Abstract

Modular HIV-1 Capsid Assemblies for Investigating Diverse Host Recognition Mechanisms

Brady J. Summers

2018

The human immunodeficiency virus-1 (HIV-1) capsid is an ordered protein shell built from a repeating array of the viral CA protein. The assembled capsid houses and protects the viral genome during early stages of infection. Once inside an infected cell, it is an expansive protein-docking platform whose finely-tuned structural integrity is critical for viral infection. Numerous host factors, both beneficial and harmful to the virus, recognize and directly bind specific capsid surfaces. Unfortunately, our mechanistic understanding of host-capsid interactions is well behind that of our knowledge of apo capsid structure. This is because many capsid binding factors only recognize high-order capsid lattice patterns and do not show appreciable affinity for soluble capsid building blocks, such as CA dimers and hexamers. Without appropriate CA-binding partners, host factor-CA complexes cannot be formed in solution, and the vast array of solution-based biochemical and structural biology techniques cannot be applied.

To overcome these limitations, I developed novel techniques to trap and stabilize CA oligomers in lattice-like fragments that are intermediates between unassembled and fully assembled capsid. These engineered lattice assemblies range in size from 50 kDa to 1 MDa (or 2-42 CA molecules) and exhibit every

unique capsid lattice interface found in infectious virions. Many of the assemblies bridge the hexamer-hexamer and hexamer-pentamer surfaces formed by the CA C-terminal domains that have been previously impossible to capture. These lattice-intermediates enable analysis of elusive capsid-host factor binding events in solution—avoiding the pitfalls associated with capsid tubes and cones.

I have used these assemblies to better define the unique capsid-binding modes of several well-studied, but poorly understood, restriction factors (TRIM5α, TRIMCyp, and MxB) and a recently discovered viral cofactor (FEZ1). Each of these factors demonstrated a unique capsid lattice-sensing ability that could only be defined with our novel CA assemblies. This work significantly expands our understanding of the diverse array of capsid-sensing motifs and targetable capsid surfaces. Besides mechanistic investigations, the advanced toolkit we have created will be valuable for searching for additional capsid-binding factors or designing specific capsid-targeting therapeutics to inhibit HIV-1 infectivity.

Modular HIV-1 Capsid Assemblies for Investigating Diverse Host Recognition Mechanisms

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Brady J. Summers

Dissertation Director: Yong Xiong

December 2018

Acknowledgements

I would like to thank my parents and siblings for their support and sacrifices throughout the years. Even though they were unsure about my decision to come to Yale to get a PhD, they have been with me my entire life and have always pushed me to succeed. I hope that I can spend more time with them and my little nieces and nephews back in Missouri when I successfully defend my thesis.

I must thank my partner, Caroline, for always supporting me through the ups and downs of a PhD. I met you the first week of graduate school and knew from the very beginning that you were the one for me. You have made these years fun and I look forward to spending the years to come with you.

I would like to thank my advisor, Yong Xiong, for providing the best possible lab atmosphere I could have asked for and for helping me develop as a scientist. I have grown and learned a lot about being a scientist during my time at Yale, and this is directly attributable to your easy-going guidance. You have always let me go for any crazy ideas!

Thank you to all the Xiong lab members, both former and current, who have guided me and been my friends over the years. You have made it a joy to come into lab. To the senior lab members, Dr. Jenny Fribourgh and Dr. Henry Nguyen, who taught me the most in the early years, thank you.

I must thank my thesis committee, Walther Mothes and Andrew Miranker, for always being engaged in my meetings, for pushing me to improve my presentations, and to think about the bigger picture. Their effort and criticisms

throughout the years are much appreciated. I think Walther is the biggest fan of my research in the entire HIV community.

Finally, I have to thank all the friends I have made playing softball and basketball throughout the years. You have given me a welcome break from the rigors of graduate school. I hope you all win lots of future IM championships!

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 The HIV-1 capsid is a highly evolved viral genome carrier

Viruses infecting all lifeforms share two common features: (1) the presence of either an RNA or DNA genome that is replicated and transmitted from cell to cell, and (2) a “capsid”—a protein shell that houses the viral genome. Capsids are ordered protein polymers composed of thousands of protein molecules. They are usually assembled from a geometric combination of one or several unique proteins. Capsids come in a variety of different architectures, most commonly icosahedral or helical, but some have evolved unique structures **(Figure 1-1A)** [29].

Retroviral capsids can be spherical, cone-like, or tube-like, but they all are organized by fullerene cone principles **(Figure 1-1B)** [30]. A common example of a fullerene cone is a soccer ball. The pattern on a soccer ball is composed of 12

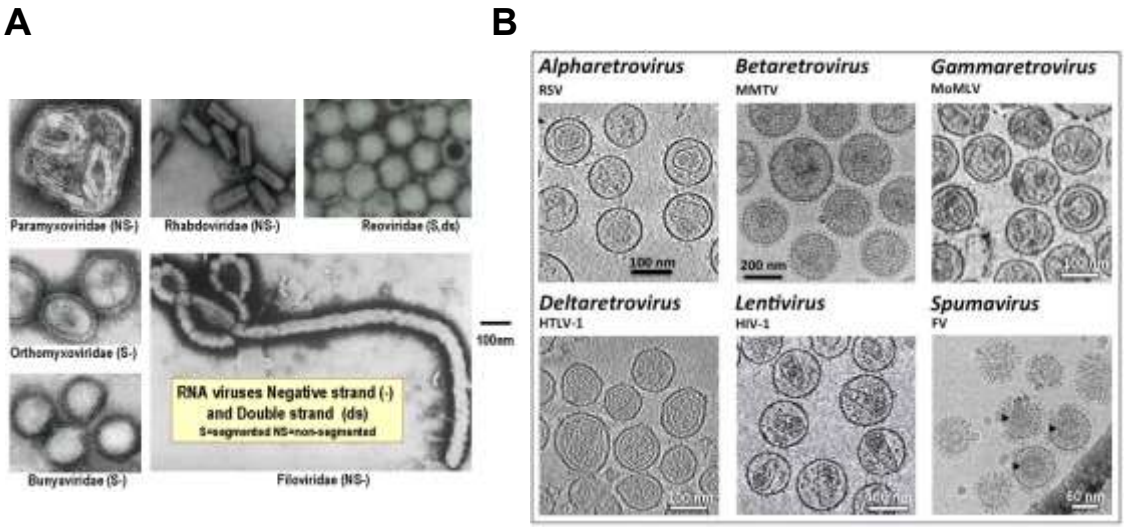


Figure 1-1. Viral capsids in nature. A) Negative-stain electron micrographs of various viral capsids. B) Representative retroviral capsids and envelopes highlighting the structural variability between genera. (A) is adapted from [10]. (B) is adapted from [11].

pentamers and 20 hexamers. A retroviral capsid, likewise, is composed of 12 pentamer subunits and a range of hexamer subunits (from 20-300+). The ratio of pentamer subunits (always 12 in total) to hexamer subunits defines the shape of the capsid. If a capsid contains fewer hexamers, it tends to be a more spherical. If it contains only hexamers, it tends to assemble as a tube [31]. All retroviral capsids are surrounded by a lipid membrane envelope derived from a previously infected cell [32].

Human immunodeficiency virus-1 (HIV-1) is a retrovirus of the lentiviral family [33]. The mature HIV-1 capsid is a fullerene cone composed of 12 pentamers and approximately 250 hexamers. It is a nearly 40 megadalton protein super-structure built from roughly 1,500 copies of the 25-kDa viral capsid protein (CA) [34-36]. Unlike many icosahedral capsids, the HIV-1 capsid is very unstable and structurally heterogeneous. In fact, a significant number of capsids are broken or not closed [8, 37, 38]. Historically, these features precluded high-resolution structural studies of the assembled capsid.

CA contains independently folded N- and C-terminal α-helical domains (termed NTD and CTD, respectively) separated by a flexible linker **(Figure 1-2A)**[39]. The structure of CA is conserved amongst retroviruses despite low sequence homology between species and genera [40-42]. NTD-NTD and NTD-CTD oligomerization interfaces define and stabilize rigid intra-hexamer/pentamer structures. CTD-CTD interfaces, including independent dimerization and trimerization motifs, are responsible for bridging adjacent hexamers and pentamers to form the cone-shaped capsid [1, 39, 43]. The CTD dimerization (or

two-fold) interface is found between two hexamers within the capsid lattice. The CTD trimerization (or three-fold) interface is found at the center of three hexamers in the lattice. Hexamers and pentamers share a quasi-equivalent structure. Only slight alterations of angle and distance between CA molecules distinguish hexamers from pentamers **(Figure 1-2B-F)** [8, 18].

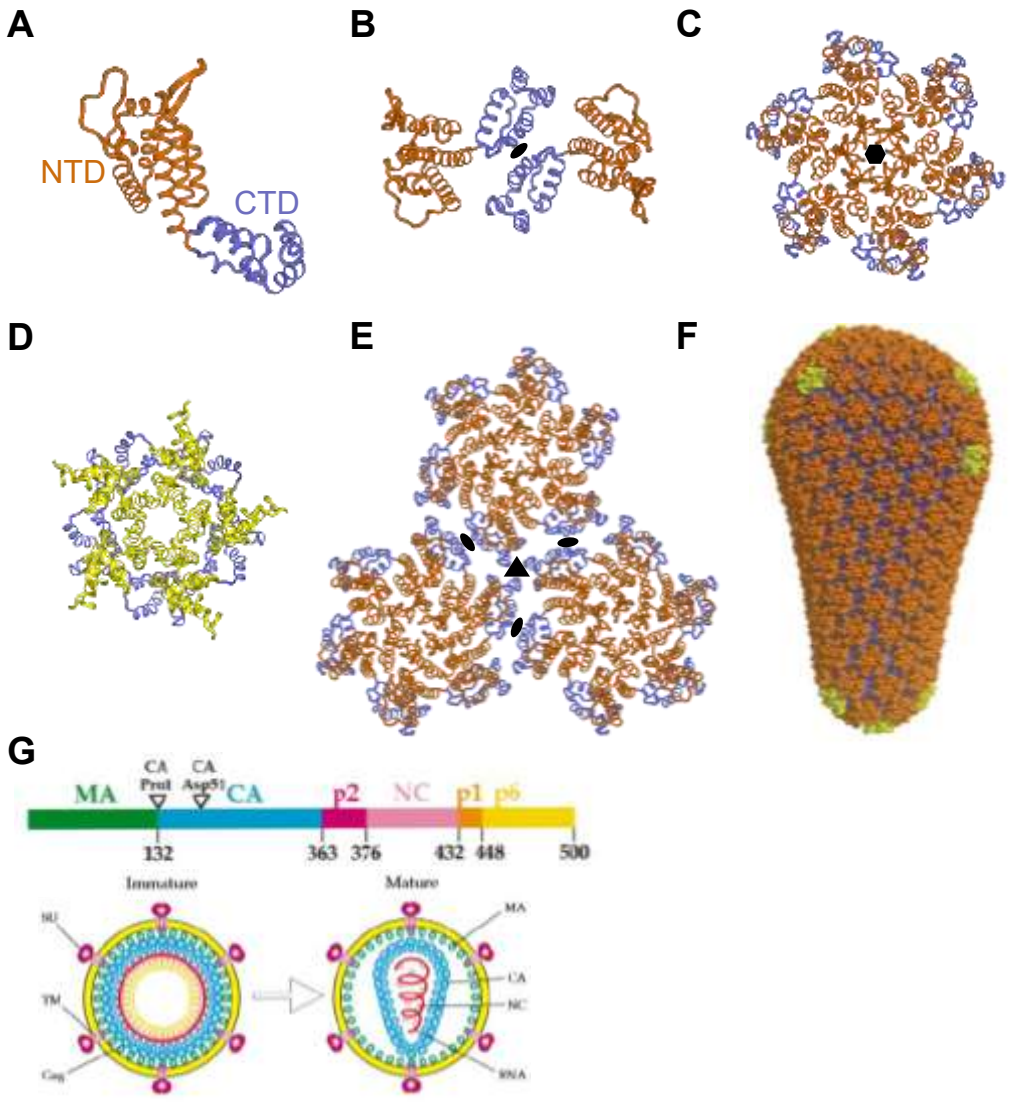


Figure 1-2. Levels of HIV-1 capsid structure. A) CA monomer. B) CA dimer. C) CA hexamer. D) CA pentamer. E) Three interacting hexamers. F) A complete capsid with hexamers and pentamers. G) Gag gene map (top) and immature versus mature capsid (bottom). (A)-(C) and (E) from PDB ID: 4XFX [4]. (D) from PDB ID: 5MCY [8]. (F) adapted from [18]. (G) adapted from [20]

How is the HIV-1 capsid assembled? In an infected cell, CA is expressed as part of the viral Gag polyprotein, which includes matrix, CA, P2, nucleocapsid, P1, and P6 domains **(Figure 1-2G, top panel)**. Gag assembles at the cell membrane into a spherical “immature” capsid before budding from the infected cell as an infectious virion. Before the virion reaches a new target cell, protease cleavage frees CA from the remaining Gag domains. In a poorly understood fashion, cleaved CA assembles into the “mature” capsid form **(Figure 1-2G, bottom panel)** (reviewed in [44]). The structural differences between immature and mature capsid are remarkable but will not be discussed further. The mature capsid is the focus of this thesis, and I will refer to it only as “capsid” hereafter.

The preeminent HIV-1 capsid structures were obtained by Briggs and colleagues by advanced cryo-electron microscopy tomography (cryo-ET) [8, 45]. They achieved high-resolution structures of both immature and mature capsids obtained from infected tissue culture cells. This work validated much of the previous decade’s worth of mature capsid structural data (some of my favorites: [1, 4, 16, 18, 39]) and provided new insights into pentamer structure. The work was a tour de force of EM experimental technique and changed perceptions of the resolution achievable by cryo-ET. The field of apo-capsid structure is considered mature at the time of this writing.

1.2 Fate of the capsid after cellular entry

In recent years, much work has been devoted to understanding what happens to the HIV-1 capsid after it enters a host cell. We know that the capsid is released into the cytosol after the viral envelope fuses with the cell membrane.

Then, inside the capsid, the single-stranded viral RNA genome is reverse transcribed into double-stranded DNA. As this happens, the capsid, likely in some state of disassembly, traffics towards nuclear pores along the microtubule network. The capsid with reverse transcribed viral DNA next docks at nuclear pores. Finally, the genome crosses the pores while the viral integrase protein integrates the viral genome into the host genome [14].

The earliest data implied that capsids immediately “uncoat,” or disassemble, after entry [46, 47]. This would leave the viral genome unprotected in the cytosol and would require some means of transport (or simple diffusion) of the genome towards the nucleus. A model of this nature was eventually deemed unlikely for several reasons, including data demonstrating that the capsid protects the genome from cytosolic DNA sensors and that the capsid is a determinant of infection of non-dividing cells [48-51]. The latter data links capsid to infection events at the nuclear membrane, which happen long after a capsid has entered the cell.

Recent advanced live-cell imaging studies have greatly expanded our knowledge of CA uncoating [52, 53]. These techniques have enabled direct visualization of capsid sub-cellular localization and timing of structural integrity loss during infection. Capsid ‘integrity’ is defined as the ability of co-packaged green fluorescent protein (GFP) to remain in the capsid interior. It appears that capsid integrity is lost early in infection—by 30 minutes post-infection [53]. However, at least partially assembled capsids were observed several hours post-infection docking at nuclear pores. These stable capsids were demonstrated to

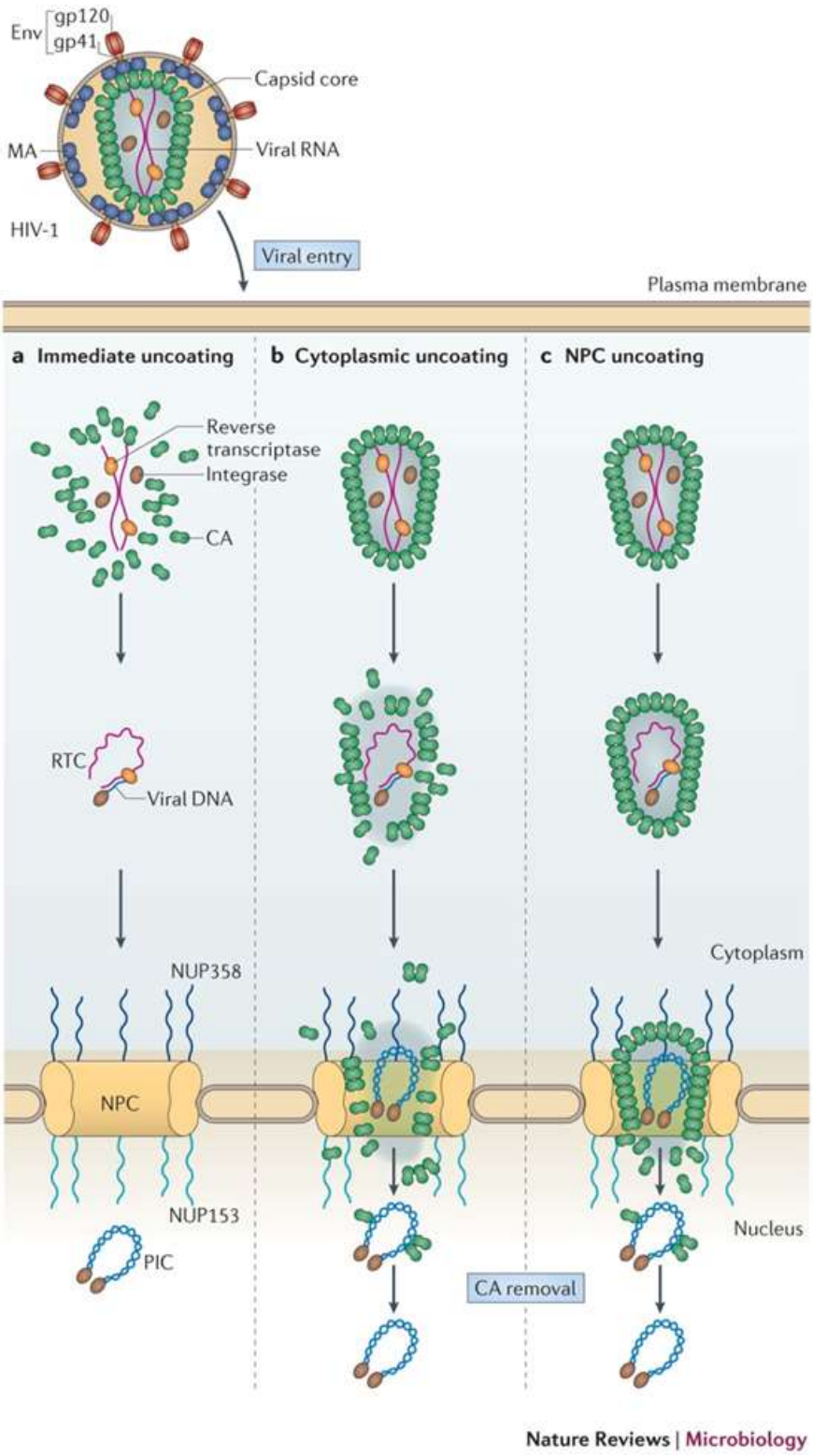


Figure 1-3. Three models of HIV-1 capsid uncoating. Adapted from [14].

lead to productive infection. This was achieved by fusing the fluorescent dsRed

protein to a capsid-binding protein (cyclophilin A, or cypA). CypA-dsRED binds to the capsid surface and fluorescent foci can be monitored over the course of infection [52]. While the timing, location, and purpose of capsid disassembly are still debated, an accumulation of experimental evidence suggests that HIV-1 capsid remains at least partially assembled and associated with the viral genome until the virus reaches the nuclear periphery.

A powerful line of evidence for the presence and requirement of assembled capsid surfaces during infection is the continued identification of capsid-binding host factors that target only assembled CA interfaces (reviewed in [17]). In fact, to date most well-studied capsid-binding host factors have either entirely lacked or displayed only partial recognition of CA monomers [5, 6, 12, 25, 26, 28]. These factors instead specifically target unique higher-order interfaces only found in the assembled CA lattice.

Capsid-binding host factors are a double-edged sword—they include both antiviral restriction factors and viral cofactors **(Figure 1-4)** (reviewed in [17]). Restriction factors are innate immune molecules that potently block infection after species-specific capsid recognition. Cofactors are host proteins hijacked by the virus and recruited to the capsid surface to promote infection. Importantly, capsid-interacting restriction factors and cofactors are both found either in the cytoplasm or associated with nuclear pores. Regardless of subcellular location, they all appear to demonstrate CA-lattice sensing ability. This highlights the importance of assembled capsid surfaces throughout the early viral life cycle.

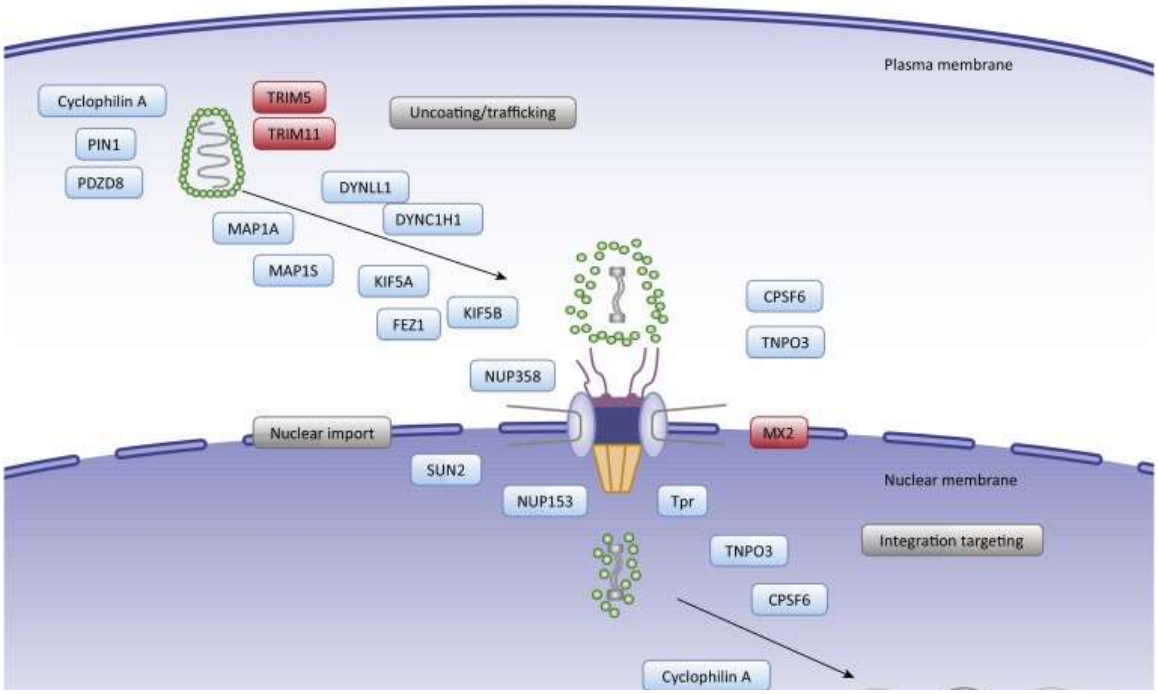


Figure 1-4. Overview of known HIV-1 capsid binding factors. Restriction factors are colored red, cofactors are colored blue. Adapted from [17]

1.3 CA-binding HIV-1 restriction factors

There are three well-studied restriction factors of HIV-1 that directly bind CA. The first, discovered in 2004, is TRIM5α [54]. TRIM5α is a potent species-specific restriction factor of numerous retroviruses including HIV-1. Upon capsid recognition, it causes premature capsid uncoating and inhibits reverse transcription [54, 55]. Like other tripartite motif-containing (TRIM) family members, it is a multi-domain E3 ligase [56]. It contains an N-terminal RING domain that is able to bind an E2 enzyme loaded with ubiquitin to catalyze formation of unlinked ubiquitin chains [57]. Adjacent to the RING domain is a B-box domain, which is a small zinc-binding domain critical for high-order TRIM5α oligomerization [58, 59]. A large coiled-coil domain enables TRIM5α dimerization [7]. Finally, a C-terminal PRY/SPRY domain has evolved as a capsid recognition

motif **(Figure 1-5A, top panel)** [55, 60]. It is believed that the SPRY domain is the only part of TRIM5α that directly recognizes CA.

TRIM5α has a spectacular CA binding mode. Individual SPRY domains bind CA weakly, but TRIM5α uses its oligomerization abilities to form a huge, hexagonal network over the capsid surface [25, 26, 61, 62]. This network allows multiple SPRY domains to simultaneously bind the capsid, creating an apparent high-affinity interaction through avidity **(Figure 1-5: B,C)**. The creation of a TRIM5α network triggers the activation of its RING domain, which catalyzes the production of unlinked K63 ubiquitin chains. These chains activate a signaling cascade that turns on the interferon response and creates an anti-viral state within the cell [63, 64].

Any particular TRIM5α gene is capable of recognizing capsids of numerous retroviral species and genera, despite low CA sequence homology between them [65, 66]. Human TRIM5α apparently does not recognize HIV-1 capsid, but it does restrict viruses like murine-leukemia virus and equine infectious anemia virus [67, 68]. *Rhesus macaque* TRIM5α, however, is a potent HIV-1 restriction factor and is used in most in vitro analysis of TRIM5α function [55, 69-71]. TRIM5α can also bind capsid lattices of drastically different curvature, including in vitro-assembled flat sheets to entirely spherical gamma-retroviral capsids [61, 72]. This flexible binding mode likely enables it to recognize a broader array of retroviruses compared to a hypothetical rigid/specific binding mode. The mechanism by which SPRY domains bind CA is unclear and is a major question in the field **(Figure 1-**

5D). SPRY does not appear to recognize CA hexamers or dimers, which suggests it targets a unique inter-hexamer surface [26, 62, 73].

Soon after the TRIM5α discovery, a related restriction factor, termed

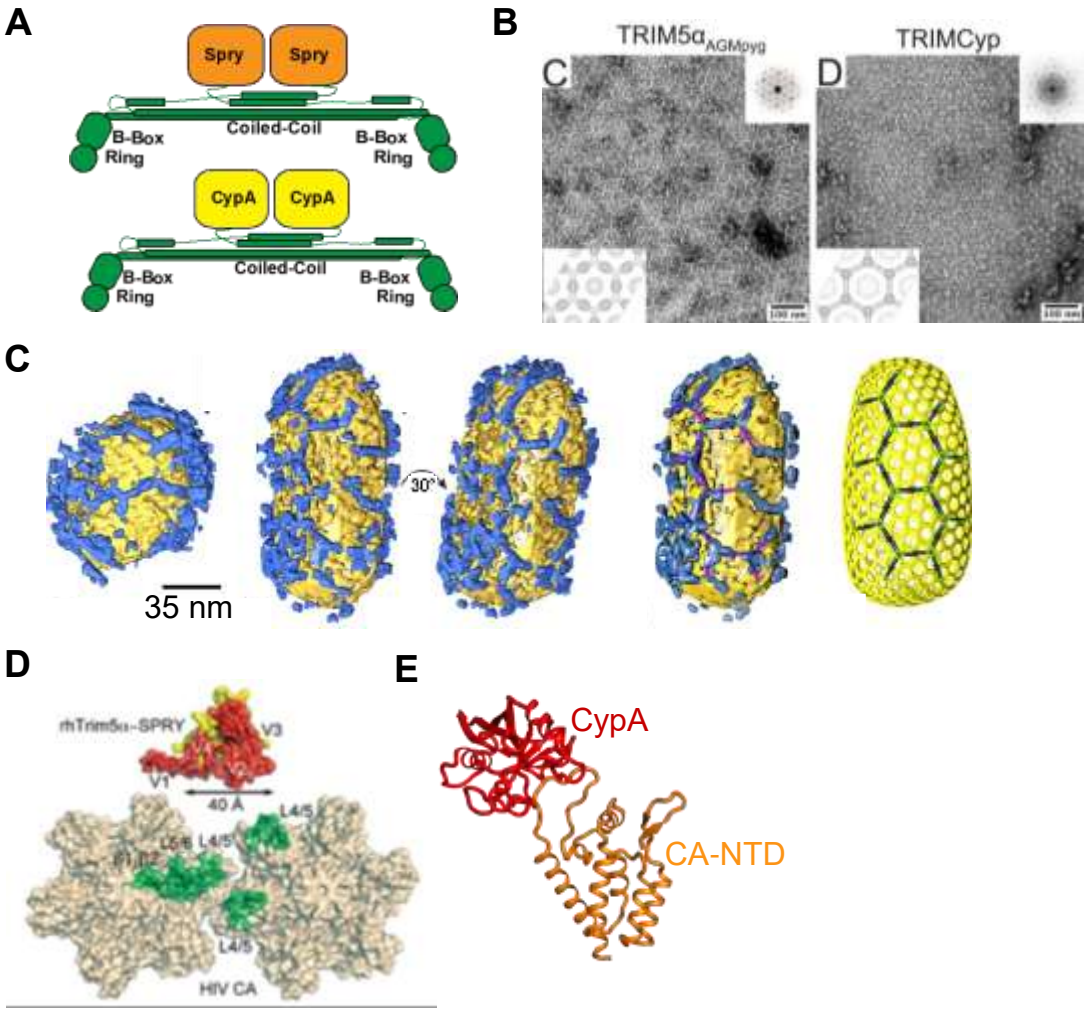


Figure 1-5. TRIM5α and TRIMCyp form hexagonal nets over the capsid surface. A) Domain architecture of TRIM5α (provided by Katie Digianantonio). In TRIMCyp, the SPRY domains are replaced with cyclophilin domains. B) Negative-stain electron micrographs demonstrating the lattice-forming abilities of TRIM5α and TRIMCyp in the absence of CA. C) Cryo-EM tomograms and model of TRIM5α bound to a stabilized capsid. D) Model of TRIM5α SPRY domain with a potential CA binding surface mapped in green on a di-hexamer interface. E) Model of cellular CypA bound to the CA NTD (PDB ID: 1AK4)[24]. (A)-(C) adapted from [25]. (D) adapted from [26].

TRIMCyp, was identified [74]. TRIMCyp is essentially identical to TRIM5α except that it contains a C-terminal cyclophilin domain in place of the PRY/SPRY domain **(Figure 1-5A, bottom panel)**. The TRIMCyp gene was created twice by retro-transposition events—once each in Old and New World monkeys [75-78]. Humans and most primates have a TRIM5α gene, yet only a select few primates have TRIMCyp genes. Like TRIM5α, TRIMCyp forms a hexagonal lattice over the capsid surface **(Figure 1-5B)**[25, 61].

The cyclophilin domain is responsible for CA recognition and binds CA in a similar manner to cellular cyclophilin A (CypA) **(Figure 1-5E)** [79-81]. CypA is a well-studied viral cofactor that binds CA and is co-packaged for efficient infection [24, 82-84]. TRIMCyp has a much higher CA affinity than TRIM5α because its cyclophilin domain binds CA with a higher affinity than SPRY domains [80, 81, 85]. It is unclear how the two cyclophilin domains of dimeric TRIMCyp cooperate and recognize two CA molecules in the CA lattice.

The most recently identified CA-binding restriction factor is MxB [86-88]. It is a dynamin-like GTPase that predominately resides at the cytoplasmic face of the nuclear envelope. Like TRIM5α and TRIMCyp, MxB is able to restrict a variety of simian retroviruses, but it appears to be less broadly effective than TRIM5α with retroviruses of different genera. It does not block reverse transcription but does reduce viral nuclear entry and integration [89-91]. The exact mechanism by which MxB restricts lentiviruses is still poorly understood. Unlike the related MxA, which restricts influenza, MxB restriction is not dependent on GTPase activity [86-89]. Instead, restriction depends on three things: a direct MxB-CA interaction,

a nuclear localization signal near the MxB N-terminus, and the ability of MxB to

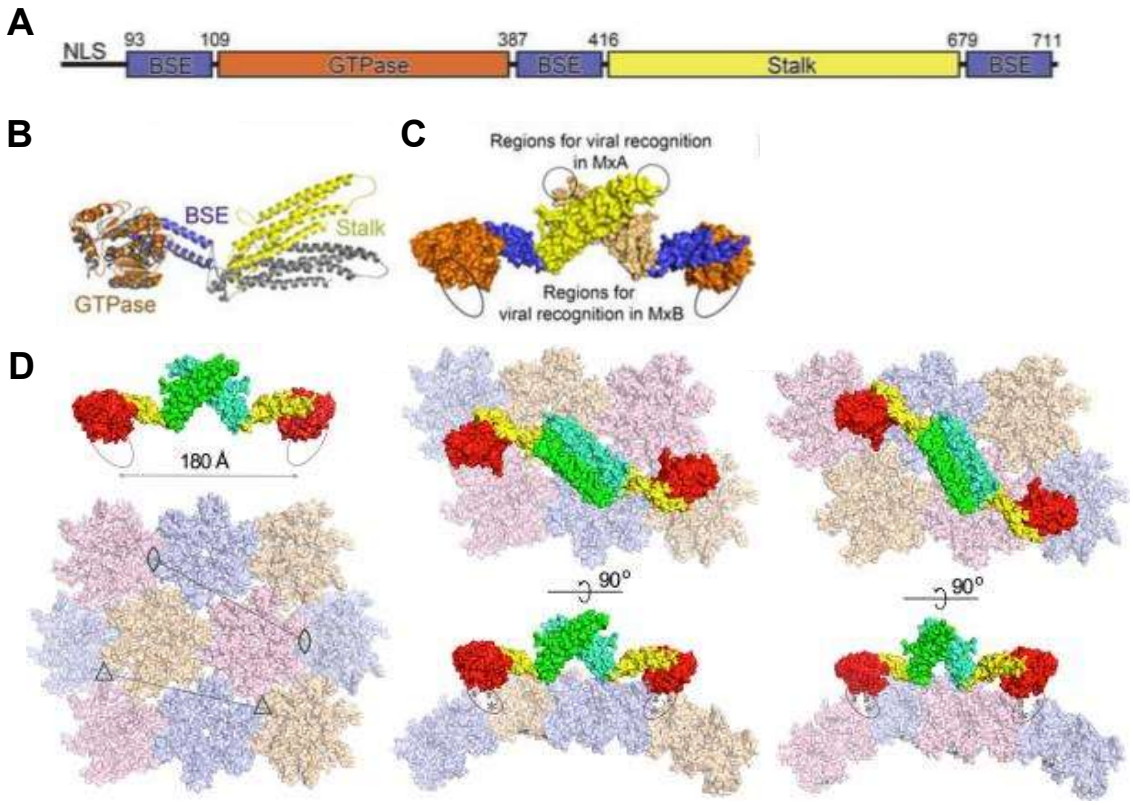


Figure 1-6. MxB is a dynamin-like GTPase that targets the capsid lattice. A) Domain map of human MxB. B) Overlay of MxA (gray) and MxB(color) aligned at their GTPase domains. C) Surface view of the MxB dimer. D) Models of MxB dimer spanning numerous hexamers to target either two- or three-fold CA surfaces. (A)-(C) adapted from [15]. (D) adapted from [12].

oligomerize into high-order structures [12, 92-98].

MxB is a multi-domain protein. It contains a GTPase domain, bundle signaling element (BSE), and stalk **(Figure 1-6A, B)**[12, 96]. The first 83 residues of MxB (which are predicted to be unstructured) are critical for HIV-1 restriction [90, 91]. Residues 11-13 (a triple arginine motif) are necessary for CA recognition, and this recognition ability can be transferred to human SAMHD1 protein if the N-terminal 25 residues of MxB are appended onto SAMHD1 [99,

100]. These data suggest that the MxB CA-binding motif is likely an unstructured peptide instead of a folded domain. We and others have shown that MxB, like TRIM5α, likely targets a unique, unknown inter-hexamer surface since no binding is observed with CA hexamers or dimers [12, 92, 101].

