STRUCTURAL ANALYSIS OF THE SP DOMAIN AND A PUTATIVE SIX-HELIX BUNDLE IN ROUS SARCOMA VIRUS ASSEMBLY

A Dissertation

Presented to the Faculty of the Graduate School

Of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Di Luo Bush

August 2013

STRUCTURAL ANALYSIS OF THE SP DOMAIN AND A PUTATIVE SIX-HELIX BUNDLE IN ROUS SARCOMA VIRUS ASSEMBLY

Di Luo Bush, Ph.D.

Cornell University 2013

Purified retroviral Gag proteins can assemble in vitro to form immature virus-like particles (VLPs). By electron cryo-tomography, Rous sarcoma virus (RSV) VLPs show an organized hexameric lattice consisting chiefly of the capsid (CA) domain, with periodic stalk-like densities below the lattice. I hypothesize that the structure underlying these densities is formed by amino acid residues immediately downstream of the folded CA, namely the short spacer peptide SP, along with a dozen flanking residues. These 24 residues comprise the SP assembly (SPA) domain, and I propose that neighboring SPA units in a Gag hexamer coalesce to form a six-helix bundle. Using in vitro assembly, alanine scanning mutagenesis, and biophysical analyses, I have further characterized the structure and function of SPA. Most of the amino acid residues in SPA could not be mutated individually without abrogating assembly, with the exception of a few residues at the N- and C-termini as well as three hydrophilic residues within SPA. I interpret these results to mean that the amino acids that do not tolerate mutations contribute to higher-order structures in VLPs. Hydrogen-deuterium exchange analyses of unassembled Gag in comparison with assembled VLPs showed strong protection at the SPA region, consistent with a higher-order structure. Circular dichroism revealed that a 29mer SPA peptide shifts from a random coil to a helix in a concentration-dependent manner. Analytical ultracentrifugation showed concentration-dependent self-

association of the peptide into a hexamer. Taken together, these results provide the first convincing evidence for the formation of a critical six-helix bundle (6HB) in RSV Gag assembly. I have modeled the RSV 6HB using structural and sequence homology with an existing synthetic 6HB and discovered a salt-bridge that may help stabilize the helical bundle.

In an attempt to assess the structure of the immature retroviral core, which is made solely of Gag, I also crystallized several proteins containing the minimum regions of Gag required for immature assembly. Due to the inherent flexibility of Gag, which is a multi-domain protein, no crystal structure has been reported thus far of a Gag-like protein. I have shown that it is possible to crystallize a “minimal Gag”, but that the crystals are difficult to optimize to the quality required for X-ray diffraction. I have also crystallized the aforementioned SPA peptide in order to determine the atomic-level structure of the putative 6HB. Unfortunately, my diffraction results show that the crystal packing is disordered. Although much more optimization will likely be required to produce crystals that can diffract to high resolution, my work has shown that it is possible to crystallize both a Gag-like protein and a peptide corresponding to the SP domain.

BIOGRAPHICAL SKETCH

Di was born on March 21st, 1985 in Wuhan, China. At the age of seven, Di immigrated to Montreal, Canada and grew up a Canuck. Di was always fascinated by microbes and illnesses, spending her time reading old medical encyclopedias and growing mold in her parents' fridge "for fun". Owing perhaps to her fondness of maple syrup, lumberjacks and toonies, Di remained in Montreal for her undergraduate education and attended McGill University, where she majored in Microbiology and Immunology in 2007 while working in the laboratory of Dr. Robert Murgita on the structure of α-fetoprotein. Di left Canada to broaden her perspective and to continue her education. She was also trying to find another legitimate excuse for growing microbes. Di enrolled in the field of Biochemistry, Molecular and Cell Biology at Cornell University and met Dr. Volker Vogt, with whom she pursued her Ph.D. studies on retroviral assembly. During her time at Cornell, Di made an earnest effort to stop saying "eh", thereby evading all manners of comments on her northern heritage. She also met the love of her life, Jim Bush, while paddling on Cayuga Lake, with whom she now has a most treasured daughter, Dagny Bush. Following her dissertation, Di will be moving to Birmingham, Alabama with her family for a post-doctoral position with Dr. Peter Prevelige, where she will likely learn various biophysical techniques and study P22 phage assembly. In her free time, Di enjoys outdoor activities, food, and travel. She no longer enjoys hockey, but is willing to give NASCAR a fair chance.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Volker Vogt, for the personal investment he places in teaching students to become better scientists, writers, and human beings than when they first walked through his door, and for leading through example. I could not have asked for a better mentor.

Second, I would like to thank my colleague Robert Dick, for all of the spirited conversations, scientific and otherwise, that we enjoy over coffee. These years in grad school would not have been the same without you.

Third, I would like to thank the undergraduates whom I have had the privilege of working with and mentoring. It was a learning experience for all of us.

Last, but most of all, I would like to thank my husband and daughter for their unwavering support. You are there to motivate me, you are there to comfort me, you are my sunshine. I could not have accomplished any of this without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

25CASP: Gag-derived protein including the CA domain, 25 N-terminal residues and SP domain

25CASP(CC): 25CASP protein with two exogenous cysteines

6HB: six-helix bundle

ΔMBDΔPR: Gag protein lacking the MBD and PR functional domains

AIDS: acquired immunodeficiency syndrome

ALV: avian leukosis virus

Au_{11}: maleimide-undecagold

AUC: analytical ultracentrifugation

BLV : bovine leukemia virus

CA: capsid protein or capsid domain of Gag polyprotein

CA_{CTD}: C-terminal domain of CA

CA_{NTD}: N-terminal domain of CA

Ccmk2: a carboxysome shell protein known to form hexamers

CD: circular dichroism

CHESS: Cornell High Energy Synchrotron Source

CMV: cucumber mosaic virus

ds: double-stranded

EALV: endogenous avian leukosis virus

EAV: endogenous avian virus

EM : electron microscopy

ER : endoplasmic reticulum

ERV: endogenous retrovirus

ESCRT: endosomal sorting complexes required for transport

FIV : feline immunodeficiency virus

HDX: hydrogen-deuterium exchange

IN: integrase protein or integrase domain of Gag-Pol polyprotein

HIV-1: human immunodeficiency virus type 1

HTHP: hexameric tyrosine-coordinated heme protein

HTLV: human T-cell leukemia virus

ISD: internal scaffolding domain

JSRV: Jaagsikte sheep retrovirus

LTR: long-terminal repeat

MA: matrix protein or matrix domain of Gag polyprotein

MBD: membrane-binding domain of the MA protein or MA domain of Gag

MES: 2-(N-morpholino)ethanesulfonic acid

MHR: major homology region

MLV: murine leukemia virus

MMTV: mouse mammary tumor virus

MPMV: Mason-Pfizer monkey virus

MS: mass spectrometry

NC: nucleocapsid protein or nucleocapsid domain of Gag polyprotein

NMR: nuclear magnetic resonance

PAGE: poly-acrylamide gel electrophoresis

PDB: Protein Data Bank

PIC: pre-integration complex

$PI_{(4,5)}P_2$: phosphatidylinositol-(4,5)-bisphosphate

PM: plasma membrane

PR: protease protein or protease domain of the Gag-Pol polyprotein (except in RSV)

PS: phosphatidylserine

RSV: Rous sarcoma virus

RT: reverse transcriptase protein or reverse transcriptase domain of Gag-Pol polyprotein

S4G: SSSSG linker sequence

SDS: sodium dodecyl sulfate

ss: single-stranded

SP: spacer peptide, a cleavage product or domain of the Gag polyprotein in RSV

SPA: SP assembly domain

SU: surface protein, a furin-cleaved envelope glycoprotein or the surface protein domain of the gp140 glycoprotein

SUMO: small ubiquitin-like modifier

TEM : transmission electron microscopy

TFE: trifluoroethanol

TM: trans-membrane protein, a furin-cleaved envelope glycoprotein or the transmembrane domain of the gp140 glycoprotein

Ulp1: a SUMO protease

UTR: untranslated region

VLP: virus-like particle

WDSV: walleye dermal sarcoma virus

WEHV: walleye epidermal hyperplasia virus

CHAPTER ONE

INTRODUCTION

GENERAL INTRODUCTION

Retroviruses are a family of enveloped, single-stranded (ss) RNA viruses that infect host species from fish to humans and are current and relevant human pathogens. The most infamous member of this family is human immunodeficiency virus type 1 (HIV-1), the causative agent of acquired immunodeficiency syndrome (AIDS), which has killed an estimated 25 million people since its initial characterization in the early 1980s, including 1.7 million individuals who died of AIDS-related causes in 2011. Another retroviral human pathogen is human T-cell leukemia virus type I (HTLV-I). Although infection with HTLV-1 does not carry the same rate of mortality as HIV-1, it has been associated with diseases such as myelopathy and parasitic superinfection, conditions that severely compromise the host's quality of life, in addition to adult T-cell leukemia.

As their name suggests, retroviruses possess an enzyme known as reverse transcriptase (RT), which converts their genomic ssRNA into double-stranded (ds) DNA through a complex series of biochemical reactions. Though RT is the retroviral namesake, integrase (IN) is the viral signature. Following reverse transcription, IN is transported into the nucleus along with the newly synthesized dsDNA and incorporates the viral dsDNA into a host chromosome, literally leaving a genetic imprint in the host. Infected cells actively and persistently produce retroviral proteins along with their own genetic products, resulting in newly produced virions that continuously re-infect the host.

Despite the discovery of the first retroviruses over a century ago by Ellerman and Bang, and Rous (Nobel Prize 1966), a cure for retroviral infections has remained elusive. The prognosis for infected individuals, however, is far from grim. Around the same time as the discovery of RT in 1970 by Temin, Baltimore, and Dulbecco (shared Nobel Prize 1975) came a shift of interest in biology from the organism as a whole to mechanisms at the molecular level. Since then, scientists have dissected various model retroviruses down to the molecular and atomic level and have acquired a great deal of knowledge on the retroviral genome, genetic products, life cycle, and interactions with their hosts. As a result of this in depth understanding of the mechanisms underlying retroviral infection, many antiretroviral treatments exist today to keep HIV-1 infection under control and prevent it from progressing to AIDS. Furthermore, new therapies are currently being developed to address unexploited targets in the retroviral life cycle.

CLASSIFICATION OF RETROVIRUSES

The retroviral family

The family *Retroviridae* is divided into two subfamilies: *Orthoretrovirinae* and *Spumavirinae*. The orthoretroviral family is composed of six genera of pathogenic retroviruses. *Alpharetrovirus* is a genus of avian retroviruses that includes the earliest discovered retroviruses: avian leukosis virus (ALV) and Rous sarcoma virus (RSV). The genus *Betaretrovirus* infects mammals and is represented by the Mason-Pfizer monkey virus (MPMV), mouse mammary tumor virus (MMTV) and Jaagsikte sheep retrovirus (JSRV). Members of the *Gammaretrovirus* genus also infect mammals, and the prototypic species is murine leukemia virus (MLV). *Deltaretrovirus* is a genus that

includes the human pathogen HTLV-I and its close relative HTLV-II, as well as bovine leukemia virus (BLV). *Epsilonretrovirus* is a genus containing two species of fish retroviruses: the walleye dermal sarcoma virus (WDSV) and walleye epidermal hyperplasia virus (WEHV). Finally, *Lentivirus,* so named because of its slow disease progression, includes viruses that cause immunodeficiency in cats (feline immunodeficiency virus, FIV), cows (bovine immunodeficiency virus, BIV), and primates (simian immunodeficiency virus, SIV and HIV types 1 and 2). Viruses belonging to the subfamily spumaviridae are sometimes called foamy viruses due to the large vacuoles that form in infected cells. Spumaviruses have not been conclusively linked to disease in hosts from which they have been isolated. Many differences exist between spumaviruses and orthoretroviruses at the level of the viral life cycle and at the molecular level. Therefore, the topics covered in the remainder of this thesis will apply exclusively to orthoretroviruses, unless otherwise noted. An excellent overview of the molecular biology of spumaviruses can be found in the review by Maxine Linial [1].

Retroviruses are also related to some genetic elements capable of reverse transcription. These are known as retroelements and include endogenous retroviruses (ERV) and long-terminal repeat (LTR) retrotransposons. Most ERVs are proviruses that often have debilitating mutations in one or more viral proteins and can no longer produce infectious virions or leave the cell, or they are integrated in non-expressed chromosomal regions. However, some ERVs can still become expressed and cause infection in their hosts, usually in the presence of certain host factors or “activating” agents such as radiation and carcinogens, leading to diseases such as spontaneous lymphomas in AKR mice (endogenous MLV) or leukosis in chickens (endogenous avian

leukosis virus, EALV and endogenous avian virus, EAV) [2]. LTR retrotransposons are similar to ERVs in that they encode their own RT and IN and multiply within the host genome by creating new copies of themselves. Although they do not produce infectious virus, retroelements may continue to alter their hosts through changes in gene expression and chromosomal rearrangement mediated by the LTR, a hallmark DNA footprint of reverse transcription and integration generated as a result of the series of biochemical acrobatics undertaken by RT. The LTR is a repeated sequence that flanks the retroviral genome at both ends and varies in length from approximately 300 bp (alpharetroviruses) to over 1200 bp (spumaviruses). The LTR may act as an alternative enhancer/promoter site for nearby genes and can also mediate recombination, deletion and duplications in the genome simply because of their high copy number. As they are found across the eukaryotic domain, retroelements are believed to be the result of retroviral integration into the host germ line over the course of evolution.

ORTHORETROVIRAL LIFE CYCLE

Early stage

The retroviral life cycle is roughly divided into two halves, termed “early” and “late”. The early stage includes viral binding, entry, uncoating, reverse transcription, trafficking, nuclear entry, and integration (Figure 1.1). The virus life cycle begins when an infectious virion binds the host cell surface via relatively non-specific interactions. Once attached to the cell plasma membrane (PM), viral envelope glycoproteins bind specifically to their cognate receptors and mediate fusion between viral and cellular membranes. Viral contents, including the capsid that contains two copies of the

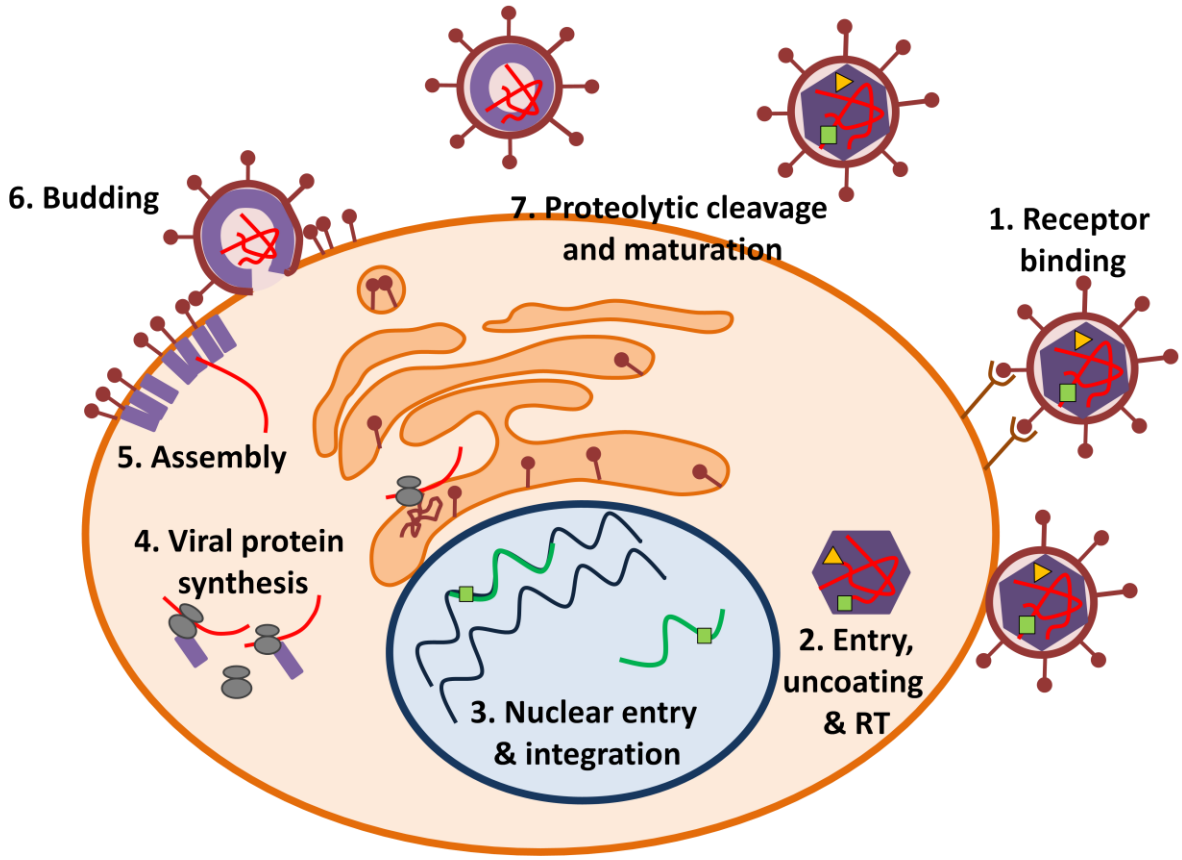


Figure 1.1: The retroviral life cycle. The early stage of the retroviral life cycle is illustrated in steps 1 – 3, while the late stage is in steps 4 – 7.

retrovirus ssRNA genome spill out into the cytoplasm, where capsid uncoating and reverse transcription take place. Reverse transcription leads to the formation of the pre-integration complex (PIC): a poorly characterized, integration-competent association of viral enzymes, reverse-transcribed DNA, and other proteins, including proteins from the host (e.g. LEDG F for HIV-1) that may assist in the retroviral life cycle [3, 4]. The PIC is trafficked through the cytoplasm, likely by association with cytoskeletal motor proteins and reaches the nucleus. Depending on the species of retrovirus, nuclear entry takes place in different ways. Some PICs are actively transported through the nuclear pore, possibly through interactions with importins that mediate nuclear entry (RSV, HIV-1), while others must wait for the breakdown of the nuclear membrane during cell division (MLV). Once inside the nucleus, the dsDNA is inserted into the host genome by integrase, which has a preference for highly transcribed regions. The integrated virus, or the provirus, is the template from which retroviral RNA can be transcribed [5, 6].

Late stage

The late stage of the retroviral life cycle includes the steps following integration: transcription, nuclear export, translation, assembly, budding, and maturation (Figure 1.1). Viral DNA is transcribed as a single long RNA transcript that is alternatively spliced. All retroviruses contain three canonical genes: *gag, pol,* and *env*, each coding for a functionally distinct group of proteins required for the retroviral life cycle. The structural proteins, matrix (MA), capsid (CA), and nucleocapsid (NC) are part of the Gag polyprotein. The *pol* reading frame codes for the essential viral enzymes, including IN, RT, and protease (PR), and *env* codes for the envelope glycoproteins, generally known

as the surface protein (SU) and the trans-membrane protein (TM). Gag is translated as a single polyprotein and the Gag-Pol polyprotein is initiated from the same start codon as Gag, but a ribosomal frame-shift or termination suppression at the *gag-pol* junction occurring approximately 5 – 10% of the time results in the production of the longer Gag-Pol. Both polyproteins are translated from the genome-sized transcript. In the cytosol, Gag binds to genome-sized viral RNA transcripts and trafficks to the PM where it assembles into the viral core and includes one molecule of Gag-Pol for every 10 – 20 Gag monomers. Env is produced from a spliced transcript as a long precursor protein and hijacks the endoplasmic reticulum (ER)/Golgi-mediated trafficking system. Within the ER, Env oligomerizes into a trimer and is glycosylated. As it progresses through the Golgi, Env is cleaved by a cellular protease into its constituent subunits SU and TM, which remain non-covalently associated. SU mediates receptor-binding, while TM contains the fusogenic peptide required for the fusion of viral and cellular membranes. The mature trimeric envelope glycoprotein is then transported to the cell surface, where it joins the assembling Gag and Gag-Pol to form the immature retrovirus.

At least two molecules of Gag bind to viral RNA with their NC domains, bringing the proteins close together and initiating the first step in assembly. Gag also associates with the PM via the MA domain, leading to oligomerization and assembly of the viral protein core at the PM. The formation of the immature core generates membrane curvature, ultimately resulting in a virus particle that buds from the cellular membrane. Betaretroviruses are a curious exception, as they assemble fully within the cytoplasm at a perinuclear location and are transported to the PM via recycling endosomes. The immature virus buds from the PM by exploiting the cellular machinery known as

endosomal sorting complexes required for transport (ESCRT), normally used for the formation of multivesicular bodies and membrane abscission during cytokinesis. During or rapidly following budding, PR is activated through an unknown mechanism and cleaves Gag and Gag-Pol to liberate the viral structural proteins and enzymes. The viral core undergoes a morphological change in which the immature spherical core takes on the appearance of a mature capsid. At this stage, maturation is complete and the retrovirus is ready to infect another cell.

Accessory genes

Retroviruses may contain genes other than *gag*, *pol*, and *env*. These “accessory” genes have evolved myriad functions to assist the virus at different steps of its life cycle. Betaretroviruses, deltaretroviruses, epsilonretroviruses, lentiviruses, and spumaviruses have genomes that contain accessory genes. For this reason, these viruses are deemed to have “complex” life cycles. Alpharetroviruses and gammaretroviruses do not have accessory genes and are classified as “simple” retroviruses. Accessory genes are a subject of current and ongoing research and are beyond the scope of this introduction.

VIRION STRUCTURE

Immature virion organization

In recent years, development and refinement of high-resolution structure determination techniques have lead to a wealth of new information on the late stages of the retroviral life cycle, particularly on the way retroviruses organize their cores. The retroviral core exists in two forms, immature and mature, that are structurally and

morphologically distinct. Figure 1.2 shows the immature core, which is made of an array of hexameric Gag polyproteins forming a spherical shell around the viral genome. Gag is the retroviral structural polyprotein (Figure 1.3) and comprises three domains common to orthoretroviruses: MA, CA, and NC, each of which will be discussed in following sections. In the immature core, Gag is organized radially with MA bound to the inner viral membrane and the C-terminus of Gag pointing toward the inside of the virion [7]. Although MA may be oligomeric in crystal structures, an organized layer consisting of the MA domain has not been observed using cryo-EM [8, 9]. The CA domain forms an ordered lattice composed of hexameric rings, and the NC domain complexed with RNA can be seen as a diffuse density near the center of the immature core.

Mature virion organization

Retroviral maturation is a rapid process and very few immature virions are normally observed among wild-type virions collected from cells. Maturation takes place when PR cleaves Gag into its constituent domains, leading to disassembly of the spherical core, followed by reassembly of each liberated domain into its mature conformation [10, 11]. The liberated MA remains bound to the viral membrane. A striking morphological transformation takes place in CA, which is the major constituent of the organized lattice in both the immature core and the mature capsid. The spherical core formed by the CA domain of Gag rearranges itself into distinct capsid geometries including conical (HIV-1), cylindrical (MPMV), and polyhedral (RSV), as observed by EM. The NC-RNA layer condenses into a dense globular mass, possibly to fit the NC-RNA complex into the spatially restricted confines of the mature capsid [14] (Figure 1.2).

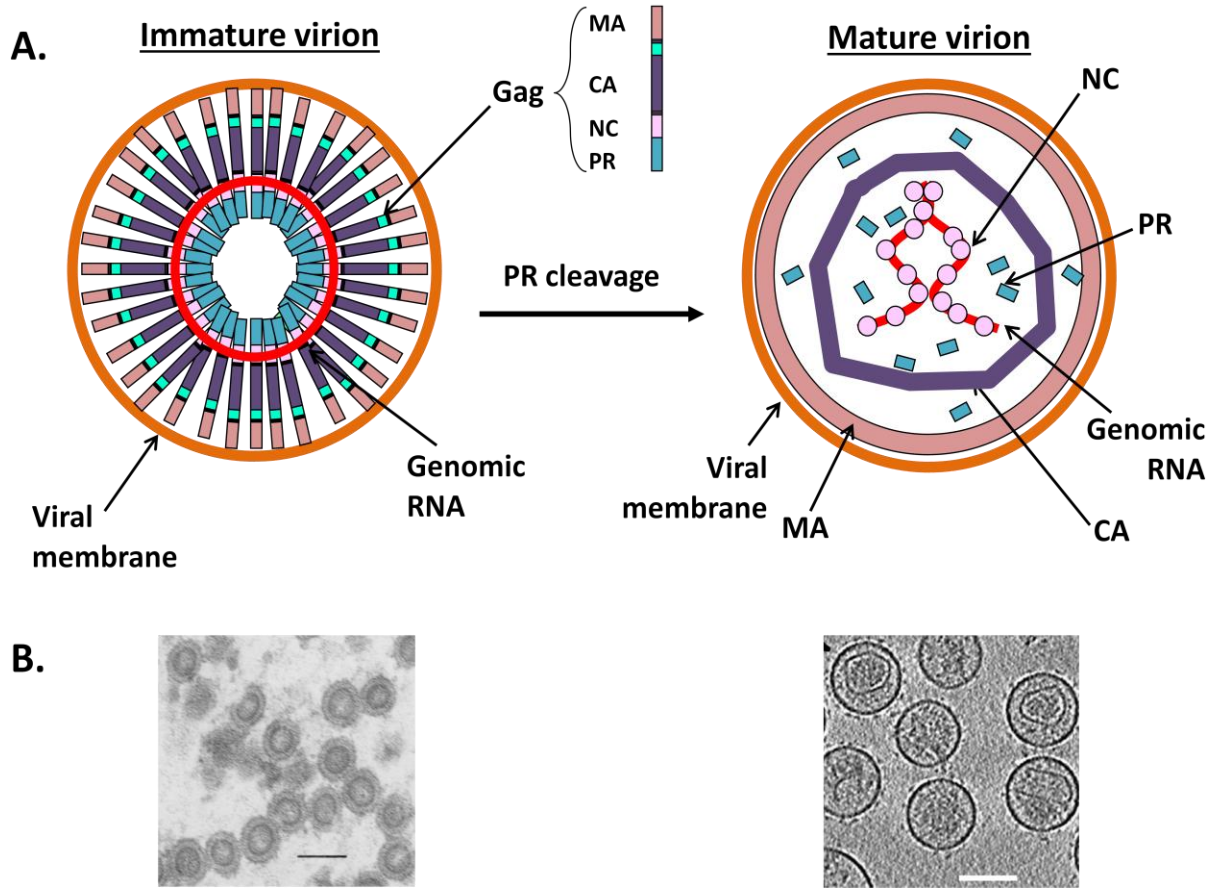


Figure 1.2: The morphology and structural organization of immature and mature retroviral virions are distinct. A. The immature virion is composed of a lattice of radially arranged Gag molecules bound to the viral genomic RNA (red) and enveloped by the viral membrane (orange). In RSV, Gag contains four major domains: MA (pink), CA (violet), NC (light pink), and PR (blue). Following PR cleavage, the liberated structural proteins reassemble into their mature conformation. B. Authentic immature virions have a spherical core made up mainly of the CA domain of Gag. The core consists of an electron-dense ring around an electron-lucent center, as seen using thin-section EM (left). Authentic mature virion capsids are morphologically distinct from immature cores. In RSV, the capsids are polyhedral, as seen using cryo-EM (right). EM images are adapted from Butan et al. and Campbell et al. [12, 13].

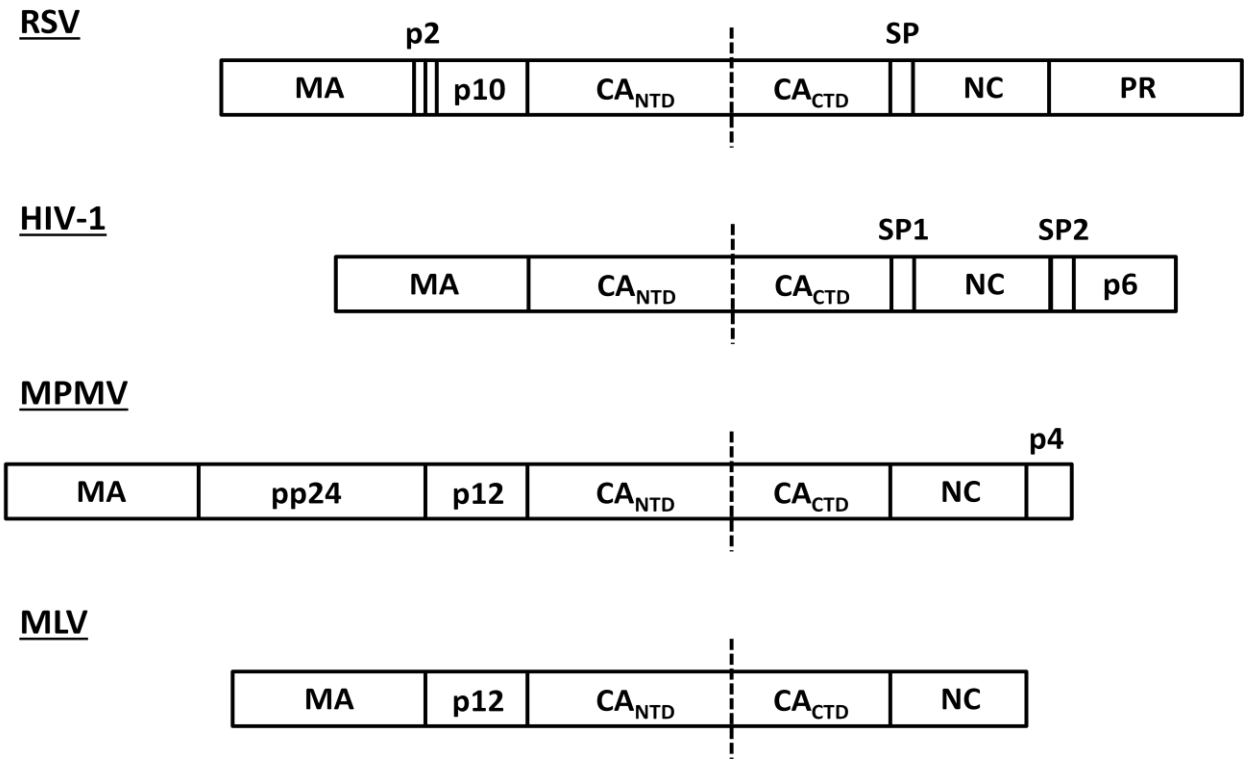


Figure 1.3: Schematic diagrams of retroviral Gag proteins from RSV, HIV-1, MPMV, MLV and their constituent domains. Inter-domain cleavage sites are indicated by vertical lines. All orthoretroviral Gag proteins have in common the MA, CA, and NC domains. CA is a two-domain protein and the location of its inter-domain junction is shown as a dashed line.

Virus-like particles and self-assembly

Virus-like particle (VLP) is a general name for viral protein-derived assemblies that are composed solely of the structural proteins required to form the viral core. VLPs are useful for the study of viral core structure as they resemble authentic virions in morphology. In retrovirology, the term VLP generally refers to assemblies of Gag, which is the only protein required to make the protein shell of immature retroviruses. VLPs can be produced by transfecting Gag into cells, which results in the release of spherical immature cores enveloped in membrane into the medium. VLPs also self-assemble in vitro using purified Gag protein under the appropriate pH and salt conditions (discussed in detail in the section on in vitro assembly), and in the presence of certain negatively-charged molecules such as oligonucleotides or inositol phosphates [15, 16]. In vitro assembled VLPs are organizationally identical to authentic virions, but they lack a viral membrane [10]. In this dissertation, I will use the term VLP to describe immature, Gag-derived assemblies, and “capsid assembly” for CA-based assemblies.

Inhibition of maturation

To facilitate studies of the immature core in the laboratory, viral maturation can be inhibited in a number of ways. Immature virions can be harvested in cells by mutating the catalytic Asp in the active site of PR, resulting in non-infectious, immature virions containing a spherical core and a full complement of viral proteins [17, 18]. VLPs can also be made in cells by transfecting a cellular expression vector containing Gag in the absence of Pol. This approach is not suitable for RSV and other avian retroviruses, which encode PR as part of the Gag polyprotein. When using these viruses to generate

VLPs, PR must be inactivated by mutation or prevented from being expressed by introducing an upstream stop codon. Mutating the cleavage sites in Gag is also a viable means of preventing maturation, but despite several analyses, mutations that inhibit PR processing remain poorly characterized, except in the case of HIV-1 [14, 19-22].

Small molecules that inhibit maturation, either by binding to PR or to the Gag lattice, have been developed as antiretroviral compounds. PR inhibitors have been hugely successful and are often used as part of antiretroviral treatment programs. Gag-binding inhibitors are a newer class of antiretrovirals generally referred to as maturation inhibitors. The most extensively studied member of this class is bevirimat, a betulinic acid-derived compound that has been shown to bind to the assembled immature Gag lattice rather than to individual Gag proteins, and that prevents PR-mediated cleavage of the CA-SP1 domain junction [23-26]. However, bevirimat was not successful in clinical trials as resistance develops rapidly [26-29]. Other HIV-1 Gag-binding compounds are still actively under study, including the Pfizer compound PF-46396 that binds a putative pocket formed by SP1 and the major homology region (MHR) [30].

GAG POLYPROTEIN

MA domain

The MA domain associates with the PM via its N-terminal portion, also called the membrane-binding domain (MBD). MA employs several distinct mechanisms for membrane-binding: 1) a patch of positively charged amino acids in the MBD common to all MA crystal structures and structural models [31] and 2) an N-terminal myristoylation present in some retroviruses, believed to help anchor MA into the PM. The positively

charged amino acids, also known as the basic patch, interact extensively with the inner leaflet of the PM, which is enriched in negatively charged phospholipids such as phosphatidylserine (PS) and phosphatidylinositol-(4,5)-bisphosphate ($PI_{(4,5)}P_2$). The N-terminal myristate has been studied in depth using HIV-1 MA and is generally accepted to be a switch for enhanced membrane binding following $PI_{(4,5)}P_2$ binding, which triggers myristate exposure and insertion into the membrane bilayer and is essential for HIV-1 viral release [17, 32-36]. $PI_{(4,5)}P_2$ binding in HIV-1 was structurally characterized by Saad et al., who showed that HIV-1 MA has a specific hydrophobic $PI_{(4,5)}P_2$ binding pocket that is not shared among non-myristoylated retroviruses such as RSV [32, 37].

Gag-driven multimerization is also an essential component to MA binding to the PM in both RSV and HIV-1, and may contribute to either myristate exposure or enhanced electrostatic association [35, 38, 39]. A recent set of experiments by Dick et al. examined the different states of oligomerization and their effects on MA-binding to the PM. The authors showed that monomeric MA protein is mainly cytoplasmic, while artificially generated dimeric and hexameric MA are clearly membrane-bound when transfected into cells. These results can be recapitulated in an in vitro system of purified protein and liposomes, in which monomeric MA is largely non-liposome-bound, while hexameric MA is primarily liposome-bound (Dick et al., in preparation).

The structure of MA and MA fragments has been solved for several retroviruses, including HIV-1 (non-myristoylated and myristoylated), RSV, MPMV, MLV, EIAV, SIV and HTLV-II [32, 38, 40-50]. They have in common four core alpha helices, although a fifth helix was observed in an unpublished dimeric structure of RSV MA and may be

involved in dimerization (Richard Kingston, personal communication). HIV-1 MA has also been crystallized in a trimeric form (PDB ID: 1HIW) [40], although it is not clear whether these multimerization states reflect biologically relevant interfaces or artifacts of crystal packing.

