeficiency syndrome (AIDS). This infectious disease has high mortality rates, encouraging HIV-1 to receive extensive research interest from scientists of multiple disciplines. Group-specific antigen (Gag) polyprotein precursor is the major structural component of HIV. This protein has 4 major domains, one of which is called the capsid (CA). These proteins join together to create the peculiar structure of HIV-1 virions. It is known that retrovirus capsid arrangements represent a fullerene-like structure. These caged polyhedral arrangements are built entirely from hexamers (6 joined proteins) and exactly 12 pentamers (5 proteins) by the Euler theorem. Different distributions of these 12 pentamers result in icosahedral, tubular, or the unique HIV-1 conical shaped capsids. In order to gain insight into the distinctive structure of the HIV capsid, we develop and analyze mathematical models to help understand the underlying biological mechanisms in the formation of viral capsids.

The pentamer clusters introduce declination and hence curvature on the capsids. The HIV-1 capsid structure follows a (5,7)-cone pattern, with 5 pentamers in the narrow end and 7 in the broad end. We show that the curvature concentration at the narrow end is about five times higher than that at the broad end. This leads to a conclusion that the narrow end is the weakest part on the HIV-1 capsid and a conjecture that "the narrow end closes last during maturation but opens first during entry into a host cell."

Models for icosahedral capsids are established and well-received, but models for tubular and conical capsids need further investigation. We propose new models for the tubular

and conical capsid based on an extension of the Caspar-Klug quasi-equivalence theory. In particular, two and three generating vectors are used to characterize respectively the lattice structures of tubular and conical capsids. Comparison with published HIV-1 data demonstrates a good agreement of our modeling results with experimental data.

It is known that there are two stages in the viral capsid assembly: nucleation (formation of a nuclei: hexamers) and elongation (building the closed shell). We develop a kinetic model for modeling HIV-1 viral capsid nucleation using a 6-species dynamical system. Numerical simulations of capsid protein (CA) multimer concentrations closely match experimental data. Sensitivity and elasticity analysis of CA multimer concentrations with respect to the association and disassociation rates further reveals the importance of CA dimers in the nucleation stage of viral capsid self-assembly.

Keywords: CA protein, capsid, cone, curvature, dynamical systems, hexamer, HIV-1, icosahedron, pentamer, sensitivity analysis, tube

I would like to thank my advisors, my committee members, as well as my friends and family for the support they have given me these past four years. Without their encouragement, this thesis would not be possible.

This dissertation is typset in LaTeX using a document class designed by Leif Anderson.

CHAPTER 1

INTRODUCTION

Viruses are macromolecular organisms that are constituted by at least two parts: genetic materials (DNA or RNA) and a protein shell that protects the genes, called the capsid. Some viruses contain an additional layer around the capsid called the envelope [17]. To enter a cell, a virus goes through a fusion process with the cell membrane. Then the nucleic acid (and other materials) enter the cell. With some viruses, the genome is completely released from the capsid during or after penetration. In others, such as retroviruses, the first stages of the viral replication cycle occur inside the capsid. The major goal of the virion is to replicate itself using the invaded cell's material. New viral genomes and viral components are produced. This can happen in a number of ways depending on the family of virus. Once the new materials have been produced, they are formed into new virions. The viral material collects near the surface of the cell, which then undergoes a maturation process where the capsid reassembles around the DNA or RNA. Once a cell is infected by a virus, it continues to undergo DNA or RNA synthesis and mitosis, polluting the host with infected cells.

It is well known that viruses are virulent to their host, but viruses can also be used to benefit the society, since the capsid serves as a platform for synthetic manipulation. Capsid reassembly properties have been employed to build a new generation of batteries that act as powerful and highly efficient fuel cells [15]. Virus-like capsids have been created to attach and selectively release the anticancer drug doxorubicin in cancer patients [17]. They have also been used as pest control agents, as well as applied to gene therapy [55]. It is also quite interesting to see that viral capsids and other protein cages can be used as containers for polymers and nano-particles to make new synthetic materials [20, 22, 35, 57].

1.1. AIDS, HIV-1 and Retrovirus

AIDS is caused by the human immunodeficiency virus (HIV), which was unknown until the early 1980's [63]. Since then, AIDS has become one of the most devastating infectious diseases to have emerged in recent history, being spread around the world to infect millions of people. The target of HIV infection is the immune system, which is gradually destroyed. Once a person becomes infected with HIV, he/she is at high risk for other illness and death from infections.

The type of HIV that is the cause for almost all infections is known as HIV-1. There are two main different types of HIV: HIV-1 and HIV-2. HIV-2 is less virulent and is not as widely spread [1, 63]. The major modes of spread are as a sexually transmitted disease, needle sharing, and perinatal infection [1, 63]. HIV-1 is not spread by casual contact or by insect vectors. In this work, unless otherwise specified, HIV refers to HIV-1.

After initial entry of HIV and establishment of infection, the viral replication generally occurs within the inflammatory cells at the site of infection. The replication quickly shifts to the lymphoid tissues of the body, including lymph nodes, spleen, liver, and bone marrow. Primary infection may go unnoticed in at least half of the cases, or the patient may present with signs and symptoms of a flu-like illness, consisting of fever, malaise, and/or a rash [1]. At this time the disease is mild, and will subside over 1 to 2 months. This is followed by a long clinical "latent" period. On average, an HIV-infected person may live up to 8 or 10 years before the development of the signs and symptoms of AIDS [63].

To date, there is no cure for HIV infection, although treatments for HIV have been developed. There have also been major advances in working towards a vaccine and daily medication, called Pre-Exposure Prophylaxis (PrEP), for prevention. These advances have

reduced the risk of HIV infection in people who are at high risk by up to 92 % by keeping the virus from establishing a permanent infection. Yet, there is still much unknown about this virus that causes AIDS, and only treatment of the symptoms with those already infected with AIDS or HIV. We aim to better understand the HIV's life cycle in order to aid in determining more efficient ways to treat the virus and disease.

1.2. HIV and Gag Proteins

In order to examine HIV's replication cycle, we first need to understand the virus's composition. HIV is composed of several main proteins, such as Env, Pol, Gag, and others [1, 7, 8, 64]. The focus of this dissertation is aimed at Gag proteins, which is one of the main structural components.

Gag. Gag (group specific antigen) is essential in the formation of new HIV virions. *In vivo*, HIV Gag proteins have numerous and complex roles during the life cycle. These roles include assembly and virion maturation. *In vitro*, virus-like particles (VLPs) can sufficiently assemble with just the Gag proteins of HIV.

Gag has 4 major domains: matrix (MA), capsid (CA), nucleocapsid (NC), and p6 (see Figure 1.1). Each domain has its own function in the formation of a mature virion. The MA domain binds to the interior of the plasma membrane (envelope). The CA domain forms the capsid shell which assembles around the RNA. The NC domain binds to the viral RNA. Lastly, the p6 is involved in the virion's budding from the host cell [7, 69].

Retroviral Life Cycle. HIV can only replicate inside of cells. In fact, HIV hijacks the cell to use the host's material during replication. HIV utilizes a protein called CD4, carried on the surface of a cell and receptors to enter the CD4 T cells and macrophages [1, 4]. The virus is able to invade the cell by sticking to the CD4, allowing the viral envelope
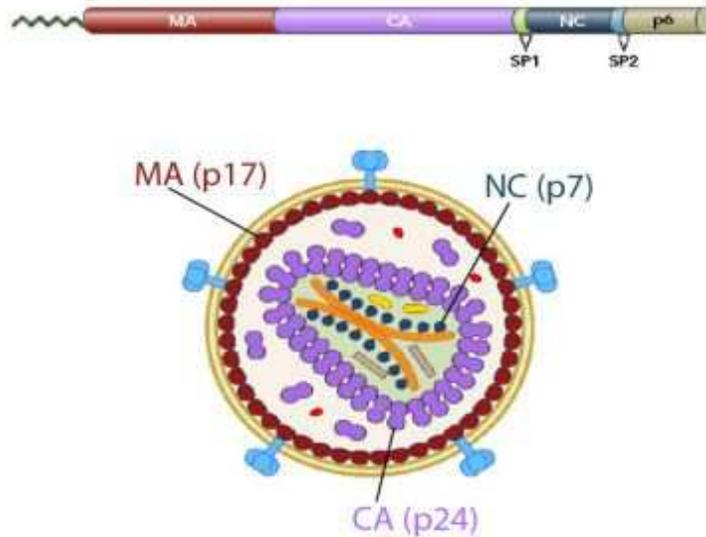
FIGURE 1.1. HIV Gag. The Gag domains MA, CA, NC, and p6 are the main structural components of the HIV particle. The MA domain binds to the interior of the viral membrane with a spherical shape. The CA domain forms the viral conical capsid. The NC domain attaches to the RNA. p6 helps the virion bud from the host cell [4].

and plasma membrane to fuse. Once the virus and cell are fused together, the genomic RNA and other proteins are released into the cell cytoplasm to start the translation process [1].

Once inside the cell, the HIV enzyme called reverse transcriptase converts the viral RNA into DNA. This DNA is transported to the cell's nucleus, where it is integrated into the human DNA by the HIV enzyme integrase [1]. The host cell transcribes the viral DNA into viral messenger RNA (mRNA), which then travels to the cytoplasm. Complete copies of HIV genetic material are contained among the strands of the mRNA. New Gag proteins are synthesized from the mRNA [47] and travel to the plasma membrane to join the assembly of new immature virions.

The new viral particles are then released from the cell, by a process known as 'budding'. Many viral particles can bud from a cell over the course of time, then begin the process

of maturation. During maturation, each newly formed immature virion develops the capsid core, which contains RNA and other proteins [7, 69] (see Figure 1.2). This process is explained in more detail in Chapter 2 section 3.
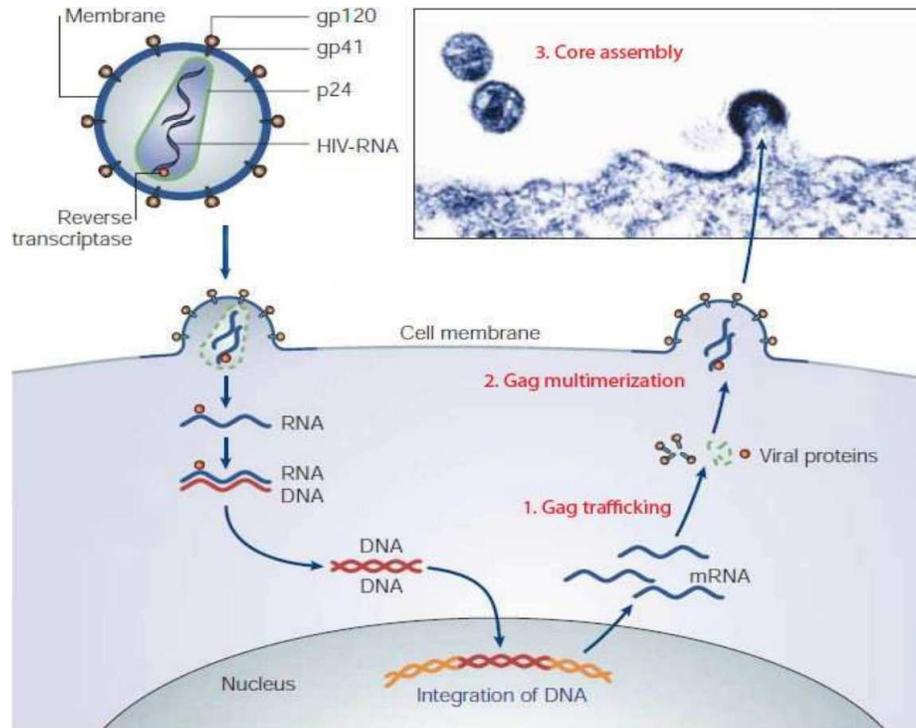
FIGURE 1.2. Retroviral Life Cycle. HIV fuses with the host cell's plasma membrane and releases its RNA and enzymes in the cytoplasm. HIV's RNA is translated to DNA by the enzyme reverse transcriptase. The viral DNA is integrated into the host cell's DNA by the enzyme integrase. The host cell translates 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 [68].

1.3. MATHEMATICAL MODELING FOR VIRUS LIFE CYCLE

Quantitative description of the HIV life cycle, in particular mathematical models, will help us understand the mechanisms of the virus. In this section, we discuss two developed mathematical models that shall provide quantitative characterization of detailed aspects of HIV reproduction.

Gag Trimerization at Plasma Membrane. A mathematical model for Gag trimerization at the plasma membrane was created in [52]. It is assumed that Gag proteins arrive at certain "Gag hotspots," where HIV virion assembly takes place. Gag proteins arrive as monomers at a constant rate, then three monomers join to form a trimer. These trimers join to form higher order mulimers, which bind at the plasma member as a new immature HIV virion assembles (see Figure 1.2).

The model consists of a nonlinear dynamical system of two ordinary differential equations. The existence and stability of a unique equilibrium is analytically shown and verified numerically [52]. In addition, a condition on the model parameters that shift the Gag monomer-trimer equilibrium towards the trimer state is derived. The lower bound for the equilibrium association constant K_a for Gag monomers and trimers is also calculated.

Viral Protein Trafficking and Binding. Quantitative results for intracellular trafficking and assembly of gag proteins have critical importance for gaining insights into the processes of virus replication and for developing novel control strategies [4, 64].

Our recent work [76] has established a model for integrating the simultaneous treatments of gag monomers and trimers in the dynamical process of transport and binding. The model characterizes the dynamics of virus trafficking and the transformation between monomeric and trimeric states by coupling different types of differential equations.

Numerical simulation results show that the gag protein trimers will accumulate at the membrane of the cell. Numerical results on the time when the first new virions appear near the cell membrane (T_a) are in very good agreement with published experiment data. Sensitivity analysis of T_a to the model parameters indicates that the diffusion and transport process affects the time of initial appearance of HIV-1 virions on the cell membrane.

1.4. Contribution of This Thesis

This thesis contains three major parts. Here we highlight the efforts and contributions of each.

Curvature Concentration on HIV Conical Cores. Viral capsids follow the fullerene-like structure with exactly 12 pentamers by the Euler theorem. Different distributions of the pentamers result in different shapes of capsids. These pentamers introduce declination and curvature on the capsids [79]. Our model intends to provide an explicit and quantitative characterization of curvature on virus capsids.

The discrete setting of the Gauss-Bonnet Theorem is applied to viral capsids for calculating the angle defect at each hexamer and pentamer. For the HIV (5,7)-cone, it is shown in [42] that the curvature concentration at the narrow end is about *five times higher* than that at the broad end. This leads to a conclusion that the narrow end is the weakest part on the HIV-1 capsid and a conjecture that "the narrow end closes last during maturation and opens first during entry into a host cell."

The modeling results should be helpful for better understanding the HIV-1 capsid structure and the underlying biology. Curvature formalism is novel to the structural virology field and can be used to rank the stability of (related) capsids.

Generating Vectors for the Lattice Structures of Viral Capsids. Virus capsids are best described by fullerene-like structures. A fullerene is any molecule composed entirely of carbon with a distinct cage-like structure defined by a simple 3-valent, n-vertex polyhedron. It is known that viral capsids could be categorized into three major types: icosahedron, tube, and cone [6, 79]. Though, there are irregular viral capsid shapes which do not fall into

there three categories. Mathematical models for the three main lattice structures help understand the underlying biological mechanisms in the formation of viral capsids. While the models for icosahedral capsids are established and well-received, tubular and conical capsids are not yet fully understood.

Our work [66] establishes a unified approach for the three common capsid shapes by extending the Caspar and Klug theory [14] and overcomes the flaw of incomplete closure when existing models are inappropriately applied [54]. In particular, one generating vector is needed to build an icosahedron, while two and three generating vectors are used to characterize respectively the lattice structures of tubular and conical capsids.

Comparison of our models with published HIV-1 data demonstrates a good agreement of our modeling results with experimental data, validating the new model for a tubular and conical capsid.

Viral Capsid Nucleation. The major goal of this part is to develop models for viral capsid assembly. Existing work has modeled viral capsid assembly using one large-size dynamical system, combining the two substages: nucleation and elongation [30].

Our approach focuses on nuclei growth (nucleation), relatively independent of capsid elongation. Investigating the nucleation stage first gives this model a unique advantage for characterizing conditions required to start capsid formation and producing the building blocks for the mature capsid. It also allows us to examine the favorable and unfavorable conditions for nucleation.

Since some biological parameters in these models are difficult to measure in experiments, mathematical analysis enables us to characterize the most important or sensitive parameters. A 6-species dynamical system model is created based on [37, 85], parameters are estimated

to fit biological data, sensitivity and elasticity analysis of CA multimer concentrations with respect to model parameters is performed. The elasticity analysis confirms the biological experiments that the dimer intermediate is vital for capsid protein self-assembly.

CHAPTER 2

BIOLOGICAL PROBLEMS: HIV-1 STRUCTURE AND

ASSEMBLY

2.1. HIV VIRUS

The human immunodeficiency virus type 1 (HIV-1) is a retrovirus that causes the acquired immunodeficiency syndrome (AIDS). Due to the exceptionally high mortality rates through AIDS and the unique structure of HIV-1 virions, HIV-1 virus is an active research area, see, e.g., [9, 26, 56, 84] and references therein. In this dissertation, we refer to HIV-1 as HIV.

AIDS is a relatively new disease, from the evolutionary point of view. Humans have not yet been able to adapt to it. HIV attacks T-lymphocytes and macrophages. In particular, HIV infects and kills CD4$^+$ T helper cells [1, 63], which allow the immune response to fight against invading pathogens. Individuals left infected and untreated usually develop AIDS between 8-10 years after infection [63]. AIDS causes most people to have very weak immune systems; they eventually die of infection due to the body's inability to heal itself.

HIV is retrovirus and member of the lentivirus family. The virion consists of three parts: single-stranded genes made from RNA, a protein shell (capsid) that protects the genome and an envelope composed of lipids. A retrovirus has the additional ability to mutate easily, in large part due to the error rate of the reverse transcriptase enzyme, which introduces a mutation approximately once per 2000 incorporated nucleotides [1, 63]. This presents a big dilemma since high mutation rates lead to the emergence of HIV variants within the infected person's cells that can escape immune attack or can resist drug therapy. It is already

difficult for the immune system to fight off HIV. Over time, different tissues of the body may harbor differing HIV variants. High mutation rates create a challenge in developing effective vaccines [1].

To date, HIV is the only known virus forming a conical capsid. The formation of this conical capsid (Figure 2.1) occurs in the maturation stage and is essential for the virion to become infectious. Our work is focused on the capsid in two ways: the mature capsid structure and the dynamics of assembly.

2.2. HIV CAPSID STRUCTURE

HIV-1 conical core along with other virus capsids are best described by a simple 3-valent, n-vertex polyhedral surface. Geometrically, the capsid forms a closed surface, consisting of hexagons and exactly 12 pentagons, according to the Euler theorem. The CA proteins join together to create the hexagon and pentagon shapes by six (hexamer) and five (pentamer)
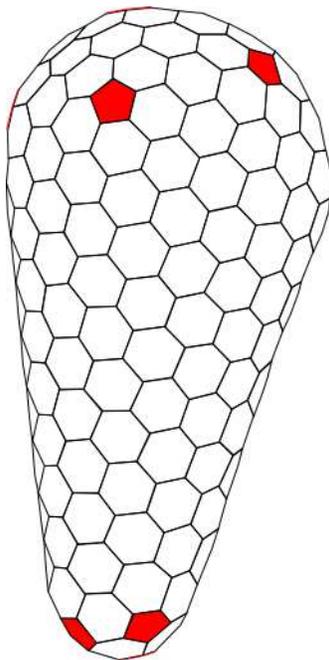
FIGURE 2.1. Illustration of HIV-1 conical capsid.

proteins. The binding among the CA hexamers and pentamers are further related to the hinge between the C-terminal domain (CTD) and N-terminal domain (NTD) of the CA protein [13, 26].

Curvature on Viral Capsids. Existing biological work [25, 44, 61, 79, 81, 82] suggest that the pentamers introduce sharp declinations in the HIV-1 capsid. It is discussed in [26] that the angle between adjacent hexamers vary from around 135° between two hexamers connected to the same pentamers at either end of the HIV-1 cone, to around 180° in the more flat region in the middle of the cone (Figure 2.1). It is concluded in [61] that the rigid-body rotations around these assembly interfaces seem to be sufficient for explaining the curvature variation on the HIV-1 cone. Recent experimental and modeling studies show that the HIV-1 capsid narrow end might not close, if conditions are unfavorable [80]. There arises a need for explicit and quantitative characterization of curvature on virus capsids.

Generating Vectors for the Lattice Structure of Viral Capsids. The icosahedral viral capsid has been extensively studied due to the highly symmetric nature. The model for icosahedral capsids is established and well-received, but models for tubular and conical capsids need further investigation. In [54], Nguyen et al. used two generating vectors for the lattice structure of tubes and cones. However, an important piece of information is missing for each type of capsid.

(i) For tubes, the previous model has limitations in the height. Multiple well known tubular viral capsids cannot be completely described by this model. With changes in definition of the leading scalars, we show more viral capsids are better defined by our new model presented in this dissertation. (ii) The conical model presented in [54] has a flaw that could lead to improper closure of the capsids when misused. An additional generating vector is

needed to guarantee the closure observed in biological experiments. There is a need for more accurate models for tubular and conical viral capsids. Not only will this lead to a characterization of more general capsids, but give insight on the position of the twelve pentamers needed for closure.

2.3. HIV Maturation

New HIV virions that form near the plasma membrane of a host cell escape the cell during the process of budding. Virions undergo a maturation process in order to become infectious and invade other cells as shown in Figure 2.2. During this stage the virus proteins assemble into a strong shell, called the capsid, with two substages: nucleation and elongation. The capsid acts as a protective shell for the DNA or RNA inside the virus and the capsid is in
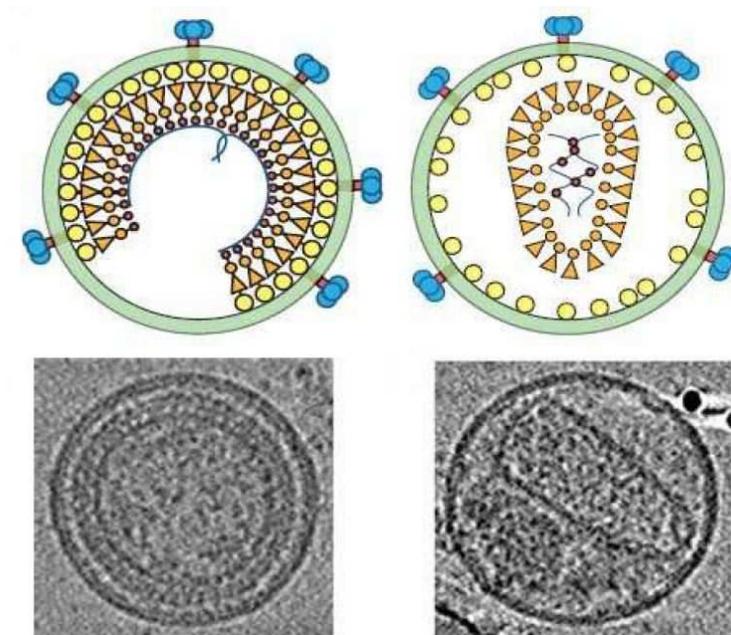
FIGURE 2.2. Immature and Mature HIV Virions. About $5,000$ Gag proteins are spread radially and uniformly inside the immature HIV virion (left). In the mature HIV virion (right), Gag's MA domain (yellow) is attached to the inner layer of the virion's membrane, while about $1,500$ of the available $5,000$ CA proteins form the capsid (outer shell) of the virion's core. The core packages two strands of viral RNA and other proteins [69].

its weakest stage during maturation. After the capsid matures, the virus is able to attack new cells and replicate its DNA or RNA, polluting the host with infected cells. Therefore, it is of great interest to understand the formation of the capsid, with the goal of developing innovative antiviral therapies that can break or control capsid development.

Icosahedral Capsid. Models for the assembly of simple icosahedral capsids have been established in the past, but their extension to retrovirus has been over simplified, due to the complicated shape of a retroviral capsid. While significant progresses have been made with regard to understanding the assembly mechanism and structure of HIV-1 capsid [16, 19, 24, 29, 30, 61, 84], there are many questions yet to be answered.

Nucleation and Elongation. Previous work has modeled the whole process of viral capsid assembly using one large-size dynamical system [21, 31], combining both the nucleation and elongation phases of the capsid assembly. It has been observed in biological experiments and exploited in simulations that separate modeling and simulations of nucleation and elongation stages shall help alleviate the difficulty in the aforementioned approach. Focusing on each stage will bring different perspectives for modeling viral capsid assembly. Simplifying the models to study nucleation and elongation separately allows us to examine the favorable and unfavorable conditions for each stage. Since some biological parameters in these models are difficult to measure in experiments, mathematical analysis will enable us to characterize which parameters are the most important or sensitive. Therefore, it is of great interest to understand the nucleation process on its own, before the nuclei form together in the elongation stage, to form a complete capsid.