1.4 CA-binding HIV-1 cofactors

There are numerous CA-binding host factors that act as viral cofactors (reviewed in [17]). HIV-1 recruits these cofactors to its capsid surface to enable efficient infection. Cofactors have a variety of roles. They may aid in capsid stabilization/destabilization, hide the capsid surface from immune detection, enable trafficking towards nuclear pores, or contribute in many other proposed roles. The most well-studied viral cofactors are cyclophilin A (CypA), CPSF6, Nup153, and Nup358 [17]. Both CypA and Nup358 use their cyclophilin domains to bind the cyclophilin-binding loop on the CA surface [24, 27, 28, 84, 102]. Crystal and/or cryo-EM structures have been solved of these factors in complex with CA or CA tubes **(Figure 1-7: G-I)**. CPSF6 and Nup153 use an “FG” motif containing peptide to recognize a binding site between two CA monomers within a hexamer [5, 6]. High-resolution crystal structures of peptide fragments of CPSF6 and Nup153 in complex with either CA hexamers or WT CA have been solved **(Figure 1-7: A-F)**. Dozens of other, less understood, potential cofactors have been identified. Many of these, including fasciculation and elongation factor zeta 1 (FEZ1), may link the virus to the microtubule-trafficking network [17].

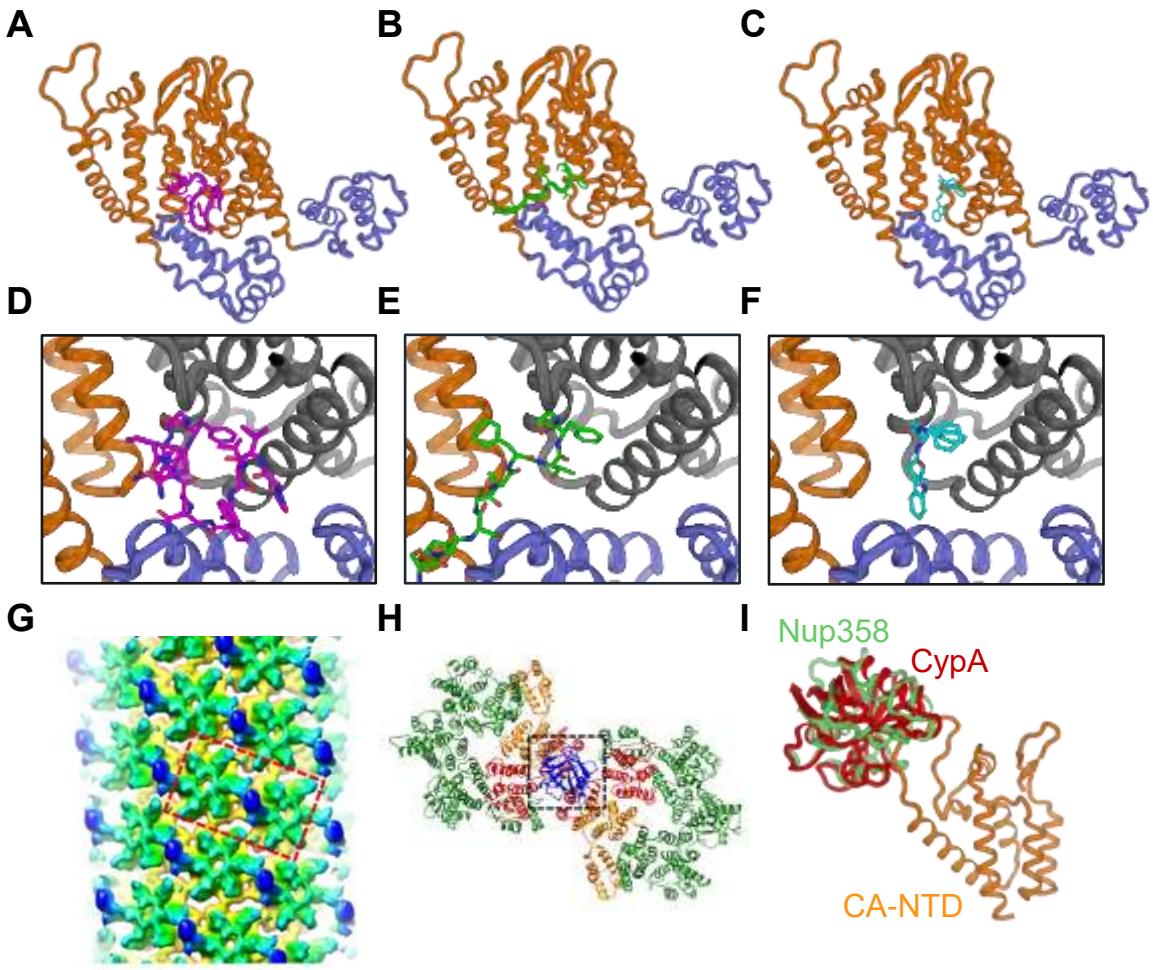


Figure 1-7. Solved structures of CA-binding cofactors. (A)-(C) are crystal structures of CPSF6 (PDB ID: 4U0A), Nup153 (PDB ID: 4U0C), and PF74 (PDB ID: 4U0E), respectively, bound to the NTD-CTD intra-hexamer surface. CPSF6 peptide represented in magenta, Nup153 peptide in green, and PF74 in cyan [5, 6]. (D)-(F) are closer views of the FG motif that drives the CA interaction. (G) Cryo-EM structure of CypA bound to CA tubes (CypA in blue). (H) Model of CypA bridging two hexamers, built from cryo-EM map (CypA in blue). (I) Crystal structure of CypA (PDB ID: 1AK4 [24]) and Nup358 (PDB ID: 4LQW [27]) bound to CA NTD. (G) and (H) adapted from [28].

FEZ1 is a kinesin-adaptor protein and may be the first definitive link between HIV-1 CA and the microtubule trafficking network [21]. It simultaneously binds cargo molecules and kinesin to promote intracellular trafficking [103]. When FEZ1 was knocked down in relevant HIV-1 targeted cell lines, myriad effects were observed: viral infection was significantly reduced, capsids failed to efficiently traffic to the nucleus, and viral DNA integration was decreased [21]. As such, FEZ1 was defined as a critical HIV-1 co-factor required for proper trafficking of viral cores. It was shown that FEZ1 (expressed in a mammalian cell-line) co-pelleted with in vitro assembled CA-NC tubes [21]. The motif in FEZ1 responsible for CA interaction was not identified, nor was the site on the CA surface targeted by FEZ1. FEZ1 is likely intrinsically unstructured, but definitive structural analyses of it have yet to be published [23, 104].

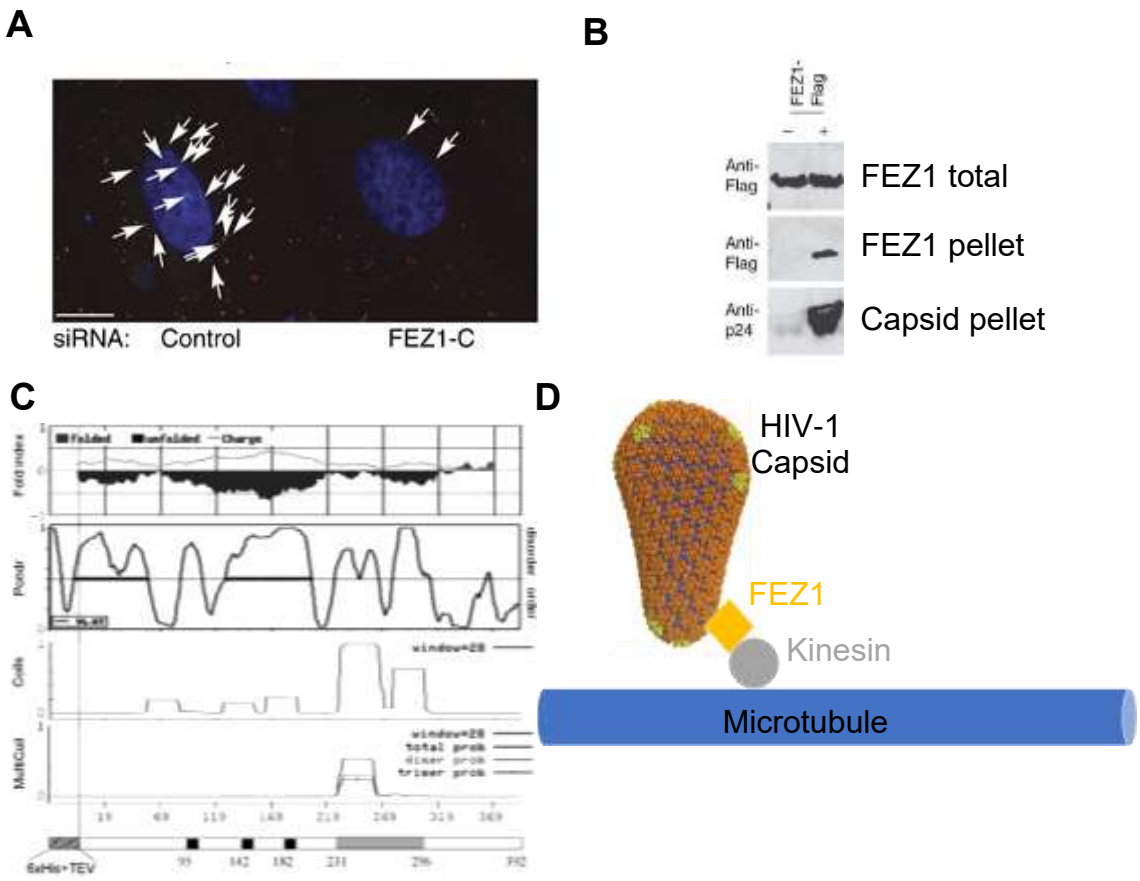


Figure 1-8. FEZ1 is CA-binding cofactor that aids in capsid trafficking. A) FEZ1 knockdown (right panel) reduces the amount of capsids observed in the nuclear periphery. B) Flag-tagged FEZ1 from cell lysate co-pellets with capsid tubes. C) FEZ1 is predicted to be largely unstructured. D) Model of FEZ1 linking the HIV-1 capsid to the microtubule trafficking machinery. (A) and (B) adapted from [21]. (C) adapted from [23]. Capsid in (D) adapted from [18].

1.5 CA behavior and oligomerization in vitro

Recombinant HIV-1 CA can be abundantly overexpressed and purified from *E. coli* using standard protein production methods. In the Xiong lab, 50-100+ mg of highly pure CA is recovered from 1.5L of cells grown in Terrific broth. CA behaves as a weak dimer in solution, with a reported dimerization dissociation constant of ~18 uM [105]. If incubated in high salt conditions (50 mM TRIS pH 8, 1M NaCl) CA spontaneously polymerizes into megadalton, insoluble tube-like structures **(Figure 1-9A)** [106]. If the DNA-binding nucleocapsid (NC) domain is present at the C-terminus of CA (CA-NC), tubes can be formed by the addition of DNA oligos at normal ionic strength **(Figure 1-9D)** [107]. These structures faithfully recreate the hexameric CA lattice observed in infectious particles [8]. Interestingly, in vitro assembled tubes only contain CA hexamers [1, 39]. Because of this, open-ended tube structures are primarily formed instead of closed, core-like structures. It is unclear why pentamer incorporation happens during infection, but is less common in vitro.

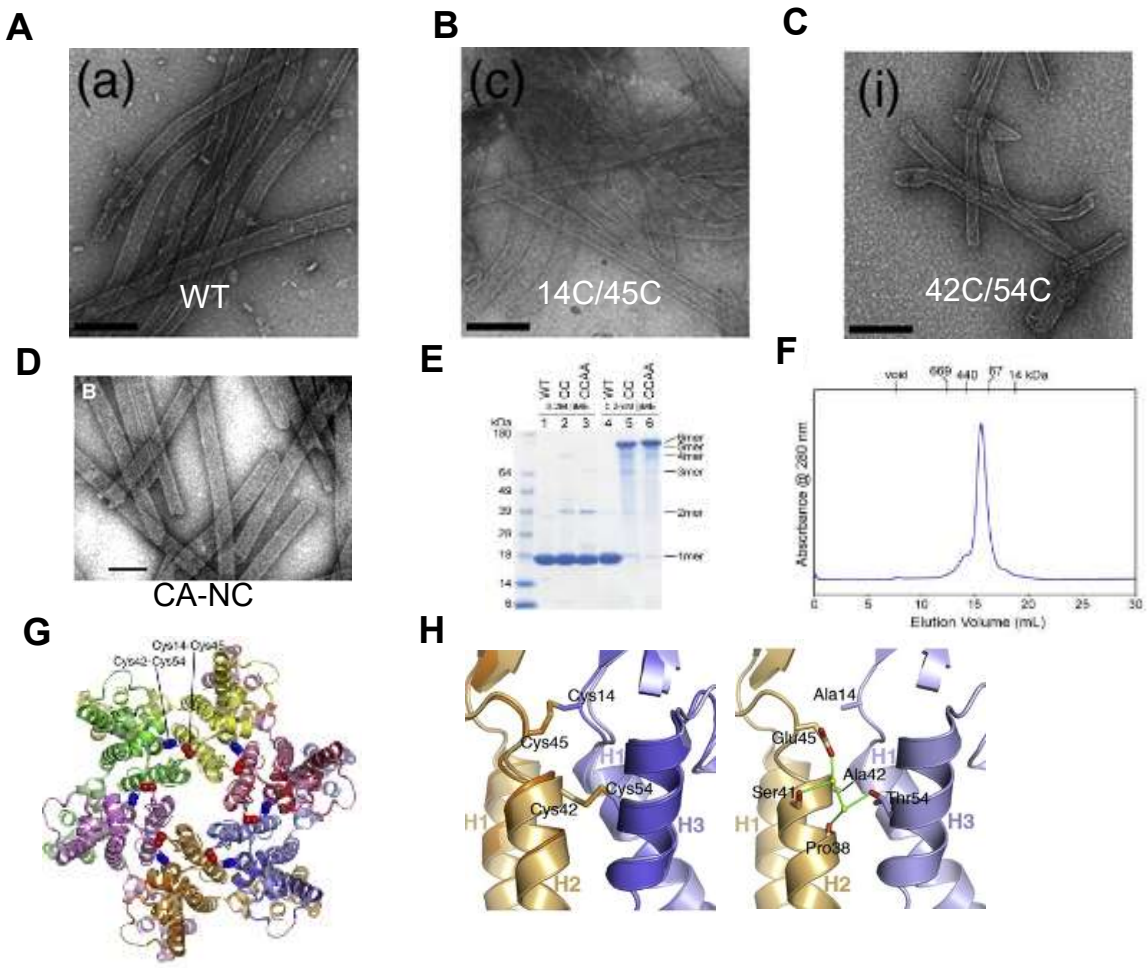


Figure 1-9. HIV-1 CA oligomerization and disulfide-bond stabilization in vitro. (A)-(D) shows representative negative-stain EM images of various in vitro assembled CA tubes (each tube-type is labeled appropriately). (E) is a non-reducing SDS-PAGE gel demonstrating hexameric disulfide-bond formation. (F) is a size-exclusion chromatogram of soluble, disulfide-bonded hexamers. (G) shows location of disulfide-bonds within a hexamer. (H) is a closer view of hexamer disulfide-bond orientation in stick view. (A)-(C) and (G)-(H) adapted from [19]. (D) adapted from [22]. (E) and (F) adapted from [16].

In 2009, engineered cysteine pairs that allow for disulfide-bond stabilization of CA hexamers (either 14C/45C or 42C/54C) and pentamers (21C/22C) were published [16, 19]. Hyper-stable CA tubes are formed with the addition of hexamer-stabilizing cysteine pairs **(Figure 1-9: B, C, E)**. Only disulfide-linked and CA-NC tubes are stable at low ionic strength, making them more valuable biochemical tools than native CA tubes (which require 1M NaCl for stability). If the CA dimer interface is concomitantly abolished (W184A/M185A), discrete, soluble disulfide-bonded hexamers and pentamers can be purified in high-yield **(Figure 1-9: F, G, H)**. This groundbreaking work led to the first high-resolution CA hexamer and pentamer crystal structures [16, 18]. Importantly, disulfide-bonded hexamers are not structurally different from native hexamers [1, 8].

When I began my thesis work in 2013, recombinant CA could be used in only a handful of oligomeric states. These included: 1) the weak native dimer, 2) soluble disulfide-linked hexamers/pentamers, and 3) insoluble native, CA-NC, or disulfide-linked tubes.

1.6 Aims and Scope

Extensive data supports the notion that the capsid is an expansive protein-docking platform whose finely-tuned structural integrity is critical for viral infection. Despite the advances in the field towards understanding the apo-capsid structure, the mechanisms by which host factors recognize the assembled capsid have remained poorly understood. This is largely due to a variety of technical constraints—many of which are inherent to self-assembling protein super-

structures. An experimental balance must be struck between insoluble, actively growing CA polymers and their unassembled CA building blocks. Unfortunately, both fully-assembled infectious capsid cores and in vitro assembled CA tubes can be highly heterogeneous, insoluble, and unstable. Additionally, many host factors have low capsid-binding affinity and/or form natural high-order oligomers themselves. These properties make it difficult to obtain the uniform host protein distribution along the surface of CA tubes and capsid cones required for high-resolution structural studies. Since most capsid-binding host factors do not recognize soluble, unassembled capsid building blocks—like native CA dimers or engineered disulfide-crosslinked hexamers—the vast arrays of solution-based biochemical and structural techniques cannot be used to explore capsid-host factor interactions.

To address these concerns, I developed novel techniques to trap and stabilize CA oligomers in lattice-like fragments that are intermediates between unassembled and fully assembled capsids. These lattice-intermediates enable analysis of elusive capsid-host factor binding events in solution—avoiding the challenges associated with capsid tubes and cones. Powerful solution-based biochemical assays and structural biology techniques, previously unworkable, can now be applied to many capsid-targeting factors and therapeutics. To achieve this, I incorporated numerous protein engineering techniques to create stable, discrete capsid “assemblies” that contain the critical lattice interfaces required for the binding of many host factors. These engineered lattice

assemblies range in size from 50 kDa to 1 MDa (or 2-42 CA molecules) and exhibit the same lattice interfaces found in the capsids of infectious virions.

The lattice assemblies allow for in-depth mechanistic studies of the known capsid-binding factors. I have used these assemblies to better define the unique capsid-binding modes of several well-studied, but poorly understood, restriction factors (TRIM5α, TRIMCyp, and MxB) and a recently discovered viral cofactor (FEZ1). Each of these factors demonstrated a unique capsid lattice-sensing ability that could only be defined with our novel CA assemblies. This work significantly expands our understanding of the diverse array of capsid-sensing motifs and targetable capsid surfaces. Besides mechanistic investigations, the advanced toolkit provides a rapid pipeline between first identification of a new factor and a thorough analysis of its binding mode, as well as the design and screening of specific capsid-targeting therapeutics to inhibit HIV-1 infectivity.

2 Trapping small CA oligomers in a capsid lattice-like architecture

2.1 Introduction

There are numerous known (and perhaps many undiscovered) HIV-1 capsid-binding host factors that target specific interfaces within CA hexamers. I will refer to these as hexamer-targeting factors. A more detailed description of known hexamer-targeting factors is in section 1.4.

There are numerous possible binding modes for any hexamer-targeting factor. For example, a factor could simply bind at each CA monomer without targeting a specific intra-hexamer surface. This is likely the case for the viral cofactor cyclophilin A (see section 1.4). Conversely, a factor could bind at intra-hexamer surfaces only. This is the case for FG motif-containing peptides (found in cofactors CPSF6 and Nup153), which bind in the NTD-CTD groove found between two CA monomers within a hexamer (see section 1.4). Intra-hexamer surfaces include any lateral interfaces between two or more adjacent CA monomers and the six-fold interface at the center of a hexamer. Possible binding modes are greatly expanded if oligomerization-induced avidity is considered. For example, a dimeric host factor could have two FG motif-containing peptides. Thus, each peptide could simultaneously bind at a distinct site within a hexamer, greatly increasing its affinity compared to monomeric FG motif-containing peptides.

To enable more thorough binding-mode mapping and structural biology studies of hexamer-targeting factors, I have successfully produced specific

“partial-hexamer” subassemblies. I created two partial-hexamer assemblies best described as 1/3- and 1/2-hexamers, which are composed of two and three CA monomers, respectively. The CA monomers are arranged with six-fold symmetric rotations—just as they are in complete CA hexamers. To stabilize and solubilize these partial-hexamers, I co-opted established CA intra-hexamer cysteine pairs (14C/45C and 42C/54C) and dimer interface mutations (184A/185A, or -AA in short) and used the previously established hexamer assembly procedure (see section 1.5).

Importantly, I was also able to give partial-hexamer subassemblies native oligomerization properties by returning their wild-type dimerization motifs (W184/M185). This enabled controlled oligomerization that captured elusive inter-hexamer surfaces that may be targeted by host factors like MxB and TRIM5α (discussed in chapter 5). I was able to build an assembly, termed hexamer-2-foldon, that contains both the native dimeric and trimeric inter-hexamer interfaces. This assembly overcomes major hurdles in the field and can be used alongside traditional disulfide-linked hexamers in host factor binding studies. Taken together, these novel assemblies enable a rapid pipeline between first identification of a CA-binding host factor and a thorough analysis of its binding mode.

2.2 Results

2.2.1 Design and purification of the 1/3-hexamer subassembly

To produce 1/3-hexamers, I separately purified two monomeric CA constructs ($CA_{14C/184A/185A}$ and $CA_{45C/184A/185)}$ each containing a single cysteine

mutation. Each construct contains only half of what is needed for assembling the complete cross-linked hexamer **(Figure 2-1A)**. Disulfide-linked higher-order oligomers were not expected if either construct was incubated alone in hexamer assembly conditions, as each construct alone does not contain a complementary cysteine pair. As designed, however, a disulfide-bonded CA dimer was formed when I assembled $CA_{14C/184A/185A}$ and $CA_{45C/184A/185A}$ together in a 1:1 ratio. Its 50 kDa molecular weight corresponded to that of a 1/3-hexamer **(Figure 2-1B)**. The assembly efficiency of 1/3-hexamers was high; roughly 80-90% of assembled CA was in a disulfide-bonded dimer state. The remaining CA appeared as a monomer.

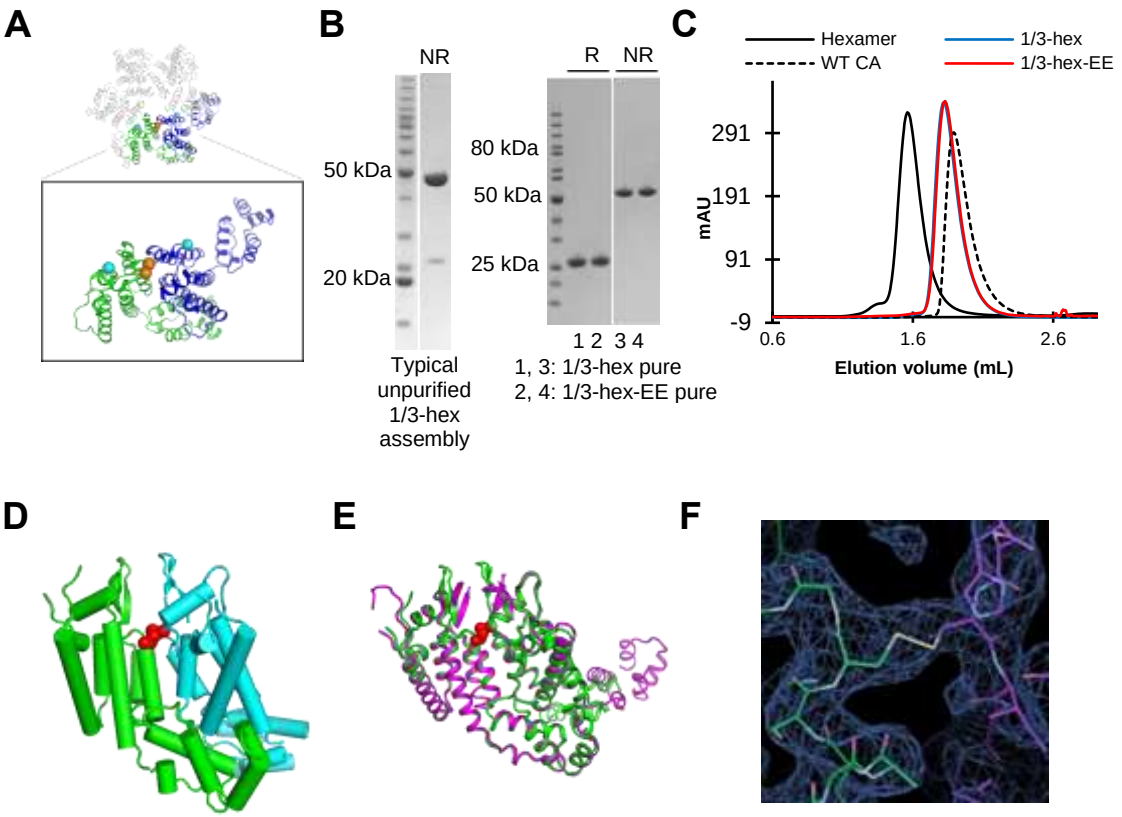


Figure 2-1. 1/3-hexamer design, purification, and structure. A) Cartoon model showing the design of 1/3-hexamers, with engineered cysteines (14C/45C in orange spheres) and anti-polymerization mutations (42E/54E, cyan spheres) (based on PDB ID: 3H47 [16]). B) SDS-PAGE gels showing purity of 1/3-hexamer assemblies. Left panel: 1/3-hexamers before purification. Right panel: 1/3-hexamers and 1/3-hexamers-EE after purification, non-reducing and reducing conditions. C) Elution profile of 1/3-hexamers compared to hexamers and WT CA dimers on S200 5/150. D) Structure of 1/3-hex-EE with disulfide side-chains shown as red spheres. The green molecule contains the 45C mutation, and the cyan molecule contains the 14C mutation. E) Alignment of 1/3-hexamer-EE crystal structure (green) and disulfide-bonded hexamer fragment (magenta)(PDB 3H47). F) Clear electron density in 2Fo-Fc map demonstrating 14C/45C disulfide bond within 1/3-hexamer (contour level 1 sigma).

To ensure that these 1/3-hexamers remain discrete and do not proceed to assemble into natural hexamers, I further introduced A42E and T54E mutations at their exposed intra-hexamer surfaces (constructs used: $CA_{14C/42E/184A/185A}$ and $CA_{45C/54E/184A/185A,}$ assemblies with -EE suffix contain 42E/54E mutations). These mutations were designed to cause charge-charge repulsion if two 1/3-hexamer

Assembly name	CA construct #1	CA construct #2
1/3-hexamer	CA-14C/184A/185A	CA-45C/184A/185A
1/3-hexamer-EE	CA-14C/42E/184A/185A	CA-45C/54E/184A/185A
1/3-hexamer-EE-mpro	CA-14C/42E-mpro-184A/185A	CA-45C/54E/184A/185A

Table 2-1. Constructs used in 1/3-hexamer assembly

subunits come in close, hexamer-like proximity. Both 1/3-hexamers and 1/3-hexamers-EE demonstrated the appropriate ~50 kDa solution molecular weight as observed by size exclusion chromatography (SEC) **(Figure 2-1C)**. It appears the 42E/54E mutations are not required to keep 1/3-hexamers discrete at the concentrations tested. **Table 2-1** provides the specific constructs used to produce 1/3-hexamer assemblies.

To validate that 1/3-hexamer has the correct CA lattice architecture, I solved a crystal structure of 1/3-hexamer-EE at 3.5 Å resolution **(Figure 2-1D)**. Crystallization statistics are shown in **Table 2-2.** Final crystallization statistics will be provided in a soon to be published manuscript.

Data Collection	
Wavelength (Å)	0.9798
Space Group	P212121
Cell Dimensions	
a, b, c (Å)	74.80, 97.43, 150.89
α, β, γ (°)	90, 90, 90
Molecules/asymmetric unit	2 (4 CA monomers)
Resolution (Å)	50.0-3.40 (3.49-3.40)
R_{merge}	0.14 (0.58)
$I/\sigma I$	6.3 (1.7)
Completeness (%)	98.1 (99.1)
Redundancy	2.1 (2.1)
Unique reflections	15753
Refinement	
Number of nonhydrogen atoms	6405
R_{work}/R_{free} (%)	21.1/27.5 (30.2/40.4)
Average B factor	76
Root mean-squared deviation (rmsd)	
Bond lengths (Å)	0.015
Bond angles (°)	1.8
Ramachandran analysis	
Preferred regions (%)	96.6
Allowed regions (%)	3.14
Outliers (%)	0.25

Table 2-2. Data collection and refinement statistics for 1/3-hexamer-EE. Statistics in parentheses indicate those for the highest resolution shell. These statistics are current as of Sept. 2018 but may change upon publication.

As designed, two disulfide-bonded CA monomers arranged as a discrete 1/3-hexamer formed the asymmetric unit. Clear electron density for the single disulfide bond was observed **(Figure 2-1F)**. This 1/3-hexamer is closely superimposable onto prior disulfide-linked hexamer structures with an overall root-mean-square deviation (RMSD) of 1.5 Å **(Figure 2-1E)**. The only deviation is in the positioning of one flexible CTD that does not participate in intra-hexamer contacts. This domain is normally stabilized in a complete hexamer by NTD-CTD contacts with its neighboring CA but is not involved in such interactions in 1/3-hexamers. 1/3-hexamer subunits did not form complete hexamers even with symmetry-related molecules. This is presumably due to the designed hexamer-preventing 42E/54E mutations, which did not otherwise alter the CA monomer structure.

2.2.2 Design and purification of the ½-hexamer subassembly

I produced 1/2-hexamer assemblies using an extension of the approach used to make 1/3-hexamers. I incubated a 1:1:1 mixture of three heterologous CA constructs ($CA_{14C/184A/185A}$, $CA_{45C/54C/184A/185A}$, and $CA_{42C/184A/185A}$) built from a combination of 14C/45C and 42C/54C cysteine pairs **(Figure 2-2A)**. Again, I did not expect individual constructs to form disulfide-bonded species when incubated alone in hexamer assembly conditions. When mixed, however, I observed a range of disulfide bonded CA species. The primary product was a disulfide-bonded CA 3-mer that corresponds to the molecular weight of a 1/2-hexamer **(Figure 2-2B)**.

In the design, a disulfide-bonded CA 3-mer could only occur if the three disparate CA constructs made intra-hexamer contacts in a precise order--$CA_{45C/54C/184A/185A}$ in the middle position, linked on one face to $CA_{14C/184A/185A}$ by the 14C/45C disulfide pair, and on the other face to $CA_{42C/184A/185A}$ by the 42C/54C disulfide pair. There were numerous unproductive and partially

Assembly name	CA construct #1	CA construct #2	CA construct #3
½-hexamer	14C/184A/185A	45C/54C/ 184A/185A	42C/184A/185A
½-hexamer-EE	14C/42E/ 184A/185A	45C/54C/ 184A/185A	42C/54E/ 184A/185A
½-hexamer-EE-mpro	14C/42E-mpro-184A/185A	45C/54C/ 184A/185A	42C/54E/ 184A/185A

Table 2-3. Constructs used in ½-hexamer assembly

productive reactions, which made 1/2-hexamer assembly less efficient than the simpler 1/3-hexamer assembly. Despite this complexity, we readily purified both 1/2-hexamers and 1/2-hexamers-EE (the latter using $CA_{14C/42E/184A/185A}$, $CA_{45C/54C/184A/185A}$, and $CA_{42C/54E/184A/185A}$) to homogeneity, and they both displayed the correct 1/2-hexamer solution molecular weight (~75 kDa) when analyzed by SEC **(Figure 2-2C)** and SEC-MALS (data not shown). A summary of constructs used in ½-hexamer assembly is described in **Table 2-3**.

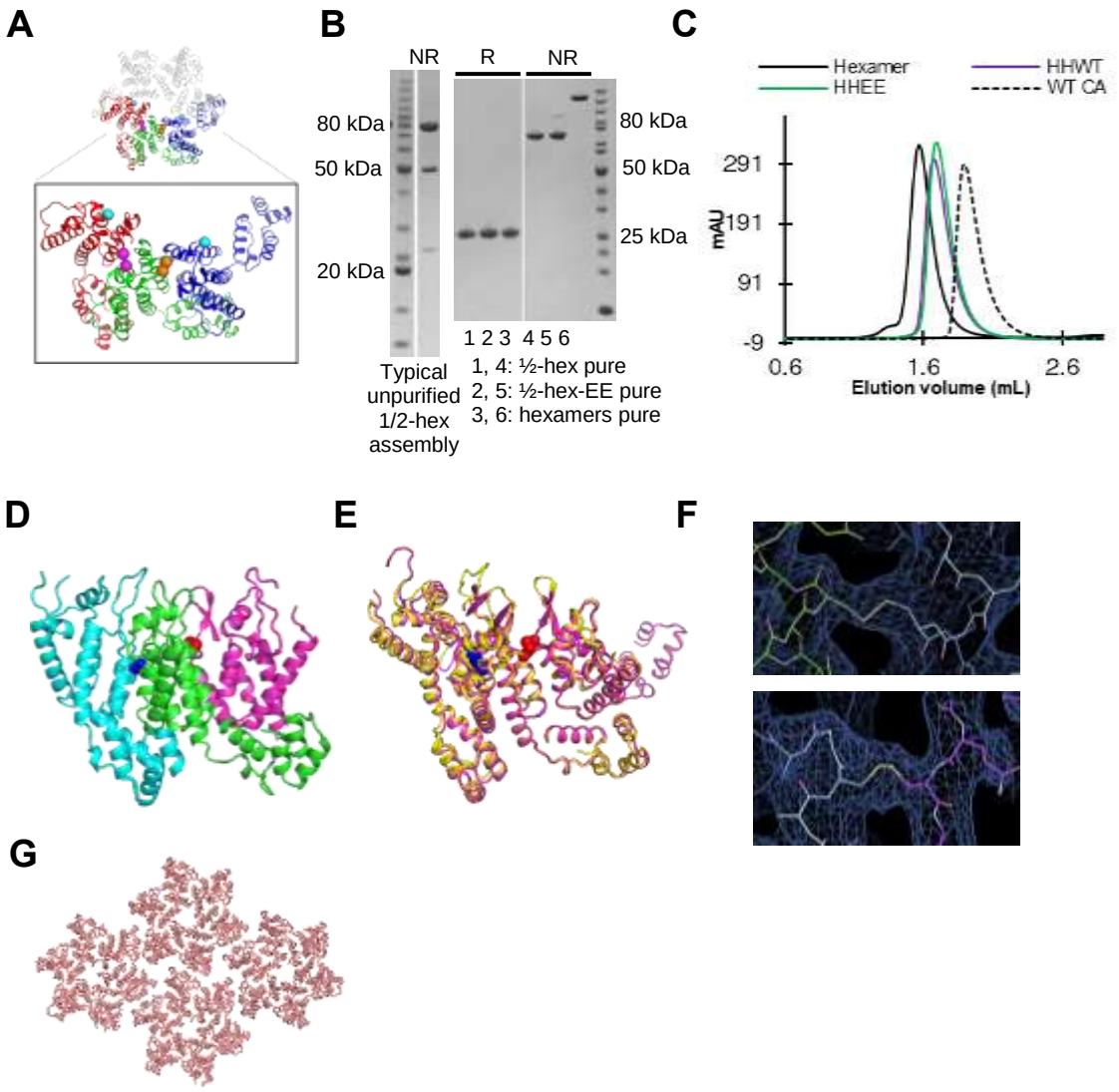


Figure 2-2. 1/2-hexamer design, purification, and structure. A) Cartoon model showing the design of 1/2-hexamers, with engineered cysteines (14C/45C, orange spheres, 42C/54C magenta spheres) and anti-polymerization mutations (42E/54E, cyan spheres) (based on PDB ID: 3H47 [16]). B) SDS-PAGE gels showing purity of 1/3-hexamer assemblies. Left panel: 1/3-hexamers before purification. Right panel: 1/3-hexamers and 1/3-hexamers-EE after purification, non-reducing and reducing. C) Elution profile of 1/3-hexamers compared to hexamers and WT CA dimers on S200 5/150. D) Structure of ½-hexamer-ΔCTD with 42C/54C disulfide bond side-chains shown as blue spheres, 14C/45C shown as red spheres. The cyan CA contains the 45C mutation, the green CA contains both 45C and 54C mutations, and the magenta CA contains the 14C mutation. E) Alignment of 1/2-hexamer-EE-ΔCTD crystal structure (yellow) and disulfide-bonded hexamer fragment (magenta) (PDB 3H47). F) Clear electron density in 2Fo-Fc map demonstrating both 42C/54C and 14C/45C disulfide bond within ½-hex-EE-ΔCTD (contour level 1 sigma). G) 1/2-hexamer crystal packing creates flat lattice similar to native capsid surface

I determined the crystal structures of 1/2-hexamer and 1/2-hexamerEE-ΔCTD (The ΔCTD suffix indicates that the flexible CA CTD was specifically removed proteolytically for crystallization by cleavage with the mpro enzyme). As designed, 1/2-hexamer-EE-ΔCTD crystallized as a discrete 1/2-hexamer and did not form complete hexamers with symmetry-related molecules **(Figure 2-2D)**. Its structure and ½-hexamer architecture overlaid closely with prior disulfide-bonded hexamer structures with an overall RMSD of 1.2 Å **(Figure 2-2E)**. Clear electron density was observed for both of the 14C/45C and 42C/54C disulfide bonds **(Figure 2-2F)**. The 42E/54E mutations and proteolytic removal of a CA CTD did not affect the structure. The crystallization statistics are summarized in **Table 2-4.** Final crystallization statistics will appear in a soon to be published manuscript.