CA domain

CA is the main component of the retroviral core and mediates the majority of protein-protein interactions in both the immature and mature cores. Based on comparative analyses of high-resolution structures of both the immature and mature cores, it has been inferred that during maturation, CA disassembles and reassembles into the mature viral capsid [10]. The detailed protein-protein interactions that lead to the assembly of the retroviral core will be discussed in detail in the following sections.

CA is a two-domain protein with independently folding N- and C-terminal domains, commonly known as CA_{NTD} and CA_{CTD}, connected by a short flexible linker (Figure 1.4). CA_{NTD} is a globular domain consisting of seven α-helices of varying length, while CA_{CTD} is composed of four α-helices. The atomic-level structure of CA has been determined using constructs in various states of oligomerization, including CA_{NTD} alone, CA_{CTD} alone, fragments of CA in complex with other molecules such as Fab fragments and cyclophilin A, and full-length CA as a monomer, dimer, and hexamer [51-66]. Like MA, the structure of CA, though not the amino acid sequence, is highly conserved across all retroviral species [56, 62]. In the free CA protein, the first residue is invariably a proline that forms a salt bridge with a conserved aspartic acid residue in helix 3 (D_{52} in RSV and D_{51} in HIV-1) [52, 59]. The formation of this salt bridge is one of the events

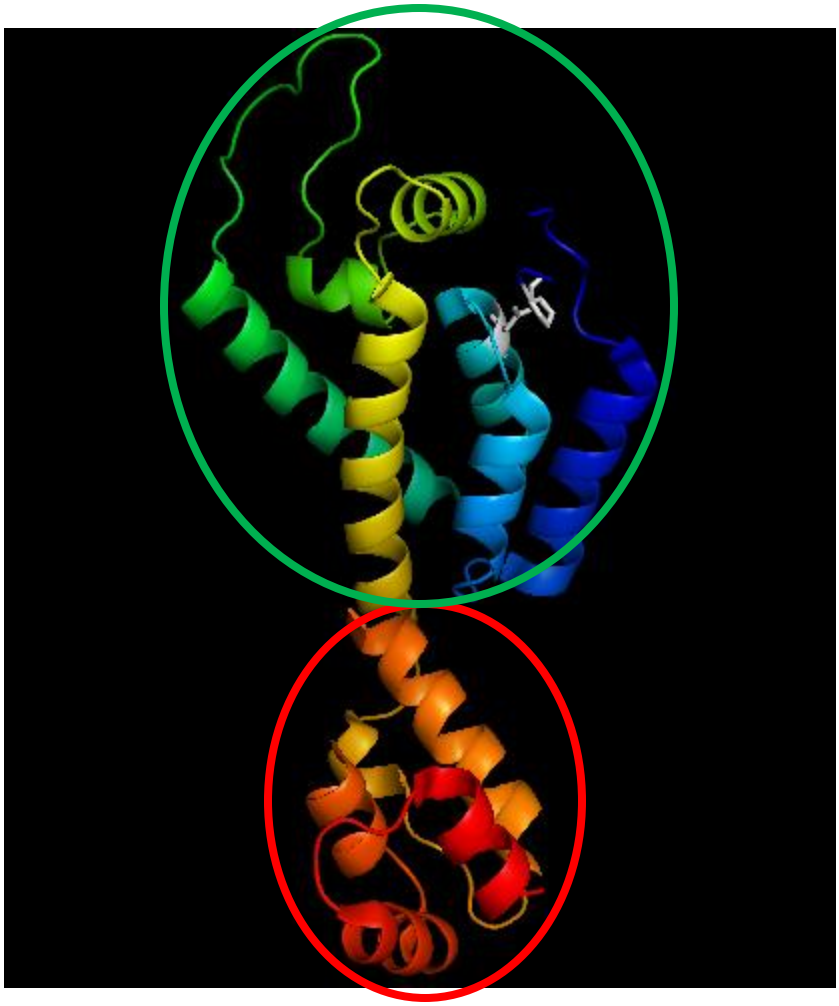


Figure 1.4: CA is a two domain protein consisting of CA_{NTD} and CA_{CTD}. CA is structurally conserved among orthoretroviruses, although this structure of HIV-1 CA is the only crystal structure of full-length CA. The seven N-terminal helices make up CA_{NTD} (green oval), and the four C-terminal helices, the CA_{CTD} (red oval). The N-terminal salt bridge is highlighted in stick view in white. Image adapted from PDB structure ID 3MGE [67].

that signal CA to assemble into the mature capsid.

Much of the structural work on CA has been done using HIV-1 proteins. An NMR solution structure of HIV-1 CA_{NTD} alone, including residues 1 – 151 of the CA protein was first determined by Gitti et al. and revealed for the first time the presence of the proline-mediated salt bridge [52]. This structure was closely followed by a crystal structure of a dimeric full-length HIV-1 CA in complex with a Fab fragment, but an ordered CA_{CTD} was not observed in the crystal and only the structure of CA_{NTD} was solved [53]. A structure of the dimeric HIV-1 CA_{CTD} encompassing residues 151 – 231 of CA was determined to 1.7 Å resolution. The authors identified W_{184} and M_{185} residues as critical residues for dimerization and confirmed this finding using viral constructs, thereby validating the crystallographic dimer interface as one found in authentic virions [54]. The CA_{CTD} structure was later confirmed using an extended construct embodying residues 146 – 231, which dimerizes similarly to full-length CA_{CTD} [55]. The first full-length structure of retroviral CA was determined using EIAV CA and confirmed multiple sequence alignments that predicted conservation among retroviral CA at the level of secondary structure [65]. This was followed by the unambiguous docking of known CA domain structures into a 9 Å cryo-crystallographic structure of full-length CA in the viral lattice [60]. The first non-lentiviral CA structure to be solved was RSV [56, 57, 59], revealing that conservation of CA structure may span all retroviral genera. MLV, MPMV, and JSRV CA_{NTD} structures were later solved as well [58, 61-63]. The first hexameric structure of CA_{NTD} was determined by Mortuza et al. using N-tropic MLV, reflecting the hexameric rings observed in the assembled capsid lattice [58]. Finally, an X-ray

structure of the full-length CA in a cross-linked hexameric form was solved by Pornillos et al., fully revealing the atomic interactions that tie together the mature CA lattice [64].

Although there is little amino acid conservation among retroviral structural proteins, a unique region of high sequence conservation named the MHR is found at the beginning of CA_{CTD}. The MHR spans approximately 20 amino acid residues, is found across all retroviruses except spumaviruses and has even been identified in retroelements. The structure of the MHR overlaps the extended strand at the N-terminus of CA_{CTD} and most of the first helix of CA_{CTD} (eighth helix of CA) [55, 57]. The function of the MHR is unknown, but mutations in the MHR often abrogate viral infectivity and compromise assembly [68-71]. Recently, the MHR has been implicated in the formation of an inter-domain interface with the spacer region downstream of CA_{CTD} [30], suggesting that it may be involved in immature assembly.

CA_{CTD} also contains a single buried but essential cysteine in alpharetroviruses, and a pair of cysteines in nearly all other orthoretroviruses [56] except gammaretroviruses. Mutation of the single cysteine (C_{192}) in RSV wrecks assembly [72], but the residue is not reactive in intermolecular thiol cross-linking experiments [73], consistent with its being buried in the structure of CA_{CTD}. Therefore, C_{192} is preserved in the otherwise cysteine-free Gag constructs used in the cysteine cross-linking experiments described in this dissertation.

NC domain

The NC domain of Gag binds and protects the viral genomic RNA. NC is the smallest of the three canonical Gag domains and appears to be unstructured aside from

the presence of one or two “zinc fingers”, zinc-coordinated cysteine-rich structural motifs ($CX_2CX_4HX_4C$) that bind specifically to a highly structured stem-loop-rich region located in the 5’ UTR of the viral genomic RNA known as Ψ, or psi [74]. Zinc fingers are characteristic of all retroviral NC domains except spumaviruses which lack the NC domain altogether [1], and different orthoretroviral species may have one or two zinc finger motifs. In RSV, both zinc fingers are not required for immature assembly, as the cysteines in the zinc fingers can be mutated to alanines and serines without compromising assembly [72, 75, 76]. The preference of the NC domain for psi binding is what directs the virus to specifically package genome-sized retroviral RNA, rather than the abundant cellular RNA and spliced retroviral RNA in the cytoplasm [74]. NC also carries a net positive charge that contributes to nucleic acid binding, and basic residues flanking and in between the two zinc fingers in both RSV and HIV-1 are required for psi binding [75, 77].

It has been proposed that the initiating event in Gag assembly is driven by the binding of several molecules of Gag to the dimeric retroviral genomic RNA via their NC domains, and that this pre-formed complex binds to the PM and nucleates subsequent virion assembly [15, 78-80]. Using RSV Gag in an in vitro assembly system, Ma et al. showed that the shortest length of oligonucleotide required for efficient assembly of Gag is 16 nt, which provides enough binding space for two units of the NC domain. Thus, the lowest multimeric state required to trigger assembly is a dimer [79]. In fact, the NC domain of RSV and HIV-1 can be replaced by a leucine zipper dimerization domain and still support Gag assembly, validating the aforementioned conclusion [81-83].

Minor domains

The minor cleaved domains of Gag differ significantly across retroviral species in their size and function. Among the best-understood are the minor domains of RSV, HIV-1, MPMV and MLV Gag. The domain organization in RSV Gag is as follows: MA-p2-p10-CA-SP-NC-PR (Figure 1.3). Both the p2 and p10 domains contain short amino acid motifs collectively known as late domains. Three classes of late domains are found in retroviruses with the following consensus sequences: PT/SAP, PPXY, and $LYPX_nL$. These motifs each recruit distinct components from the cellular ESCRT machinery and at least one late domain is required for efficient retroviral budding. Late domain deletion mutants produce virus particles that remain tethered to the cell surface, unable to pinch off the membrane attachment that holds them to the PM [84]. In RSV, a PPPY motif is found in p2 and a LYPSL is located in p10. The p10 domain is also a “molecular switch” in Gag assembly, in that it promotes immature lattice formation by creating an interface that is not present in the mature capsid [72]. SP is the spacer peptide, located between CA and NC and is also a molecular switch hypothesized to form an ultrastructural element required for immature assembly. The role of SP in Gag assembly is the main focus of this thesis and will later be described in detail. PR is encoded as a domain of Gag by RSV, rather than as part of Gag-Pol like in most other retroviruses.

HIV-1 Gag includes, from N- to C-terminus the following six domains: MA-CA-SP1-NC-SP2-p6 (Figure 1.3). SP1 is hypothesized to play a similar role as RSV SP in promoting the structure of the immature core [9]. The role of SP2 is unclear, though it has been proposed to play a role in regulating the kinetics of Gag processing by PR

[85]. The p6 domain contains the late domain PTAP and a LXXLF motif that seems be homologous to the $LYPX_nL$ motif [86]. The p6 domain also binds Vpr, one of the six HIV-1 accessory proteins. Vpr plays a number of roles in the HIV-1 life cycle, including facilitating PIC nuclear entry, binding promoter regulating elements to increase HIV-1 RNA production, and arresting the cell cycle at G2 to promote virus production [87].

The domains of MPMV Gag are MA-pp24-p12-CA-NC-p4 (Figure 1.3). The pp24 domain harbors the PPPY and PSAP late domains [88, 89] and contains a highly basic stretch of amino acids that may play a role in Gag localization to the nuclear pore and viral genome packaging [90]. Several studies have attempted to elucidate the role of the MPMV p12 domain in Gag assembly, and have characterized an internal scaffolding domain (ISD) believed to promote higher-level Gag oligomerization at low concentration [91-93]. The ISD appears to constitute a functional domain, as an HIV-1 Gag- MPMV p12 chimera can assemble in a reticulocyte system that is ordinarily restrictive to HIV-1 Gag assembly, but permissive to MPMV Gag assembly [94]. Since MPMV assembles inside the cytoplasm rather than at the PM, it has been hypothesized that the ISD may be the mechanism by which MPMV overcomes the requirement for high Gag concentration for assembly, which is fulfilled by association with the PM in other retroviruses [91]. The function of the p4 domain is currently unknown.

MLV Gag has a single minor cleavage product, p12, located between MA and CA (Figure 1.3). MLV p12 appears to be a multifunctional domain. It is a structural switch for immature assembly, similar to RSV p10, as deletion of p12 results in tubular, rather than spherical VLPs [95, 96]. The PPPY late domain is also found in MLV p12. In

addition, the p12 domain also plays an interesting and unique role in PIC trafficking. Certain mutations in p12 abolish viral infectivity, apparently by preventing the PIC from correctly localizing to the host cell nucleus [97, 98]. This phenotype is caused, at least in part, by disrupting p12-mediated attachment of the PIC to mitotic chromosomes, allowing the PIC to gain access to the nucleus during cell division [99, 100].

Gag conformation

The conformation of Gag has been extensively explored by Datta et al., who have shown that in solution, monomeric Gag can exist in an elongated, linear conformation (MLV), similar to that found in assembled immature virions, or as a compact “horseshoe” structure (HIV-1) in which the MA and NC domains are near one another and Gag is folded over at one of its flexible regions (HIV-1) [101, 102]. The significance of the folded-over monomer is a matter of active study, as HIV-1 Gag must take on its elongated conformation before immature assembly can be possible. It has been suggested that this conformational change may reflect an additional regulatory step in Gag assembly, as HIV-1 Gag must simultaneously bind membrane and nucleic acid before taking on a linear conformation favorable for protein-protein interactions that drive oligomerization [103]. Interestingly, HIV-1 Gag on its own fails to assemble in vitro into VLPs resembling bona fide virions. However, upon addition of inositol pentakisphosphate (IP_5) or inositol hexaphosphate (IP_6) derivatives which are highly negatively-charged molecules, correct VLP assembly can be induced [16, 104]. These findings suggest that the folded-over conformation of Gag may prevent premature assembly within the cytoplasm prior to an electrostatics-mediated release.

RETROVIRAL CORE ARCHITECTURE

Stoichiometry of assembly

The number of Gag proteins required to assemble an immature core has been studied in detail using RSV and HIV-1. Using RSV, Vogt and Simon were the first to employ scanning TEM to measure the mass of purified viral particles. After making assumptions for the percentage of lipids, RNA, and carbohydrate from the glycosylated Env, and contaminating vesicles, the authors estimated that 73% of the total viral protein is Gag, which corresponds to approximately 1500 Gag molecules per virion [105]. This number was adjusted downward to 1200 in a different scanning TEM study [76], then adjusted upward following a later cryo-EM study, which measured lattice spacing between hexamers and the radius of the virions to calculate the number of Gag proteins per virion, which is approximately 2500 [106]. In HIV-1, the number of Gag proteins required for an immature core is much higher, at 5000 molecules per virion as measured by cryo-EM. This difference between the two viruses was attributed to the smaller size of RSV VLPs. However, these calculations are based on the assumption that the immature core forms a complete sphere. Since Wright et al. first discovered that the core lattice in immature virions harvested from cells is only, on average, 40% complete in HIV-1, a follow-up study by de Marco et al. found that this feature is conserved among retroviruses [9, 107]. Therefore, the actual number of Gag molecules may be much lower than the values calculated by lattice spacing and virion size.

Interestingly, mature HIV-1 capsids have been estimated to contain no more than 1250 units of CA [108, 109]. Even considering the incomplete Gag lattice, this number is

significantly lower than 40% of the estimated 5000 Gag molecules in immature HIV-1 cores. The substantial difference in the number of proteins required to assemble an immature or mature core led to the hypothesis that at least in HIV-1, the virus disassembles the immature core following proteolytic cleavage and reassembles the capsid, and does not simply rearrange protein-protein contacts in the immature lattice to transform it into a mature lattice [10]. Not all of the liberated CA is used to assemble the mature capsid, as measurements performed on capsid assemblies using hydrogen-deuterium exchange (HDX) coupled with mass spectrometry (MS) have shown that in mature assemblies, CA is present in both unassembled and assembled forms [110].

Mature CA capsid

Ganser et al. first proposed that the mature CA capsid is organized based on the same underlying principles governing the structure of fullerene cones [111]. A fullerene cone is an elemental carbon structure built from a layer of hexameric carbon subunits that form a curved sheet. The cone is closed at the ends by the incorporation of 12 pentameric “defects” into the hexameric sheet, based on the Euler characteristic. Similarly, a retroviral capsid can be made of hexameric and pentameric units of CA. Depending on where the pentamers are located in the viral capsid, the capsid takes on different geometries (Figure 1.5). For example, the vast majority of HIV-1 capsids incorporate five pentamers at the narrow end of the cone, and seven pentamers at the broad end [111]. Although other arrangements are possible, they result in altered capsid geometries. An even distribution of pentamers on both ends of a hexameric tube would result in a pill-shaped capsid, as observed in MPMV, while a scattered distribution of

pentamers over the hexameric lattice would result in a roughly spherical capsid, like that of MLV [112]. A preponderance of later data showed that the proposed fullerene capsid geometry is common to most retroviruses [108, 113-116] and reflects the lattice organization in both authentic virions as well as capsid assemblies [11, 109]. Once CA hexamers and pentamers were directly visualized using X-ray crystallography, the fullerene cone HIV-1 capsid was modeled in its entirety [116].

CA hexamers

The CA hexamer is the basic building block of the retroviral capsid (Figure 1.6). Early on, rings that appear roughly hexameric could be observed using TEM in stained virions. However, the existence of the CA hexamer was first confirmed by cryo-EM followed by image reconstruction of in vitro assembled CA tubes, which revealed a network of CA hexamers at 20 – 30 Å resolution [108]. The spacing between hexameric rings is 9.6 nm [109]. By fitting in the X-ray structures of the two domains of CA, it was determined that CA_{NTD} forms the bulk of the hexamer, while CA_{CTD} is located beneath the hexameric ring and links neighboring hexamers to one another. Using a novel technique combining cryo-EM and 2D crystallography, Ganser-Pornillos et al. were able to generate a higher resolution density map into which existing CA_{NTD} and CA_{CTD} structures could be docked unambiguously. This study validated three major interaction surfaces in the mature lattice. First, the CA_{NTD}-CA_{NTD} interface is the hexamerization interface, with helices 1, 2, and 3 interacting with neighboring CA_{NTD} subunits to form the central hole. This finding confirms alanine mutagenesis and HDX results that showed these helices to be critical for capsid assembly and protected from exchange

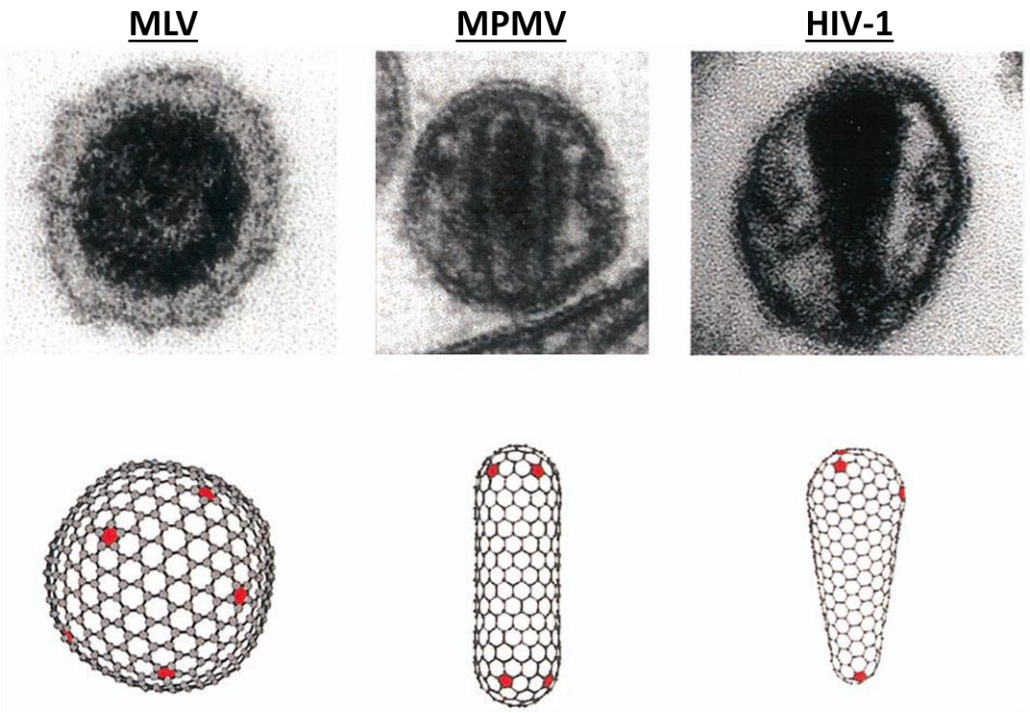
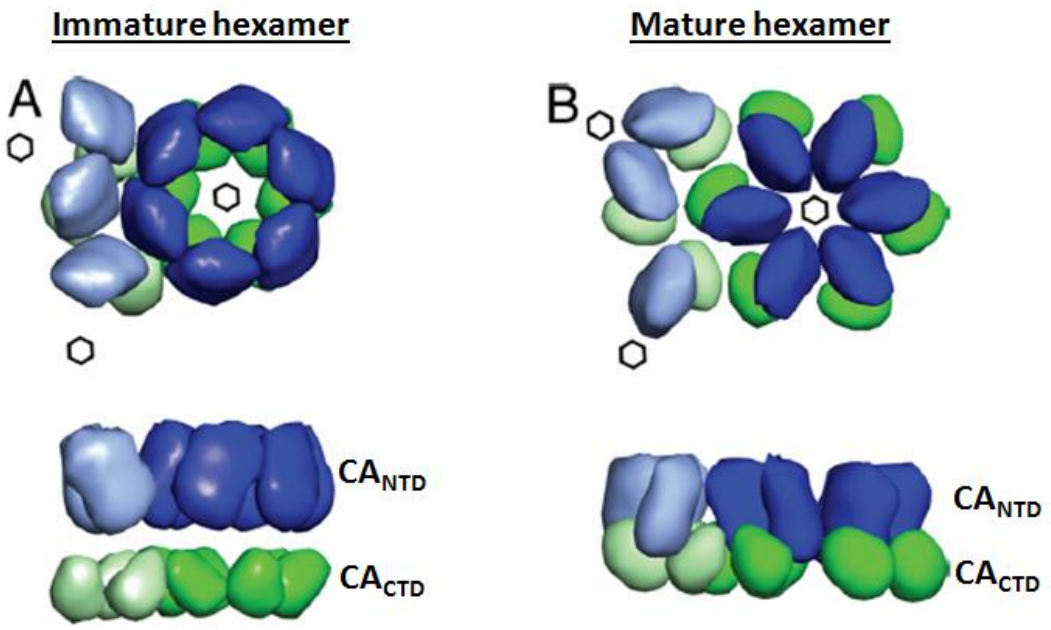


Figure 1.5: Retroviral capsids are fullerene cones. MLV, MPMV, and HIV-1 virions harvested from cells are shown above and differ obviously in capsid morphology (top). The fullerene cone model for each capsid morphology is shown as a network of hexameric rings (interconnected grey spheres) interspersed with pentameric defects (red). The location and distribution of pentamers determines the morphology of the fullerene cone (bottom). Figure adapted from Ganser-Pornillos et al. [112].



C.

Interaction	Immature		Mature	
	Intra-hexamer	Inter-hexamer	Intra-hexamer	Inter-hexamer
CA-NTD–CA-NTD	0	39	23	0
CA-NTD–CA-CTD	0	0	20	0
CA-CTD–CA-CTD	5	6	0	7

Figure 1.6: Immature and mature hexamers differ in protein-protein contacts. The immature hexamer has a large central hole that is absent in the mature (A & B, top). The positioning of CA_{CTD} relative to the CA_{NTD} ring is lower in the immature hexamer than in the mature (A & B, bottom). The specific contacts that hold together the immature and mature hexamers also differ drastically (C). The cryo-EM densities of the immature and mature hexamers were based on data from HIV-1 virions, as modeled by Briggs et al. [8] and the table of putative contacts was based on a cryo-EM model of MPMV tubes, as modeled by Bharat et al. [117]

when assembled [112, 118, 119]. Second, the CA_{CTD}-CA_{CTD}interface mediates mature lattice formation by dimerizing interactions in helix 9, as previously uncovered in an X-ray structure of HIV-1 CA_{CTD} [54, 55]. The dimers tie two adjacent hexamers together, resulting in lattice formation. Third, a CA_{NTD}-CA_{CTD} interface formed by helices 4, 8, 9, 10, and 11 mediates interactions between adjacent CA molecules within the hexamer, confirming a previous prediction made by Lanman et al. using chemical cross-linking and HDX [110, 118].

The atomic-level structure of the HIV-1 CA hexamer was finally solved by Pornillos et al. through elegantly placed mutations and thiol cross-linking guided by the lower resolution Ganser et al. structure. Two cysteine mutations were placed at positions 14 and 45 in the CA protein ($A_{14}C$ and $E_{45}C$) to stabilize the CA_{NTD}-CA_{NTD} interactions that form the hexameric ring, while two alanine mutations were made to positions 184 and 185 in CA ($W_{184}A$ and $M_{185}A$), disrupting the well-characterized lattice-forming dimerization interface. This engineered CA construct was cross-linked into soluble hexamers amenable to crystallization. The hexameric interface between CA_{NTD} subunits is highly solvated, suggesting a likely mechanism for the quasi-equivalence between CA hexamers and pentamers (further discussed in the following section), as the water molecules in the hydrogen-bonding network can easily adapt to the altered protein conformations required for pentamer formation. The extensive CA_{NTD}-CA_{CTD} interface is also water-rich and possesses a network of intermolecular helix-capping interactions, whereby each helix in the CA_{NTD}-CA_{CTD} interaction surface is capped by another residue in the adjacent molecule [64]. A later cryo-EM study also revealed a novel CA_{CTD}-CA_{CTD} interaction at the local three-fold axis between three CA

hexamers mediated by neighboring CA_{CTD} dimers. Although the cryo-EM images were only reconstructed to 16 Å resolution, the putative interacting residues were confirmed using cysteine mutations and disulfide cross-linking [120].

These high-resolution structures of the CA hexamer have provided a model for the generation of curvature on the hexameric lattice, where flexibility in the CA_{CTD} layer can tilt the relatively rigid CA_{NTD} layer sufficiently to allow curvature, regardless of whether pentameric defects are introduced, while preserving the helix-capping interactions in the CA_{NTD}-CA_{CTD} interface [64, 121]. A study by Bailey et al. also showed that RSV CA has significantly fewer constraints in the CA_{NTD}-CA_{CTD} interaction surface, suggesting a possible mechanism for the formation of different core morphologies [121].

CA pentamers

Based on the fullerene geometry hypothesis, closure of the mature capsid requires incorporation of 12 pentameric defects, which had been inferred to exist since the fullerene cone hypothesis was first proposed [111]. However, it was not until recently that pentamers were directly visualized by Cardone et al. in cryo-EM reconstructed RSV capsid assemblies. RSV CA was induced to assemble into small 17 nm particles in 0.5 M phosphate buffer. These particles were reconstructed and found to be T = 1 containing 12 pentamers. Docking CA crystal structures into the cryo-EM densities revealed that CA pentamers and hexamers are quasi-equivalent, meaning that the formation of a pentamer required only the removal of one subunit from the hexamer, and that the molecular interactions in both pentamers and hexamers are similar (Figure 1.7) [113]. Hyun et al. later produced RSV T = 1 icosahedral capsid assemblies using

low pH rather than phosphate buffer to trigger assembly and confirmed that these were also the product of 12 CA pentamers [122]. The observation of icosahedral VLPs is unique to RSV, and although icosahedral capsid assemblies do not reflect authentic retroviral capsid assembly, they allowed the direct visualization of CA pentamers and confirmed the fullerene cone hypothesis for capsid assembly. Structural conservation of the CA pentamer among different retroviruses was confirmed by Pornillos et al., who used a thiol cross-linking technique similar to that used for CA hexamers to solve the structure of the HIV-1 CA pentamer [116].

Immature Gag core

Unlike the mature capsid which differs significantly in morphology across retroviral species, the immature Gag core has a more uniform spherical appearance [107]. Based on cryo-EM and image reconstruction, the immature core lattice is made up entirely of Gag hexamers. The lattice is continuous, but incomplete, as regions lacking protein density are clearly visible beneath the membrane layer (Figure 1.8) [8]. The incomplete core allows for curvature in the purely hexameric lattice, overcoming the requirement for the incorporation of pentamers to close the core. This observation provides a simple explanation for the relatively uniform appearance of immature retroviral cores. An incomplete core also fits into the retroviral life cycle, as the necessary proteins for viral maturation and subsequent steps are incorporated into the virion as part of the structural polyprotein, and the viral genomic RNA also enters the nascent virion via Gag-binding. Core closure is only necessary during maturation, so that the virus may enclose all of the required components for a new round of infection.

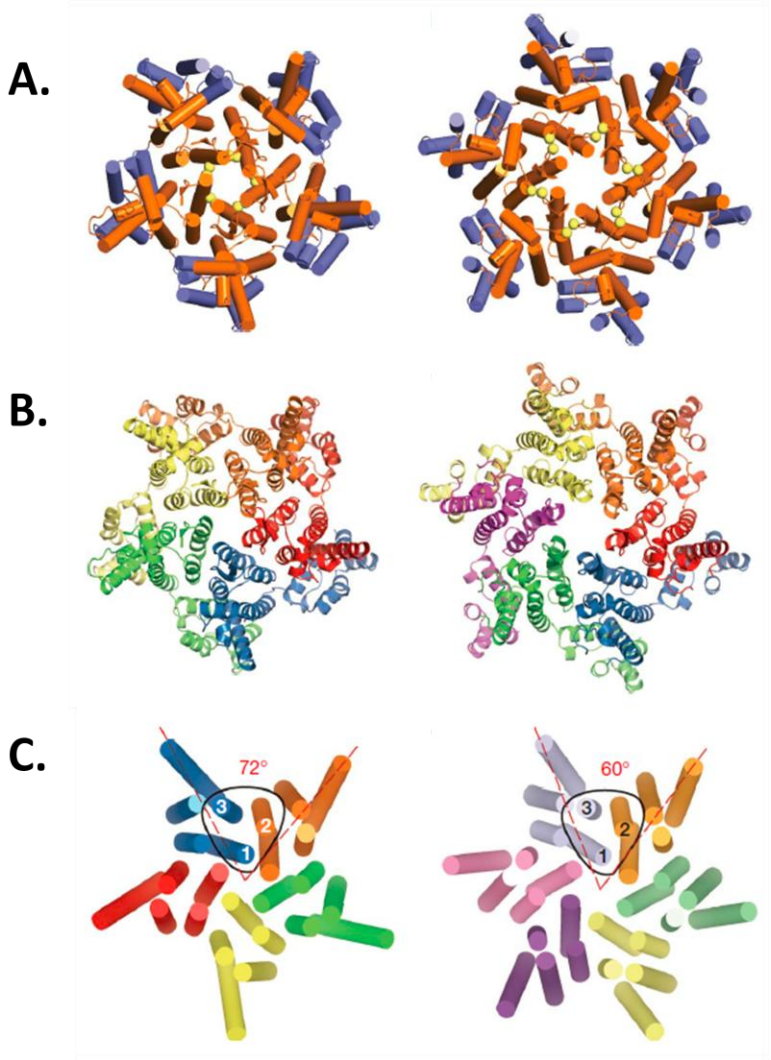


Figure 1.7: Pentamer and hexamers of mature CA are quasi-equivalent. A) The HIV-1 pentamer (left) and hexamer (right) are shown. The α-helices are represented as rods, with the locations of the thiol cross-linkages that stabilized the pentamers and hexamers for crystallography highlighted in yellow. B) The pentamer and hexamer are shown in cartoon representation. C) The quasi-equivalence of the pentamer and hexamer is shown. A simple change in the angle between interacting surfaces explains the shift between pentamer and hexamer assembly. Figure adapted from Pornillos et al. [116].

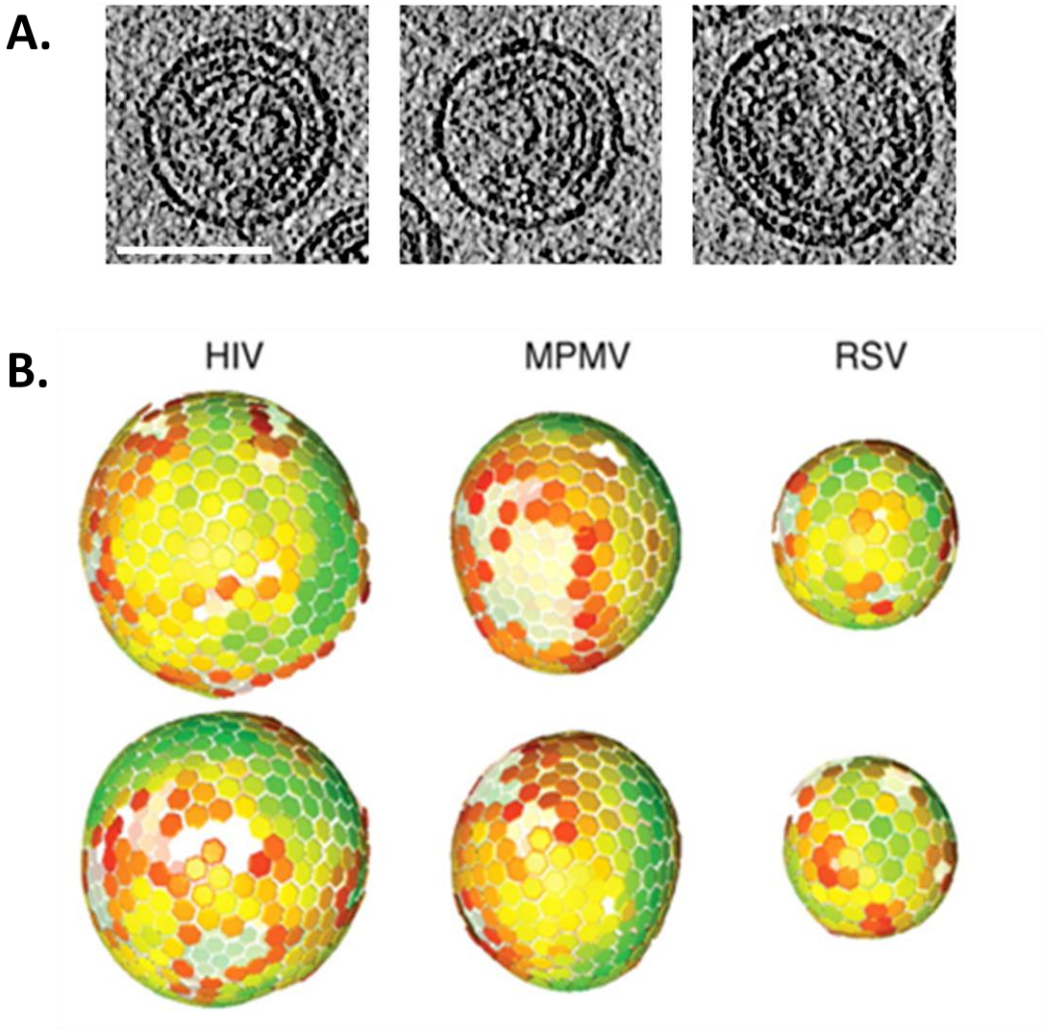


Figure 1.8: The immature core of retroviruses is made of an incomplete lattice of Gag hexamers. A. Three representative HIV-1 virions are shown above. The incomplete immature core layer is visible under the complete membrane layer. B. Each sphere represents a VLP that was reconstructed based on cryo-EM tomography and sub-tomographic averaging. Hexamers are colored based on the calculated cross-correlation and the confidence of each hexamer assignment ranges from red (low confidence) to green (high confidence). Figure adapted from Briggs et al. and de Marco et al. [8, 107].