Dimer Pathways. Previous models also consider a simplified pathway that only allows association or dissociation of one monomer at a time [21, 31]. That is, the capsid

only changes between n-mer and $(n+1)$-mer. However, there is strong evidence that dimers form with other dimers [16]. The findings in [60] suggest that even higher order subunits can assemble with each other. Therefore, there is a need for research focused on exploring models where larger intermediates can bind with each other.

CHAPTER 3

Curvature Concentration on Viral Capsids

Viral capsids are best described by a fullerene-like structure. Generally speaking, the capsids are composed of a varying amount hexamers and exactly 12 pentamers. Small-size virus capsids tend to conform to the preferred icosahedral symmetry [14, 34]. This symmetry allows the 12 pentamer groups to be evenly placed on the surface of a sphere. Although most virus capsids follow the fullerene-like structure, not all virus capsids follow the icosahedral symmetry. Tubular (spherocylinder) viral capsids have been observed for Cowpea Chlorotic Mottle Virus and Alfalfa Mosaic Virus, among others. The cone-shaped HIV capsid is composed of exactly 12 pentamers and approximately 218 hexamers [9, 61, 79]. The murine leukemia virus (MuLV) and Rous sarcoma virus (RSV) also exhibit asymmetry or irregularity in their capsid structures [32].

3.1. HIV Conical Cores

HIV is the only known virus to date with a conical core. The cone angle of the HIV capsid has been measured in experiments by dehydrating the core of a virus-like particle (VLP) onto a carbon grid [24]. It is found that the angle was quantified into the five allowed values prescribed by the Euler formula

$$
\sin(\theta/2) = 1 - P/6, \tag{1}
$$

where θ is the cone angle and P is the number of pentamers at the narrow end of the cone, as shown in Figure 3.1. The five angle values (and the corresponding P values) are $\theta = 112.9°(P = 1)$, $\theta = 83.6°(P = 2)$, $\theta = 60°(P = 3)$, $\theta = 38.9°(P = 4)$, $\theta = 19.2°(P = 5)$.
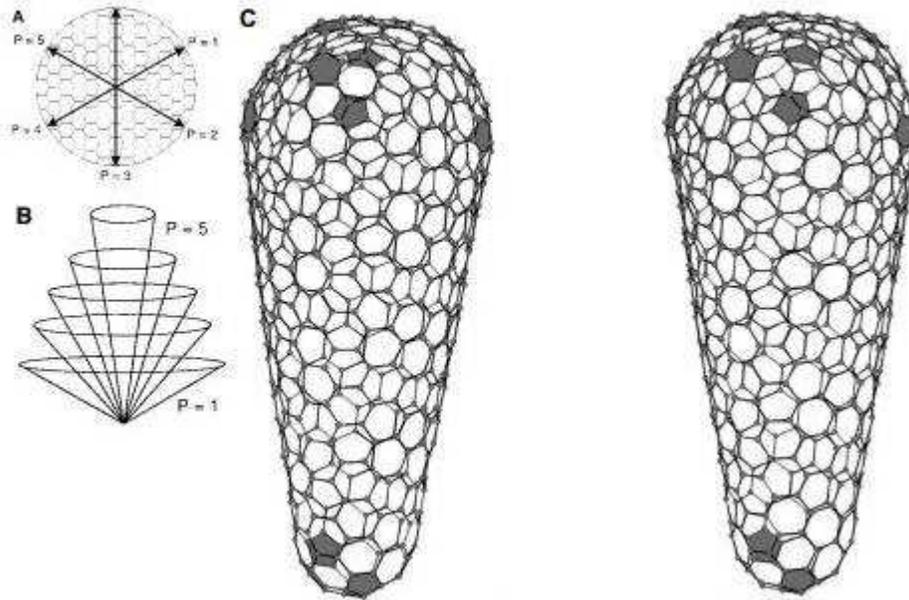
FIGURE 3.1. *A:* Angles of cones derived from conical hexagonal lattices. *B:* The five allowable angles resulting from the conical constructs described in (*A*). Allowable cone angles are $\theta = 112.9°(P = 1)$, $\theta = 83.6°(P = 2)$, $\theta = 60°(P = 3)$, $\theta = 38.9°(P = 4)$, $\theta = 19.2°(P = 5)$. *C:* 19.2° fullerene cone composed of 252 hexagons and 12 pentagons [24].

For convenience, we name these cones as (1,11)-, (2,10)-, (3,9)-, (4,8)-, (5,7)-patterns. It is also found that the viral core and most synthetic cones exhibited cone angles of approximately 19 degrees [24]. In other words, most HIV-1 cones are in the (5,7)-pattern, but (4,8)-cones and other unusual types of VLPs have also been observed in experiments [6, 8, 11].

Recently, the unique cone structure of the HIV-1 capsid has been intensively investigated. It is suggested in [13] that the asymmetry and quasi-equivalence exhibiting in conic and tubular capsids are related to the hinge between the C-terminal domain (CTD) and N-terminal domain (NTD) of the capsid protein. In [79], it is presented that the 12 pentamers introduce sharp declinations on the HIV-1 capsid. A line of hexamers connecting two declinations is presented to illustrate the continuously varying curvature. Dihedral angles along this line are also calculated. These dihedral angles between two subunits (hexamer/hexamer or hexamer/pentamer) are defined as bite angles. It is discussed in [26] that the bite angle

17

between adjacent hexamers vary from around 135° between two hexamers connected to the same pentamers at either end of the HIV-1 cone to around 180° in the more flat region in the middle of the cone. This is also examined in [26], especially the different angles between the subunits in CA pentamers or hexamers and the approximate pivot point for rotations. It is concluded in [61] that the rigid-body rotations around these assembly interfaces seem to be sufficient for explaining the curvature variation on the HIV-1 cone. It is also concluded in [84] that incorporation of CA pentamers into the surface hexameric lattice induces acute surface curvature. These studies deepen our understanding of viral capsid structure and viral assembly mechanism and motivate inhibitors for the formation of critical CA-CA interfaces in the capsid assembly [27].

There arises a need for a quantitative characterization of curvatures on virus capsids. This chapter and preprint [42] is the first, as to the authors' best knowledge, to address such a need. Furthermore, the concept of curvature concentration is introduced and this quantity is calculated for the narrow and broad ends of HIV-1 conical capsids. For the HIV (5,7)-, (4,8)- conical capsids, the results in this chapter show that the narrow end always has the highest curvature concentration, which is an indication that the narrow end is the weakest region on the HIV-1 capsid.

3.2. Mathematical Background: Curvature and Discrete Curvature

To understand curvatures on a surface, we need the concept of curvatures on a curve. In three-dimensional space, the curvature of a curve at a given point is a measure of how fast the curve changes its direction at that point. A formal mathematical definition for curvature

is given by

$$(2) \qquad \kappa = \left| \frac{d\vec{T}}{ds} \right|,$$

where $\vec{T}$ is the unit tangent vector and s is the arc length [62].

Continuous Curvatures on a Smooth Surface   Let M be a smooth surface and p be a point on M. The curvatures are characterized and quantified by the shape operator S, defined as

$$(3) \qquad S = \cup\{\pm S_p : p \in M\}, \qquad S_p(\vec{\mathbf{u}}) = -D_{\vec{\mathbf{u}}}\vec{N}(p),$$

where $\vec{N}$ is the unit normal vector field defined in an open neighborhood of the point p on the given surface M, $\vec{\mathbf{u}}$ is any tangent vector to M at p, namely, a tangent vector to a curve that passes through the point p but is entirely on the surface M. Note that $S_p(\vec{\mathbf{u}})$ defines the negative directional derivative of M at p along the vector $\vec{\mathbf{u}}$. Intuitively, $S_p(\vec{\mathbf{u}})$ explains how the surface M "curves" around the point p. The shape operator of M at p derived from $-\vec{\mathbf{u}}$ is $-S_p$, since it essentially reverses only the direction. Therefore, the shape operator of M is the union of all S_p at the given point p on M [62]. For a non-planar surface, the Gaussian curvature is given by the determinant of the operator S.

The principal curvatures of a surface at a given point are the two eigenvalues of the shape operator S discussed above. Denoted as κ_1 and κ_2, the principal curvatures measure the maximum and minimum bending of the surface at a given point, as shown in Figure 3.2. Mathematically, the Gaussian curvature is

$$(4) \qquad K = \kappa_1 \kappa_2,$$
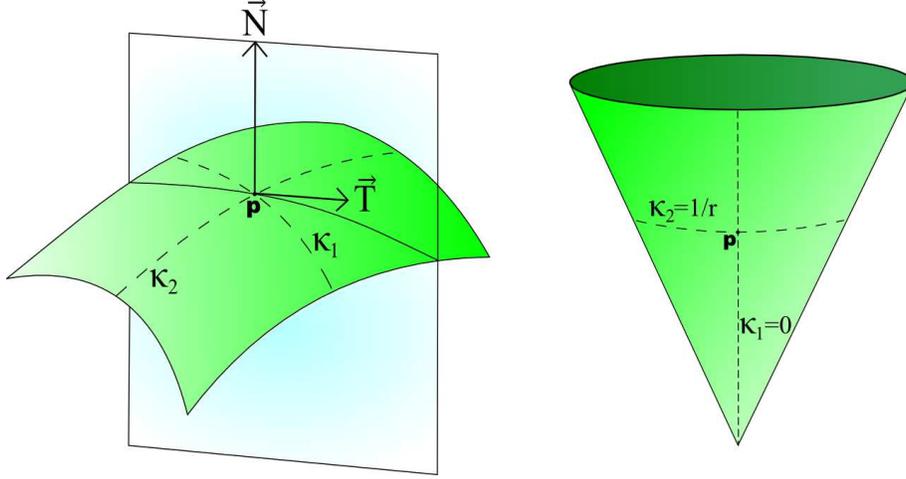
FIGURE 3.2. *Left*: There are two principal curvatures κ_1 and κ_2 at a point p on a smooth surface. *Right*: For a point on a conic surface, the minimum principal curvature is $\kappa_1 = 0$ whereas the maximum curvature is $\kappa_2 = 1/r$ with r being the radius of the circular section on which the point is located.

the product of the two principal curvatures. These principal curvatures can also be used to define other terms, e.g., the mean curvature

$$(5) \qquad\qquad H = \frac{1}{2}(\kappa_1 + \kappa_2).$$

For example, the Gaussian curvature of a plane at any point is zero, as the plane will not bend in any direction, that is, $\kappa_1 = \kappa_2 = 0$ and hence $K = \kappa_1 \kappa_2 = 0$. Cylinders and cones also have zero Gaussian curvature, since the minimum principal curvature $\kappa_1 = 0$ in each shape is zero. The case for a cone is illustrated in Figure 3.2 (*right*).

For a smooth surface M, the Gauss-Bonnet Theorem (see [62]) asserts that the integral of the Gaussian curvature on the surface is equal to 2π times the Euler characteristic

$$(6) \qquad\qquad \iint_M K(p)dp = 2\pi\chi_M.$$

Triangulation and the Euler Characteristic of a Polyhedral Surface   In the discrete setting, a smooth surface is replaced by a polyhedral surface. The concept of the discrete Gaussian curvature on a polyhedral surface is based on the triangulation of such a surface. Triangulation, in this case, is equivalent to the idea of tiling, see [45, 46], in which the tiles or subsections within one polygon are related by the theory of quasi-equivalence. For convenience, we assume each tile is equivalent, resulting in similar triangles within each hexagon and pentagon along the polyhedral surface. Since each polygon is cut into similar triangles, we call this process triangulation.

Consider a polyhedral surface as a set of polygons (in the space) joined together at their edges with varying dihedral angles. The most natural way to cut a polyhedral surface into subsections, is to divide the non-triangular shape into the least amount of similar triangles. For example, to triangulate a hexagon, one would cut it into six equal pieces or equilateral triangles, with a common vertex at the center of the hexagon, as shown in Figure 3.3 (*left*). Pentagons can be cut in a similar fashion, with five similar triangles having a common vertex
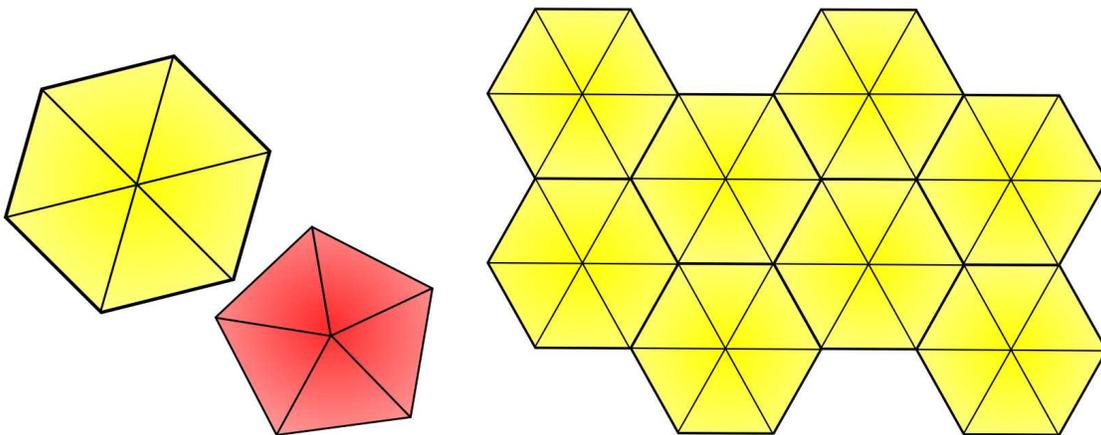
FIGURE 3.3. *Left*: A hexagon (in color yellow) and a pentagon (in color red) are each cut by similar triangles. *Right*: A hexameric lattice is triangulated by cutting each hexagon into six equilateral triangles, all sharing the vertex at the center.

at the center of each pentagon, see also Figure 3.3 (*left*). This allows the dihedral angles along the vertices of the hexagons to be equivalent and angles along the vertices of the pentagons to be equivalent.

Let M be a polyhedral surface. Denote by V the number of vertices, E the number of edges, and F the number of faces. The Euler characteristic of a closed polyhedral surface is given as

$$(7) \qquad \chi_M = V - E + F,$$

regardless of how the surface is bent. Any closed convex polyhedral surface has an Euler characteristic $\chi_M = 2$, see [46, 62]. This characteristic is independent of the choice of subsections, triangles, or tiles, since it is assumed that each polygon is a planar object.

Discrete Gaussian Curvature on a Polyhedral Surface In the discrete setting, the Gauss-Bonnet Theorem (see [62]) holds analogously

$$(8) \qquad \sum_{v \in D} K_v = 2\pi \chi_M; \qquad K_v = 2\pi - \sum_i \theta_i,$$

where D is a triangulated region on a given polyhedral surface M, v is a vertex in D, and θ_i are the interior angles at v. Each θ_i is an angle of a triangle adjoined at v. K_v is called the angle defect at v, which describes the discrete Gaussian curvature at the point.

Viral capsids are examples of closed convex polyhedral surfaces. Their Euler characteristic is $\chi = 2$, and so the sum of the discrete Gaussian curvatures is 4π. For icosahedral capsids, the curvature is distributed uniformly over the capsid due to its spherical-like shape. Non-icosahedral capsids do not share this property. A question then arises: how is the total curvature of 4π distributed throughout the capsid?

Many known viral capsids have a fullerene-like structure, which is a caged polyhedral surface composed of CA proteins grouped as hexamers and pentamers (6 and 5 proteins respectively). The number of hexamers varies for each capsid, depending on the size of the capsid. However, the number of pentamers always equals 12. This specific number of pentamers is required by the Euler Theorem to guarantee closure with no holes.

Some viral capsids have icosahedral or cylindrical symmetry. For the former, the pentamers are evenly spaced. For the latter, the pentamers are split into a (6,6)-pattern: 6 pentamers at the bottom, 6 pentamers at the top, and a large number of hexamers between the two ends.

HIV is unique in its cone shape, mostly in the (5,7) pattern, that is, 5 pentamers at the narrow end and 7 pentamers at the broad end. In [11], it is demonstrated that HIV-1 VLPs could have a larger cone angle resulting in a (4,8)-cone shape. Mathematically, there are five possible cones: (5,7), (4,8), (3,9), (2,10), (1,11), as shown in Equation (1). However, extreme distributions such as the (2,10)- and (1,11)-cones are rarely seen in the nature [28].

Literature suggests that the pentamers are isolated for stability, each surrounded by a ring of hexamers. Triangulation of the hexamers and pentamers in the most natural way leads to only two possible cases for curvature calculations, due to the consistent interior angles in each polygon.

Case 1. The first case occurs when a vertex v of a pentamer P is surrounded by two hexamers H, as shown in Figure 3.4. The triangulation produces six equilateral triangles for each hexamer and five similar triangles for the pentamer. At the shared vertex, there are a total of six interior angles, four from the hexamers and two from the pentamer triangulation.
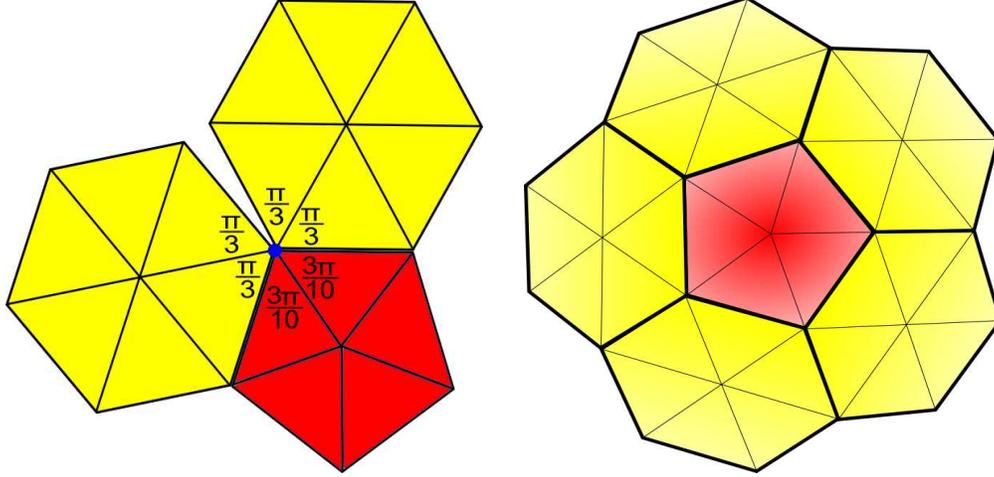
FIGURE 3.4. *Case 1*: A pentamer is surrounded by hexamers. Around a pentamer vertex, each triangle inside the pentamer has an interior angle $\theta = 3\pi/10$ whereas each triangle inside the hexamers has an interior angle $\theta = \pi/3$.

The interior angles of the triangles in the hexamers have values $\theta_i = \pi/3$ for $i = 1, 2, 3, 4$. The interior angles in the pentamers have values $\theta_i = 3\pi/10$ for $i = 5, 6$. Then the angle defect or the discrete Gaussian curvature at v is given by

$$(9) \qquad K_v(P) = 2\pi - \sum_i^6 \theta_i = 2\pi - 4\left(\frac{\pi}{3}\right) - 2\left(\frac{3\pi}{10}\right) = \frac{\pi}{15}.$$

Notice that the curvature calculation will be the same for each vertex of the pentamers P, since each vertex of P is also connected to two hexamers with the same triangulation. Each of the five vertices of P contributes a curvature of $\pi/15$, so the total discrete Gaussian curvature for the entire pentamer P is $\pi/3$.

Case 2. The second case occurs when a hexamer is surrounded by six other hexamers. Connecting the hexamers creates a flat surface (plane), as shown in Figure 3.3. Considering the same triangulation used in case one, the interior angles at each vertex v are given by

$\theta_i = 60° = \pi/3$ for $i = 1, 2, 3, 4, 5, 6$. The discrete Gaussian curvature at v is then

$$(10) \qquad K_v(H) = 2\pi - \sum_i^6 \theta_i = 2\pi - 6\frac{\pi}{3} = 0.$$

This is expected, as discussed in the examples in Section 3.2, since at least one principal curvature is zero along a plane and cone. Given exactly 12 pentamers and N_H hexamers, the total discrete Gaussian curvature of the caged cone is

$$(11) \qquad 12 * 5 * K_v(P) + N_H * 6 * K_v(H) = 4\pi,$$

which agrees with the discrete version of the Gauss-Bonnet Theorem (8).

For a cone-shaped capsid, the total curvature is independent of the number of hexamers in the cone and the curvature is nonzero only at pentamer positions. This implies the position of each pentamer is related to a location of high curvature on the capsid, introducing sharp declinations on the capsid as observed in [61, 79].

3.4. CURVATURE CONCENTRATIONS ON THE (5,7)-CONE

Consider the HIV-1 conical capsid as a polyhedral surface M, consisting of hexamers H and pentamers P. We assume that each vertex of P is surrounded by a pentamer and two hexamers.

The (5,7)-pattern has been widely reported in experimental observations. This means 5 pentamers at the narrow end of an HIV-1 capsid and 7 pentamers at the broad end of the capsid. This is the case when the capsid has a cone angle 19.2° [6, 8, 11, 24].

The total discrete curvature of the broad end (or the top) is $K_{7P} = 7\pi/3$, whereas the total discrete curvature of the narrow end (or the bottom) is $K_{5P} = 5\pi/3$. The middle region

of the cone is assumed to have only hexamers, so the total discrete curvature of this part is zero. This could be better understood when considering the middle region of the HIV-1 capsid as a right cone. The principal curvatures at any point on the right cone are given by $\kappa_1 = 0$ and $\kappa_2 = 1/r$ (r is the radius of the circular section on which the point is located), as shown in Figure 3.2. The Gaussian curvature at any point on a right cone is zero.

Another useful metric is the curvature concentration, i.e., curvature per area on a given surface.

General Formulas. To calculate the curvature concentration, the sum of the areas of the hexamers and pentamers in each region (the narrow end or the broad end) is considered. For the (5,7)-cone, there are 5 pentamers in the narrow end and 7 pentamers in the broad end. Assume that

(A1). These pentamers are isolated;

(A2). For each end, each vertex of a pentamer is surrounded by the pentamer and two hexamers;

(A3). There are H_n hexamers in one particular end (narrow end or broad end);

(A4). a is the side length of pentamers or hexamers.

Then direct mathematical calculations yield

$$(12) \qquad SA_{5P}(a, H_n) \;=\; a^2 \left(H_n \frac{3\sqrt{3}}{2} + \frac{5}{4}\sqrt{5(5 + 2\sqrt{5})} \right),$$

$$(13) \qquad SA_{7P}(a, H_n) \;=\; a^2 \left(H_n \frac{3\sqrt{3}}{2} + \frac{7}{4}\sqrt{5(5 + 2\sqrt{5})} \right).$$

Parameter Estimates. The overall height of a HIV-1 capsid was found to be 119±11nm [11], and the mean diameters of the broad and narrow ends are 56nm and 27nm, respectively [8]. Using the measurements reported in [9, 11, 61, 79], it is estimated that

- 4% of the hexamers lay at the narrow end;
- 36% of the hexamers are at the broad end;
- the remaining 60% are in the middle region;

with an average of 218 hexamers in each capsid.

Recent cryo-EM results [84] indicate a larger diameter at the narrow end, with an estimate of 6% of the hexamers at the narrow end, 33% at the broad end, and 61% in the middle region, with an average of 216 hexamers. The diameter of a hexameric unit was found to be approximately 9.8nm with a 3.2nm spacing between units in VLPs [11]. For the fullerene-like structure model, it is assumed that each subunit (pentamer or hexamer) has a side length 6.5nm.

Calculations. For the narrow end of a HIV-1 capsid, the surface area is calculated using Equation (13), with a range of 216 to 218 hexamers

$$(14) \qquad SA_{5P}(6.5, H_n) \approx 1,460 \pm 219\text{nm}^2.$$

Similarly, the surface area of the broad end is estimated as

$$(15) \qquad SA_{7P}(6.5, H_n) \approx 8,796 \pm 384\text{nm}^2.$$

The curvature concentration is defined as the ratio of the discrete Gaussian curvature per surface area in each region. For the broad and narrow ends, we have curvature concentrations

$$(16) \qquad C_{K_{7P}} \approx 8.34 \times 10^{-4}, \qquad C_{K_{5P}} \approx 3.59 \times 10^{-3},$$

which show that $C_{K_{5P}}$ is about five times higher than $C_{K_{7P}}$. That is, the curvature concentration at the narrow end is about five times higher than that at the broad end.