Data Collection	**½-hexamer-EE-ΔCTD**	**½-hexamer**
Wavelength (Å)	0.9798	0.9792
Space Group	P212121	R3
Cell Dimensions		
a, b, c (Å)	65.61, 84.10, 248.54	89.93, 89.93, 173.75
α, β, γ (°)	90, 90, 90	90, 90, 120
Molecules/asymmetric unit	2	2 CA monomers
Resolution (Å)	50.0-3.40 (3.49-3.40)	71.1-3.0 (3.1-3.0)
R_{merge}	0.13 (1.1)	0.047 (1.42)
$I/\sigma I$	5.7 (1.0)	13.6 (0.7)
Completeness (%)	99.0 (99.1)	98.9 (98.1)
Redundancy	3.1 (3.1)	3.5 (3.3)
Unique reflections	19697	10397 (1053)
Refinement		
Number of nonhydrogen atoms	8369	3358
R_{work}/R_{free} (%)	22.5/27.5 (37.8/38.4)	25.6/27.1 (44.2/49.0)
Average B factor	148	141.9
Root mean-squared deviation (rmsd)		
Bond lengths (Å)	0.012	0.006
Bond angles (°)	1.6	1.09
Ramachandran analysis		
Preferred regions (%)	96.6	98.4
Allowed regions (%)	3.0	0.70
Outliers (%)	0.37	0.94

Table 2-4. Data collection and refinement statistics for ½-hexamer constructs. Statistics in parentheses indicate those for the highest resolution shell. These statistics are current as of July 2018 but may change upon publication.

While the 1/2-hexamer without 42E/54E mutations behaved as a discrete half-hexamer in solution, it crystallized as a complete hexamer together with symmetry-related molecules. It formed a 2-dimensional hexamer lattice almost identical to that observed in prior disulfide-linked hexamer and native CA structures **(Figure 2-2G)**. This confirms that the designed 1/2-hexamer generates

the overall correct capsid architecture. On the basis of these data, we conclude that 1/3- and 1/2-hexamer assemblies can reliably complement complete hexamers host factor capsid-binding studies. Although no difference was observed in solution with and without the hexamer-preventing 42E/54E mutations, it is potentially better to use these mutations to ensure discrete assemblies in various conditions.

2.2.3 Design and purification of naturally oligomeric 1/3- and ½-hexamer assemblies

Both 1/3- and ½-hexamer assembly and purification were relatively efficient reactions. However, these assemblies lack the native dimerization that creates inter-hexamer surfaces potentially targeted by MxB or TRIM5α (see section 1.3). Therefore, I attempted to return native dimerization (i.e. W184/M185 instead of 184A/185A) to specific CA molecules within each assembly.

I purified two different “dimeric” 1/3-hexamer assemblies. Since a 1/3-hexamer is composed of two CA molecules, and a dimeric 1/3-hexamer would

Assembly name	CA construct #1	CA construct #2	CA construct #3
Tetramer-1	14C/184A/185A	45C	
Tetramer-2	14C	45C/184A/185A	
Tetramer-1-EE	14C/42E/ 184A/185A	45C/54E	
Tetramer-2-EE	14C/42E	45C/54E/ 184A/185A	
Butterfly	14C/42E/ 184A/185A	45C/54C	42C/54E/ 184A/185A
Triskelion/9-mer	14C/42E	45C/54C	42C/54E/ 184A/185A

Table 2-5. Basic constructs used to form oligomeric 1/3- and ½-hexamer assemblies

contain four CA molecules, I will refer to these assemblies as either “tetramer-1” **(Figure 2-3A)** or “tetramer-2” **(Figure 2-3B). Table 2-5** summarizes the constructs used to make tetramer-1 and tetramer-2, as well as their -EE versions.

Both tetramers assembled and purified efficiently, similar to standard 1/3-hexamer subunits. However, they did not form stable tetramers. On SEC, the tetramers generally eluted as asymmetric peaks at a molecular weight between 50 kDa 1/3-hexamers and 100 kDa tetramers **(Figure 2-3: C, D)**. Only at high concentrations did they appear as complete tetramers. This intermediate molecular weight was supported by SEC-MALS (data not shown).

Although I formed a complex of tetramer-1 with CypA and solved its crystal structure, the native dimerization interface within tetramers was broken. Data collection and refinement statistics are in **Table 2-6.** We observed a discrete 1/3-hexamer with a CypA molecule bound at each CypA binding loop **(Figure 2-3: E, F)**. We had hoped to observe CypA bridging both halves of the tetramer, as had been shown in a published cryo-EM structure of CypA bound to CA tubes **(Figure 1: G, H)** (see section 1.4). Our structure did not provide any novel biological insights. The weak dimerization of the tetramer likely made it too unstable and flexible to crystallize readily. This limits the use of the tetramers for structural and binding studies of capsid-binding host factors.

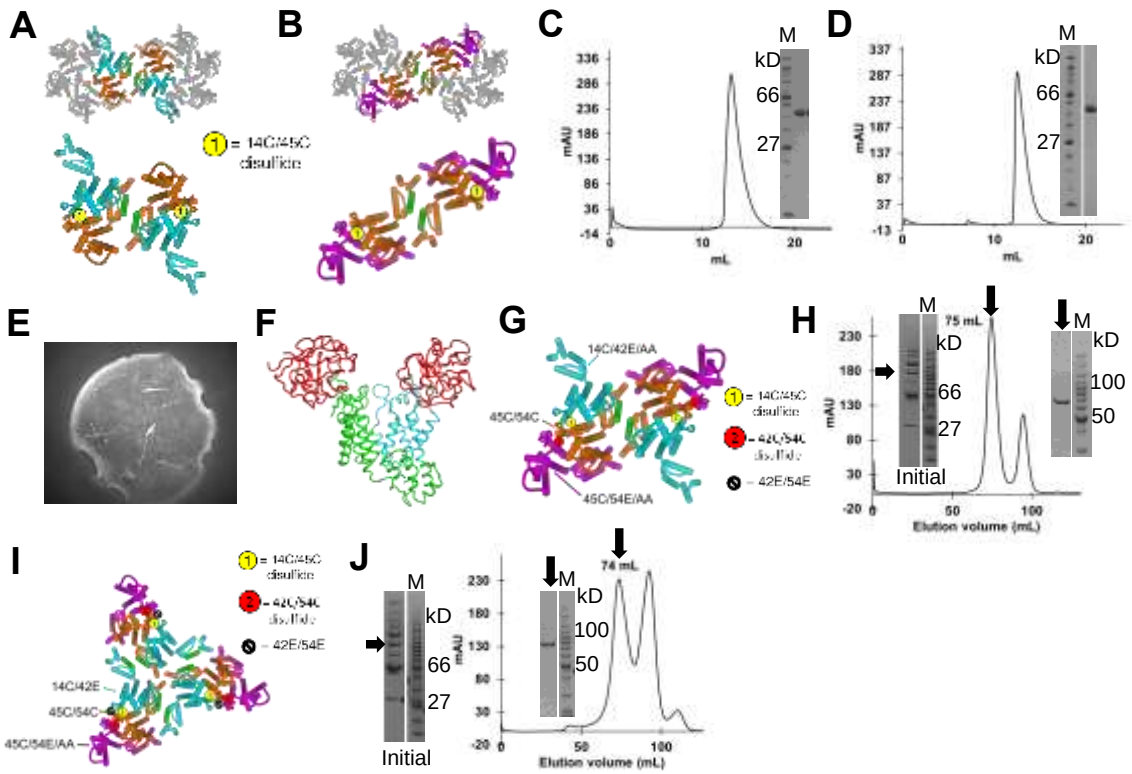


Figure 2-3. Design and purification of oligomeric 1/3- and ½-hexamer assemblies. A) Cartoon model of tetramer-1 design. Green helices contain native dimerization interfaces. B) Cartoon model of tetramer-2 design. Green helices contain native dimerization interfaces. C) SEC chromatogram and SDS PAGE of purified tetramer-1. D) SEC chromatogram and SDS PAGE of purified tetramer-2. E) Crystals under UV light of tetramer-1-CypA complex. F) Crystal structure of tetramer-1-CypA complex. CypA is colored red, 45C-containing CA molecule is green, and 14C-containing CA molecule is cyan. G) Design of the capsid butterfly. H) Initial SDS PAGE and final SDS-PAGE/SEC chromatogram of capsid butterfly. Black arrows indicate the desired CA species and/or the fraction of the SEC run used in SDS-PAGE analysis. I) Design of the capsid 9-mer or triskelion. J) Initial SDS-PAGE and final SDS-PAGE/SEC chromatogram of capsid 9-mer. All cartoon models from PDB ID: 4XFX [4].

I also attempted to create a complete inter-hexamer dimer interface between two ½-hexamers by placing a native dimer interface in their central CA position **(Figure 2-3G)**. This would hypothetically allow the dimerization of the 75 kDa ½-hexamer into a 150 kDa assembly which resembles a butterfly in appearance, inspiring its name. The constructs used in butterfly assembly are summarized in **Table 2-5**. Unfortunately, the efficiency of disulfide-formation was substantially reduced than standard ½-hexamers. A CA 3-mer was not the major product of the reaction. There was an increase in disulfide-bonded CA 2-, 4-, and 5-mers. The small amount of CA 3-mers did not purify readily and eluted as only a 75-kDa ½-hexamer species **(Figure 2-3H)**. This construct was abandoned after several tries by myself and Wei Wang (a postdoctoral fellow in the lab).

I also attempted to introduce two native dimerization interfaces into ½-hexamers. This hypothetically would enable the trimerization of 75 kDa ½-hexamers into 225 kDa trimers that contain both the complete dimer and trimer

Data Collection	
Wavelength (Å)	1.1
Space Group	P41212
Cell Dimensions	
a, b, c (Å)	115.1, 115.1, 308.8
α, β, γ (°)	90, 90, 90
Molecules/asymmetric unit	4 CA monomers
Resolution (Å)	49.2-3.2 (3.3-3.2)
R_{merge}	0.184 (2.87)
$I/\sigma I$	9.5 (0.9)
Completeness (%)	98.2 (93.6)
Redundancy	8.5 (8.7)
Unique reflections	35007 (3406)
Refinement	
Number of nonhydrogen atoms	10514
R_{work}/R_{free} (%)	27.6/34.9 (39.9/45.0)
Average B factor	125.9
Root mean-squared deviation (rmsd)	
Bond lengths (Å)	0.020
Bond angles (°)	1.92
Ramachandran analysis	
Preferred regions (%)	74.5
Allowed regions (%)	17.0
Outliers (%)	8.5

Table 2-6. Data collection and refinement statistics for tetramer-1-CypA. Statistics in parentheses indicate those for the highest resolution shell. These statistics are current as of July 2018 but may change. These statistics were generated with minimal and incomplete refinement.

inter-hexamer surfaces **(Figure 2-3I)**. This construct was termed the CA 9-mer or the triskelion. Unfortunately, it suffered the same problems as the butterfly. Disulfide-formation was inefficient, and the samples were heterogeneous. I did succeed in generating a sufficient quantity for SEC, unfortunately, the protein

eluted mostly as a 75-kDa ½-hexamer, not the 225-kDa assembly we had hoped **(Figure 2-3J)**. This construct was also abandoned after some effort.

2.2.4 Design and purification of hexamer-2-foldon

I sought to develop a soluble capsid platform containing the native two- and three-fold CTD-CTD oligomerization surfaces between CA hexamers. Disulfide-linked CA hexamers and partial-hexamers only remain discrete because their inter-hexamer dimerization motifs are effectively abolished by mutations (W184A/M185A). I wanted to make an alternative hexamer—one that encompasses the six CA molecules immediately surrounding the CA three-fold interface. My alternative hexamer is formed by three pairs of CA dimers arranged with three-fold symmetry, also known as a CA "trimer of dimers." Importantly, a CA trimer of dimers contains all native inter-hexamer lattice interfaces. We envisioned the CA trimer of dimers as a powerful counterpart to traditional hexamers when analyzing capsid-host interactions. For this reason, and for simplicity, I will refer to it as "hexamer-2". I devised a strategy to combine engineered disulfides, native CA dimerization motifs, and solubilizing mutations to create a functional hexamer-2 CA assembly **(Figure 2-4A)**.

For design purposes it is simpler to think of hexamer-2 as a "trimer of 1/3-hexamers", where a "1/3-hexamer" is an adjacent pair of two of the six molecules in a traditional hexamer. I built hexamer-2 off the previously described 1/3-hexamer-EE (see **section 2.2.1**), which is stabilized by a single-disulfide bond between the two CA subunits and contains 42E/54E mutations to prevent hexamer formation. Both of its CTD domains are mutated to prevent

dimerization. I hypothesized that returning native CTD dimerization to both CA positions in 1/3-hexamer-EE would enable its natural trimerization into the desired CA “trimer of dimers”.

I assembled and purified 1/3-hexamer-EE with two native dimerization motifs to form native hexamer-2 **(Figure 2B)**. All constructs used are described in **Table 2-7.** I analyzed its solution molecular weight using SEC **(Figure 2-4C)** and validated these results with SEC-MALS (data not shown). Native hexamer-2 displayed a broad elution profile and a molecular weight ranging between 80-100

Assembly name	CA construct #1	CA construct #2
Native hexamer-2	14C/42E	45C/54E
Hexamer-2-foldon	14C/42E	45C/54E(1-226)-foldon
Hexamer-2-foldon(204D/221*)	14C/42E/204D(1-221)	45C/54E(1-226)-foldon

Table 2-7. Basic constructs used in hexamer-2 and hexamer-2foldon

kDa. The observed weight is intermediate to the weights of 1/3-hexamer building blocks (~50 kDa) and complete hexamer-2 trimers (~150 kDa). This suggests the hexamer-2 inter-hexamer interfaces are engaged, but too weak to provide stable oligomerization at the concentrations used for analysis. This was surprising since the “trimer of dimers” CA structure has been computationally modeled to be the most stable lattice interface. It was further proposed as the key nucleating structure during capsid maturation [108-110].

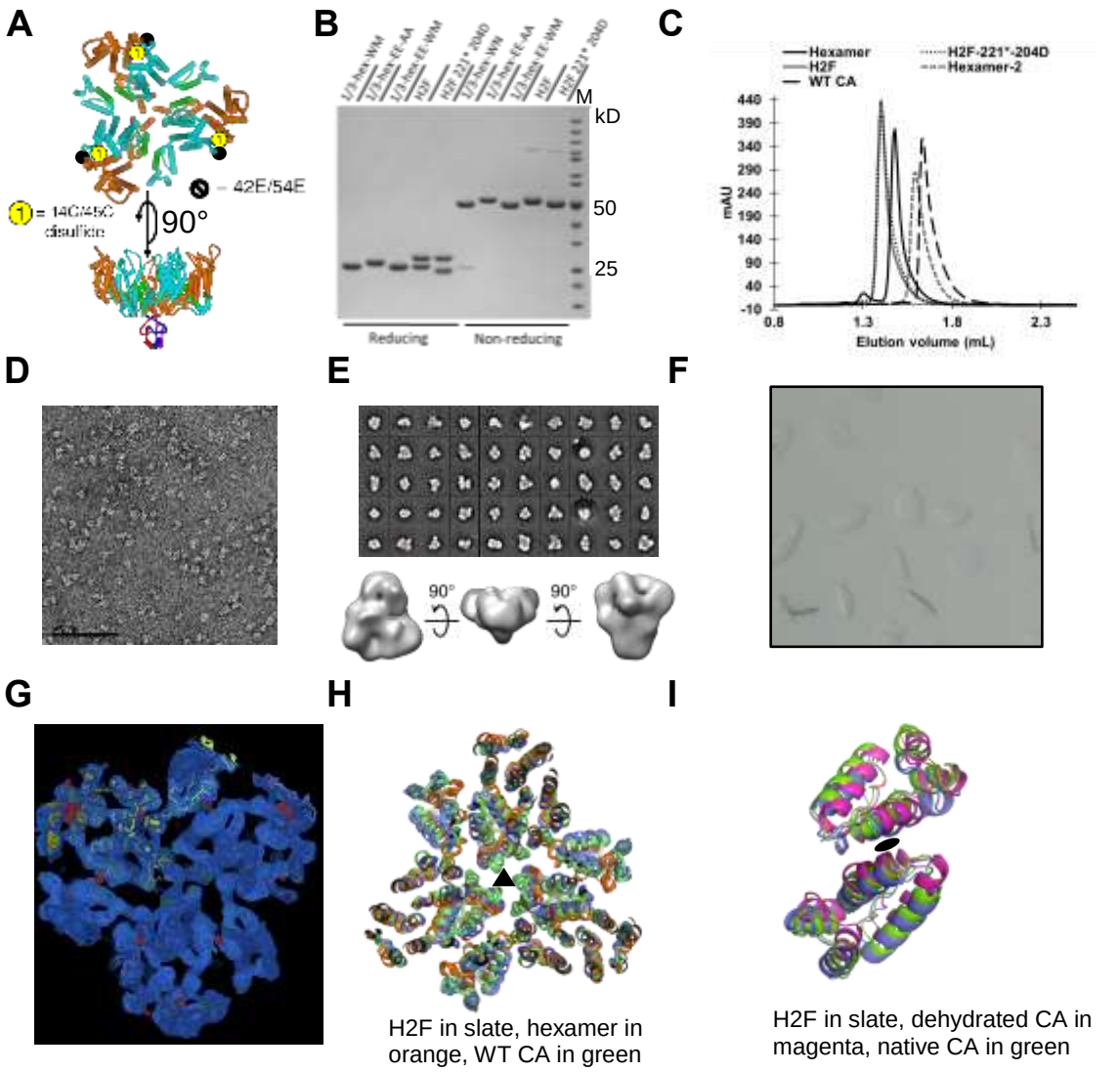


Figure 2-4. Design and purification of hexamer-2-foldon. A) Cartoon model demonstrating foldon (PDB ID: 4NCU [13]) C-terminally fused to hexamer-2 (PDB ID: 4XFX [4]). The 14C-containing CA molecule is colored cyan, the 45C-containing molecule is colored orange. B) SDS-PAGE demonstrating correct disulfide-bond formation of hexamer-2-foldon assemblies. C) Size exclusion chromatogram demonstrating the relative solution molecular weights of hexamer-2-foldon and comparable capsid assemblies. D) Negative-stain EM micrograph of hexamer-2-foldon. E) 2D class averages and 3D reconstruction of hexamer-2-foldon from negative-stain EM. F) Appearance of hexamer-2-foldon crystals. G) Hexamer-2-foldon asymmetric unit and symmetry mates with electron density (2Fo-Fc 1σ). H) Overlay of hexamer-2-foldon, disulfide-bonded hexamer (PDB ID: 3H47 [16]), and WT CA (PDB ID: 4XFX [4]) crystal structures. I) Overlay of dimerized CTD domains of hexamer-2-foldon, native WT CA (PDB ID: 4XFX), and native dehydrated CA (PDB ID: 4XFY [4]) crystal structures.

The hexamer-2 trimeric interface is stabilized by the bacteriophage T4 foldon domain

I sought to stabilize the weakly oligomeric native hexamer-2 by using template-directed trimerization. I searched the protein data bank for a small, naturally trimeric protein with dimensions matching the central three CA C-termini within hexamer-2. I settled on the 3-kDa bacteriophage T4 foldon domain as this domain has been used to stabilize HIV Env and other weakly trimeric proteins for biochemical and structural analysis [13, 111, 112]. In the design, the foldon domain is directly fused to the inner CA lattice surface and, thus, is unlikely to block host factor binding to the outer exposed CA surface **(Figure 2-4A).**

I removed several flexible CA C-terminal residues to create a more rigid fusion between CA and foldon and fused foldon to the C-terminus of the CA construct used to assemble native hexamer-2 to create hexamer-2-foldon (construct denoted $CA_{(1\text{-}226)/45C/54E\text{-}foldon}$). Upon purification, I mixed and assembled $CA_{(1\text{-}226)/45C/54E\text{-}foldon}$ at a 1:1 ratio with either $CA_{14C/42E}$ or $CA_{(1\text{-}221)14C/42E/A204D}$. $CA_{(1\text{-}221)14C/42E/A204D}$ was designed to prevent hexamer-2-foldon propagation by using a previously reported CA lattice mutant (CA_{A204D}) [1], and to better facilitate structural biology by removing ten flexible CA C-terminal residues.

The major products observed on non-reducing SDS PAGE were disulfide-linked 1/3-hexamers, as designed. Upon purification, these separated into stoichiometric 25 kDa ($CA_{14C/42E}$ or $CA_{(1\text{-}221)14C/42E/204D}$) and 28 kDa ($CA_{(1\text{-}226)/45C/54E\text{-}foldon}$) species in reducing conditions **(Figure 2-4B)**. Strikingly, monodisperse ~150 kDa species for both hexamer-2-foldon and hexamer-2-

foldon$_{(1\text{-}221)/204D}$ were observed by SEC **(Figure 2-4C)**. This demonstrates that the ~50-kDa disulfide-linked fragments within hexamer-2-foldon form stable trimers. We analyzed the global architecture of hexamer-2-foldon and hexamer-2-foldon$_{(1\text{-}221)/204D}$ by negative-stain electron microscopy and observed triangular shaped assemblies that matched our design predictions at low resolution **(Figure 2-4: D,E)**

The crystal structure of hexamer-2-foldon validates its correct CA lattice architecture

I solved a crystal structure of hexamer-2-foldon$_{(1\text{-}221)/204D}$ to 4.5 Å resolution (statistics summarized in **Table 2-8**). Both hexamer-2-foldon and hexamer-2-foldon$_{(1\text{-}221)/204D}$ formed similar thin, plate-like hexagonal crystals in numerous commercially available conditions **(Figure 2-4F)**. The structure was solved by molecular replacement with a native CA structure (PDB ID 4XFX). The asymmetric unit contained two disulfide-linked CA monomers arranged as 1/3-hexamers. Upon application of crystallographic three-fold symmetry, the six CA monomers centered around the capsid three-fold interface are observed **(Figure 2-4G)**. Crystal packing created a two-dimensional CA lattice very similar to that of native hexameric CA, but the 42E/54E mutations prevented close packing of traditional hexamers. Large solvent channels were present between lattice layers with enough area to contain the trimeric foldon domain. To our surprise, only partial density was observed in the predicted foldon location, and peptide backbone could not be reliably modeled. This suggests that the linkage between CA and foldon is flexible, and that CA architecture within hexamer-2-foldon is not

Data Collection	
Wavelength (Å)	0.9798
Space Group	P321
Cell Dimensions	
a, b, c (Å)	95.03, 95.03, 122.26
α, β, γ (°)	90, 90, 120
Molecules/asymmetric unit	2 CA monomers
Resolution (Å)	50.0-4.20 (4.31-4.20)
R_{merge}	0.11 (0.90)
$I/\sigma I$	4.4 (1.4)
Completeness (%)	94.7 (89.2)
Redundancy	3.0 (2.6)
Unique reflections	5023
Refinement	
Number of nonhydrogen atoms	3252
R_{work}/R_{free} (%)	28.8/31.6 (50.9/41.8)
Average B factor	191
Root mean-squared deviation (rmsd)	
Bond lengths (Å)	0.012
Bond angles (°)	1.3
Ramachandran analysis	
Preferred regions (%)	96.4
Allowed regions (%)	3.36
Outliers (%)	0.24

Table 2-8. Data collection and refinement statistics for hexamer-2-foldon(221*/204D). Statistics in parentheses indicate those for the highest resolution shell. These statistics are current as of Sept. 2018 but may change upon publication.

rigidly or unnaturally constrained by the presence of the foldon domain. Flexibility between CA and foldon likely contributed to the two-dimensional growth, high-solvent content, and weak diffraction of hexamer-2-foldon crystals.

I compared the hexamer-2-foldon structure to crystal structures of native CA (PDB ID: 4XFX), native dehydrated CA (PDB: 4XFY) and soluble, disulfide-linked hexamers (PDB ID: 3H47)—all of which crystallize in a flat CA lattice comparable to native capsids. In all structures individual 1/3-hexamer subunits

align very closely, especially at CA domains making intra-hexamer contacts **(Figure 2-4H)**. When analyzed as rigid bodies, the dimerized CTDs **(Figure 2-4I)** and six NTDs of hexamer-2-foldon also align well with the global architecture of native and dehydrated native CA structures. In fact, hexamer-2-foldon appeared almost as a mixture between the two native CA structures, with certain parts aligning better to native CA and others to dehydrated native CA. The most apparent difference observed in hexamer-2-foldon is the slight rigid body motion of its CTD dimer pair mediated by the flexible hinge between NTD and CTD domains. In the context of a native CA lattice this motion is not allowed due to NTD-CTD packing, but solubilizing mutations in hexamer-2-foldon prevent such packing. The rotated CTD domains are stabilized by crystal contacts. This causes CA helix 10, which packs at the three-fold CA interface, to be in an open conformation most similar to that found in disulfide-linked CA hexamer structures, and, to a lesser extent, native dehydrated CA **(Figure 2-4H)**.

2.3 Discussion

Previous studies of CA-host factor interactions have lacked soluble CA oligomers that capture specific lattice interfaces found in complete capsids. This has greatly reduced the speed and ease of understanding these interactions. I have made substantial advances towards relieving these experimental burdens by developing novel partial-hexamer assemblies that are easy to produce in high purity and large scale. I have made two classes of partial-hexamer assemblies: 1) assemblies that contain only intra-hexamer CA interfaces, and 2) assemblies that contain a combination of intra- and inter-hexamer surfaces. The assemblies

that contain only intra-hexamer surfaces, specifically 1/3- and ½-hexamers with mutations at all CA dimerization interfaces (184A/185A), are excellent tools for better binding mode mapping of host factors that target complete hexamers, like TRIMCyp (discussed in Chapter 4), FEZ1 (discussed in Chapter 4), CPSF6, and Nup153 (see section 1.4).

Generally, most experiments failed when I attempted to introduce native dimerization motifs into 1/3- and ½-hexamers. 1/3-hexamers with native dimerization motifs, referred to as tetramers, assembled and purified well, but their dimerization was potentially too weak to enable their use as an effective tool. Similarly, the native hexamer-2 needed the additional artificial trimerization aid of foldon to be successfully assembled. When I returned native dimerization motifs to ½-hexamer assemblies, the resultant disulfide formation was extremely inefficient. The impurity of these reactions does not make them feasible tools. Much effort was put into them, however, since assemblies like the CA 9-mer (or triskelion) would contain the important CA three-fold interface and would be spectacular examples of our design proficiency.

The assemblies that contain both intra- and inter-hexamer surfaces, specifically hexamer-2-foldon, are groundbreaking in that they contain the most experimentally difficult interfaces to capture. Hexamer-2-foldon contains all unique inter-hexamer surfaces, with the most challenging being the complete three-fold CA surface. This surface is mapped to be targeted by the restriction factor MxB. If the CA interface targeted by MxB is like other targeted CA surfaces, then numerous new host factors may be identified that also bind at this

site. We are in an excellent position to rapidly test new host factors in this regard. We believe that hexamer-2-foldon is the perfect complement to the previously published disulfide-bonded hexamers. Its high yield and excellent solution behavior further buoy its significance.

I believe that the slight differences observed between the hexamer-2-foldon structure and other capsid structures presented here are simply an example of the malleability required in the native CA lattice to provide curvature, especially since the observed differences are most prominent at CA regions known to be flexible (NTD-CTD hinge, CTD dimer interface). As such, I expect the inter-hexamer interfaces within hexamer-2-foldon assemblies to sample an ensemble of related structures in solution and to only form a rigid structure (like observed) in crystallographic conditions. I believe the solution biophysical analysis and multi-faceted structural validation described here supports the correct architecture of hexamer-2-foldon assemblies and their use in future analysis of capsid-host factor interactions.

Each designed assembly described in this chapter absolutely necessitates the use of previously published engineered cysteines. The 14C/45C pair is most heavily relied on because it is the most specific and efficient. The 42C/54C is needed for more complex assemblies that contain ½-hexamer building blocks. In its publication and in our hands the 42C/54C pair is less efficient than its counterpart but is still effective. I only briefly began working on a method to introduce a third disulfide pair in order to cross-link four CA monomers in a hexamer-like fashion. I believe the efficiency of such a reaction would be too low

to be effective, as evidenced by the large drop in efficacy going from one disulfide bond in 1/3-hexamers to two disulfide-bonds in ½-hexamers.

2.4 Future directions

All in all, I believe I tested most of the design space using combinations of disulfide pairs and dimer interface mutations. As assemblies grew larger and included more disulfide bonds, their formation efficiency was drastically reduced. Certain assemblies, like the CA 9-mer and butterfly, could potentially be purified in yields large enough for negative-stain or cryo-EM analysis. However, their use in solution-based biochemical assays would be limited due to their low yield and poor stability.

Certain assemblies, like the tetramers and 9-mer, may benefit from template-based oligomerization like used in the hexamer-2-foldon. A simple, small dimeric or trimeric domain, like a coiled-coil, could be C-terminally fused to either the tetramer or 9-mer constructs to improve their dimerization/trimerization affinity. In this light, I have done very little screening or optimization of different artificial oligomerization domains. The foldon domain was the only one I attempted, which happened to work very well with hexamer-2. There are dozens of foldon-comparable domains that could be used to stabilize dimers, trimers, tetramers, etc. The SpyTag/SpyCatcher isopeptide bond system, which I will talk about in more detail in Chapter 3, could also be used to stabilize these assemblies. An early version of hexamer-2 used this system, but the final assembly appeared too flexible (data not shown).

A potentially interesting future avenue is in measuring the oligomerization affinity of certain capsid assemblies with native interfaces. For instance, the dimerization affinity of the two tetramer assemblies (which are dimerized 1/3-hexamers) could be measured. A similar analysis could be undertaken with the ½-hexamer assemblies that reform complete hexamers at high concentrations, or with the hexamer-2 assembly that is weakly trimeric without the stabilizing foldon domain. If one could accurately measure their oligomerization affinities the resultant values could better aid in computational modeling of mature capsid assembly.

2.5 Experimental procedures

2.5.1 Cloning, expression, and purification

Cloning was performed using standard molecular biology techniques. Point mutations were made using the QuikChange protocol from Stratagene and the KOD Hot Start polymerase. Overlapping PCR with Pfu polymerase was used to generate CA-foldon fusion constructs. The original CA14C/45C/184A/185A construct in the pET11a vector was obtained from Christopher Aiken at the University of Vanderbilt. The foldon DNA sequence was purchased as a large primer from IDT.

All CA constructs described in this chapter were cloned into pET11a and lacked any type of affinity tag. For cloning and plasmid purification DH5α XL-10 Gold cells (derived from Agilent stock) were used.

To express CA, I used chemically competent *E. coli* BL-21(DE3) cells transformed using the heat shock method with the CA plasmid of interest.

Generally, 1.5L flasks of Terrific broth were used for protein expression. Cells were grown to an OD of 0.6-1 and induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 22-25°C for 12-16 hrs. This expression procedure generally yielded >50 mg of purified protein.

Cells were harvested by centrifugation and lysed by a microfluidizer. The lysate was centrifugally clarified at 13500 rpm for 35 minutes. CA protein was precipitated from the supernatant by adding up to ~25-30% w/v ammonium sulfate to the solution at 4°C for 30min-1hr with slow stirring (~100 rpm). CA-foldon fusion proteins required 35% w/v ammonium sulfate precipitation. For untagged CA, the ammonium sulfate pellet was resuspended in 30-50 mL of 25 mM HEPES pH 7, 0.1 mM TCEP (called SA buffer). The protein solution was dialyzed for several hours or overnight into 2L of SA buffer to remove excess salts. The dialyzed solution was applied to a 5mL HiTrap-SP column and protein was eluted over a salt gradient. CA usually eluted between 20-100 mM NaCl and was very pure. SEC is not necessary for further purification. The SP column elution was concentrated and dialyzed overnight into 50 mM TRIS pH 8, 75-100 mM NaCl and 30 mM BME (termed CA storage buffer) and frozen the next day. The CA-foldon fusion was applied to a HiTrap Q anion exchange column after ammonium sulfate precipitation/dialysis and eluted around 250 mM NaCl. The elution peak was approximately 50% pure, but was pooled, concentrated and dialyzed into CA storage buffer. The purity after each step was analyzed by SDS-PAGE.

2.5.2 CA assembly, disulfide bond formation, and purification

1/3-hexamer base assemblies used purified CA (in CA storage buffer) and were assembled at 1:1 molar ratios usually between 10-40 mg/mL of total protein. Mixtures were dialyzed overnight (using Thermo Slide-a-lyzer dialysis cassettes) in 50 mM TRIS pH 8, 1M NaCl (and often 10-30 mM BME). Mixtures were dialyzed for a second night in 50 mM TRIS pH 8. Each dialysis step can be extended for 2-3 days at the experimenter's convenience. For ½-hexamer-type assemblies purified CA constructs were mixed at 1:1:1 molar ratios. Assemblies were purified using HiTrap Q anion exchange columns. 1/3- and ½-hexamer type assemblies usually eluted between 100-200 mM NaCl. A monoQ was occasionally needed to separate contaminates from ½-hexamers. Hexamer-2-foldon assemblies eluted from an anion exchange column at approximately 250 mM NaCl. Each assembly was polished using a S200PG or S200GL SEC column ran in 50 mM TRIS pH 8, 300 mM NaCl. All assemblies could be concentrated to at least 50 mg/mL and frozen at -80°C for months to years.

Several mpro cleavage sequences between the CA NTD and CTD were tested, however our traditional sequence (CA144-RMYS—KLQAGF—IRQG-CA157) cut with the most efficiency. The mpro cleavage sequence drastically reduced disulfide-bonded hexamer formation in the 14C/45C/AA background, but did little to alter the assembly of 1/3- and ½-hexamers. The CTD cleavage is efficient in monomeric CA but is reduced in 1/3- and ½-hexamers. Cleavage reactions went at least overnight at 4 degrees and were usually not complete.