Gag hexamers

Although the mature hexamer consists of proteolytically processed CA and the immature hexamers are primarily composed of the CA domain of Gag, respectively, structural constraints present in the CA domain but absent in proteolytically processed CA create an immature hexamer that is substantially different from the mature. Even when observed at low resolution, using cryo-EM reconstruction, there are several notable differences between the two hexamers that led to the inference that the mature and immature hexamers differ in specific protein-protein contacts (Figure 1.6). The observed differences include the following. The Gag hexamers have a large central hole absent from CA hexamers. The orientation of CA_{NTD} subunits that make up each Gag hexamer differs from that found in the CA hexamer. The CA_{CTD} layer is located further beneath the CA_{NTD} layer in Gag hexamers and is entirely tucked beneath the CA_{NTD} layer, while the CA_{CTD} layer in CA hexamers forms a ring around the CA_{NTD} layer just beneath it [8, 9]. The spacing between each immature hexamer is also smaller, measuring 8.0 nm rather than 9.6 nm, as found in the mature hexamer [10]. One interaction between the p10 and CA domains of RSV Gag was found to favor immature hexamer formation by creating an interface that pries apart the hexamer, providing a putative mechanism for the formation of the large central hole in the immature hexamer, but absent in the mature (Figure 1.9) [72]. The SP1 domain of HIV-1 Gag has also been proposed to play a similar structural role in promoting immature interactions in the CA domain [9], which will be further discussed in the following section.

A recently published study using cryo-EM tomography and crystal structure docking by Bharat et al. took advantage of the helical symmetry found in MPMV tubes to build an 8 Å model of the immature lattice. This model revealed some key distinctions between protein contacts found in the immature and mature cores, most notable being that contacts between adjacent CA_{NTD} subunits form dimeric inter-hexamer interfaces, rather than intra-hexamer interfaces, while the only contacts that appear to hold together the hexamer are located in CA_{CTD}. Also, the extensive CA_{NTD}-CA_{CTD} interface found in the CA hexamer is not present in the Gag hexamer. Another interesting comparison is the number of residues involved in intra-hexamer contacts. In the mature hexamer, the total number of residues that form contacts in the hexamer is 43, while it is only five in the immature hexamer (Figure 1.6) [117]. This discrepancy in the number of contacts suggests that, as is the case with RSV p10, other domains flanking CA may be involved in forming contacts that promote the immature hexamer. A prime candidate to fulfill this role, at least in RSV and HIV-1, is the downstream spacer peptide SP.

SP in the immature hexamer

The spacer peptide is a minor cleavage product of Gag in alpharetroviruses and lentiviruses. In RSV Gag, it is known as the SP domain and is located immediately downstream of CA, while in HIV-1, the homologous peptide is called SP1. Kräusslich et al. were the first to propose that HIV-1 SP1 may play a part in forming a transient assembly domain with HIV-1 CA_{CTD} that results in a critical intermolecular interaction during immature assembly [123]. This hypothesis was further extended by Accola et al., who proposed that the disruption of a helix spanning the HIV-1 CA-SP1 junction is

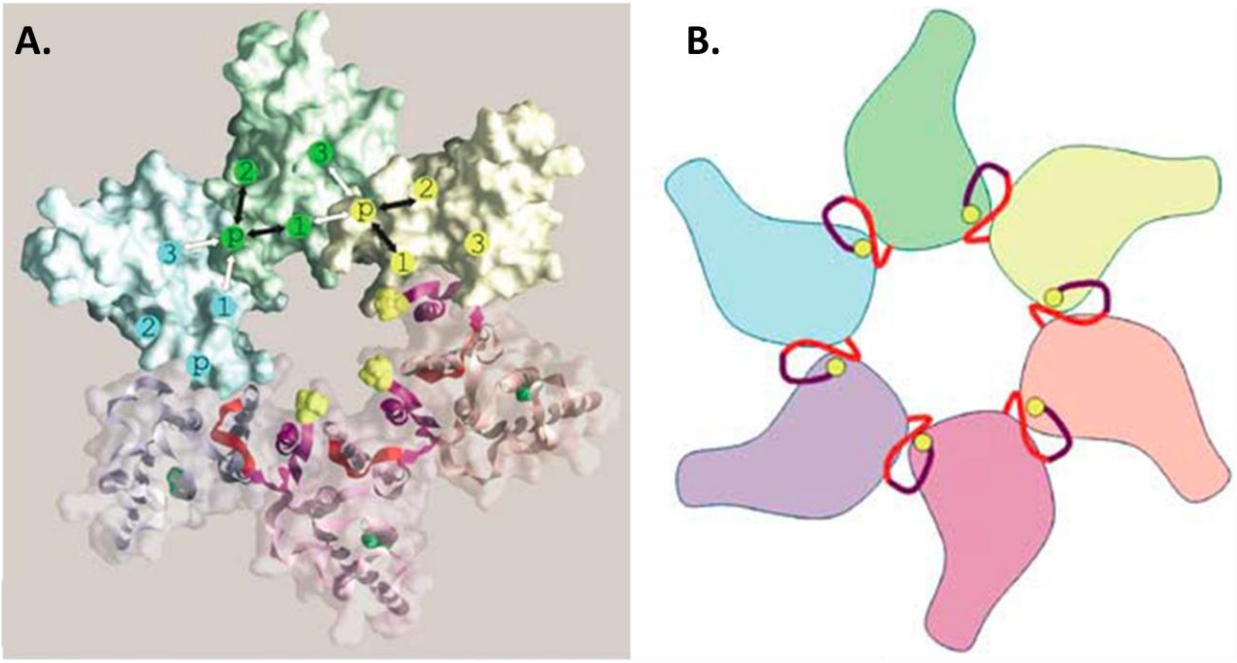


Figure 1.9: The RSV p10 domain forms an intermolecular interface between Gag molecules. A. Three CA_{NTD} subunits that make up the immature hexamer are shown in space-fill model, while the other three subunits are in ribbon model superimposed over space-fill. The p10 domain of one subunit forms intermolecular interactions with helices 1 and 2, and intramolecular interactions with helices 1 and 3 of the neighboring subunit. B. Cartoon representation of inter-hexameric interactions modeled based on the crystallographic dimer by Nandhagopal et al. [59]. The p10 domain is predicted to interact with residues that form part of the hairpin loop created by the N-terminal salt bridge in PR cleaved CA. Figure adapted from Phillips et al. [72].

required for proper mature assembly [124]. Later studies using cryo-EM tomography reconstructions of immature RSV and HIV-1 VLP lattices revealed a pillar-like density located at the center of each immature hexamer below the densities assigned to CA_{CTD} (Figure 1.10) [9, 107]. This pillar-like density was originally hypothesized by Wright et al. to consist of a six-helix bundle (6HB) that contributes to the stability of an immature Gag hexamer [9]. For both RSV and HIV-1, secondary structure predictions assign amino acids including and immediately flanking the spacer peptide domain as a single amphipathic helix [107, 124-126]. This structural prediction is not restricted to alpharetroviruses and lentiviruses. In gammaretroviruses, a highly charged region at the very end of CA has also been inferred to form an alpha helix [127], though there is no cleaved spacer peptide. Amino acids in this region can be deleted without affecting MLV Gag assembly, as long as the deletions allow conservation of the original helical phase.

The NMR structure of a peptide corresponding to HIV-1 SP1 and eight adjoining amino acids shows that this region has the ability to form an amphipathic helix under certain conditions, though the biological relevance of this structure has been questioned as it was solved in the presence of 30% trifluoroethanol (TFE), a helix-promoting solvent [125]. A more recent study using a different HIV-1 peptide encompassing the last eight amino acids of CA and the N-terminal two-thirds of SP1 showed a shift in the circular dichroism (CD) spectrum from unstructured to α-helix when peptide concentration was increased to levels comparable with the local protein concentration in an authentic assembling virion. This finding was interpreted to mean that SP1 acts as a conformational switch that shifts between two states, remaining mostly unstructured until virion assembly increases the local concentration of Gag, at which point SP1

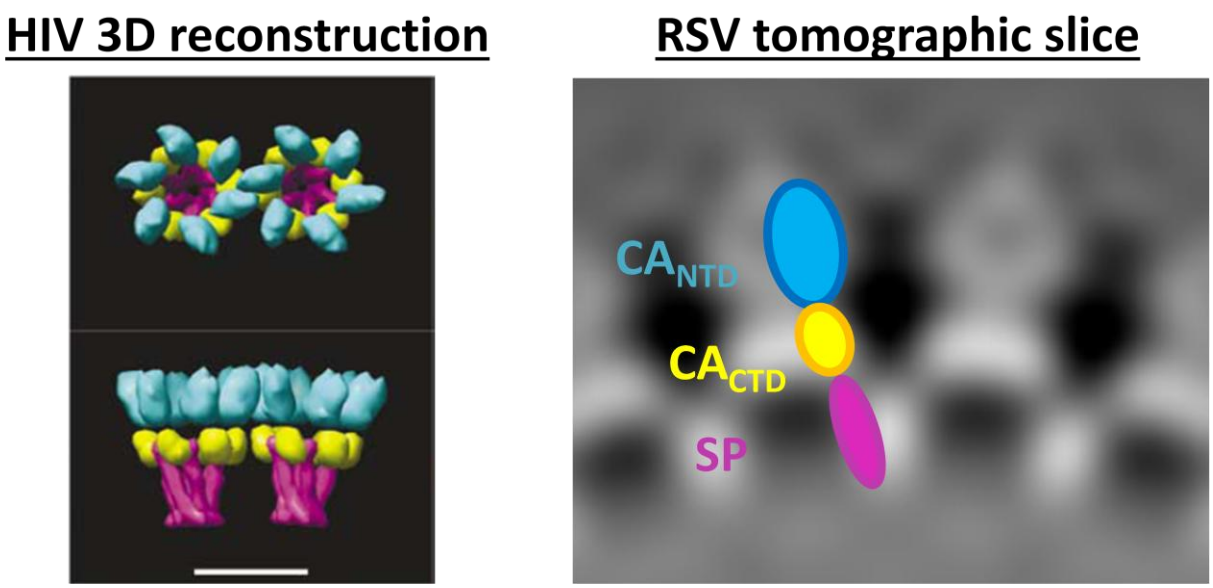


Figure 1.10: SP forms a pillar beneath the organized CA lattice in authentic virions and VLPs. Reconstructed electron densities obtained from cryo-EM tomography of authentic immature HIV-1 virions are shown on the left. The densities assigned to CA_{NTD} are shown in blue, those assigned to CA_{CTD} are shown in yellow, and the pillar-like density below was assigned to the putative SP1 domain helical bundle. A tomographic slice of an RSV VLP is shown on the right. Electron dense areas are white, while areas without electron density are black. The densities assigned to one unit of Gag are highlighted (CA_{NTD} in blue, CA_{CTD} in yellow, and SP in purple). Figure adapted from Wright et al. and de Marco et al. [9, 107].

adopts a more helical conformation to provide interaction surfaces for the assembling immature virion [102]. Our lab defined a similar critical region for immature assembly in RSV, starting just downstream of the structured CA_{CTD} and extending into SP [126]. Although RSV SPA and HIV-1 SP1 share little sequence homology, their proposed role as a molecular switch in stabilizing the immature hexamer is likely to be similar.

RETROVIRAL ASSEMBLY

Assembly in cells

Gag is the only protein required for the production of VLPs in cells. When transfected into cells, expression of Gag results in the production of VLPs that are physically indistinguishable from virions produced by infected cells. The fact that Gag is both necessary and sufficient for VLP assembly provides a means to study individual steps in immature retroviral assembly and maturation, which is often accomplished by introducing mutations in Gag and observing their effects on different steps in the immature assembly and maturation processes. For example, mutations that affect the curvature of the viral core lattice result in the accumulation of flat or dome-shaped sheets of Gag at the PM and can be easily visualized using thin-section TEM [128]. Another striking example comes from late domain deletion mutants, where the expression of these mutant Gag results in tethered VLPs and gives the surface of the transfected cell a blistered appearance when examined by scanning EM. Gag assembly in cells is a useful technique in the study of the late steps of the retroviral life cycle and Gag interactions with different cellular factors and components.

However, specific Gag-Gag and CA-CA interactions that drive self-assembly of retroviral cores are more difficult to tease out using cell-mediated Gag assembly. It is challenging to determine whether a mutant phenotype is the result of a disruption of bona fide Gag-Gag interactions, or the interaction between cellular factors and Gag. Conversely, if mutations lead to a wild-type immature assembly phenotype, it cannot be concluded that the mutation has no effect on immature assembly. In order to study both immature and mature assemblies at the molecular level, a better, more well-defined system was required, resulting in several groups independently developing in vitro VLP and capsid self-assembly systems.

General in vitro assembly

Much of the current knowledge on retroviral assembly emerged as a result of the development of in vitro assembly systems. These systems use purified retroviral structural proteins derived from Gag or CA to recapitulate immature and mature viral assembly in a cell-free environment. The resulting VLPs are identical to authentic virions when assessed using cryo-EM and radial density measurements [10]. Immature in vitro assembly using Gag or Gag-derived proteins generally results in the production of spherical VLPs similar in morphology to immature virions, although MPMV is a notable exception, as immature MPMV assembly can produce tubes as well as spheres, both with an immature Gag lattice. Capsid assemblies differ somewhat from bona fide retroviral capsids, but appear to maintain the same lattice organization. Mature in vitro assembly using CA generates capsid assemblies of various morphologies, including cones, tubes, or polyhedral structures under specific pH and salt conditions. VLPs and

capsid assemblies can be easily purified by centrifugation and visualized using negative-stain EM. Given the high degree of similarity between in vitro assembled VLPs and authentic virions, in vitro assembly has been used extensively to tease apart the protein-protein interaction involved in self-assembly. Currently, in vitro assembly is well established using Gag-derived and CA proteins from three model retroviruses: HIV-1, MPMV, and RSV, each of which will be described in detail in the following sections.

HIV-1 in vitro assembly

The first reported retroviral in vitro assembly used HIV-1 CA. Knowing that plant virus capsids, including that of tobacco mosaic virus, can self-assemble in vitro, Ehrlich et al. examined purified CA protein using negative stain EM. They noticed that the protein appeared to form branching fibrous structures approximately 200 nm in length as well as smaller spheres and rods. Using a high concentration of CA (10 mg/mL), the purified protein was dialyzed against various pH and salt conditions and it was found that HIV-1 CA assembles into structured tubes in acidic buffer with high salt (acetate buffer pH 6, 1.0 M NaCl) [129]. Tubular capsid assemblies have the same lattice organization as authentic mature virions and are inferred to represent a mature-like lattice in in vitro assembly, except in the case of MPMV [108, 109, 117].

The optimal conditions required for CA purification and in vitro capsid assembly were later defined by Gross et al., who showed that CA assembles most efficiently in alkaline high salt buffer (1M NaCl, pH 8.0) and at higher protein concentrations (4 mg/mL) [130]. Gross et al. also reported that a five amino acid N-terminal extension of CA is sufficient to convert the particle phenotype from tubular to spherical. The spherical

particles reported in this study did not have the electron-lucent hole observed in authentic immature virions and are likely some aberrant form of assembly. However, the observation led to the proposal that formation of the N-terminal salt bridge is essential for tubular assembly and, by inference, for the assembly of the immature lattice [131].

Immature HIV-1 assembly was shown for the first time by Campbell and Rein, who used an HIV-1 Gag lacking the C-terminal p6 domain to generate small spherical VLPs approximately 25 – 30 nm in diameter in high pH (pH 8.0) and low salt buffer (0.1 M NaCl) in the presence of RNA [104]. As authentic HIV-1 virions are normally 100 – 120 nm in diameter, these VLPs were obviously defective at some step in assembly. Campbell et al. later determined that binding to IP_5 and other inositol phosphate derivatives can trigger HIV-1 Gag assembly into VLPs of the correct diameter, and they isolated the region of Gag that binds to IP_5 to amino acids 16 – 99 [16]. As discussed above, HIV-1 Gag assembly may be regulated through conformational changes in the Gag monomer [102]. Therefore, binding to IP_5 may render the Gag monomer available for higher order oligomerization by releasing Gag from its horseshoe conformation into the assembly-ready extended conformation.

MPMV in vitro assembly

The MPMV in vitro assembly system was initially developed by Klikova et al., who were the first to show that immature assembly can also be recapitulated in vitro and who provided the first evidence that in vitro assembled particles have the same protein-protein interactions as virions collected from infected cells. They denatured the Gag protein found in inclusion bodies using 8 M urea and restored its function by

controlled dialysis. In basic Tris buffer (pH 8.8), approximately 50% of the Gag protein had changed from fully soluble to pelletable by centrifugation. The pelleted sample was run through a sucrose gradient and the majority of pelleted Gag was found in the fraction corresponding to a density of 1.19 g/mL. When this fraction was visually examined using negative-stain EM, spherical particles with an electron-dense ring surrounding an electron-lucent core were observed [132]. This morphology was very similar to immature betaretrovirus particles collected from infected cells [133], suggesting that in vitro assembled VLPs form via the same mechanisms that drive authentic immature assembly.

Since authentic MPMV Gag virions assemble in the cytoplasm, upon Gag expression in *E. coli*, assembled VLPs are abundantly observed in inclusion bodies. The morphology of these intracellular VLPs mimic in vitro assembled VLPs and the technique is often used to assess MPMV assembly phenotypes. Using the *E. coli* mediated assembly system, CA alone does not form any observable capsid assemblies, suggesting that the MPMV mature assembly pathway may differ fundamentally from that of HIV-1 and RSV [134]. However, a purified MPMV CA-NC construct lacking the N-terminal proline (Pro(-)CA-NC) can assemble in vitro into both tubes and spheres [117, 135], and this result can be recapitulated using the intracellular *E. coli* assembly system [134]. Since isolated MPMV virions have capped cylindrical capsids, it was initially believed that, as for HIV-1, tubular MPMV assembly reflects mature assembly, but the recent cryo-EM study by Bharat et al. showed that, despite their morphology, Pro(-)CA-NC tubes actually have an immature lattice [117].

RSV in vitro assembly

Campbell and Vogt worked out the first RSV in vitro assembly system using a Gag-derived protein containing the CA-NC region. CA-NC can assemble into irregular spherical particles when a concentrated sample of the protein (10 mg/mL) is dialyzed into acidic buffer (pH 5.5). However, upon addition of RNA, CA-NC (pH 5.5) and the HIV-1 Gag-derived CA-NC-p6 (pH 8.0) assembled into tubes with a constant diameter of 30 nm or 50 nm, respectively. In the presence of RNA, these constructs assembled more efficiently than purified protein alone, requiring only 1 mg/mL of protein instead of 10 mg/mL. This was the first study to show that nucleic acid plays an important structural role in retroviral assembly by enhancing the efficiency of assembly [15]. Campbell and Vogt also developed a highly robust immature assembly system using a Gag construct lacking the MBD of MA and the PR domain, henceforth referred to as ΔMBDΔPR. This protein is easily and abundantly purified from *E. coli*, as it remains in the soluble fraction of the cell lysate and can be separated from most cellular proteins by charge, using the positive charges in NC. ΔMBDΔPR assembles optimally in slightly acidic conditions (pH 6.5) at low salt concentration (0.1 M NaCl) with approximately 10% wt/wt nucleic acid [13].

A system for RSV capsid assembly was first described by Purdy et al. to study the role of the MHR [71]. CA was discovered to assemble at a concentration of 2 mg/mL in the presence of 0.5 M sodium phosphate. It was found that capsid assembly involves a lag time that can be shortened by increasing CA concentration, increasing buffer pH, or by changing the type of buffer, with sodium phosphate dibasic being the most

favorable for assembly [136]. The authors took this finding to mean that capsid assembly requires nucleation, specifically the formation of CA protein dimers. Aside from assembling into particles that resemble authentic capsids, CA assembly in vitro also generated small icosahedral particles that allowed for CA pentamers to be visualized for the first time [113].

THESIS OUTLINE

In this dissertation, I will discuss my work which centers on the structural contribution of the SPA domain to immature assembly. Chapter two outlines the materials and methods used to perform the experiments described in this dissertation. Chapter three first addresses the structure of SPA using mutations coupled with in vitro assembly as the readout. Secondary structure prediction algorithms assign a high helical propensity to the SPA domain. When individual residues were mutated to alanine, in vitro assembly was abrogated with a majority of the mutated proteins, but not all of them. The locations of the tolerated mutations suggested that the predicted SPA helix may form a higher-order structure in the assembled VLP, and this hypothesis was strongly supported by HDX data generated by my collaborators in the Prevelige lab at the University of Alabama at Birmingham. I followed up on the SPA region by performing CD analyses on a peptide that covers the SPA domain and showed that it became more structured with increasing concentration, supporting our hypothesis that the SP region is unstructured in the Gag monomer, but becomes structured in the assembling VLP. The peptide was found to associate as a hexamer when analyzed by sedimentation velocity, again by our collaborators in the Prevelige Lab. These data

support the hypothesis that SP forms a six-helix bundle that ties together the immature hexamer.

Chapter four describes the various protein constructs that I used in my attempts to crystallize the immature RSV core, as well as my ongoing efforts to obtain a crystal structure of the SPA crystal. I showed in this chapter that despite the inherent flexibility in the two-domain CA and its flanking sequences in Gag, it is possible to crystallize a protein spanning the minimal regions required for immature assembly. I also crystallized the SPA peptide, but the crystals proved to be of insufficient quality for diffraction. Finally, chapter five consists of a discussion of the contribution of in vitro assembly to the study of retroviral assembly, the implication of molecular switches in assembly, and some future experiments of relevance to the projects presented in this dissertation.

CHAPTER TWO

MATERIALS AND METHODS

DNA constructs for SP mutation analyses

All alanine point mutations and insertion/substitution mutations were made using site-directed mutagenesis and the previously described ΔMBD.ΔBlpI.pET3xc construct [72], which was derived from the ΔMBD.pET3xc plasmid [13]. The mutations D_{472}A/C – I_{475}A/C were made by replacing the BlpI – SacII fragment of ΔMBD.ΔBlpI.pET3xc with the products of site-directed mutagenesis. A_{476}S/C – A_{478}S/C were made by replacing the BlpI – KpnI fragment, while M_{479}A/C – N_{492}A/C were made by replacing the SacII – KpnI fragment.

The insertion mutant 3Ala was made by replacing the BlpI/SacII fragment with the product of site-directed mutagenesis. 7Ala1 was made using a version of ΔMBD.ΔBlpI.pET3xc containing a MluI site after the first four residues of the NC domain. The plasmid was digested using MluI and double-stranded oligonucleotides coding for seven alanines and containing the appropriate sticky ends were ligated in. 7Ala2 and 7AA were made by replacing the SacII/KpnI fragment with the products of site-directed mutagenesis. The substitution mutant CMV1 was made by replacing the BlpI/KpnI fragment, CMV2 and SP1ins were made by replacing the SacII/KpnI fragment, and SP1sub was made by replacing the BlpI/SacII fragment.

All cysteine point mutations were made in the same manner as described for alanine mutations, but used the ΔMBD.ΔBlpI (-11C) pET3xc plasmid as PCR template and vector backbone. All PCR products were amplified using ΔMBD.ΔBlpI.pET3xc as

the template. All final constructs used for protein expression were confirmed by sequencing.

DNA constructs for “minimal Gag” constructs

SUMO-25CASP was made using ΔMBD.ΔBlpI.pET3xc as the PCR template, an upstream mutagenic primer that included a BamHI site, and a downstream mutagenic primer containing a HindIII site. The PCR fragment included the 25 required residues from p10, all of CA, all of SP, and the first four residues of NC. Using BamHI and HindIII, the PCR fragment was inserted into a digested pSUMO backbone immediately downstream of the SUMO tag. SUMO-25CASP(CC) was made in the same way as 25CASP, but the ΔMBD.ΔBlpI (-11C) pET3xc construct containing the E51C/T20C mutation was used as the PCR template [72]. SUMO-Ccmk2-25CASP was made using 25CASP.pSUMO as the backbone and a Ccmk2 PCR fragment containing a BamHI site at the N-terminus and an EagI site at the C-terminus was generated using a Ccmk2.pET22b plasmid (generously shared by Todd Yeates) as vector and two mutagenic PCR primers containing the aforementioned restriction sites [137]. All final constructs used for protein expression were verified by sequencing.

Protein expression and purification using pET3xc vector

All ΔMBDΔPR-derived mutants were expressed in *E. coli* BL21 cells. For each protein, 500 mL of culture was grown to log phase and induced using 0.5 mM IPTG for 4 hours at 37ºC. Cells were pelleted by centrifugation in a Sorvall GSA rotor and resuspended in 25 mL high salt buffer (20 mM Tris, pH 8.0, 500 mM NaCl) containing 10 mM DTT, 1 mM PMSF, and 20 μM $ZnCl_2$, lysed by sonication and the lysate was

spun at 45,000 rpm for 3 hours using a Beckman Ti50.2 rotor to spin down ribosomes. ΔMBDΔPR and mutants are mostly found over-expressed in the supernatant. The over-expressed proteins were first precipitated using saturated ammonium sulfate to a final concentration range of 25% to 30%, and spun at 9000 rpm for 10 minutes using a Sorvall SA-600 rotor to spin down the precipitated protein. The pellet was then resuspended in 10 mL low salt buffer (20 mM Tris, pH 8.0, 100 mM NaCl), filtered by syringe (0.45 µm pore), and further purified using a 1 mL cation exchange column (Amersham SP) mounted to an ÄKTA prime FPLC pump (Pharmacia Amersham). The protein was bound to the column with low salt buffer at a rate of 1.0 mL/minute, the column was washed with 10 mL low salt buffer, and the protein was eluted as a single fraction with high salt buffer. Concentrations of the purified proteins were determined by spectrophotometric measurements (260/280 nm) using either a Pharmacia Biotech Ultrospec 2000 or a Thermo Scientific NanoDrop 2000.

Protein expression and purification using pSUMO vector

All SUMO-linked proteins were expressed in *E. coli* BL21 cells. For each protein, 1 L of culture was grown to log phase and induced using 0.5 mM IPTG for 4 hours at 30ºC. Cells were pelleted by centrifugation in a Sorvall GSA rotor and resuspended in 50 mL binding buffer (50 mM Na2HPO4, pH 8.0, 1 M NaCl), lysed by sonication, and the lysate was spun at 9000 rpm for 30 minutes using a Sorvall SA-600 rotor. The supernatant was bound to 1 mL of Ni-NTA beads (QIAGEN Ni-NTA agarose beads, approximately 2 mL of bead slurry) for a minimum of 4 hours (up to overnight) at 4 ºC while gently rocking. The mixture was poured onto a chromatography column and the

beads were allowed to settle. The lysate was allowed to flow through at a rate of approximately 1 mL/minute and was collected for later analysis. The beads were washed using 20 mL wash buffer (50 mM Na2HPO4, pH 8.0, 1 M NaCl, 10 mM imidazole), also at a rate of 1 mL/minute, and the SUMO-tagged protein was eluted from the beads using 5 – 10 bead volumes of elution buffer (50 mM Na2HPO4, pH 8.0, 1 M NaCl, 250 mM imidazole).

In vitro assembly and EM

When protein concentrations were equal to or above 5 mg/mL, the protein was assembled by dilution to 1 mg/mL in 50 mM MES, pH 6.5 after additions of 10% wt/wt of the ssDNA GT50 (a 50mer oligonucleotide composed of 25 GT repeats). Assembly reactions were complete after incubating at room temperature for 15 minutes. When protein concentration was below 5 mg/mL, assembly was performed by first diluting the protein to 1 mg/mL in 50 mM MES, 100 mM NaCl, pH 6.5. GT50 was added (10% wt/wt) and 50 µL of the protein-oligonucleotide mix was dialyzed overnight at 4ºC against 50 mM MES, 100 mM NaCl, pH 6.5. Ten microliters of each assembly reaction was spotted onto a carbon-Formvar-coated copper grid for one minute, and then stained using a 2% uranyl acetate solution for one minute. The samples were examined using a Morgagni 268 transmission electron microscope.

SPApep sequence and circular dichroism

The 28 or 29 amino acid SPApep, which encompasses residues P_{469} – D_{496} of RSV Gag with or without an exogeneous cysteine was purchased from the Tufts University Core Facility (http://www.tucf.org/peptidesynthesis-f.html). The peptide was

synthesized using FastMoc Chemistry on an ABI 431 Peptide Synthesizer and further purified by HPLC. For TFE-dependent structural studies, the purified, lyophilized peptide was re-suspended in distilled H_2O to a concentration of 10 mg/mL. Peptide concentration was calculated by measuring absorbance at 205 nm and using an extinction coefficient of 31 for 1 mg/mL [138]. SPApep was diluted into 10 mM sodium phosphate buffer, pH 6.5, 0 – 30% TFE just prior to recording the CD spectrum. Data were collected in the region of 190 – 250 nm with a 1 mm path length Hellma quartz cell and an AVIV 202 CD spectrometer (25ºC, 1 nm wavelength step, 10 s averaging time). For concentration-dependent studies in aqueous buffer, SPApep was diluted into buffer without TFE and measured as described above. For CD measurements at high peptide concentrations (10 mg/mL), a shorter path length (0.1 mm) quartz cell was used.

The percent helicity was calculated using an Excel version of PEPFIT, generously provided by Michael Amon [139, 140]. For each CD spectrum, the best fit was obtained by selecting a counter setting of 4 for the initial round of analysis and eliminating reference curves that do not contribute to the spectrum. A second round of fitting was performed using a counter setting of 1. For each peptide concentration, the percent α-helix value was taken from the best fit.

Modeling the 6HB using Swiss PDB Viewer

The model was built using Swiss PDB Viewer. The raw amino acid sequence of the 28 residue SPApep (without the exogenous cysteine) was imported into Swiss PDB Viewer. The program automatically assigned residues L_{470} – R_{495} as α-helical and showed SPApep as a single helix. SPApep helix was spatially aligned with each

individual chain of CC_{hex} (PDB ID 3R3K) using the Magic Fit tool and the alignments were manually adjusted using the Clustal Omega sequence homology. The six resulting SPApep helices were then combined and subjected to energy minimization three times to minimize side-chain clashes.

VLP purification via centrifugation

In vitro assembled VLPs were purified for cryo-EM studies by centrifugation through a 10% - 60% sucrose gradient to equilibrium. The sucrose-containing solutions were prepared using assembly buffer (50 mM MES, pH 6.5, 100 mM NaCl) and stored at 4ºC. Gradients were built immediately before use in Beckman thinwall Ultra-clear 4 mL centrifugation tubes. The overall volume pipetted into each tube was 3.6 mL. The volume of each gradient step was calculated to be (3.6 mL – assembly volume) divided by eight. VLP assemblies were layered above the sucrose gradient and spun in a Beckman SW60 rotor at 50,000 rpm for 3 hours. The VLP fraction formed a visible band that was collected manually (200 – 400 µL) after suctioning the sucrose solution layered above. The collected fraction was diluted into assembly buffer to a total of 3.6 mL (minimum of 1:5 dilution) and VLPs were pelleted by centrifugation in the SW60 rotor at 50,000 rpm for 1 hour. The VLPs were then resuspended in roughly 50 µL assembly buffer.

CHAPTER THREE

MUTATIONAL ANALYSIS OF THE RSV SP ASSEMBLY DOMAIN

expanded from Bush et al. (2013), manuscript in preparation

INTRODUCTION

The orthoretroviral Gag protein self-associates to form the immature retroviral core during assembly. Gag is a polyprotein consisting of the major domains MA, CA, and NC and several minor domains that differ among retroviruses. During the late stages of the virus life cycle, retroviral particles bud from the infected cell and maturation takes place. The transition from the immature Gag core to the mature CA capsid is brought about by PR-mediated cleavage of Gag during or soon after budding. Proteolysis of Gag releases its constituent protein domains. The liberated CA protein then reassembles to form the shell of the mature capsid [10, 106].

Both immature Gag assembly and mature CA assembly can be recapitulated in vitro using purified proteins. Mature in vitro capsid assembly has been established using purified CA protein from lentiviruses [129, 130] and alpharetroviruses [57, 71, 113, 141]. The products of mature assembly can take on a variety of morphologies, including cones (HIV-1), tubes (RSV, HIV-1) or smaller polyhedral structures (RSV) under specific pH and salt conditions. Although the tubes are not outwardly similar to mature capsids, they conserve the CA lattice observed in retroviral protein cores [108, 111]. Immature VLPs have been successfully assembled using purified Gag-derived proteins from members of the lentiviridae [10, 104, 131, 142-144], alpharetroviridae [13, 15, 76, 145, 146] betaretroviridae [132, 134, 147], and gammaretroviridae [102, 148] genera.

Immature assembly usually yields spherical particles, with the exception of MPMV, which is able to form tubes under certain conditions [117, 149]. VLPs formed by in vitro assembly are structurally comparable to virions collected from infected cells [8, 10, 106, 111, 150, 151], lacking only the viral membrane.

The immature Gag viral core and the mature CA capsid have a similar protein lattice composed of hexameric subunits. The immature Gag lattice curves into an incomplete sphere in the immature viral core [8, 14, 107], while the mature CA lattice closes into a fullerene structure by incorporating 12 pentameric CA subunits into the mature lattice [108, 111-113]. In both the immature and mature lattices, the two independently-folding domains of CA, the CA_{NTD} and the CA_{CTD}, are the main constituents of both the pentameric and hexameric lattice subunits [9, 60, 107, 113, 116, 117, 121]. Although the general lattice geometry is similar in immature and mature retroviral cores, the details of the CA lattice show several differences. These differences in protein-protein contacts in part result from intermolecular interactions that are absent in the cleaved CA protein. In the case of RSV, mutational studies in Gag including disulfide cross-linking followed by structural modeling showed that one such critical interaction takes place between p10, the domain immediately upstream of CA, and CA_{NTD}, and serves to stabilize the immature hexamer [13, 59, 72, 146].

Another critical intermolecular interaction for immature assembly is found immediately downstream of the CA domain. In RSV, the cleaved SP is found in this region. Mutational studies have defined a 24 amino acid region including SP and flanking residues to be required for proper immature assembly, as insertion mutations in

this region result in tubular assembly [126]. By cryo-EM tomography, RSV VLPs show an organized hexameric lattice consisting chiefly of the CA domain, with periodic stalk-like densities below the lattice [107]. I hypothesize that the structure underlying these densities is formed by amino acid residues immediately downstream of the folded CA, namely the last eight amino acids of CA, all of SP, and the first four residues of NC. I have named this region the SP assembly (SPA) domain, and I propose that neighboring SPA domains in a Gag hexamer coalesce to form a six-helix bundle (6HB).