Curvatures on (4,8)- and Other Type Cones As discussed in the previous sections, there are five possible cone angles for a cone composed of only hexamers and pentamers, according to the Euler formula. Although 19.2° is the most common cone angle for HIV-1 cores, larger cone angles between 30° and 40° have also been reported in experimental data [6, 11, 24]. This implies that HIV-1 cores could form into a cone with $4P$ at the narrow end and $8P$ at the broad end. Both (4,8)- and (3,9)-cones have been seen in graphite nanocones, although it is thought (2,10)- and (1,11)-cones will not form, due to the high strain at the narrow end [28].

Based on the same assumptions (A1)-(A4) listed in the previous subsection, the surface areas of the narrow ($4P$) and broad ($8P$) ends of a conic capsid are estimated by adding the surface areas of hexamers and pentamers in each region as follows,

$$(17) \qquad SA_{4P}(a, H_n) \;=\; a^2 \left(H_n \frac{3\sqrt{3}}{2} + \sqrt{5(5 + 2\sqrt{5})} \right),$$

$$(18) \qquad SA_{8P}(a, H_n) \;=\; a^2 \left(H_n \frac{3\sqrt{3}}{2} + 2\sqrt{5(5 + 2\sqrt{5})} \right),$$

where again H_n is the number of hexamers in that region and a is the side length of the pentamers or hexamers.

For the (4,8)-cone, based on the above formulas, the approximations for the surface areas are as follows

(19) $$SA_{8P} \approx 8,485\text{nm}^2, \qquad SA_{4P} \approx 1,059\text{nm}^2.$$

Then the curvature concentrations of the broad and narrow ends are respectively

(20) $$C_{K_{8P}} \approx 9.87 \times 10^{-4}, \qquad C_{K_{4P}} \approx 4 \times 10^{-3}.$$

Therefore, $C_{K_{4P}}$ is about four times higher than $C_{K_{8P}}$.

Similar calculations can be performed for curvature concentrations of other cone types. As the number of P (pentamers) in the narrow end decreases, the surface area and total curvature for that region will also decrease. As the number of P in the broad end increases, the surface area and total curvature of that region will increase. This implies that the curvature concentration at the narrow end will always be greater than the concentration at the broad end, making the narrow end of the conic capsid the weakest region of the capsid, regardless of the cone angle.

Relation to Declination. Existing work [25, 44, 61, 79, 81, 82] suggest that the pentamers introduce sharp declinations on the HIV-1 capsid, as shown in Figure 3.5. This agrees with our curvature calculations. These sharp declinations occur because the pentamers are the sources of curvature on the closed capsid. In the middle region of the HIV-1 capsid, the dihedral angles (angles between the hexamer-hexamer planes) vary but are close to 180°, implying there is little to no curvature in that region.
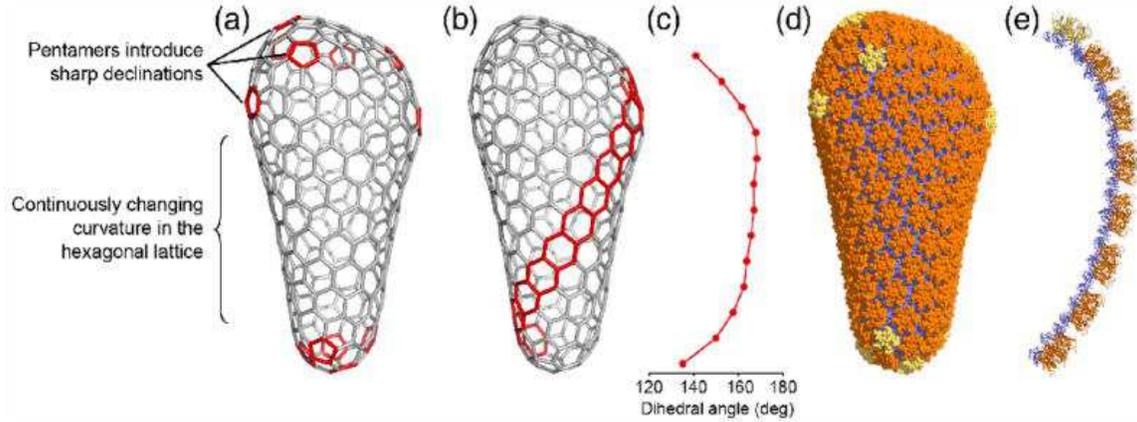
FIGURE 3.5. On a caged (5,7)-cone, the 12 pentamers (red) introduce sharp declinations to close the capsid. The graph of the dihedral angles between the hexamers shows measurements close to 180°, implying similarity between the middle region and a rolled plane [79] (with permission from Elsevier for reuse of the figure).
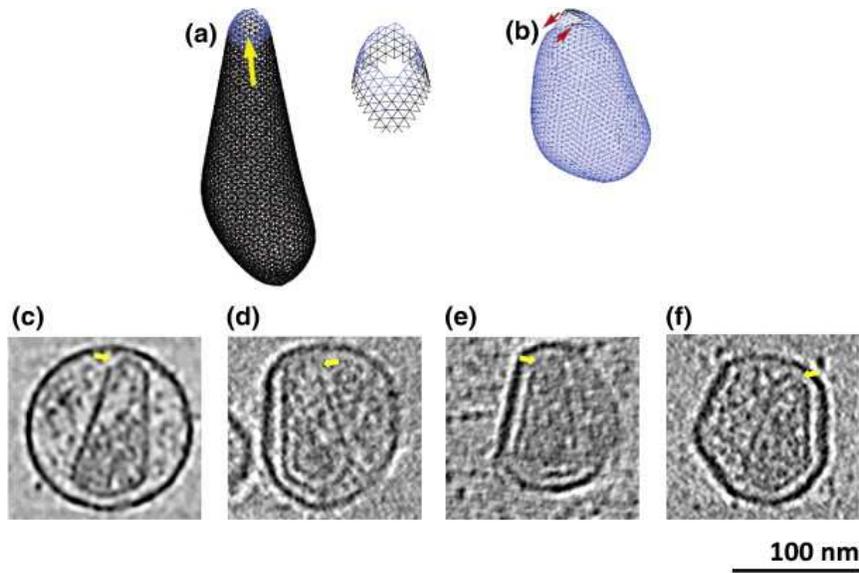
FIGURE 3.6. Modeling and experimental results in [80] show that the HIV-1 capsid narrow end might not close, if conditions are unfavorable [80] (with permission from Elsevier for reuse of the figure).

3.5. DISCUSSION

Note that the curvature at a point depends on the interior angles of the triangles meeting there. A dihedral angle in this case is the angle between two triangles in the triangulation of pentamers or hexamers. By the definition of the discrete Gaussian curvature, for the

30

curvature to be 0 at one point, the dihedral angle should be 180° (equivalent to $\kappa_1 = 0$) along one principal direction. For a cone, this direction usually occurs along the edge straight up to the broad end, as shown by the κ_1 dotted line in Figure 3.2 (*right*).

Conjecture: HIV-1 Narrow End Closed Last But Opened First. The above curvature calculations demonstrate that the narrow end of an HIV-1 capsid has the highest curvature concentration and hence is the weakest part on the capsid. Recent experimental and modeling studies show that the HIV-1 capsid narrow end might not close, if conditions are unfavorable [80], as shown in Figure 3.6. These studies lead to our conjecture: the narrow end of the HIV-1 capsid might be closed last during viral maturation but opened first during entry into a host cell.

CHAPTER 4

GENERATING VECTORS FOR VIRAL CAPSID LATTICE

STRUCTURES

It is well known that CA hexamers and pentamers form a lattice structure that folds into a viral capsid. It has been observed that a viral capsid takes an icosahedral, tubular, conical or irregular shape [17, 43, 54].

4.1. LATTICE STRUCTURES OF VIRAL CAPSIDS

The icosahedral viral capsid has been extensively studied due to the highly symmetric nature. This is the preferred geometry for viral capsids, since the symmetry allows 60 CA proteins or 12 pentamers to be placed on the surface in an equivalent manner. There are many viruses whose capsids have more than 60 CA proteins, in fact around thousands of proteins. In these cases, not all of the subunits (CA proteins) can be placed in equivalent positions. Caspar and Klug's theory of quasi-equivalence [14] addresses this issue by classifying icosahedral shells by similar protein neighborhoods rather than subunits. The Caspar-Klug quasi-equivalence theory allows capsids with multiples of 60 subunits, indicated by the triangulation (T) number, to form with icosahedral symmetry. The quasi-equivalence is demonstrated in the experiment results reported in [84], see Figure 4.1 here in this dissertation.

There are also many viruses that have tubular or lozenge-like capsids. Mature HIV-1 cores have cone-shaped capsids [6, 24]. Unlike the icosahedral viral capsids, the structures of tubular and conical capsids are not yet fully understood.
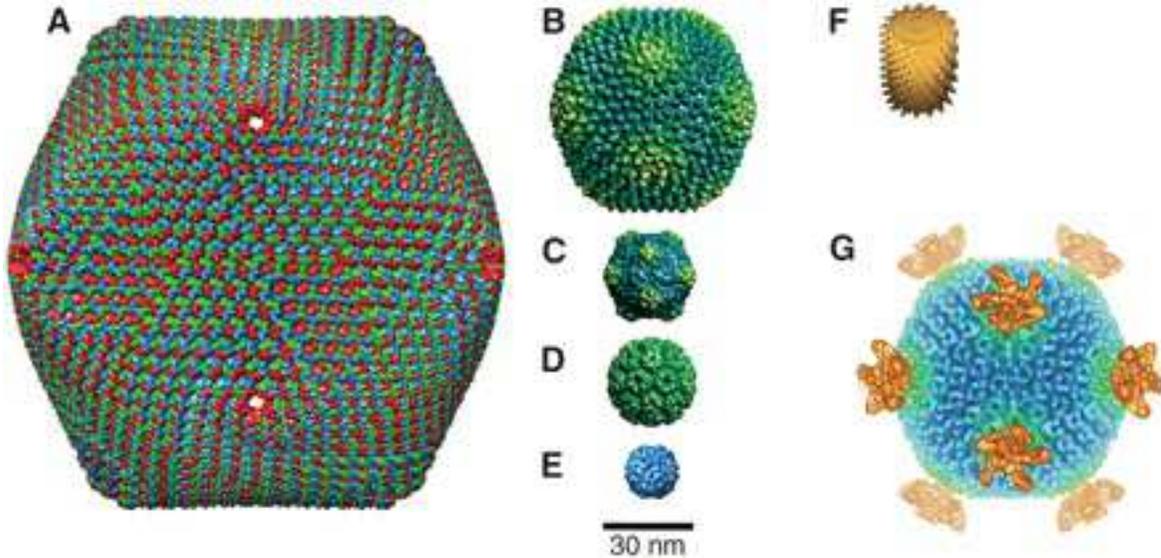
FIGURE 4.1. Illustrations of viral capsids, including both icosahedral and helical viruses [17].

There have been models for the lattice structures of tubular and conical capsids, see [43, 54] and references therein. However, there are flaws in the Nguyen models [54], where the tubular model is too restricted and cannot produce the wide arrange of tubular viral capsid and missed the requirement on the cone height which could result in incomplete cones if the model is inappropriately applied.

In this chapter, we propose new models for tubular and conical capsids in a unified fashion based on an extension of the Caspar-Klug quasi-equivalence theory. The new models are easier than the existing models. When applied to the HIV-1 (5,7)-cone (5 pentagons in the narrow end and 7 pentagons in the broad end), the capsid properties derived from our models show good agreement with published experimental data. This demonstrates the correctness and usefulness of the new models.

The rest of this chapter is organized as follows. Section 4.2 briefly reviews the concepts of the T-number and generating vector for the icosahedral viral capsid. Section 4.3 presents

a mathematical model for tubular viral capsids using two generating vectors and three pa-
rameters. Section 4.4 presents a model for conical viral capsids that uses three generating
vectors and four parameters. Further details on the (5,7)- and (4,8)-cones are examined. Ac-
cording to the Euler theorem, there are 3 other possible cone angles for a hexagonal lattice,
but models for the (1,11)-,(2,10)-, and (3,9)-cones are not investigated, since they are rarely
seen in nature without overlapping in a spiral fashion [28]. Section 4.5 compares modeling
results to published experimental data on the HIV-1 conical capsid. Section 4.6 concludes
the chapter with some remarks.

4.2. Generating Vector and T-number for Icosahedral Viral Capsids

This section briefly reviews the concepts of the T-number and generating vector for an
icosahedral viral capsid. This will be helpful for understanding the models for tubular and
conical viral capsids to be discussed in the following sections.

About half of the virus species are found to have an icosahedral capsid [43]. The geometric
structure (symmetry and periodicity) of icosahedral capsids can be well characterized by the
Caspar-Klug quasi-equivalence theory [14].

By the Euler theorem, for a convex polyhedron made of hexagons and pentagons, there
are exactly 12 pentagons. When these 12 pentagons are evenly distributed, an icosahedron
forms, which can be circumscribed into a sphere.

To understand the concepts of the generating vector and T-number, we start with a flat
hexagonal lattice consisting entirely of identical hexagons. As shown in Figures 4.2 and 4.3,
we choose the center of one hexagon as the origin and set the lengths of the basis vectors
$\vec{a}_1, \vec{a}_2$ as 1. It is obvious that the angle between $\vec{a}_1$ and $\vec{a}_2$ is 60° and hence their inner
product (dot product) is $\langle \vec{a}_1, \vec{a}_2 \rangle = \frac{1}{2}$. The generating vector, as shown in Figures 4.2 and

4.3, is defined as a linear combination of the two basis vectors

$$(21) \qquad \vec{A} = h\vec{a}_1 + k\vec{a}_2,$$

where h, k are non-negative numbers (but not both zero).
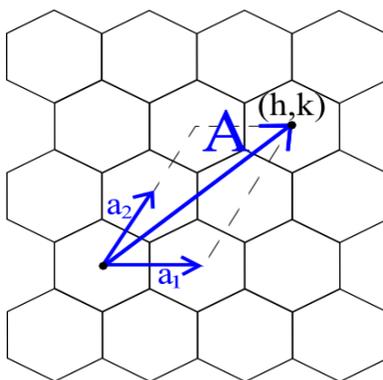
FIGURE 4.2. Two basis vectors $\vec{a}_1, \vec{a}_2$ and one generating vector $\vec{A}$ for a hexagonal lattice. In this illustration, $h = 1, k = 2$ and hence $T = h^2 + hk + k^2 = 7$. The two hexagons where the starting and ending points of the generating vector $\vec{A}$ reside will be replaced by two pentagons when the lattice is folded into an icosahedron.
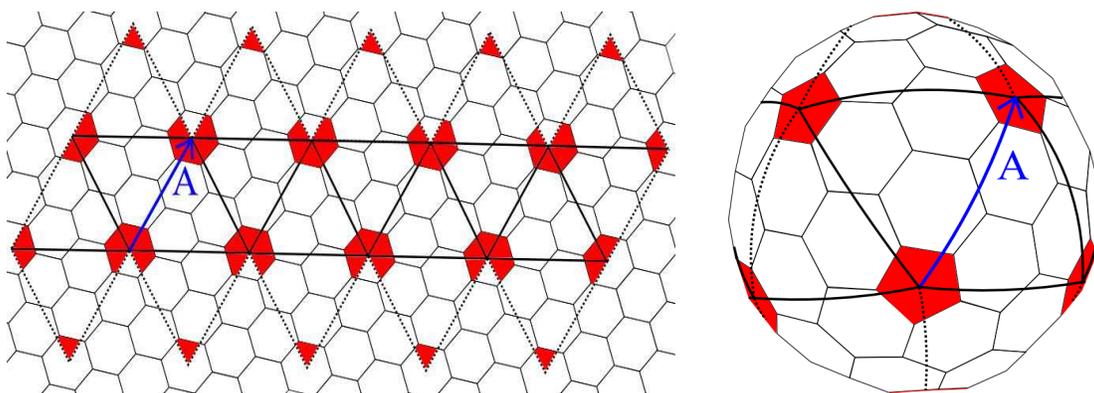
FIGURE 4.3. *Left*: A lattice with $(h, k) = (1, 2)$ and $T = h^2 + hk + k^2 = 7$. The dotted lines indicate where to cut the lattice to fold it. *Right*: The icosahedral capsid obtained from folding the lattice shown in the left panel. Pentagons are shown in red. An example for $T = 7$ is the widely studied bacteriophage HK97 [48].

Therefore, we have

$$(22) \quad |\vec{A}|^2 = \langle h\vec{a}_1 + k\vec{a}_2, h\vec{a}_1 + k\vec{a}_2 \rangle = h^2 \langle \vec{a}_1, \vec{a}_1 \rangle + 2hk \langle \vec{a}_1, \vec{a}_2 \rangle + k^2 \langle \vec{a}_2, \vec{a}_2 \rangle$$

$$= h^2 + hk + k^2 =: T,$$

which is the so-called T-number. Geometrically, the T-number can be understood as the squared length of each triangle edge in the construction of the icosahedron. This relates (h, k) to the area of a single triangle by the formula: Area $= \frac{\sqrt{3}}{4}T$.

There are 20 equilateral triangles used in the construction of an icosahedron, placed symmetrically on a flat hexagonal lattice, as shown in Figure 4.3. The triangle size depends on the T-number, with a varying number of hexagons within. Each vertex of a triangle lands at the center of a hexagon, which is the very position of a pentagon when folded in three dimensions. The pentagon is formed by cutting a 60° wedge from a hexagon then adjoining the two cut edges. This creates a convex five-sided polygon, whose center is no longer on the hexagonal plane. Clearly, the T-number measures the distance squared between the centers of two nearby pentagons.

Figure 4.4 represents another commonly used approach for illustrating the icosahedral viral capsid. See also Figure 1 of [54]. The lattice structure shown in Figure 4.4 has $(h, k) = (2, 2)$ and hence a triangulation number of $T = h^2 + hk + k^2 = 12$. Each triangle side has length $\sqrt{12}$, and the vertices lie at the centers of hexagons that will be replaced by pentagons when folded.

In structural virology, icosahedral capsids are usually described by $T(h, k)$. However, there is no guaranteed uniqueness for $T \geq 49$. For example, $(7, 0)$ and $(5, 3)$ both give $T = 49$, see [43]. To classify these virus capsids uniquely, Caspar and Klug [14] proposed a
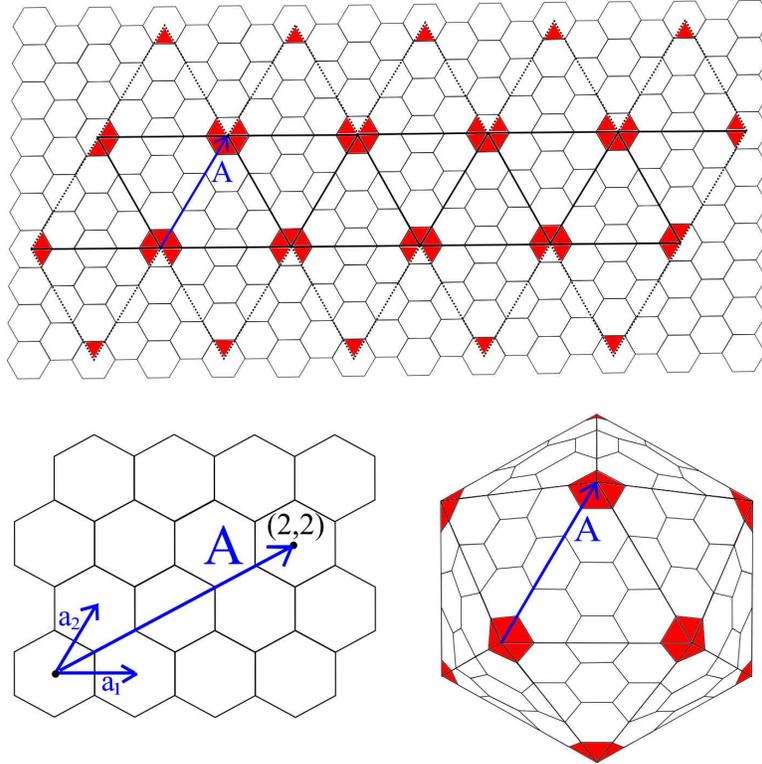
FIGURE 4.4. Another commonly used approach for illustrating icosahedral viral capsids: *Top*: The centers of the pentagons form a coarse triangular mesh; *Bottom Left*: Two basis vectors and one generating vector with $(h, k) = (2, 2)$ and $T = h^2 + hk + k^2 = 12$; *Bottom Right*: The lattice folds into an icosahedron with the pentagon centers being the triangle vertices. However, the folded pentagons and the flat triangles shown in (*bottom right*) need careful reading. Actually all pentagons are planar objects. The coarse triangles are used to locate the pentagons. The hexagons do not lie on the virtual triangles.

reorganization of the T-number, in terms of the P classes. Any class with $P > 3$ (starting with $T = 7$) is skewed so that (h, k) generates a chiral structure mirrored by that created with (k, h) [14, 43]. Both chiral structures can be produced from this model by recreating the folding from a mirrored lattice. Further details are excluded from this chapter so we may focus on the construction of the tube and conical capsids.

4.3. Generating Vectors for Tubular Viral Capsids

A tubular (spherocylinder) viral capsid has been observed for Cowpea Chlorotic Mottle Virus and Alfalfa Mosaic Virus, among others. As reported in [6, 11], HIV-1 cores could also exhibit a tube-like capsid.

A tubular viral capsid consists of only CA protein hexamers and pentamers. The Euler theorem guarantees exactly 12 pentamers on the capsid, assuming it is a convex polyhedron. The tubular structure can be considered as cutting an icosahedron in half and extending the middle region by a hexagonal cylinder. Each end cap is a truncated icosahedron with exactly 6 pentagons and a varying number of hexagons, determined by the T-number.

The tubular model also follows the Caspar-Klug quasi-equivalence theory. However, describing the lattice structure of a tubular viral capsid needs two generating vectors: one to describe the equal distance between the pentagons and the other for the varying height.

First, we define a vector $\vec{A} = h\vec{a_1} + k\vec{a_2}$ for the folding of the two end caps. This is the same as for the construction of an icosahedron. The two caps will be displaced some distance from each other, in the direction perpendicular to $\vec{A}$ by a new generating vector $\vec{B}$. The vector $\vec{B} = h\vec{b_1} + k\vec{b_2}$ is so defined that its basis vectors $\vec{b_1}$ and $\vec{b_2}$ are respectively orthogonal to $\vec{a_1}$ and $\vec{a_2}$. Specifically, we have, as shown in Figure 4.5 (*left*),

$$(23) \qquad\qquad\qquad\qquad \vec{b_1} = \vec{a_1} - 2\vec{a_2},$$

$$(24) \qquad\qquad\qquad\qquad \vec{b_2} = 2\vec{a_1} - \vec{a_2}.$$

The folding template for a tubular viral capsid is shown in Figure 4.5. This is similar to the template for an icosahedron. The difference is exhibited in the ten triangles located at
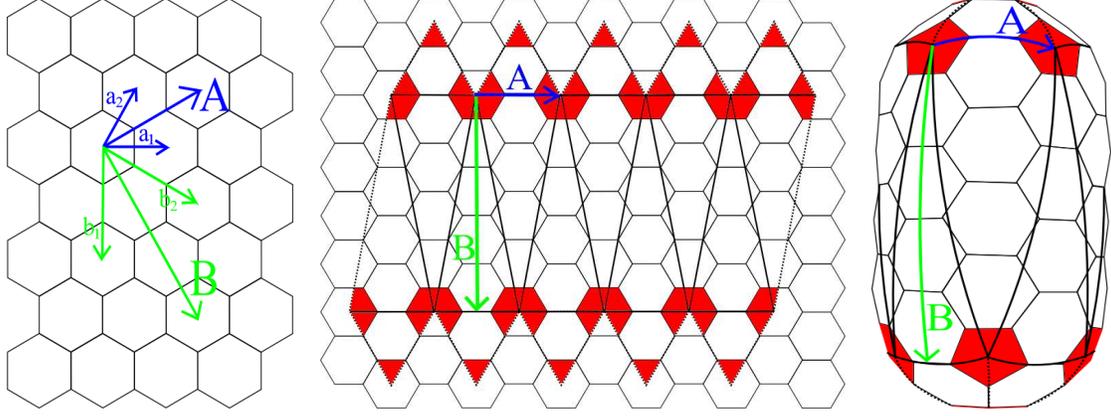
FIGURE 4.5. *Left*: Basis vectors and generating vectors. *Middle*: A lattice structure with $(h, k) = (1, 1), T = h^2 + hk + k^2 = 3, \gamma = \frac{9}{6}$. The dotted lines indicate where to cut the lattice to fold it. Pentagons are shown in red. *Right*: The lattice folds into a tubular capsid.

the center of the folding. They are no longer equilateral. This is due to the displacement of the end caps, creating an elongated hexagonal tube. Triangle vertices still lie at the centers of the hexagons that will be replaced by pentagons when folded in three dimensions.