Anion exchange using either a HiTrap Q or monoQ column was required to separate uncut and cut species.

2.5.3 Crystallization studies

All small assemblies were screened for crystallization using the microbatch under-oil method using a 2:1 ratio of paraffin to silicon oil. Protein concentrations ranged from 0.5-3 mg/mL. Most trays were screened at room temperature. Common screens tested were PACT, ProComplex, JCSG 1, MCSG I, MCSG II, Classics, and Morpheus. 1 uL of protein solution was mixed with 1 uL of precipitant solution.

All small assemblies crystallized in numerous conditions and optimized well. 1/3-hexamer-EE crystallized at room temperature between 1-2 mg/mL in 0.2 M Calcium Acetate Hydrate, 0.1 M MES: NaOH, pH 6, and 25 % (w/v) PEG 8000. The crystals were frozen in paratone oil. Data was collected at the NE-CAT beamline 24ID-E at the Advanced Photon Source. ½-hex-EE-ΔCTD crystallized at room temperature at 1.5 mg/mL in 0.1M Sodium Citrate pH 5 and 8% PEG 8000. The crystals were frozen in paratone oil. Data was collected at NE-CAT beamline 24ID-C at the Advanced Photon Source. ½-hexamers crystallized at room temperature at 1 mg/mL in 0.1 M PCB buffer pH 8 25% (w/v) PEG 1500 and were cryo-protected in 25% ethylene glycol. Data was collected at NE-CAT beamline 24-IDC at the Advanced Photon Source. Hexamer-2-foldon$_{(1\text{-}221/A204D)}$ crystallized at room temperature at 0.75 mg/mL in 0.2M NaCl, 0.1M HEPES pH 7.5, and 12% PEG 8000. Crystals were cryo-protected in 25% glycerol. Data was collected at Brookhaven National Laboratory beamline AMX.

I mixed one tetramer-1 assembly with three CypA molecules to make a complex for crystallization. The complex crystallized after several weeks at room temperature at approximately 5 mg/mL in 0.2M Potassium thiocyanate, 0.1M BisTris propane pH 6.5, and 20% PEG 3350. Crystals were frozen in 25% glycerol. Data was collected at NE-CAT beamline 24-IDC at the Advanced Photon Source. SDS-PAGE analysis of crystal drops confirmed the degradation of the flexible CTD from tetramer-1.

Both HKL2000 and XDS were optionally used for data processing [113, 114]. Molecular replacement CA search models were made from either a disulfide-hexamer structure (PDB ID: 3H47 [16]) or a native CA structure (PDB ID: 4XFX [4]). The tetramer-CypA structure was solved using the CA NTD-CypA structure (PDB ID: 1AK4 [24]). Molecular replacement was performed using the CCP4 program Phaser [115, 116]. Iterative rounds of refinement in REFMAC and PHENIX were carried out, along with model building in COOT [117-119]. Alignment RMSD values were generated using the LSQ align module in COOT. Figure images were generated in Pymol and Coot [120].

2.5.4 Negative-stain EM analysis

To observe assemblies using negative-stain EM, we glow discharged a 400 mess Cu grid, carbon coated, for 30 seconds at 25 mA. Sample was applied for 30 seconds and excess was removed by blotting with filter paper. We performed negative staining by applying 2% uranayl acetate to the grid, blotting immediately, applying again, incubating 30 seconds, and finally blotting residual stain with filter paper. We then collected 50-100 images at 73k magnification on

the Tecnai T12 120 kV or Talos L120C 230 kV microscopes. We analyzed images and performed 2-D classification using Relion [121]. All negative-stain experiments in this chapter were performed by Kaifeng Zhou.

3 Design and purification of large, multi-hexamer assemblies

3.1 Introduction

Some of the least understood capsid-binding host factors are those that span multiple hexamers for CA recognition. Frequently, these factors form large oligomers over the capsid surface to take advantage of avidity and the repeating nature of the CA lattice. Specifically, the detailed mechanism of TRIM5α CA recognition as a chief block to interspecies retroviral transmission remains elusive. Other factors, like full-length, native versions of TRIMCyp, MxB, and CPSF6 also (or likely) span multiple hexamers (see section 1.3 and 1.4).

Large (>150 kDa), soluble capsid oligomers may be required as binding partners to obtain high-resolution structures of these factors in complex with the capsid lattice. I successfully co-opted established protein engineering strategies to stabilize such large, multi-hexamer capsid assemblies. My major breakthrough was in the use of the SpyCatcher/SpyTag system to link neighboring disulfide-bonded hexamers via isopeptide bond [122]. SpyCatcher and SpyTag are a 13-kDa protein and 13-amino acid peptide, respectively, derived from the *Streptococcus pyogenese* fibronectin-binding protein FbaB. Upon mixture, SpyCatcher and SpyTag are rapidly and stably linked by a natural isopeptide bond **(Figure 3-1A, left)**.

By using this approach, I was able to produce two, three, and seven hexamer platforms (referred to as di-, tri-, and hepta-hexamers, respectively). The relatively high-yield and efficiency of these assemblies makes them an

excellent tool to study CA-host factor complexes that could not be captured before.

3.2 Results

3.2.1 Use of the SpyCatcher/SpyTag isopeptide bond system to stabilize di- and tri-hexamer assemblies

My approach to building large capsid platforms was predicated on the simple ‘mix and match’ strategy described in Chapter 2 to build small capsid assemblies. My initial goal was to build hexamers composed of a mixture of two different CA constructs. I will refer to these as “mixed hexamers.” Mixed hexamers could be built from at least two types of CA proteins—the traditional hexamer subunit, $CA_{14C/45C/AA}$, and a different, but complementary CA construct (complementary in that it also contains 14C/45C engineered cysteines, in addition to other features). $CA_{14C/45C/AA}$ is unable to participate in inter-hexamer interactions because its dimer interface is mutated (184A/185A). I hoped that mixed hexamers could be formed where one (or more) of the six CA subunits within the hexamer retains a native dimerization motif and can thus participate in inter-hexamer interactions. These native dimerization motifs would enable controlled, multi-hexamer oligomerization. Moreover, this dimerization-competent CA molecule could be directly fused at its free C-terminus to any number of useful protein domains to stabilize its weak dimerization affinity.

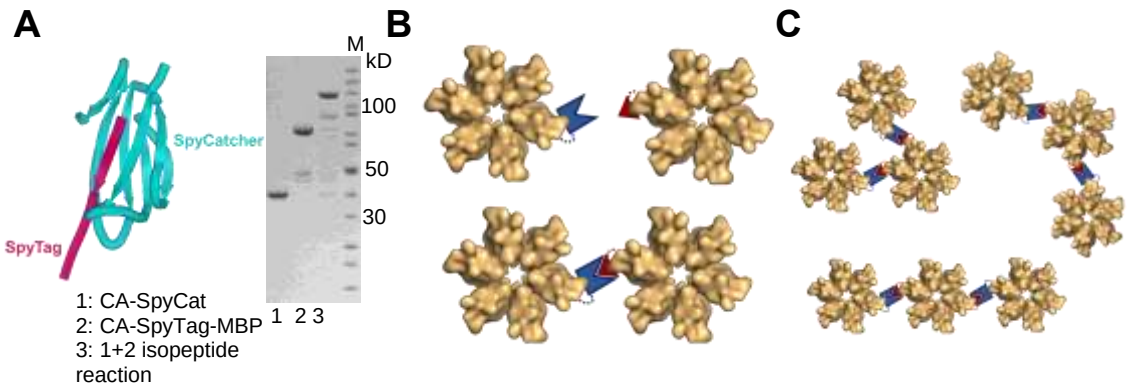


Figure 3-1. SpyCatcher/SpyTag assembly design. A) Left panel: Crystal structure demonstrating the SpyTag (magenta) isopeptide bonded to SpyCatcher (cyan) (PDB ID: 4MLS [2]). Right panel: CA fusions to SpyTag and SpyCatcher form a covalent bond as observed on SDS PAGE. B) Di-hexamer design. SpyCatcher domains are blue and SpyTag peptides are red triangles. Capsid NTD domains are shown in surface colored tan. C) Tri-hexamer design with coloring as in (B). Capsid NTDs were derived from PDB ID: 4XFX [4].

In this light, I made CA C-terminal fusions to SpyTag and SpyCatcher motifs. A CA construct with a C-terminally fused SpyCatcher ($CA_{14C/45C}$-SpyCat) rapidly reacted with a CA construct with a C-terminally fused SpyTag ($CA_{14C/45C}$-SpyTag-MBP) to form a covalent dimer **(Figure 3-1A, right)**. MBP was required after the SpyTag to enable visualization on SDS-PAGE (since SpyTag is only a few kilodaltons in size). I placed a five amino acid Gly-Ser linker between CA and SpyTag/Catcher to provide freedom of motion. Other linkers were tested (a rigid fusion and a ten amino acid Gly-Ser stretch) and did not show obvious differences. The 14C/45C engineered cysteines within $CA_{14C/45C}$-SpyCat and $CA_{14C/45C}$-SpyTag enabled their incorporation into mixed hexamers with $CA_{14C/45C/AA}$. Importantly, the CA-Spy fusion molecules have native dimerization motifs **(Figure 3-1B, C)**. I mixed and assembled a ratio of one CA-Spy fusion molecule with either 4 or 6 $CA_{14C/45C/AA}$ molecules and purified the assembly mixture using anion exchange chromatography. I observed several distinct elution peaks that corresponded to disulfide-bonded hexamers containing zero, one, two or more CA-Spy fusion molecules as observed by SDS-PAGE **(Figure 3-2A,C)**. This result clearly demonstrated the effectiveness of “mixed-hexamer” incorporation of the CA-Spy fusion molecules.

Notably, the amount of zero, one, two, or more hexamer incorporation events closely followed theoretical predictions using a simple binomial distribution model **(Figure 3-2A, B)**. Thus, the ratio of incorporation events could be fine-tuned to experimental needs. Hexamers with one incorporated

SpyCatcher (termed hexamer-1SpyCat) and hexamers with either one or two SpyTag-MBP (termed hexamer-1SpyTag-MBP and hexamer-2SpyTag-MBP, respectively) were purified in this fashion.

I incubated hexamer-1SpyCat and hexamer-1SpyTag-MBP at a 1:1 ratio to initiate isopeptide bond formation between the two hexamers (denoted di-hexamers hereafter). I also mixed hexamer-1SpyCat and hexamer-2SpyTag-MBP at a 2:1 ratio to covalently link three hexamers (denoted tri-hexamers hereafter). For both mixtures the SpyCatcher/SpyTag isopeptide reaction occurred efficiently at all tested concentrations (generally between 1-20 mg/mL of total protein). Covalent ~300 kDa (putative di-hexamer MW) and ~450 kDa (putative tri-hexamer MW) assemblies were purified to homogeneity using SEC as observed by non-reducing SDS PAGE and SEC-MALS **(Figure 3-2D, E).**

We used negative-stain EM to validate the correct structure of the purified assemblies **(Figure 3-2F, G)**. Strikingly, both di-hexamers and tri-hexamers displayed similar hexamer orientation and spacing to that of hexamers observed in infectious virions. In stark contrast, wide spacing and apparent flexibility was observed between hexamers when the native dimerization motifs on the CA-Spy fusion proteins were mutated **(Figure 3-2H)**. This supports a model where hexamers within di-hexamers and tri-hexamers are engaged via the native dimerization interfaces incorporated in the CA-SpyTag/Catcher fusion molecules.

Tri-hexamer assemblies contained three orientations of hexamer building blocks—referred to as triangular, branched, or linear. These orientations are a result of the random incorporation of two CA-SpyTag-MBP molecules into two of

the six positions of a hexamer **(Figure 3-1C)**. Triangular, branched, and linear orientations should represent 40%, 40%, and 20% of the total, respectively, if CA-SpyTag-MBP incorporation is truly random (based on binomial distribution principles). The observed representations were ~27%, 30%, and 33%, indicating a slight assembly preference of linear tri-hexamers over adjacent and branched tri-hexamers. Unfortunately, I was unable to separate these different tri-hexamer species so all future tri-hexamer experiments were performed with this mixture.

Wei Wang (a post-doctoral fellow in the lab) also used this approach to stably link the previously reported disulfide-bonded CA pentamers with hexamers. He assembled both pentamer-hexamer and hexamer-pentamer-hexamer assemblies comparable to di-hexamers and tri-hexamers (data not shown).

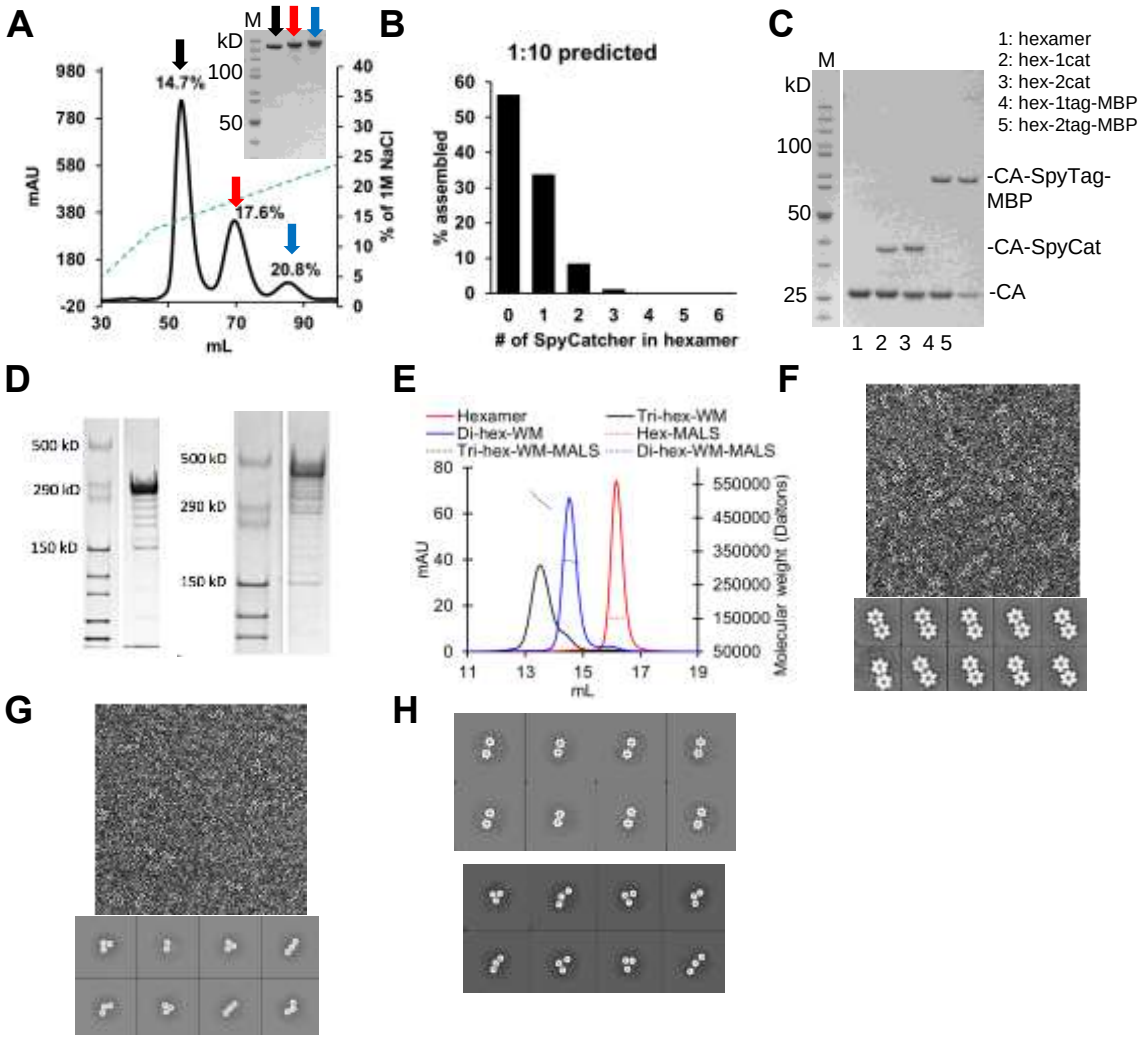


Figure 3-2. Di-hexamer and tri-hexamer assembly and purification. A) Anion exchange column showing separation of hexamers with zero, one, or two SpyCatcher domains (non-reducing SDS PAGE of each peak above). B) Binomial distribution model of SpyCatcher incorporation events based on 1 CA-SpyCat: 10 CA molecule assembly ratio used in (A). C) Reducing SDS PAGE showing ratio of CA-SpyCat and CA-SpyTag-MBP incorporation into hexamers. D) Non-reducing Tris-acetate gel demonstrating disulfide-bond and isopeptide-bond stabilization of di-hexamers (left) and tri-hexamers (right). E) SEC-MALS analysis of di-hexamers, tri-hexamers, and hepta-hexamers. F) Di-hexamer negative-stain EM raw images and class averages (scale bar not provided). G) Tri-hexamer negative-stain EM raw images and class averages. H) Negative-stain class averages of di-hexamers (top) and tri-hexamers (bottom) with dimer interface mutations.

3.2.2 Megadalton capsid assemblies are stabilized by a CA-binding nanobody

The SpyTag/SpyCatcher approach theoretically allowed for the simple assembly of megadalton, symmetric 7-hexamer assemblies (referred to as hepta-hexamers). Hepta-hexamers are assembled with a central hexamer containing SpyTag-MBP fusions at all six CA positions (termed hexamer-6SpyTag-MBP). Hexamer-6SpyTag-MBP is then reacted and surrounded by six hexamers each having one SpyCatcher fusion (hexamer-1SpyCat) **(Figure 3-4A)**. I aimed to produce hepta-hexamers with native dimerization motifs on all SpyCatcher/Tag fusion molecules to promote correct capsid architecture. In this design, the exposed CA molecules on the outside of hepta-hexamers are $CA_{14C/45C/AA}$ and lack the ability to form inter-hexamer interactions. This assembly proved challenging, however, as most initial hepta-hexamer assemblies were incomplete (containing between four and seven hexamers) and heavily aggregated during purification **(Figure 3-4B, C).**

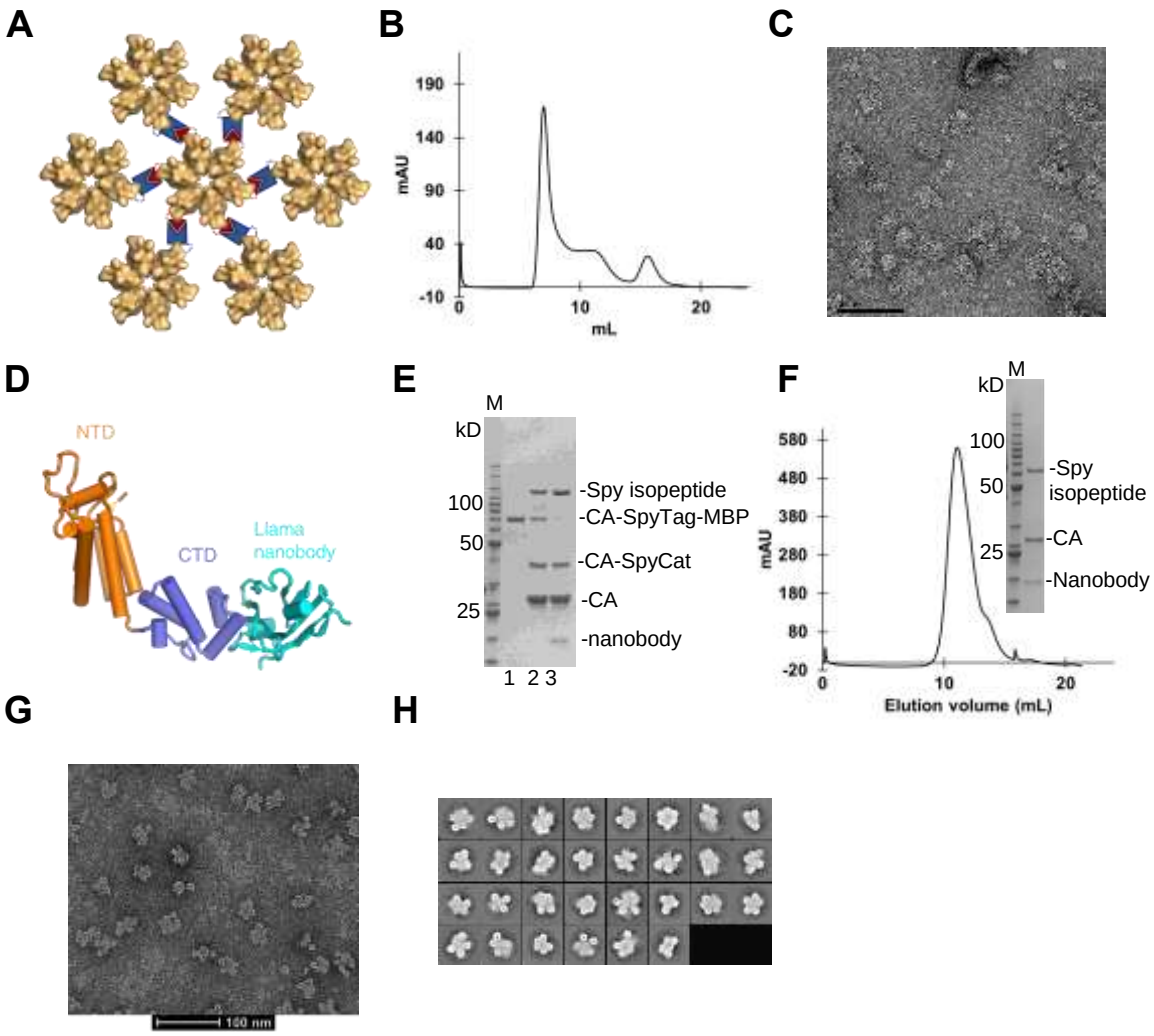


Figure 3-3. Hepta-hexamer design and purification. A) Model of hepta-hexamer design. Blue represents SpyCatcher, red triangle represents SpyTag. CA NTDs are colored tan in surface view. B) Representative Superose 6 (S6) chromatogram of hepta-hexamer purification without nanobody. Most protein eluted near the void volume of the column, indicating aggregation. C) Representative negative-stain electron micrograph of aggregating hepta-hexamer from (B). D) Model of the llama nanobody bound to CA (PDB ID: 5O2U [9]. E) SDS-PAGE demonstrating SpyTag-Catcher reaction with and without the nanobody. Lane 1 is hexamer-6SpyTag-MBP alone. Lane 2 is hex-6tag-MBP+hex-1cat w/o nanobody. Lane 3 is hex-6tag-MBP+hex-1cat with nanobody. F) S6 chromatogram showing improved behavior of hepta-hexamer with bound nanobody. SDS PAGE of peak inset. G) Negative-stain EM analysis of hepta-hexamer with bound nanobody. H) Negative-stain EM 2D class averages of hepta-hexamer with bound nanobody. CA NTD models are derived from PDB ID: 4XFX [4].

To address these problems, I incorporated a newly described anti-capsid nanobody in the assembly procedure [9]. This nanobody binds the CA CTD with a reported 300 nM dissociation constant **(Figure 3-4D)**. Based on the published nanobody-CA CTD structure, I hypothesized that the nanobody would not interfere with CA hexamer formation but would block inter-hexamer propagation and prevent non-specific propagation of multi-hexamer assemblies.

I assembled hexamer-6SpyTag-MBP with each CA position containing a native dimerization motif (containing six copies of $CA_{14C/45C\text{-}SpyTag\text{-}MBP}$) in the presence of stoichiometric amounts of nanobody. Whereas hexamer-6SpyTag-MBP without nanobody is poorly behaved and difficult to purify, the same construct in the presence of nanobody remains soluble and purifies well. I also added nanobody to the previously purified hexamer-1SpyCat (a ratio of three nanobodies to one hexamer). I mixed one hexamer-6SpyTag-MBP with an excess of hexamer-1SpyCat (usually 8-12 fold) in the presence of nanobody. The SpyTag/SpyCatcher reaction appeared to go to completion after overnight incubation **(Figure 3-4E)**. The same reaction without nanobody failed to go to completion. Putative hepta-hexamers with bound nanobody remained soluble and eluted as a single, albeit broad, peak on SEC **(Figure 3-4F)**. Strikingly, upon negative-stain EM analysis, it appeared that most assemblies contain seven hexamers **(Figure 3-4G)**. In 2D class averages, their architecture was heterogenous, however, as only a fraction appeared 6-fold symmetric and many contained hexamers that were loosely packed **(Figure 3-4H)**. With more

optimization to their assembly and behavior, I believe hepta-hexamers can be a valuable tool to study capsid-host factor interactions.

3.2.3 Design and purification of the “snowflake”

I attempted to build a large capsid assembly composed of a central hexamer surrounded by six ½-hexamers. It was informally named the capsid “snowflake,” and its design was similar to hepta-hexamers described in Chapter 3.2.2 **(Figure 3-4A)**. Each CA monomer within the central hexamer contained a native dimerization motif and a C-terminal SpyTag-MBP. This was purified with bound nanobody to prevent aggregation. I will refer to this assembly as hexamer-6SpyTag-MBP-nanobody. The ½-hexamer building block was purified separately, and its assembly followed standard ½-hexamer procedure (see Chapter 2). It was formed from a 1:1:1 ratio of $CA_{14C/42E}$, $CA_{45C/54C\text{-}SpyCatcher}$, and $CA_{42C/54E}$ to create a disulfide-bonded ½-hexamer with a single SpyCatcher domain. Importantly, each CA monomer within this ½-hexamer contained a native dimerization motif. I will refer to this assembly as ½-hexamer-EE-SpyCat. ½-hexamer-EE-SpyCat was mostly aggregated on initial purification, but upon addition of nanobody its solution behavior improved greatly **(Figure 3-4B, C)**.

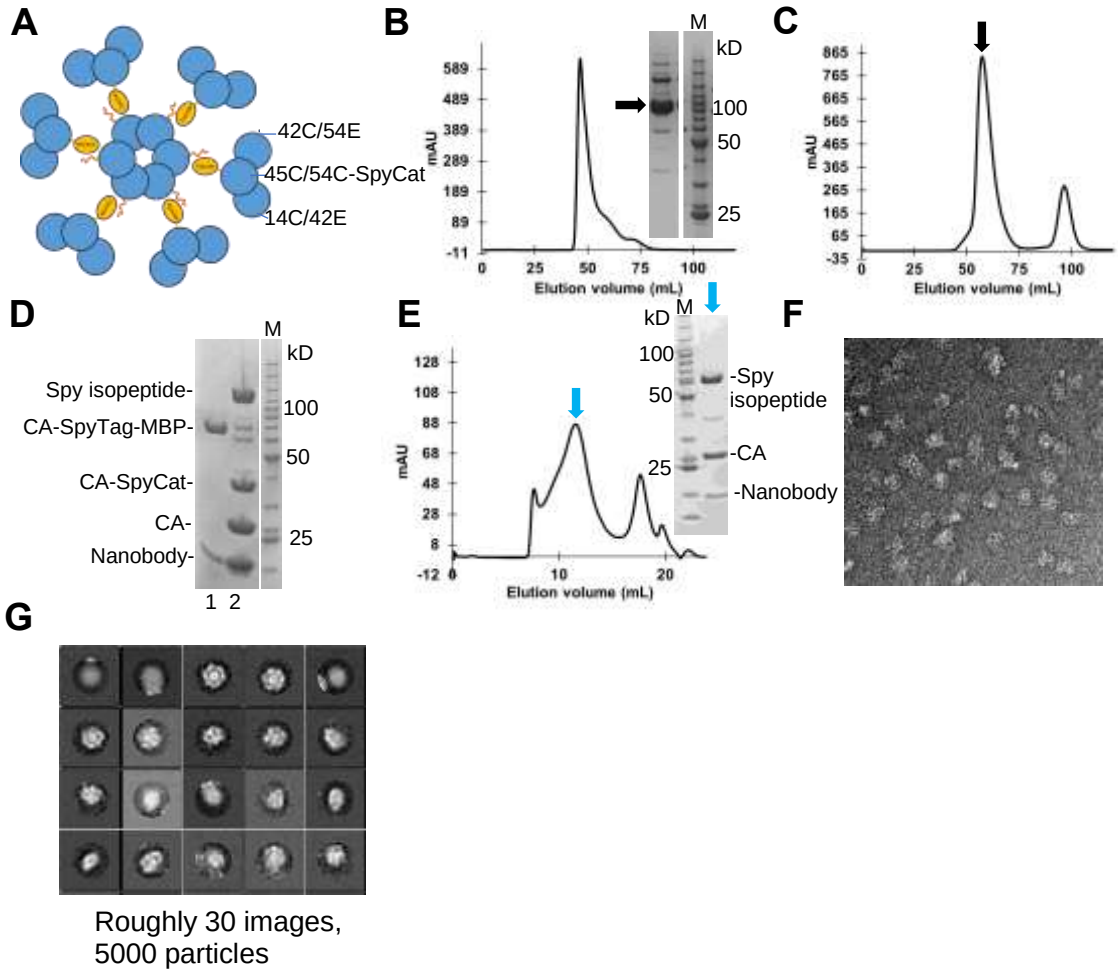


Figure 3-4. Design and purification of the capsid “snowflake.” A) A simple cartoon model demonstrating the snowflake design. The center hexamer is hexamer-6SpyTag, the outer fragments are ½-hexamer-EE-SpyCat. B) SEC chromatogram of the ½-hexamer-EE-SpyCat lacking the anti-capsid nanobody that prevents aggregation. The assembly eluted at a volume indicating aggregation. The non-reducing SDS-PAGE of the peak fraction is inset. The black arrow highlights the correct molecular weight species. C) SEC chromatogram of the ½-hexamer-EE-SpyCat in the presence of nanobody. This assembly did not aggregate. D) Efficient snowflake SpyCatcher/SpyTag isopeptide reaction products as observed on reducing SDS-PAGE. E) Final SEC chromatogram of the snowflake with bound nanobody. The reducing SDS-PAGE of the peak fraction is inset. F) Negative-stain electron micrograph of the snowflake peak fraction. G) Negative-stain EM 2D class averages of the snowflake.

I reacted a ratio of one hexamer-6SpyTag-MBP-nanobody with twelve ½-hexamer-EE-SpyCat-nanobody. The reaction went to completion efficiently after overnight incubation **(Figure 3-4D)**. The MBP tags on the central hexamer were removed by protease cleavage and the putative snowflake was purified using SEC. Depending on the preparation, the putative snowflake either eluted mostly at the correct molecular with some aggregation or primarily as an aggregate **(Figure 3-4E)**. The assembly was prone to precipitation upon concentration. Nonetheless, we performed negative-stain EM analysis of a correct molecular weight fraction **(Figure 3-4F)**. We observed large capsid assemblies that were reminiscent of our design model. Two-dimensional class averages were obtained that clearly showed the correct snowflake architecture in a fraction of the assemblies (**Figure 3-4G)**. At the time of this writing, I have not further optimized snowflake assembly and purification to make it more amenable in host factor binding and structural biology experiments.

3.2.4 Alternative production and stabilization of di-hexamers and tri-hexamers

My initial efforts to stabilize di-hexamers focused entirely on the usage of engineered cysteines. Instead of placing a SpyCatcher or SpyTag at one of the six positions within a mixed hexamer I incorporated a CA molecule with an additional engineered cysteine, either W184C or A204C (the constructs used were $CA_{14C/45C/AA/184C}$ or $CA_{14C/45C/AA/204C}$). A hexamer with a single 184C or 204C incorporation could, in theory, react with another hexamer containing a 184C or 204C incorporation to form a disulfide-bonded di-hexamer **(Figure 3-5A)**.

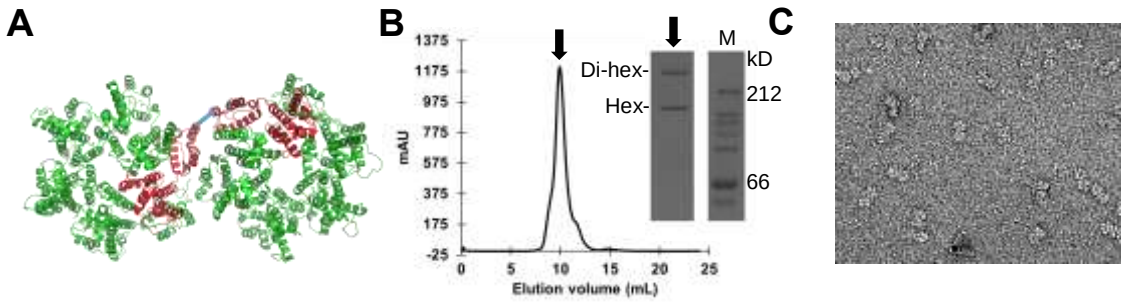


Figure 3-5. A204C disulfide-linked di-hexamers. A) Cartoon model showing position of the A204C asymmetric disulfide linking two hexamers (PDB ID: 3J34 [1]). Red CA molecules contain the A204C mutation, and the blue line represents a disulfide-bond between these molecules. B) SEC chromatogram and corresponding non-reducing SDS-PAGE of A204C linked di-hexamers. C) Negative-stain electron micrograph showing assembled and pure di-hexamer species (no scale bar provided).

The A204C disulfide bond has been described previously [1]. The Ahn lab at the University of Pittsburgh aided in the initial design and studies of the W184C disulfide. It was modeled from the high-resolution capsid tube cryo-EM structure [1]. Hexamers containing a single W184C or A204C incorporation have essentially no molecular weight difference from a hexamer without the additional cysteine so I could not specifically purify these altered hexamers. Instead, I mixed and assembled one W184C or A204C containing CA molecule with 10-20 molecules of $CA_{14C/45C/AA}$. I purified the assembly products on SEC (**Figure 3-5B**). The primary product was single hexamers, but a higher-molecular weight shoulder was present that could be purified and contained cross-linked di-hexamers. This procedure was generally very inefficient although disulfide-bonded di-hexamer architecture looked very promising on negative-stain EM (**Figure 3-5C**). The A204C disulfide generally assembled more efficiently than the 184C disulfide.

I used a similar approach to attempt to link three-hexamers with disulfide-bonds. I made mixed hexamers that incorporate both 207C and 216C mutations in a single molecule within a hexamer. This cysteine pair was previously published and is capable of linking three CA molecules at the three-fold CA interface [43]. However, the efficiency of this cysteine pair's reactivity was low and no disulfide-bonded tri-hexamer product was obtained (data not shown).

Because the production of di- and tri-hexamers stabilized by W184C, A204C, P207C, and T216C engineered disulfides was not efficient, I abandoned

their usage in favor of the SpyCatcher/SpyTag isopeptide bond system described above.

3.3 Discussion

Previous attempts to solve high-resolution structures of host factors that span multiple hexamers (most prominently TRIM5α) have failed (at least in part) due to a lack of a soluble CA binding partner. These host factors efficiently bind insoluble CA tubes but show no affinity towards CA hexamers. I believe the isopeptide-bond stabilized multi-hexamer assemblies described in this chapter are the first step to overcoming this challenge.

I successfully assembled and purified di-, tri-, and hepta-hexamer assemblies that faithfully recreate the capsid hexameric lattice. Importantly, these assemblies appear to be stabilized by native inter-hexamer dimer interfaces. The assemblies were relatively efficient to purify—from 4.5 L of *E. coli* cells I obtained approximately 2-5 mg of purified assembly. Both di- and tri-hexamers displayed excellent solution behavior. Hepta-hexamers, on the other hand, only remained monodispersed in the presence of an anti-capsid nanobody. Similarly, the capsid "snowflake," which is a hexamer surrounded by six ½-hexamers, required the nanobody during purification to prevent aggregation. This is the first demonstrated use of the nanobody in aiding in vitro capsid assembly.