Using in vitro assembly, alanine scanning mutagenesis, and biophysical analyses, I have further characterized the structure and function of SPA. Most of the amino acid residues in SPA could not be mutated individually without abrogating assembly. I interpret these results to mean that the amino acids that do not tolerate mutations contribute to higher-order structures in VLPs. HDX analyses of unassembled Gag in comparison with assembled VLPs showed strong protection at the SPA region, consistent with a higher-order structure. Circular dichroism revealed that a 29mer SPA peptide shifts from a random coil to a helix in a concentration-dependent manner. Analytical ultracentrifugation (AUC) showed concentration-dependent self-association of the peptide into a hexamer. Taken together, these results provide strong evidence for the formation of a crucial 6HB in Gag assembly.

RESULTS

The RSV SP assembly domain is highly sensitive to alanine point mutations

To guide structural studies of the RSV SPA domain, a string of amino acids including SP and adjoining sequences was submitted to PsiPred

(http://bioinf.cs.ucl.ac.uk/psipred/) for secondary structure prediction. Similar to previously reported results generated using an earlier version of PsiPred [126], 21 residues, including the C-terminal five residues of CA, all of SP, and the N-terminal four residues of NC, were predicted to form an α-helix (Figure 3.1). I have termed this predicted structure the SPA helix. Although the region surrounding SP and its role in immature assembly have been studied previously using five- or two-amino acid insertion mutations coupled with baculovirus-mediated expression of Gag in insect cells [126], that study only defined the boundaries of the SPA domain and showed the entire domain to be indispensable for assembly.

My interest is the structural contribution of the predicted SPA helix to the immature core. More specifically, I wanted to know if the SPA helix is able to form a 6HB. To further probe the structural function of the SPA helix, I performed alanine scanning mutations on its 21 constituent residues (existing alanines were mutated to serines). Each mutant was made in a well-characterized, truncated version of Gag known as ΔMBDΔPR (Figure 3.1) that has been previously shown to assemble robustly in vitro and to form VLPs nearly identical to wild-type immature virions [13, 72, 76, 80, 106]. The mutant proteins were expressed in E. coli, purified, and placed in assembly conditions in the presence of a long DNA oligonucleotide. Negative stain transmission electron microscopy (TEM) was used to evaluate the ability of the proteins to assemble. The mutants were scored as either positive or negative for assembly based on the presence or absence of VLPs, and each protein purification and assembly was performed at least twice. Table 3.1 summarizes the results of this mutagenesis. Assembly-positive mutants were defined as mutations that resulted in at least five VLPs

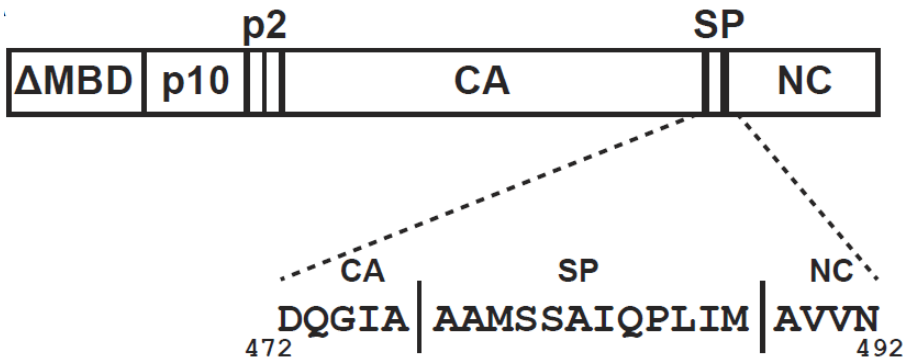


Figure 3.1: The SPA helix and its location within ΔMBDΔPR. The SPA helix sequence was determined from secondary structure prediction and is enumerated from the first residue of RSV Gag.

Mutation	Assembly	Mutation	Assembly
$D_{472}A$	+	$I_{483}A$	-
$Q_{473}A$	+	$Q_{484}A$	-
$G_{474}A$	+	$P_{485}A$	-
$I_{475}A$	-	$L_{486}A$	-
$A_{476}S$	-	$I_{487}A$	-
$A_{477}S$	+	$M_{488}A$	-
$A_{478}S$	-	$A_{489}S$	-
$M_{479}A$	-	$V_{490}A$	-
$S_{480}A$	+	$V_{491}A$	+
$S_{481}A$	+	$N_{492}A$	+
$A_{482}S$	-		

Table 3.1: Alanine scan mutants and their assembly results. All mutations were made within ΔMBDΔPR. Mutations that allow assembly are indicated with a “+” symbol. Mutations that abrogate assembly are indicated with a “-” symbol.

per EM grid. Assembly of the wild type protein ΔMBDΔPR was used as a control and invariably resulted in many VLPs per field (Figure 3.2). The assembly-negative phenotype usually appeared as large regions of protein aggregation with no sign of VLPs (representative example shown in Figure 3.2). Some assembly-positive mutants could assemble at near wild-type efficiency, but more often produced fewer particles than ΔMBDΔPR assembly (Figure 3.2). None of the mutations resulted in other morphologies, for example tubular assembly.

Residues around and including RSV SP form an amphipathic helix when plotted into a helical wheel [107]. We generated a helical wheel using the extended SPA helix sequence and found that this longer sequence also forms an amphipathic helix with a large hydrophobic face and a smaller hydrophilic face (Figure 3.3). With the exception of $D_{472}A$ (the first residue on the helical wheel), all the mutations that did not disrupt assembly (indicated by asterisks) are located on the hydrophilic face of the helix. Furthermore, it appears that every residue on the hydrophilic face, except Q_{484}, can be mutated without compromising assembly. This finding is in agreement with the idea that the SPA helix contributes to immature assembly by forming an amphipathic helix that mediates Gag-Gag interaction through hydrophobic contacts.

Cysteine point mutation results mostly mirror alanine mutations

In addition to the alanine and serine mutations, I also mutated each residue in SPA helix to cysteine. Cysteine mutants were all made in ΔMBDΔPR(-11C), in which 11 of the 12 native cysteines were mutated to alanine or serine, with the exception of one buried and presumed non-reactive cysteine in CA_{CTD} [72]. The rationale behind the

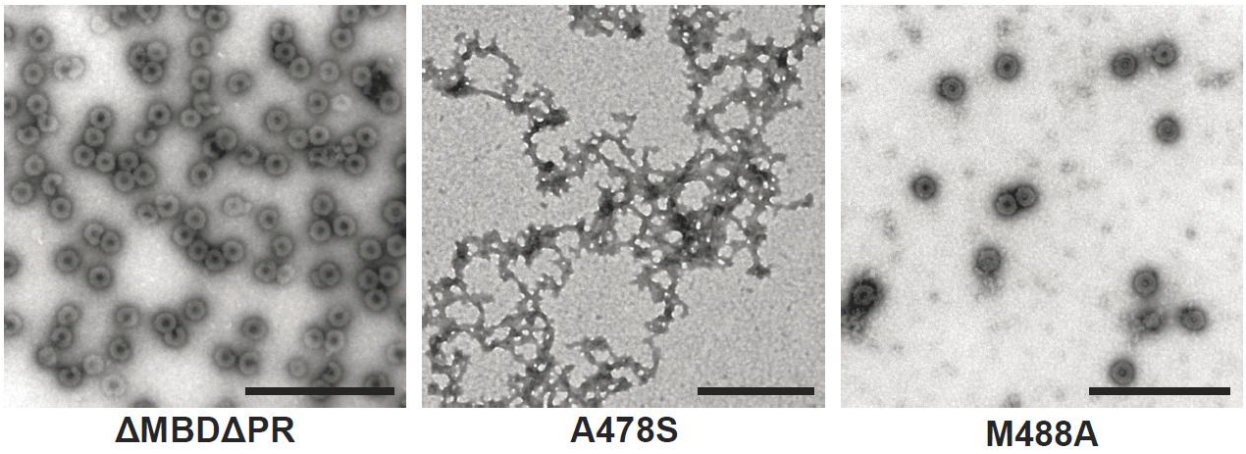


Figure 3.2: Examples of observed in vitro assembly using negative stain EM. ΔMBDΔPR (wild type protein, control). Assembly-negative mutant (A478S). Assembly-positive mutant (M488A). Scale bars are 500 nm.

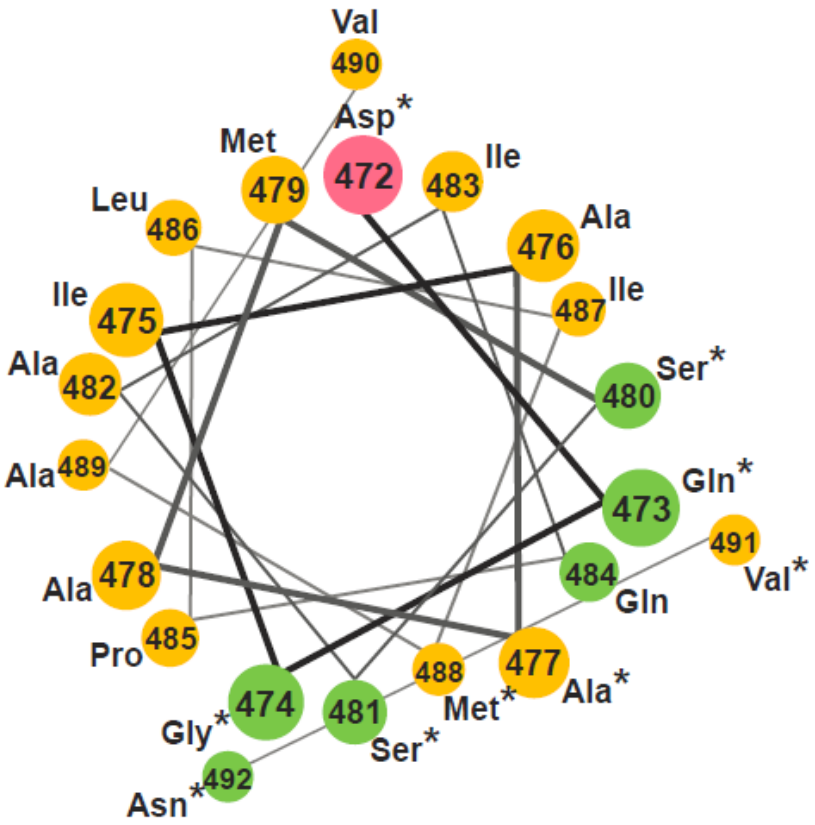


Figure 3.3: Helical wheel projection of SPA helix sequence. Each residue is enumerated according to its position within full-length RSV Gag. Polar, uncharged residues are green, negatively charged residues are pink, and non-polar/hydrophobic residues are yellow. Residues indicated by asterisks can be mutated to alanine without abrogating *in vitro* assembly.

cysteine mutagenesis was two-fold. First, for most residues in the SPA helix, changing the native amino acids to cysteine is a more disruptive structural change than changing them to alanine and should result in a loss of assembly. However, the more conservative mutations (serine, valine, and methionine to cysteine) may lead to different assembly outcomes when compared to alanine mutations. Second, the presence of a free sulfhydryl group is conducive to labeling and cross-linking experiments.

The cysteine mutation results mostly mirror those from the alanine scan, with a few notable exceptions (Table 3.2). The first three residues (D_{472}, Q_{473}, G_{474}) at the N-terminus of SPA helix did not tolerate mutation to cysteine, but did tolerate mutation to alanine. The opposite was observed with V_{490}. D_{472} is likely the helix-capping residue, and mutation to cysteine may have been too disruptive to its role in stabilizing the N-terminus of the predicted helix, although it is difficult to envision why mutation to alanine is more readily tolerated. Although they are located on the hydrophilic face of the predicted helix, Q_{473}C, and G_{474}C are assembly-defective mutations. This is likely because the nature of the side-chains is dramatically changed and any side-chain interactions mediated by these residues outside of those found in the presumed 6HB may be disrupted in the assembling VLP. V_{490}C is a more conservative mutation, which may explain why it is a tolerated mutation in assembly, while V_{490}A is not.

Insertion or substitution mutations in SPA mostly abrogate assembly

Whether the SPA helix can be replaced by another helical segment is unknown. It is also unknown if the helix can be extended by a turn or two without affecting its function in the protein. I assumed that if the role of the SPA helix is simply a structural

Mutation	Assembly	Mutation	Assembly
D472C	-	I483C	-
Q473C	-	Q484C	-
G474C	-	P485C	-
I475C	-	L486C	-
A476C	-	I487C	-
A477C	+	M488C	+
A478C	-	A489C	-
M479C	-	V490C	+
S480C	+	V491C	+
S481C	+	N492C	+
A482C	-		

Table 3.2: Cysteine point mutants and their assembly results. All mutations were made within ΔMBDΔPR(-11C). Mutations that allow assembly are indicated with a "+" symbol. Mutations that abrogate assembly are indicated with a "-" symbol.

one, meaning that merely the presence of an α-helix suffices to satisfy its contribution to the virion, then it should be replaceable with a helical segment from another protein or the length of the helix should be modifiable without severely compromising assembly. However, if the SPA helix forms homotypic higher-order structures like a 6HB or interactions with more distant parts of Gag, then it may need to be amino acid-specific. To address this question, I tested a panel of nine insertion or extension mutants for their ability to assemble. A description of each mutant can be found in Table 3.3.

The insertion mutants are as follows: 3Ala, 7Ala1, 7Ala2 and 7AA. Of these mutations, the three alanine insertion mutants were made based on publications showing that alanine-based peptides are stably α-helical [152] and that alanine is helix-stabilizing [153]. In particular, 3Ala was based on results from Keller et al., who showed that insertion of up to three alanine residues between residues D472 and Q473 does not affect Gag assembly in a baculovirus system [126]. The 7AA mutant elongates the SPA helix by repeating the last seven amino acids (LIMAVVN) of the predicted helix. The substitution mutants are: CMV1, CMV2, SP1sub, and SP1ins. CMV1 and CMV2 substitute part or all of SPA helix with a helical segment of the cucumber mosaic virus (CMV) capsid. In the assembled CMV capsid, this helical segment is known to form a parallel 6HB, as determined by X-ray crystallography [154]. SP1sub and SP1ins, respectively, substitute or insert a sequence from HIV-1 SP1 with high predicted helical propensity into the SPA helix.

Of these eight mutations, only three are assembly-competent: 3Ala, 7AA, and SP1sub. Why these three mutants assembled but others did not is unclear, but it may

Name	Description	Assembly
3Ala	Inserts three Ala residues between D472 and Q473	+
7Ala1	Inserts seven Ala residues after predicted SPA helix	-
7Ala2	Inserts seven Ala residues between S480 and S481	-
7AA	Inserts LIMAVVN after M488	+
CMV1	replaces entire SPA predicted helix with CMV helix	-
CMV2	replaces A478 – V491 with CMV helix	-
SP1sub	replaces Q473 – A477 with KARVLAE from HIV-1 SP1	+
SP1ins	Inserts ARVLAEA from SP1 between M488 and A489	-

Table 3.3: Insertion and substitution mutants in SPA helix and their assembly results. All mutations were made within ΔMBDΔPR. Mutations that did not disrupt assembly are indicated with a "+" symbol. Mutations that abrogate assembly are indicated with a "-" symbol.

be that they are the least disruptive to the helical propensity of the SPA domain. It is also possible that SPA helix forms part of a higher-order structure and mutations that knock the helix out of phase may prohibit that structure from forming. Finally, it is worth noting that both 3Ala and SP1sub insert residues after D_{472}, the first residue in the predicted helix, and 7AA inserts residues at the end of SPA helix. Perhaps changes to helical structures are less deleterious when they are made at the helix ends.

A conserved predicted helix is found downstream of the structured CA_{CTD}

The structure of CA is highly conserved among retroviruses [56, 62]. The CA_{NTD} is composed of seven α-helices that fold into a globular shape and is linked to the CA_{CTD} via a flexible linker. CA_{CTD} consists of four α-helices, and X-ray structures have shown that the region downstream of the last (11^{th}) helix is unstructured. However, given that the unstructured region is highly sensitive to mutation, I wanted to determine if there is actually any structural conservation that was not detected by previous studies. To this end, I collected known Gag sequences from 16 orthoretrovirus species and submitted a structurally conserved portion of Gag, from the initiating proline in CA to the first zinc-finger in NC, to Psipred for secondary structure prediction. Interestingly, the program identified an additional predicted 12^{th} helix downstream of the known structured CA_{CTD} in all cases.

Lentiviruses and alpharetroviruses have a cleaved spacer peptide between the CA and NC domains of Gag. In these two genera, the predicted 12^{th} helix either covers the entire spacer region or spans the junction between CA and spacer. Betaretroviruses and deltaretroviruses have no cleaved spacer, but the predicted 12^{th} helix also spans

the cleavage site between CA and NC. Gammaretroviruses also have no cleaved spacer, but there is a long stretch of highly charged amino acids immediately downstream of the structured CA_{CTD} that has been inferred to form a α-helix based on mutational studies [127]. Indeed, the predicted helix spans the length of the highly charged region and either overlaps the CA-NC cleavage site or is located just upstream of the cleavage site. The data suggest that there may be a conserved helix located downstream of the known structured CA_{CTD} that is not present in monomeric Gag, but that takes shape in the assembled Gag lattice.

A model of the putative 6HB in assembled VLPs

When RSV VLPs are studied using electron microscopy, the immature Gag hexamer appears to consist of six Gag molecules interacting laterally [107]. This arrangement is a conserved property of Gag assembly and has been confirmed for other retroviruses that have been studied at this level, notably MPMV and HIV-1 [8, 9, 14, 23, 107, 117, 155, 156]. As the results from my mutations indicate that the SPA helix can form an amphipathic helix in immature assembly, I hypothesized that six of these helices organize into a parallel 6HB, as originally suggested for HIV-1 by Wright et al. [9], and that opposing sides of the large hydrophobic face are involved in contacts that tie the bundle together (Figure 3.4, opposing sides designated I and II). In this model, the hydrophilic face of each helix is oriented toward the outside of the helical bundle and therefore is not involved in inter-helix contacts, explaining its lack of sensitivity to mutation.

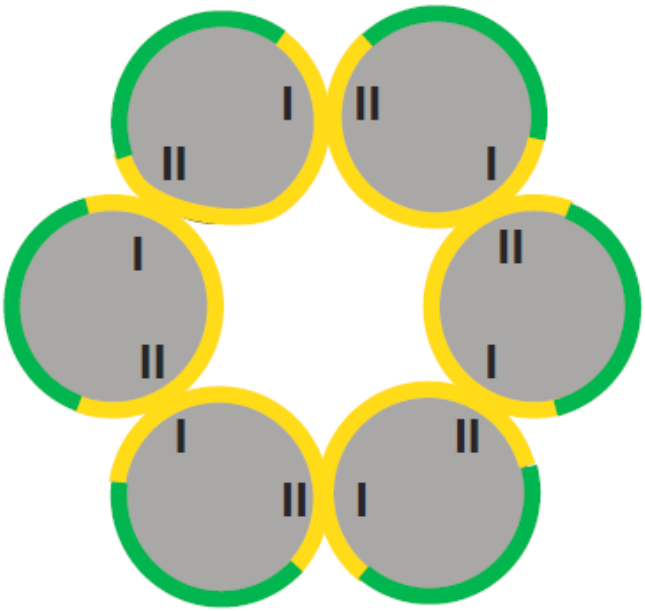


Figure 3.4: Conceptual model of the putative RSV SPA 6HB. The numbers “I” and “II” designate opposite, but putatively interacting sides of the large hydrophobic face. The approximate location of polar, uncharged residues on each helix is green and that of non-polar/hydrophobic residues is yellow.

The SPA domain is highly protected against HDX in assembled VLPs

The data from my point mutations suggest that the SPA helix may be capable of forming homotypic interactions with neighboring SPA helices in the context of an assembled RSV VLP. To further probe the structural contribution of the SPA helix in assembly, my collaborators in the Prevelige Lab at the University of Alabama at Birmingham used HDX MS to examine this segment of Gag, comparing free ΔMBDΔPR protein with assembled VLPs. HDX MS allows the monitoring of protein structure and dynamics by measuring the rates of exchange of protein backbone amide hydrogens with hydrogens from the solvent. By diluting protein into a deuterated buffer, these exchange events result in a 1 Da mass shift (H to D) for each exposed amide, which is readily detectable with mass spectrometry. The rate of this exchange is dependent on secondary, tertiary, and quaternary structure. Regions of a protein involved in more stable folds and interactions exchange more slowly than regions that form less stable structural elements. While this technique cannot provide explicit structural detail, comparisons of exchange patterns provide valuable insight into the dynamics, stability, and intermolecular interfaces of protein assemblies, especially in those cases where the atomic level structure of a protein is unknown.

My collaborators compared the HDX profiles of ΔMBDΔPR in unassembled (free) and assembled (VLP) form, which I purified, assembled, and shipped to them on multiple occasions. The exchange plots illustrate several notable differences in HDX across the two states. Most important, comparisons of peptides from the SP domain showed very high protection following assembly (Figure 3.5). This level of protection is

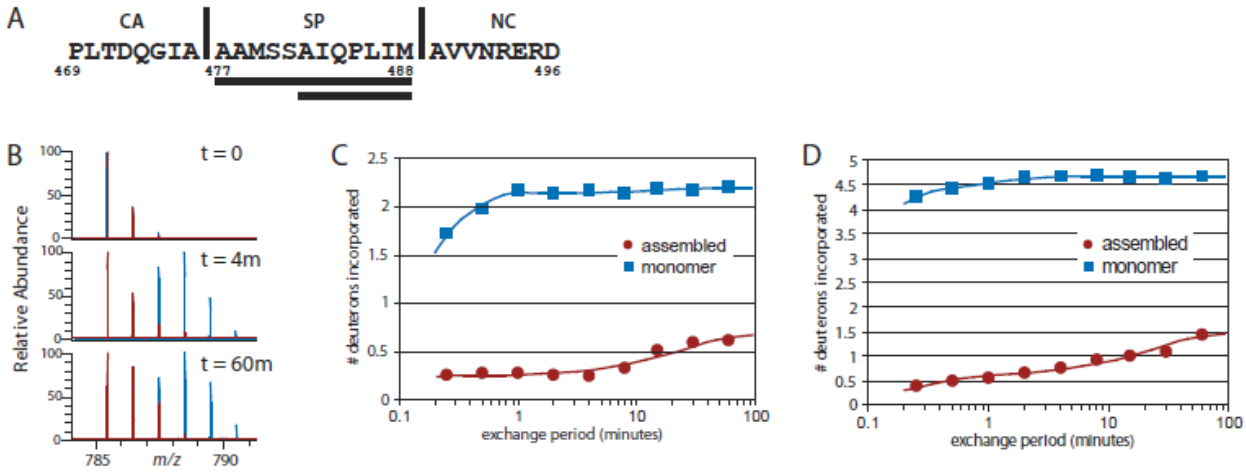


Figure 3.5: HDX plots comparing free ΔMBDΔPR to assembled VLPs. A. The two peptic fragments corresponding to regions in RSV SP are underlined. Vertical lines correspond to natural protease cleavage sites in Gag. B. Mass to charge spectra for the shorter peptide 482 – 488 in unassembled ΔMBDΔPR protein (blue) and assembled VLP (red). C. Plot of deuterium incorporation in peptide 482 – 488 and D. peptide 477 – 488 at different time-points.

suggestive of a very stable ultrastructural element, and thus is consistent with the proposed helical bundle motif [107]. Although our collaborators were not able to find a peptide that spanned the entire length of the SPA helix, they were able to examine two relevant peptides: a longer fragment spanning residues 477 - 488 (AAMSSAIQPLIM) and a shorter one including residues 482 – 488 (AIQPLIM). Both showed very similar exchange and protection patterns. This result indicates that the structural stability afforded by the assembled lattice is likely to entail the complete SP domain and not just the N- or C-terminal portion. Furthermore, the low level of protection of the same region observed in the unassembled protein suggests that this region may be unstructured before assembly, consistent with previously published structures of CA or CA fragments that showed an unstructured coil downstream of CA_{CTD} helix 11 [53, 54, 57].

SPApep becomes more structured with increasing peptide concentration

To investigate the mechanism underlying the structural transformation of the SPA helix during assembly, I performed circular dichroism (CD) spectroscopy on a 29 amino acid peptide (PLTDQGIAAAMSSAIQPLIMAVVNRERDC) named SPApep. It includes the last eight residues of CA, all of SP, the first eight residues of NC, and an extra C-terminal ectopic cysteine for potential labeling experiments. Preliminary experiments in the presence and absence of 1 mM tris(2-carboxyethyl)phosphine (TCEP, a reducing agent) revealed no difference in the CD spectra. Therefore, I inferred that the additional cysteine residue does not interfere with peptide structure (data not shown).

I measured the CD spectra of SPApep at various concentrations in aqueous buffer, from 0.033 mM (0.1 mg/mL) to 3.3 mM (10 mg/mL). These experiments were

repeated by my collaborators in the Prevelige lab, as they have the 0.1 mm path length Hellma quartz cuvette required to take CD measurements at high concentrations and far-UV wavelengths. Previously published work showed that the CD spectra of a similar peptide based on the HIV-1 SP1 region became more helical as peptide concentration was increased [102]. Due to the structural similarity between HIV-1 and RSV VLPs [106, 107], I predicted that SPApep should exhibit similar behaviors when its concentration is increased. Indeed, the CD spectrum of SPApep began to shift from coil to helix at 1.6 mM (5 mg/mL) and became significantly more helical at 3.3 mM (Figure 3.6). At 3.3 mM, a noticeable maximum at 197 nm and local minima at 206 nm and 222 nm appeared in the spectrum, with the calculated helicity of the peptide increasing to 23% (Figure 3.6). SPApep became more α-helical at a concentration similar to what was previously reported using the HIV-1 SP1 peptide [102]. My finding suggests that the SPA helix region becomes more structured within an assembling virus particle, because the local Gag concentration might reach values around 4 mM, as roughly calculated using an average immature viral diameter of 129 nm and 2500 Gag proteins per virion [106].

To definitively rule out any contribution of cysteine cross-linking leading to unwanted nucleation, my collaborators at the Prevelige lab repeated the CD measurements using a version of SPApep that did not contain the exogenous cysteine and saw nearly identical CD spectra for the cysteine-free peptide (Figure 3.7).

SPApep becomes more structured with increasing TFE concentration

To further probe the concentration-dependent structural transition of SPApep, I analyzed the conformation of SPApep in the presence and absence of trifluoroethanol

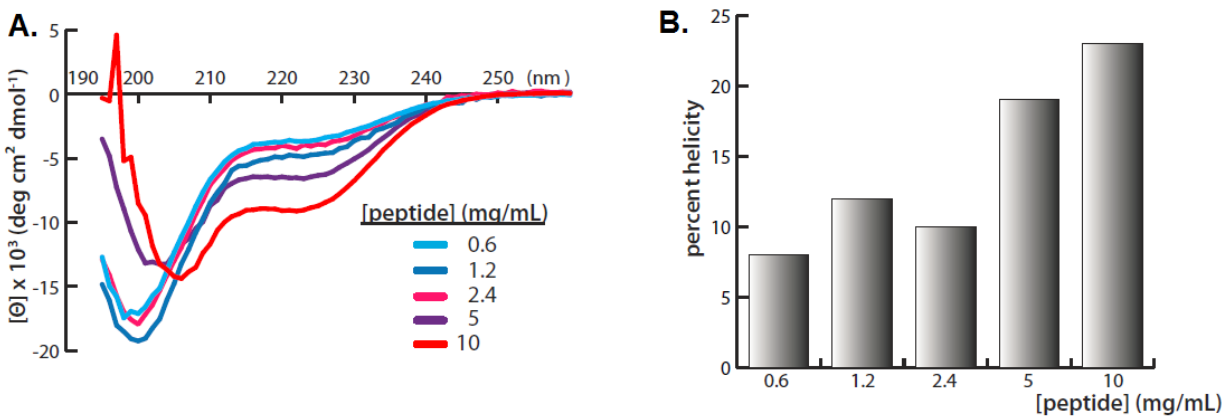


Figure 3.6: CD spectra of SPApep at various concentrations. A) At lower concentrations (0.6 mg/mL – 2.5 mg/mL), SPApep displays CD spectra similar to an unstructured coil. A structural shift begins at 5 mg/mL, and by 10 mg/mL, SPApep begins to display two characteristic local minima at 208 nm and 222 nm, indicative of helical structure. B) The percent helicity of SPApep was quantified using PEPfit.

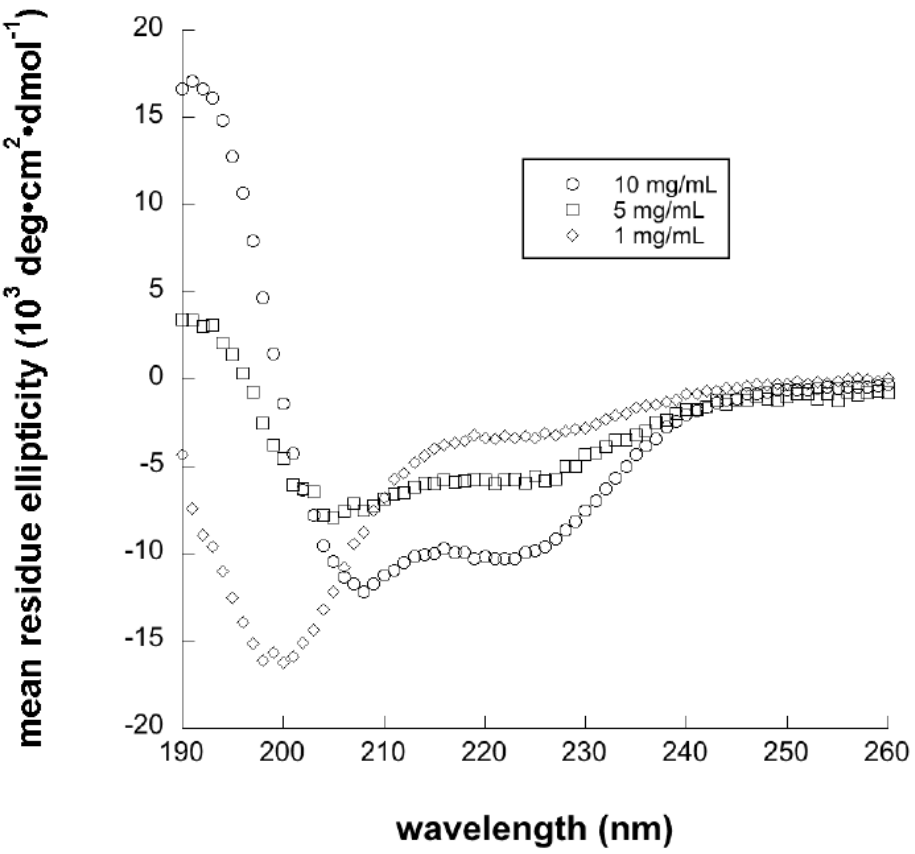


Figure 3.7: CD spectra of SPApep (without cysteine) in aqueous buffer. The CD spectra at 1 mg/mL (diamond), 5 mg/mL (square), and 10 mg/mL (circle) are very similar to the spectra measured using SPApep containing an exogenous cysteine. These experiments were performed by our collaborators in the Prevelige Lab.

(TFE) with various peptide concentrations. It had been shown that another peptide that included HIV-1 SP1 takes on an amphipathic α-helical structure in 30% TFE, as determined by NMR [125]. Although TFE is often used as a helix-stabilizing solvent in NMR and CD studies [157], it can induce non-native α-helical structures . This becomes particularly relevant with peptides, where nonlocal interactions that stabilize the native protein conformation are nonexistent [158]. I diluted SPApep into solvent containing 0% – 30% TFE to a final concentration of either 0.033 mM or 0.20 mM and measured the CD spectra. At 0.033 mM, SPApep showed significantly increased helicity in 20% TFE, which is lower than the concentration used to induce helix formation in the aforementioned HIV-1 SP1 study. At a peptide concentration of 0.20 mM, a clear coil to helix transition took place in 15% TFE (Figure 3.8). In summary, these results reveal that with SPApep, both the increase in peptide concentration and addition of TFE can independently induce helix formation, and that these two variables have additive effects.

The SP assembly peptide associates into a hexamer in a concentration-dependent manner

My collaborators in the Prevelige lab also employed sedimentation velocity AUC to directly probe SPApep oligomerization in solution, using the cysteine-free version of SPApep. An increase in s-value with increasing concentration is expected for a species in rapid association/dissociation equilibrium, as higher peptide concentration favors the associated state. At all three concentrations measured, 1 mg/mL, 5 mg/mL, and 10 mg/mL (0.33 mM, 1.7 mM, and 3.3 mM, respectively), the c(s) distribution displays a peak at approximately 0.2 s (Figure 3.9, A). At 0.33 mM the peak is skewed, trailing

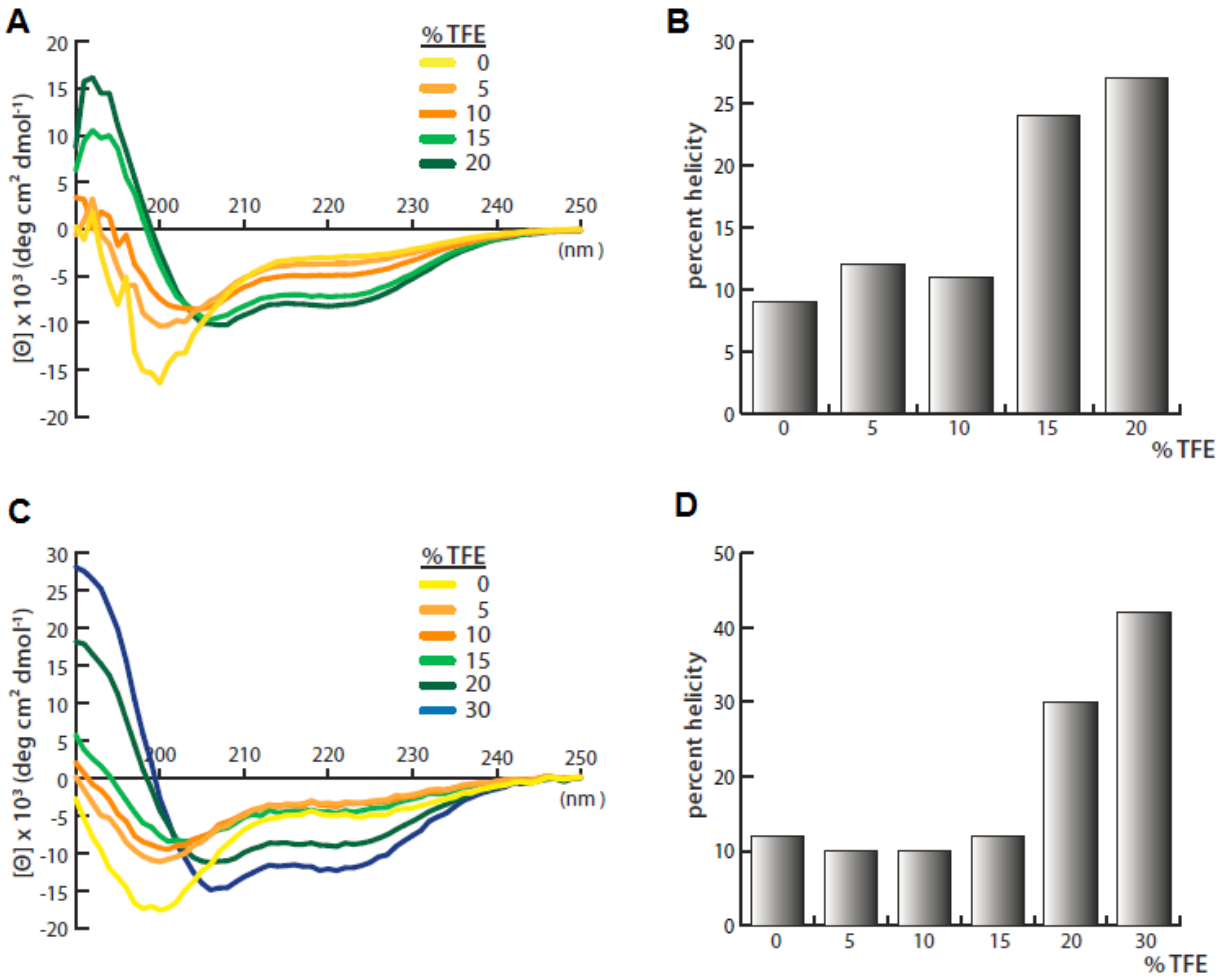


Figure 3.8: CD spectra of SPApep in TFE. A) The CD spectra of SPApep at 0.6 mg/mL concentration in varying amounts of TFE. At 0.6 mg/mL (0.2 mM), SPApep begins to show helical characteristics when 15% TFE is added to the solution. C) The CD spectra of SPApep at 0.1 mg/mL concentration in varying amounts of TFE. At 0.1 mg/mL (0.033 mM), SPApep begins to show helical characteristics when 20% TFE is added. B) & D) are the calculated percent helicity of SPApep in increasing TFE concentrations.