In summary, the tubular folding template is constructed by two generating vectors

$$(25) \qquad \vec{A} = h\vec{a_1} + k\vec{a_2},$$

$$(26) \qquad \vec{B} = \gamma(h\vec{b_1} + k\vec{b_2}),$$

where

$$(27) \qquad \gamma = r\frac{\gcd(h, k)}{2T},$$

and r is an integer. Here $\gcd(h, k)$ is the greatest common divisor of h and k. When $\gamma = 1/2$, this model reproduces an icosahedron described in the previous section.

The derivation for γ is intuitive. The only limitation on the height of a tubular capsid comes from the construction requirement that $\vec{A}$ must start and end at the center of a

hexagon. There are multiple positions along the direction of $\vec{A}$ that satisfy this requirement. If h and k are not co-prime, then these positions do not create unique structures. If fact, similar positions occur along a triangle edge for every length of $\frac{\sqrt{T}}{\gcd(h,k)}$. Moving in the direction perpendicular to $\vec{A}$, there is a hexagon meeting the requirements for every integer increment of $\frac{\sqrt{3}}{2}\frac{\gcd(h,k)}{\sqrt{T}}$. Relating this to $|\vec{B}|$, we have

$$(28) \qquad\qquad |\vec{B}| = \gamma\,\sqrt{3T} = r\,\frac{\sqrt{3}\,\gcd(h,k)}{2\sqrt{T}},$$

and hence

$$(29) \qquad\qquad \gamma = r\,\frac{\gcd(h,k)}{2T}.$$

The model proposed in this chapter is similar to but more general than the tubular model introduced in [54]. In [54], the two generating vectors are defined as

$$(30) \qquad\qquad \vec{A} \;=\; n(h\vec{a_1} + k\vec{a_2}),$$

$$(31) \qquad\qquad \vec{B} \;=\; m(h\vec{b_1} + k\vec{b_2}),$$

where (h,k) are the same as those for the icosahedral capsid, (n,m) are two non-negative integers, and b_1, b_2 are similarly defined (but have opposite directions) as in Equations (23) and (24). For this model, scaling $h\vec{a_1} + k\vec{a_2}$ by a constant n is somewhat unnecessary, since all variations can be accounted for by varying h and k. Restricting the scaling constant in $\vec{B}$ to an integer excludes the model from covering several types of virus capsids, for instance, the occasional tubular shape of HIV-1 [6] and the bacteriophage $\phi 29$ [12, 71]. To use the model in [54] for creating the bacteriophage $\phi 29$ capsid shown Figure 4.6 (c), n must be
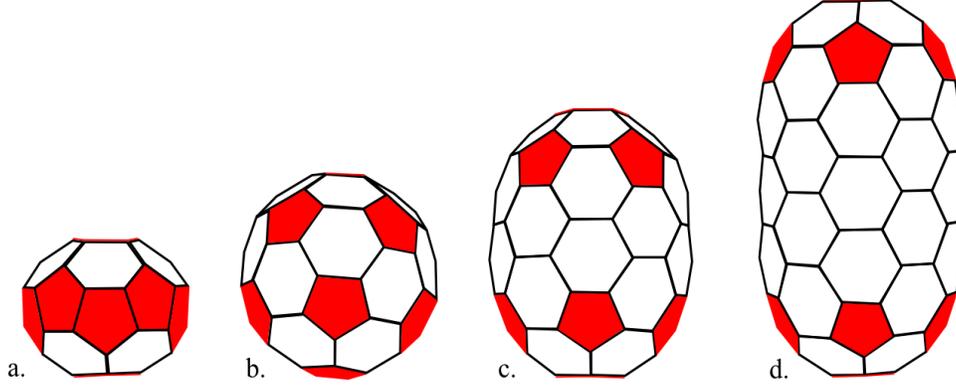
FIGURE 4.6. Tubular (spherocylindrical) capsid with $T = 3$ and varying γ. Pentagons are shown in red. From left to right: a: $\gamma = \frac{1}{6}$, unlikely to occur in nature, but it is the smallest possible tube that can be created by this model. b: $\gamma = \frac{3}{6}$, examples include Cowpea Chlorotic Mottle Virus or Norwalk Virus [12]. c: $\gamma = \frac{5}{6}$, seen in bacteriophage $\phi 29$ [12, 71]. d: $\gamma = \frac{8}{6}$, seen in Alfalfa Mosaic Virus [43].

defined as $n = 1$ due to the end caps, and $m = \frac{5}{3}$ due to the height. Similarly, to use the model is [54] for creating the Alfalfa Mosaic Virus capsid shown Figure 4.6 (d), n must be defined as $n = 1$ due to the end caps, and $m = \frac{8}{3}$ due to the height. Neither of these are valid since both n and m are required to be integers in the model described in [54].

4.4. GENERATING VECTORS FOR CONICAL CAPSIDS

A simple cone can be produced by rolling a section of a sheet around its apex and joining the two open sides. However, a cone created with a hexagonal lattice will not have infinitely many cone angles. With the hexagonal lattice, the hexagon/pentagon units along the closure line must match.

The Euler theorem implies that there are five possible cone angles for a hexagonal lattice as shown below

(32) $$\sin(\theta/2) = 1 - P/6,$$

where θ is the cone angle and P is the number of pentagons located in the narrow end of the cone. The five angle values (and the corresponding P values) are $\theta = 112.9°(P = 1)$, $\theta = 83.6°(P = 2)$, $\theta = 60°(P = 3)$, $\theta = 38.9°(P = 4)$, $\theta = 19.2°(P = 5)$, see [24]. For convenience, we name these cones as (1,11)-, (2,10)-, (3,9)-, (4,8)-, (5,7)-cones. In the notations for $(i, j)-$, i is the number of pentagons in the narrow region, j is the number of pentagons in the broad region, and $i + j = 12$. Most HIV-1 cones are in the (5,7)-pattern, but (4,8)-cones have also been observed in experiments [6, 9, 11].

As far as what has been discovered, HIV-1 is the only virus with a conical capsid, although similar phenomena have been observed in carbon nanocones [67].

Generating Vectors for the (5,7)-Cone   A (5,7)-cone has the smallest allowed cone angle formed from a hexagonal lattice. In this subsection, we consider generating vectors or a folding template for the (5,7)-cone.

For consistency, we consider a generating vector $h\vec{a_1} + k\vec{a_2}$ scaled by two non-equal integers to generate the triangles needed for the two end caps of the cone. Without loss of generality, we assume $\alpha < \beta$ are such two integers. Let $\vec{A} = \alpha(h\vec{a_1} + k\vec{a_2})$ generate the five smaller equilateral triangles needed to fold the 5 pentamers in the narrow end. A parallel vector $\vec{B} = \beta(h\vec{a_1} + k\vec{a_2})$ is used to generate the six larger identical triangles needed to create the 7 pentamers in the broad end (Figure 4.7).

These equilateral triangles are determined by the T-number and the two additional constants α and β. These triangle vertices have a slightly different meaning than those in the icosahedral and tubular models. In the icosahedral and tubular models, triangle vertices lie
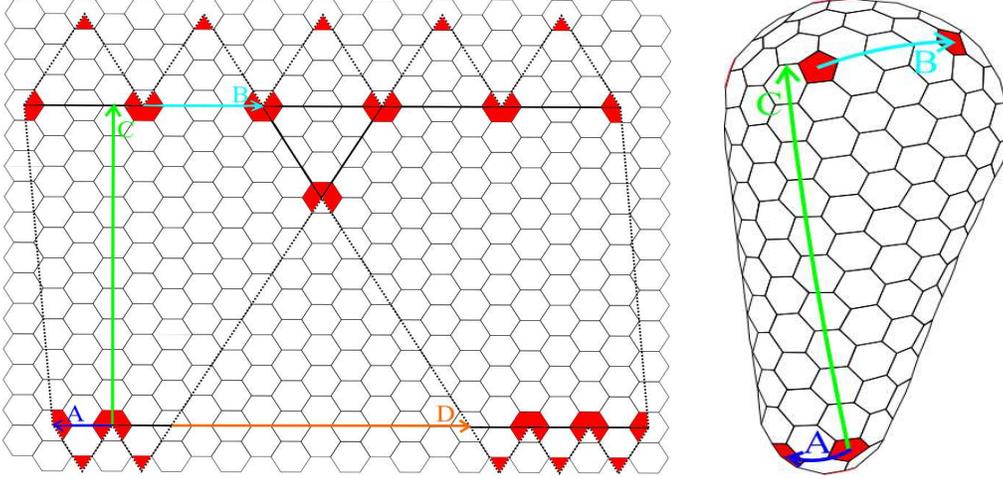
FIGURE 4.7. *Left*: Three generating vectors are needed for folding a lattice into a conical capsid: $\vec{A}$ generates the side length of the triangles in the narrow end of the cone; $\vec{B}$ generates the side length for the triangles in the broad end; $\vec{C}$ ensures the unique height needed for closure. The dotted lines indicate where to cut the lattice to fold it into a 3-dimensional cone. Pentagons are shown in red. *Right*: The lattice on the left panel with $(h,k) = (1,1)$, $T = 3$, $(\alpha,\beta) = (1,2)$ folds into a (5,7)-cone.

at the centers of hexagons that will be replaced by pentagons when folded in three dimensions. For the conical capsid, triangle vertices are not necessarily located in pentagons. To clarify, pentagons are shown in red in Figure 4.7.

However, more information is needed to form a closed (5,7)-cone. Since the generating vectors are defined on a hexagonal lattice, the model must be positioned correctly to ensure only hexagons and pentagons are produced during the folding. This occurs when the outer closure lines in the middle region of the lattice are parallel, as shown by the dotted lines in the far left and far right of Figure 4.7 (*left*). Without this requirement, the folding cannot close correctly. The parallel lines ensure hexamers are matched along the closure line, leading to a 0° declination that produces perfect hexamers in the middle region of the cone. This necessary addition to the Nguyen model [54] is further explained in the Discussion section.

To ensure parallel outer edges, a third generating vector, $\vec{C}$, is needed. Let $\vec{C} = q_0(h\vec{b_1} + k\vec{b_2})$, where $\vec{b_1}$ and $\vec{b_2}$ are the vectors defined in Equations (23) and (24). Clearly, $\vec{C}$ is perpendicular to both $\vec{A}$ and $\vec{B}$.

We introduce a vector $\vec{D} = \rho_0(h\vec{a_1} + k\vec{b_2})$ to generate the distance between the two neighboring triangles in the bottom of the lattice that do not share a common vertex, as shown orange in Figure 4.7. This shall allow a parallelogram to be formed in the middle of the cone. The top and bottom sides of the parallelogram should have the same length, that is,

$$5\beta\sqrt{T} = 5\alpha\sqrt{T} + \rho_0\sqrt{T}, \tag{33}$$

and hence

$$\rho_0 = 5(\beta - \alpha). \tag{34}$$

Note that the vector length $|\vec{C}|$ can be determined in two ways: either from $q_0|h\vec{b_1}+k\vec{b_2}| = q_0\sqrt{3T}$ or from the sum of the height of the triangles generated by $\vec{B}$ and $\vec{D}$. Setting them equal yields

$$q_0\sqrt{3T} = \frac{\sqrt{3}}{2}\left(\beta\sqrt{T} + 5(\beta - \alpha)\sqrt{T}\right), \tag{35}$$

and hence

$$q_0 = \frac{1}{2}(6\beta - 5\alpha). \tag{36}$$

In summary, three generating vectors are needed for the folding template of a (5,7)-cone:

$$(37) \qquad \vec{A} = \alpha(h\vec{a_1} + k\vec{a_2}),$$

$$(38) \qquad \vec{B} = \beta(h\vec{a_1} + k\vec{a_2}),$$

$$(39) \qquad \vec{C} = \frac{1}{2}(6\beta - 5\alpha)(h\vec{b_1} + k\vec{b_2}),$$

where $\alpha < \beta$ are two non-negative integers and $\vec{b_1}$ and $\vec{b_2}$ are defined in Equations (23) and (24). Note that when $\alpha = \beta$, this cone model reproduces an icosahedron, with $T = \alpha^2 (h^2 + hk + k^2)$.

Following the lattice construction, the surface area of the conical capsid is calculated by summing the areas of the three regions (the broad end, the middle region, the narrow end) on the lattice used in the folding.

$$(40) \qquad SA = 5\left(\frac{\sqrt{3}}{4}|\vec{B}|^2\right) + \left(5|\vec{B}||\vec{C}| - \frac{\sqrt{3}}{4}|\vec{D}|^2\right) + 5\left(\frac{\sqrt{3}}{4}|\vec{A}|^2\right) (\text{unit}^2)$$
$$=: 15T\sqrt{3}\,(2\beta^2 - \alpha^2)\,a^2\,(\text{nm}^2),$$

where a multiple of $\sqrt{3}a$ gives a conversion between (unit) and (nm), and a is the side length of a single hexamer in nanometers.

It is assumed that there are exactly six CA proteins in a hexamer with area $3a^2\sqrt{3}/2(\text{nm}^2)$, given the side length a. Therefore, there are $4\sqrt{3}/3a^2$ CA proteins per (nm^2). The total number of CA proteins for this conical model is given by the surface area multiplied by the number of CA proteins per area, or

$$(41) \qquad CA = 60T\left(2\beta^2 - \alpha^2\right).$$

Subtracting the 60 proteins used to create the pentagons and dividing by 6 (CA proteins per hexamer) yields the total number of hexamers, N_H, on the (5,7)-conical capsid:

$$(42) \qquad\qquad N_H = 10T\left(2\beta^2 - \alpha^2\right) - 10.$$

Generating Vectors for the (4,8)-Cone and Other Types Cones Literature [8, 11, 24, 77] showed evidence that the cores of HIV-1 virus like particles (VLPs) have cone angles between 30° and 50°, which is an indication of existence of the (4,8)-conical capsid.

To maintain symmetry, the center of the narrow end should be a hexagon instead of a pentagon as seen in the previous (sub)sections. This way, four pentagons can be evenly spaced around it, The center of the narrow end is created by joining the smaller triangle tips on the lattice. For the icosahedral, tubular, and (5,7)-conical models, the narrow end has 5 triangles. When five 60° triangle tips are joined, an angular defect is produced. This results in the formation of a pentagon. However for a (4,8)-cone, six (rather than five) small triangles are used in the narrow end. When six 60° triangle tips are joined, there is no angular defect. This results in the formation of a hexagon. Similar to the (5,7)-conical model, triangle vertices are no longer guaranteed positions of pentagons when folded in three dimensions. This results in a total of 4 pentagons being created in the narrow end.

The broad end is very similar to that of the (5,7)-cone. The only difference is the addition of a 7th equilateral triangle. This allows eight pentagons to be formed in this region. Pentagon positions are shown in red in Figure 4.8.

Similar to the (5,7)-cone model, the (4,8)-cone model requires three generating vectors to ensure proper closure. Vector $\vec{A}$ will determine the size of the smaller triangles located in the narrow end, a parallel vector $\vec{B}$ will determine the size of the larger triangles located

in the broad end, and a perpendicular vector $\vec{C}$ determines the unique height, which insures that the outer closure lines in the middle region of the cone are parallel. This guarantees proper closure of the cone with only hexagons and pentagons.

We now explain this in detail. To enforce parallel outer edges, we define $\vec{C} = q_1(h\vec{b_1} + k\vec{b_2})$. The constant $q_1 = \frac{1}{4}(7\beta - 6\alpha)$ can be derived in a similar way to that for deriving q_0 in the (5,7)-cone model.

We use vector $\vec{D} = \rho_1(h\vec{a_1} + k\vec{b_2})$ to describe distance between two neighboring triangles that do not have a common vertex. This is similar to the vector $\vec{D}$ defined in the (5,7)-cone model. Note that in this cone model, there are two pairs of neighboring triangles that do not share vertex, as shown in Figure 4.8. This shall allow a parallelogram to be formed in the middle of the cone. The top and bottom sides of the parallelogram should have the same length. Therefore, we have

$$(43) \qquad\qquad 5\beta\sqrt{T} = 6\alpha\sqrt{T} + 2\rho_1\sqrt{T},$$
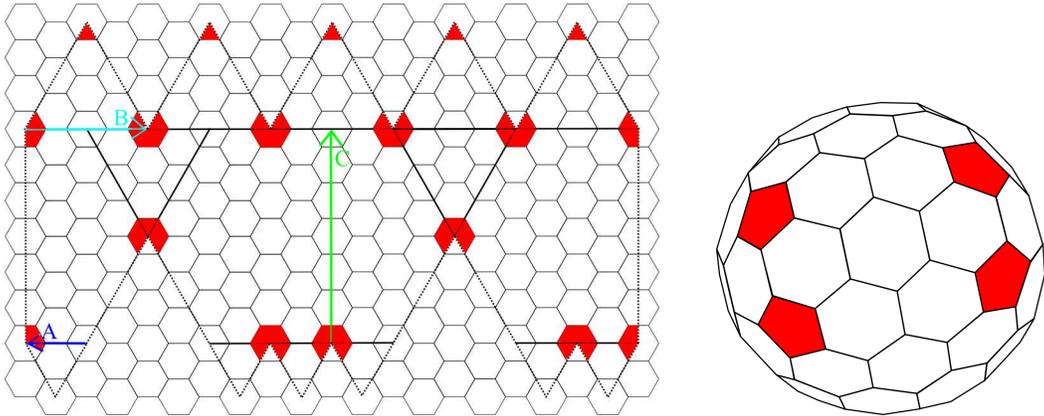
FIGURE 4.8. *Left*: Model for the (4,8)-cone using 3 generating vectors. The dotted lines indicate where to cut the lattice to fold it into a 3-dimensional cone. Pentagons are shown in red. *Right*: A view of the bottom or narrow end of the (4,8)-cone.

and hence

$$(44) \qquad \rho_1 = \frac{1}{2}(5\beta - 6\alpha).$$

Similar to the (5,7)-cone model, there are two ways for expressing the length of the generating vector $\vec{C} = q_1(h\vec{b_1} + k\vec{b_2})$. Specifically, we have

$$(45) \qquad q_1\sqrt{3T} = \frac{\sqrt{3}}{2}\left(\beta\sqrt{T} + \frac{1}{2}(5\beta - 6\alpha)\sqrt{T}\right),$$

and hence

$$(46) \qquad q_1 = \frac{1}{4}(7\beta - 6\alpha).$$

In summary, for a (4,8)-cone, the folding template is determined by two pairs of integers (h, k) and (α, β) (β needs to be strictly even), and three generating vectors:

$$(47) \qquad \vec{A} = \alpha(h\vec{a_1} + k\vec{a_2}),$$

$$(48) \qquad \vec{B} = \beta(h\vec{a_1} + k\vec{a_2}),$$

$$(49) \qquad \vec{C} = \frac{1}{4}(7\beta - 6\alpha)(h\vec{b_1} + k\vec{b_2}),$$

where $\vec{b_1}$ and $\vec{b_2}$ are defined in Equations (23) and (24).

Based on the above (4,8)-cone model, we can derive the surface area (SA), the number of proteins (CA), and the number of hexamers (N_H) as follows.

(50)
$$SA = 5 \left(\tfrac{\sqrt{3}}{4} |\vec{B}|^2 \right) + \left(5|\vec{B}||\vec{C}| - \tfrac{\sqrt{3}}{2} |\vec{D}|^2 \right) + 6 \left(\tfrac{\sqrt{3}}{4} |\vec{A}|^2 \right) (\text{unit}^2)$$
$$= \tfrac{3\sqrt{3}}{8} T \left(55\beta^2 - 24\alpha^2 \right) a^2 \ (\text{nm}^2),$$

(51)
$$CA = \frac{3T}{2} \left(55\beta^2 - 24\alpha^2 \right),$$

(52)
$$N_H = \frac{T}{4} (55\beta^2 - 24\alpha^2) - 10.$$

Folding Templates for the (3,9)-, (2,10)-, (1,11)-Cones. According to the Euler theorem, there are 3 other possible cone angles for a hexagonal lattice, corresponding to a narrow end with 3, 2, or 1 pentagon(s). The $1P$ and $2P$ cones would induce higher strain due to their non-spherical shapes, thus are unlikely to form in nature [28]. The $3P$ cone, with a cone angle 60°, is the preferred cone angle for the helical cone for graphite [78]. Since these cones overlap, they would not follow the same construction rules as the isometric models investigated in this chapter. Therefore, it is unnecessary to construct isometric cones for the remaining cone angles.

4.5. COMPARISON OF MODELING RESULTS TO HIV-1 DATA

The formulas for the surface area, number of CA proteins, and number of hexamers on the (5,7)-cone have already been established in Equations (40-42). Other common measurements for the (5,7)-cone such as the broad end diameter D_b, the narrow end diameter D_a, and the

overall height H, are respectively

$$(53) \qquad\qquad D_b = \frac{5\beta\sqrt{3T}}{\pi} a \text{ (nm)},$$

$$(54) \qquad\qquad D_a = \frac{5\alpha\sqrt{3T}}{\pi} a \text{ (nm)},$$

$$(55) \qquad\qquad H = \frac{5\sqrt{3T}}{2\pi} \left(\frac{\beta - \alpha}{\tan\theta} + (\beta + \alpha) \right) a \text{ (nm)},$$

where a is the side length of a hexamer, α and β are the constants associated respectively with the scaling of $\vec{A}$ and $\vec{B}$, T is the triangulation number, and $\theta = 19.2°/2$.

Equations (53) and (54) can be derived from the lattice construction. Note that the circumference of the broad end is given by $5|\vec{B}| = 5\beta\sqrt{3T}a$ (nm). Similarly, for the narrow end, the circumference is $5|\vec{A}| = 5\alpha\sqrt{3T}a$ (nm).

Equation (55) follows from summing the height of the cone with the radii of both hemispheres. We use the commonly known equation for the opening angle of a right cone, that is, $\phi = 2\arctan\left(\frac{r}{h}\right)$, where r is the radius and h is the height. This leads to $H = \frac{1}{2}\left(\frac{D_b - D_a}{\tan\theta}\right) + \frac{D_b}{2} + \frac{D_a}{2}$, where $2\theta = \phi$.

Next we compare modeling results with the experimental data on HIV-1 VLPs reported in [6, 11].

Comparison with Data in [11]. In [11], it is found that most HIV-1 VLPs cores exhibit a conical shape with an average cone angle of $22.3° \pm 6°$, although about 7% VLPs exhibit tubular morphology and few show amorphous morphology. Measurements for the overall height, the broad end diameter, and the cone angle were performed for 267 conical cores. Among the conical capsids with a single core, [11] found that the hexamer diameter is 9.8(nm) with a 3.2(nm) spacing between repeating hexameric/pentameric units.

The spacing in the assumption for the hexagon size should be taken into consideration. This implies that a has a value $(9.8 + 3.2)/2 = 6.5$(nm). Taking into account of how small the radius is found to be in this region compared to the length of a [8], it is assumed that the pentagons are grouped closely together yet still isolated. To explain this tight grouping, we take $(h, k) = (1, 1)$ for our model defined by Equations (37-39). Similar construction of the narrow end can be found in [9, 24, 61, 74]. We then take $(\alpha, \beta) = (1, 2)$. $\alpha = 1$ also follows from the narrow end described in [9, 24, 61, 74], while $\beta = 2$ is determined from the size of the broad-end diameter in [11].

With this value for a, the model for the (5,7)-cone with $(h, k) = (1, 1)$ and $(\alpha, \beta) = (1, 2)$ produces a conical capsid with properties listed in the 3rd column of Table 4.1.

Comparison with Data in [6]. Among the experimental data on 26 HIV-1 VLPs reported in [6] are

- 16 VLPs exhibit conical morphology;

- 3 VLPs have tubular morphology;

- The rest have irregular shape.

Among the VLPs with conical morphology, the measurements are found in the 2nd column of Table 4.2.

To compare the experimental data in [6] to the theoretical results derived from the models proposed in this chapter, we only require the side length of the hexamers and pentamers.

TABLE 4.1. Comparison of modeling results with experimental data in [11].

	Experimental data	Modeling results
Cone angle	22.3° ± 6°	19.2°
Cone overall height	119.3(nm) ± 11(nm)	134(nm)
Broad-end diameter	60.7(nm) ± 8(nm)	62(nm)

TABLE 4.2. Comparison of modeling results with experimental data in [6].

	Experimental data	Modeling results
Mean angle	20.1°	19.2°
Mean height	143(nm) (standard deviation 10.8(nm))	127(nm)
Mean surface area	21,000 (nm^2) (standard deviation 9000(nm^2))	20,612(nm^2)
Number of hexamers	206 hexamers for each capsid (1300 CA monomers)	200

[6] reported the surface area of the capsid with an estimated 200 hexamers. Assuming the hexamers and pentamers have the same side length a, one reaches an estimate $a = 6.1465$ (nm) for the given surface area. This value is similar to the a value found from the data in [11]. So we choose the same parameter values for our model described in Equations (37-39). Then the model for the (5,7)-cone with $(h, k) = (1, 1)$ and $(\alpha, \beta) = (1, 2)$ yields the results shown in the 3rd column of Table 4.2.