The hepta-hexamers and snowflake contained co-purified sub-stoichiometric amounts of nanobody that appears necessary to keep both assemblies soluble. The amount and position of bound nanobody on each assembly is likely very different and contributes to their structural heterogeneity.

For instance, if any nanobody is present on CA molecules within the interior of a hepta-hexamer the bound nanobody would sterically clash with neighboring hexamers and prevent their close hexamer packing. Our hope was that the nanobody would be outcompeted at inter-hexamer surfaces within hepta-hexamers by the native CA oligomerization motifs and the stabilization created by disulfide- and isopeptide-bonds. This appeared to be the case in a majority of hepta-hexamer and snowflake assemblies since their hexamer-packing was consistent with known CA structures.

3.4 Future directions

The tri-hexamers assembled here can be in three different orientations (symmetric, linear, or branched) due to the random incorporation of SpyTag molecules into a hexamer. It would be beneficial, however, to obtain pure symmetric tri-hexamer assemblies. I did not succeed in separating the three tri-hexamer orientations using standard purification techniques, presumably because the properties of the different tri-hexamers species (primarily size and charge) are too similar. To make symmetric tri-hexamers a new design approach is needed. Katie Digianantonio in the lab is attempting to do this by fusing trimeric coiled-coils to CA constructs. She is also using this technique to stabilize tetra-hexamer assemblies.

With regards to hepta-hexamer and snowflake assemblies, I expect that the use of these to solve high-resolution host factor-capsid structures will require significant optimization of protein concentrations/conditions and nanobody-CA ratio to increase the amount of correctly packed and discrete assemblies. It is

clear, however, that the llama-derived anti-capsid nanobody is an effective tool to prevent CA oligomerization in vitro and may be useful in future attempts to build otherwise aggregation-prone CA lattice assemblies.

3.5 Experimental procedures

3.5.1 Cloning, expression, and purification

Cloning was performed using standard molecular biology techniques. Point mutations were made using the QuikChange protocol from Stratagene and the KOD Hot Start polymerase. Overlapping PCR with Pfu polymerase was used to generate certain CA-SpyTag-MBP or SpyCatcher fusions. Other fusions were produced by Gibson assembly. The SpyCatcher DNA sequence was provided by the Regan lab at Yale.

All CA constructs described in this chapter were cloned into pET11a. For cloning and plasmid purification DH5α XL-10 Gold cells (derived from Agilent stock) were used. All CA constructs lacking an additional domain or affinity tag were purified as described in section 2.5.1.

All CA-SpyTag fusion proteins contained a C-terminal mpro cleavage site, maltose binding protein (MBP) tag, and 6xHis affinity tag. All CA-SpyCatcher fusions contained a C-terminal 6xHis tag or mpro-6xHis tag. To express these constructs, I used chemically competent *E. coli* BL-21(DE3) cells transformed using the heat shock method with the CA plasmid of interest. Generally, 1.5L flasks of Terrific broth were used for protein expression. Cells were grown to an OD of 0.6-1 and induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside

(IPTG) at 22-25°C for 12-16 hrs. This expression procedure generally yielded >50 mg of purified protein.

To purify CA-Spy fusion constructs cells were harvested by centrifugation and lysed by a microfluidizer in lysis buffer (50 mM TRIS pH 8, 300 mM NaCl, and 0.1 mM TCEP). The lysate was centrifugally clarified at 13500 rpm for 35 minutes. CA-SpyTag/Catcher-6xHis constructs were applied to a 10-mL gravity nickel column, washed with several column volumes of lysis buffer, and eluted with 50 mM TRIS pH 8, 400 mM NaCl, 400 mM imidazole, 0.1 mM TCEP. Generally speaking, buffer composition was flexible as long as strong reducing agent was present. Eluted CA protein was diluted 5X in 50 mM TRIS pH 8, 0.1 mM TCEP and applied to a HiTrap-Q anion exchange column. A linear NaCl gradient was applied. CA-SpyTag/SpyCat fusion molecules eluted between 50 and 150 mM NaCl. The Q column elution was concentrated and dialyzed overnight into 50 mM TRIS pH 8, 75-100 mM NaCl and 30 mM BME (CA storage buffer) and frozen. Each step in the purification process was monitored by SDS-PAGE.

The llama nanobody gene was synthesized using GeneArt. It was cloned into pET28 with a C-terminal mpro-6xHis tag. The publication that produced the nanobody described a high affinity nanobody and a lower affinity nanobody with $_{56}$FDP$_{58}$>$_{56}$GYA$_{58}$ mutations. For this thesis, I only used the low affinity nanobody. We inadvertently introduced an S74R mutation in the weaker nanobody during cloning, but the mutation is on the nanobody surface away from its CA binding site.

The weak nanobody was grown in BL-21(DE3) cells and induced overnight with 0.5 mM IPTG at 18°. It was purified in oxidizing conditions by gravity nickel column, dialysis to remove excess salt, a combination of HiTrap Q and HiTrap SP flowthrough collection, and S75PG or S200PG columns in 50 mM TRIS pH 8, 300 mM NaCl. It frequently precipitated during concentration, but this appeared reversible if the protein was allowed to warm to room temperature.

3.5.2 Multi-hexamer isopeptide bond formation and purification

Hexamers with incorporated SpyTag-MBP or SpyCatcher were assembled as described in section 2.5.2 with alternating dialysis in high salt conditions and low salt conditions. To obtain the highest amount of one and two SpyTag/Catcher incorporations a ratio of one CA-SpyTag/Catcher to four CA14C/45C/AA was used. The binomial distribution function in Microsoft Excel was used to predict likelihood of hexamer incorporation of CA-Spy fusion molecules. After dialysis, the assembly mixtures were applied directly to a HiTrap-Q column without reducing agent. Hexamers with various amounts of SpyTag/Catcher incorporation were eluted by a linear NaCl gradient. Assemblies generally eluted between 100-200 mM NaCl.

SpyCatcher/Tag reactions generally proceeded overnight as isopeptide formation was occasionally not complete after only a few hours. The reaction was performed in 50 mM TRIS pH 8 and a range of NaCl concentrations without apparent difficulties (from 50 mM to 300 mM NaCl). Only after the reaction was complete was the MBP tag on CA-SpyTag-MBP removed by mpro digestion. Di-

hexamers, tri-hexamers, and hepta-hexamers were purified using an S200PG followed by S6 column in 50 mM TRIS pH 8, 300 mM NaCl.

Large molecular weight assemblies were analyzed by performing SDS-PAGE with NuPAGE 3-8% Tris-Acetate gels purchased from ThermoFischer following their provided protocol.

3.5.3 Negative stain electron microscopy analysis

We used the same procedure as described in section 2.5.4. Equal number of particles were picked for comparing di-hexamers/tri-hexamers with and without 184A/185A mutations at their dimer interface. Experiments were performed by both Kaifeng Zhou and Katie Digianantonio.

4 Analysis of the hexamer-targeting CA-binding host factors TRIMCyp and FEZ1

4.1 Introduction

Many host factors likely recognize specific surfaces within HIV-1 CA hexamers (described in section 1.3 and 1.4). My design of partial-hexamer subassemblies described in sections 2.2.1 and 2.2.2 enables a more thorough understanding of the binding modes of these hexamer-targeting factors. Previously, a detailed understanding of their CA binding mode was limited due to a lack of such CA assemblies.

In this chapter I describe detailed hexamer-binding mode analyses of two host factors, TRIMCyp and FEZ1. TRIMCyp is a potent restriction factor of HIV-1 and uses its two C-terminal cyclophilin domains to directly bind CA. FEZ1 is a newly identified cofactor of HIV-1. It appears to be recruited to the CA surface to

directly link CA with kinesin motors. Both TRIMCyp and FEZ1 are described in more detail in section 1.3 and 1.4.

I developed expression and purification protocols to produce TRIMCyp constructs that demonstrate good solution behavior and bind CA. Along with Dr. Katie Digianantonio (a postdoctoral fellow in the lab) we found that TRIMCyp has a flexible binding mode classified by its ability to bind pairs of CA molecules separated by considerable distances and with different orientations. Description of this binding mode was only possible with partial-hexamer and multi-hexamer assemblies described in previous chapters.

Along with Ivy Huang (a fellow graduate student in the lab), we purified full-length and truncation fragments of human FEZ1. We demonstrated that FEZ1 fragments can bind CA tubes with high affinity. This was the first demonstration of a direct FEZ1-CA interaction. We further used our FEZ1 constructs in combination with partial-hexamer assemblies to describe the novel mode by which FEZ1 targets hexamers. A CA-binding mode like the one demonstrated by FEZ1 has not been described for any other capsid-binding host factors.

4.2 TRIMCyp results

4.2.1 Expression and purification of dimeric TRIMCyp constructs

Before biochemical and structural analysis of TRIMCyp could begin, I needed to be able to produce well-behaving, capsid-binding constructs of TRIMCyp. Most virological studies use a version of TRIMCyp from the New World monkey species *Aotus trivirgatus*, as this was the first TRIMCyp gene described and potently restricts HIV-1 [74]. I received this gene from Jinwoo Ahn

at the University of Pittsburgh. Additionally, I purchased a synthesized gene fragment of the Old World monkey *Macaca fascicularis* TRIMCyp. TRIMCyp$_{Macaca}$ also potently restricts HIV-1 [75, 123]. TRIMCyp$_{Macaca}$ has a shorter linker region between the coiled-coil and cyclophilin domains compared to TRIMCyp$_{Aotus}$ **(Figure 4-1A)**. Thus, I hypothesized it would be more amenable to crystallization.

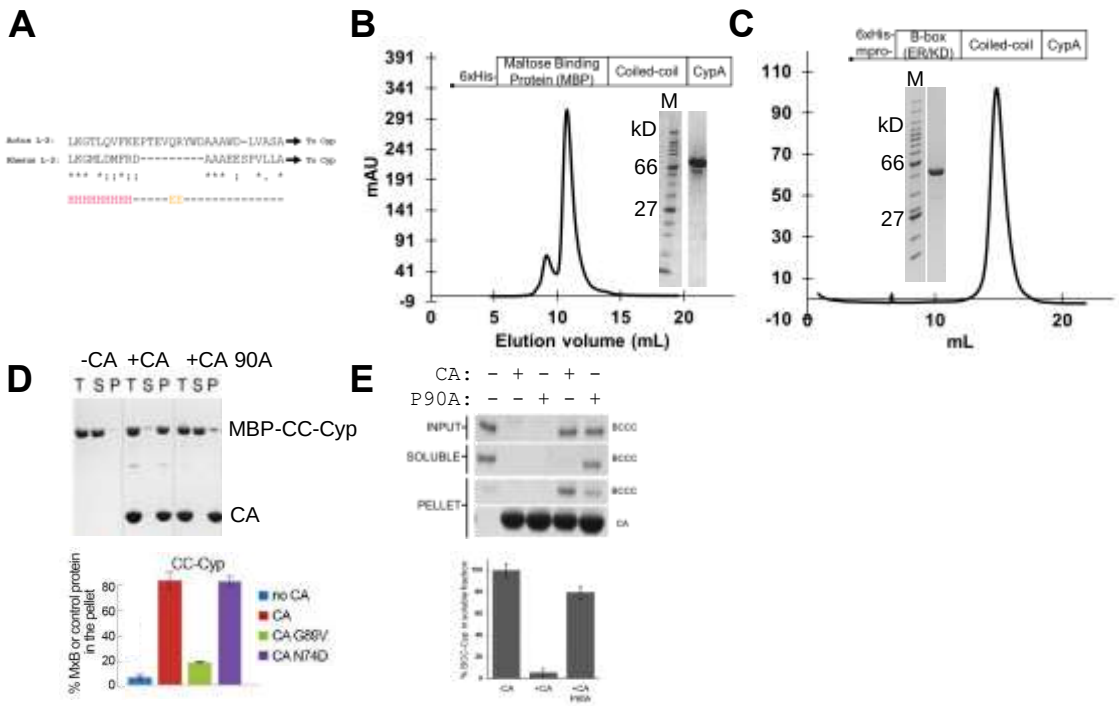


Figure 4-1. Purification and CA tube-binding of two different TRIMCyp genes. A) Sequence alignment of the L2 region of two TRIMCyp genes showing a 9-aa deletion in the *M. fascicularis* gene (lower sequence) compared to the *A. trivirgatus* sequence (upper). B) SEC chromatogram of *Aotus trivirgatus* 6xHis-MBP-CC-Cyp with SDS-PAGE inset. C) SEC chromatogram of *Macacca fascicularis* 6xHis-mpro-B-box (EK/RD)-CC-Cyp with SDS-PAGE of peak fraction inset. D) Cross-linked CA tube co-pelleting assay with *A. trivirgatus* MBP-CC-Cyp and quantification below (adapted from [12]). E) Cross-linked CA tube co-pelleting assay with *M. fascicularis* 6xHis-B-CC-Cyp and quantification below.

I purified TRIMCyp constructs containing the coiled-coil and cyclophilin domains (termed CC-Cyp) with N-terminal MBP fusions from both monkey species **(Figure 4-1B, *Aotus* version)**. These proteins were unstable and precipitated if the MBP tag was proteolytically removed. I added a mutated B-box domain (E120K/R121D, referred to hereafter as EK/RD) to produce 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$. These mutations were previously described and prevent B-box self-oligomerization [7, 58]. 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ expressed and purified better than MBP tagged CC-Cyp constructs **(Figure 4-1C)**. For all TRIMCyp constructs the cyclophilin domain tended to be proteolytically removed during and after purification. Thus, protease inhibitors, speed, and cold temperatures aided in TRIMCyp purification.

All TRIMCyp constructs co-pelleted strongly with disulfide-crosslinked CA tubes. To test if this binding was specific, I made a point mutation to CA and assessed binding. The CA P90A mutation is known to abolish or strongly reduce cyclophilin-CA interaction [85]. CA binding of both MBP-CC-Cyp$_{Aotus}$ and 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ was reduced when co-pelleted with P90A mutant CA tubes **(Figure 4-1D, E)**. For soluble binding assays I used either MBP-CC-Cyp$_{Aotus}$ or 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$, as these were the best behaving constructs from their respective monkey species.

4.2.2 Mapping the CA binding mode of TRIMCyp$_{Macaca}$ using novel capsid assemblies

How does TRIMCyp bind within a CA hexamer?

A soluble hexamer-TRIMCyp complex had not been described in the literature previously. Therefore, I tested the ability of my TRIMCyp constructs to bind CA hexamers. Both MBP-CC-Cyp$_{Aotus}$ and 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ strongly bound CA hexamers in SEC coelution assays **(Figure 4-2A, *Macaca* version)**. Based on stoichiometric analysis of the SEC complex and the accompanying SDS-PAGE, approximately two 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ dimers can bind one hexamer. It was unclear, however, how the two cyclophilin domains of TRIMCyp bind to the six CA molecules within a hexamer.

The novel partial-hexamer assemblies described in Chapter 2 allowed for a finer mapping of capsid binding by TRIMCyp. I used SEC co-elution assays to determine if 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ can recognize 1/3- or 1/2-hexamers. These constructs present only two or three adjacent CA subunits, respectively, present in a full hexamer, and may not be sufficient for B-CC-Cyp binding. Indeed, 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ co-eluted with both 1/3- and 1/2-hexamer-EE, which was comparable to its binding with complete hexamers **(Figure 4-2B, C)**.

We next wondered whether the tethered cyclophilin domains on 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ enable higher affinity binding to CA compared to single cyclophilin domains. Coelution of monomeric CypA domains with the same CA constructs at similar concentrations was significantly worse than dimeric cyclophilin domains on 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ (data not shown). The affinity of the monomeric CypA-hexamer interaction from isothermal titration calorimetry experiments (ITC) (~40 uM or weaker) likely precludes strong SEC

coelution **(Figure 4-2E, H)**. Coelution was also weak between 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ and native dimeric CA **(Figure 4-2D)**. We expect the observed apparent increase in CA affinity of dimeric 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ compared to monomeric CypA is due to avidity. Our data supports a model in which the two cyclophilin domains of 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ simultaneously bind two rigid CA molecules present in 1/3-, 1/2-, and complete hexamers.

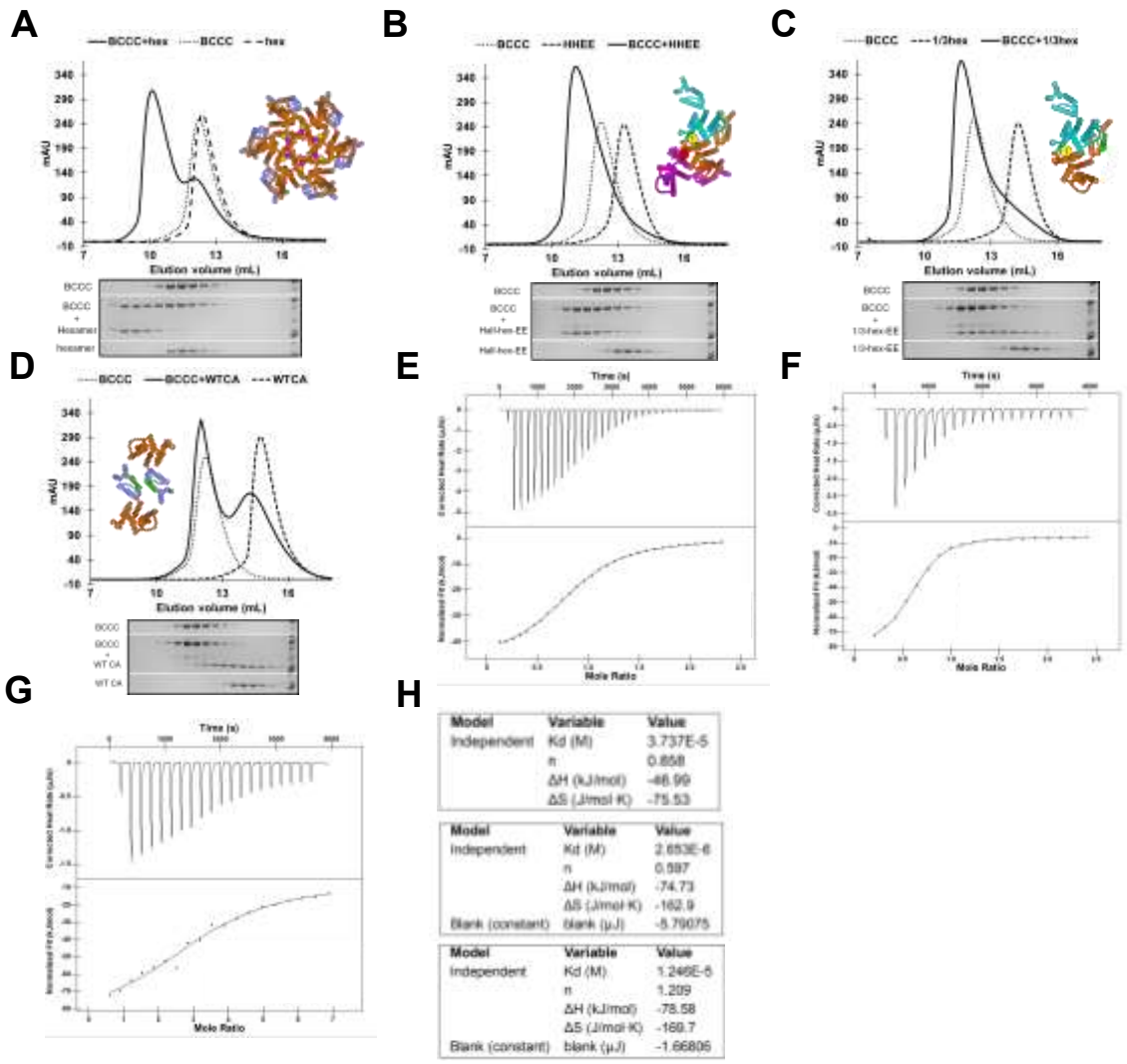


Figure 4-2. Small capsid assembly binding by *M. fasc* 6xHis-B-CC-Cyp. A) SEC coelution of disulfide-crosslinked hexamers and BCCCyp with SDS-PAGE below. B) SEC coelution of ½-hexamer-EE and BCCCyp with SDS-PAGE below. C) SEC coelution of 1/3-hexamer-EE and BCCCyp with SDS-PAGE below. D) SEC coelution of WT CA and BCCCyp with SDS-PAGE below. E) Representative ITC isotherm of BCCCyp-1/3-hexamer-EE complex. F) Representative ITC isotherm of BCCCyp-1/2-hexamer-EE complex. G) Representative ITC isotherm of *M. fasc* CypA with cross-linked hexamers. H) Representative fitted data from (E) top panel, (F) middle panel, (G) lower panel. All capsid models from PDB ID: 4XFX [4].

Next, we sought to determine the thermodynamic properties of TRIMCyp avidity using ITC **(Figure 4-2: F,G, H)**. In preliminary studies, we used 1/3-hexamer-EE and ½-hexamer-EE and assessed binding to 6xHis-B(EK/RD)-CC-Cyp_{Macaca}. The affinity of 6xHis-B(EK/RD)-CC-Cyp_{Macaca} to either construct was approximately 6-10 uM. When compared to the ~40 uM affinity of monomeric $CypA_{Macaca}$ described earlier, we observed at best an approximately 8-fold increase in B(EK/RD)-CC-Cyp_{Macaca} affinity to these constructs. We have not yet obtained triplicate data for 6xHis-B(EK/RD)-CC-Cyp_{Macaca} interaction with either 1/3-hexamer-EE or ½-hexamer-EE.

Even though 6xHis-B(EK/RD)-CC-Cyp_{Macaca} bound 1/3- and ½-hexamers with a much higher affinity than the monomeric CypA-hexamer interaction, we expected its absolute affinity to be higher. In SEC co-elution assays we observed complete binding even when protein concentration was below that used in the ITC experiments. Because of this apparently weaker ITC interaction, the ITC experiments required significant quantities of protein to achieve quality isotherms. We are considering numerous explanations (including binding conditions and data processing) to explain the weaker than expected binding affinity observed in ITC. Nonetheless, this is the first attempt at achieving a quantitative binding constant for TRIMCyp, and the early results are promising for future work.

Specific P90A mutations in ½-hexamer-EE enable a finer TRIMCyp binding-mode mapping

Our ability to make specific mutations in the 1/3- and 1/2-hexamer assemblies enabled a more in-depth mechanistic examination of capsid-TRIMCyp interactions. As previously reported, P90A mutations significantly reduced TRIMCyp-CA interaction. Therefore, we made single and multiple P90A mutations in the 1/2-hexamer-EE background to more finely map TRIMCyp binding. As a negative control, we mutated all CA positions of ½-hexamer-EE to P90A and assessed 6xHis-B(EK/RD)-CC-Cyp_{Macaca} binding. As expected, we observed no coelution of 6xHis-B(EK/RD)-CC-Cyp_{Macaca} with this construct **(Figure 4-3A).** Similarly, the 1/2-hexamer-EE with two P90A mutations, and thus only one native CypA binding loop, also demonstrated no 6xHis-B(EK/RD)-CC-Cyp_{Macaca} co-elution **(Figure 4-3B)**. With only one native CypA loop, 6xHis-B(EK/RD)-CC-Cyp_{Macaca} cannot bind this construct avidly.

In contrast, if 1/2-hexamer-EE contained two native CypA loops, complete 6xHis-B(EK/RD)-CC-Cyp_{Macaca} coelution was observed. A single P90A mutation at either the center (WT-90A-WT) or lateral (90A-WT-WT) CA position did not reduce 6xHis-B(EK/RD)-CC-Cyp_{Macaca} coelution as compared to complete hexamers or native 1/2-hexamer-EE **(Figure 4-3C, D)**. This suggests that 6xHis-B(EK/RD)-CC-Cyp_{Macaca} can avidly bind two CA molecules within a hexamer that are either adjacent or separated by one CA molecule. Furthermore, it can do so with little preference for one mode over the other.

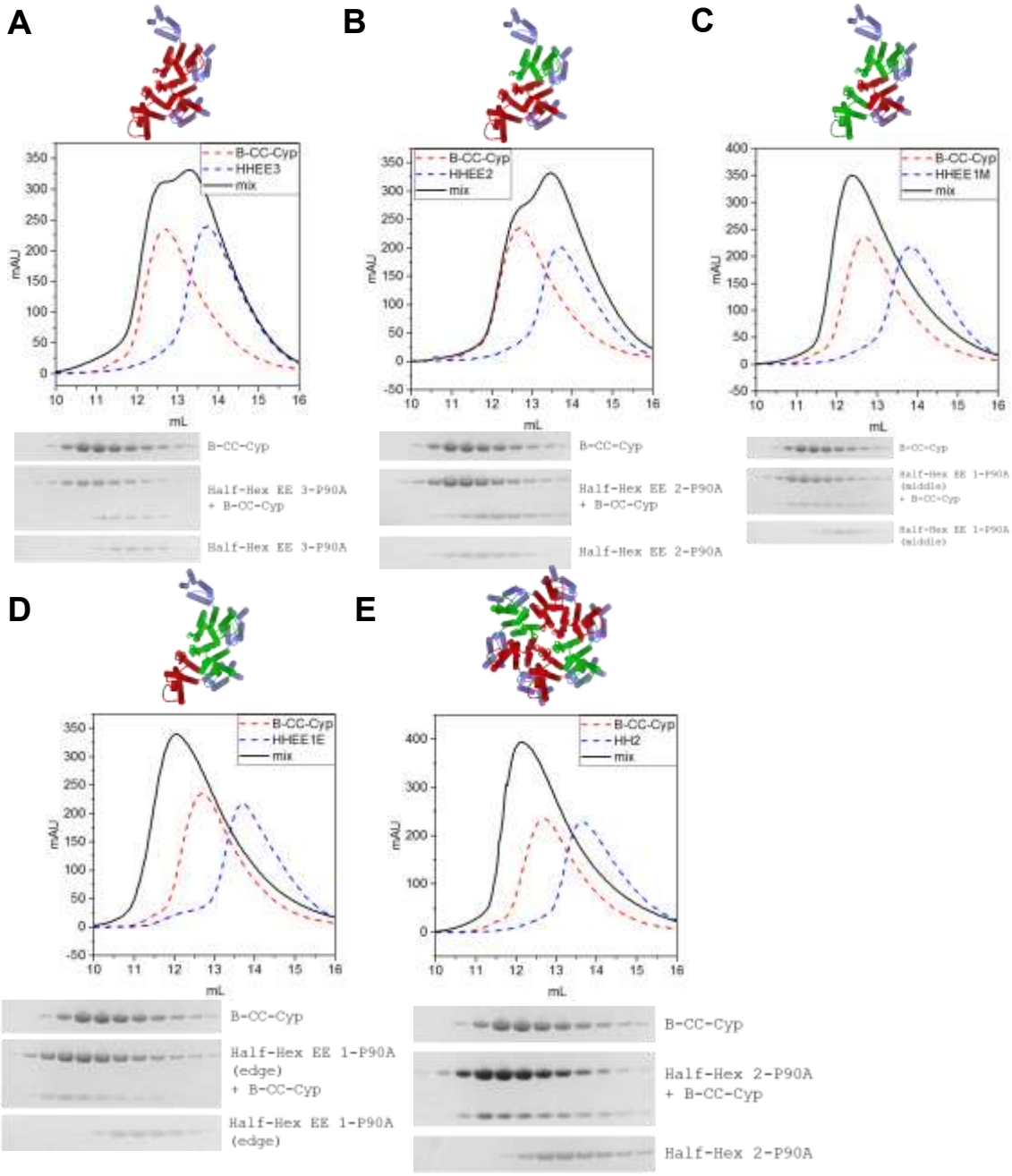


Figure 4-3. Using site-specific P90A mutations to map BCCCyp binding. Cartoon ½-hexamers and hexamers over each panel is colored as such: NTD with native CypA loop is green, NTD with P90A loop is red, CTD colored in slate. A) SEC coelution profile of BCCCyp with a ½-hexamer with all CA positions P90A. B) SEC coelution profile of BCCCyp with ½-hexamer with two P90A mutations. C) SEC coelution of BCCCyp with a ½-hexamer with a single central P90A mutation. D) SEC coelution of BCCCyp with a ½-hexamer with a single lateral P90A mutation. E) SEC coelution profile of BCCCyp with a ½-hexamer with two P90A mutations, thus strong binding only occurs if BCCCyp bridges two ½-hexamers into a hexamer-like structure. Corresponding SDS-PAGE gels are below each panel. All capsid models from PDB ID: 4XFX [4].

The aforementioned experiments tested CA binding of 6xHis-B(EK/RD)-CC-Cyp_{Macaca} within a ½-hexamer, but not explicitly the binding of 6xHis-B(EK/RD)-CC-Cyp_{Macaca} between two half hexamers. To test this possibility, we used the ½-hexamer construct that re-formed complete hexamers upon crystallization (section 2.2.2). This ½-hexamer lacks 42E/54E mutations ('-EE') that prevent re-hexamerization. In this 1/2-hexamer background, we made two P90A mutations (WT-90A-90A). We expected weak 6xHis-B(EK/RD)-CC-Cyp_{Macaca} coelution with this construct if it remained a ½-hexamer since it only contains one native CypA loop. However, we observed significant coelution between these constructs. The complex eluted in similar elution volume to that of the complex between hexamers and 6xHis-B(EK/RD)-CC-Cyp_{Macaca} **(Figure 4-3E)**. This data suggests that two 1/2-hexamers are forming a full hexamer, and presenting two native CypA loops. I can conclude that a 6xHis-B(EK/RD)-CC-Cyp_{Macaca} dimer may be able to bind two opposite, 180° rotated CA molecules within a hexamer.

Can TRIMCyp bridge multiple hexamers within the capsid lattice?

Although we demonstrated that TRIMCyp constructs can flexibly, yet avidly bind two CA molecules within a hexamer, a question still remained. Can TRIMCyp bridge multiple hexamers or is its binding restrained to one hexamer? To answer this question, we took advantage of the mixed protein composition of our di-hexamer assemblies. The edge subunits of di-hexamers are composed of $CA_{14C/45C/184A/185A}$ molecules, whereas the dimer interface between hexamers is composed of reacted CA-SpyTag and CA-SpyCatcher fusion molecules (see

section 3.2.1). Thus, each di-hexamer is composed of three different CA molecules, and each can be independently mutated to P90A.

We aimed to use this property of di-hexamers to test the ability of TRIMCyp to bind between two hexamers. As a negative control, we first mutated all di-hexamer CA positions to P90A to attempt to abolish TRIMCyp binding. Indeed, no co-elution was observed between this di-hexamer and 6xHis-B(EK/RD)-CC-Cyp_{Macaca} **(Figure 4-4A, D)**. Next, we reverted the CA-

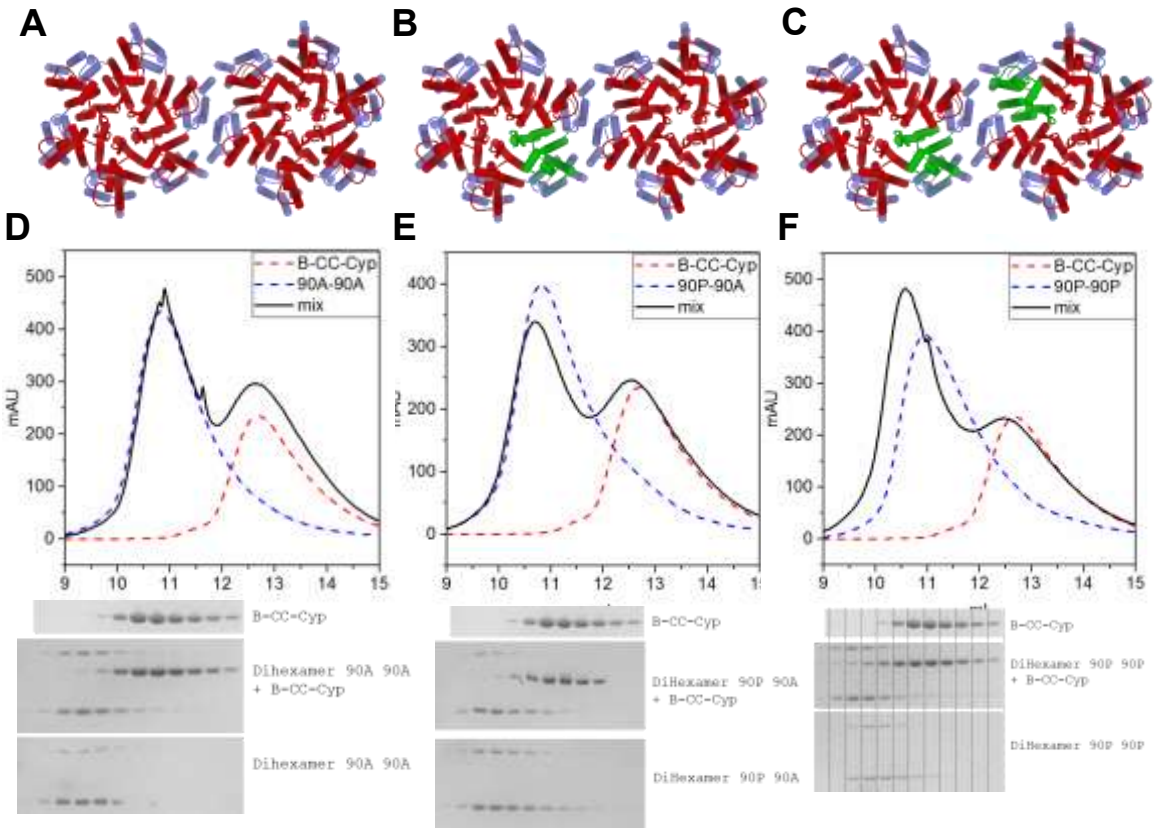


Figure 4-4. Using site-specific P90A mutations to map BCCCyp binding on di-hexamers. (A)-(C) Cartoon models representing the different di-hexamers tested. Each is colored as such: NTD with native CypA loop is green, NTD with P90A loop is red, CTD colored in slate. D) SEC coelution profile of BCCCyp with di-hexamers with P90A mutation at every CA position. E) SEC coelution profile of di-hexamers with a single wild-type CypA loop (all others are P90A). F) SEC coelution profile of BCCCyp with di-hexamers containing two native CypA loops (all others are P90A). Corresponding SDS-PAGE gels for coelution assays are below each figure. All capsid models from PDB ID: 4XFX [4].

SpyTag fusion molecule from P90A back to its native form (P90), so that the di-hexamers would have only a single WT CypA binding loop. Only very weak 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ co-elution was observed with this di-hexamer construct **(Figure 4-4B, C)**. Finally, we reverted both CA-Spy fusions to WT CypA loops. These newly assembled di-hexamers would have precisely two WT CypA loops, one on each hexamer centered at the native dimerization interface. 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ strongly co-eluted with these di-hexamers **(Figure 4-4C, F)**.

From these results, we propose that TRIMCyp is capable of avidly binding two CA molecules on neighboring hexamers. This further reinforces the idea that the cyclophilin domains of TRIMCyp are not rigidly oriented relative to their coil-coiled domains and can avidly bind two CA molecules regardless of their distance and orientation.

4.2.3 Purification of natively oligomeric TRIMCyp constructs containing the RING and B-box domains

Self-oligomerization of TRIM5 RING and B-box domains is critical for the hexagonal assembly of the TRIM5 lattice on the capsid surface. However, at the beginning of my dissertation work it was unclear what the native oligomeric states of these domains were. I attempted to simultaneously co-express MBP-RING-B-CC-Cyp$_{Macaca}$ and 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ in the hopes that the two constructs would form coiled-coil-mediated heterodimers. The RING and B-box domains of MBP-RING-B-CC-Cyp$_{Macaca}$ are fully wild-type and should form natural oligomers. The 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ construct lacks a RING

domain and has a mutated B-box domain so it is not capable of further oligomerization **(Figure 4-5: A)**.