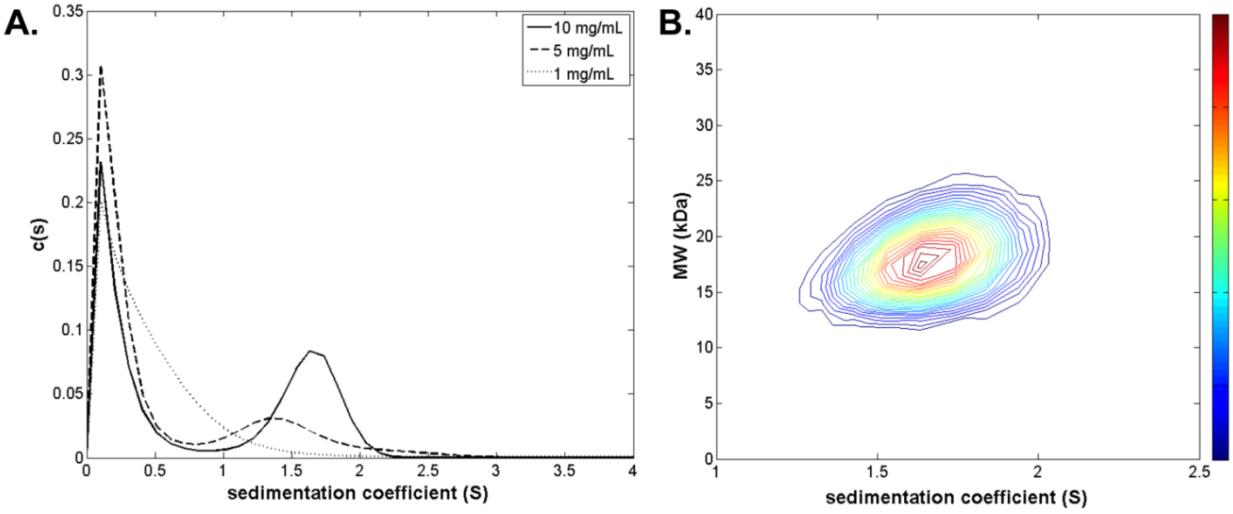


Figure 3.9: Sedimentation velocity analysis of SPApep at 10 mg/mL. A) Sedimentation velocity AUC shows that SPApep associates as its concentration is increased. B) c(s,ffo) model fit of the 3.3 mM (10 mg/mL) species indicating that its molecular weight is nearly identical to that of a hexameric species of SPApep. These data were generated by our collaborators in the Prevelige Lab.

towards higher s-values and as the loading concentration is increased a second peak at approximately 0.2 s (Figure 3.9, A). At 0.33 mM the peak is skewed, trailing towards higher s-values and as the loading concentration is increased a second peak becomes evident. The failure to achieve baseline separation, as well as the concentration dependent shift of the second peak, is a consequence of the fact that the species are in rapid equilibrium.

The sedimentation rate is influenced by both the molecular weight and the shape of the molecule. Information about the shape (frictional coefficient) of the molecule is required to accurately calculate the molecular weight. Determination of the shape factor can be accomplished by modeling the diffusion-induced broadening of the experimentally measured boundary profiles. To determine the molecular weight of the fastest sedimenting species, the data from the 3.3 mM sample were fit with a two-dimensional c(s,ffo) model (Figure 3.9, B). Consistent with the one-dimensional fit, the predominant species has a weight-average sedimentation coefficient of 1.65 S. The weight-average molecular weight of this species was determined to be 17854 Da, very close to the expected mass of a 6HB derived from SPApep (17808 Da) and providing strong evidence that SPApep forms a 6HB in solution.

A computer-generated model of the putative SPA 6HB

Currently, three parallel, homotypic 6HBs have been described at atomic resolution. The first 6HB was discovered and solved as part of a 3.2 Å x-ray structure of the CMV capsid (PDB ID 1F15) [154]. The second 6HB forms the central pore of a hexameric tyrosine-coordinated heme protein (HTHP) from *Silicibacter pomeroyi* and is

part of a 2.5 Å x-ray structure (PDB ID 2OYY) [159]. The most recent 6HB comes from the assembly of a synthetic peptide named CC_{hex}, which was designed with the goal of engineering protein pores (PDB ID 3R3K) [160]. Each of the constituent helices in the CMV 6HB is made of 17 residues. In the HTHP 6HB, each helix is made of 21 residues. In the CC_{hex} 6HB, each helix is made of 32 residues. I used Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) to align the amino acid sequence of each of these helices with RSV SPA, and found CC_{hex} to be most similar in sequence. Using the Clustal Omega sequence homology as a guide, I modeled the cysteine-free SPApep onto the structure of CC_{hex} with Swiss PDB-Viewer (http://www.expasy.org/spdbv/) to generate an idea of what the SPApep 6HB would look like.

The Swiss PDB Viewer model of RSV SPApep is strikingly similar to our predicted model of the RSV 6HB (Figure 3.4 & 3.10). Residues that do not tolerate mutations are found either on the inside of the helical bundle or mediate contacts between helices. Mutable residues do not appear to contact neighboring helices. Molecular surface calculations reveal a seam of interacting surfaces that run down the length of each pair of helices (Figure 3.11). Each individual helix interacts with two adjacent helices using two opposing sides of the helix, creating the same heterotypic I/II interaction illustrated in Figure 3.5 around the entire 6HB (Figure 3.11). The putative interacting residues are listed in Table 3.4 and highlighted in red in Figure 3.11 and correspond closely to residues that do not tolerate point mutations. The exceptions are D_{472}, Q_{473}, and S_{480}, which can all be mutated to alanine, and S_{480} which can even be mutated to cysteine. As previously observed, mutations at the beginning of the

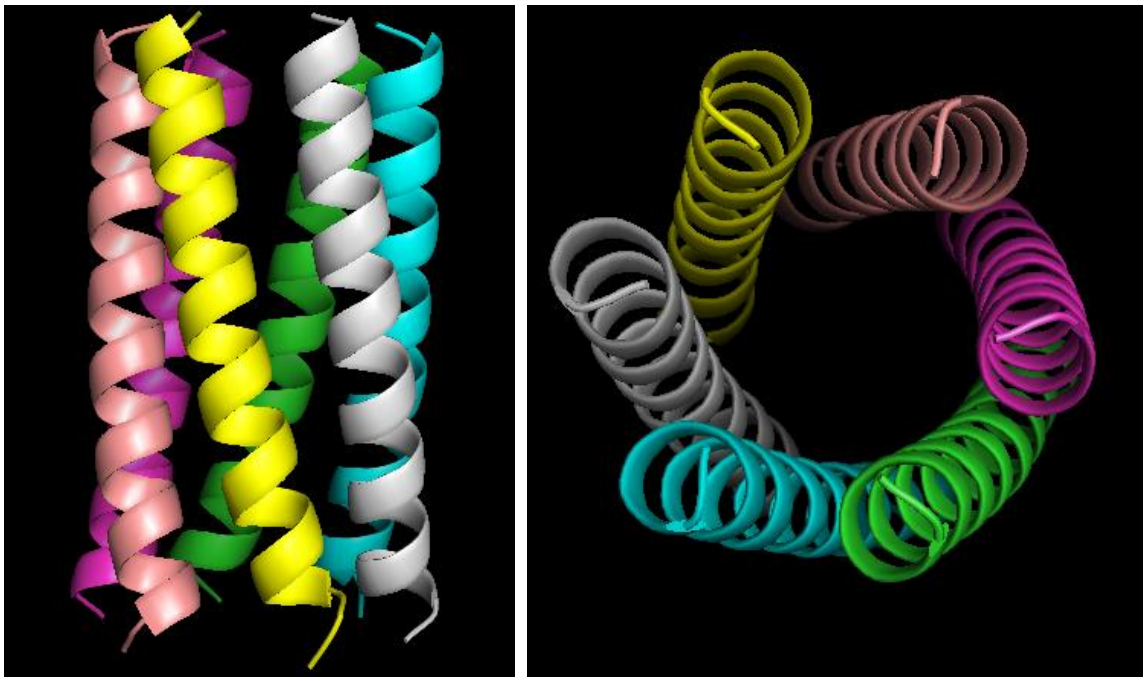


Figure 3.10: Structural alignment model of RSV SPA 6HB. The lateral view of the putative 6HB is on the left, while the top-down view is on the right.

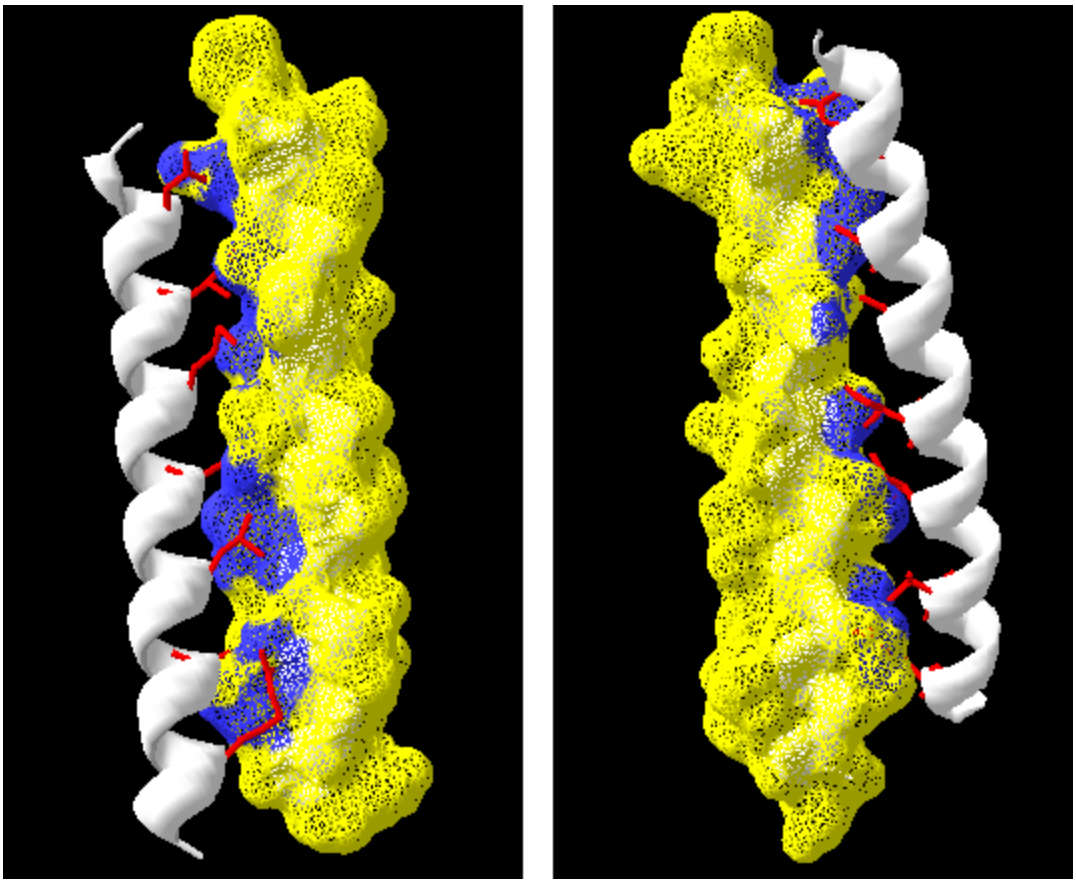


Figure 3.11: Hydrophobic seam between neighboring interacting helices in 6HB. The molecular surface of one helix is outlined in yellow mesh. The adjacent helix is represented in cartoon form. The interaction surface between the two helices is colored blue and the putative interacting residues are colored red and listed in Table 3.4. Helix A is represented in cartoon form on the left, and helix B is represented in cartoon form on the right.

Helix A	Helix B
D_{472}	Q_{473}
I_{475}	A_{476}
M_{479}	S_{480}
A_{482}	I_{483}
L_{486}	I_{487}
A_{489}	V_{490}
R_{493}	E_{494}

Table 3.4: Putative interacting residues in neighboring helices.

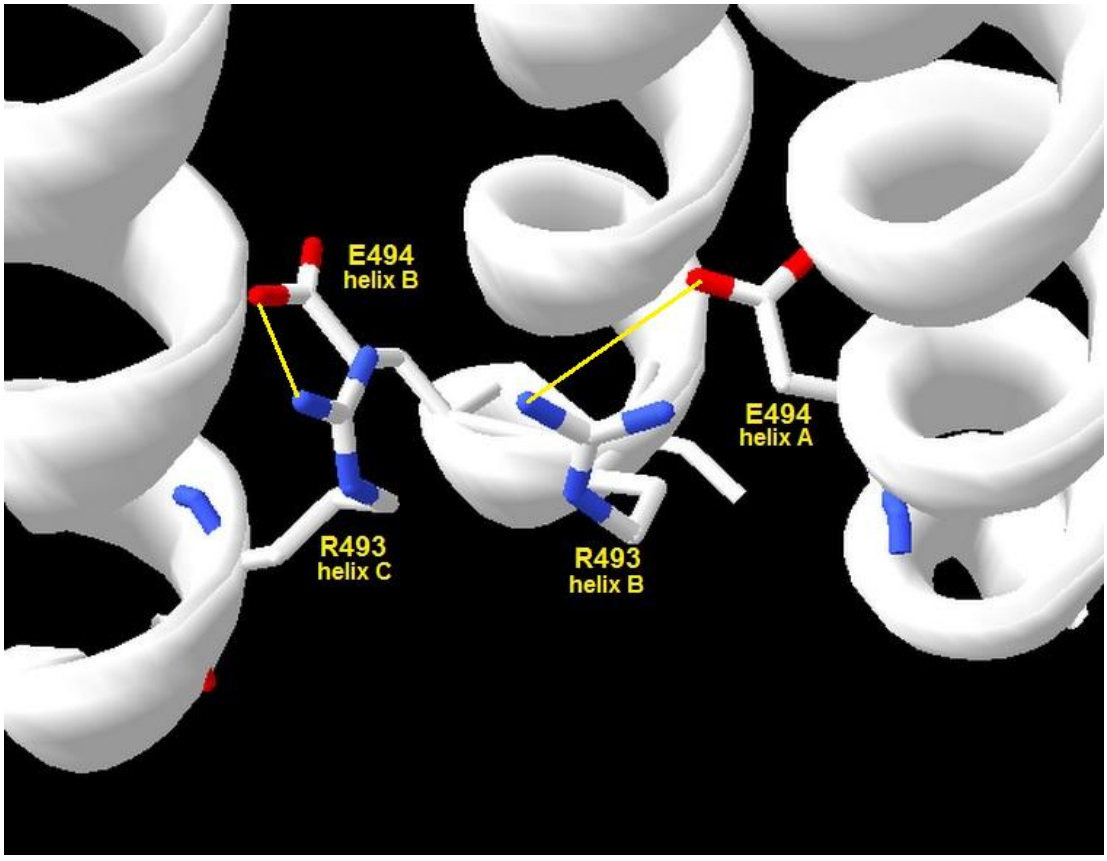


Figure 3.12: A salt bridge between residues R493 and E494 of neighboring helices may stabilize the putative 6HB. The yellow line illustrates the putative salt bridge that may be formed.

predicted SPA helix may be more easily tolerated. S_{480}, on the other hand, appears to fit into a wide groove in the adjacent helix formed at least in part by M_{479}. Therefore, it appears that substitution with alanine or cysteine is tolerated, as they are both small and relatively non-polar. The Swiss PDB Viewer model also reveals a salt bridge at the base of the model 6HB between the side chains of R_{493} and E_{494} (Figure 3.12), two residues that are positioned approximately 4 Å from each other. Furthermore, charged residues immediately downstream of the predicted spacer helix have been implicated in assembly in a number of retroviruses [76, 147, 161]. Despite applying a de novo approach to model the putative RSV 6HB, the end result corresponds closely to the model generated from my point mutations and supports the hypothesis that a 6HB plays a critical structural role in immature assembly.

DISCUSSION

The data presented in this study strongly support the hypothesis that a predicted amphipathic helix overlaps the RSV SPA domain. I identified residues within the SPA helix that are critical for immature assembly and that may be involved in a putative 6HB that promotes immature virus formation. The existence of this helical bundle is supported by results of biophysical analyses demonstrating that a stable, self-associating, hexameric element is able to take on helical conformation in a concentration-dependent manner in the SPA region of assembled VLPs. My results, in conjunction with data from our collaborators in the Prevelige lab, strongly support a model of the SPA domain forming a 6HB in immature assembly, but do not prove that the model is correct.

One proposed structure for the RSV SPA helix is a 6HB. Although parallel, homotypic 6HBs are relatively rare, one naturally-occurring 6HB exists in the CMV capsid [154] and another at the core of HTHP [159]. The only example of a free-standing 6HB is from a study of synthetic helical bundles illustrating the fundamental concepts behind helical pore formation [160]. Similar to the three existing structures, I have modeled the RSV putative SPA helix bundle with hydrophobic interactions on the inside and along helix-helix contacts and with hydrophilic residues on the outside (Figure 2B, Swiss PDB Viewer model). Hydrophobic interactions are known to contribute critical contacts in well-defined α-helical tertiary structures such as leucine zippers and coiled-coils. Our mutation and modeling data suggest that if the SPA domain is a 6HB in immature virions, the interactions that hold the bundle together would form a hydrophobic seam along the length of two neighboring helices, similar to previously described 6HBs.

Interestingly, charged residues downstream of the predicted SPA helix also play a role in assembly. Keller et al. mapped the SPA domain to include the first four residues of NC and excluded the four charged residues that followed [126]. However, there are indications that at least a couple of these four residues play a structural role in assembly. The two arginine residues were previously shown to be critical for assembly of a truncated, purified version of ΔMBDΔPR, since they did not tolerate mutation to alanine [76]. A similar requirement for at least two basic residues downstream of an SP-like domain was identified in MPMV [147], while in HIV-1, two arginine residues near the N-terminus of the NC domain) are required for proper VLP formation [161]. I included these four residues in SPApep to enhance peptide solubility, as I hypothesize

that they play a role in stabilizing the higher-order structure of the peptide hexamer. In fact, my Swiss PDB Viewer model of the putative RSV 6HB places R_{493} and E_{494} at the seam between two SPApep helices, suggesting that these two residues form an inter-helix salt bridge.

What might a 6HB be doing under the organized CA lattice? One hypothesis is that the 6HB is a transient structure that helps stabilize the upstream CA domain into its immature conformation. However, no helix has ever been observed in the region downstream of CA_{CTD}, and my data as well as data presented by Datta et al. suggest that helical structure in that region only forms in a concentration-dependent or chemically-induced manner [102]. The metastable nature of this region may be the key to the function of SPA in RSV assembly and maturation. I speculate that during the early steps of assembly, Gag converges to the PM, increasing the local concentration and allowing SPA to shift from unstructured to helical. The CA domain is highly flexible due to the floppy CA_{NTD}-CA_{CTD} linker and may sample many possible conformations during initial dimerization and multimerization. Formation of the 6HB may force the CA_{CTD} to expose surfaces favorable for immature assembly, and these downstream spatial constraints could propagate upstream into CA_{NTD}. The CA_{NTD} hexamer is then locked in place by its interactions with the upstream p10 residues, securely fastening both ends of the Gag hexamer. The 6HB itself likely also has considerable flexibility, which may explain why even at peptide levels approaching virion Gag levels, the time-averaged percent helicity of the SPApep does not exceed 23%. This may reflect significant “breathing” of the helix to allow protease to release CA during maturation. The relatively high level of order at the CA-SP junction may also explain why in both

RSV and HIV-1, cleavage of SP and SP1 from CA are the last steps in maturation [162, 163].

Can one 6HB functionally replace another? Although the answer is ambiguous, it is heavily weighted towards no. The first experiments that attempted to address this question used an assembly-competent chimeric BIV Gag in which domains downstream of the p3 spacer were replaced with their cognate HIV-1 Gag domains. When p3 was substituted with SP1, virus production was abolished. However, p3 can be replaced by short, five residue-long sequences derived from the spacer domains from feline immunodeficiency virus, EIAV, and an N-terminal NC fragment from Visna virus without severely compromising virus production and reverse transcriptase activity [164]. A similar set of experiments using RSV Gag showed that sequences downstream of the structured CA_{CTD} cannot be replaced by homologous sequences from HIV-1 without compromising VLP assembly [126], but when I exchanged five amino acids at the beginning of the SPA helix with seven residues downstream of the structured HIV-1 CA having the highest predicted helical propensity (SP1sub mutant), the protein assembled in vitro. In MPMV, when the SP-like region at the N-terminus of NC is replaced by HIV-1 SP1, no virus assembly is observed [147].

The amino acid specificity of the spacer region across retroviral genera suggests that there may be contacts beside the putative homotypic inter-helix interactions that exist between SP and other Gag domains. In fact, using the maturation inhibitor PF-46396, Waki et al. mapped a novel binding pocket formed by the major homology region (MHR) of HIV-1 CA and SP1 and predicted that these two distant regions may be in

close proximity in an assembled virus [30]. This inference is supported by data from Bowzard et al. that showed the rescue of a lethal MHR mutation in RSV CA by a second-site suppressor mutation in SP [70]. Given the apparent conservation of a SP-like domain and the absolute conservation of the MHR across all retroviral genera, perhaps this putative interaction surface between SP and the MHR should be further investigated for its role in stabilizing the immature retrovirus.

CHAPTER FOUR

CRYSTALLOGRAPHY OF A MINIMAL GAG PROTEIN AND THE SPA PEPTIDE

INTRODUCTION

Short sequences or structural motifs that flank the CA domain of Gag have come to be known as molecular switches. They are thus named because in their absence, a morphological transformation from immature assembly to mature-like assembly is triggered. Upstream and downstream switches have been characterized for RSV, HIV-1, and to a lesser extent, MPMV and MLV. In HIV-1 and MPMV, the upstream switch appears to be an N-terminal extension of the CA domain, or a mutation of the N-terminal proline of the mature CA, either of which prevents formation of the conserved proline-aspartic acid salt bridge in CA [131, 134]. In MLV, deletion of p12, the late domain-containing minor cleavage product between the MA and CA domains, results in mature-like tubular assembly [95]. Substitution of p12 with RSV p2 or HIV-1 p6, both of which are also late domain-containing minor domains, partially rescues the immature assembly phenotype [96]. This result suggests that MLV may require specific late domain-associated sequences to form immature VLPs, and that a hindrance to the formation of the N-terminal salt bridge is insufficient. In RSV, the last 25 residues of the p10 domain, which is located immediately upstream of the CA domain, are essential for immature assembly. Molecular modeling and thiol cross-linking experiments performed in our lab showed that several of these residues form an inter-molecular interface unique to the immature hexamer that serves to lock the CA_{NTD} in its immature conformation [72].

The downstream switches in RSV and HIV-1 are the cleaved SP and SP1 domains, respectively, which have been proposed to form a six-helix bundle (6HB) that ties together the immature hexamer [9]. Secondary structure predictions assign a helical structure to both the SP and SP1 domains, and cryo-EM has shown a structured density downstream of the ordered CA_{CTD} in immature VLPs [9, 107]. This feature may be conserved in MLV as well, as MLV CA has a highly charged region at the C-terminus of the CA_{CTD} that is predicted to form an α-helix. This predicted helix may contribute to a higher-order structure such as a 6HB, as deletion mutations in this region do not affect immature assembly as long as they remove complete turns of the helix [127]. In MPMV, a 15 residue stretch at the N-terminus of NC has also been found to be crucial for immature assembly [147], and my data from Chapter 3 suggest that these residues form part of a predicted 12^{th} helix spanning the junction between the CA and NC domains.

The immature Gag hexamer has been examined extensively using cryo-EM tomography and sub-tomogram averaging. Both HIV-1 and RSV have been studied using these techniques and the most complete report on the immature core lattice is an 8 Å structure of MPMV VLPs obtained using a combination of cryo-EM tomography and fitting in high-resolution crystal structures of CA_{NTD} and CA_{CTD} [8, 9, 107, 117]. Although these studies have provided much insight on the differences in organization between the immature and mature lattices, there is no high-resolution structure of the immature lattice complete with both the upstream and downstream molecular switches.

Currently, there is no evidence to suggest that a common structural element in all retroviral Gag proteins underlies the upstream switch in immature retroviral assembly,

but the downstream switch appears to be conserved, at least in RSV and HIV-1. The structural contribution that the spacer region makes to immature assembly is unclear for two reasons. First, there is no structure of the spacer region in the available crystal structures of CA and downstream domains [54, 56, 57, 65]. Second, in the absence of a crystal structure, it is impossible to assign visible densities in the immature lattice, such as those observed in HIV-1 and RSV [107], to any specific part of Gag. Although the SPA 6HB hypothesis is strongly supported by the data presented in Chapter 3 of this dissertation, a crystal structure of the putative 6HB would provide conclusive evidence.

In this chapter, I will describe my efforts to obtain atomic level structural information on the immature RSV Gag lattice and the putative SPA domain 6HB. To do this, I have designed, expressed, and crystallized two minimal versions of Gag known as 25CASP and 25CASP(CC), as well as N- and C-terminal deletion mutants of 25CASP. I have tried to template crystal formation by linking 25CASP to the hexamerizing protein Ccmk2, and by linking the hexamer template HTHP to the SPA sequence. Finally, I performed a crystallization trial on the SPA peptide described in the previous chapter and found several crystallization conditions. So far, SPApep has shown the most promise in delivering a structure that may help us to understand the role that the spacer region plays in immature retroviral assembly.

RESULTS

A minimal Gag protein (25CASP) can be purified using nickel-affinity and size-exclusion chromatography

Multi-domain proteins are often difficult to crystallize, as their inherent flexibility often prevents them from packing regularly. To reduce flexibility in Gag, I made a "minimal" RSV Gag construct that contains solely the domains that we know to be essential for immature assembly: the last 25 amino acids of p10, all of CA, and the SPA domain (Figure 4.1). This "minimal Gag" construct was named 25CASP. A version of this protein was also made using a previously characterized version of ΔMBDΔPR(-11C) with two exogenous cysteines in p10 and CA_{NTD} that cross-link the immature CA_{NTD} hexamer [72]. This protein was named 25CASP(CC). The rationale is that these two cysteines may help stabilize any hexamers that form during crystallization. Both 25CASP and 25CASP(CC) have theoretical molecular weights of approximately 30 kD, but their migration through SDS-polyacrylamide gels closely mimic that of the 25 kD molecular weight marker.

As 25CASP was a novel protein that had never been expressed in *E. coli*, I attached a small ubiquitin-like modifier (SUMO) tag to its N-terminus to facilitate the purification. The SUMO vector was developed to enhance protein expression and solubility in *E. coli*, and a highly specific protease, Ulp1, cuts the tag efficiently to liberate the desired protein product. Additionally, in this system both SUMO and Ulp1 are tagged with a poly-histidine tag. In essence, a SUMO-tagged protein is purified using two-step nickel-affinity purification. In the first step, the SUMO-tagged protein is bound to the nickel-

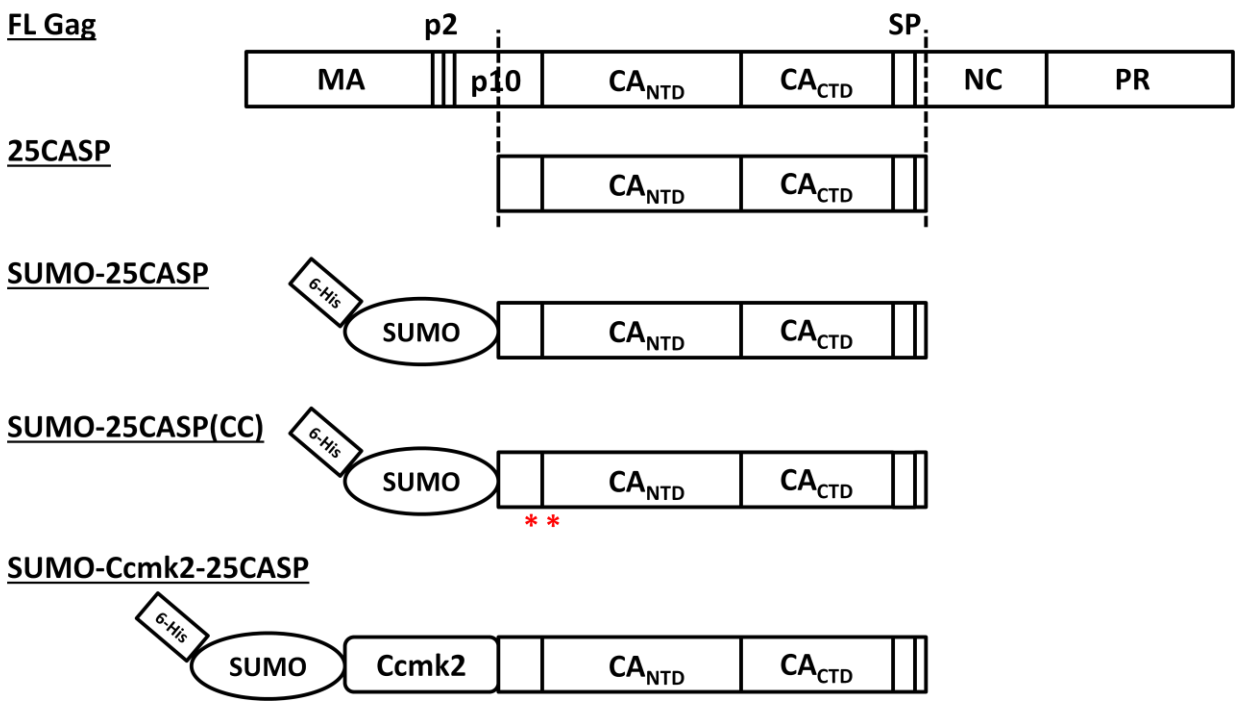


Figure 4.1: Schematic diagrams of Gag-derived constructs used for X-ray crystallography. Full-length RSV Gag is designated FL Gag. 25CASP contains the parts of Gag deemed to be essential for immature assembly. SUMO-25CASP is 25CASP fused to a SUMO tag, while SUMO-25CASP(CC) contains two exogenous cysteines, locations indicated by the two red asterisks. SUMO-Ccmk2-25CASP is the hexamer-templated construct.

affinity column and separated from other *E. coli* proteins; in the second step, SUMO is proteolytically cleaved from the protein of interest, and SUMO and Ulp1 are bound to the nickel-affinity column, while the protein of interest flows through (Invitrogen SUMO protease manual, Cat. No. 12588-018). SUMO-25CASP is predicted to be a 44 kD protein, but migrates through poly-acrylamide gels only slightly above the 37 kD marker.

Figure 4.2 shows a typical purification process for 25CASP. SUMO-25CASP was highly expressed and efficiently cleaved by Ulp1 in an overnight dialysis into imidazole-free buffer. The cleaved 25CASP remained highly soluble in a high salt buffer (20 mM sodium phosphate dibasic, pH 8.0, 500 mM NaCl, 10 mM imidazole) and could be purified to approximately 95% purity after a second pass over a nickel-affinity column. 25CASP remained soluble when dialyzed into a low salt buffer (20 mM Tris, pH 8.0, 100 mM NaCl) for gel filtration and eluted from a size-exclusion column as a single peak with a small shoulder. Fractions corresponding to the peak were collected and assayed for purity by SDS-PAGE. The fractions with the highest ratio of 25CASP to contaminants were pooled and concentrated to >5 mg/mL in the low salt buffer for crystallography trials (Figure 4.3).

25CASP forms micro-crystals upon initial screening and can be improved to needle clusters in a hanging drop

My initial crystallization screen used a total of 480 conditions in sitting drops, dispensed using an ARI Phoenix liquid handling system. Of these conditions, only one was found to foster crystal growth for 25CASP (0.1 M sodium citrate, pH 5.5, 0.2 M lithium sulfate, 15% ethanol), and a different condition led to crystal growth for

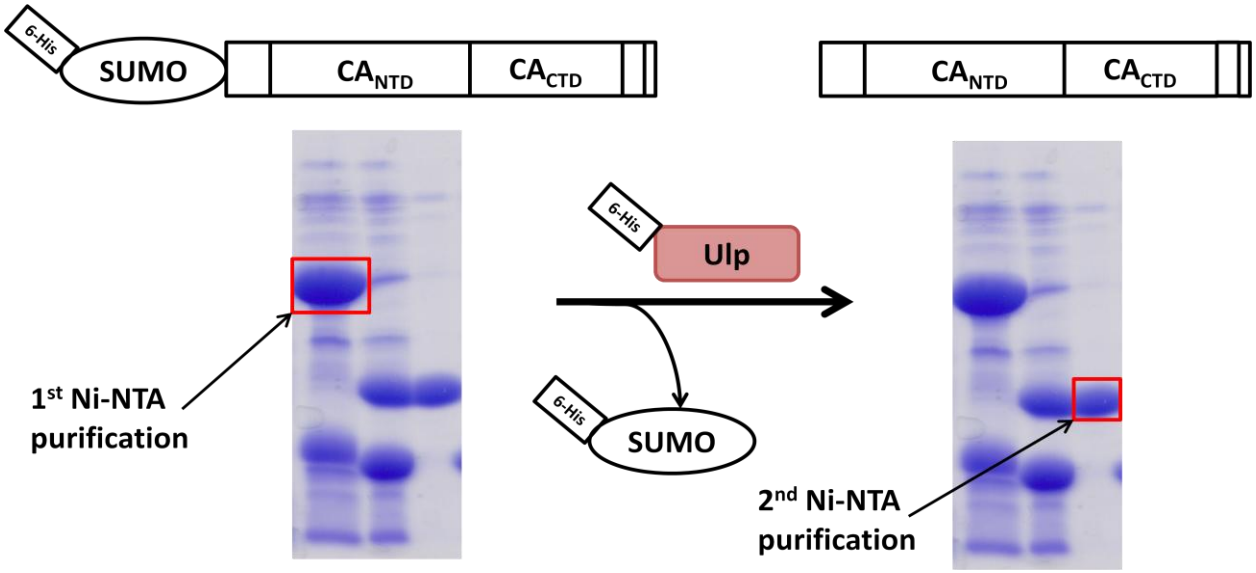
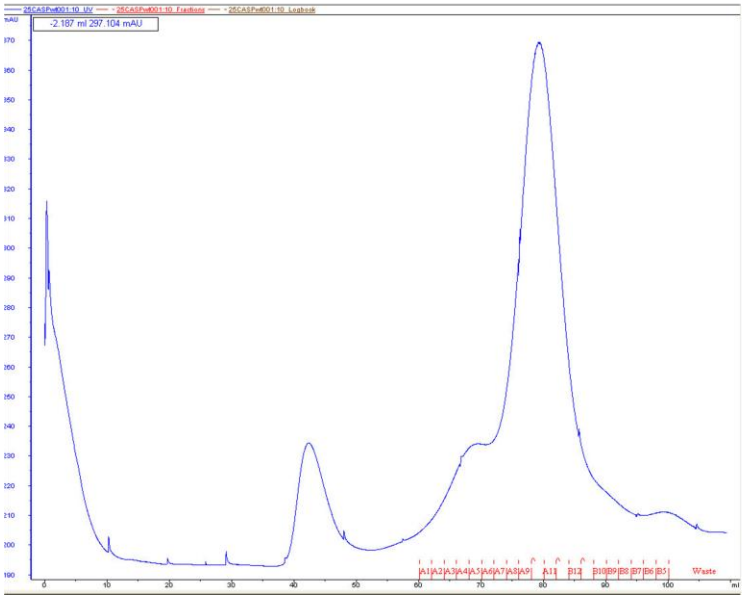


Figure 4.2: Purification of 25CASP using a SUMO-tagged protein. On the left, the full-length SUMO fusion protein is eluted off the nickel-affinity column (left, boxed in red). After dialysis with Ulp1 protease, the SUMO protein is cleaved off and both SUMO and Ulp1 are bound to a second nickel-affinity column. The flow-through of the second nickel-affinity column is the protein of interest, in this case, 25CASP (right, boxed in red).