The comparison with the data in these two papers demonstrate good agreement of our modeling results with experimental data.

4.6. DISCUSSION

Flaw in the Nguyen Model. In [54], Nguyen et al. used two generating vectors for the lattice structure of the HIV-1 (5,7) conical capsid. The two generating vectors are

$$(56) \qquad \vec{A} = n(h\vec{a_1} + k\vec{a_2}),$$

$$(57) \qquad \vec{B} = m(h\vec{a_1} + k\vec{a_2}),$$

where h, k are two non-negative integers used to determine the T-number and n, m are two other non-negative integers similar to those in our model. However, an important piece of

information is missing from this model: the unique height required for closure of the cone. The required height allows the two outer middle closure lines to be parallel, which in return match the cut hexagons to form into a pentagon. If these lines are not parallel, then either a more than 60° declination (shown in yellow near the top of Figure 4.9 (*left*)) or a less than 60° declination (shown in yellow near the bottom of Figure 4.9 (*left*)) will be produced. Both cases result in incomplete pentagons formed on the capsid. Shown in Figure 4.9 is an example of failure in closure.

The geometric models proposed in this chapter are useful for investigating the discrete curvatures and curvature concentrations on the HIV-1 conical capsids [42]. It is suggested in [13] that the asymmetry and quasi-equivalence exhibiting in tubular and conical capsids are related to the hinge between the C-terminal domain (CTD) and N-terminal domain (NTD) of the capsid protein. This information could be utilized to study the elastic energy on the capsid and the relationship between curvature and elastic energy.
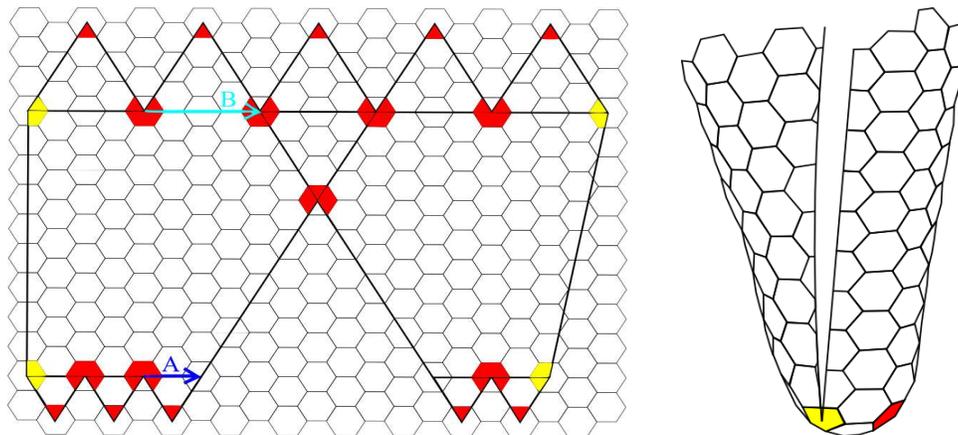
FIGURE 4.9. *Left*: An illustration of a lattice with $(h, k) = (1, 1)$, $T = 3$, $(n, m) = (1, 2)$ according to the Nguyen model with only two generating vectors [54], where the two generating vectors $\vec{A}$ and $\vec{B}$ are defined in Equations (56) and (57). Pentamer positions are shown in red. Incomplete pentagons are shown in yellow. *Right*: The lattice folds into an incomplete (5,7)-cone, the partial hexagons along the outer edges do not match, since the unique height is not enforced in the Nguyen model.

Modeling Viral Capsid Assembly

In this chapter, we explore an inexpensive approach for modeling and simulations of viral capsid assembly. Based on the biological evidence presented in Chapter 2, the assembly process is viewed in two stages: nucleation and elongation. At the early stage of viral capsid assembly, lower order CA proteins nucleate into hexamers simultaneously in many locations within the virion. Then these hexamers further assemble into the viral capsid. Pentamers might form at the places where it is difficult for a hexamer to form.

We focus on the nucleation stage by investigating the kinetics of nucleation. Specifically, a 6-species dynamical system model is developed by considering all possible pathways of association and dissociation. Then biological evidence [10, 16, 29, 49, 79] are used to reduce the model. Published biological experimental data [60] are utilized to estimate the model parameters representing the association and disassociation rates. Furthermore, sensitivity and elasticity analysis are performed to determine which association / dissociation terms are required during the nucleation stage.

5.1. HIV-1 Maturation

Retrovirus capsid assembly has been observed to undergo two stages: nucleation and elongation [37], as shown in Figure 5.1. For HIV-1, these occur during the process of maturation, which is the formulation of the cone-shaped core. In principle, the HIV-1 capsid is composed of two types of units: hexamers and pentamers. Hexamers are the primary units (Figure 5.2), where the amount of hexamers range between 200-260 in each capsid [6],
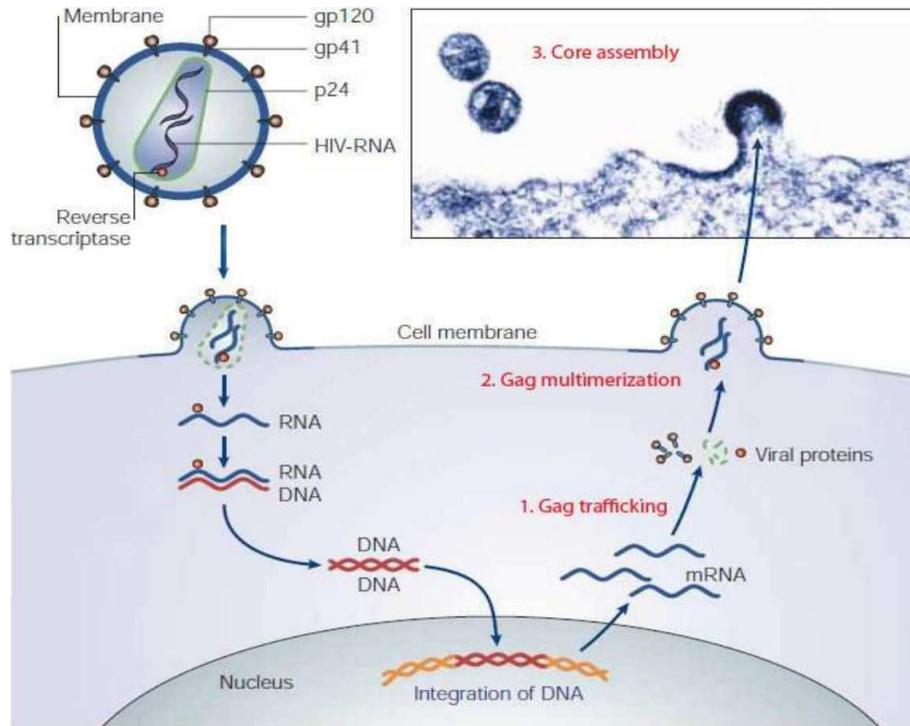
FIGURE 5.1. Retroviral Lifecycle. HIV fuses with the host cell's membrane and releases it's RNA and enzymes in the cytoplasm. HIV's RNA is translated to DNA, integrated into the host cell's DNA, then translated into viral mRNA. The mRNA produces the Gag proteins which travel to the membrane. The virions leave the cell during the budding process, develop a protective capsid shell and become infectious during maturation [52, 79].
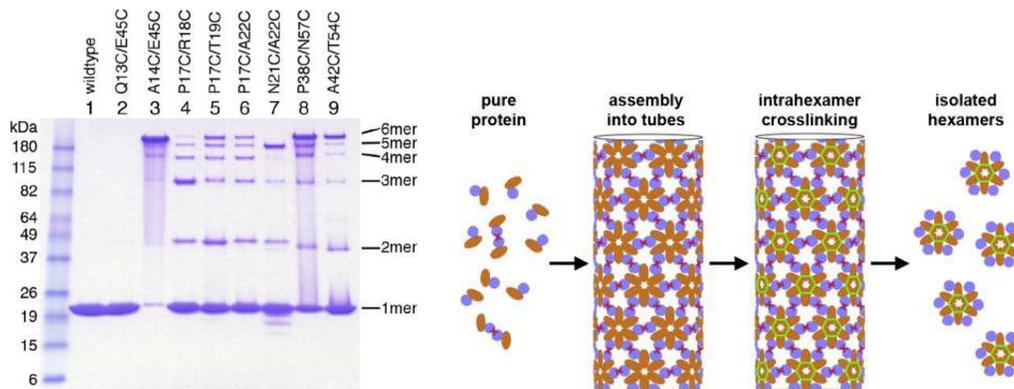
FIGURE 5.2. *Left:* HIV-CA n-mer profiles during nucleation. In most experiments reported here, CA subunits (monomers) and 6-mers were the most prevalent [79]. *Right:* HIV-CA 6-mers (hexamers). Simulations to obtain soluble HIV-1 CA hexamers for 3D crystallization [79].

depending on the size of the conical core. It is known that there are exactly 12 pentamers for each closed retroviral capsid [17, 74, 79] as shown in Figure 5.3.

The basic building blocks of a HIV viral capsid are capsid (CA) protein subunits, called monomers. Two monomers form a dimer, a monomer and a dimer form a trimer, so on so forth, until the first nucleus forms. In the case of HIV-1, we could assume the first nucleus is a hexamer, since they are most prevalent as shown in Figure 5.3, though under certain conditions the formulation of pentamers are favored [6, 11]. The formation of the first nucleus completes the nucleation stage. It has been hypothesized that the elongation stage begins as more hexamers or pentamers are added to the growing lattice until the CA proteins have formed the closed protective shell, i.e., a capsid consisting of hexamers and pentamers.

Modeling of viral capsid assembly and more broadly, viral life cycles, is an active and challenging research area in mathematical biology. Previous work has modeled the whole process of viral capsid assembly using one large-size dynamical system [85]. But these models
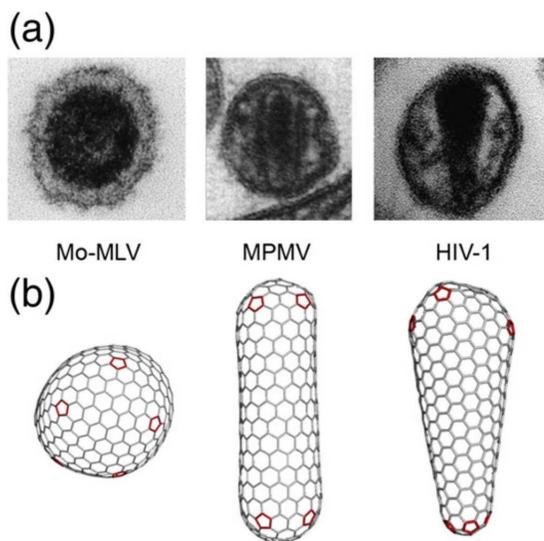
FIGURE 5.3. Three different capsid lattices: 12 pentamers (red) close the curved hexagonal lattice in Moloney murine leukemia virus (Mo-MLV), Mason-Pfizer monkey virus (MPMV) and HIV-1 [79].

consider a simplified pathway that allows association or dissociation of one capsomer unit at a time. This approach might be conceptually simple, but the implementation of the models and computer simulations are difficult, since the dynamical system is of size 1500 or larger. Additionally, these simplified pathways ignore biological evidence that non-monomer subunits, such as dimers and higher order n-mers, can assemble with each other [59, 60]. This past approach has also overlooked the effects and experimental evidence of the nucleation and elongation stages.

5.2. Existing Work

There is a wide range of biological and mathematical work published on HIV-1 capsid assembly. This section will briefly review this work as well as some of the drawbacks associated with each model.

Ganser's Group. Ganser and co-workers studied the structure of the HIV-1 conical capsid from a biological perspective by creating virus-like particles (VLPs) via election cryotomography (CryoEM) [6], shown in Figure 5.4. The reconstructions revealed that the structures and positions of the conical cores within each VLP are unique, though they still exhibited several consistent features. One being the conical core with an average cone angle of 19.2°. Another being the positioning of the base of the capsid to the envelope/MA layer. [6] also found multiple and nested capsids. These results support the fullerene cone model, indicating that maturation involves a free re-organization of the capsid shell and not a continuous condensation of proteins. This implies the assembly process may be template-driven (See Chapter 4) and the particular cone-shaped capsid is strongly favored.
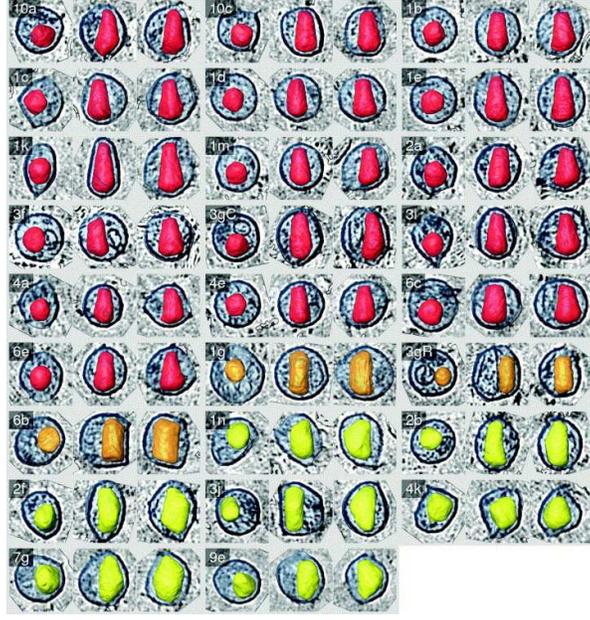
FIGURE 5.4. HIV-1 Virus-Like Particles (VLPs) [6].

Zlotnick's Group. Zlotnick and his group have been working on characterizing viral protein assembly for almost two decades [85]. Their work focuses on using differential equations for the rates of change of the intermediate concentrations of the capsid. These equations model the capsid assembly as a polymerization reaction, assuming only one monomer at a time associates or dissociates from the growing structure. Association or disassociation of higher order intermediates is not considered.

In most of their work, only one forward rate constant k_f is used. The backward rate constants $k_b n$ vary depending on the size (n) of the growing capsid and are calculated by the equation: $k_b n = k_f / K_\alpha n$, where $K_\alpha n$ is an association constant. The full equations used are

$$(58) \qquad \frac{d[polymer_n]}{dt} = k_f([polymer_{n-1}] - [ploymer_n])[freesubunit] - k_b[polymer_n].$$

In simulations, these equations were used to model the formation of a spherical capsid with T=1, or in other words a dodecahedron, where there are exactly 12 pentamers and no

hexamers. Numerical solutions were computed with STELLA, using a forth-order Runge-Kutta method. Both equilibrium assembly (EQ) and kinetically limited (KL) models were considered. In the EQ model, intermediates form and break apart and the reaction is able to continue, though kinetic traps are likely. In the KL model, early reaction is slow. The two models have their own advantages, EQ does not require nucleation and is more appropriate for an assembly process with weak association energies but KL is able to avoid kinetic trapping.

In [21], rate equations are used to model assembly of a dodecahedron (12 pentamers, no hexamers) and icosahedral capsids (30 tetramers). The system of rate equations is formulated as

$$(59) \qquad \frac{d[m]}{dt} = f_m s_m [u][m-1] - f_{m+1} s_{m+1} [u][m] + b_{m+1} [m+1] - b_m [m],$$

where s_m is a degeneracy statistical factor, and f_m, b_m are forward and backward rates respectively, for the m-th species.

Zlotnick and Katen reiterate in [37] that virus capsid subunits interact through weak contact energies, which leads to a dynamic globally stable structure. Assembly is still modeled as a polymerization, although assembly is now divided into the two stages: nucleation and elongation. There is a lag phase, which ends when the first nucleus forms, then rapid growth during elongation. During elongation, concurrent nucleation and elongation events can happen, leading to multiple capsids as seen *in vitro*. The system equations are similar to [21], with subunit s, elongation k_e and disassociation k_d, expressed as

$$(60) \quad \frac{d[nuc+n]}{dt} = k_{e,n-1}[nuc+n-1][s] + k_{d,n+1}[nuc+n+1][s] - k_{e,n}[nuc+n][s] + \cdots .$$

They also investigate conditions for kinetic trapping. When forward rates are too high, the intermediates form too rapidly, creating too many intermediates without forming closed capsids. Also, if nucleation happens too quickly in comparison to elongation, too many metastable intermediates are formed. Lastly, "off-path assemblies" can also occur, causing metastable intermediates as well.

Hagan's Group. In [30], Hagan presents a review of the theoretical and computational methodologies that have been used to model the assembly of viral capsids. A new model is created in attempt to eliminate the kinetic trapping found in [21] from the extreme differences in the time needed for nucleation versus elongation.

The reaction system of capsid protein subunits, with total concentration ρ_T, that start at $t = 0$ is given by

$$(61) \qquad 1 \underset{b_{\text{nuc}}}{\overset{f\rho_1}{\rightleftharpoons}} 2 \underset{b_{\text{nuc}}}{\overset{f\rho_1}{\rightleftharpoons}} \cdots \underset{b_{\text{nuc}}}{\overset{f\rho_1}{\rightleftharpoons}} n_{\text{nuc}} \underset{b_{\text{elong}}}{\overset{f\rho_1}{\rightleftharpoons}} \cdots \underset{b_{\text{elong}}}{\overset{f\rho_1}{\rightleftharpoons}} N,$$
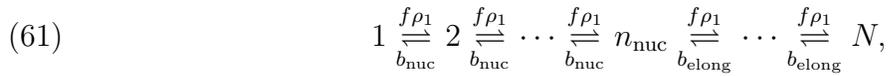
where N is the number of subunits (monomers) in a capsid, ρ_1 is the concentration of subunits not yet assembled, and b_i is the dissociation rate constant for each stage: nucleation and elongation.

Time is also distinguished between the two stages: nucleation (τ_{nuc}) and elongation (τ_{elong}), where

$$(62) \qquad \tau_{\text{elong}} = \frac{n_{\text{elong}}}{f\rho_1 - b_{\text{elong}}} - \left(\frac{b_{\text{elong}}}{f\rho_1 - b_{\text{elong}}}\right)^2 \left(\frac{b_{\text{elong}}}{f\rho_1}\right)^{n_{\text{elong}}},$$

and $n_{\text{elong}} = N - n_{\text{nuc}}$ with n_{nuc} being the average nucleus size.

Similarly, the formula for τ_{nuc} can be derived as

$$(63) \qquad \tau_{\text{nuc}} = \frac{n_{\text{nuc}}}{f\rho_1 - b_{\text{nuc}}} - \left(\frac{b_{\text{nuc}}}{f\rho_1 - b_{\text{nuc}}}\right)^2 \left(\frac{b_{\text{nuc}}}{f\rho_1}\right)^{n_{\text{nuc}}-1}.$$

However, free subunits are depleted by assembly, so the net nucleation rate never reaches this value but asymptotically approaches zero as time approaches equilibrium. Again this system becomes kinetically trapped at a larger concentration ρ_{kt}. Kinetic traps arising from depletion of free subunits has also been seen in experiments on CCMV and HBV [30].

Hagan and Elrad [31] consider rate equation models for capsid assembly developed by Zlotnick's group, disregarding malformed capsids. It again assumes only monomers can associate and dissociate. They define only one forward rate and one backward coefficient for each intermediate. The model is given by

$$(64) \qquad \begin{cases} \frac{dc_1}{dt} = -2f_1 c_1^2 + b_2 c_2 + \sum_{n=2}^{N} -f_n c_n c_1 + b_n c_n \\[2mm] \frac{dc_n}{dt} = f_{n-1} c_1 c_{n-1} - f_n c_1 c_n - b_n c_n + b_{n+1} c_{n+1} \end{cases}$$

for $n = 2, ..., N$, where c_n is the concentration of intermediates with n subunits, f_n, b_n are forward and backward rates respectively for intermediate n.

Despite the simplifications, the rate equations show good agreement with median assembly times of experimental assembly kinetics data. Using this model, they show that the nucleus size can be determined from the concentration dependence of the assembly half-life. They also determine that elongation time is dependent on the length of the lag phase.

Each of these models presented focus on the assembly of empty icosahedral viral shells. Specifically they best describe capsids with T=1, or 12 pentamers and no hexamers. There

is a need to examine higher order capsid structure dynamics as well as cargo-containing capsids. Each model also reiterates the time difference between the two maturation stages. Since there is a need for separate modeling and simulations of nucleation and elongation stages, we first consider models for nucleation only. This work will further shed light on the elongation stage for additional shapes and symmetries of capsids.

5.3. Dynamical System Models for Nucleation

Our approach in modeling HIV-1 assembly is to use dynamical systems. Previous models assume that only single monomers bind or unbind. There is also strong evidence [13, 16, 29] that dimers form with other dimers. Moreover, non-monomer subunits can assemble with each other [59, 60]. Therefore, this research is focused on exploring models where larger intermediates can bind with each other.

The assembly models we consider is related to Zlotnick's model [85], which is a dynamical system similar to a population model for interaction species. In our case, each species represents a particular capsid intermediate or n-mer.

5.3.1. A Nucleation Model of Three Intermediates. A system of three differential equations, based on the concentrations of subunits, intermediates, and capsids will be examined. This model only considers the interaction of monomers, (c_1), dimers, (c_2), and hexamers, (c_6), also referred to as nuclei.

Assumptions.

- Nucleation ends with 6-mer formation. [60, 59] observed little to no existence of $c_n, n > 6$;

- One forward rate for each species;

- Multimers can dissociate in the same way they are formed in association.

Variables and Parameters.

- c_n is the concentration of the n-mer intermediate;

- f_{ij} is the association rate of c_i and c_j;

- f_{222} is the association rate for trimer-of-dimer;

- b_{ij} is the rate of c_i dissociating into two intermediates with c_j being the largest intermediate of the disassociated terms, b_{62} is for the special case 6-mer dissociates into three dimers.

Intermediates: Association and Dissociation. We consider one pathway for the formation of hexamers: Three dimers form together to create a hexamer as shown in Figure 5.5.

$$(65) \qquad c_1 + c_1 \underset{b_{21}}{\overset{f_{11}}{\rightleftharpoons}} c_2, \quad c_2 + c_2 + c_2 \underset{b_{62}}{\overset{f_{222}}{\rightleftharpoons}} c_6.$$

Full Equations. The assembly model is a dynamical system of first-order, autonomous, nonlinear ordinary differential equations. The rate of change of the concentration
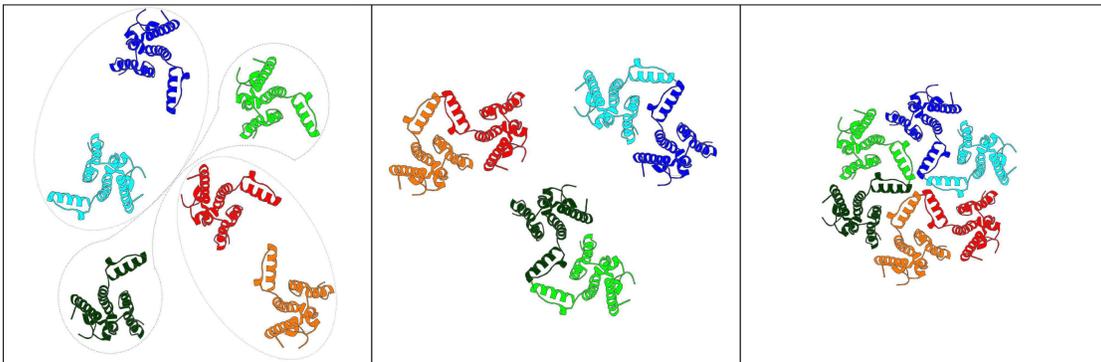
FIGURE 5.5. Illustration of the second pathway (trimer-of-dimers) for hexamer assembly. Protein illustrations are drawn according to PDB 3H47 HIV-1 CA monomer [58].

of a specific n-mer (monomer, dimer, hexamer) is described in each equation, with respect to time. The equations for the nucleation phase are given by

(66)
$$\begin{cases} \frac{dc_1}{dt} = 2b_{21}c_2 - 2f_{11}c_1^2 \\[2ex] \frac{dc_2}{dt} = f_{11}c_1^2 + 3b_{62}c_6 - b_{21}c_2 - 3f_{222}c_2^3 \\[2ex] \frac{dc_6}{dt} = f_{222}c_2^3 - b_{62}c_6 \end{cases}$$

Unique Equilibrium. We set the Jacobian determinant to zero as follows

(67)
$$\begin{vmatrix} -4f_{11}c_1 & 2b_{21} & 0 \\[1.5ex] 2f_{11}c_1 & -b_{21} - 9f_{222}c_2^2 & 3b_{62} \\[1.5ex] 0 & 3f_{222}c_2^2 & -b_{62} \end{vmatrix} = 0.$$

This implies that the equations are not independent. We reduce the system by imposing the mass conservation condition (the total concentration of the subunits is constant):

(68) $\quad c_1(t) + 2c_2(t) + 6c_6(t) = c_1(0), \quad \forall t \geq 0 \quad \Rightarrow c_6(t) = \dfrac{c_1(0) - c_1(t) - 2c_2(t)}{6}.$

The reduced system is then

(69)
$$\begin{cases} \frac{dc_1}{dt} = -2f_{11}c_1^2 + 2b_{21}c_2 \\[2ex] \frac{dc_2}{dt} = f_{11}c_1^2 - 3f_{222}c_2^3 - b_{21}c_2 + \frac{1}{2}b_{62}\left(c_1(0) - c_1 - 2c_2\right) \end{cases}$$

64

The Jacobian determinant for the reduced system is

$$
(70) \qquad \begin{vmatrix} -4f_{11}c_1 & 2b_{21} \\ 2f_{11}c_1 - \frac{1}{2}b_{62} & -b_{21} - 9f_{222}c_2^2 - b_{62} \end{vmatrix} = 36f_{11}f_{222}c_1c_2^2 + 4f_{11}b_{62}c_1 + b_{21}b_{62} \neq 0.
$$

Therefore, we conclude that a unique equilibrium exists. Let

$$
B = b_{21} + b_{62} + 9f_{222}c_2^2 + 4f_{11}c_1,
$$

$$
(71) \qquad C = 36f_{11}f_{222}c_1c_2^2 + b_{21}b_{62} + 4b_{62}f_{11}c_1,
$$

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

then the eigenvalues of the Jacobian matrix are

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

It is assumed that the equilibrium concentrations c_1 and c_2 are both non-negative and that the forward and backward rates are all positive. With these assumptions, we have $B > 0$, which guarantees the equilibrium to be asymptotically stable if $D \leq 0$.