Both proteins expressed well, and I was able to purify MBP-RING-B-CC-Cyp/6xHis-B(EK/RD)-CC-Cyp heterodimers by using (in order): nickel affinity, MBP affinity, anion exchange, and SEC. This enabled us to observe the oligomerization state of the single RING and B-box domains on MBP-RING-B-CC-Cyp_{Macaca}. The purified heterodimer was a large, concentration-dependent oligomer **(Figure 4-5: B, B-CC-Cyp/B(EK/RD)CC-Cyp version)**. By using a combination of SEC and SEC-MALS, it appeared that the heterodimer further oligomerized into a dimer, and perhaps a trimer at high concentrations, mediated by the wild-type RING and B-box domains **(Figure 4-5: C, R-B-CC-Cyp/B(EK/RD)CC-Cyp)**.

We analyzed the oligomeric heterodimer using negative-stain EM **(Figure 4-5: D)**. Noodle-like two- and three-fold spokes were observed that perhaps correlate to fragments of the TRIM5 hexagonal lattice. The oligomers appeared flexible and were generally hard to observe due to the thin diameter of the coil-coils.

Soon after these experiments were performed crystal structures and solution oligomeric state analysis of dimeric and trimeric B-box domains were published using a similar approach [124-126]. The experimenters used truncated coiled-coil domains attached to native B-box domains. They proposed that the trimeric B-box domains form the three-fold spoke in the hexagonal TRIM5 lattice. Additionally, a crystal structure of a dimeric RING domain was solved [127].

These data supported my preliminary findings and confirmed that I was on the right track to better understand TRIM5 oligomerization.

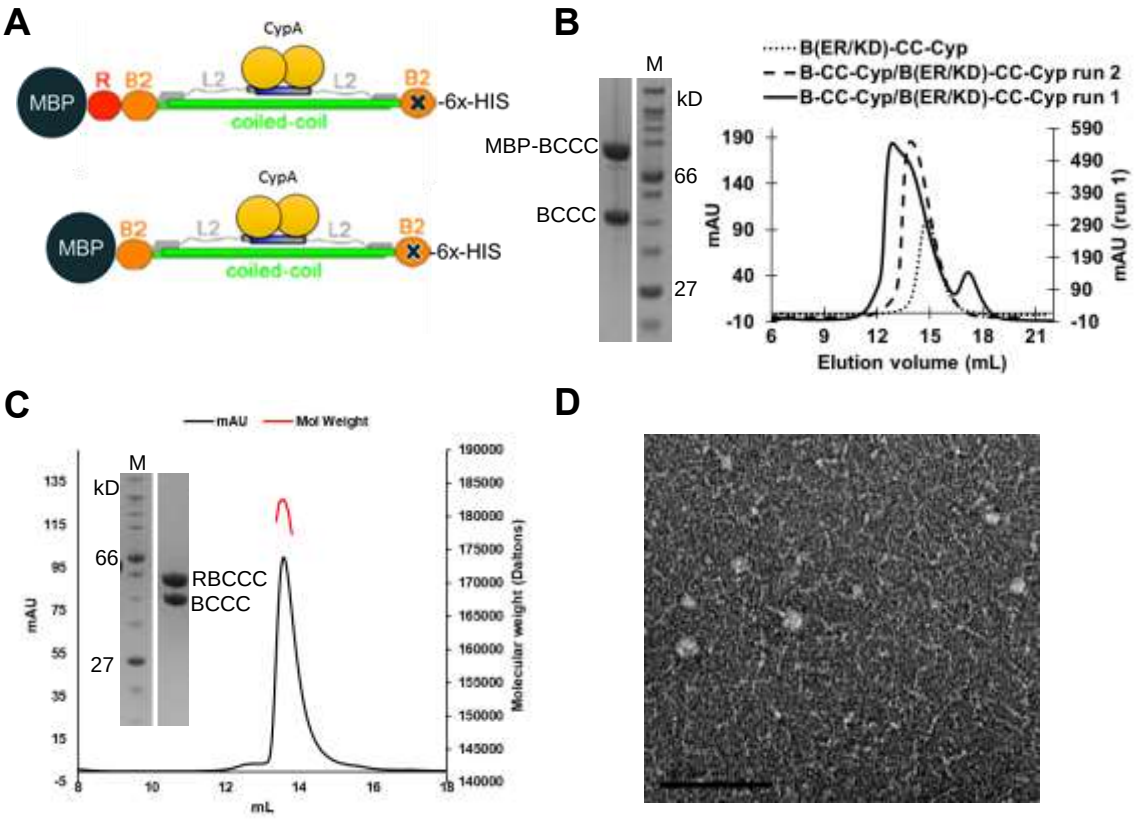


Figure 4-5. Capturing natively oligomeric TRIMCyp constructs. A) Model of two different TRIMCyp heterodimers. B) SEC overlays of first high concentration B-CC-Cyp/B(EK/RD)CC-Cyp heterodimer run, second lower concentration run, and B(EK/RD)CC-Cyp homodimer run. SDS-PAGE showing heterodimer purity is inset. C) SEC-MALS analysis of R-B-CC-Cyp/B(EK/RD)CC-Cyp heterodimer, with SDS-PAGE inset. D) Negative-stain EM analysis of R-B-CC-Cyp/B(EK/RD)CC-Cyp heterodimer demonstrating triangular spokes.

4.2.4 TRIMCyp-CA structure determination attempts

With the aforementioned biochemistry and well-behaving TRIMCyp constructs in hand, I felt prepared to pursue the structure of TRIMCyp and CA. I performed significant crystallization screening using either MBP-CC-Cyp$_{Aotus}$ or 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ in complex with CA hexamers. The complexes were usually assembled and purified using SEC prior to screening. Drops showed ~30-50% precipitation when total protein concentration was between 0.5-1.5 mg/mL, indicating good screening concentrations. Unfortunately, several unforeseen roadblocks hindered crystallization of the complex. For one, hexamers crystallize readily on their own, so at least 50-100 conditions would have apparent apo-hexamer crystals. Two, the relatively weak TRIMCyp-hexamer binding affinity (high nanomolar to low micromolar) meant that the complexes were likely to dissociate frequently. Three, based on our analysis in section 4.2.2, we believe the cyclophilin domains of TRIMCyp are flexible in relation to the coiled-coil domains, thus greatly limiting chances of crystallization. Four, the cyclophilin domains tended to be proteolyzed in crystallization conditions. Finally, disulfide-bonded TRIMCyp oligomers formed in non-reducing conditions over time.

I performed similar crystallization screening with 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ in complex with 1/3-hexamer-EE and ½-hexamer-EE. Additionally, I used ½-hexamer-EE assemblies with P90A mutations at specific sites to attempt to make more homogenous TRIMCyp-CA complexes. Unfortunately, none of my crystallization trials produced crystals of a TRIMCyp-CA complex.

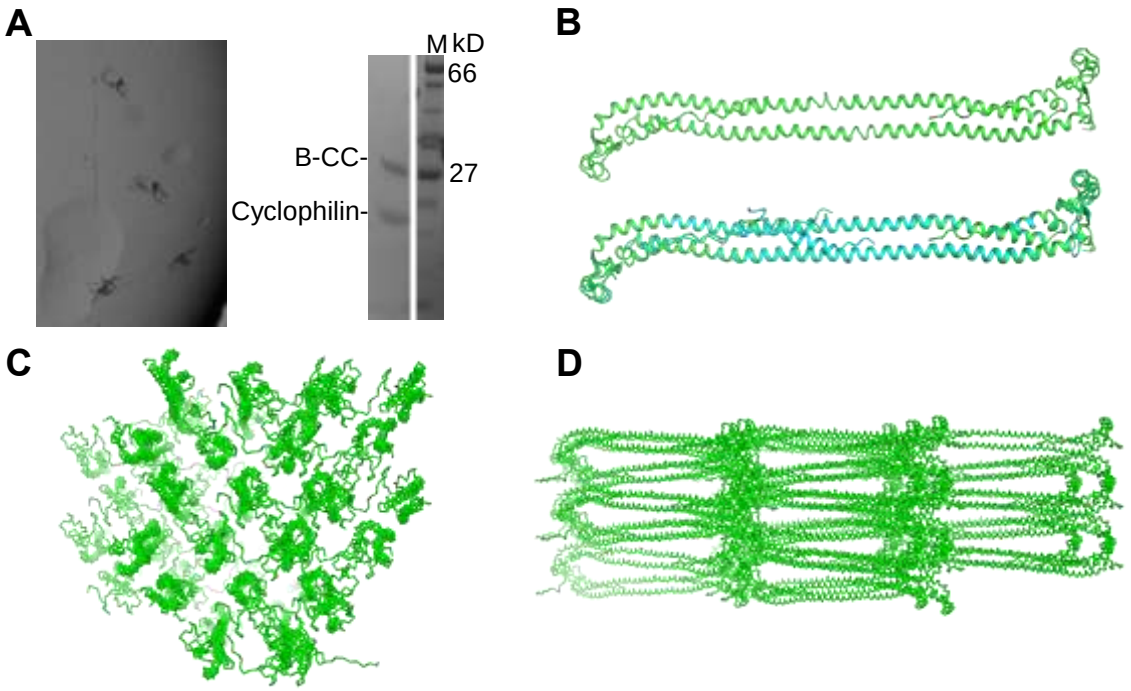


Figure 4-6. Crystal structure of dimeric *M. fascicularis* B-box(EK/RD)-coiled-coil. A) Left panel: initial crystals hits of B-box(EK/RD)-coiled-coil. Right panel: SDS PAGE showing cyclophilin is cleaved from coiled-coil. B) Crystal structure *M. fasc* B-box(EK/RD)-coiled-coil overlaid with prior *M. mulatta* TRIM5 coiled-coil structure (PDB ID: 4TN3 [7]). C) Crystal packing of B-box(EK/RD)-coiled-coil structure demonstrating small channels for potential CypA domains. D) A rotated view of crystal packing from (C).

Because the cyclophilin domains of 6xHis-B(EK/RD)-CC-Cyp$_{Macaca}$ were so prone to degradation, I frequently found broom-like crystals containing only the B-box-coiled-coil domains **(Figure 4-6: A)**. I optimized these crystals and solved a 2.5 Å structure. Collection and refinement statistics are summarized in **Table 4-**1. The structure aligned well with the prior TRIM5 B-box-coiled-coil structure (RMSD of 1.6Å) **(Figure 4-6: B)** [7]. There was likely not enough room

Data Collection	
Wavelength (Å)	1.11
Space Group	P1211
Cell Dimensions	
a, b, c (Å)	50.06, 50.11, 139.34
α, β, γ (°)	90, 99.31, 90
Molecules/asymmetric unit	2
Resolution (Å)	49.1-2.5 (2.6-2.5)
R_{merge}	0.109 (0.729)
$I/\sigma I$	8.4 (1.6)
Completeness (%)	96.5 (96.9)
Redundancy	2.8 (2.6)
Unique reflections	17868 (676)
Refinement	
Number of nonhydrogen atoms	2924
R_{work}/R_{free} (%)	26.6/28.8 (33.5/31.7)
Average B factor	43.4
Root mean-squared deviation (rmsd)	
Bond lengths (Å)	0.004
Bond angles (°)	0.75
Ramachandran analysis	
Preferred regions (%)	91.7
Allowed regions (%)	5.9
Outliers (%)	0.37

Table 4-1. Data collection and refinement statistics for *Macacca fascicularis* TRIMCyp B-box(EK/RD)-coiled coil structure. Statistics in parentheses indicate those for the highest resolution shell. These statistics are current as of July 2018 but may change upon publication.

in the solvent channels to accommodate cyclophilin domains **(Figure 4-6: C, D)**. Despite this, I attempted to grow crystals in conditions where the cyclophilin domains were not degraded, but to no avail.

The structure does not appear to provide significant novel biological information. It does confirm the prior TRIM5 B-box-coiled-coil structure derived from *Macaca mulatta.* There are approximately ten mutations between the two species within the constructs crystallized, but these do not lead to apparent structural differences.

4.3 FEZ1 results

4.3.1 Expression and purification of FEZ1

To begin analysis of the putative CA-binding cofactor FEZ1, along with Ivy Huang, we expressed full-length and truncated FEZ1 fragments with an N-terminal 6xHis-tag. Expression and purification were generally standard, except that FEZ1 fragments containing FEZ1(198-392) were prone to degradation and required protease inhibitors and speed. All FEZ1 constructs were very soluble, likely due to their high percentage of negative charges. They eluted at higher molecular weights than predicted on SEC indicating they are extended and unfolded **(Figure 4-7: A, B)**. Ivy performed SEC-MALS analysis and confirmed that the tested FEZ1 constructs are monomeric (data not shown).

Ivy Huang purified numerous FEZ1 truncations and tested their ability to bind CA tubes. More details of these experiments will be in a soon to be published paper. Of most significance to this thesis, FEZ1(92-198) bound CA tubes with apparent similar affinity as full length FEZ1(1-392) **(Figure 4-7: C, D)**.

Importantly, this was the first demonstration of a direct capsid interaction using purified FEZ1. The ionic strength of the buffer was critical to observe FEZ1-CA interaction. Low ionic strength buffer conditions (50-75 mM NaCl) provided the most complete binding in pelleting assays. We chose to use the FEZ1(92-198) truncation fragment in SEC binding tests due to its higher yield and better solution behavior compared to full-length FEZ1.

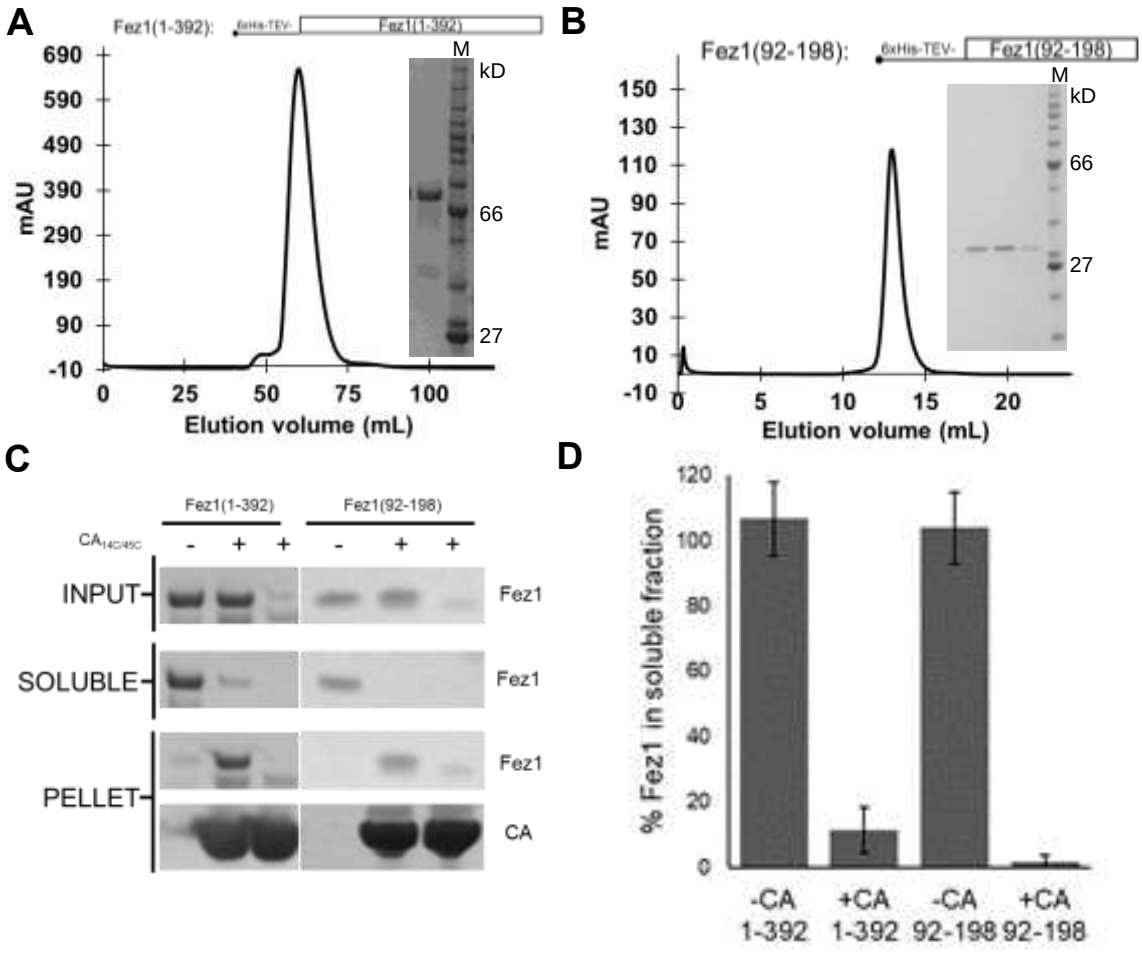


Figure 4-7. Purification of FEZ1 constructs and CA tube co-pelleting. A) SEC chromatogram of 6xHis-FEZ1(1-392) purification with SDS-PAGE of peak inset. B) SEC chromatogram of 6xHis-FEZ1(92-198) with SDS-PAGE inset. C) FEZ1 co-pelleting with cross-linked CA tubes. D) Quantification of pelleting in (C).

4.3.2 FEZ1 targets the hexameric CA lattice

To our surprise, in SEC coelution assays FEZ1(92-198) did not co-elute with CA dimers, but displayed strong and complete co-elution with hexamers **(Figure 4-8: A, D)**. Based on our observations, FEZ1 appeared to bind CA hexamers with potentially the highest affinity among the currently known capsid-interacting proteins. Analysis of the FEZ1-hexamer complex by SDS-PAGE revealed that approximately one FEZ1 molecule bound one complete CA hexamer. A lack of FEZ1 co-elution with CA dimers, but strong recognition of hexamers, demonstrates a capsid-lattice sensing ability of FEZ1. This is the first identification of FEZ1 as a novel capsid pattern sensor that likely targets CA hexamers.

Although FEZ1 bound hexamers with a high affinity that was comparable to TRIMCyp, it diverged from TRIMCyp in its behavior towards 1/3- and 1/2-hexamers. FEZ1 displayed essentially no co-elution with 1/3-hexamer-EE or 1/2-hexamer-EE **(Figure 4-8: B)**. We further tested FEZ1 binding with our native 1/2-hexamer. Without 42E/54E mutations, this 1/2-hexamer readily crystallized as a complete hexamer even though it appeared discrete in solution (see section 2.2.2). Surprisingly, we observed a strong FEZ1 coelution with this assembly **(Figure 4-8: C)**. The large elution profile shift by the FEZ1-1/2-hexamer complex was comparable to the shift observed in the hexamer-FEZ1 complex. Based on predicted SEC molecular weights, we believe FEZ1 may be able to bridge two 1/2-hexamers in a hexamer-like orientation. These data suggest that FEZ1 may require a complete hexameric surface for binding.

Further analysis of the FEZ1-CA interaction will be in Ivy Huang's thesis and a soon to be published paper.

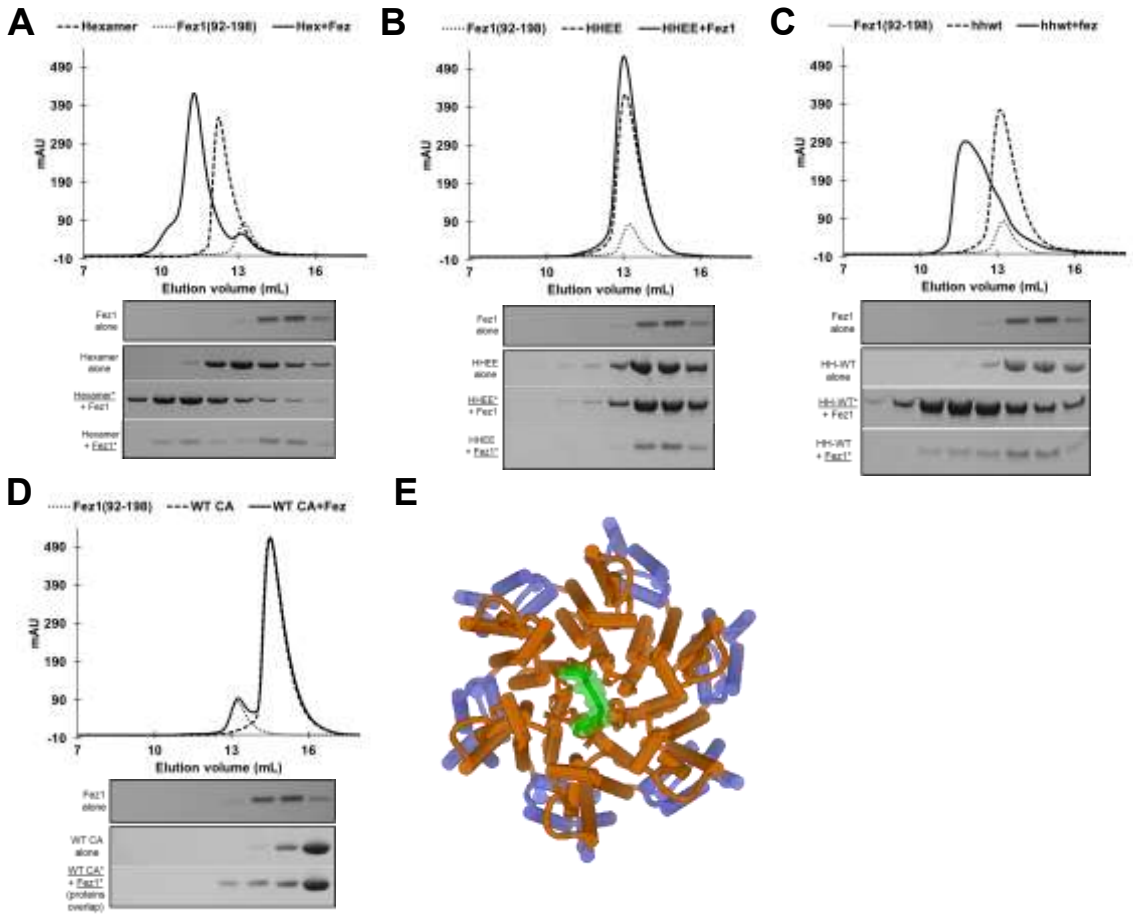


Figure 4-8. FEZ1 targets the six-fold hexamer surface. A) Hexamer-FEZ1(92-198) SEC coelution assay and corresponding SDS-PAGE. Underlined proteins indicate the band shown on the adjacent SDS-PAGE analysis. B) ½-hexamer-EE-FEZ1(92-198) SEC coelution assay and corresponding SDS-PAGE. C) ½-hexamer-FEZ1(92-198) SEC coelution assay and corresponding SDS-PAGE. D) WT CA-FEZ1(92-198) SEC coelution assay and corresponding SDS-PAGE. E) Model of FEZ1 peptide in green surface binding to a CA hexamer (PDB ID: 4XFX [4]).

4.3.3 MRI-1 may bind CA hexamers with a similar mechanism as FEZ1

Ivy Huang identified a poly-glutamate stretch of FEZ1 (residues 178-188) potentially responsible for its CA binding ability. This mapping data will be further described in a soon to be published work. I used BlastP to attempt to identify other human proteins that have a similar sequence [128]. The highest similarity protein was modulator of retrovirus infection-1 (MRI-1) **(Figure 4-9: A)**. When overexpressed this protein can improve HIV-1 infectivity in certain cell-lines in a poorly understood fashion [129]. Like FEZ1, it is cytoplasmic and is predicted to be natively unstructured [130].

I received the MRI-1, MRI-2, and MRI-3 genes from professor Sarah Slavoff at Yale University. I cloned, expressed, and purified 6xHis-MRI-1, along with similar constructs of MRI-2 and MRI-3 **(Figure 4-9: B)**.

In a preliminary analysis, in similar conditions to FEZ1 (25 TRIS pH 8, 50 mM NaCl), 6xHis-MRI-1 co-pelleted strongly with disulfide-crosslinked CA tubes **(Figure 4-9: C)**. In an SEC coelution assay, the presence of 6xHis-MRI-1 shifted the elution profile of CA hexamers to a larger molecular weight **(Figure 4-9: D)**. This suggests that MRI-1 directly interacts with hexamers in vitro.

Interestingly, MRI-2 lacks the potential capsid-interacting poly-glutamate motif and could serve as an important non-capsid binding control. Due to time constraints, I was unable to test MRI-2 or MRI-3 in CA binding assays.

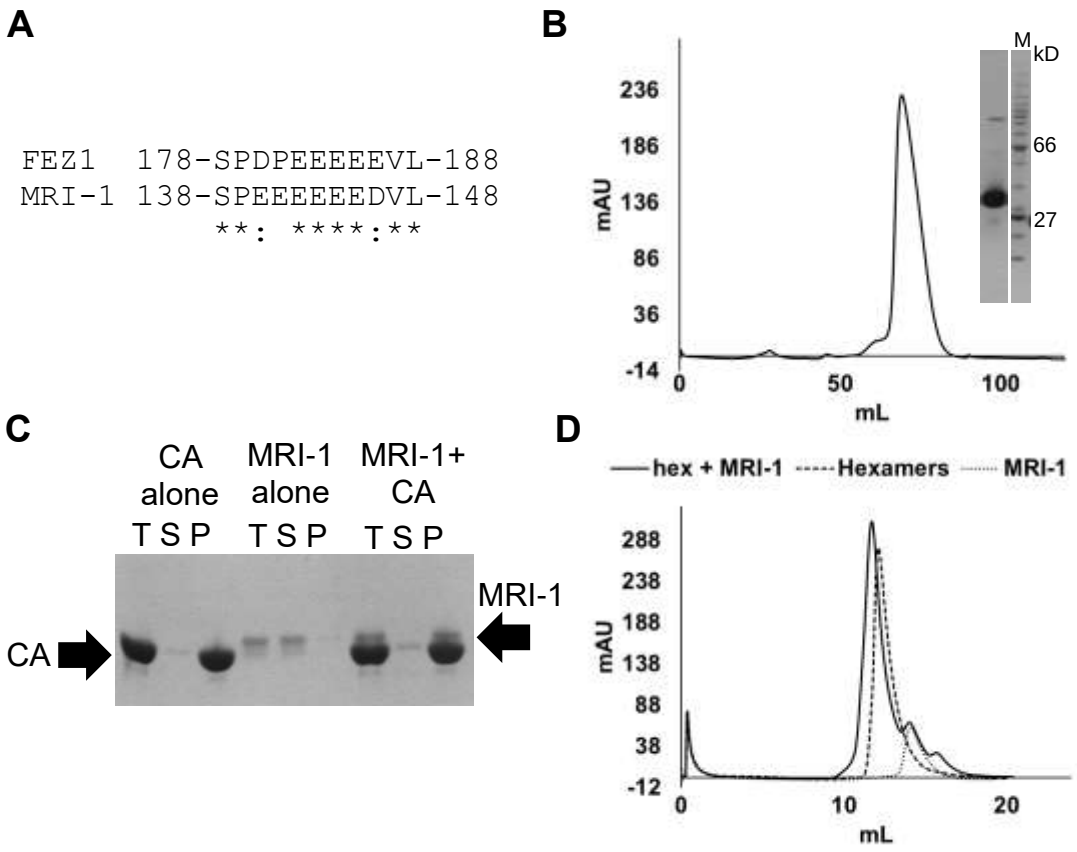


Figure 4-9. MRI-1 binds CA. A) Sequences alignment of the FEZ1 motif perhaps responsible for CA binding and a similar stretch in MRI-1. B) SEC purification of MRI-1 and SDS-PAGE. C) MRI-1 co-pellets with CA tubes. D) MRI-1 coelutes with hexamers on SEC

4.3.4 Preliminary structure determination of hexamer-CPSF6$_{313-327}$ and hexamer-PF-74

Early in my dissertation work I wanted to determine if the host factor CPSF6 or the small molecule inhibitor PF-74 altered the capsid lattice. At the time, a structure of a CPSF6 peptide (CPSF6$_{313-327}$) in complex with the CA NTD, as well as an NTD-PF-74 complex, had been solved [3, 131]. Binding by these two factors did not appear to alter capsid structure. However, there were unpublished reports that CPSF6 stabilized the capsid lattice during infection. Conversely, there were reports demonstrating a capsid destabilizing effect of PF-74 [132]. I sought to determine the structure of both factors in complex with disulfide-bonded hexamers. By using full-length capsid arranged in a lattice architecture, I hoped to observe lattice-altering effects caused by the binding of the two factors.

I grew hexamer-CPSF6$_{313-327}$ co-crystals at 2 mg/mL hexamers with 3 mM CPSF6$_{313-327}$ peptide (peptide purchased from ChinaPeptide) at room temperature using the microbatch under oil method. The best crystals grew in a precipitant solution of 0.2 M sodium formate, 20% PEG 3350. Simultaneously, I grew hexamer-PF-74 co-crystals with 2 mg/mL hexamers and 1 mM PF-74 (purchased from Sigma). The best crystals formed in 0.2 M sodium fluoride, 20% PEG 3350. Crystals of both were cryo-protected in 25% glycerol and frozen. Data was collected at NE-CAT beamline APS-IDE at the Advanced Photon Source. Preliminary and partial crystallization statistics are in **Figure 4-10: A (PF74**

only). The structures were solved using molecular replacement with a model derived from PDB ID: 3H47 [16].

From early analysis, it appeared that the hexamer-CPSF6$_{313-327}$ complex was identical to that observed in the NTD only structure (data not shown). There were differences in the hexamer-PF-74 structure, however. The PF-74 indole group was flipped almost 180°, and additional hydrogen bonds between it and CA were observed **(Figure 4-10: B)**.

Unfortunately, shortly after I solved these structures several groups published the identical structures [5, 6]. The published structures validated my preliminary findings. Importantly, these publications quantitatively measured that both CPSF6$_{313-327}$ and PF-74 bind hexamers with at least an order of magnitude higher affinity than the NTD alone. This was the first demonstration that CPSF6 is a hexamer-lattice targeting factor.

I naturally stopped working on this project when the competing structures were published. Fortunately, I only spent approximately three weeks from the time I received the peptide and PF-74 until structure determination. The experience helped me garner valuable crystallization knowledge, however.

A

Crystallographic Statistics	
Spacegroup:	P6
Unit cell a,b, c (Å):	86.9, 86.9, 55.9
Unit cell α,β, γ (°):	90.0, 90.0, 120.0
Resolution range:	50.0-2.10
Unique Reflections:	13300(520)
Completeness:	93.2 (74.2)
Redundancy:	2.4 (2.0)
I/sigma:	12.8 (0.66)
Rmerge:	0.056 (0.942)
Rwork/Rfree:	0.213/0.262
Bond lengths:	0.017
Bond angles:	2.281

B

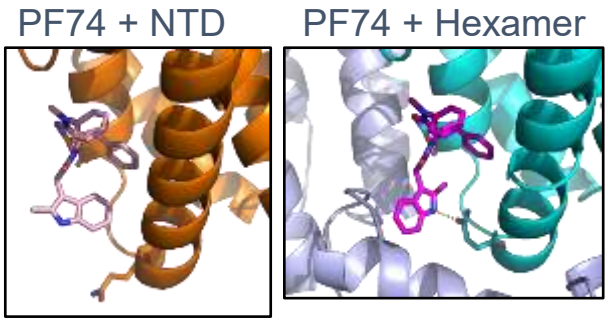


Figure 4-10. Hexamer-PF74 co-crystallization and structure. A) Preliminary crystallization statistics of the hexamer-PF74 structure. B) My hexamer-PF74 structure viewed side-by-side with the prior NTD only structure (PDB ID: 2XDE [3]).

4.4 Discussion

I developed expression and purification procedures to produce large amounts of well-behaving TRIMCyp constructs (including coiled-coil and cyclophilin domains) from two monkey species: *Aotus trivirgatus* and *Macaca fascicularis.* These bound CA tubes with high affinity and were excellent tools to use in solution-based biochemical assays and structural biology experiments. Along with Ivy Huang, we identified human FEZ1 truncation fragments that bound CA tubes with high affinity. This was the first demonstration of a direct FEZ1-CA interaction.

We demonstrated that dimeric TRIMCyp constructs and FEZ1 were able to form stable complexes in solution with CA hexamers. Neither TRIMCyp nor FEZ1 formed stable complexes with native CA dimers. This was the first description of a hexamer-targeting binding mode of these two factors.

While both TRIMCyp and FEZ1 strongly bound CA hexamers, we were able to show that they use very distinct binding modes. TRIMCyp appears to use its two cyclophilin domains to bind simultaneously any two CA molecules within a hexamer. TRIMCyp can also bind CA molecules across neighboring hexamers. It appears that TRIMCyp has an extremely flexible binding-mode. TRIMCyp was able to avidly bind numerous pairs of CA molecules separated by different distances and orientations.

FEZ1, conversely, appears to only recognize complete, or largely complete, CA hexamers. It showed no affinity towards 1/3-hexamer-EE or ½-

hexamer-EE. FEZ1 appeared capable of stabilizing two native 1/2-hexamers in a complete hexamer-like conformation. This suggests that FEZ1 may cross the midpoint of the CA hexamer as part of its binding path. Whether FEZ1 remains disordered or adopts a stable three-dimensional structure upon CA binding remains to be determined. A binding mode of this nature has not been demonstrated in previously analyzed capsid-binding host proteins. The FEZ1 binding mode may represent a new class of co-factor binding mechanisms that specifically sense CA hexamers.

Interestingly, recent publications reported that dNTPs, NTPs, and the small molecule IP6 bind the hexamer center with high affinity, and that dTNPs may be transported through the small hexamer pore into the capsid interior [133-135]. Along with the preliminary MRI-1-hexamer binding data, this suggests that the six-fold hexamer center may be a capsid-binding hotspot.

4.5 Future directions

Obtaining high-resolution structures of TRIMCyp, FEZ1, or MRI-1 in complex with a hexamer or partial-hexamer subassembly are of the highest priority. The numerous small capsid assemblies described in previous chapters provide many possibilities for crystallization or cryo-EM analysis.

Solving these structures will likely be challenging, however. As alluded to in Section 4.2.4, the likely flexible cyclophilin domains of TRIMCyp may impede structural studies. FEZ1 has its own unique issues. We believe that one FEZ1 molecule binds one hexamer. In preliminary crystallization trials by Ivy Huang (personal communication) hexamers crystallize readily in the presence of a FEZ1

peptide. However, the crystals generally have three- or six-fold crystallographic symmetry. This means that a FEZ1 peptide in the crystal lattice could be in one of three or six orientations in relation to each hexamer, and, thus, FEZ1 peptide density may be averaged out or blurred. This may make it very challenging to observe unambiguous FEZ1 electron density. FEZ1 may be better suited to structure determination using cryo-EM or solid-state NMR.

Future optimization of TRIMCyp-1/2-hexamer ITC conditions may better inform the thermodynamic properties of TRIMCyp-CA binding. With more thorough analysis, it may be possible to identify a preferred CA pair recognized by a TRIMCyp dimer.

4.6 Materials and methods

4.6.1 Cloning

The *Aotus trivirgatus* CC-Cyp gene fragment was a gift from Jinwoo Ahn at the University of Pittsburgh. I cloned it into pMAT9S to provide it with a N-terminal 6xHis and MBP tag. An mpro protease cleavage site was present between MBP and CC-Cyp. I synthesized the *Macaca fascicularis* CC-Cyp domain DNA sequence with Genscript. I used overlapping PCR to append RING, B-box and cyclophilin domains to the coiled-coil domain (sequences derived from TRIM5α and CypA genes from the NIH AIDS Reagent Program). I cloned these fragments into pMAT9S and pRSF vectors to make N-terminal MBP tagged and N-terminal His tagged constructs, respectively. The human FEZ1 cDNA sequence was purchased from Dharmacon and cloned into the pET28 vector with an N-terminal His-tag, thrombin cleavage site, and TEV cleavage site to

make full-length FEZ1 (1-392) or truncations. I received MRI-1, MRI-2, and MRI-3 genes from Sarah Slavoff at Yale. I cloned these into the pRSF vector with an N-terminal 6xHisTag. All capsid clones used in this chapter were cloned as described in sections 2.5 and 3.5.