Gel filtration readout



SDS-PAGE of collected fractions

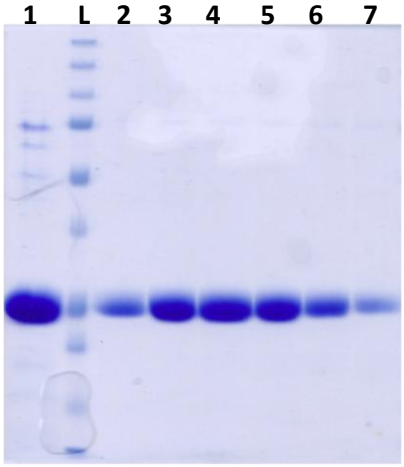


Figure 4.3: Gel filtration and SDS-PAGE of 25CASP. 25CASP mostly elutes off the size-exclusion column as a single peak after approximately 80 mL (left). The injected protein has several higher molecular weight contaminants or oligomers (lane 1), but these are removed from the eluted fractions (lanes 2 – 7). Lane L is the molecular weight ladder.

25CASP(CC) (0.1 M sodium/potassium phosphate, pH 6.2, 10% PEG-3350). For both proteins, the crystals formed were tiny two-dimensional micro-crystals approximately one micron in length. Both conditions were optimized by varying the pH, as well as the protein, salt, and precipitant conditions found in the initial hit, as well as scaling up the drop size in a hanging drop. Simply by altering the conditions of the initial hit, 25CASP was optimized to needle clusters approximately 1 mm in diameter, but further optimization using different precipitants did not enhance crystal growth (Figure 4.4). Unfortunately, these needle clusters were too small for diffraction. The initial hit for 25CASP(CC) contained a phosphate buffer known to produce salt crystals. I changed the buffer to 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer for the optimization and observed no crystal formation, suggesting that the microcrystals seen in the initial hit may have been salt crystals.

N-terminal and C-terminal deletions of 25CASP greatly affect protein solubility

When proteins fail to crystallize during optimization, the logical next step is to optimize the protein construct itself. A common method to optimize constructs is by truncating the protein by a few amino acids at the N- or C-terminus. However, as 25CASP only includes the domains of Gag that are necessary for assembly, the number of possible N-terminal and C-terminal mutations is limited.

To generate proteins that may be more amenable to crystallization, I generated the following three mutants: ΔN3, ΔC2, ΔN3C2. ΔN3 removes the first three amino acids in 25CASP (P_{215}, G_{216}, P_{217}), ΔC2 removes the last two (V_{491}, N_{492}), and ΔN3C2 removes all five aforementioned residues. I proceeded with the two-step nickel-affinity

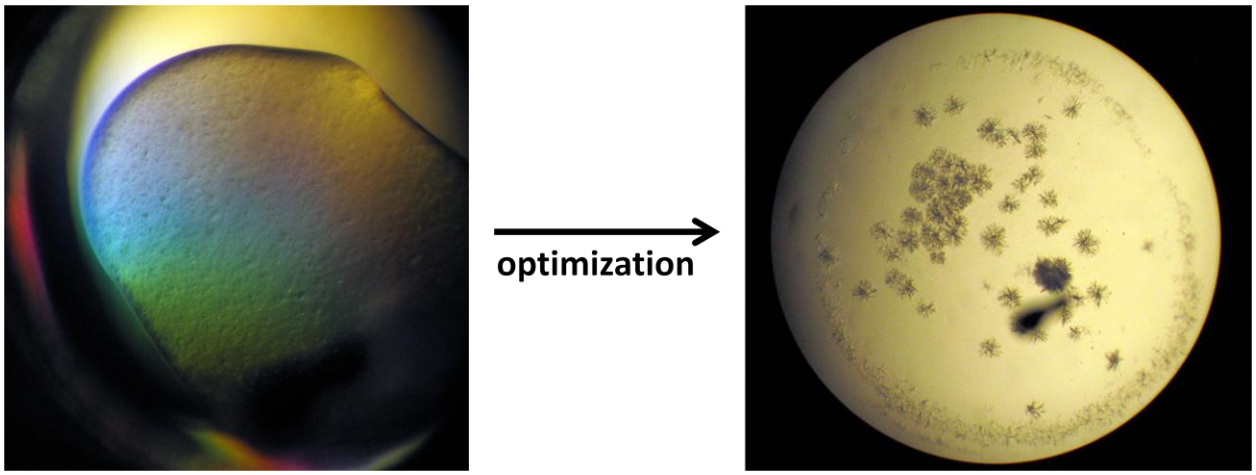


Figure 4.4: Optimization of 25CASP. A single hit in the initial screen produced tiny crystals (left). After optimizing the buffer pH (0.1 M sodium citrate, pH 5.0 – 6.5), salt concentration (lithium sulfate, 0 M – 0.3 M), and concentration of precipitant (ethanol, 5 – 15%), the crystals grew into larger needle clusters (right).

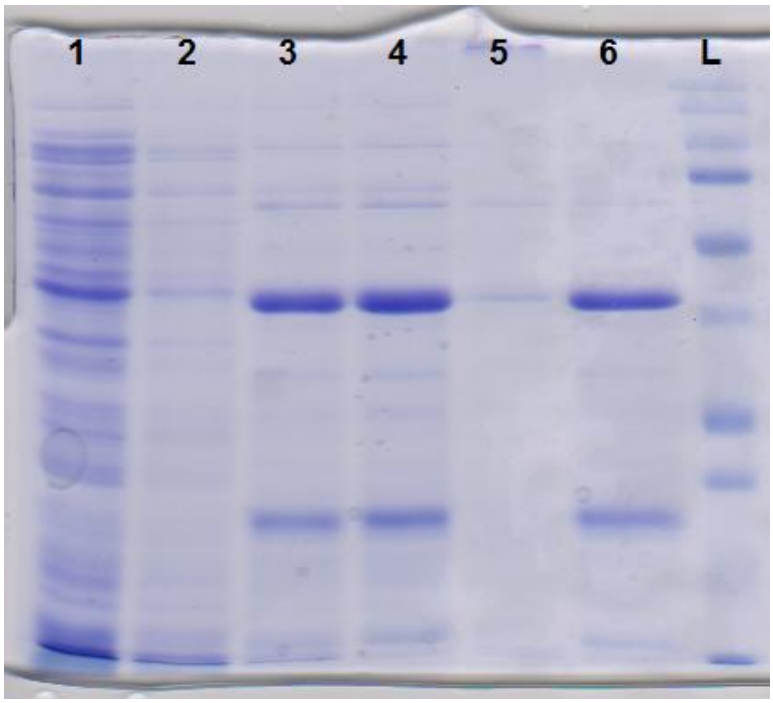


Figure 4.5: SUMO-25CASPΔN3C2 cannot be cleaved by Ulp1. The contents of the lanes are as follows: 1) nickel-affinity column flow-through; 2) column wash; 3) elution; 4) dialysis with Ulp1; 5) second nickel-affinity column flow-through; 6) column elution; L, molecular weight ladder. A proportionate amount of each fraction was run on each lane.

purification described in the first section of this chapter, and all three mutant proteins are well-expressed in *E. coli* as a SUMO fusion and can be easily purified. However, when I dialyzed and cleaved the SUMO fusion proteins with Ulp1, the SUMO tag was not removed (Figure 4.5). This unlikely result was repeated several times with all three constructs, and only the ΔC2 construct was able to be cleaved by Ulp1. Since Ulp1 recognizes a structural element in SUMO rather than a primary amino acid sequence, it is my belief that ΔN3 and ΔN3C2 mis-fold and conceal the Ulp1 cleavage site in SUMO. I continued to work with ΔC2, but encountered additional expression and purification problems, which eventually led me to abandon the construct optimization portion of this project.

A hexamer-templated minimal Gag protein aggregates and precipitates during purification

Previous work by Pornillos et al. has shown that templated crystal formation is a viable way of generating retroviral CA crystals [64]. Knowing that Gag assembles into a hexamer in VLPs, I used a protein with known hexamer structure to induce hexamer formation in 25CASP. Ccmk2 is a carboxysome shell protein from a cyanobacterium [137]. The crystal structure of the protein hexamer is shown in Figure 4.6. The C-termini of the Ccmk2 hexamer are approximately 5 – 7 Å apart. I reasoned that these distances should be close enough to bring six units of Gag together, but they may be too close and impose unnecessary constraints on Gag folding and inter-molecular interactions. Therefore, I placed three alanines between the template and 25CASP to introduce some flexibility. The final construct was named SUMO-Ccmk2-25CASP and has a

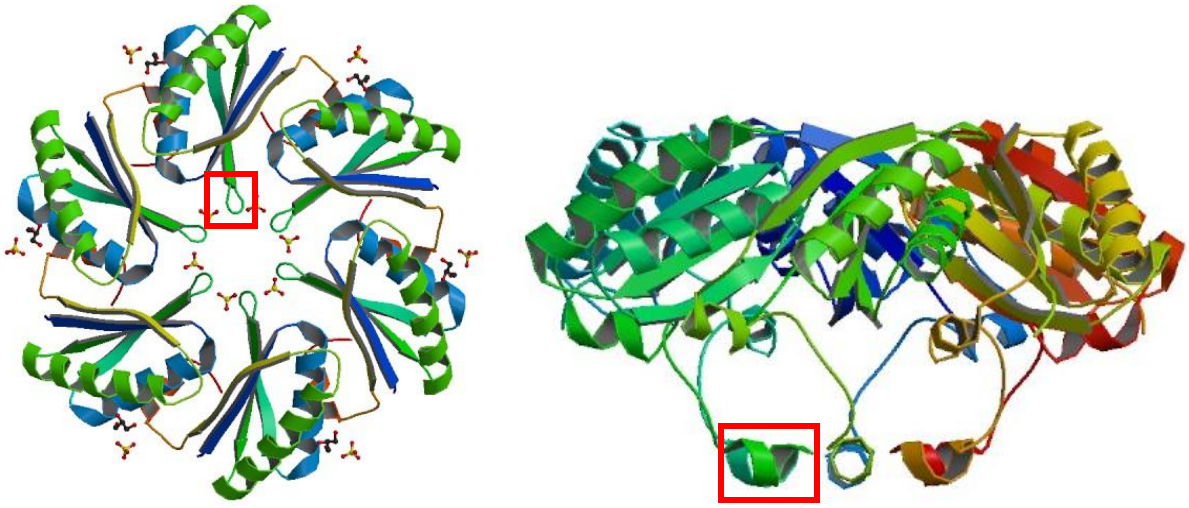


Figure 4.6: Crystal structure of Ccmk2 hexamer. The C-terminal end of each monomer points toward the inside of the pore formed by the hexamer. In each view of the structure, the C-terminal end of one monomer is boxed in red.

predicted molecular weight of 46 kD, but the observed size on poly-acrylamide gels is closer to 50 kD.SUMO-Ccmk2-25CASP proved quite difficult to purify. The protein was not highly expressed in *E. coli* and it often precipitated out of solution during the purification process, most often when the concentration of sodium chloride was reduced or when SUMO was cleaved (Figure 4.7, lane 7 is the precipitated fraction). Given that Ccmk2 forms part of the bacterial carboxysome, a highly organized protein structure in many ways similar to a viral capsid, it is likely that the Ccmk2 domain tries to assemble, leading to protein precipitation. By maintaining a 1 M concentration of sodium chloride in all buffers, I was able to purify Ccmk2-25CASP successfully. Presumably, the high salt concentration helped to disrupt electrostatic interactions between monomers that may mediate carboxysome assembly. Figure 4.7 shows an example of a successful Cckm2-25CASP purification. Ccmk2-25CASP is approximately 42 kD in molecular weight, but migrates through poly-acrylamide gels like the 37 kD marker.

Figure 4.8 shows the elution profile of Ccmk2-25CASP from a size-exclusion column. The high peak in the void volume indicates that there is considerable protein aggregation as the protein proceeds through the column. However, most of the protein seems to have remained in the second peak, as evidenced by the SDS-PAGE gel. Fractions from the second peak were pooled, concentrated, and subjected to an initial crystal screen. I found several conditions that produced plate-shaped two dimensional crystals. The condition that produced the best crystals was 0.1 M sodium/potassium phosphate, pH 6.2, 0.2 M sodium chloride, 20% PEG-1000. These conditions were reproduced and expanded using MES buffer and plate-shaped crystals grew in a hanging drop. Figure 4.9 shows the morphology of these crystals.

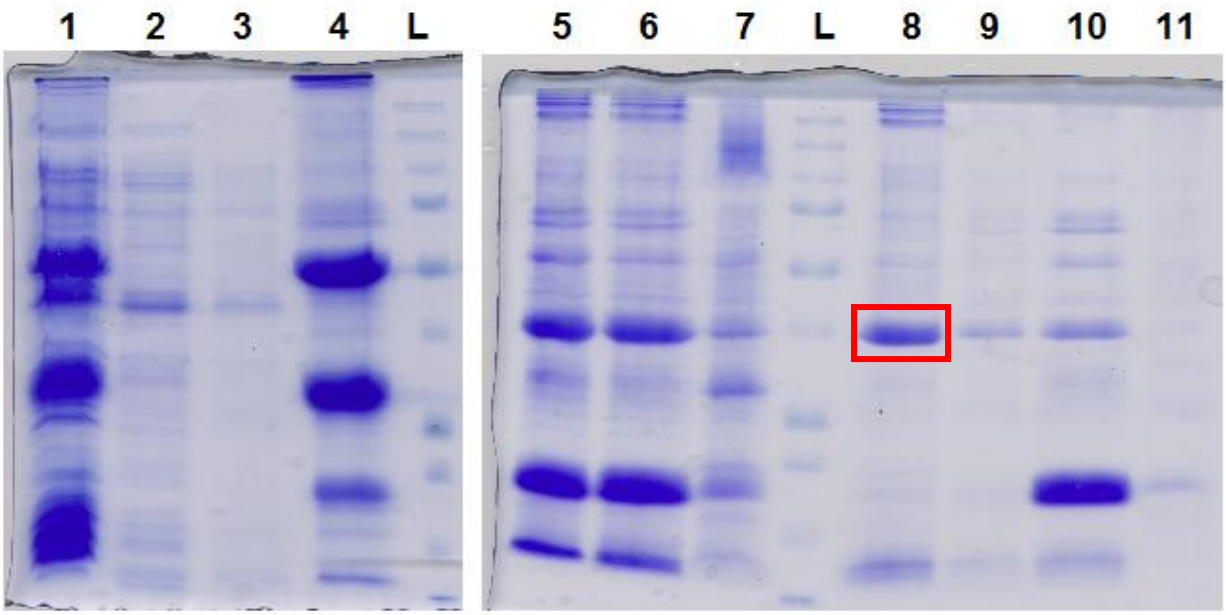
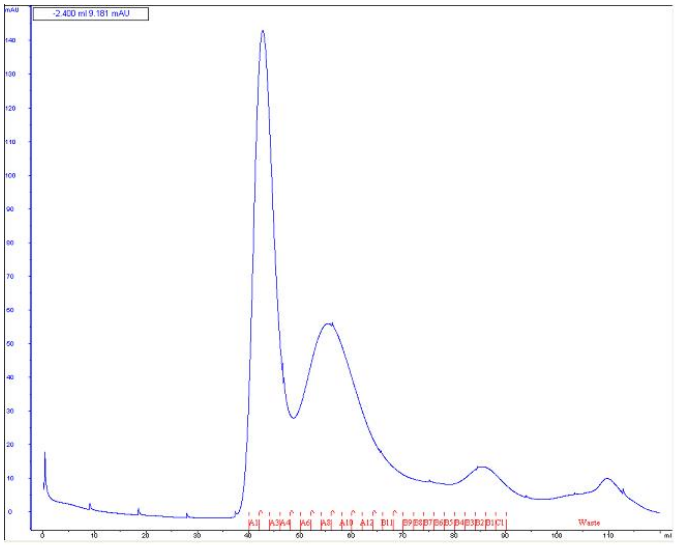


Figure 4.7: Nickel-affinity purification of SUMO-Ccmk2-25CASP. The contents of each lane are as follows: 1) E. coli lysate; 2) nickel-affinity column flow-through; 3) column wash; 4) eluted SUMO-Ccmk2-25CASP; 5) dialysis with Ulp1; 6) supernatant fraction; 7) precipitated fraction; 8) second nickel-affinity column flow-through; 9) column wash; 10) elution; 11) bead fraction; L, molecular weight ladder. A proportionate amount of each fraction was run on each lane. The red box indicates the purified Ccmk2-25CASP protein that was later run on a size-exclusion column.

Gel filtration readout



SDS-PAGE of collected fractions

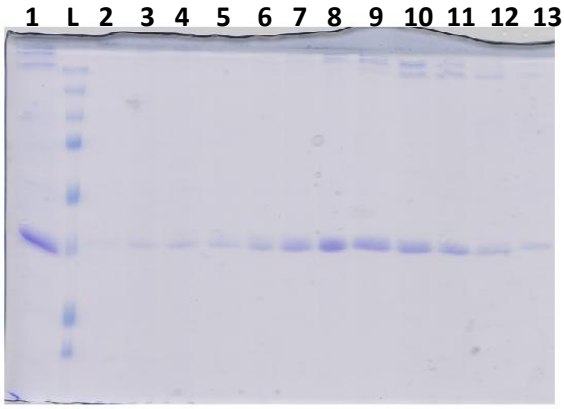


Figure 4.8: Gel filtration and SDS-PAGE of Ccmk2-25CASP. Ccmk2-25CASP mostly elutes as a single broad peak. There is substantial aggregation on the column, as evidenced by the left-most peak (left). When the fractions are collected and assayed by SDS-PAGE, most of the protein is found in lanes 7 – 12 (right), which correspond to the broad, lower peak.

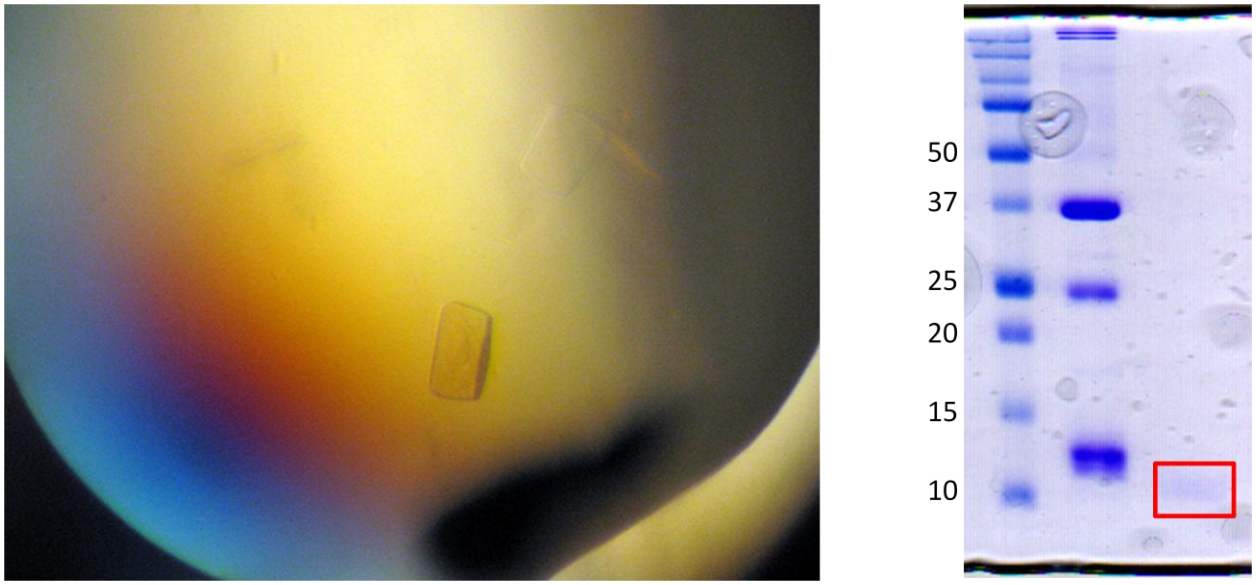


Figure 4.9: Plate crystals are actually cleaved Ccmk2. The observed crystals formed after the purified Ccmk2-25CASP (right, second lane) was placed into a hanging drop. When the crystals were run on SDS-PAGE, their molecular weight corresponded to breakdown products of Ccmk2 (right, red box).

The Ccmk2-25CASP fusion protein crystals appear very similar to the description of Ccmk2 crystals described by Kerfeld et al. [137]. Furthermore, when the size-exclusion column-purified Ccmk2-25CASP was concentrated and assessed for purity with SDS-PAGE, approximately half of the full-length protein seemed to have been degraded into 25 kD and 12 kD products, corresponding roughly to the size of 25CASP and Ccmk2, respectively. To verify that the crystals are actually composed of the entire fusion protein and not just a proteolytically cleaved fragment, I picked several crystals and evaluated them using SDS-PAGE (Figure 4.9). No band was observed at 37 kD, 25 kD or 12 kD, but a very faint band corresponding to a product of approximately 11 kD was visible (Figure 4.9, red box). This unexpected result suggests that the plate-shaped crystals are breakdown products of Ccmk2, rather than 25CASP, and that there may have been protease contamination in the purified sample.

A hexamer-templated SPA protein purifies as a doublet at low concentration

As previous attempts to crystallize the entire immature hexamer have been largely unsuccessful, I turned my focus to the elusive role of SP in the hexamer. As my point mutation data suggested that the SPA domain forms a 6HB in immature assembly, I built a construct that may help template 6HB formation. The template protein is the hexameric tyrosine-coordinated heme protein HTHP, which has a parallel, homotypic 6HB at its core (Figure 4.10). I added a flexible linker sequence (SSSSG) between HTHP and the SPA domain sequence, including the four charged amino acid residues putatively involved in forming an inter-helix salt bridge downstream of SPA. I named the construct HTHP-S4G-SPA, and its predicted molecular weight was approximately 13

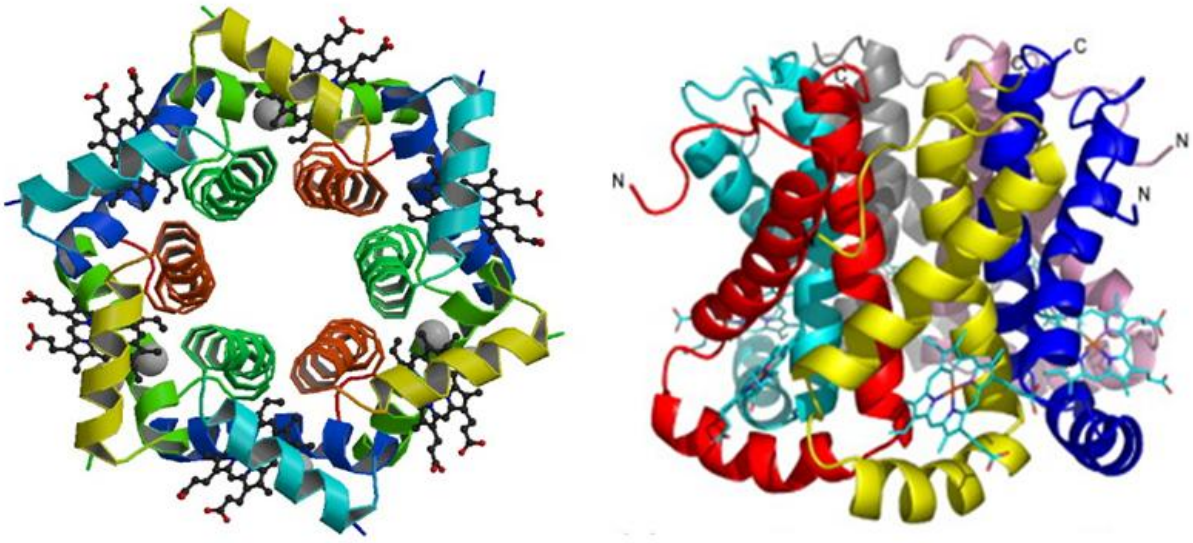


Figure 4.10: Crystal structure of HTHP hexamer. At the center of the hexamer is the parallel 6HB, depicted in orange and green (left). The N- and C-termini of each monomer are indicated by the letters N and C, respectively (right).

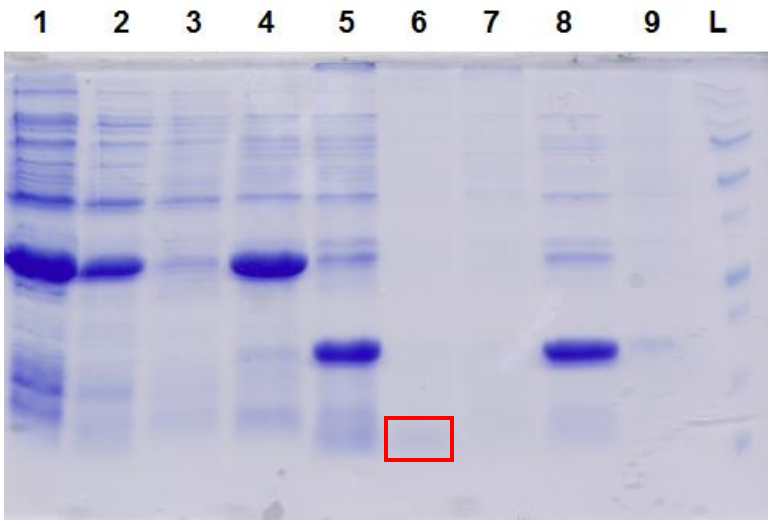


Figure 4.11: HTHP-S4G-SPA is purified at low concentration, despite high expression in *E. coli*. The contents of each lane are as follows: 1) *E. coli* lysate containing SUMO-HTHP-S4G-SPA; 2) nickel-affinity column flow-through; 3) column wash; 4) elution; 5) dialysis with Ulp1; 6) second nickel-affinity column flow-through; 7) column wash; 8) elution; 9) collected beads; 10) Precision Plus ladder. The purified HTHP-S4G-SPA is boxed in red.

kD. SUMO was linked to the N-terminus of this construct to facilitate purification.

SUMO-HTHP-S4G-SPA expressed abundantly in E. coli (Figure 4.11). However, after cleavage of SUMO, the amount of HTHP-S4G-SPA recovered was very low. Nonetheless, I further purified the protein on a gel exclusion column, where it eluted in three distinct peaks. The leftmost peak is almost certainly aggregated protein, while the central peak is likely an oligomerized form. The peak to the right is probably the monomer. The eluted fractions corresponding to the central and right-most peaks are shown in lanes 5 – 10 in Figure 4.12. Although the protein injected onto the size-exclusion column appears to be pure, HTHP-S4G-SPA elutes as a doublet, suggesting that there is substantial breakdown of the protein during purification. Combined with the low levels of protein purified, I chose to not pursue this construct.

SPApep crystallizes as disc and plate crystals

Our collaboration with the Prevelige Lab generated AUC data indicating that SPApep assembles into a higher-order structure at high concentrations, which suggests that SPApep may not be unstructured in solution as was previously believed. Using purified peptide remaining from previous CD experiments, I attempted a crystal screen using SPApep. Unlike 25CASP, SPApep crystallized rapidly (immediately after mixing in some conditions) under a multitude of conditions into mostly disc-shaped crystals. I found 11 conditions with the largest crystals and least amount of precipitation, and compared the crystallization conditions for common elements.

Table 4.1 shows the list of conditions under which SPApep initially crystallized. Nearly all of the best conditions contained some percentage (8% - 25%) of polyethylene

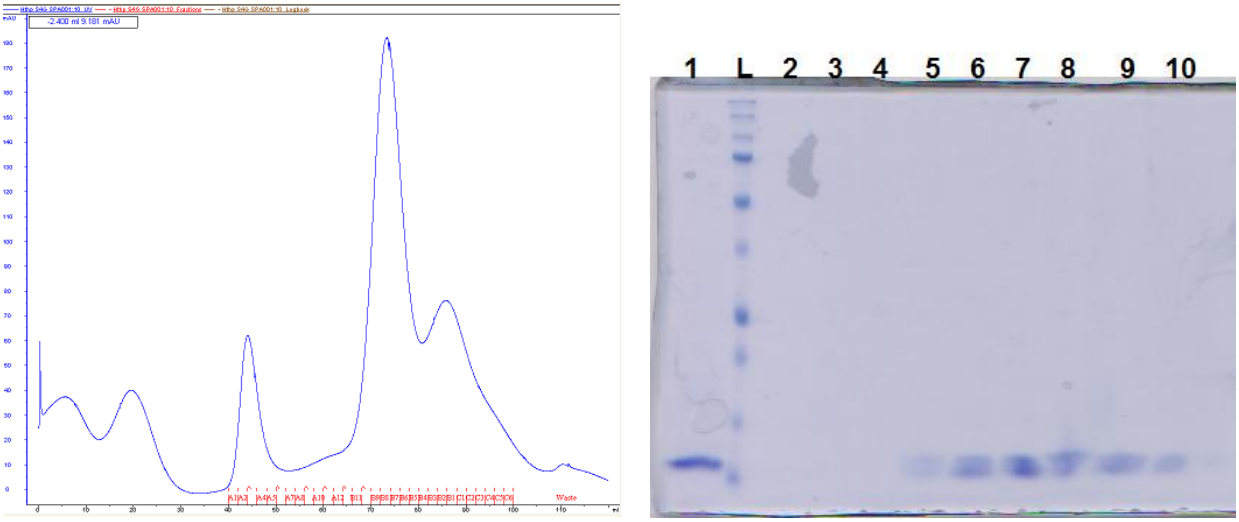


Figure 4.12: HTHP-S4G-SPA purifies as a doublet at low concentration. The elution profile indicates that HTHP-S4G-SPA is a polydisperse protein (left). This is confirmed by SDS-PAGE, where the eluted fractions of HTHP-S4G-SPA run as doublets.

Screen name	Well	Condition
Wizard I + II	A6	0.1 M sodium citrate, pH 5.5, 20% PEG-3000
Hampton Index	E9	0.05 M bis-tris, pH 6.5, 30% pentaerythritol ethoxylate, 0.05 M ammonium sulfate
Hampton Index	G6	0.1 M bis-tris, pH 5.5, 25% PEG-3350, 0.2 M ammonium acetate
Hampton HR2-110/2-112	D1	0.1 M sodium acetate, pH 4.6, 8% PEG-4000
JCSG-Plus	B2	20% PEG-3350, 0.2 M sodium thiocyanate
JCSG-Plus	B9	0.1 M sodium citrate, pH 5.0, 20% PEG-6000
JCSG-Plus	G8	20% PEG-3350, 0.15 M DL-malic acid
PACT Premier	B9	0.1 M MES, pH 6.0, 20% PEG-6000, 0.2 M lithium chloride
PACT Premier	D4	0.1 M MMT buffer, pH 7.0, 25% PEG-1500
PACT Premier	E4	20% PEG-3350, 0.2 M potassium thiocyanate
PACT Premier	E9	20% PEG-3350, 0.2 M potassium/sodium tartrate

Table 4.1: Initial screen hits for SPApep crystals.

glycol (PEG), ranging in molecular weight from PEG-1500 to PEG-6000, though PEG-3350 was most common. The buffered conditions were all at or below pH 7.0, and most of the hits contained 0.05 M – 0.2 M of some salt. Based on the list of hits, I set up optimization plates that tested pH values ranging from 4.5 – 6.5 using sodium citrate and MES buffer, 0 – 27% PEG-3350, and 0 – 0.3 M lithium chloride. Figure 4.13 shows SPApep crystals obtained from a first round of optimization that appear disk-shaped or cylindrical, consistent with the most commonly observed morphologies in the initial screen. However, disc-shaped crystals without a visible flat edge may reflect irregular crystal packing (Yuxin Mao, personal communication). Indeed, the disc-shaped crystals did not diffract to sufficiently high resolution.

I re-examined my initial screening trays and discovered six conditions under which I found crystals with morphologies other than the original disc shape. The two conditions with the best looking crystals were very similar: 0.1 M sodium acetate trihydrate, pH 4.6, 0.2 M ammonium acetate, 20% PEG-4000 and 0.1 M sodium acetate trihydrate,pH 4.5, 25% PEG-3350. More importantly, the crystals were shaped like flat plates and had distinct edges (Figure 4.13). As the two conditions found in the initial screen had the same buffer at virtually the same pH, I varied only the concentration of PEG-3350 and ammonium acetate in my optimization. Larger crystals of similar plate-like morphology were observed in the hanging drops. These were taken to the Cornell High Energy Synchrotron Source (CHESS) for x-ray diffraction on three separate occasions. Unfortunately, the plate crystals produced low quality diffraction patterns unsuitable for structure determination (Figure 4.14). This is likely due to irregular packing of the individual unit cells within the plate crystals. Therefore, the logical next step is to return

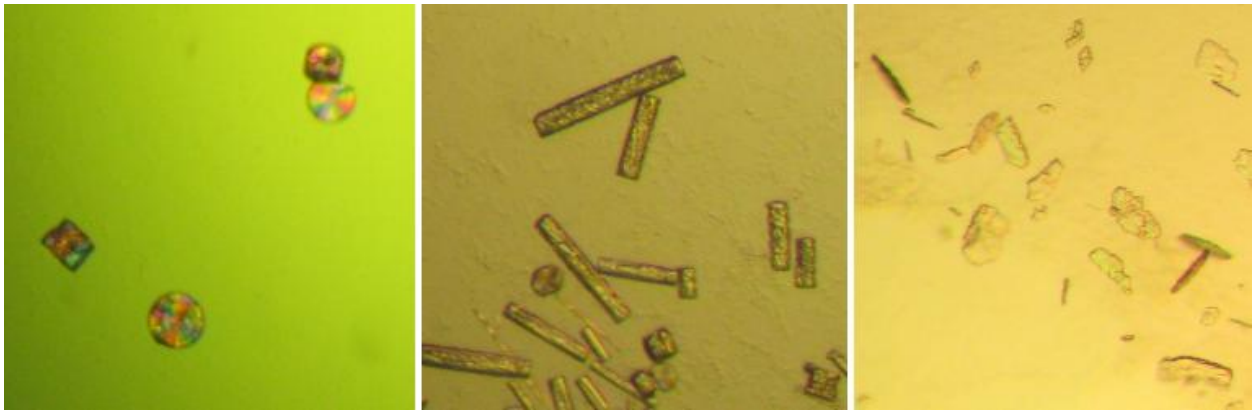


Figure 4.13: SPApep crystals take on three different morphologies. Disc-shaped crystals were the first observed crystal form in MES buffer (left). In some drops, longer rod-shaped crystals were observed together with disc crystals (center). Using sodium acetate buffer, plate crystals eventually formed (right).