Suppose that $D > 0$. Then, the equilibrium is unstable when $D > B^2$. But $D = B^2 - 4C < B^2$, since $C > 0$ by the assumptions. Therefore, the equilibrium is asymptotically stable.

***In Silico:* Parameters.** We chose arbitrary values for the forward and backward coefficients f_{ij} and b_{ij}. Literature suggests that hexamers are the most stable structure [61], thus we assume their dissociation into smaller intermediates, b_{62}, is small. It is also known that monomers are the least stable, since they readily form together to create dimers, thus we choose f_{11} to be large. Similarly to hexamers, the dissociation rate of dimers, b_{21} is small,

though we can assume $b_{62} < b_{21}$, since hexamers are more stable than dimers [79]. Lastly, it is know that trimers of dimers tend to form together to create hexamers[29], thus f_{222} is chosen to be large in magnitude.

In Silico: **Initial Values.** The initial concentration of c_2 and c_6 are set to 0. We set $c_1(0) = 1300$ since on average, 1300 copies of CA are needed for one HIV-1 conical core [6].

Observations. Figure 5.6 shows the n-mer (monomer, dimer, hexamer) values described by the dynamical system from time $t = 0$ to $t = 100$. These solutions are for the
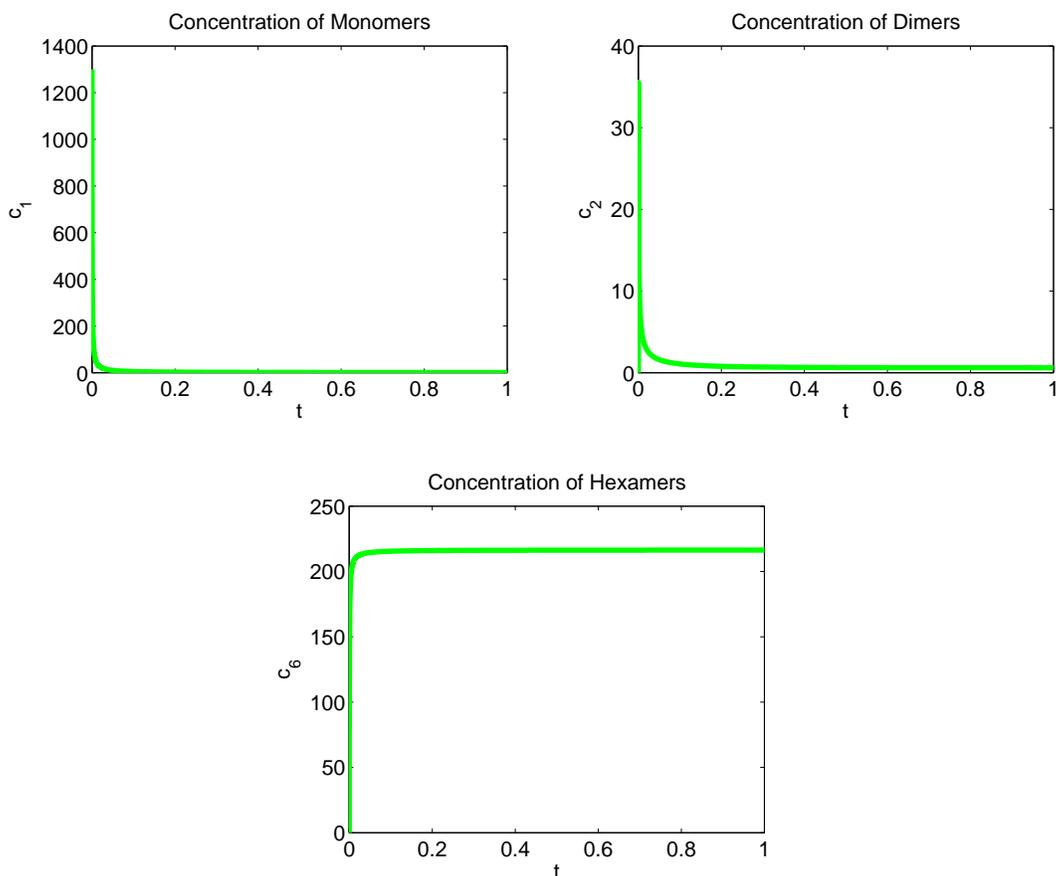
FIGURE 5.6. *In silico* for $c_1(0) = 1300$: Concentration of n-mers from time $t = 0$ to $t = 100$, where c_1 is the concentration of monomers, c_2 for dimers, and c_6 for hexamers.

66

case when the parameter values are set as $f_{11} = 1$, $b_{21} = 0.1$, $f_{222} = 1$, and $b_{62} = 0.01$, with initial conditions $c_1(0) = 1300$ and $c_2(0) = c_6(0) = 0$.

The concentration results are as expected. The concentration of monomers quickly decreases over time, as the dimers and hexamers are composed from them. There is an initial spike on the concentration of dimers, as monomers first join to form dimers. Then the dimer concentration starts to decrease as trimer-of-dimers form into hexamers. Clearly, the hexamer concentration is expected to increase over time, which corresponds to the monomers and dimers forming complete hexamers.

This model is not complicated, but clearly demonstrates one underlying process of hexamer formation. Do to the simplicity of this model, we skip the sensitivity and elasticity analysis so we may move forward to a more in-depth model for capsid nucleation.

5.3.2. FULL SIX-INTERMEDIATE NUCLEATION MODEL. As mentioned before, among the existing work [21, 31, 52], a natural and straightforward approach considers only one pathway of assembly: only one CA protein (monomer) can assemble with another subunit at a time, That is, from n-mer to $(n+1)$-mer. Similarly the dissociation is from $(n+1)$-mer to n-mer. However, there is strong evidence [13, 16, 29] that dimers interact with other dimers. The findings in [59, 60] suggest that non-monomer subunits can assemble with each other. Stability analysis in [16] predicts that dimer is an important CA intermediate in self assembly.

Based on the aforementioned work, we start with a new model by considering all possible pathways for forming a nucleus, also referred to as a hexamer or 6-mer. Dissociation is also important, due to high concentrations of intermediates left after nucleation, more terms are

added to describe the multitude of disassociations for c_n. This model and following sections are incorporated into preprint [65].

Assumptions, Variables, and Parameters. The assumptions, variables, and parameters in this model are the same as those listed in the three-intermediate nucleation model.

Full Equations. Here is a system of six ordinary differential equations or a dynamical system of size six describing the kinetics in the association and dissociation, based on the above assumptions.

$$(73) \quad \begin{cases} \dfrac{dc_1}{dt} = b_{65}c_6 + b_{54}c_5 + b_{43}c_4 + b_{32}c_3 + 2b_{21}c_2 \\ \qquad\quad - f_{15}c_1c_5 - f_{14}c_1c_4 - f_{13}c_1c_3 - f_{12}c_1c_2 - 2f_{11}c_1^2 \\[2mm] \dfrac{dc_2}{dt} = f_{11}c_1^2 + 3b_{62}c_6 + b_{64}c_6 + b_{53}c_5 + 2b_{42}c_4 + b_{32}c_3 \\ \qquad\quad - b_{21}c_2 - 3f_{222}c_2^3 - f_{24}c_2c_4 - f_{23}c_2c_3 - 2f_{22}c_2^2 - f_{21}c_1c_2 \\[2mm] \dfrac{dc_3}{dt} = f_{12}c_1c_2 + 2b_{63}c_6 + b_{53}c_5 + b_{43}c_4 - b_{32}c_3 - 2f_{33}c_3^2 - f_{23}c_2c_3 - f_{13}c_1c_3 \\[2mm] \dfrac{dc_4}{dt} = f_{13}c_1c_3 + f_{22}c_2^2 + b_{64}c_6 + b_{54}c_5 - b_{43}c_4 - b_{42}c_4 - f_{24}c_2c_4 - f_{14}c_1c_4 \\[2mm] \dfrac{dc_5}{dt} = f_{14}c_1c_4 + f_{23}c_2c_3 + b_{65}c_6 - b_{54}c_5 - b_{53}c_5 - f_{15}c_1c_5 \\[2mm] \dfrac{dc_6}{dt} = f_{15}c_1c_5 + f_{33}c_3^2 + f_{222}c_2^3 + f_{24}c_2c_4 - b_{65}c_6 - b_{64}c_6 - b_{63}c_6 - b_{62}c_6 \end{cases}$$

5.3.3. REDUCED SIX-INTERMEDIATE NUCLEATION MODEL. The full six-intermediate nucleation model is a six-species dynamical system with 20 parameter values, described in the previous section. The model considers all possible pathways of two binding intermediates and one triple bond (trimer-of-dimers) in the association leading to and dissociation down

from hexamers. Published biological data suggest that this model could be simplified. This section formulates a reduced model from Equation (79). The assumptions, variables, and parameters are consistently defined.

Intermediates: Association and Dissociation. The hexamer pathways are based on the findings presented in [49]. The first pathway [P1] along which monomers join one at a time was adopted in [21, 31, 52]. "The symmetric appearance (of a hexamer) is suggestive of symmetric head-to-head dimers," as shown in Figure 5.7, promoting the trimer-of-dimer assembly seen in the second pathway [P2]. This is also advocated in [10, 16, 29] and here illustrated in Figure 5.5. The third pathway [P3] for a hexamer considered in our reduced model is established based on the discussion in [7, 13, 27, 38, 39, 49, 74]. In particular, [74] asserts that CA prefers to form both dimers and tetramers. This pathway could also be considered the "slow" formation of trimer-of-dimers. Instead of three dimers joining
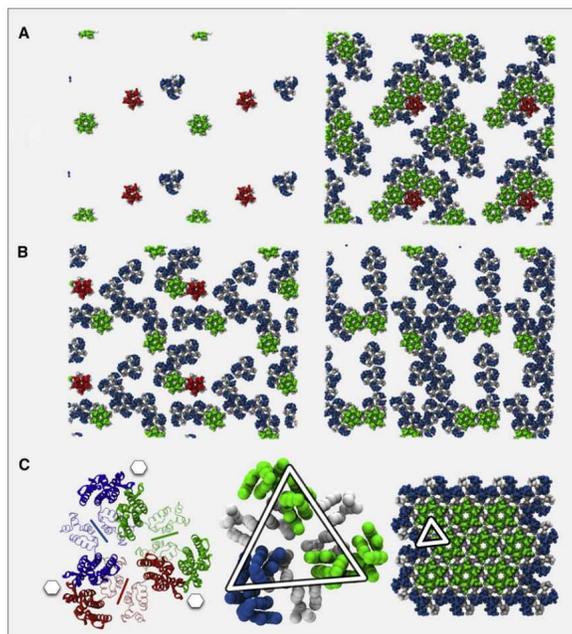
FIGURE 5.7. Self-assembly snapshots for the HIV-1 CA dimer lattice. Hexamers are denoted in green, pentamers in red, and trimer-of-dimers in blue. Monomers not associated with the three listed structures are omitted from this figure for clarity [16].

together simultaneously, two dimers may first form a tetramer quickly followed by a third dimer joining to create the hexamer. Assuming only these two pathways will eliminate the parameter f_{33} and the corresponding backward rate b_{63} from the model.

Single monomers join:

$$(74) \qquad c_1 + c_1 \underset{b_{21}}{\overset{f_{11}}{\rightleftharpoons}} c_2, \quad c_1 + c_2 \underset{b_{32}}{\overset{f_{12}}{\rightleftharpoons}} c_3, \quad c_1 + c_3 \underset{b_{43}}{\overset{f_{13}}{\rightleftharpoons}} c_4, \quad c_1 + c_4 \underset{b_{54}}{\overset{f_{14}}{\rightleftharpoons}} c_5, \quad c_1 + c_5 \underset{b_{65}}{\overset{f_{15}}{\rightleftharpoons}} c_6.$$

Trimer-of-dimers as illustrated in Figure 5.5:

$$(75) \qquad\qquad\qquad c_1 + c_1 \underset{b_{21}}{\overset{f_{11}}{\rightleftharpoons}} c_2, \quad c_2 + c_2 + c_2 \underset{b_{62}}{\overset{f_{222}}{\rightleftharpoons}} c_6.$$

Single binding dimers:

$$(76) \qquad\qquad\qquad c_1 + c_1 \underset{b_{21}}{\overset{f_{11}}{\rightleftharpoons}} c_2, \quad c_2 + c_2 \underset{b_{42}}{\overset{f_{22}}{\rightleftharpoons}} c_4, \quad c_2 + c_4 \underset{b_{64}}{\overset{f_{24}}{\rightleftharpoons}} c_6.$$

In addition to hexamer pathways, pentamer pathways must be examined since pentamers are required for formation of a closed viral capsid [6, 8, 23, 61]. Both pathways for pentamer formation occur as either a sub-pathway or union of hexamer pathways. Note only considering these two pentamer pathways allows the elimination of the term $f_{23}c_2c_3$, and its corresponding backward rate term $b_{53}c_5$, from the full model.

This implies there are mainly two pathways for a pentamer:

Single monomers join (seen as part of pathway [P1]):

$$(77) \qquad\qquad c_1 + c_1 \underset{b_{21}}{\overset{f_{11}}{\rightleftharpoons}} c_2, \quad c_1 + c_2 \underset{b_{32}}{\overset{f_{12}}{\rightleftharpoons}} c_3, \quad c_1 + c_3 \underset{b_{43}}{\overset{f_{13}}{\rightleftharpoons}} c_4, \quad c_1 + c_4 \underset{b_{54}}{\overset{f_{14}}{\rightleftharpoons}} c_5.$$

Dimers and monomer (seen as the union of pathways [P1] and [P3]):

$$(78) \qquad c_1 + c_1 \underset{b_{21}}{\overset{f_{11}}{\rightleftharpoons}} c_2, \quad c_2 + c_2 \underset{b_{42}}{\overset{f_{22}}{\rightleftharpoons}} c_4, \quad c_1 + c_4 \underset{b_{54}}{\overset{f_{14}}{\rightleftharpoons}} c_5.$$

It is clear that these pathways reduce the emphasis on the trimers. Even though trimers of MA proteins are predominately observed during the assembly of immature virions [7, 76], there is not much evidence that the CA proteins prefers trimer formation [2].

Full Equations. The above discussion leads to a reduced 6-species model:

$$(79) \quad \begin{cases} \dfrac{dc_1}{dt} = b_{65}c_6 + b_{54}c_5 + b_{43}c_4 + b_{32}c_3 + 2b_{21}c_2 \\[2mm] \qquad\quad -f_{15}c_1c_5 - f_{14}c_1c_4 - f_{13}c_1c_3 - f_{12}c_1c_2 - 2f_{11}c_1^2 \\[4mm] \dfrac{dc_2}{dt} = f_{11}c_1^2 + 3b_{62}c_6 + b_{64}c_6 + 2b_{42}c_4 + b_{32}c_3 \\[2mm] \qquad\quad -b_{21}c_2 - 3f_{222}c_2^3 - f_{24}c_2c_4 - 2f_{22}c_2^2 - f_{21}c_1c_2 \\[4mm] \dfrac{dc_3}{dt} = f_{12}c_1c_2 + b_{43}c_4 - b_{32}c_3 - f_{13}c_1c_3 \\[4mm] \dfrac{dc_4}{dt} = f_{13}c_1c_3 + f_{22}c_2^2 + b_{64}c_6 + b_{54}c_5 - b_{43}c_4 - b_{42}c_4 - f_{24}c_2c_4 - f_{14}c_1c_4 \\[4mm] \dfrac{dc_5}{dt} = f_{14}c_1c_4 + b_{65}c_6 - b_{54}c_5 - f_{15}c_1c_5 \\[4mm] \dfrac{dc_6}{dt} = f_{15}c_1c_5 + f_{222}c_2^3 + f_{24}c_2c_4 - b_{65}c_6 - b_{64}c_6 - b_{62}c_6 \end{cases}$$

This reduced 6-species model will be used for numerical simulations of CA protein nucleation. Sensitivity and elasticity of the intermediate concentrations $c_n(n = 1, \ldots, 6)$ to the forward and backward rates will be analyzed also (See the section "Results").

5.4. MATLAB Implementation

5.4.1. An Optimization Method for Model Parameter Fitting.

To obtain values of the model parameters based on published experimental data, we adopt the Particle Swarm Optimization (PSO) method [18]. This is a method for optimizing continuous non-linear functions. PSO has an open source MATLAB implementation, which will be used in this dissertation to optimize the values of the 16 parameters in the reduced model for viral capsid nucleation under certain constraints on the forward and backward rates.

PSO is a numerical method based on the stochastic optimization technique developed by Eberhart and Kennedy [18] in 1995. Since then, it has been widely used in many research fields, for example, neural network, telecommunications, design, control, signal processing, power systems, and data mining.

PSO shares similarities with other optimization techniques, for example, the Genetic Algorithm (GA). Compared to GA, PSO is easier to implement and has fewer parameters to adjust for reaching an optimal solution. PSO is also able to take real numbers as particles, in contrast, GA needs to change to binary encoding or special genetic operators. Other advantages of PSO include making no assumptions about the problem being optimized and obtaining global optimum solutions.

PSO optimizes a problem by having a population of candidate solutions (particles). It iteratively tries to improve the solutions with regard to additional constraints by updating generations until the target is met. In each iteration, the solutions are updated by tracking two values. One is the best solution or fitness ($\mathbf{p}$) each parameter has achieved, the other is the best value obtained by any other particle in the population ($g\mathbf{1}$).

After finding the two best values up to that time, the solutions update their velocities and positions by the following formulas:

(80) $$\mathbf{v}(i+1) \;=\; w\mathbf{v}(i) + c_1 r_1[\mathbf{p}(i) - \mathbf{x}(i)] + c_2 r_2[g\mathbf{1}(i) - \mathbf{x}(i)],$$

(81) $$\mathbf{x}(i+1) \;=\; \mathbf{x}(i) + \mathbf{v}(i+1),$$

where

- w is the initial inertia weight with a default value 0.9;

- $\mathbf{v}(i)$ is the particle velocity at iteration i;

- c_1, c_2 are the local and global best influence weights, respectively, typically set to $c_1 = c_2 = 2$;

- r_1, r_2 are random variables between $(0, 1)$;

- $\mathbf{x}(i)$ is the particle position at iteration i;

- $\mathbf{p}, g\mathbf{1}$ are defined as stated before.

A pseudo code for the procedure is shown as follows.

Begin $i := 0$;

For each particle

   Initialize the particle $\mathbf{P}(i) = \{x_1, x_2, ..., x_N\}$;

   Calculate the fitness value of $\mathbf{P}(i)$;

   If fitness value $(\mathbf{p})$ is better than $\mathbf{p}$ in history, replace $\mathbf{p}$;

End.

Choose the particle with the best fitness value and set as g;

For each particle

Calculate the new velocities and positions (Equations (80-81));

$i := i + 1;$

End.

5.4.2. SENSITIVITY & ELASTICITY ANALYSIS. Sensitivity analysis examines how a system's behavior respond to the changes in its parameters. Sensitivity analysis is useful for identifying important parameters that require additional investigation or insignificant parameters that could be eliminated from the model [73, 76].

Sensitivity is computed by finding the derivatives of each solution variable with respect to each parameter. In other words, the sensitivity of the ith variable (c_i) with respect to the kth parameter (p_k) is defined as

$$(82) \qquad S_{i,k} = \frac{\partial c_i}{\partial p_k}, \quad i = 1, ..., N, \ k = 1, ..., K,$$

where N is the size of the system and k is the dimension of the parameter space.

Redefining each equation in the ODE system to be

$$(83) \qquad \frac{dc_i}{dt} = h_i(\mathbf{c}, \mathbf{p}), \quad i = 1, ..., N; \ \mathbf{p} \in \mathbb{R}^k,$$

gives the sensitivity of all variables (c_i) with respect to all parameters when the following ODE system is solved:

$$(84) \qquad \frac{dS_{i,k}}{dt}(t) = \left(\sum_{n=1}^{N} \frac{\partial h_i}{\partial c_n} S_{n,k}(t) \right) + \frac{\partial h_i}{\partial p_k}(t), \qquad S_{i,k}(0) = 0.$$

Sensitivity analysis can yield misleading results when the parameter values vary in a large range of magnitude. Elasticity describes the rate of change of the relative change in the size of the variable with respect to the relative size of the parameter. The elasticity of the ith variable with respect to the kth parameter is defined as

$$(85) \qquad E_{i,k}(t) = \frac{p_k}{c_i(t)} \frac{\partial c_i}{\partial p_k}(t).$$

SENSAI [72] is a freely available MATLAB package for performing a forward sensitivity and/or elasticity analysis on parametrized systems of nonlinear first-order differential equations. SENSAI evaluates the Jacobian

$$(86) \qquad \frac{\partial h_i}{\partial c_n}, \quad i, n = 1, ..., N,$$

and the partial derivatives with respect to the parameters

$$(87) \qquad \frac{\partial h_i}{\partial p_k}, \quad i = 1, ..., N, \ k = 1, ..., K,$$

symbolically using MuPAD, then solves Equation (84) in MATLAB.

5.5. RESULTS

In this section, we first describe the experimental data used in comparison for our model. Next, we describe the constraints on parameters used to find a good initial guess needed for the PSO toolbox, when finding the appropriate parameter values. Then, the results from the PSO parameter fitting are discussed. Lastly numerical simulations are preformed in MATLAB.

5.5.1. USE OF BIOLOGICAL EXPERIMENTAL DATA. It is known from the discussion in [59, 60, 79] that the structures of CA hexamers are very difficult to obtain because of the weak interactions holding the hexamers together. Mutant CA hexamers are utilized for investigation.

[59] compared each mutant hexamer to the HIV-1 CA hexamer given by the Protein Data Bank (PDB) code 3dik. It was found that four mutants assembling into tubes "appeared similar in morphology to the wild-type tubes". Of the four, only two mutants (A14C/E45C in lane 3 and A42C/T54C in lane 9) had enriched 6-mer bands, which is favorable for hexamer bonding to create the full lattice.

[59] states that A14C/E45C produces hexamers that are the most similar to wild-type HIV-1 hexamers, and adding two more mutations gives the more favorable results to construct A14C/E45C/W184A/M185A. However, no data is reported for this construct.

[60] presented a similar study, creating mutant CA protein that faithfully mimic the hexamer properties of HIV-1 capsid. It was found that the same two mutants A14/E45 and A14C/E45C/W184A/M185A produced the most realistic results. In this case, it was found that the latter mutant assembled less efficiently than A14C/E45C alone. Results were also compared the PDB code 3dik.

Both [59, 60] considered hexamers stabilized by engineering disulfide cross-link (the mutation) A14/E45 with similar results. [60] gives more information about the protein concentration and timing.

In [60], Crosslinked CA A14C/E45C hexamers were prepared by 10 mg/mL protein into assembly buffer. The buffer was given first with 200 mM β-mercaptoethanol (βME), then 0.2 mM βME, and 20 mM Tris (pH 8). Each step was performed for 8 hours.

Data was taken from the image D line 5 in [60] (shown Figure 5.8 (*right*) in this dissertation) by using the image processing software ImageJ. Each *i*-mer was measured five times to alleviate any discrepancies due to human error. The average of these measurements are used as our equilibrium concentrations.