4.6.2 Expression and purification

To express all TRIMCyp constructs I used chemically competent *E. coli* BL-21(DE3) cells transformed using the heat shock method with the plasmid of interest. Generally, 1.5L flasks of Terrific broth were used for protein expression. Cells were grown to an OD of 0.6-1 and induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 18°C overnight. This expression procedure generally yielded 5-10 mg of purified protein.

Cells were harvested by centrifugation and lysed by a microfluidizer with Roche protease inhibitor tablets. The lysate was centrifugally clarified at 13500 rpm for 35 minutes. All TRIMCyp constructs were purified by a gravity nickel column, followed by anion exchange (proteins generally eluted between 200-300 mM NaCl), mpro digestion for 2+ hours if the MBP tag is to be removed, and a polishing GE S200PG. The final SEC column was run in 50 mM TRIS pH 8, 300 mM NaCl, 0.5 mM phenylmethane sulfonyl fluoride (PMSF). The purity after each step was analyzed by SDS-PAGE.

To purify TRIMCyp heterodimers (such as MBP-RING-B-box-CCCyp with 6xHis-B-box(EK/RD)-CC-Cyp) the MBP tagged TRIMCyp construct in the pMAT9S vector (without N-terminal 6xHis-tag) was co-transformed with the 6xHis-tagged construct in the pRSF vector. The proteins were expressed as with

other TRIMCyp constructs. They were purified using a gravity nickel column, followed by an MBP affinity column, then anion exchange column, mpro digestion for 2+ hours to remove MBP tag, and a final S200PG run in 50 mM TRIS pH 8, 300 mM NaCl, 0.5 mM PMSF.

FEZ1 and MRI-1 constructs were expressed and cells were lysed using the same protocol as with TRIMCyp constructs. They were purified using a gravity nickel column, anion exchange column (protein eluted at more than 300 mM NaCl), and S200PG run in 50 mM TRIS pH 8, 300 mM NaCl.

4.6.3 CA-tube co-pelleting assays

14C/45C disulfide-stabilized CA tubes were assembled by overnight dialysis in 50 mM TRIS pH 8, 1M NaCl at approximately 15 mg/mL. This was followed by overnight dialysis into 50 mM TRIS pH 8. Approximately 1-5 uM purified host factors were mixed with approximately 75-150 uM CA tubes and incubated with shaking at room temperature for 30 minutes. Tubes were pelleted by centrifugation at 4 °C at 14000 rpm for 10 minutes. Total, soluble, and pellet fractions were taken at appropriate times and analyzed via SDS-PAGE. FEZ1 and MRI-1 co-pelleting was performed in 25 mM TRIS pH 8, 50 mM NaCl. TRIMCyp pelleting assays were performed in 25 mM TRIS pH 8, 150 mM NaCl.

4.6.4 Size-exclusion chromatography (SEC) coelution assays

Host factors were mixed with CA assemblies for 30 minutes to 1 hour on ice in the same buffer used for CA-tube co-pelleting assays. For TRIMCyp assays, mixtures were in a 500 uL volume with 36.6 uM TRIMCyp monomers and 27.5 uM CA monomers (of the appropriate CA assembly). For FEZ1 and

MRI-1 binding tests I used 200 uL reaction volumes with 32.6 uM FEZ1/MRI and 97.7 uM CA monomer (of the appropriate assembly). All binding tests were performed on GE S200 10/300 GL columns.

4.6.5 ITC experiments

All ITC experiments were performed using a TA Instruments NanoITC machine in the Yale Chemical Biology Instrumentation Center. Data was analyzed using the provided NanoAnalyze software. All curves were fit with an independent one-site binding model. $CypA_{Macaca}$ ITC experiments were performed in 25 mM phosphate buffer pH 7, 100 mM NaCl at 20 °C. TRIMCyp experiments shown here were performed in 25 mM TRIS pH 8, 150 mM NaCl at 4 °C. TRIMCyp constructs precipitated in phosphate buffer pH 7, necessitating the use of TRIS pH 8 buffer.

4.6.6 Negative stain EM

All negative stain EM experiments were performed as described in section 2.5.4. Oligomeric TRIMCyp heterodimer constructs were in 50 mM TRIS pH 8, 150 mM NaCl. At lower salt concentrations the heterodimers heavily precipitated.

4.6.7 Crystallization

Crystals of the *Macaca fascicularis* B-box(EK/RD)-coiled-coil domains were grown using the microbatch under oil method at room temperature at 1.5 mg/mL. A 1 uL drop of protein solution was mixed with a 1 uL drop of precipitant solution. Crystals formed in 0.1 M sodium nitrate, 0.1 M HEPES pH 7.5, 8% v/v ethylene glycol, 10% w/v PEG 8000 after one night. Crystals were cryo-protected

in 25% ethylene glycol. Data was collected at the NE-CAT beamline APS-IDC at the Advanced Photon Source.

Both HKL2000 and XDS were optionally used for data processing. Molecular replacement search models were made from the prior TRIM5 B-box(EK/RD)-coil-coil-lysozyme structure (PDB XXX). Molecular replacement was performed using the CCP4 program Phaser. Iterative rounds of refinement in REFMAC and PHENIX were carried out, along with model building in COOT. Alignment RMSD values were generated using the LSQ align module in COOT. Figure images were generated in Pymol and Coot. Citations for the crystallographic software used can be found in Section 2.5.3.

4.7 Chapter contributions

All negative-stain EM experiments in this chapter were performed by Kaifeng Zhou. TRIMCyp SEC coelutions assays and ITC experiments were performed by myself and Katie Digianantonio. Ivy Huang performed much of the cloning and purification of FEZ1 constructs. FEZ1 SEC coelution assays were performed by myself and Ivy Huang.

5 Analysis of the inter-hexamer targeting CA-binding host factors MxB and TRIM5α

5.1 Introduction

TRIM5α and MxB are intracellular immune factors that can potently restrict HIV-1 and other retroviruses. Both factors directly recognize the HIV-1 capsid lattice in order to block infection (see section 1.3).

Only the TRIM5α SPRY domains are believed to directly interact with CA. Its RING, B-box, and coiled-coil domains contribute to significant self-oligomerization. MxB is a dynamin-like GTPase, and only its unstructured N-terminus (approximately 83 residues), and, in particular, the first 25 amino acids, is believed to directly interact with CA. The remaining globular domains (including the bundle-signaling element, stalk, and GTPase domains) contribute to self-oligomerization (see section 1.3).

It has been established by us and others that neither TRIM5α nor MxB recognize disulfide-linked CA hexamers or native CA dimers in vitro. This led to the hypothesis that both factors recognize unique inter-hexamer surfaces only present in the fully-assembled CA lattice (see section 1.3). However, this hypothesis could not be tested or proven since these inter-hexamer interfaces could not be captured in discrete, stable forms to enable solution-based biochemical or structural analysis. With the design of novel CA lattice assemblies that for the first time recreate these interfaces in solution (described in Chapters 2 and 3), I had in hand the tools to begin deciphering the binding modes of TRIM5α and MxB.

In this chapter I describe the purification of minimal, well-behaving constructs of TRIM5α and MxB that strongly co-pellet with in vitro assembled CA tubes. These constructs were tested in solution-based biochemical assays against the previously designed small and large capsid assemblies. I found that MxB specifically recognizes the three-fold CA interface present in hexamer-2-foldon and tri-hexamer assemblies. This is a novel CA binding mode not described in other known host factors. TRIM5α requires numerous complete CA hexamers for binding. Binding could only be observed using an artificially oligomeric SPRY domain with large hepta-hexamer assemblies. Its relatively weak binding affinity, in combination with its requirement for multiple hexamers, continues to make it a challenging host factor to study. However, we believe our work has laid a solid framework to continue studying this host factor and gets us closer to understanding its detailed mechanism at high resolution.

5.2 MxB Results

5.2.1 Optimization of CA-binding MxB constructs

Using a full-length MxB construct (along with oligomerization-ablating mutations) the lab previously showed that MxB binds in vitro assembled cross-linked CA tubes in co-pelleting assays [12]. This binding was abolished if the first 25 or 83 residues was removed. Due to the poor expression and behavior of full-length MxB constructs, we (including Ben Slater and Sarah Smaga) began designing fusions of the first 35 or 83 MxB residues to solubility tags including glutathione-S-transferase (GST), MBP, and GCN4. These constructs behaved and expressed much better than full-length MxB. However, the MxB(1-35)-GST

fusion precipitated heavily upon freeze/thaw. We successfully purified MxB(1-35)-MBP and MxB(1-83)-MBP along with corresponding CA-binding 11RRR13>11AAA13 mutations **(Figure 5-1, A-C)**. Certain special requirements were needed when purifying and using these constructs since the MxB peptide portion is extremely positively charged and tends to stick to DNA/RNA and the negatively charged resin of agarose-based columns (more discussion of this in the methods section below). I made MxB(1-35)-GCN4 and GCN4-MxB(1-35) fusions with dimeric and trimeric GCN4. These GCN4 oligomers were designed to increase MxB avidity when binding to CA tubes and soluble assemblies. The MxB-GCN4 fusions had similar behavior and purification problems as MxB-MBP fusions.

With purified MxB peptide fusions in hand, our first discriminatory test of each construct was in co-pelleting assays with 14C/45C CA tubes. All constructs co-pelleted strongly **(Figure 5-1D, E for MxB-MBP fusions)** and in a NaCl-concentration dependent manner. Lower ionic strength conditions generally improved MxB-CA interaction. We used between 1-10 uM of MxB with 75-150 uM CA to achieve complete binding. In the MxB(1-83)-MBP background, a triple alanine mutation ($MxB_{11}RRR_{13}>_{11}AAA_{13}$) known to reduce MxB-CA interaction significantly reduced MxB-CA co-pelleting **(Figure 5-1F).**

We generally observed much stronger CA co-pelleting with our MxB peptide fusions compared to the full-length MxB used previously. We believe this is due to the superior solution behavior of our fusions and optimization of their binding conditions. Also, artificial MxB oligomers (including GST and GCN4

fusions) qualitatively demonstrated higher CA-binding affinity than their monomeric MBP-fusion counterparts (data not shown).

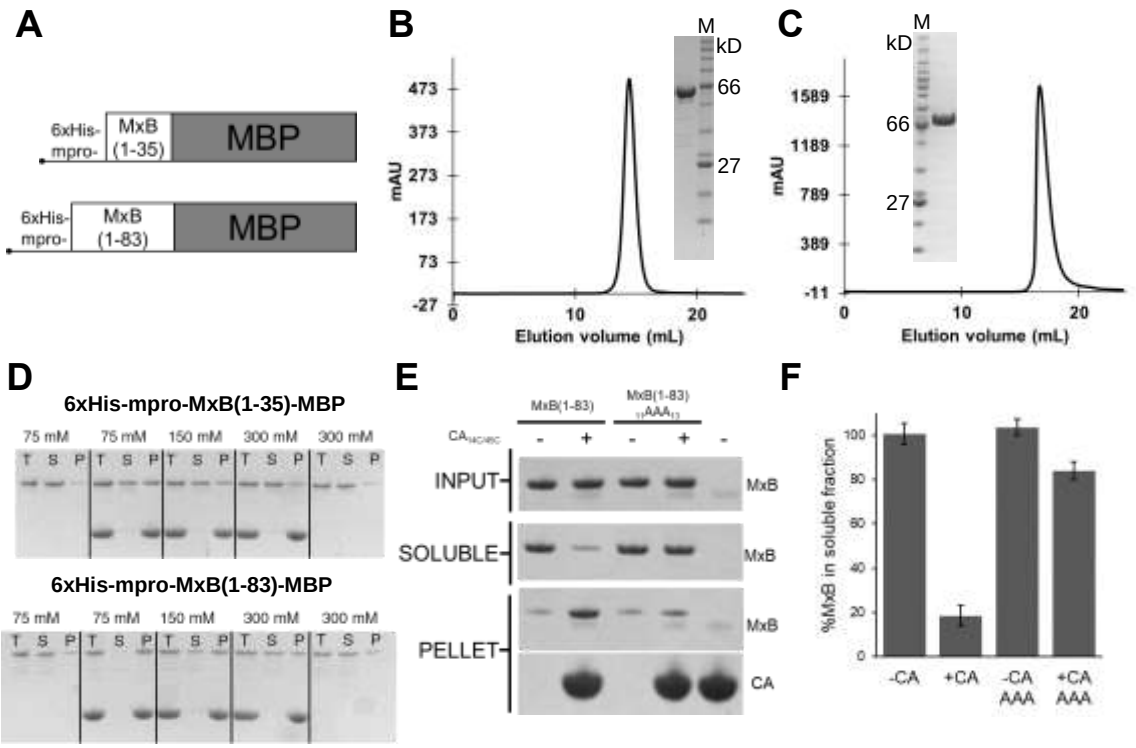


Figure 5-1. Purification of CA-binding MxB constructs. A) Construct design of MxB-MBP fusions. B) SEC chromatogram and SDS-PAGE of MxB(1-35)-MBP. C) SEC chromatogram and SDS-PAGE of MxB(1-83)-MBP. D) NaCl concentration dependence of MxB in CA co-pelleting assay. Top panel is MxB(1-35)-MBP, bottom panel is MxB(1-83)-MBP. E) CA pelleting assay of MxB(1-83)-MBP versus MxB(1-83-$_{11}$AAA$_{13}$)-MBP. F) Quantification of (E).

For the solution-based binding assays described in sections 5.2.2 I used the MxB(1-83)-MBP construct as this construct appeared to have a slightly higher CA tube binding affinity than MxB(1-35)-MBP **(Figure 5-1D, E)**. Sarah Smaga initiated experiments to understand this difference in binding, and the results will be described in her dissertation.

5.2.2 MxB peptide fusions specifically recognize the hexamer-2-foldon

To determine the MxB binding site on capsid I performed extensive binding analysis of MxB(1-83)-MBP with the numerous CA assemblies described in Chapters 2 and 3.

In SEC-binding assays I observed no co-elution of MxB(1-83)-MBP with disulfide-linked CA hexamers or wild-type dimeric CA **(Figure 5-2A, B, D)**. This underscores the requirement of inter-hexamer surfaces for MxB binding. Strikingly, I observed significant co-elution of MxB(1-83)-MBP with hexamer-2-foldon$_{(1\text{-}221)204D}$ **(Figure 5-2C, D)**. This co-elution was abolished in the MxB(1-83)-$_{11}$AAA$_{13}$-MBP mutant background **(Figure 5-2E, F)**. Co-elution was also abolished when the foldon domain was not present to stabilize hexamer-2 **(Figure 5-2D, chromatogram not shown)**. Hexamer-2 (lacking the foldon domain) is only weakly oligomeric at the concentrations used in this assay. Significantly, hexamer-2-foldon constructs are the only small capsid assemblies that contain a complete three-fold CA interface.

I further analyzed the hexamer-2-foldon$_{(1\text{-}221)204D}$-MxB(1-83)-MBP interaction via isothermal titration calorimetry (ITC). I optimized buffer conditions

to keep MxB(1-83)-MBP soluble over the course of overnight dialysis and ITC experiments. I found that a certain buffer (25 mM phosphate buffer pH 7, 75 mM NaCl, 5% glycerol) and performing ITC experiments at 6 °C were optimal. Under these conditions MxB(1-83)-MBP bound hexamer-2-foldon$_{(1\text{-}221)204D}$ with a 9.6 ± 1.4 uM dissociation constant **(Figure 5-2G, I)**. Approximately one MxB molecule bound one hexamer-2-foldon assembly. The binding event was entropically driven and enthalpically unfavorable. These thermodynamic parameters are consistent with ion displacement in a binding mode driven by electrostatic interactions. MxB(1-83)-$_{11}$AAA$_{13}$-MBP association with hexamer-2-foldon$_{(1\text{-}221)204D}$ was significantly weaker **(Figure 5-2H, I)**.

These data clearly and for the first time positively demonstrate that an MxB peptide fragment targets assembled inter-hexamer CA surfaces. Importantly, a triple alanine mutation that dramatically reduces MxB viral restriction also significantly reduces MxB lattice recognition in vitro. This data directly links MxB antiviral and capsid-binding properties to a specific CA high-order motif.

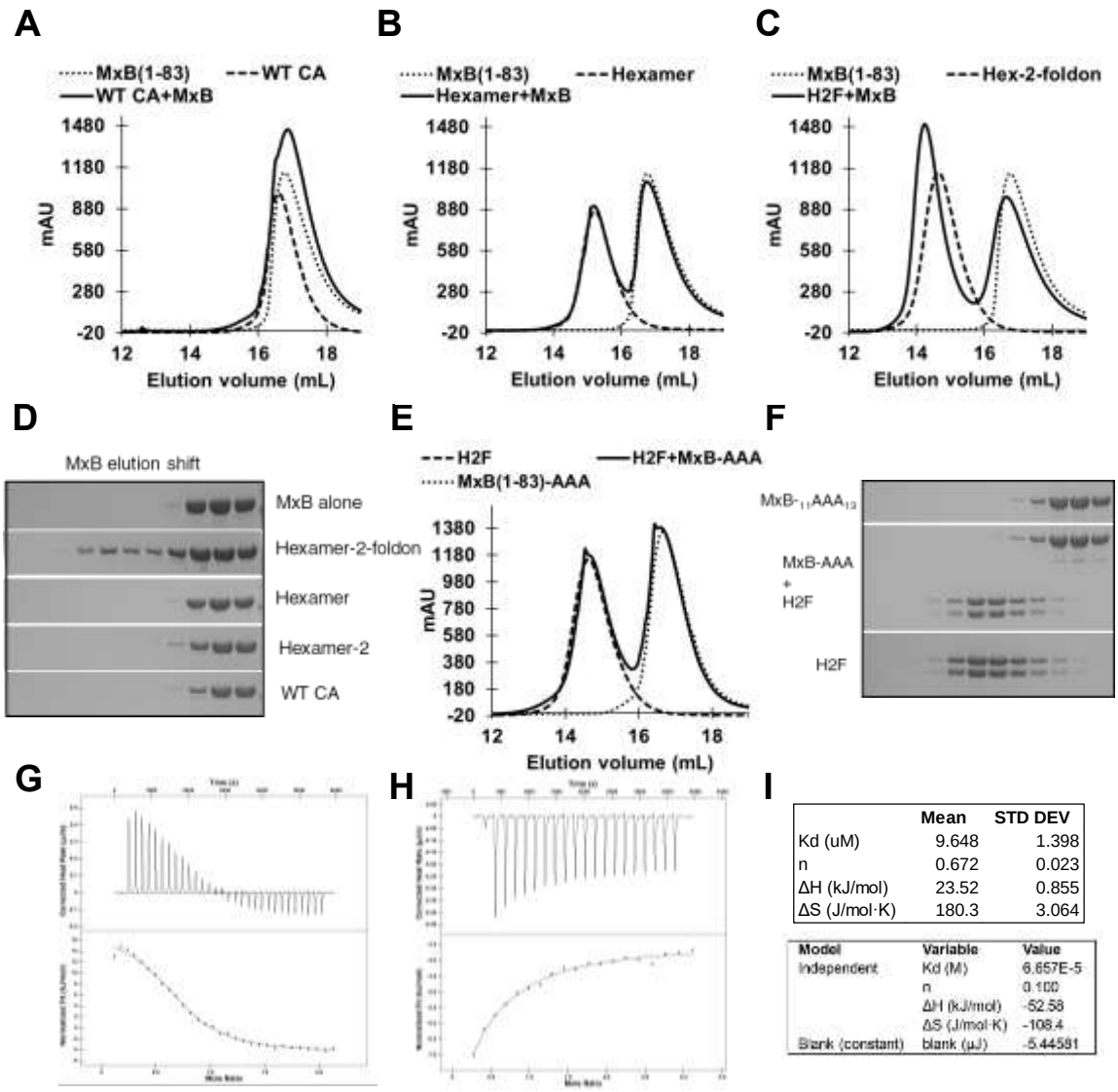


Figure 5-2. MxB recognizes three-fold CA surface within hexamer-2-foldon. A) SEC coelution of MxB(1-83)-MBP with WT CA dimers. B) SEC coelution of MxB(1-83)-MBP with traditional disulfide-linked hexamers. C) SEC coelution of MxB(1-83)-MBP with hexamer-2-foldon$_{(1\text{-}221)204D}$. D) SDS PAGE of SEC coelution assays showing the shift in elution position of MxB(1-83)-MBP with various capsid oligomers. E) SEC coelution of MxB(1-83)-$_{11}$AAA$_{13}$-MBP with hexamer-2-foldon$_{(1\text{-}221)204D}$. F) SDS-PAGE gel of (E). G) Representative ITC isotherm of MxB(1-83)-MBP with hexamer-2-foldon$_{(1\text{-}221)204D}$. H) Representative ITC isotherm of MxB(1-83)-$_{11}$AAA$_{13}$-MBP with hexamer-2-foldon$_{(1\text{-}221)204D}$. I) Chart summarizing values from ITC experiments. Top panel is triplicate data from MxB(1-83)-MBP. Bottom panel is representative data of MxB(1-83)-$_{11}$AAA$_{13}$-MBP.

5.2.3 MxB preferentially recognizes tri-hexamer assemblies

Hexamer-2-foldon is the only small assembly that contains a complete CA three-fold interface. The ITC binding data suggests this interface is occupied by only one MxB molecule. It was important, then, that MxB(1-83)-MBP bind the previously described tri-hexamer assemblies (section 3.2.1). Roughly one-third of the purified tri-hexamers are symmetric and contain a complete three-fold interface. The remaining tri-hexamers are either linear or branched and only contain di-hexamer surfaces.

In SEC co-elution assays, I observed coelution of MxB(1-83)-MBP with tri-hexamers **(Figure 5-3C, D)**. MxB(1-83)-MBP also coeluted weakly with di-hexamers and did not show any co-elution with hexamers **(Figure 5-3A, B, D)**. These data support our model of MxB targeting the tri-hexamer surface. They also support the correct architecture of tri-hexamer assemblies.

However, MxB(1-83)-MBP coelution with tri-hexamer assemblies was significantly weaker than with hexamer-2-foldon. I believe this is due to several factors. For one, hexamer-2-foldon is very pure, well-behaved, and every assembly contains a complete three-fold surface. Conversely, at most only one-third of tri-hexamers contain a three-fold interface. Two, the yield of tri-hexamers was always much lower than that of hexamer-2-foldon and frequently limited the concentrations used in SEC analysis. Third, and maybe most importantly, MxB(1-83)-MBP strongly non-specifically bound to GE S200GL and Superose 6 columns. I consistently observed hexamer-2-foldon and tri-hexamer shifts to

larger molecular weight volumes when MxB was present, but very little or no co-eluting MxB could be observed via SDS PAGE. I believe that at low concentrations where MxB is only weakly bound to CA, MxB(1-83)-MBP is more likely to stick to the resin of the column. I eventually began using a Yarra SEC3000 HPLC column which uses a silica-based resin instead of an agarose-based resin. MxB(1-83)-MBP appeared to adsorb less to this column and I was able to better observe MxB(1-83)-MBP coelution with capsid assemblies.

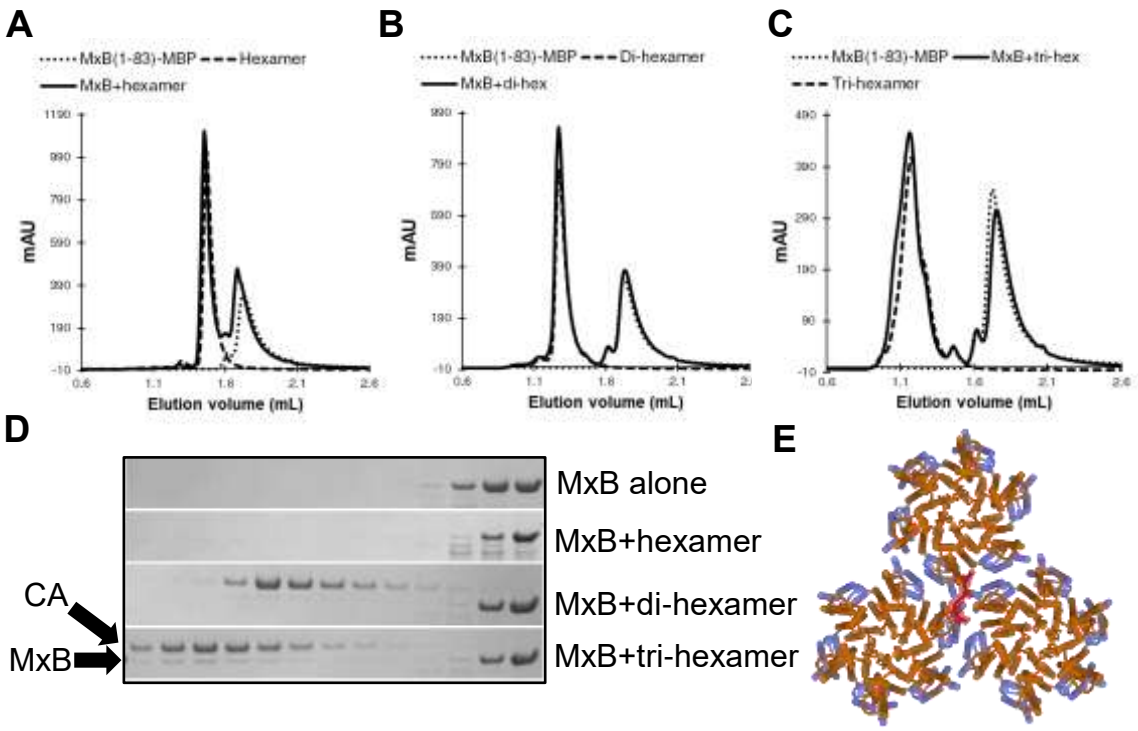


Figure 5-3. MxB recognizes the three-fold CA surface of multi-hexamer assemblies. (A)-(C) SEC coelution assays with MxB(1-83)-MBP and hexamers, di-hexamers, and tri-hexamers, respectively. D) SDS-PAGE highlighting the SEC elution change of MxB(1-83)-MBP depending on the capsid oligomer present. E) Cartoon model of a MxB peptide (red) binding at the three-fold surface between hexamers. CA model based off PDB ID: 4XFX [4].

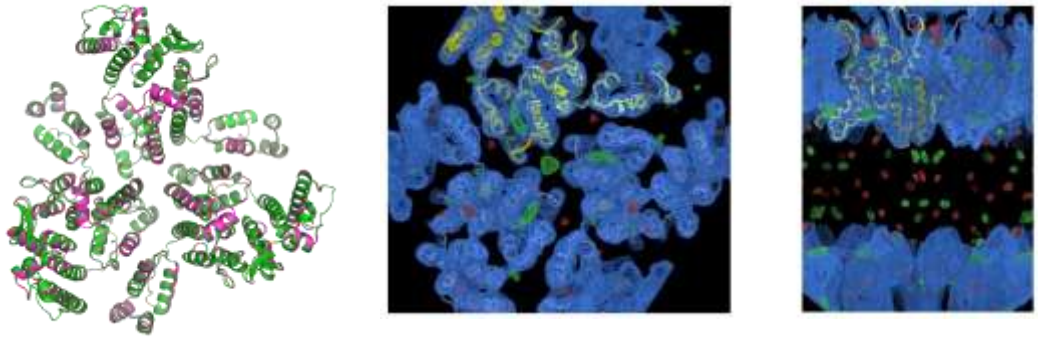
5.2.4 MxB-CA crystallization trials

I attempted to solve a crystal structure of an MxB peptide in complex with CA. I primarily used commercially synthesized MxB peptides (either MxB(5-16) or MxB(5-24) in any CA-co-crystallization screening. I believed that our MxB-MBP fusions would be too flexible and low affinity for effective crystallization. The small peptides could be used at high concentrations (generally 2-10 mM) in vast molar excess to CA to overcome weak binding affinity.

The only CA constructs I tested extensively in co-crystallizations assays were disulfide-linked hexamers, WT CA, and hexamer-2-foldon (all between 0.5 and 1.5 mg/mL in drops). I hoped that hexamers and WT CA would form a flat capsid lattice creating the necessary inter-hexamer surfaces for MxB binding. Hexamer-2-foldon contains a single MxB peptide binding site so I hoped it could crystallize in a discrete complex with an MxB peptide.

A significant issue arose with all crystallization trials: the CA constructs only crystallized in conditions where they also crystallize in apo form. For hexamers and hexamer-2-foldon, this was in 25-100+ unique conditions. All crystals were hexagonal in morphology, like their apo forms. I collected datasets of hexamers and WT CA in P6 lattices with co-crystallized MxB peptide. However, no electron density was present for potential peptides in the solved structures (data not shown).

A B C



D

MxB(5-24): 5-HKPWPY<u>RRR</u>SQFSSRKYLK-24
MxB(5-16): 5-HKPWPY<u>RRR</u>SQF-16

Figure 5-4. Crystal structure of hexamer-2-foldon with MxB peptide in crystallization solution. A) Alignment of hexamer-2-foldon structure (with MxB peptide present in crystallization solution) with the apo-hexamer-2-foldon structure described in Chapter 2. B) Electron density of hexamer-2foldon+MxB (2Fo-Fc 1σ, Fo-Fc 2.9σ. C) Crystal packing of H2F+MxB crystal showing the large gap between lattice layers likely occupied by the flexible foldon domain. D) Sequences of MxB peptides used in crystallization trials. The triple arginine motif critical for CA recognition is underlined.

The apo hexamer-2-foldon crystallized with only three-fold symmetry which potentially made it more amenable to co-crystallization with the MxB peptide (see section 2.2.4). I hoped that MxB peptide binding to hexamer-2-foldon could break its three-fold symmetry allowing us to observe MxB peptide electron density. I solved a crystal structure of hexamer-2-foldon$_{(1\text{-}221)204D}$+MxB(5-24) to 5.5 Å resolution **(Figure 5-4A)(Table 5-1)**. Unfortunately, this crystal also had three-fold crystallographic symmetry centered on the hexamer-2-foldon trimeric interface. Strong MxB peptide density was not observed. Areas with positive difference density were present but did not have corresponding 2Fo-Fc density and were likely noise **(Figure 5-4B)**. Like in the apo-hexamer-2-foldon structure there were large gaps between crystal-lattice layers that could accommodate a foldon domain **(Figure 5-4C)**. No foldon electron density was apparent, however.

Based on what we now know about MxB peptide binding (one MxB per CA three-fold interface) I expect that observation of clear MxB electron density in any of the structures solved would have been unlikely. Due to the symmetry mismatch between peptide binding and the crystallographic lattice, any peptide density likely would be averaged or blurred out in relation to the CA lattice.

Data Collection	
Wavelength (Å)	0.98
Space Group	R32
Cell Dimensions	
a, b, c (Å)	94.96, 94.96, 440.74
α, β, γ (°)	90, 90, 120
Molecules/asymmetric unit	2 CA monomers
Resolution (Å)	19.9-5.5 (5.7-5.5)
R_{merge}	0.23 (3.18)
$I/\sigma I$	5.4 (0.6)
Completeness (%)	89.8 (71.5)
Redundancy	9.0 (6.9)
Unique reflections	2467 (77)
Refinement	
Number of nonhydrogen atoms	3374
R_{work}/R_{free} (%)	23.0/27.0 (45.0/46.6)
Average B factor	257.23
Root mean-squared deviation (rmsd)	
Bond lengths (Å)	0.016
Bond angles (°)	1.33
Ramachandran analysis	
Preferred regions (%)	95.1
Allowed regions (%)	4.21
Outliers (%)	0.7

Table 5-1. Data collection and refinement statistics for MxB(5-24)-hexamer-2foldon(221*/204D) structure. Statistics in parentheses indicate those for the highest resolution shell. These statistics are current as of July 2018 but were not finalized

5.3 TRIM5α Results

5.3.1 Optimization of CA-binding TRIM5α constructs

The lab previously published the purification and structure of a MBP-$SPRY_{Rhesus}$ fusion protein derived from the rhesus macaque TRIM5α gene [26]. However, its affinity towards CA tubes was weak. The affinity of single SPRY domains towards CA has been estimated to be in the high micromolar to millimolar range. This made it nonideal to use in SEC coelution assays with the capsid assemblies described in Chapters 2 and 3. We were given purified CC-$SPRY_{Rhesus}$ protein expressed and purified in a baculovirus system from Jinwoo Ahn at the University of Pittsburgh. This construct contains the TRIM5α coiled-coil and SPRY domains. In its publication, CC-$SPRY_{Rhesus}$ appeared to bind CA tubes with a much higher affinity than MBP-$SPRY_{Rhesus}$ [71]. Unfortunately, we were significantly limited by the low amount of protein provided by Dr. Ahn, which made thorough binding analysis difficult.

To overcome this limitation, Xiaoyun Ji in the lab developed an expression and purification protocol to recombinantly express MBP-B(EK/RD)-CC-$SPRY_{Rhesus}$ in *E. coli*. This construct contains a mutated B-box domain (EK/RD) to prevent homo-oligomerization (described in section 4.2) and both the coiled-coil and SPRY domains to enable capsid recognition. Katie Digianantonio (a post-doctoral fellow in the lab) and I optimized the protocol to produce protein of higher purity and yield **(Figure 5-5A)**.

MBP-B(EK/RD)-CC-SPRY$_{Rhesus}$ and MBP-tag removed B(EK/RD)-CC-SPRY$_{Rhesus}$ co-pelleted efficiently with 14C/45C CA tubes **(Figure 5-5C)**. Highest affinity binding was observed at low ionic strength (75 mM NaCl).

To attempt to make an even higher affinity capsid binder, I fused the SPRY$_{Rhesus}$ domain onto the C-terminus of the trimeric PCNA protein (denoted PCNA-SPRY$_{Rhesus}$). This construct expressed in higher yield and demonstrated better solution behavior than MBP-B(EK/RD)-CC-SPRY$_{Rhesus}$ **(Figure 5-5B)**. PCNA-SPRY$_{Rhesus}$ bound CA tubes with high completeness at low micromolar concentrations **(Figure 5-5D)**.