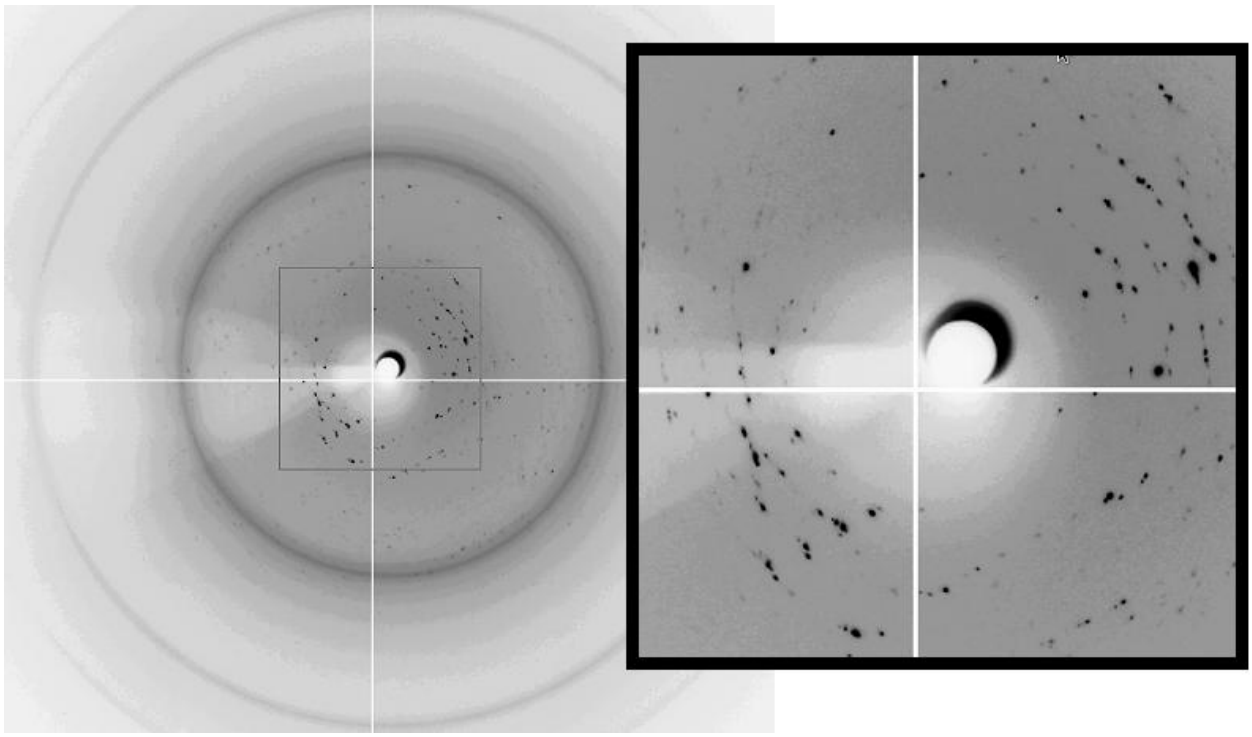


Figure 4.14: Preliminary diffraction pattern of plate-shaped SPApep crystals. The overall diffraction pattern is on the left, and a magnified pattern is shown on the right.

to the initial screen and select conditions for optimization that enable crystals of a different morphology to form.

DISCUSSION

One of the most important challenges in the field of retrovirology is to generate a high-resolution structure of the immature viral core, as it would allow a more complete understanding of the mechanism behind PR-driven maturation, as well as provide structural targets for potential maturation inhibitors. The recent 8 Å structure of MPMV VLPs obtained using cryo-EM tomography, sub-tomogram averaging and atomic-level structure fitting has proven to be highly informative, showing that an immature Gag-like protein assembles into a hexamer using completely novel interfaces when compared to the mature CA hexamer [117]. In this paper, specific residues involved in immature hexamer contacts were inferred based on the best fit of existing crystal structures, but the resolution of the reconstructed structure is insufficient for the conclusive assignment of interacting residues. Therefore, there is still a pressing need for a high-resolution structure of the immature core lattice.

So far, there has been no published crystal structure of a Gag-derived protein that contains the entire CA domain and both the upstream and downstream molecular switches for assembly. This is likely due to the inherent flexibility of the two-domain CA in combination with the molecular switches which may only acquire structure at the high concentrations favorable for assembly, at least in the case of RSV SP. With the data presented in this chapter, I have shown that a minimally flexible construct that contains the RSV CA domain and both molecular switches, termed 25CASP, can be purified to

high concentration and crystallized. Though the small crystals I obtained were not useable for X-ray crystallography, they prove that a Gag-derived protein containing all of the elements required for immature assembly is amenable to crystallization.

An alternative approach to the initial sparse matrix screening for crystal formation would be to assay various conditions for hexamer formation using the 25CASP(CC) double cysteine mutant. Using known conditions for both Gag and CA assembly, I could adjust salt concentrations and pH to generate conditions where hexamers can be observed using non-reducing SDS-PAGE as readout. This method is relatively quick, as over a dozen conditions can be assayed at once. Once an optimal condition for hexamer formation has been found, I would run the protein onto a gel filtration column to purify the hexameric fraction and subject that population of proteins to sparse matrix screening. The two cysteine mutations, located in p10 and CA_{NTD}, were characterized by Phillips et al. to be hexamer-linking cysteines [72]. In principle, the resulting disulfide cross-linking would stabilize the hexamer and minimize flexibility.

Conceptually, using a hexamer template to stabilize the immature retroviral Gag hexamer should help reduce flexibility and promote crystallization. In fact, Pornillos et al. showed that this approach is viable to crystallize CA [64]. The main problem I encountered with the hexamer-templated Ccmk2-25CASP was proteolytic cleavage in the linker region. To solve this issue, the proteins must be purified and kept in protease inhibitor at all times. Another problem was in purifying the protein, as the templated minimal Gag often aggregated on the gel filtration column and eluted in the void volume (Figure 4.8). One possible remedy for this is to vary the length of the linker and assay

the solubility of the resulting proteins using gel filtration (Owen Pornillos, personal communication).

Crystallization of SPApep was another big step toward solving the structure of the RSV downstream molecular switch. I have shown that SPApep can crystallize under a number of conditions and that the resulting crystals are diffraction-capable. What remains to be done here is to continue optimizing crystal growth, focusing specifically on the plate-shaped crystals, which have shown superior diffraction patterns compared to the rod-shaped or disc-shaped crystals. The main difficulty is in growing a sufficiently large crystal that is also a single crystal, as many conditions appear to favor nucleation rather than growth, resulting in an abundance of small crystals that are insufficient for X-ray diffraction. Several methods exist to slow down the growth of crystals in a hanging drop, including lowering the protein concentration in the drop (did not work) and slowing down the exchange between the drop and the well solution by covering the well solution with silicon oil (did not try).

Overall, I have shown that crystallography is a viable method to produce a high-resolution structure of the immature core lattice. With more time and further optimization, diffraction quality crystals can be produced for both the immature lattice and the putative SP 6HB.

CHAPTER FIVE

PERSPECTIVES

Studying retroviral assembly using in vitro assembly

Currently, the best-studied immature retroviral assembly systems are those using some variation of the Gag protein from HIV-1, RSV, and MPMV. Purified Gag-derived proteins from these three retroviruses can assemble in vitro into morphologically similar spherical particles with an electron-lucent center [13, 76, 80, 104, 132, 136, 143, 146, 147, 150, 151], though MPMV Gag can be coaxed into tubes bearing an immature lattice [117]. CA-derived proteins assemble into various morphologies that may be reflective of authentic mature capsids in RSV and HIV-1 capsid assemblies, but not in MPMV [134]. In vitro assembly has proven to be a very useful tool for the study of retroviral core structures, as amino acids critical for core formation can be easily identified through site-directed mutagenesis followed by in vitro assembly as the readout. In vitro assembly can also be a useful method for mapping protein-protein interactions, similar to higher-resolution structure analysis techniques such as X-ray crystallography, but has the added advantage of generating data that are almost certain to be biologically relevant. In vitro assembly can also generate VLPs free of cellular debris or contamination from unassembled protein or mis-folded aggregates, and these VLPs can be further analyzed on the atomic level using cryo-EM, HDX MS, or similar structural analyses that require high sample purity.

The development of retroviral in vitro assembly systems has led to the discovery of the assembly requirements of the immature core and mature capsid. The ability of

Gag and CA to self-assemble, the role of RNA in initiating assembly, the overall core lattice architecture, as well as the structural determinants for the switch between immature and mature assemblies were all discovered using in vitro assembly. The technique of in vitro assembly continues to prove useful for studying retroviral assembly, as evidenced by the data presented in this dissertation.

Molecular switches for assembly

The morphological similarity between immature VLPs is likely due to the structural homology between CA_{NTD} and CA_{CTD} in all retroviral CA proteins described so far [56]. However, specific contacts that promote immature assembly differ between retroviral species and can be found in the regions immediately adjacent to the CA domain. These have come to be known as conformational switches or molecular switches for immature assembly [72, 102, 150], since their presence seems to constrain the CA domain into its immature conformation. For some retroviruses, the upstream switch requires nothing more than residues upstream of the CA domain that prevent the formation of a critical salt bridge between the first proline and a conserved aspartic acid residue of the processed CA protein [52, 57]. Disruption to this beta-hairpin through an N-terminal extension results in the assembly of immature-like VLPs [131, 134]. The downstream switch seems to be amino acid specific, possibly due to its role in forming a higher-order structure, as mutations generally lead to disruption of immature assembly [102, 123, 124, 126, 127, 147, 150]. Although HIV-1 does not have a well-characterized upstream switch, a Gag-derived protein encompassing only CA through NC (CA-NC) assembles into tubes and cones in vitro [15, 111, 130]. RSV CA-NC also assembles

into tubes in vitro [15]. Tubes are generally believed to be reflective of mature assembly, as free CA assembles into tubes [71, 129-131, 165]. Similarly, when the downstream switch is deleted or disrupted, HIV-1 and RSV Gag-derived proteins may assemble into tubes and cones [102, 123, 124, 126, 150]. Therefore, having a single upstream or downstream switch is not sufficient to elicit immature assembly.

Perhaps the best-described molecular switches governing mature and immature assembly are found in RSV. The upstream switch consists of 25 amino acids immediately upstream of CA in p10. When p10 is deleted, the mutant Gag assembles in vitro into tubes [13, 146]. The switch region was narrowed down to the last 25 amino acids in p10 using a series of truncations [146]. In the crystal structure of an N-terminally extended CA_{NTD} dimer, six amino acids in p10 make side-chain contacts with CA_{NTD} [59, 72]. These six residues are critical for immature VLP assembly. As shown by disulfide cross-linking analysis, this switch segment of p10 acts as a bridge between neighboring CA_{NTD} domains, essentially locking the CA_{NTD} domains into an immature hexamer, which differs from the mature CA hexamer in its inter-subunit contacts and closer spacing in the lattice [72]. To date, RSV is the only retrovirus in which a necessary and well-defined amino acid sequence upstream of CA has been identified for immature assembly. Since in HIV-1 CA is directly adjacent to MA, HIV-1 is not believed to require an upstream switch other than the cleavage of CA from MA, allowing the formation of the N-terminal beta-hairpin. Both MPMV and MLV have a 12 kD cleavage product upstream of CA called p12. MPMV p12 does not appear to be crucial for immature assembly; similar to HIV-1, disruption of the N-terminal beta-hairpin suffices for the formation of spherical particles [134]. On the other hand, MLV p12

deletion mutants result in tubular, rather than spherical assembly, suggesting that MLV p12 may interact with CA in a manner similar to that in RSV [95, 96].

The downstream molecular switch in RSV is the SPA domain. The cleaved spacer peptide between CA and NC is common to both RSV and HIV-1, but is not found in MPMV or MLV. Both RSV and HIV-1 spacer peptides are predicted to be part of a 12^{th} helix in CA downstream of the last helix of CA_{CTD} and are exquisitely sensitive to mutation [102, 123, 124, 126, 150, 164, 166-168]. Cryo-ET studies of in vitro assembled HIV-1 and RSV VLPs showed a stalk-like density underneath the organized CA-CTD lattice, which can be fitted with the structure of a parallel 6HB [9, 107]. Although MLV Gag does not have a cleaved peptide between CA and NC, the C-terminus of MLV CA is predicted to form a highly charged helix and mutational studies have shown that this charged helix is required for proper immature assembly and infectivity [127]. MPMV is also predicted to have a 12^{th} helix spanning CA-NC, which overlaps with a previously characterized immature assembly domain similar to RSV SPA located at the N-terminus of the MPMV NC domain [147]. Though the recent high-resolution cryo-ET structure of MPMV VLPs did not show a prominent stalk-like density under the CA_{CTD} layer, the authors point out a ring of densities adjacent to helix 11 in CA_{CTD} [117], which might reflect an alternative higher-order structure involving the predicted helix. In fact, when I aligned the amino acid sequences of representative viruses from every orthoretroviral genus, I found a predicted helix downstream of the structured CA_{CTD} in all retroviruses except epsilonretroviridae. In alpha- and lentiviruses, the predicted helix overlaps with the C-terminus of CA and part or all of the spacer peptide. In beta-, gamma-, and deltaretroviruses, the predicted helix covers the cleavage site between CA and NC.

My hypothesis is that this conserved predicted helix plays a crucial role in immature assembly through homologous interactions with other Gag molecules that expose surfaces favorable for the immature hexamer. When PR cleaves the helix, either by a single cut at the CA-NC junction or two cuts at the CA-SP-NC junctions, the helix loses its structural stability and the downstream molecular switch is flipped to mature. To my knowledge, this is the first time that a model has been proposed for a conserved downstream molecular switch in retroviral assembly.

Future directions

My mutational study of the SPA domain is mostly complete. I have found the specific amino acid residues required for in vitro immature VLP assembly and have provided convincing evidence that the SPA domain forms a 6HB. An important remaining question to address is whether this putative 6HB is a conserved feature among retroviruses, or if it is only a feature of RSV. Some evidence supports both possibilities, as comparative cryo-EM tomography studies of RSV and HIV-1show that both viruses have a conserved stalk below the organized CA_{CTD} layer, but MPMV does not [107, 117]. However, the region downstream of the MPMV CA_{CTD} is sensitive to deletion [147]. It is also predicted to form a helix, which may be part of an observed alternative higher-order structure that promotes immature assembly [117].

One method by which this question can be addressed is by measuring the helical propensity and multimerization states of peptides corresponding to the sequence of the predicted 12^{th} helix in HIV-1, MLV, and MPMV, an experiment that is similar to what I described for RSV SPApep in this dissertation. One could synthesize a peptide

spanning the predicted helical sequence along with several flanking amino acids to preserve any putative upstream and downstream residues that make structural contributions. The peptide structure could be measured using CD and the multimerization state, using AUC. These relatively simple and straightforward experiments would quickly allow one to determine whether the region downstream of CA_{CTD} has a conserved structural function.

Another curious question is whether the salt bridge found in the model of the putative RSV 6HB, formed by R_{493} of one helix and E_{494} of a neighboring helix, is a conserved feature among retroviruses. Although previous work by Keller et al. has shown that an insertion mutation made between the first four residues of NC and the sequence of this predicted salt bridge does not affect in vitro assembly in a baculovirus/insect cell system [126], in vitro assembly would be a more stringent way to assess the contribution of the predicted salt bridge to immature assembly. As previously discussed, cellular assembly systems may contain factors that rescue a mutant that would otherwise not assemble in vitro, and it is always possible that the highly abundant protein expression in insect cells may override the need for the charged residues.

To test the salt bridge hypothesis, one could first mutate both R_{493} and E_{494} to alanines and evaluate VLP formation using in vitro assembly. If the residues do not tolerate mutation to alanine, which one would expect if the predicted salt bridge were required for assembly, then the residues could be mutated to amino acids of opposite but complementary charges, for example $R_{493}E$ and $E_{494}R$, and the resulting protein assessed for assembly. If VLPs form following the charge reversal, this data would

provide convincing evidence for an inter-helix salt bridge in the putative SPApep 6HB. Finally, I could also make the mutations $R_{493}C$ and $E_{494}C$, assemble a 1:1 mixture of the two proteins and measure any resulting VLPs for hexamer cross-linking.

Finally, one could continue to optimize the existing SPApep crystals. Having already overcome the first two of three major hurdles in X-ray crystallography, 1) generating a protein that is soluble at high concentrations, 2) obtaining initial crystal hits, and 3) obtaining diffraction quality crystals, I predict that a structure of SPApep will soon be forthcoming. I am in the process of performing an additive screen on top of the initial screen condition that produced plate-shaped crystals, which will allow me to identify conditions favorable to the formation of larger, more regularly packed crystals that may ultimately lead to the structure of the putative SPA 6HB. As I have a rather detailed model on hand, I anticipate that the structure should be easy to solve. If there are any concerns about the phasing of the helices, it is always possible to use a peptide containing seleno-methionine in the place of one of the two existing methionines.

An atomic-level structure of the RSV SPA 6HB would greatly contribute to our current understanding of immature retroviral assembly, as it could be fitted into the existing lower-resolution cryo-EM densities of immature cores and novel interaction interfaces may be uncovered. This type of information may open the door to the intelligent design of maturation inhibitors. My work, as presented in this dissertation, is a significant step forward toward a high-resolution structure of the immature retroviral core.

APPENDIX I

UNDECAGOLD LABELING OF AN SP CYSTEINE MUTANT

I initiated this project in collaboration with the Briggs Lab at the European Molecular Biology Lab (EMBL) in Heidelberg, Germany. The goal was to label an assembly-capable RSV ΔMBDΔPR(-11C) cysteine mutant in SP using a reagent known as maleimide-undecagold (Au_{11}), assemble VLPs in vitro, take images of the VLPs using cryo-EM tomography, and look for additional electron densities in the reconstructed images. Our hypothesis was that if the RSV SPA domain did correspond to the stalk-like density below the organized CA_{CTD} layer, we would be able to see the additional Au_{11} density on the stalk. I decided to use the $S_{480}C$ mutant for labeling, because I found this protein to assemble at near wild-type levels when assembling point mutants in vitro.

As described in previous publications using Au_{11}, Au_{11}-labeled protein migrate slower than unlabeled protein by approximately 5 kD on SDS-PAGE. Approximately 50 - 75% of $S_{480}C$ is Au_{11}-labeled when measured by SDS-PAGE (Figure I.1). I used dialysis into high-salt buffer (20 mM Tris, pH 7.5, 500 mM NaCl) to dialyze out unused Au_{11} and assembled the mixture of unlabeled and labeled $S_{480}C$. The resulting assembly mixture was evaluated by TEM and VLPs were observed to form as efficiently as assembly usin unlabeled $S_{480}C$ (Figure I.2). The assembly mixture was centrifuged through a sucrose gradient in order to purify VLPs, which formed a strikingly gold-colored band in the gradient, thereby confirming that the labeled protein was incorporated into VLPs (Figure I.3). When the Au_{11}-labeled VLPs were assessed by cryo-EM, however, there was an

abundance of electron-dense material in the background that interfered with image reconstruction. I believe that the background material may either be gold clusters that have dissociated from the Au_{11} complex, or it could be the product of spontaneous VLP disassembly. More experiments are needed to determine the source of the electron-dense material and to find optimal buffer conditions that would prevent gold dissociation or VLP disassembly.

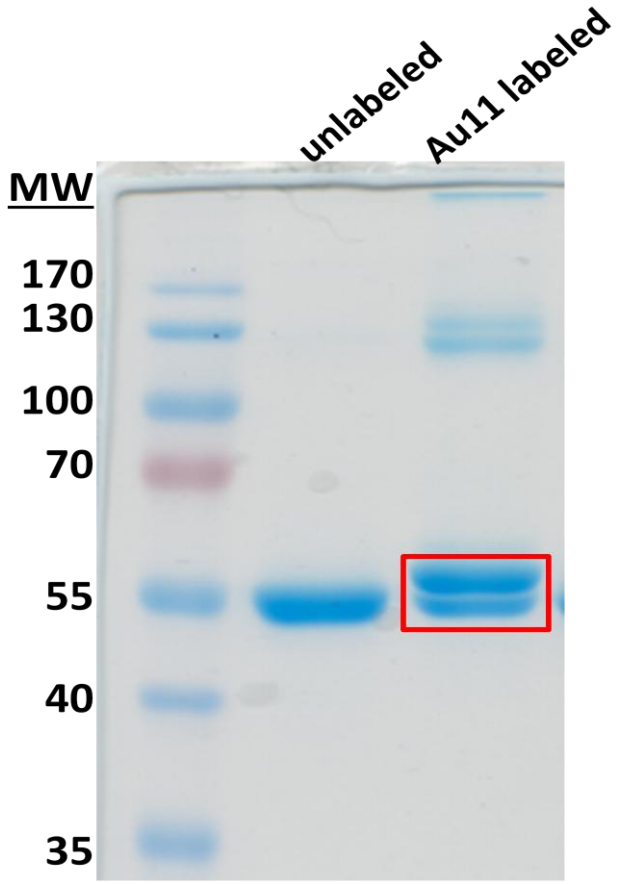


Figure I.1: Undecagold-labeled $S_{480}C$ can be differentiated from unlabeled protein using SDS-PAGE. The doublet boxed in red is both the labeled and unlabeled $S_{480}C$ protein.

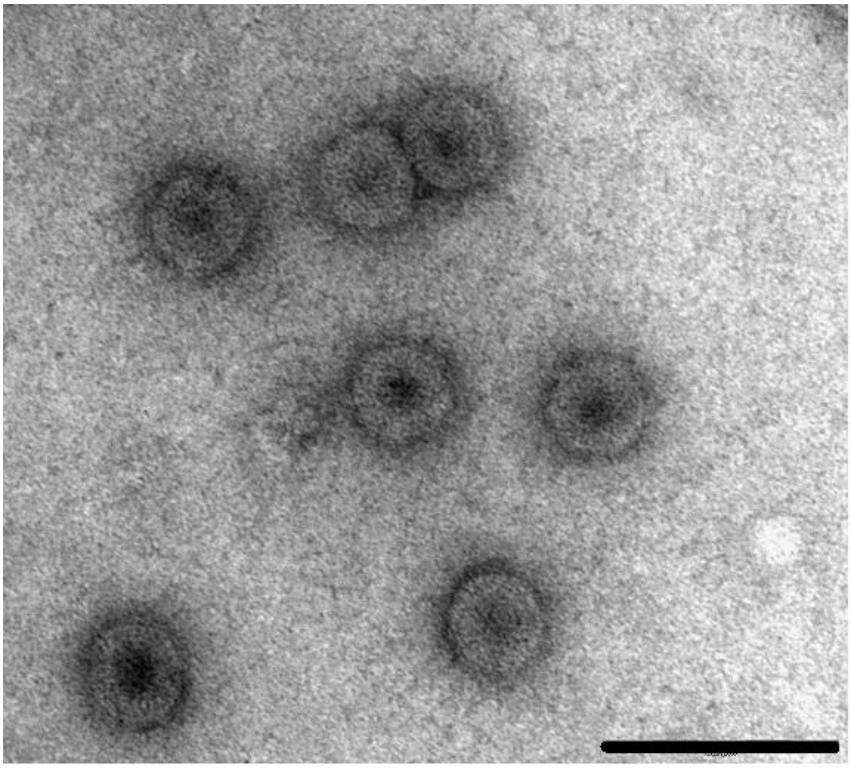


Figure I.2: Au11-S_{480}C assembles as efficiently as unlabeled S_{480}C. The scale bar is 500 nm.

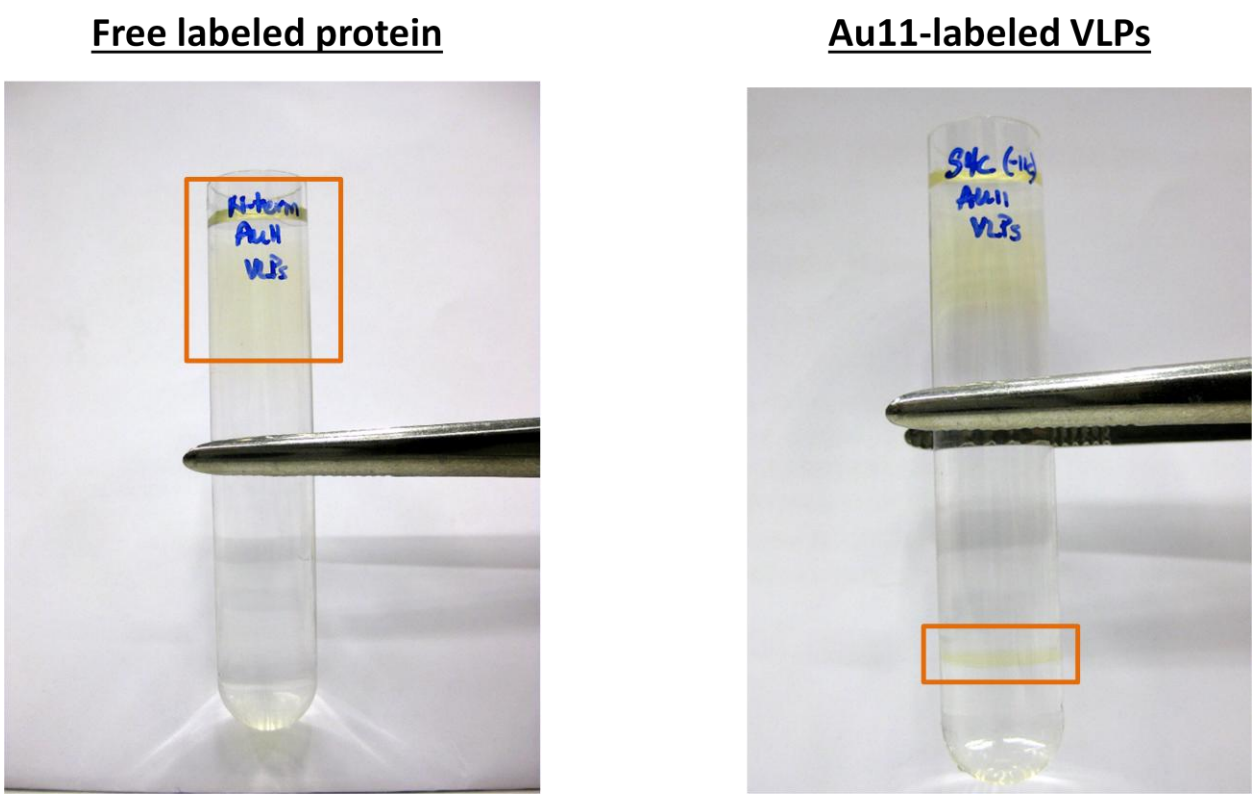


Figure I.3: VLPs assembled using Au11-S_{480}C form a gold-colored band after centrifugation through a sucrose gradient. On the left, the labeled protein does not assemble and most of the gold-colored material is found at the top of the sucrose gradient. On the right, assembled Au_{11}-S_{480}C forms a gold band at approximately the same density as unlabeled VLPs.

REFERENCES

1. Linial, M.L., *Foamy viruses are unconventional retroviruses.* J Virol, 1999. **73**(3): p. 1747-55.
2. Weiss, R.A., *The discovery of endogenous retroviruses.* Retrovirology, 2006. **3**: p. 67.
3. Cherepanov, P., G. Maertens, P. Proost, B. Devreese, J. Van Beeumen, Y. Engelborghs, E. De Clercq, and Z. Debyser, *HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells.* J Biol Chem, 2003. **278**(1): p. 372-81.
4. Maertens, G., P. Cherepanov, W. Pluymers, K. Busschots, E. De Clercq, Z. Debyser, and Y. Engelborghs, *LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells.* J Biol Chem, 2003. **278**(35): p. 33528-39.
5. Coffin, J.M., S.H. Hughes, and H. Varmus, *Retroviruses*. 1997, Plainview, N.Y.: Cold Spring Harbor Laboratory Press. xv, 843 p.
6. Nisole, S. and A. Saib, *Early steps of retrovirus replicative cycle.* Retrovirology, 2004. **1**: p. 9.
7. Fuller, S.D., T. Wilk, B.E. Gowen, H.G. Krausslich, and V.M. Vogt, *Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle.* Curr Biol, 1997. **7**(10): p. 729-38.
8. Briggs, J.A., J.D. Riches, B. Glass, V. Bartonova, G. Zanetti, and H.G. Krausslich, *Structure and assembly of immature HIV.* Proc Natl Acad Sci U S A, 2009. **106**(27): p. 11090-5.
9. Wright, E.R., J.B. Schooler, H.J. Ding, C. Kieffer, C. Fillmore, W.I. Sundquist, and G.J. Jensen, *Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells.* EMBO J, 2007. **26**(8): p. 2218-26.
10. Briggs, J.A., M.N. Simon, I. Gross, H.G. Krausslich, S.D. Fuller, V.M. Vogt, and M.C. Johnson, *The stoichiometry of Gag protein in HIV-1.* Nat Struct Mol Biol, 2004. **11**(7): p. 672-5.

11. Benjamin, J., B.K. Ganser-Pornillos, W.F. Tivol, W.I. Sundquist, and G.J. Jensen, *Three-dimensional structure of HIV-1 virus-like particles by electron cryotomography.* J Mol Biol, 2005. **346**(2): p. 577-88.
12. Butan, C., D.C. Winkler, J.B. Heymann, R.C. Craven, and A.C. Steven, *RSV capsid polymorphism correlates with polymerization efficiency and envelope glycoprotein content: implications that nucleation controls morphogenesis.* J Mol Biol, 2008. **376**(4): p. 1168-81.
13. Campbell, S. and V.M. Vogt, *In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: identification of the p10 domain as a morphological determinant in the formation of spherical particles.* J Virol, 1997. **71**(6): p. 4425-35.
14. de Marco, A., B. Muller, B. Glass, J.D. Riches, H.G. Krausslich, and J.A. Briggs, *Structural analysis of HIV-1 maturation using cryo-electron tomography.* PLoS Pathog, 2010. **6**(11): p. e1001215.
15. Campbell, S. and V.M. Vogt, *Self-assembly in vitro of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1.* J Virol, 1995. **69**(10): p. 6487-97.
16. Campbell, S., R.J. Fisher, E.M. Towler, S. Fox, H.J. Issaq, T. Wolfe, L.R. Phillips, and A. Rein, *Modulation of HIV-like particle assembly in vitro by inositol phosphates.* Proc Natl Acad Sci U S A, 2001. **98**(19): p. 10875-9.
17. Gottlinger, H.G., J.G. Sodroski, and W.A. Haseltine, *Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1.* Proc Natl Acad Sci U S A, 1989. **86**(15): p. 5781-5.
18. Stewart, L., G. Schatz, and V.M. Vogt, *Properties of avian retrovirus particles defective in viral protease.* J Virol, 1990. **64**(10): p. 5076-92.
19. Wiegers, K., G. Rutter, H. Kottler, U. Tessmer, H. Hohenberg, and H.G. Krausslich, *Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites.* J Virol, 1998. **72**(4): p. 2846-54.

20. Pettit, S.C., G.J. Henderson, C.A. Schiffer, and R. Swanstrom, *Replacement of the P1 amino acid of human immunodeficiency virus type 1 Gag processing sites can inhibit or enhance the rate of cleavage by the viral protease.* J Virol, 2002. **76**(20): p. 10226-33.
21. Wyma, D.J., J. Jiang, J. Shi, J. Zhou, J.E. Lineberger, M.D. Miller, and C. Aiken, *Coupling of human immunodeficiency virus type 1 fusion to virion maturation: a novel role of the gp41 cytoplasmic tail.* J Virol, 2004. **78**(7): p. 3429-35.
22. Coren, L.V., J.A. Thomas, E. Chertova, R.C. Sowder, 2nd, T.D. Gagliardi, R.J. Gorelick, and D.E. Ott, *Mutational analysis of the C-terminal gag cleavage sites in human immunodeficiency virus type 1.* J Virol, 2007. **81**(18): p. 10047-54.
23. Keller, P.W., C.S. Adamson, J.B. Heymann, E.O. Freed, and A.C. Steven, *HIV-1 maturation inhibitor bevirimat stabilizes the immature Gag lattice.* J Virol, 2011. **85**(4): p. 1420-8.
24. Li, F., R. Goila-Gaur, K. Salzwedel, N.R. Kilgore, M. Reddick, C. Matallana, A. Castillo, D. Zoumplis, D.E. Martin, J.M. Orenstein, G.P. Allaway, E.O. Freed, and C.T. Wild, *PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13555-60.
25. Nguyen, A.T., C.L. Feasley, K.W. Jackson, T.J. Nitz, K. Salzwedel, G.M. Air, and M. Sakalian, *The prototype HIV-1 maturation inhibitor, bevirimat, binds to the CA-SP1 cleavage site in immature Gag particles.* Retrovirology, 2011. **8**: p. 101.
26. Sakalian, M., C.P. McMurtrey, F.J. Deeg, C.W. Maloy, F. Li, C.T. Wild, and K. Salzwedel, *3-O-(3',3'-dimethysuccinyl) betulinic acid inhibits maturation of the human immunodeficiency virus type 1 Gag precursor assembled in vitro.* J Virol, 2006. **80**(12): p. 5716-22.
27. Adamson, C.S., S.D. Ablan, I. Boeras, R. Goila-Gaur, F. Soheilian, K. Nagashima, F. Li, K. Salzwedel, M. Sakalian, C.T. Wild, and E.O. Freed, *In vitro resistance to the human immunodeficiency virus type 1 maturation inhibitor PA-457 (Bevirimat).* J Virol, 2006. **80**(22): p. 10957-71.