5.5.2. CONSTRAINTS ON THE FORWARD AND BACKWARD RATES. Before using the PSO toolbox to optimize the parameters, an initial guess $P(1)$ must be contributed. The choice of PSO parameters can have a large impact on optimization performance. The following size order relations on the forward and backward rates help find a good initial guess and set bounds for each parameter.

Constraints on the Forward Rates. The models presented in [21, 37, 85] assume that only one protein is added (could associate) at a time and all forward rates are equivalent. In [52], it is assumed f_n (equivalent to f_{1n} in our model) increases monotonically with n. In [59], it is found that monomers assemble spontaneously into a hexamer lattice tube,
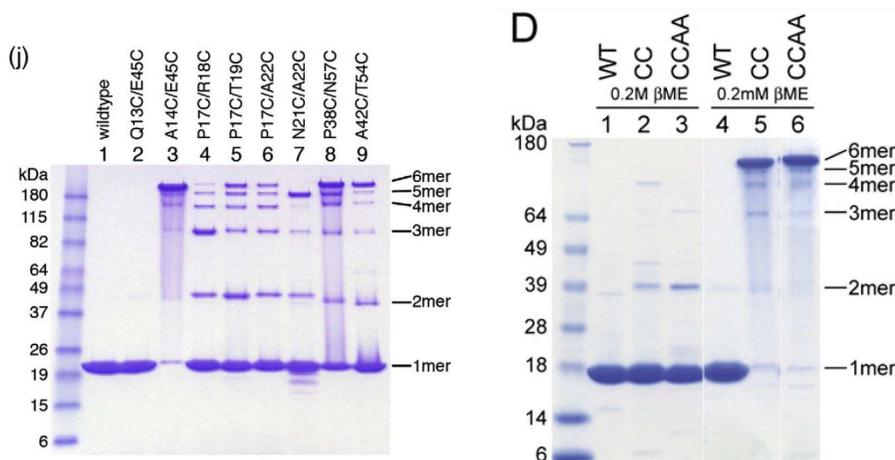
FIGURE 5.8. *Left:* SDS-PAGE profiles of the assembly reactions [59] (reprinted with permission from Elsevier). *Right:* WT stands for wild type, CC corresponds to A14C/E45C, and CCAA is A14C/E45C/W184A/M185A [60] (reprinted with permission from Elsevier).

indicating that the CA proteins tend to form hexamers. Based on these studies, we assume the forward rates f_{1n} increase as the size of the intermediate increases.

It is expected for f_{11} to be very small, since the subunit-subunit interactions are inherently weak [37, 79]. The pentamer subunit is the least stable intermediate, so f_{15} will be very large compared to the others [79].

We adopt a similar size order relation as seen in [52], excluding the rates which may not react the same in our model due to the addition of binding intermediates:

$$(88) \qquad\qquad f_{11} \leq f_{12} \ll f_{15}.$$

[79] discusses the stability of intermediates and claims that a hexamer is more stable than a tetramer and a tetramer is more stable than a pentamer. We assume that stability helps drive intermediate formation and state

$$(89) \qquad\qquad f_{22} \leq f_{24} \ll f_{15}.$$

For the reduced nucleation model presented in this paper, all the forward rates except f_{222} have the physical dimension $T^{-1}L^3M^{-1}$, where T is time given in seconds, L^3 in milliliters cubed, and M in milligrams. The forward rate f_{222} (for trimer-of-dimer) is the only rate that has a physical dimension $T^{-1}(L^3M^{-1})^2$. It cannot be simply compared to the other forward rates. [16] notes that the trimer-of-dimers structure is crucial for lattice formation, and [13, 29] found hexameter formation occurs with increased CA dimer concentration, so it is expected f_{222} to be large.

Constraints on Backward Rates. All the backward rates have the physical dimension T^{-1}.

The discussion in [13, 16, 29] implies that it is less likely for a dimer to dissociate. We assume that b_{21} will be the smallest backward rate. Additionally, the instability of pentamers [79] implies that the rate of b_{65} should be low compared to that of other hexamer disassociations. These lead to the assumptions that

$$b_{21} \le b_{65} \le b_{64}, \tag{90}$$

$$b_{21} \le b_{65} \le b_{62}. \tag{91}$$

5.5.3. RESULTS OF MODEL PARAMETER FITTING. We perform parameter fitting, using the PSO Toolbox, for our reduced 6-species model based on the discussion in [59, 60, 79] about HIV-1 hexamer formation and the experimental data reported in [60].

The initial guess and bounds are constructed using the relationships defined in the above section. The PSO toolbox solves the ODE system with these parameters bounds with the additional condition that the chosen parameter values should produce a solution with concentrations close to those measured from the experimental data in Section 5.5.1. PSO is run 10 times due to the randomness involved in Equation (80). Weights are set to the conventional values, with $c_1 = c_2 = 2$ and $w = 0.9$. Iterations are terminated after the max number of iterations ($i = 2000$) or by achieving the minimum global error gradient

$$|g(i+1) - g(i)| < 1 \times 10^{-25}. \tag{92}$$

We choose the set of parameters that minimize the error between the experimental data and the numerical solution. The optimized parameters yield the lowest relative error (0.0125) are listed in Table 5.1 and ODE solution is compared to the data in Figure 5.9.

5.5.4. RESULTS OF MULTIMER CONCENTRATIONS $(c_1, c_2, c_3, c_4, c_5, c_6)$. The stability of equilibria for this model were considered. First, we reduced the system according to the mass conservation law and our initial condition $\vec{c}0 = (1300, 0, 0, 0, 0, 0)$, which states

$$(93) \qquad\qquad c_1 + 2c_2 + 3c_3 + 4c_4 + 5c_5 + 6c_6 = 1300.$$

TABLE 5.1. Optimal Parameter Values Chosen for Discussion.

$f_{11} = 0.000556$	$f_{12} = 0.004504$	$f_{13} = 0.000867$	$f_{14} = 0.038226$
$f_{15} = 0.179675$	$f_{22} = 0.013196$	$f_{222} = 0.159765$	$f_{24} = 0.061905$
$b_{65} = 0.193838$	$b_{64} = 0.256905$	$b_{62} = 0.993826$	$b_{54} = 0.056015$
$b_{43} = 0.728455$	$b_{42} = 0.719905$	$b_{32} = 0.717905$	$b_{21} = 0.019094$
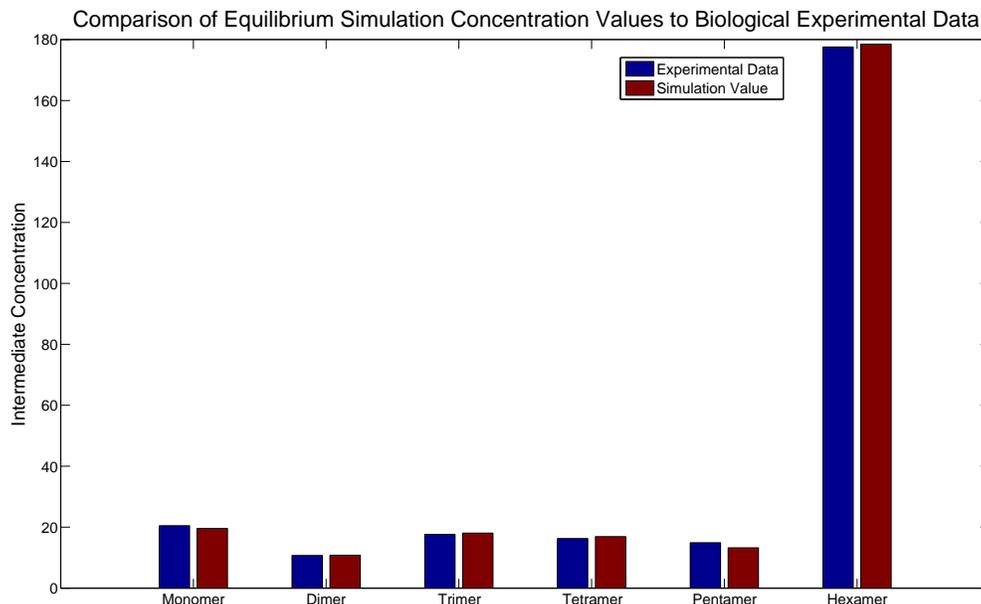
FIGURE 5.9. Concentrations of the intermediates after $t = 24 \times 3600$ seconds, with initial value $(c_1(0), c_2(0), c_3(0), c_4(0), c_5(0), c_6(0)) = (1300, 0, 0, 0, 0, 0)$. ODE solutions with optimized parameters are shown in dark red, data from [60] are shown in dark blue.

Using the optimized parameters, the equilibria of the mass conserving model are found using the solve function in MATLAB. Seventeen solutions were found, though the negative and imaginary equilibrium points are discarded since they are not biologically relevant. This reduces the number of physically possible equilibrium points to one. The Jacobian of the system is computed and evaluated at the equilibrium point. Each eigenvalue is found to have a negative real part, implying that the equilibrium point shown in Figure 5.10 is stable.

The monomer concentration c_1 quickly decreases as the CA proteins bind with c_i concentrations to form c_{i+1} intermediates. Note that there is a large initial spike in the dimer concentration c_2, implying many monomer proteins bind together to form dimers first, as discussed in [7, 13, 27]. The quick decrease in c_2 indicates the importance of the dimers in
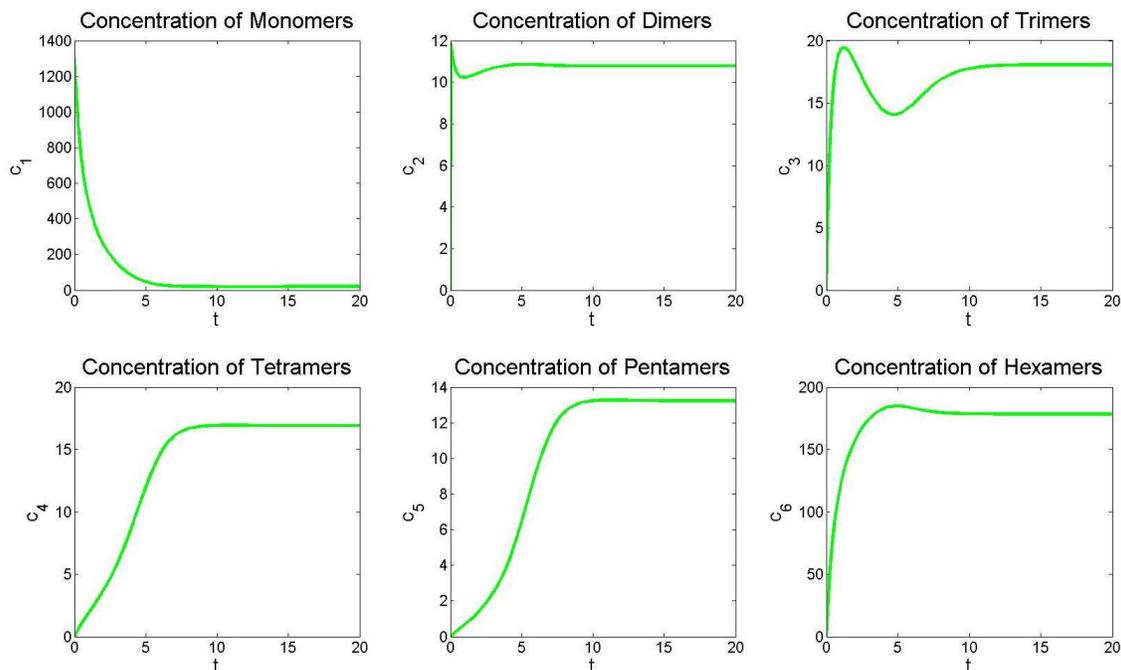
FIGURE 5.10. Simulation results. Concentrations of each intermediate c_n from $t = 0$ to $t = 20$ seconds, with $\vec{c}(0) = (c_1(0), c_2(0), c_3(0), c_4(0), c_5(0), c_6(0)) = (1300, 0, 0, 0, 0, 0)$. Simulations were performed until $t = 24 \times 3600$, although they are not shown here due to early convergence of the solution. Convergent concentrations agree with the experimental data in [60] as shown in Figure 5.9.

building higher order n-mers. It is interesting to see a spike in the trimer concentration c_3, decrease, then gradually increase to equilibrium. This will be addressed in the embedded modeling section. Furthermore, the concentrations $c_n(n = 4, 5, 6)$ are gradually increasing as expected.

5.5.5. RESULTS OF ELASTICITY ANALYSIS. Sensitivity and elasticity analysis is performed for the concentration of n-mer c_n (n=1,2,3,4,5,6) with respect to the association and dissociation rates (forward and backward rates) using the SENSAI MATLAB package [72]. There are a total of 16 forward and backward rates, as shown in Figures 5.11 and 5.12.

The sensitivity of parameters to intermediate concentrations is first considered. These parameter values (see Table 5.1) vary along three orders of magnitude indicating that a scaling of the parameter values is necessary and elasticity may be a more appropriate choice for analysis.

Sensitivity is quantified as a derivative. For six concentrations $c_i, i = 1, .., 6$ and sixteen parameters $p_k, k = 1, ..., 16$, a total of 96 derivatives are calculated over time. A scaling is then applied as shown in Equation (85) to define the elasticity.

We first look at the elasticity of parameters to concentrations. Elasticity is considered at the following times: $t = 1 \times 10^{-5}, 0.03, 0.1, 1, 2, 4, 7, 12$. We consider the values at $t = 12$ to be equilibrium values. There are rapid changes in the concentration of monomers for $t < 1$ and so we consider elasticity at three other times before $t = 1$, then three other times after but before equilibrium.

The elasticity results tell an expected story. Near the beginning (Figure 5.11), concentrations are most elastic to the forward rates, especially f_{11}. This is intuitive since the c_1 concentration is rapidly decreasing as the monomers are forming into dimers and trimers,
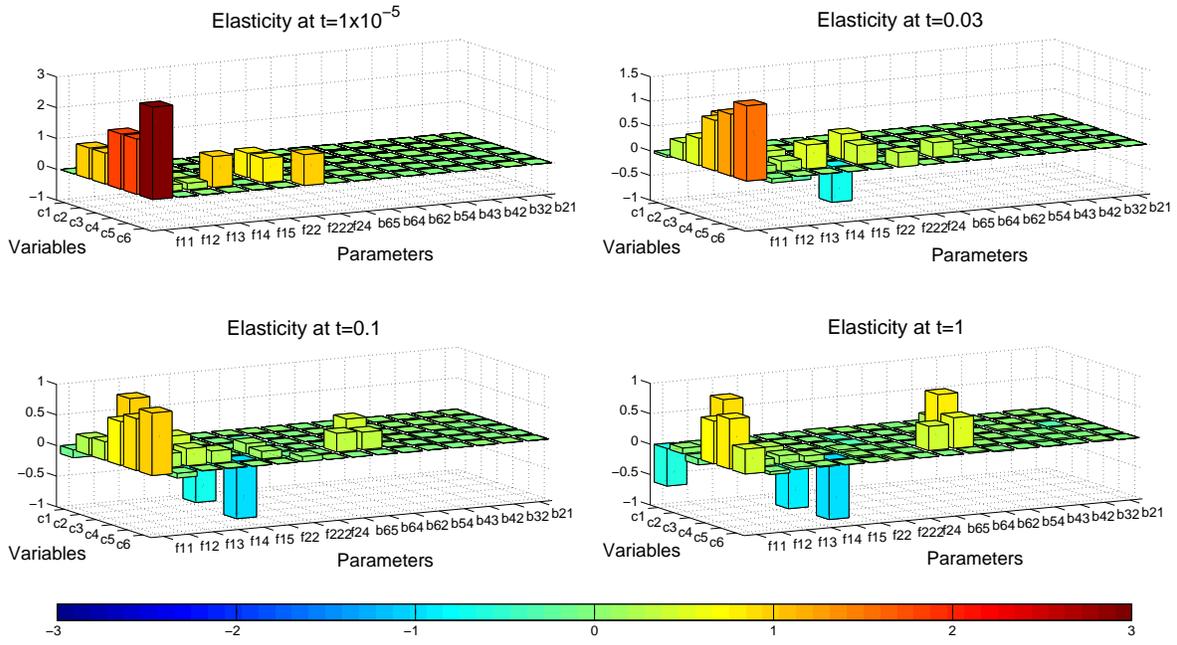
FIGURE 5.11. Elasticity of the n-mer concentration c_n with respect to the association and dissociation rates are evaluated at four times: $t = 1 \times 10^{-5}, 0.03, 0.1, 1$.
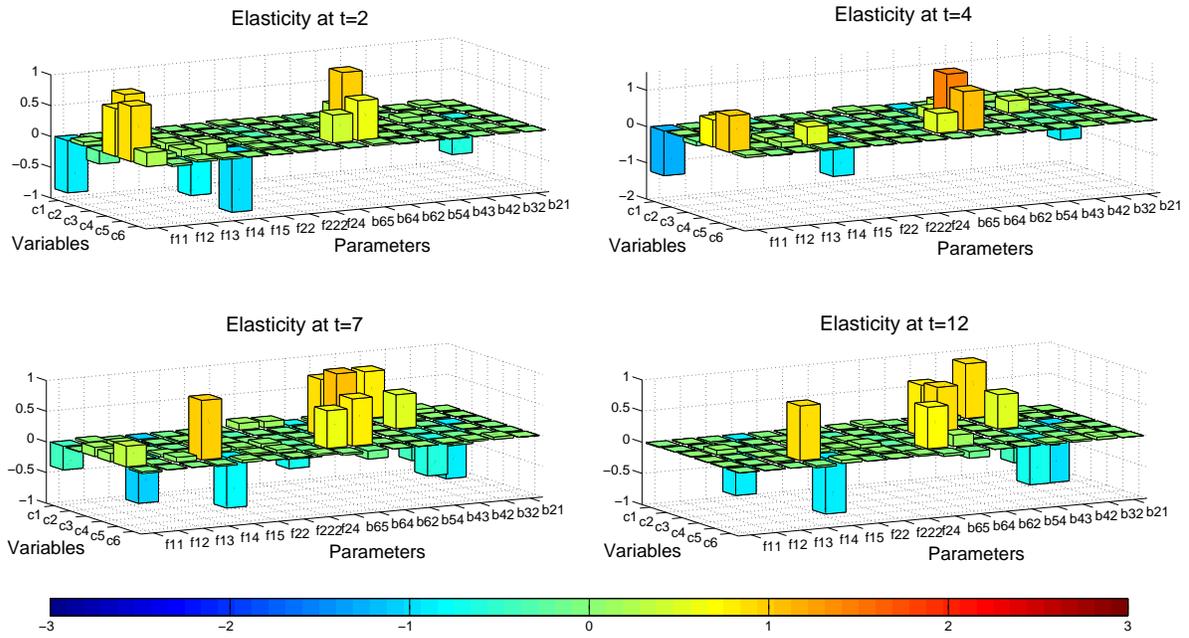

FIGURE 5.12. Elasticity of the n-mer concentration c_n with respect to the association and dissociation rates are evaluated at four times: $t = 2, 4, 7, 12$.

as demonstrated in the spikes of c_2 and c_3 concentrations in Figure 5.10. As time increases, concentrations become less elastic to these forward rates and become more elastic towards higher intermediate forward rates, such as f_{14} and f_{15} as show in Figure 5.11 (*bottom*).

There is an equivalent increase in elasticity to the backward rates (Figure 5.12). It is interesting to note that elasticity to parameters b_{65} and b_{64} appear first out of the backward rates, (Figure 5.11), and remain evident throughout the rest of the time period. Since hexamers are assumed to be the most stable intermediate, these results could provide insight on when hexamers disassemble.

Elasticity to association rates f_{1i}, $i = 1, ..., 6$.   The monomer concentration c_6 shows the largest elasticity to the forward rate f_{11} at the start of nucleation. Other concentrations also show elasticity to f_{11} at times, as expected since f_{11} is the parameter needed for nucleation to begin. These elasticities decrease as time increases, except for concentrations c_1, c_4 where some fluctuation is seen (see Figure 5.11 for c_1 and Figure 5.12 for c_1, c_4). All other intermediate concentrations follow a similar decreasing in magnitude pattern for each forward rate f_{12}, starting from larger elasticity then decreasing over time.

Elasticity of c_5 to f_{14} is seen at the start (Figure 5.11), dissipates, then gradually increases as time approaches equilibrium (Figure 5.12). Concentration c_5 also shows consistent elasticity towards parameter f_{15}, implying these two forward rates f_{14}, f_{15} (and therefore pathway listed in Equation (74)) may be important in the assembly of a pentamer and hexamer. Minimal elasticity is seen for any concentration with respect to f_{13}.

Elasticity to association rates f_{22}, f_{222} **and** f_{24}.   Concentrations c_3, c_4 both show elasticity towards parameter f_{22} at the start of nucleation (Figure 5.11) then elasticity decreases as time increases. A similar pattern is seen for c_6 with respect to f_{222} as time

approaches equilibrium. These results could lead to insights on how important the dimer intermediate is during assembly (pathways listed in Equations (75) and (76)). The elasticity of concentrations to parameter f_{24} is minimal.

Elasticity of parameters to the backward rates. As shown in Figures 5.11 and 5.12, the magnitude of elasticities with respect to the backward rates tend to increase as the magnitude of elasticities with respect to the forward rates decrease. Elasticity to backward rates b_{65} appears first (Figure 5.11) and stays evident as time increases. Concentration c_3 has consistent elasticity past $t = 2$ and c_4 has consistent elasticity with respect to b_{43} from $t = 7$ to equilibrium. These results reemphasize that higher order concentrations may prefer to disassemble one monomer at a time.

Concentrations c_4 and c_5 show elasticity to parameter b_{64}. This is expected for c_4, since the backward rate b_{64} is representative of a hexamer disassociating into a tetramer and dimer. The elasticity for c_5 with respect to b_{64} may be indicative of a pentamer being integrated into the lattice, from a hexamer, as discussed in [79]. Minimal elasticity is seen for any concentration with respect to parameters $b_{62}, b_{54}, b_{42}, b_{21}$.

5.5.6. MODEL SENSITIVITY & EMBEDDED MODELS. Consistent low elasticity over time could imply that certain parameters are not important in the model for capsid nucleation. These parameters may not give additional or important information in our model. To test this claim, embedded models are analyzed to further characterize which parameters are most important for recreating the dynamics seen in biological experiments. Parameters with low elasticity are removed from the model, one at a time, to analyze its importance in the model. The parameter will be deemed important only if the equilibrium solution changes or the time to equilibrium changes drastically.

Since elasticities of concentrations with respect to parameters varies significantly during the transient phase, the largest magnitude of the elasticity for every concentration c_n with respect to parameter p_k for $0 < t < 200$ is shown in Figure 5.13. We look for parameters with low elasticity for all concentrations c_n. The parameters of question are taken to be $f_{13}, f_{24}, b_{62}, b_{54}, b_{42}$, and b_{21}.

Each parameter is removed from the model, one at a time. The ODE system is then reduced and resolved. Equilibrium solution is evaluated and the relative error between the new equilibrium (X_r) and the original model equilibrium (X) is calculated. The results from the embedded modeling are listed in Table 5.2. It is found that parameters f_{13}, b_{54}, b_{21} can be eliminated from the model individually with a negligible change to the equilibrium concentrations.

Then, the process was repeated by removing sets of parameters. The relative error of removing parameter sets are listed in Table 5.3. It holds that all three parameters f_{13}, b_{54}, b_{21}
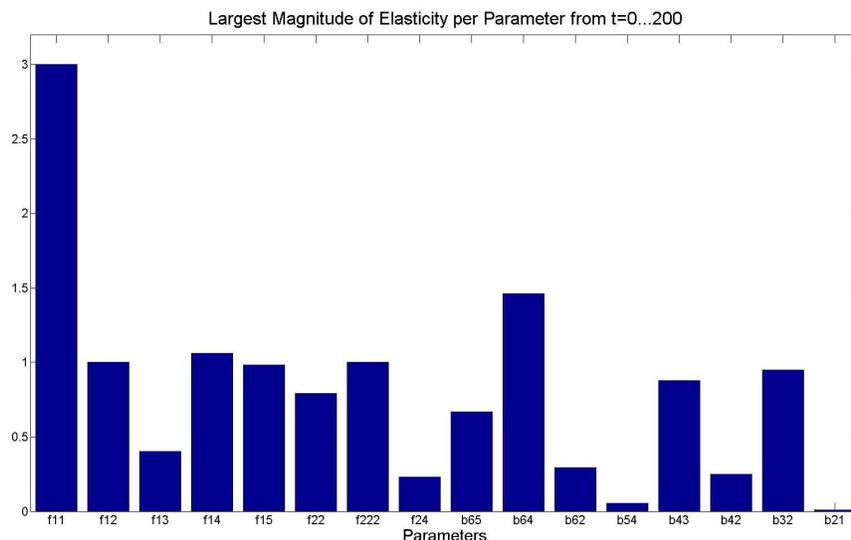

FIGURE 5.13. Largest magnitude of elasticity over all time of n-mer concentrations c_n with respect to the parameters (represented by the magnitude of the derivative). Low elasticity is seen for parameters $f_{22}, f_{24}, b_{62}, b_{54}, b_{42}$ and b_{21}.

86

TABLE 5.2. Relative Error by Removing Individual Parameters.

Parameters	f_{13}	f_{24}	b_{62}	b_{54}	b_{42}	b_{21}
$\frac{\|\|X_r - X\|\|}{\|\|X\|\|}$	0.0034	0.0479	0.0314	0.0075	0.0537	0.0020

TABLE 5.3. Relative Error by Removing Parameter Sets.

Parameters	f_{13}, b_{54}	f_{13}, b_{21}	b_{54}, b_{21}	f_{13}, b_{54}, b_{21}
$\frac{\|\|X_r - X\|\|}{\|\|X\|\|}$	0.0048	0.0021	0.0095	0.0068

can be eliminated from the model simultaneously with a negligible change to the equilibrium concentrations. This implies these parameters may not be important for nucleation.

Discussion. By removing parameters b_{54}, b_{21} concentrations c_5, c_2 are no longer able to disassociate in this model. Similarly, by removing parameter f_{13} there is only one pathway for a tetramer assembly, given by pathway listed in Equation (76), by two dimers. It is interesting to note that all three of these parameters are found only in the traditional pathway, Equation (74), used in previous models [31, 85]. Removal of these parameters disrupts this pathway. Calculating the probability of each pathway would be insightful to the usefulness of the traditional pathway compared to the two novel pathways presented in this chapter for hexamer assembly: single binding dimers and the trimer-of-dimer pathway (Equations (75) and (76)).

5.6. REMARKS: BIOLOGICAL IMPLICATIONS

In summary, it is quite interesting to note that although no concentrations are elastic with respect to the parameters removed in the final model (f_{13}, b_{54}, b_{21}), the corresponding backward/forward rates b_{43}, f_{15}, f_{11} seem to be important in this model.

Importance of CA Dimers. This chapter focuses on the nucleation stage of viral capsid assembly. It is different than the existing work [21, 31, 52] that consider mainly one pathway and add/delete one capsomer unit at a time. Our model considers more pathways

for association and dissociation and provides more information about the assembly. It is now revealed by the model that CA dimers indeed play an important role in the nucleation stage, as reflected in the initial spike in the numerical simulations and analysis showing that f_{22}, f_{24}, f_{222} are important parameters for HIV-1 nucleation. This agrees with the findings in [7, 13, 27, 74].

Model Predictability. Parameters f_{11}, f_{12}, b_{64} exhibit elasticity in the monomer and hexamer concentrations c_1, c_6. These association or dissociation rates correspond respectively to two monomers forming a dimer, a monomer and dimer producing a trimer, and a hexamer breaking apart into a tetramer and dimer. Looking at the elasticity at different times also gives insight on when each pathway is the most important. After the initial spike of dimers, the intermediates become more sensitive to f_{222} implying the importance of three dimers forming a hexamer. These results imply that the most important pathways for hexamer formation are single monomers joining together and triple binding dimers (pathways in Equations (74) and (75) discussed in Section 5.3.3). These results demonstrate that our model has predictability to a certain level.

CHAPTER 6

FURTHER WORK ON MODELING VIRAL CAPSID ASSEMBLY

6.1. CURVATURE AND BEYOND

The curvature characterization presented in this thesis provides a new viewpoint and could be useful to the field of structural virology, especially in ranking the stability of related capsids for pleomorphic capsids such as HIV. The conjecture that capsid stability is inversely related to curvature concentration is worth of further exploration, given the matching observations in HIV electron micrographs [6, 8, 11, 23].

More Experimental Data Desired for Other Types of HIV-1 Conical Cores. The majority of the existing work on the HIV-1 cone model focuses on the (5,7)-pattern, even though (4,8)-cones have been observed in experiments [11]. The results in this dissertation show that the (4,8)-cone has also high contrast of curvature concentrations for the narrow and broad ends. It should be interesting to examine possibility and stability (or instability) of formations for different types of cones. It will be helpful for modeling research on HIV-1 cone structure if more detailed and specific experimental data on different types of cones are available. Typically, both the (5,7)- and the (4,8)-cone statistics get grouped together during averaging [6, 11]. Yet having independent information on the (4,8)- and other types of cones will help identify favorable or unfavorable conditions for formation of HIV-1 cone structure.

Relation of Curvature to Elastic Energy. The modeling research presented in this dissertation shows that the narrow end of the HIV-1 conical core has the highest curvature concentration for the (5,7)- and (4,8)-patterns. This high curvature concentration is tightly related to the stress and elastic (bending) energy at the narrow/broad ends. Similar

conjectures have been analyzed to find that the degree of localization of Gaussian curvature depends on elastic stiffness [33]. Also, the role of nonzero spontaneous curvature drives the energy of the 12 pentameric declinations producing the shape of the capsid [54, 53]. This is further related to the binding of CTD, NTD of the CA protein. Studying the excess energy shall further shed light on HIV-1 cone structure [67].

Capsids of Other Retroviruses. The concepts and methodology presented in this dissertation could be applied to other types of viral capsids, e.g., murine leukemia virus (MuLV) and Rous sarcoma virus (RSV) [32]. The derivation for generating vectors on lattice structures can be extended to other retroviral capsid shapes. This information could then be utilized to study the elastic energy on the capsid and the relationship between curvature and elastic energy. This is a direction for further research.

6.2. Further Modeling of Viral Capsid Nucleation

Biological experiments indicate that separate modeling and simulations of the nucleation and elongation stages shall help bring in different perspectives for modeling viral capsid assembly. Simplifying the models to study nucleation and elongation separately allows us to closely examine the favorable and unfavorable conditions for each stage. Possible topics for future research on HIV-1 capsid nucleation are identified as follows.

Stochastic Dynamical Systems. Clearly, there exists randomness in the nucleation stage of viral capsid assembly. The temperature, pH-value, and many other factors in the environment of assembly affect the association and dissociation rates and hence the formation of CA hexameters and pentamers. Further research includes investigation of the stochastic features of nucleation in which stochastic dynamical systems will be an indispensable tool.

Embedded Modeling. The models established in this dissertation may be modified and used to further explore the importance of the added pathways. The models should be compared to the original full model formulated in Chapter 5. Embedded models with a statistically significant difference in model parameters, sensitivity, or equilibrium to the full model, will indicate which parameters need to be further examined. This will also shine light on which pathways could be important for describing nucleation.

6.3. Modeling of Viral Capsid Elongation

Retrovirus capsid assembly has been observed to undergo two stages during maturation: nucleation and elongation. The investigation of nucleation cannot be completely isolated from the whole process of viral capsid assembly. There have been kinetic models for full viral capsid assembly, though these models describe simple icosahedron-shaped capsids [21, 30, 31, 85]. Their extension to retrovirus has been over simplified, due to the complicated shape of a retroviral capsid.

Cascaded Stochastic Dynamical Systems (CSDS). It is our postulation that at the early stage of viral capsid assembly, hexamer formation happens simultaneously in many locations within the virion. Then these hexamers further assemble into the viral capsid. Pentamers might form at the places where it is difficult for a hexamer to form. This is the elongation stage. In other words, the products of nucleation serves as a feeding for the elongation stage. Research involving cascade of kinetics and cascaded stochastic dynamical systems (CSDS) shall be an exploratory tool for further investigation.

Role of viral RNA in assembly. Many single-stranded RNA viruses, such as HIV, self-assemble their capsids around their genomes. The roles that the RNA plays in this assembly process have mostly been ignored. Data from molecular approaches suggest there

is a strong interaction between the genomic RNA and the CA proteins [75]. The RNA may guide the capsid into a conformation that is compatible with the underlying geometry of the virion. This is a direction for further mathematical modeling of the capsid assembly.

Bibliography

1. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell*, Garland Science, (2008).

2. A. Alfadhli, D. Huseby, E. Kapit, D. Colman, E. Barklis, *Human immunodeficiency virus type 1 matrix protein assembles on membranes as a hexamer*, J. Virol., **81**(2007), pp. 1472–1478.

3. G.D. Bailey, J.K. Hyun, A.K. Mitra, R.L. Kingston, *A structural model for the generation of continuous curvature on the surface of a retroviral capsid*, J. Mol. Biol., **417**(2012), pp. 212–223.

4. M. Balasubramaniam, E.O. Freed, *New insights into HIV assembly and trafficking*, Physiology, **26**(2011), pp. 236-251.

5. J.E. Baschek, H.C.R. Klein, U.S. Schwarz, *Stochastic dynamics of virus capsid formation: direct versus hierarchical self-assembly*, BMC Biophys., **5**(2012), pp. 1–18.

6. J. Benjamin, B.K. Ganser-Pornillos, W.F. Tivol, W.I. Sundquist, G.J. Jensen, *Three-dimensional structure of HIV-1 virus-like particles by electron cryotomography*, J. Mol. Biol., **346**(2005), pp. 577–588.

7. L. Briant, B. Gay, C. Devaux, N. Chazal, *HIV-1 assembly, release, and maturation*, World J. of AIDS, **1**(2011), pp. 111–130.

8. J.A.G. Briggs, K. Grünewald, B. Glass, F. Förster, H-G. Kräusslich, S.D. Fuller, *The mechanism of HIV-1 core assembly: Insights from three-dimensional reconstructions of authentic virions*, Structure, **14**(2006), pp. 15–20.

9. J.A.G. Briggs, H-G. Kräusslich, *The molecular architecture of HIV*, J. Mol. Biol., **410**(2011), pp. 491–500.

10. J.A.G. Briggs, J.D. Riches, B. Galss, V. Bartonova, G. Zanetti, H.-G. Kräusslich, *Structure and assembly of immature HIV*, PNAS, **106**(2009), pp. 11090–11095.

11. J.A.G. Briggs, T. Wilk, R. Welker, H-G. Kräusslich, S.D. Fuller, *Structural organization of authentic, mature HIV-1 virions and cores*, EMBO J., **22**(2003), pp. 1707–1715.

12. C. Büchen-Osmond, ICTVdB - The universal virus database, Version 4, ICTVdB Management, New York, 2006.

13. I.L. Byeon, X. Meng, J. Jung, G. Zhao, R. Yang, J. Ahn, J. Shi, J. Concel, C. Aiken, P. Zhang, A.M. Gronenborn, *Structural convergence between cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function*, Cell, **139**(2009), pp. 780–790.

14. D.L. Caspar, A. Klug, *Physical principles in the construction of regular viruses*, Cold Spring Harb. Symp. Quant. Biol., **27**(1962), pp. 1–24.

15. X. Chen, K. Gerasopoulos, J. Guo, A. Brown, C. Wang, R. Ghodssi, J.N. Culver, *Virus-enabled silicon anode for lithium-ion batteries*, ACS Nano., **4**(2010), pp. 5366–5372.

16. B. Chen, R. Tycko, *Simulated self-assembly of the HIV-1 capsid: Protein shape and native contacts are sufficient for two-dimensional lattice formation*, Biophys. J., **100**(2011), pp. 3035–3044.

17. T. Douglas, M. Young, *Virus: Making friends with old foes*, Science, **312**(2006), pp. 873–875.

18. R.C. Eberhart, J. Kennedy, *A new optimizer using particle swarm theory*, Proc. 6th Intl. Symp. Micro Machine & Human Sci., (1995), pp. 39–43.

19. L.S. Ehrlich, T. Liu, S. Scarlata, B. Chu, C.A. Carter, *HIV-1 capsid protein forms spherical (immature-like) and tubular (mature-like) particles in vitro: Structure switching by pH-induced conformational changes*, Biophys. J., **81**(2001), pp. 586–594.

20. O.M. Elrad, M.F. Hagan, *Encapsulation of a polymer by an icosahedral virus*, Phys. Biol., **7**(2010), pp. 045003:1–17.

21. D. Endres, A. Zlotnick, *Model-based analysis of assembly kinetics for virus capsids or other spherical polymers*, Biophys. J., **83**(2002), pp. 1217–1230.

22. A. de la Escosura, R.J.M. Nolte, J.J.L.M. Cornelissen, *Viruses and protein cages as nanocontainers and nanoreactors*, J. Mater. Chem., **19**(2009), pp. 2274–2278.

23. B.K. Ganser-Pornillos, A. Cheng, M. Yeager, *Structure of full-length HIV-1 CA: A model for the mature capsid lattice*, Cell, **131**(2007), pp. 70–79.

24. B.K. Ganser, S. Li, V.Y. Klishko, J.T. Finch, W.I. Sundquist, *Assembly and analysis of conical models for the HIV-1 core*, Science, **283**(1999), pp. 80–83.

25. B.K. Ganser-Pornillos, U.K. von Schwedler, K.M. Stray, C. Aiken, W.I. Sundquist, *Assembly properties of the human immunodeficiency virus type 1 CA protein*, J. Virol., **78**(2004), pp. 2545–2552.

26. B.K. Ganser-Pornillos, M. Yeager, O. Pornillos, *Assembly and architecture of HIV*, Adv. Exp. Med. Biol., **726**(2012), pp. 441–465.

27. B.K. Ganser-Pornillos, M. Yeager, W.I. Sundquist, *The structural biology of HIV assembly*, Curr. Opin. Stuct. Biol., **18**(2008), pp. 203–217.

28. Y. Gogotsi, *Nanomaterial Handbook*, CRC Press, (2006).

29. J.M. Grime, G.A. Voth, *Early stages of the HIV-1 capsid protein lattice formation*, Biophys. J., **103**(2012), pp. 1774–1783.

30. M.F. Hagan, *Modeling viral capsid assembly*, Adv. Chem. Phys., **155**(2014), pp. 1–68.

31. M.F. Hagan, O. Elrad, *Understanding the concentration dependence of viral capsid assembly kinetics – the origin of the lag time and identifying the critical nucleus size*, Biophys. J, **98**(2010), pp. 1065–1074.

32. J.B. Heymann, C. Butan, D.C. Winkler, R.C. Craven, A.C. Steven, *Irregular and semi-regular polyhedral models for Rous Sarcoma virus cores*, Comput. Math. Meth. Med., **9**(2008), pp. 197–210.

33. S.D. Hicks, C.L. Henley, *Irreversible growth model for virus capsid assembly*, Phys. Rev. E, **74**(2006), pp. 031912:1–17.

34. R.W. Horne, P. Wildy, *Symmetry in virus architecture*, Virology, **15**(1961), pp. 348–373.

35. Y. Hu, R. Zandi, A. Anavitarte, C.M. Knobler, W.M. Gelbart, *Packaging of a polymer by a viral capsid: The interplay between polymer length and capsid size*, Biophys. J., **94**(2008), pp. 1428–1436.

36. L. Huang, C. Chen, *Understanding HIV-1 protease autoprocessing for novel therapeutic development*, Future Med. Chem., **5**(2013), pp. 1215–1229.

37. S. Katen, A. Zlotnick, *The thermodynamics of virus capsid assembly*, Meth. Enzymol., **455**(2009), pp. 395–417.

38. J. Lanman, T.T. Lam, S. Barnes, M. Sakalian, M.R. Emmett, A.G. Marshall, P.E. Preveige, *Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry*, J. Mol. Biol., **325**(2003), pp. 759–772.

39. S. Li, C.P. Hill, W.I. Sundquist, J.T. Finch, *Image reconstructions of helical assemblies of the HIV-1 CA protein*, Nature, **407**(2000), pp. 409–413.

40. Y. Li, M. Yi, X. Zou, *Identification of the molecular mechanisms for cell-fate selection in budding yeast through mathematical modeling*, Biophys. J., **104**(2013), pp. 2282-2294.

41. J. Liu, R. Munoz-Alicea, T. Huang, S. Tavener, C. Chen, *A mathematical model for intracellular HIV-1 gag protein transport and its parallel numerical simulations*, Proc. Comput. Sci., **9**(2012), pp. 679–688.

42. J. Liu, F. Sadre-Marandi, S. Tavener, C. Chen, *Curvature concentrations on the HIV-1 capsid*, Mol. Based Math. Biol., 2015 (accepted).

43. A. Luque, D. Reguera, *The structure of elongated viral capsids*, Biophys. J., **98**(2010), pp. 2993–3003.

44. R.V. Mannige, C.L. Brooks III, *Periodic table of virus capsids: Implications for natural selection and design*, PLoS One, **5**(2010), pp. e9423:1–7.

45. R.V. Mannige, C.L. Brooks III, *Geometric considerations in virus capsid size specificity, auxiliary requirements, and buckling*, PNAS, **106**(2009), pp. 8531–8536.

46. R.V. Mannige, C.L. Brooks III, *Tilable nature of virus capsids and the role of topological constraints in the natural capsid design*, Phys. Rev. E, **77**(2008), pp. 051902:1–8.

47. N.W. Martinez, X. Xue, R.G. Berro, G. Kreitzer, M.D. Resh, *Kinesin KIF4 regulates intracellular trafficking and stability of the human immunodeficiency virus type 1 gag polyprotein*, J. Virol., **82**(2008), pp. 9937–9950.

48. E.R. May, J. Feng, C.L. Brooks III, *Exploring the symmetry and mechanism of virus capsid maturation via an ensemble of pathways*, Biophys. J., **102**(2012), pp. 606–612.

49. K. Mayo, D. Huseby, J. McDermott, B. Arvidson, L. Finlay, E. Barklis, *Retrovirus capsid protein assembly arrangements*, J. Mol. Biol., **325**(2003), pp. 225–237.

50. M.F. Moody, *Geometry of phage head construction*, J. Mol. Biol., **293**(1999), pp. 401–433.

51. M.F. Moody, *The shape of the T-even bacteriophage head*, Virology, **26**(1965), pp. 567–576.

52. R. Munoz-Alicea, *HIV-1 gag trafficking and assembly: Mathematical models and numerical simulations*, Ph.D. Dissertation, Colorado State University, (2013).

53. T.T. Nguyen, R.F. Bruinsma, W.M. Gelbart, *Continuum theory for retroviral capsids*, Phys. Rev. Lett., **96**(2006), pp. 078102:1–4.

54. T.T. Nguyen, R.F. Bruinsma, W.M. Gelbart, *Elasticity theory and shape transitions of viral shells*, Phys. Rev. E, **72**(2005), pp. 051923:1–19.

55. C.C. Payne, *Insect viruses as control agents*, Parasitology, **84**(1982), pp. 35–77.

56. A.S. Perelson, P.W. Nelson, *Mathematical analysis of HIV-1: Dynamics in vivo*, SIAM Review, **41**(1999), pp. 3–44.

57. J.K. Pokorski, N.F. Steinmetz, *The art of engineering viral nanoparticles*, Mol. Pharma., **8**(2011), pp. 29–43.

58. PDB ID: 3H47

    O. Pornillos, B.K. Ganser-Pornillos, B.N. Kelly, Y. Hua, F.G. Whitby, C.D. Stout, W.I. Sundquist, C.P. Hill, M. Yeager, *X-ray structures of the hexameric building block of the HIV capsid*, Cell, **137**(2009), pp. 1282–1292.

59. O. Pornillos, B.K. Ganser-Pornillos, S. Banumathi, Y. Hua, M. Yeager, *Disulfide bond stabilization of the hexameric capsomer of human immunodeficiency virus*, J. Mol. Biol., **401**(2010), pp. 985–995.

60. O. Pornillos, B.K. Ganser-Pornillos, B.N. Kelly, Y. Hua, F.G. Whitby, C.D. Stout, W.I. Sundquist, C.P. Hill, M. Yeager, *X-ray structures of the hexameric building block of the HIV capsid*, Cell, **137**(2009), pp. 1282–1292.

61. O. Pornillos, B.K. Ganser-Pornillos, M. Yeager, *Atomic-level modelling of the HIV capsid*, Nature, **469**(2011), pp. 424–427.

62. A. Pressley, *Elementary Differential Geometry*, 2nd ed., Springer, (2010).

63. P.H. Raven, G.B. Johnson, J.B. Losos, K.A. Mason, S.R. Singer, *Biology*, 8th ed., McGraw Hill, (2008).

64. M.D. Resh, *A myristoyl switch regulates membrane binding of HIV-1 gag*, PNAS, **101**(2004), pp. 417–418.

65. F. Sadre-Marandi, Y. Liu, J. Liu, S. Tavener, X. Zou, *Modeling HIV-1 capsid nucleation by dynamical systems*, Preprint, Submitted to Math. Biosci., (2015).

66. F. Sadre-Marandi, J. Liu, S. Tavener, C. Chen, *Generating vectors for the lattice structures of tubular and conical viral capsids*, Mol. Based Math. Biol., **2**(2014), pp. 128–140.

67. A. Siber, *Continuum and all-atom description of the energetics of graphene nanocones*, Nanotech., **18**(2007), pp. 375705:1–6.

68. V. Simon, D.D. Ho, *Hiv-1 dynamics In Vivo: Implications for therapy*, Nat. Rev. Microbiol., **1**(2003), pp. 181-190.

69. W.I. Sundquist, H.-G. Kräusslich, *HIV-1 assembly, budding, and maturation*, Cold Spring Harbor Lab Press, **2**(2012), pp. a006924:1–24.

70. J. Tan, R. Pan, L. Qiao, X. Zou, Z. Pan, *Modeling and dynamical analysis of virus-triggered innate immune signaling pathways*, PLoS ONE, **7**(2012), pp. e48114:1-15.

71. Y. Tao, N.H. Olson, W. Xu, D.L. Anderson, M.G. Rossmann, T.S. Baker, *Assembly of a tailed bacterial virus and its genome release studied in there dimensions*, Cell, **95**(1998), pp. 431–437.

72. S. Tavener, M. Mickucki, *SENSAI: Software package for sensitivity analysis*,
    `http://www.math.colostate.edu/~tavener/FEScUE/SENSAI/sensai.shtml`

73. S. Tavener, M. Mikucki, S. Field, M.F. Antolin, *Transient sensitivity analysis for nonlinear population models*, Meth. Eco. Evol., **2**(2011), pp. 560–575.

74. B.G. Turner, M.F. Summers, *Structural biology of HIV*, J. Mol. Biol., **285**(1999), pp. 1–32.

75. R. Twarock, *Mathematical virology: a novel approach to the structure and assembly of viruses*, Philos. Trans. A. Math. Phys. Eng. Sci., **364**(2006), pp. 3357–3373.

76. Y. Wang, J. Tang, F. Sadre-Marandi, J. Liu, X. Zou, *Mathematical modeling for intracellular transport and binding of HIV-1 gag proteins*, Math. Biosci., **262**(2015), pp. 198–205.

77. R. Welker, H. Hohenberg, U. Tessmer, C. Huckhagel, H.-G. Kräusslich, *Biochemical and structural analysis of isolated mature cores of human immunodeficiency virus type 1*, J. Vir., **74**(2000), pp. 1168–1177.

78. F.F. Xu, Y. Bando, D. Golberg *The tubular conical helix of graphitic boron nitride*, New J. Phy., **5**(2003), pp. 118:1–16.

79. M. Yeager, *Design of in vitro symmetric complexes and analysis by hybrid methods reveal mechanisms of HIV capsid assembly*, J. Mol. Biol., **410**(2011), pp. 534–552.

80. Z. Yu, M.J. Dobro, C.L. Woodward, A. Levandovsky, C.M. Danielson, V. Sandrin, J. Shi, C. Aiken, R. Zandi, T.J. Hope, G.J. Jensen, *Unclosed HIV-1 capsids suggest a curled sheet model of assembly*, J. Mol. Biol., **425**(2013), pp. 112–123.

81. R. Zandi, D. Reguera, *Mechanical properties of viral capsids*, Phys. Rev. E, **72**(2005), pp. 021917:1–12.

82. R. Zandi, D. Reguera, R.F. Bruinsma, W.M. Gelbart, J. Rudnick, *Origin of icosahedron symmetry in viruses*, PNAS, **101**(2004), pp. 15556–15560.

83. W. Zhang, X. Zou, *Systematic analysis of the mechanisms of virus-triggered type I IFN signaling pathways through mathematical modeling*, IEEE/ACM Trans. Comput. Biol. Bioinform., **10**(2013), pp. 771–779.

84. G. Zhao, J.R. Perilla, E.L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A.M. Gronenborn, K. Schulten, C. Aiken, P. Zhang, *Mature HIV-1 capsid structure by cryo-election microscopy and all-atom molecular dynamics*, Nature, **497**(2013), pp. 643–646.

85. A. Zlotnick, J.M. Johnson, P.W. Wingfield, S.J. Stahl, D. Endres, *A theoretical model successfully identifies features of hepatitis b virus capsid assembly*, Biochem., **38**(1999) pp. 14644–14652.