28. Adamson, C.S., M. Sakalian, K. Salzwedel, and E.O. Freed, *Polymorphisms in Gag spacer peptide 1 confer varying levels of resistance to the HIV- 1 maturation inhibitor bevirimat.* Retrovirology, 2010. **7**: p. 36.
29. Zhou, J., C.H. Chen, and C. Aiken, *Human immunodeficiency virus type 1 resistance to the small molecule maturation inhibitor 3-O-(3',3'-dimethylsuccinyl)-betulinic acid is conferred by a variety of single amino acid substitutions at the CA-SP1 cleavage site in Gag.* J Virol, 2006. **80**(24): p. 12095-101.
30. Waki, K., S.R. Durell, F. Soheilian, K. Nagashima, S.L. Butler, and E.O. Freed, *Structural and functional insights into the HIV-1 maturation inhibitor binding pocket.* PLoS Pathog, 2012. **8**(11): p. e1002997.
31. Murray, P.S., Z. Li, J. Wang, C.L. Tang, B. Honig, and D. Murray, *Retroviral matrix domains share electrostatic homology: models for membrane binding function throughout the viral life cycle.* Structure, 2005. **13**(10): p. 1521-31.
32. Saad, J.S., J. Miller, J. Tai, A. Kim, R.H. Ghanam, and M.F. Summers, *Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly.* Proc Natl Acad Sci U S A, 2006. **103**(30): p. 11364-9.
33. Saad, J.S., E. Loeliger, P. Luncsford, M. Liriano, J. Tai, A. Kim, J. Miller, A. Joshi, E.O. Freed, and M.F. Summers, *Point mutations in the HIV-1 matrix protein turn off the myristyl switch.* J Mol Biol, 2007. **366**(2): p. 574-85.
34. Chukkapalli, V., I.B. Hogue, V. Boyko, W.S. Hu, and A. Ono, *Interaction between the human immunodeficiency virus type 1 Gag matrix domain and phosphatidylinositol-(4,5)-bisphosphate is essential for efficient gag membrane binding.* J Virol, 2008. **82**(5): p. 2405-17.
35. Dalton, A.K., D. Ako-Adjei, P.S. Murray, D. Murray, and V.M. Vogt, *Electrostatic interactions drive membrane association of the human immunodeficiency virus type 1 Gag MA domain.* J Virol, 2007. **81**(12): p. 6434-45.
36. Ono, A., S.D. Ablan, S.J. Lockett, K. Nagashima, and E.O. Freed, *Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the plasma membrane.* Proc Natl Acad Sci U S A, 2004. **101**(41): p. 14889-94.
37. Chan, J., R.A. Dick, and V.M. Vogt, *Rous sarcoma virus gag has no specific requirement for phosphatidylinositol-(4,5)-bisphosphate for plasma membrane*

association in vivo or for liposome interaction in vitro. J Virol, 2011. **85**(20): p. 10851-60.

38. Tang, C., E. Loeliger, P. Luncsford, I. Kinde, D. Beckett, and M.F. Summers, *Entropic switch regulates myristate exposure in the HIV-1 matrix protein.* Proc Natl Acad Sci U S A, 2004. **101**(2): p. 517-22.
39. Ono, A., J.M. Orenstein, and E.O. Freed, *Role of the Gag matrix domain in targeting human immunodeficiency virus type 1 assembly.* J Virol, 2000. **74**(6): p. 2855-66.
40. Hill, C.P., D. Worthylake, D.P. Bancroft, A.M. Christensen, and W.I. Sundquist, *Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly.* Proc Natl Acad Sci U S A, 1996. **93**(7): p. 3099-104.
41. Massiah, M.A., D. Worthylake, A.M. Christensen, W.I. Sundquist, C.P. Hill, and M.F. Summers, *Comparison of the NMR and X-ray structures of the HIV-1 matrix protein: evidence for conformational changes during viral assembly.* Protein Sci, 1996. **5**(12): p. 2391-8.
42. Christensen, A.M., M.A. Massiah, B.G. Turner, W.I. Sundquist, and M.F. Summers, *Three-dimensional structure of the HTLV-II matrix protein and comparative analysis of matrix proteins from the different classes of pathogenic human retroviruses.* J Mol Biol, 1996. **264**(5): p. 1117-31.
43. Hatanaka, H., O. Iourin, Z. Rao, E. Fry, A. Kingsman, and D.I. Stuart, *Structure of equine infectious anemia virus matrix protein.* J Virol, 2002. **76**(4): p. 1876-83.
44. Matthews, S., P. Barlow, N. Clark, S. Kingsman, A. Kingsman, and I. Campbell, *Refined solution structure of p17, the HIV matrix protein.* Biochem Soc Trans, 1995. **23**(4): p. 725-9.
45. Rao, Z., A.S. Belyaev, E. Fry, P. Roy, I.M. Jones, and D.I. Stuart, *Crystal structure of SIV matrix antigen and implications for virus assembly.* Nature, 1995. **378**(6558): p. 743-7.
46. Riffel, N., K. Harlos, O. Iourin, Z. Rao, A. Kingsman, D. Stuart, and E. Fry, *Atomic resolution structure of Moloney murine leukemia virus matrix protein and its*

relationship to other retroviral matrix proteins. Structure, 2002. **10**(12): p. 1627-36.

47. Conte, M.R., M. Klikova, E. Hunter, T. Ruml, and S. Matthews, *The three-dimensional solution structure of the matrix protein from the type D retrovirus, the Mason-Pfizer monkey virus, and implications for the morphology of retroviral assembly.* EMBO J, 1997. **16**(19): p. 5819-26.
48. McDonnell, J.M., D. Fushman, S.M. Cahill, W. Zhou, A. Wolven, C.B. Wilson, T.D. Nelle, M.D. Resh, J. Wills, and D. Cowburn, *Solution structure and dynamics of the bioactive retroviral M domain from Rous sarcoma virus.* J Mol Biol, 1998. **279**(4): p. 921-8.
49. Prchal, J., P. Srb, E. Hunter, T. Ruml, and R. Hrabal, *The structure of myristoylated Mason-Pfizer monkey virus matrix protein and the role of phosphatidylinositol-(4,5)-bisphosphate in its membrane binding.* J Mol Biol, 2012. **423**(3): p. 427-38.
50. Vlach, J., J. Lipov, M. Rumlova, V. Veverka, J. Lang, P. Srb, Z. Knejzlik, I. Pichova, E. Hunter, R. Hrabal, and T. Ruml, *D-retrovirus morphogenetic switch driven by the targeting signal accessibility to Tctex-1 of dynein.* Proc Natl Acad Sci U S A, 2008. **105**(30): p. 10565-70.
51. Gamble, T.R., F.F. Vajdos, S. Yoo, D.K. Worthylake, M. Houseweart, W.I. Sundquist, and C.P. Hill, *Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid.* Cell, 1996. **87**(7): p. 1285-94.
52. Gitti, R.K., B.M. Lee, J. Walker, M.F. Summers, S. Yoo, and W.I. Sundquist, *Structure of the amino-terminal core domain of the HIV-1 capsid protein.* Science, 1996. **273**(5272): p. 231-5.
53. Momany, C., L.C. Kovari, A.J. Prongay, W. Keller, R.K. Gitti, B.M. Lee, A.E. Gorbalenya, L. Tong, J. McClure, L.S. Ehrlich, M.F. Summers, C. Carter, and M.G. Rossmann, *Crystal structure of dimeric HIV-1 capsid protein.* Nat Struct Biol, 1996. **3**(9): p. 763-70.
54. Gamble, T.R., S. Yoo, F.F. Vajdos, U.K. von Schwedler, D.K. Worthylake, H. Wang, J.P. McCutcheon, W.I. Sundquist, and C.P. Hill, *Structure of the carboxyl-*

terminal dimerization domain of the HIV-1 capsid protein. Science, 1997. **278**(5339): p. 849-53.

55. Worthylake, D.K., H. Wang, S. Yoo, W.I. Sundquist, and C.P. Hill, *Structures of the HIV-1 capsid protein dimerization domain at 2.6 A resolution.* Acta Crystallogr D Biol Crystallogr, 1999. **55**(Pt 1): p. 85-92.
56. Campos-Olivas, R., J.L. Newman, and M.F. Summers, *Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses.* J Mol Biol, 2000. **296**(2): p. 633-49.
57. Kingston, R.L., T. Fitzon-Ostendorp, E.Z. Eisenmesser, G.W. Schatz, V.M. Vogt, C.B. Post, and M.G. Rossmann, *Structure and self-association of the Rous sarcoma virus capsid protein.* Structure, 2000. **8**(6): p. 617-28.
58. Mortuza, G.B., L.F. Haire, A. Stevens, S.J. Smerdon, J.P. Stoye, and I.A. Taylor, *High-resolution structure of a retroviral capsid hexameric amino-terminal domain.* Nature, 2004. **431**(7007): p. 481-5.
59. Nandhagopal, N., A.A. Simpson, M.C. Johnson, A.B. Francisco, G.W. Schatz, M.G. Rossmann, and V.M. Vogt, *Dimeric rous sarcoma virus capsid protein structure relevant to immature Gag assembly.* J Mol Biol, 2004. **335**(1): p. 275-82.
60. Ganser-Pornillos, B.K., A. Cheng, and M. Yeager, *Structure of full-length HIV-1 CA: a model for the mature capsid lattice.* Cell, 2007. **131**(1): p. 70-9.
61. Mortuza, G.B., M.P. Dodding, D.C. Goldstone, L.F. Haire, J.P. Stoye, and I.A. Taylor, *Structure of B-MLV capsid amino-terminal domain reveals key features of viral tropism, gag assembly and core formation.* J Mol Biol, 2008. **376**(5): p. 1493-508.
62. Macek, P., J. Chmelik, I. Krizova, P. Kaderavek, P. Padrta, L. Zidek, M. Wildova, R. Hadravova, R. Chaloupkova, I. Pichova, T. Ruml, M. Rumlova, and V. Sklenar, *NMR structure of the N-terminal domain of capsid protein from the mason-pfizer monkey virus.* J Mol Biol, 2009. **392**(1): p. 100-14.
63. Mortuza, G.B., D.C. Goldstone, C. Pashley, L.F. Haire, M. Palmarini, W.R. Taylor, J.P. Stoye, and I.A. Taylor, *Structure of the capsid amino-terminal domain*

from the betaretrovirus, Jaagsiekte sheep retrovirus. J Mol Biol, 2009. **386**(4): p. 1179-92.

64. Pornillos, O., B.K. Ganser-Pornillos, B.N. Kelly, Y. Hua, F.G. Whitby, C.D. Stout, W.I. Sundquist, C.P. Hill, and M. Yeager, *X-ray structures of the hexameric building block of the HIV capsid.* Cell, 2009. **137**(7): p. 1282-92.
65. Jin, Z., L. Jin, D.L. Peterson, and C.L. Lawson, *Model for lentivirus capsid core assembly based on crystal dimers of EIAV p26.* J Mol Biol, 1999. **286**(1): p. 83-93.
66. Khorasanizadeh, S., R. Campos-Olivas, and M.F. Summers, *Solution structure of the capsid protein from the human T-cell leukemia virus type-I.* J Mol Biol, 1999. **291**(2): p. 491-505.
67. Pornillos, O., B.K. Ganser-Pornillos, S. Banumathi, Y. Hua, and M. Yeager, *Disulfide bond stabilization of the hexameric capsomer of human immunodeficiency virus.* J Mol Biol, 2010. **401**(5): p. 985-95.
68. Craven, R.C., A.E. Leure-duPree, R.A. Weldon, Jr., and J.W. Wills, *Genetic analysis of the major homology region of the Rous sarcoma virus Gag protein.* J Virol, 1995. **69**(7): p. 4213-27.
69. Cairns, T.M. and R.C. Craven, *Viral DNA synthesis defects in assembly-competent Rous sarcoma virus CA mutants.* J Virol, 2001. **75**(1): p. 242-50.
70. Bowzard, J.B., J.W. Wills, and R.C. Craven, *Second-site suppressors of Rous sarcoma virus Ca mutations: evidence for interdomain interactions.* J Virol, 2001. **75**(15): p. 6850-6.
71. Purdy, J.G., J.M. Flanagan, I.J. Ropson, K.E. Rennoll-Bankert, and R.C. Craven, *Critical role of conserved hydrophobic residues within the major homology region in mature retroviral capsid assembly.* J Virol, 2008. **82**(12): p. 5951-61.
72. Phillips, J.M., P.S. Murray, D. Murray, and V.M. Vogt, *A molecular switch required for retrovirus assembly participates in the hexagonal immature lattice.* EMBO J, 2008. **27**(9): p. 1411-20.
73. Pepinsky, R.B., D. Cappiello, C. Wilkowski, and V.M. Vogt, *Chemical crosslinking of proteins in avian sarcoma and leukemia viruses.* Virology, 1980. **102**(1): p. 205-10.

74. D'Souza, V. and M.F. Summers, *How retroviruses select their genomes.* Nat Rev Microbiol, 2005. **3**(8): p. 643-55.
75. Lee, E.G., A. Alidina, C. May, and M.L. Linial, *Importance of basic residues in binding of rous sarcoma virus nucleocapsid to the RNA packaging signal.* J Virol, 2003. **77**(3): p. 2010-20.
76. Yu, F., S.M. Joshi, Y.M. Ma, R.L. Kingston, M.N. Simon, and V.M. Vogt, *Characterization of Rous sarcoma virus Gag particles assembled in vitro.* J Virol, 2001. **75**(6): p. 2753-64.
77. Dannull, J., A. Surovoy, G. Jung, and K. Moelling, *Specific binding of HIV-1 nucleocapsid protein to PSI RNA in vitro requires N-terminal zinc finger and flanking basic amino acid residues.* EMBO J, 1994. **13**(7): p. 1525-33.
78. Jouvenet, N., S.M. Simon, and P.D. Bieniasz, *Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles.* Proc Natl Acad Sci U S A, 2009. **106**(45): p. 19114-9.
79. Ma, Y.M. and V.M. Vogt, *Rous sarcoma virus Gag protein-oligonucleotide interaction suggests a critical role for protein dimer formation in assembly.* J Virol, 2002. **76**(11): p. 5452-62.
80. Ma, Y.M. and V.M. Vogt, *Nucleic acid binding-induced Gag dimerization in the assembly of Rous sarcoma virus particles in vitro.* J Virol, 2004. **78**(1): p. 52-60.
81. Accola, M.A., B. Strack, and H.G. Gottlinger, *Efficient particle production by minimal Gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain.* J Virol, 2000. **74**(12): p. 5395-402.
82. Johnson, M.C., H.M. Scobie, Y.M. Ma, and V.M. Vogt, *Nucleic acid-independent retrovirus assembly can be driven by dimerization.* J Virol, 2002. **76**(22): p. 11177-85.
83. Zhang, Y., H. Qian, Z. Love, and E. Barklis, *Analysis of the assembly function of the human immunodeficiency virus type 1 gag protein nucleocapsid domain.* J Virol, 1998. **72**(3): p. 1782-9.
84. Bieniasz, P.D., *Late budding domains and host proteins in enveloped virus release.* Virology, 2006. **344**(1): p. 55-63.

85. de Marco, A., A.M. Heuser, B. Glass, H.G. Krausslich, B. Muller, and J.A. Briggs, *Role of the SP2 domain and its proteolytic cleavage in HIV-1 structural maturation and infectivity.* J Virol, 2012. **86**(24): p. 13708-16.
86. Strack, B., A. Calistri, S. Craig, E. Popova, and H.G. Gottlinger, *AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding.* Cell, 2003. **114**(6): p. 689-99.
87. Kogan, M. and J. Rappaport, *HIV-1 accessory protein Vpr: relevance in the pathogenesis of HIV and potential for therapeutic intervention.* Retrovirology, 2011. **8**: p. 25.
88. Gottwein, E., J. Bodem, B. Muller, A. Schmechel, H. Zentgraf, and H.G. Krausslich, *The Mason-Pfizer monkey virus PPPY and PSAP motifs both contribute to virus release.* J Virol, 2003. **77**(17): p. 9474-85.
89. Yasuda, J. and E. Hunter, *A proline-rich motif (PPPY) in the Gag polyprotein of Mason-Pfizer monkey virus plays a maturation-independent role in virion release.* J Virol, 1998. **72**(5): p. 4095-103.
90. Bohl, C.R., S.M. Brown, and R.A. Weldon, Jr., *The pp24 phosphoprotein of Mason-Pfizer monkey virus contributes to viral genome packaging.* Retrovirology, 2005. **2**: p. 68.
91. Knejzlik, Z., Z. Smekalova, T. Ruml, and M. Sakalian, *Multimerization of the p12 domain is necessary for Mason-Pfizer monkey virus Gag assembly in vitro.* Virology, 2007. **365**(2): p. 260-70.
92. Sakalian, M. and E. Hunter, *Separate assembly and transport domains within the Gag precursor of Mason-Pfizer monkey virus.* J Virol, 1999. **73**(10): p. 8073-82.
93. Sommerfelt, M.A., S.S. Rhee, and E. Hunter, *Importance of p12 protein in Mason-Pfizer monkey virus assembly and infectivity.* J Virol, 1992. **66**(12): p. 7005-11.
94. Sakalian, M., S.S. Dittmer, A.D. Gandy, N.D. Rapp, A. Zabransky, and E. Hunter, *The Mason-Pfizer monkey virus internal scaffold domain enables in vitro assembly of human immunodeficiency virus type 1 Gag.* J Virol, 2002. **76**(21): p. 10811-20.

95. Yuan, B., S. Campbell, E. Bacharach, A. Rein, and S.P. Goff, *Infectivity of Moloney murine leukemia virus defective in late assembly events is restored by late assembly domains of other retroviruses.* J Virol, 2000. **74**(16): p. 7250-60.
96. Lee, S.K., K. Nagashima, and W.S. Hu, *Cooperative effect of gag proteins p12 and capsid during early events of murine leukemia virus replication.* J Virol, 2005. **79**(7): p. 4159-69.
97. Yuan, B., A. Fassati, A. Yueh, and S.P. Goff, *Characterization of Moloney murine leukemia virus p12 mutants blocked during early events of infection.* J Virol, 2002. **76**(21): p. 10801-10.
98. Yuan, B., X. Li, and S.P. Goff, *Mutations altering the moloney murine leukemia virus p12 Gag protein affect virion production and early events of the virus life cycle.* EMBO J, 1999. **18**(17): p. 4700-10.
99. Elis, E., M. Ehrlich, A. Prizan-Ravid, N. Laham-Karam, and E. Bacharach, *p12 tethers the murine leukemia virus pre-integration complex to mitotic chromosomes.* PLoS Pathog, 2012. **8**(12): p. e1003103.
100. Prizan-Ravid, A., E. Elis, N. Laham-Karam, S. Selig, M. Ehrlich, and E. Bacharach, *The Gag cleavage product, p12, is a functional constituent of the murine leukemia virus pre-integration complex.* PLoS Pathog, 2010. **6**(11): p. e1001183.
101. Datta, S.A., J.E. Curtis, W. Ratcliff, P.K. Clark, R.M. Crist, J. Lebowitz, S. Krueger, and A. Rein, *Conformation of the HIV-1 Gag protein in solution.* J Mol Biol, 2007. **365**(3): p. 812-24.
102. Datta, S.A., L.G. Temeselew, R.M. Crist, F. Soheilian, A. Kamata, J. Mirro, D. Harvin, K. Nagashima, R.E. Cachau, and A. Rein, *On the role of the SP1 domain in HIV-1 particle assembly: a molecular switch?* J Virol, 2011. **85**(9): p. 4111-21.
103. Datta, S.A., F. Heinrich, S. Raghunandan, S. Krueger, J.E. Curtis, A. Rein, and H. Nanda, *HIV-1 Gag extension: conformational changes require simultaneous interaction with membrane and nucleic acid.* J Mol Biol, 2011. **406**(2): p. 205-14.
104. Campbell, S. and A. Rein, *In vitro assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain.* J Virol, 1999. **73**(3): p. 2270-9.

105. Vogt, V.M. and M.N. Simon, *Mass determination of rous sarcoma virus virions by scanning transmission electron microscopy.* J Virol, 1999. **73**(8): p. 7050-5.
106. Briggs, J.A., M.C. Johnson, M.N. Simon, S.D. Fuller, and V.M. Vogt, *Cryo-electron microscopy reveals conserved and divergent features of gag packing in immature particles of Rous sarcoma virus and human immunodeficiency virus.* J Mol Biol, 2006. **355**(1): p. 157-68.
107. de Marco, A., N.E. Davey, P. Ulbrich, J.M. Phillips, V. Lux, J.D. Riches, T. Fuzik, T. Ruml, H.G. Krausslich, V.M. Vogt, and J.A. Briggs, *Conserved and variable features of Gag structure and arrangement in immature retrovirus particles.* J Virol, 2010. **84**(22): p. 11729-36.
108. Li, S., C.P. Hill, W.I. Sundquist, and J.T. Finch, *Image reconstructions of helical assemblies of the HIV-1 CA protein.* Nature, 2000. **407**(6802): p. 409-13.
109. Briggs, J.A., T. Wilk, R. Welker, H.G. Krausslich, and S.D. Fuller, *Structural organization of authentic, mature HIV-1 virions and cores.* EMBO J, 2003. **22**(7): p. 1707-15.
110. Lanman, J., T.T. Lam, M.R. Emmett, A.G. Marshall, M. Sakalian, and P.E. Prevelige, Jr., *Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange.* Nat Struct Mol Biol, 2004. **11**(7): p. 676-7.
111. Ganser, B.K., S. Li, V.Y. Klishko, J.T. Finch, and W.I. Sundquist, *Assembly and analysis of conical models for the HIV-1 core.* Science, 1999. **283**(5398): p. 80-3.
112. Ganser-Pornillos, B.K., U.K. von Schwedler, K.M. Stray, C. Aiken, and W.I. Sundquist, *Assembly properties of the human immunodeficiency virus type 1 CA protein.* J Virol, 2004. **78**(5): p. 2545-52.
113. Cardone, G., J.G. Purdy, N. Cheng, R.C. Craven, and A.C. Steven, *Visualization of a missing link in retrovirus capsid assembly.* Nature, 2009. **457**(7230): p. 694-8.
114. Ganser, B.K., A. Cheng, W.I. Sundquist, and M. Yeager, *Three-dimensional structure of the M-MuLV CA protein on a lipid monolayer: a general model for retroviral capsid assembly.* EMBO J, 2003. **22**(12): p. 2886-92.
115. Mayo, K., J. McDermott, and E. Barklis, *Hexagonal organization of Moloney murine leukemia virus capsid proteins.* Virology, 2002. **298**(1): p. 30-8.

116. Pornillos, O., B.K. Ganser-Pornillos, and M. Yeager, *Atomic-level modelling of the HIV capsid.* Nature, 2011. **469**(7330): p. 424-7.
117. Bharat, T.A., N.E. Davey, P. Ulbrich, J.D. Riches, A. de Marco, M. Rumlova, C. Sachse, T. Ruml, and J.A. Briggs, *Structure of the immature retroviral capsid at 8 A resolution by cryo-electron microscopy.* Nature, 2012. **487**(7407): p. 385-9.
118. Lanman, J., T.T. Lam, S. Barnes, M. Sakalian, M.R. Emmett, A.G. Marshall, and P.E. Prevelige, Jr., *Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry.* J Mol Biol, 2003. **325**(4): p. 759-72.
119. von Schwedler, U.K., K.M. Stray, J.E. Garrus, and W.I. Sundquist, *Functional surfaces of the human immunodeficiency virus type 1 capsid protein.* J Virol, 2003. **77**(9): p. 5439-50.
120. Byeon, I.J., X. Meng, J. Jung, G. Zhao, R. Yang, J. Ahn, J. Shi, J. Concel, C. Aiken, P. Zhang, and A.M. Gronenborn, *Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function.* Cell, 2009. **139**(4): p. 780-90.
121. Bailey, G.D., J.K. Hyun, A.K. Mitra, and R.L. Kingston, *A structural model for the generation of continuous curvature on the surface of a retroviral capsid.* J Mol Biol, 2012. **417**(3): p. 212-23.
122. Hyun, J.K., M. Radjainia, R.L. Kingston, and A.K. Mitra, *Proton-driven assembly of the Rous Sarcoma virus capsid protein results in the formation of icosahedral particles.* J Biol Chem, 2010. **285**(20): p. 15056-64.
123. Krausslich, H.G., M. Facke, A.M. Heuser, J. Konvalinka, and H. Zentgraf, *The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity.* J Virol, 1995. **69**(6): p. 3407-19.
124. Accola, M.A., S. Hoglund, and H.G. Gottlinger, *A putative alpha-helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly.* J Virol, 1998. **72**(3): p. 2072-8.

125. Morellet, N., S. Druillennec, C. Lenoir, S. Bouaziz, and B.P. Roques, *Helical structure determined by NMR of the HIV-1 (345-392)Gag sequence, surrounding p2: implications for particle assembly and RNA packaging.* Protein Sci, 2005. **14**(2): p. 375-86.
126. Keller, P.W., M.C. Johnson, and V.M. Vogt, *Mutations in the spacer peptide and adjoining sequences in Rous sarcoma virus Gag lead to tubular budding.* J Virol, 2008. **82**(14): p. 6788-97.
127. Cheslock, S.R., D.T. Poon, W. Fu, T.D. Rhodes, L.E. Henderson, K. Nagashima, C.F. McGrath, and W.S. Hu, *Charged assembly helix motif in murine leukemia virus capsid: an important region for virus assembly and particle size determination.* J Virol, 2003. **77**(12): p. 7058-66.
128. Auerbach, M.R., K.R. Brown, A. Kaplan, D. de Las Nueces, and I.R. Singh, *A small loop in the capsid protein of Moloney murine leukemia virus controls assembly of spherical cores.* J Virol, 2006. **80**(6): p. 2884-93.
129. Ehrlich, L.S., B.E. Agresta, and C.A. Carter, *Assembly of recombinant human immunodeficiency virus type 1 capsid protein in vitro.* J Virol, 1992. **66**(8): p. 4874-83.
130. Gross, I., H. Hohenberg, and H.G. Krausslich, *In vitro assembly properties of purified bacterially expressed capsid proteins of human immunodeficiency virus.* Eur J Biochem, 1997. **249**(2): p. 592-600.
131. Gross, I., H. Hohenberg, C. Huckhagel, and H.G. Krausslich, *N-Terminal extension of human immunodeficiency virus capsid protein converts the in vitro assembly phenotype from tubular to spherical particles.* J Virol, 1998. **72**(6): p. 4798-810.
132. Klikova, M., S.S. Rhee, E. Hunter, and T. Ruml, *Efficient in vivo and in vitro assembly of retroviral capsids from Gag precursor proteins expressed in bacteria.* J Virol, 1995. **69**(2): p. 1093-8.
133. Menendez-Arias, L., C. Risco, P. Pinto da Silva, and S. Oroszlan, *Purification of immature cores of mouse mammary tumor virus and immunolocalization of protein domains.* J Virol, 1992. **66**(9): p. 5615-20.

134. Rumlova-Klikova, M., E. Hunter, M.V. Nermut, I. Pichova, and T. Ruml, *Analysis of Mason-Pfizer monkey virus Gag domains required for capsid assembly in bacteria: role of the N-terminal proline residue of CA in directing particle shape.* J Virol, 2000. **74**(18): p. 8452-9.
135. Ulbrich, P., S. Haubova, M.V. Nermut, E. Hunter, M. Rumlova, and T. Ruml, *Distinct roles for nucleic acid in in vitro assembly of purified Mason-Pfizer monkey virus CANC proteins.* J Virol, 2006. **80**(14): p. 7089-99.
136. Purdy, J.G., J.M. Flanagan, I.J. Ropson, and R.C. Craven, *Retroviral capsid assembly: a role for the CA dimer in initiation.* J Mol Biol, 2009. **389**(2): p. 438-51.
137. Kerfeld, C.A., M.R. Sawaya, S. Tanaka, C.V. Nguyen, M. Phillips, M. Beeby, and T.O. Yeates, *Protein structures forming the shell of primitive bacterial organelles.* Science, 2005. **309**(5736): p. 936-8.
138. Scopes, R.K., *Measurement of protein by spectrophotometry at 205 nm.* Anal Biochem, 1974. **59**(1): p. 277-82.
139. Reed, J. and T.A. Reed, *A set of constructed type spectra for the practical estimation of peptide secondary structure from circular dichroism.* Anal Biochem, 1997. **254**(1): p. 36-40.
140. Amon, M.A., M. Ali, V. Bender, K. Hall, M.I. Aguilar, J. Aldrich-Wright, and N. Manolios, *Kinetic and conformational properties of a novel T-cell antigen receptor transmembrane peptide in model membranes.* J Pept Sci, 2008. **14**(6): p. 714-24.
141. Bailey, G.D., J.K. Hyun, A.K. Mitra, and R.L. Kingston, *Proton-linked dimerization of a retroviral capsid protein initiates capsid assembly.* Structure, 2009. **17**(5): p. 737-48.
142. Affranchino, J.L. and S.A. Gonzalez, *In vitro assembly of the feline immunodeficiency virus Gag polyprotein.* Virus Res, 2010. **150**(1-2): p. 153-7.
143. Crist, R.M., S.A. Datta, A.G. Stephen, F. Soheilian, J. Mirro, R.J. Fisher, K. Nagashima, and A. Rein, *Assembly properties of human immunodeficiency virus type 1 Gag-leucine zipper chimeras: implications for retrovirus assembly.* J Virol, 2009. **83**(5): p. 2216-25.

144. Alfadhli, A., T.C. Dhenub, A. Still, and E. Barklis, *Analysis of human immunodeficiency virus type 1 Gag dimerization-induced assembly.* J Virol, 2005. **79**(23): p. 14498-506.
145. Krishna, N.K., S. Campbell, V.M. Vogt, and J.W. Wills, *Genetic determinants of Rous sarcoma virus particle size.* J Virol, 1998. **72**(1): p. 564-77.
146. Joshi, S.M. and V.M. Vogt, *Role of the Rous sarcoma virus p10 domain in shape determination of gag virus-like particles assembled in vitro and within Escherichia coli.* J Virol, 2000. **74**(21): p. 10260-8.
147. Bohmova, K., R. Hadravova, J. Stokrova, R. Tuma, T. Ruml, I. Pichova, and M. Rumlova, *Effect of dimerizing domains and basic residues on in vitro and in vivo assembly of Mason-Pfizer monkey virus and human immunodeficiency virus.* J Virol, 2010. **84**(4): p. 1977-88.
148. Hadravova, R., A. de Marco, P. Ulbrich, J. Stokrova, M. Dolezal, I. Pichova, T. Ruml, J.A. Briggs, and M. Rumlova, *In vitro assembly of virus-like particles of a gammaretrovirus, the murine leukemia virus XMRV.* J Virol, 2012. **86**(3): p. 1297-306.
149. Nermut, M.V., P. Bron, D. Thomas, M. Rumlova, T. Ruml, and E. Hunter, *Molecular organization of Mason-Pfizer monkey virus capsids assembled from Gag polyprotein in Escherichia coli.* J Virol, 2002. **76**(9): p. 4321-30.
150. Gross, I., H. Hohenberg, T. Wilk, K. Wiegers, M. Grattinger, B. Muller, S. Fuller, and H.G. Krausslich, *A conformational switch controlling HIV-1 morphogenesis.* EMBO J, 2000. **19**(1): p. 103-13.
151. Wilk, T., I. Gross, B.E. Gowen, T. Rutten, F. de Haas, R. Welker, H.G. Krausslich, P. Boulanger, and S.D. Fuller, *Organization of immature human immunodeficiency virus type 1.* J Virol, 2001. **75**(2): p. 759-71.
152. Marqusee, S., V.H. Robbins, and R.L. Baldwin, *Unusually stable helix formation in short alanine-based peptides.* Proc Natl Acad Sci U S A, 1989. **86**(14): p. 5286-90.
153. Rohl, C.A., W. Fiori, and R.L. Baldwin, *Alanine is helix-stabilizing in both template-nucleated and standard peptide helices.* Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3682-7.

154. Smith, T.J., E. Chase, T. Schmidt, and K.L. Perry, *The structure of cucumber mosaic virus and comparison to cowpea chlorotic mottle virus.* J Virol, 2000. **74**(16): p. 7578-86.
155. Mayo, K., D. Huseby, J. McDermott, B. Arvidson, L. Finlay, and E. Barklis, *Retrovirus capsid protein assembly arrangements.* J Mol Biol, 2003. **325**(1): p. 225-37.
156. Meng, X., G. Zhao, E. Yufenyuy, D. Ke, J. Ning, M. Delucia, J. Ahn, A.M. Gronenborn, C. Aiken, and P. Zhang, *Protease cleavage leads to formation of mature trimer interface in HIV-1 capsid.* PLoS Pathog, 2012. **8**(8): p. e1002886.
157. Buck, M., *Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins.* Q Rev Biophys, 1998. **31**(3): p. 297-355.
158. Thomas, P.D. and K.A. Dill, *Local and nonlocal interactions in globular proteins and mechanisms of alcohol denaturation.* Protein Sci, 1993. **2**(12): p. 2050-65.
159. Jeoung, J.H., D.A. Pippig, B.M. Martins, N. Wagener, and H. Dobbek, *HTHP: a novel class of hexameric, tyrosine-coordinated heme proteins.* J Mol Biol, 2007. **368**(4): p. 1122-31.
160. Zaccai, N.R., B. Chi, A.R. Thomson, A.L. Boyle, G.J. Bartlett, M. Bruning, N. Linden, R.B. Sessions, P.J. Booth, R.L. Brady, and D.N. Woolfson, *A de novo peptide hexamer with a mutable channel.* Nat Chem Biol, 2011. **7**(12): p. 935-41.
161. Sandefur, S., R.M. Smith, V. Varthakavi, and P. Spearman, *Mapping and characterization of the N-terminal I domain of human immunodeficiency virus type 1 Pr55(Gag).* J Virol, 2000. **74**(16): p. 7238-49.
162. Pettit, S.C., M.D. Moody, R.S. Wehbie, A.H. Kaplan, P.V. Nantermet, C.A. Klein, and R. Swanstrom, *The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions.* J Virol, 1994. **68**(12): p. 8017-27.
163. Pepinsky, R.B., I.A. Papayannopoulos, E.P. Chow, N.K. Krishna, R.C. Craven, and V.M. Vogt, *Differential proteolytic processing leads to multiple forms of the CA protein in avian sarcoma and leukemia viruses.* J Virol, 1995. **69**(10): p. 6430-8.

164. Guo, X., J. Hu, J.B. Whitney, R.S. Russell, and C. Liang, *Important role for the CA-NC spacer region in the assembly of bovine immunodeficiency virus Gag protein.* J Virol, 2004. **78**(2): p. 551-60.
165. Barklis, E., A. Alfadhli, C. McQuaw, S. Yalamuri, A. Still, R.L. Barklis, B. Kukull, and C.S. Lopez, *Characterization of the in vitro HIV-1 capsid assembly pathway.* J Mol Biol, 2009. **387**(2): p. 376-89.
166. Craven, R.C., A.E. Leure-duPree, C.R. Erdie, C.B. Wilson, and J.W. Wills, *Necessity of the spacer peptide between CA and NC in the Rous sarcoma virus gag protein.* J Virol, 1993. **67**(10): p. 6246-52.
167. Liang, C., J. Hu, R.S. Russell, A. Roldan, L. Kleiman, and M.A. Wainberg, *Characterization of a putative alpha-helix across the capsid-SP1 boundary that is critical for the multimerization of human immunodeficiency virus type 1 gag.* J Virol, 2002. **76**(22): p. 11729-37.
168. Morikawa, Y., D.J. Hockley, M.V. Nermut, and I.M. Jones, *Roles of matrix, p2, and N-terminal myristoylation in human immunodeficiency virus type 1 Gag assembly.* J Virol, 2000. **74**(1): p. 16-23.