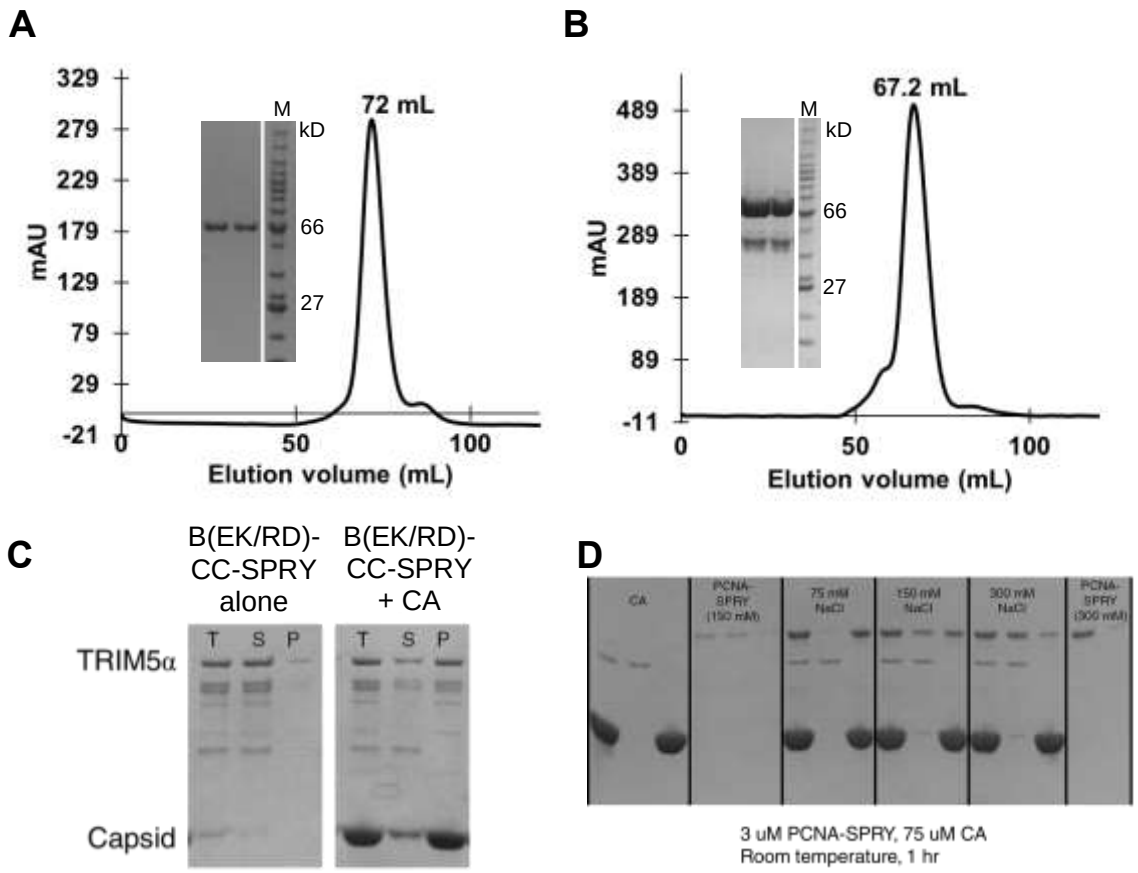


Figure 5-5. Purification and CA-binding of TRIM5α constructs. A) MBP-B(EK/RD)-CC-SPRY$_{Rhesus}$ SEC and SDS-PAGE. B) PCNA-SPRY$_{Rhesus}$ SEC and SDS PAGE. C) CA co-pelleting assay with MBP-B(EK/RD)-CC-SPRY$_{Rhesus}$. D) CA co-pelleting assay with PCNA-SPRY$_{Rhesus}$ demonstrating ionic strength dependence of binding.

5.3.2 TRIM5α coelution assays with large CA assemblies

In our above co-pelleting assays with CA tubes we were unable to observe monomeric SPRY domain binding, likely due to the low affinity of the interaction. We did observe strong tube binding of the natively dimeric MBP-B(EK/RD)-CC-SPRY$_{Rhesus}$ and artificially trimeric PCNA-SPRY$_{Rhesus}$. We used these oligomeric versions of TRIM5α to determine if they can recognize our multi-hexamer assemblies. We did this with the understanding that any binding we observe is likely contributed by either two or three SPRY domains. Additionally, each multi-hexamer assembly contains numerous different and redundant CA lattice surfaces. The complexity of the system would make it difficult to map a SPRY binding site, but we envisioned it as a necessary early step in future high-resolution structural studies.

To begin teasing out an experimentally validated TRIM5α binding mode, I performed a series of SEC binding tests using conditions where dimeric MBP-B(EK/RD)-CC-SPRY$_{Rhesus}$ and trimeric PCNA-SPRY$_{Rhesus}$ bound strongly to cross-linked CA tubes. Under these conditions, I was not able to observe coelution between MBP-B(EK/RD)-CC-SPRY$_{Rhesus}$ or PCNA-SPRY$_{Rhesus}$ with hexamers, di-hexamers, or tri-hexamers. I additionally attempted the same binding tests with CC-SPRY$_{Rhesus}$ derived from baculovirus expression systems. Again, it was difficult to observe any coelution. Representative examples of SEC binding chromatograms with tri-hexamers are in **Figure 5-6A, B, D, E.** The displayed chromatograms are examples of the “best” binding I was able to

observe with tri-hexamers. In most experiments, the chromatograms/SDS-PAGE gels of TRIM5α elution volume with and without multi-hexamers assemblies were identical. Frequently, MBP-B(EK/RD)-CC-SPRY$_{Rhesus}$ aggregated during the course of the experiments. This clouded identification of any weak coelution.

The only consistently strong coelution I observed was with PCNA-SPRY$_{Rhesus}$ and hepta-hexamer assemblies **(Figure 5-6C, F)**. These assemblies can be thought of as a soluble, capsid aggregate that we believe likely have multiple TRIM5α binding sites.

I was unable to observe binding of MBP-B(EK/RD)-CC-SPRY$_{Rhesus}$ with hepta-hexamers. However, the behavior of the MBP-B(EK/RD)-CC-SPRY$_{Rhesus}$ protein was poor in many experiments (I frequently observed a major aggregation shoulder). Either way, this suggests that the binding affinity of a dimeric TRIM5α construct to CA might be too weak to observe in coelution assays. In the case of PCNA-SPRY$_{Rhesus}$ it may be that all three SPRY domains are capable of simultaneously binding the hepta-hexamer surface. This avid binding must increase the SPRY domain affinity enough to observe coelution in this assay.

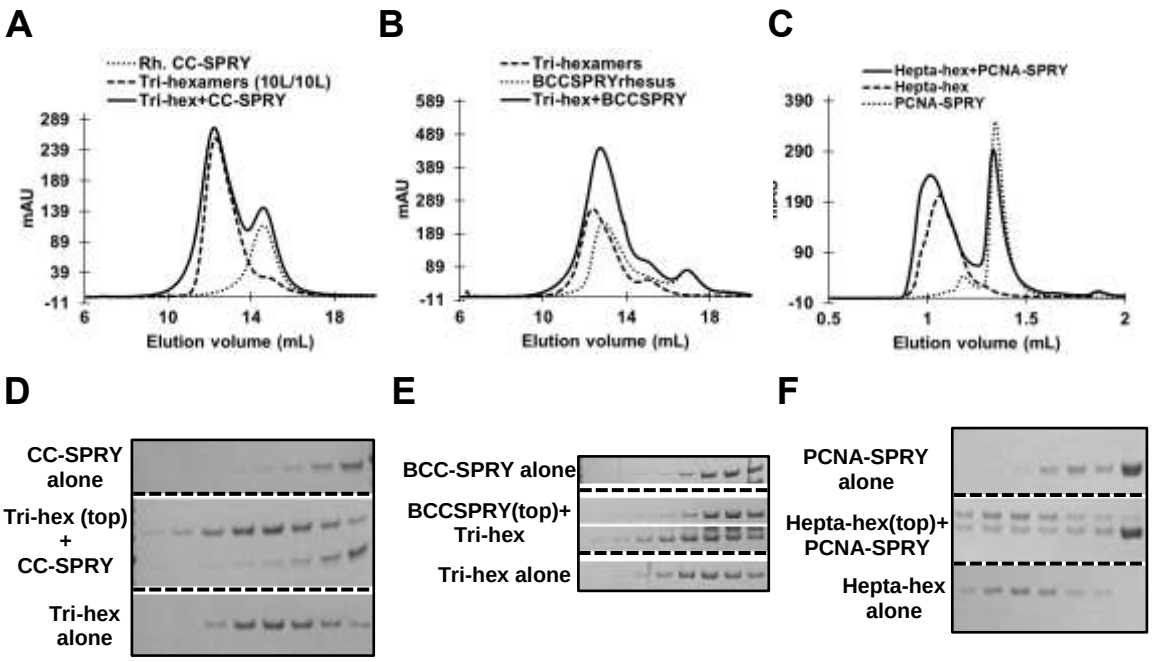


Figure 5-6. TRIM5α constructs binding to multi-hexamer assemblies in SEC coelution assays. A) Representative CC-SPRYRhesus displaying only weak coelution with tri-hexamer assemblies. B) Representative B(EK/RD)-CC-SPRYRhesus displaying only weak coelution with tri-hexamer assemblies. C) PCNA-SPRYRhesus demonstrating consistently strong co-elution with hepta-hexamer assemblies. D) Associated SDS-PAGE gel from (A). E) Associated SDS-PAGE gel from (B). F) Associated SDS-PAGE gel from (C). (A) and (B) were performed on GE Superose 6 10/300 column. (C) was performed on a Yarra SEC3000 column.

5.3.3 TRIM5α crystallization studies

I attempted to crystallize TRIM5α constructs containing both coiled-coil and SPRY domains in order to understand how SPRY domains are packed and oriented in relation to the coiled-coil domains. The architecture of these proteins may determine the capsid surfaces able to be targeted by them. I used the baculovirus expressed CC-SPRY$_{Rhesus}$ and CC-SPRY$_{human}$ provided by Dr. Jinwoo Ahn at the University of Pittsburgh in extensive crystallization screening. Generally, precipitation was observed in 30-50% of the conditions if protein concentration was between 0.75-1.5 mg/mL. I screened conditions at room temperature and 12 °C. Unfortunately, I did not find any crystallization hits.

I analyzed the stability of both constructs after overnight incubation at room temperature. Significant aggregation was observed on SEC (data not shown). The instability of the constructs, paired with potential flexibility in the SPRY domains in relation to the coiled-coil domains likely contributed to the lack of crystals.

I tried very few co-crystallization attempts using any TRIM5α construct with any of the designed multi-hexamer assemblies. I felt that chances of co-crystallization were low considering the low-binding affinity of their interactions.

5.4 Discussion

Despite being two of the most well-studied HIV-1-related host proteins, it has proven extremely difficult to obtain high-resolution structural information of the interactions between TRIM5α and MxB with capsid. A major cause of this is their specialized binding modes—both TRIM5α and MxB display essentially no

affinity towards CA hexamers and must target an interface only present in the assembled capsid lattice. Add to this their extensive self-oligomerization properties, and it is clear why the field has struggled to even identify their binding sites on capsid, let alone solve their three-dimensional structure at high-resolution (see section 1.3 for more background).

My design and production of novel CA oligomers “trapped” in a lattice-like state (described in Chapters 2 and 3) put me in an excellent position to begin deciphering the TRIM5α and MxB binding modes. I first had to develop expression and purification protocols to produce well-behaved, CA-binding MxB and TRIM5α constructs. I succeeded in this with the help of fellow graduate students and post-doctoral associates in the lab.

With minimal, CA-binding MxB and TRIM5α constructs in hand I began extensive solution-based biochemical assays to determine their binding site on capsid. I was very clearly able to demonstrate that the N-terminal 83 residues of MxB target the three-fold CA interface present in hexamer-2-foldon assemblies and tri-hexamers. This is a novel capsid recognition mode not described previously. This binding mode is supported by the identification of specific CA mutations in this region that reduce MxB restriction in viral evolution assays (CA mutations P207S, G208R, and T210K)[90].

TRIM5α proved more difficult to study. I was able to produce TRIM5α constructs that strongly co-pellet with CA tubes (including B(EK/RD)-CC-$SPRY_{Rhesus}$ and PCNA-$SPRY_{Rhesus}$), but it was very difficult to observe coelution by either of these constructs with even my largest multi-hexamer assemblies.

The only consistent and significant binding I observed was between the trimeric PCNA-SPRY$_{Rhesus}$ construct and the crude hepta-hexamer assembly. This was a combination of my highest-oligomeric state TRIM5α construct with my largest capsid oligomer. Due to the redundancy of capsid interfaces on the hepta-hexamer (for example, it contains six three-fold interfaces), I was unable to draw many conclusions about the specific binding site targeted by the SPRY domain. Instead, I obtained some level of validation that recombinant SPRY domains are, in fact, able to bind engineered multi-hexamer capsid assemblies in vitro. This may be an important first step in obtaining SPRY-CA complexes suitable for more detailed biochemical or structural analysis.

5.5 Future directions

As when I began my dissertation work, the major targets to go after are high-resolution structures of TRIM5α and MxB in complex with CA. Due to the potential symmetry mismatch of MxB peptides binding the CA lattice it may be beneficial to attempt cryo-EM or NMR based structural biology experiments. These techniques would likely have less symmetry mismatch issues.

While ITC enabled the thermodynamic characterization of MxB-hexamer-2-foldon binding, it required large amounts of protein due to the low binding affinity of the interaction. It may be in a future researcher's best interest to optimize a different assay to measure MxB-CA affinity, such as fluorescence anisotropy or microscale thermophoresis. Additionally, while the behavior of MxB-MBP fusions is better than that of full-length MxB, it is still relatively poor due to the numerous positively charged residues at the MxB N-terminus.

Perhaps neutralizing some of this positive charge (in regions that do not effect capsid binding) could significantly reduce experimental challenges in the future.

Since the CA-binding mode of MxB appears unique, it may be interesting to screen for small molecule compounds that bind and compete at the MxB binding site within hexamer-2-foldon. This could potentially unveil a new class of anti-CA drugs. Conversely, hexamer-2-foldon could be used to pull-down unidentified cellular factors that bind the CA three-fold interface in a similar fashion as MxB.

In regards to TRIM5α, optimization of the PCNA-SPRY$_{Rhesus}$-hepta-hexamer complex may yield results using cryo-EM. The complex is well over a megadalton in size (which is ideal for cryo-EM), and it may be necessary to have large oligomers of TRIM5α in complex with CA in order to form a complex stable enough for structural studies. One could build an even larger SPRY oligomer to further improve binding affinity.

The SEC coelution assays described in this thesis are not sensitive enough for extremely low affinity interactions. A recent publication used microscale thermophoresis to suggest that SPRY domains target the three-fold CA surface [136]. This group did not have access to the library of capsid assemblies presented in this thesis, so their scope and thoroughness was limited. It may be beneficial to use microscale thermophoresis to better map SPRY binding using our capsid assemblies.

The solution behavior of B(EK/RD)-CC-SPRY constructs was not ideal for solution-based assays. Despite high purity, constructs frequently formed

oligomers or aggregates on SEC. This clouded any possible coelution differences when capsid assemblies were present. Construct optimization (or the identification of better behaving TRIM5 proteins from different monkey species) is still recommended before serious biochemical analysis is performed.

5.6 Materials and methods

5.6.1 Cloning

TRIM5α genes were provided by the NIH AIDS Reagent Program. B-box, coiled-coil, and SPRY domains were cloned into pMAT9S with N-terminal 6xHis, MBP tag, and mpro protease cleavage site. The SPRY domain was fused to the C-terminus of 6xHis-PCNA in the pRSF vector. MxB peptide fragments (either residues 1-35 or 1-83) were fused to MBP and cloned into pETDuet. They contained N-terminal 6xHis-mpro cleavage sequences. Mutations were made by QuickChange site-directed mutagenesis (Stratagene).

5.6.2 Expression and purification

To express all TRIM5α constructs I used chemically competent *E. coli* BL-21(DE3) cells transformed using the heat shock method with the plasmid of interest. They were co-transformed with the pGro7 plasmid to express the GroE chaperone. Generally, 3L flasks of Terrific broth were used for protein expression. Cells were grown to an OD of 0.6-1 and induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 2 mg/mL arabinose at 18°C overnight. This expression procedure generally yielded 5-10 mg of purified protein.

Cells were harvested by centrifugation and lysed by a microfluidizer with Roche protease inhibitor tablets. The lysate was centrifugally clarified at 13500 rpm for 35 minutes. 6xHis-MBP-B(EK/RD)-CC-SPRY constructs were purified by a gravity nickel column, followed by an MBP column, then anion exchange (proteins generally eluted between 200-300 mM NaCl), mpro digestion for 2+ hours if the MBP tag is to be removed, and a polishing GE S200PG. The final SEC column was run in 50 mM TRIS pH 8, 300 mM NaCl, 0.5 mM phenylmethane sulfonyl fluoride (PMSF). The purity after each step was analyzed by SDS-PAGE. PCNA-SPRY constructs were purified similarly except only gravity nickel, anion exchange, and SEC columns were used.

MxB-MBP fusions were expressed in *E. coli* BL-21(DE3) cells transformed using the heat shock method with the plasmid of interest. Generally, 1.5-3L flasks of Terrific broth were used for protein expression. Cells were grown to an OD of 0.6-1 and induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 18°C overnight. This expression procedure generally yielded 10-20 mg of purified protein.

Cells were harvested by centrifugation and lysed by a microfluidizer with Roche protease inhibitor tablets. The lysate was centrifugally clarified at 13500 rpm for 35 minutes. Protein was purified using a gravity nickel column, a cation exchange column (HiTrap SP)(protein eluted at approximately 400 mM NaCl), and a GE S200PG run in 50 mM TRIS pH 8, 300 mM NaCl, 0.5 mM phenylmethane sulfonyl fluoride (PMSF).

GCN4-MxB fusion proteins were expressed similarly, except that a 0.1% polyethyleneimine (PEI) precipitation was performed after lysis. Excess PEI was removed using a 70% w/v ammonium sulfate precipitation. The pellet was resuspended and applied to a gravity nickel column, followed by cation exchange, and SEC as described for MxB-MBP fusions.

5.6.3 CA-tube co-pelleting assays

14C/45C disulfide-stabilized CA tubes were assembled by overnight dialysis in 50 mM TRIS pH 8, 1M NaCl at approximately 15 mg/mL. This was followed by overnight dialysis into 50 mM TRIS pH 8. Approximately 1-5 uM purified host factors were mixed with approximately 75-150 uM CA tubes and incubated with shaking at room temperature for 30 minutes. Tubes were pelleted by centrifugation at 4 °C at 14000 rpm for 10 minutes. Total, soluble, and pellet fractions were taken at appropriate times and analyzed via SDS-PAGE. MxB pelleting assays were performed in 25 mM TRIS pH 8 and either 75 or 100 mM NaCl. TRIM5α pelleting assays were performed in 25 mM TRIS pH 8, 75 mM NaCl.

5.6.4 Size-exclusion chromatography (SEC) coelution assays

Host factors were mixed with CA assemblies for 30 minutes to 1 hour on ice in the same buffer used for CA-tube co-pelleting assays. MxB-MBP binding tests with small capsid assemblies were performed using 500 uL reactions with 56 uM MxB and 78.5 uM CA monomer (of the appropriate CA assembly). A GE Superose 6 GL was used. It was important to first run MxB-MBP fusion protein over the column to coat non-specific binding surfaces with MxB. Using this

approach MxB-CA coelution was much stronger. For MxB-MBP coelution assays with multi-hexamer assemblies I used a Yarra SEC3000 column. The column was run in 25 mM phosphate buffer pH 7, 100 mM NaCl. Binding reactions were 60 uL and contained 56 uM MxB-MBP protein and 78.5 uM CA monomer (from appropriate multi-hexamer assembly).

For TRIM5α coelution assays I used either a Superose 6 column or the Yarra SEC3000 column. For Superose 6 assays I used 25 mM TRIS pH 8, 75 mM NaCl with 500 uL reactions with 60 uM or higher TRIM5α monomer and 80 uM or higher CA monomer (from multi-hexamer assemblies). I generally used as high concentrations as I could with the available protein I had. For Yarra SEC3000 assays, I used 25 mM phosphate buffer pH 7, 75 mM NaCl. Reaction mixtures were 60 uL with similar protein concentrations as in Superose 6 assays.

5.6.5 ITC experiments

All ITC experiments were performed using a TA Instruments NanoITC machine in the Yale Chemical Biology Instrumentation Center. Data was analyzed using the provided NanoAnalyze software. All curves were fit with an independent one-site binding model. For MxB(1-83)-MBP and MxB(1-83)-$_{11}$AAA$_{13}$-MBP the buffer used was 25 mM phosphate pH 7, 75 mM NaCl, and 5% glycerol. Experiments were performed at 6 °C to improve MxB behavior over the course of the run.

5.6.6 Crystallization

We used the microbatch under oil method for crystallization screening. Hexamer-2-foldon$_{(1\text{-}221)204D}$ and MxB(5-24) complexes were formed at

approximately 0.5 mg/mL hexamer-2-foldon with 2 mM MxB(5-24) in 25 mM TRIS pH 8 and 75-150 mM NaCl. To our surprise, crystals formed in this binding mixture without added precipitant when under oil overnight. We varied the protein and salt concetration to optimize the crystals, but improvement seemed random. Crystals were frozen with 25% glycerol as a cryo-protectant. Data was collected at the NE-CAT beamline APS-IDE at the Advanced Photon Source.

Both HKL2000 and XDS were optionally used for data processing. Molecular replacement search models were made from my prior apo hexamer-2-foldon structure described in Chapter 2. Molecular replacement was performed using the CCP4 program Phaser. Iterative rounds of refinement in REFMAC and PHENIX were carried out, along with model building in COOT. Alignment RMSD values were generated using the LSQ align module in COOT. Figure images were generated in Pymol and Coot. Citations for the crystallographic software used can be found in Section 2.5.3.

5.7 Chapter contributions

Cloning, expression, and purification of MxB constructs were performed by both myself and Sarah Smaga. MxB-CA co-pelleting assays were performed by both myself and Sarah Smaga. Crystallization of hexamer-2-foldon and the MxB(5-24) peptide was performed by both myself and Sarah Smaga.

Expression and purification of MBP-B(EK/RD)-CC-SPRY constructs were performed by both myself and Dr. Katie Digianantonio. TRIM5α construct CA co-pelleting assays were also performed by both myself and Dr. Katie Digianantonio.

6 Concluding Remarks

During the course of my five years of thesis work in the Xiong lab, I succeeded in creating an extensive library of stabilized HIV-1 capsid oligomers **(Summarized in Figure 6-1)**. These can be used to quickly and thoroughly determine the capsid-binding mode of known and soon to be discovered capsid-binding host factors relevant to HIV-1 infection. I used these capsid oligomers to define the binding modes of several capsid-binding host factors. Some are critical to preventing infection, others are used by HIV-1 to promote infection. By understanding the unique ways nature has evolved to recognize the capsid surface I hope that future scientists and clinicians can develop novel antiviral therapeutics targeted towards capsid.

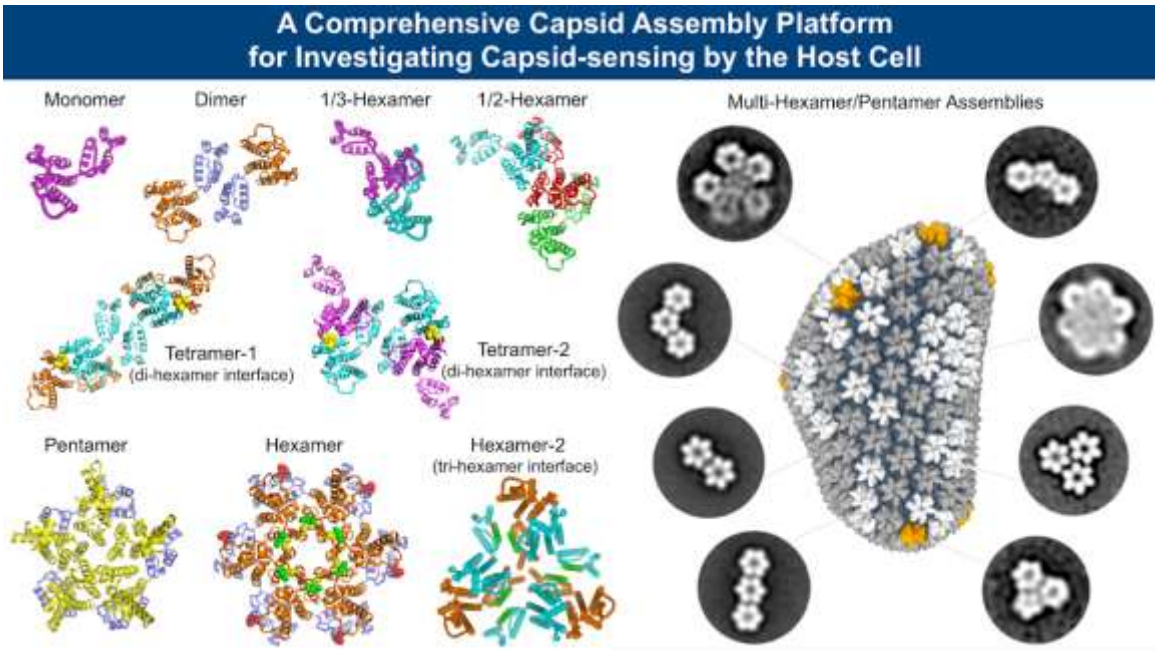


Figure 6-1 A comprehensive library of stabilized HIV-1 capsid oligomers.

7 References

1. Zhao, G., et al., *Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics.* Nature, 2013. **497**(7451): p. 643-6.
2. Li, L., et al., *Structural analysis and optimization of the covalent association between SpyCatcher and a peptide Tag.* J Mol Biol, 2014. **426**(2): p. 309-17.
3. Blair, W.S., et al., *HIV capsid is a tractable target for small molecule therapeutic intervention.* PLoS Pathog, 2010. **6**(12): p. e1001220.
4. Gres, A.T., et al., *STRUCTURAL VIROLOGY. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability.* Science, 2015. **349**(6243): p. 99-103.
5. Price, A.J., et al., *Host cofactors and pharmacologic ligands share an essential interface in HIV-1 capsid that is lost upon disassembly.* PLoS Pathog, 2014. **10**(10): p. e1004459.
6. Bhattacharya, A., et al., *Structural basis of HIV-1 capsid recognition by PF74 and CPSF6.* Proc Natl Acad Sci U S A, 2014. **111**(52): p. 18625-30.
7. Goldstone, D.C., et al., *Structural studies of postentry restriction factors reveal antiparallel dimers that enable avid binding to the HIV-1 capsid lattice.* Proc Natl Acad Sci U S A, 2014. **111**(26): p. 9609-14.
8. Mattei, S., et al., *The structure and flexibility of conical HIV-1 capsids determined within intact virions.* Science, 2016. **354**(6318): p. 1434-1437.
9. Gray, E.R., et al., *Unravelling the Molecular Basis of High Affinity Nanobodies against HIV p24: In Vitro Functional, Structural, and in Silico Insights.* ACS Infect Dis, 2017. **3**(7): p. 479-491.
10. *Introduction to Viruses*. Microbiologybook.org Retrieved 2018; Available from: http://www.microbiologybook.org/mhunt/intro-vir.htm.
11. Zhang, W., et al., *Morphology and ultrastructure of retrovirus particles.* AIMS Biophys, 2015. **2**(3): p. 343-369.
12. Fribourgh, J.L., et al., *Structural Insight into HIV-1 Restriction by MxB.* Cell host & microbe, 2014. **16**(5): p. 627-638.
13. Berthelmann, A., et al., *Versatile C(3)-symmetric scaffolds and their use for covalent stabilization of the foldon trimer.* Org Biomol Chem, 2014. **12**(16): p. 2606-14.
14. Campbell, E.M. and T.J. Hope, *HIV-1 capsid: the multifaceted key player in HIV-1 infection.* Nat Rev Microbiol, 2015. **13**(8): p. 471-83.
15. Jia, X., Q. Zhao, and Y. Xiong, *HIV suppression by host restriction factors and viral immune evasion.* Curr Opin Struct Biol, 2015. **31**: p. 106-14.
16. Pornillos, O., et al., *X-ray structures of the hexameric building block of the HIV capsid.* Cell, 2009. **137**(7): p. 1282-92.
17. Yamashita, M. and A.N. Engelman, *Capsid-Dependent Host Factors in HIV-1 Infection.* Trends Microbiol, 2017. **25**(9): p. 741-755.
18. Pornillos, O., B.K. Ganser-Pornillos, and M. Yeager, *Atomic-level modelling of the HIV capsid.* Nature, 2011. **469**(7330): p. 424-7.

19. Pornillos, O., et al., *Disulfide bond stabilization of the hexameric capsomer of human immunodeficiency virus.* J Mol Biol, 2010. **401**(5): p. 985-95.
20. von Schwedler, U.K., et al., *Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly.* EMBO J, 1998. **17**(6): p. 1555-68.
21. Malikov, V., et al., *HIV-1 capsids bind and exploit the kinesin-1 adaptor FEZ1 for inward movement to the nucleus.* Nat Commun, 2015. **6**: p. 6660.
22. Ulbrich, P., et al., *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.
23. Lanza, D.C., et al., *Human FEZ1 has characteristics of a natively unfolded protein and dimerizes in solution.* Proteins, 2009. **74**(1): p. 104-21.
24. Gamble, T.R., et al., *Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid.* Cell, 1996. **87**(7): p. 1285-94.
25. Li, Y.L., et al., *Primate TRIM5 proteins form hexagonal nets on HIV-1 capsids.* Elife, 2016. **5**.
26. Yang, H., et al., *Structural insight into HIV-1 capsid recognition by rhesus TRIM5alpha.* Proc Natl Acad Sci U S A, 2012. **109**(45): p. 18372-7.
27. Bichel, K., et al., *HIV-1 capsid undergoes coupled binding and isomerization by the nuclear pore protein NUP358.* Retrovirology, 2013. **10**: p. 81.
28. Liu, C., et al., *Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site.* Nat Commun, 2016. **7**: p. 10714.
29. C Michael Hogan, S.D., *Virus*, in *Encylopedia of Earth*, C.J. Cleveland, Editor. 2010, Environmental Information Coalition, National Council for Science and the Environment: Washington, D.C.
30. Wills JW, C.R., *Form, function, and use of retroviral gag proteins.* AIDS, 1991(5): p. 639-54.
31. Nguyen, T.T., R.F. Bruinsma, and W.M. Gelbart, *Continuum theory of retroviral capsids.* Phys Rev Lett, 2006. **96**(7): p. 078102.
32. Dick, R.A. and V.M. Vogt, *Membrane interaction of retroviral Gag proteins.* Front Microbiol, 2014. **5**: p. 187.
33. Douek, D.C., M. Roederer, and R.A. Koup, *Emerging concepts in the immunopathogenesis of AIDS.* Annu Rev Med, 2009. **60**: p. 471-84.
34. Briggs, J.A., et al., *The stoichiometry of Gag protein in HIV-1.* Nat Struct Mol Biol, 2004. **11**(7): p. 672-5.
35. Ganser, B.K., et al., *Assembly and analysis of conical models for the HIV-1 core.* Science, 1999. **283**(5398): p. 80-3.
36. Li, S., et al., *Image reconstructions of helical assemblies of the HIV-1 CA protein.* Nature, 2000. **407**(6802): p. 409-13.
37. Frank, G.A., et al., *Maturation of the HIV-1 core by a non-diffusional phase transition.* Nat Commun, 2015. **6**: p. 5854.
38. Yu, Z., et al., *Unclosed HIV-1 capsids suggest a curled sheet model of assembly.* J Mol Biol, 2013. **425**(1): p. 112-23.

39. 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.
40. Obal, G., et al., *STRUCTURAL VIROLOGY. Conformational plasticity of a native retroviral capsid revealed by x-ray crystallography.* Science, 2015. **349**(6243): p. 95-8.
41. Cornilescu, C.C., et al., *Structural analysis of the N-terminal domain of the human T-cell leukemia virus capsid protein.* J Mol Biol, 2001. **306**(4): p. 783-97.
42. Mortuza, G.B., et al., *High-resolution structure of a retroviral capsid hexameric amino-terminal domain.* Nature, 2004. **431**(7007): p. 481-5.
43. Byeon, I.J., et al., *Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function.* Cell, 2009. **139**(4): p. 780-90.
44. Mattei, S., F.K. Schur, and J.A. Briggs, *Retrovirus maturation-an extraordinary structural transformation.* Curr Opin Virol, 2016. **18**: p. 27-35.
45. Schur, F.K., et al., *Structure of the immature HIV-1 capsid in intact virus particles at 8.8 A resolution.* Nature, 2015. **517**(7535): p. 505-8.
46. Fassati, A. and S.P. Goff, *Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1.* J Virol, 2001. **75**(8): p. 3626-35.
47. Miller, M.D., C.M. Farnet, and F.D. Bushman, *Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition.* J Virol, 1997. **71**(7): p. 5382-90.
48. Gao, D., et al., *Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses.* Science, 2013. **341**(6148): p. 903-6.
49. Rasaiyaah, J., et al., *HIV-1 evades innate immune recognition through specific cofactor recruitment.* Nature, 2013. **503**(7476): p. 402-405.
50. Yamashita, M. and M. Emerman, *Capsid is a dominant determinant of retrovirus infectivity in nondividing cells.* J Virol, 2004. **78**(11): p. 5670-8.
51. Yamashita, M., et al., *Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells.* PLoS Pathog, 2007. **3**(10): p. 1502-10.
52. Francis, A.C. and G.B. Melikyan, *Single HIV-1 Imaging Reveals Progression of Infection through CA-Dependent Steps of Docking at the Nuclear Pore, Uncoating, and Nuclear Transport.* Cell Host Microbe, 2018. **23**(4): p. 536-548 e6.
53. Mamede, J.I., et al., *Early cytoplasmic uncoating is associated with infectivity of HIV-1.* Proc Natl Acad Sci U S A, 2017. **114**(34): p. E7169-E7178.
54. Stremlau, M., et al., *The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys.* Nature, 2004. **427**(6977): p. 848-53.

55. Stremlau, M., et al., *Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor.* Proc Natl Acad Sci U S A, 2006. **103**(14): p. 5514-9.
56. Reymond, A., et al., *The tripartite motif family identifies cell compartments.* EMBO J, 2001. **20**(9): p. 2140-51.
57. Yamauchi, K., et al., *Ubiquitination of E3 ubiquitin ligase TRIM5 alpha and its potential role.* FEBS J, 2008. **275**(7): p. 1540-55.
58. Diaz-Griffero, F., et al., *A B-box 2 surface patch important for TRIM5alpha self-association, capsid binding avidity, and retrovirus restriction.* J Virol, 2009. **83**(20): p. 10737-51.
59. Li, X. and J. Sodroski, *The TRIM5alpha B-box 2 domain promotes cooperative binding to the retroviral capsid by mediating higher-order self-association.* J Virol, 2008. **82**(23): p. 11495-502.
60. Stremlau, M., et al., *Species-specific variation in the B30.2(SPRY) domain of TRIM5alpha determines the potency of human immunodeficiency virus restriction.* J Virol, 2005. **79**(5): p. 3139-45.
61. Ganser-Pornillos, B.K., et al., *Hexagonal assembly of a restricting TRIM5alpha protein.* Proc Natl Acad Sci U S A, 2011. **108**(2): p. 534-9.
62. Biris, N., et al., *Structure of the rhesus monkey TRIM5alpha PRYSPRY domain, the HIV capsid recognition module.* Proc Natl Acad Sci U S A, 2012. **109**(33): p. 13278-83.
63. Pertel, T., et al., *TRIM5 is an innate immune sensor for the retrovirus capsid lattice.* Nature, 2011. **472**(7343): p. 361-5.
64. Kutluay, S.B., D. Perez-Caballero, and P.D. Bieniasz, *Fates of retroviral core components during unrestricted and TRIM5-restricted infection.* PLoS Pathog, 2013. **9**(3): p. e1003214.
65. Hatziioannou, T., et al., *Restriction of multiple divergent retroviruses by Lv1 and Ref1.* EMBO J, 2003. **22**(3): p. 385-94.
66. Wilson, S.J., et al., *Rhesus macaque TRIM5 alleles have divergent antiretroviral specificities.* J Virol, 2008. **82**(14): p. 7243-7.
67. Perron, M.J., et al., *TRIM5alpha mediates the postentry block to N-tropic murine leukemia viruses in human cells.* Proc Natl Acad Sci U S A, 2004. **101**(32): p. 11827-32.
68. Yap, M.W., et al., *Trim5alpha protein restricts both HIV-1 and murine leukemia virus.* Proc Natl Acad Sci U S A, 2004. **101**(29): p. 10786-91.
69. Black, L.R. and C. Aiken, *TRIM5alpha disrupts the structure of assembled HIV-1 capsid complexes in vitro.* J Virol, 2010. **84**(13): p. 6564-9.
70. Sebastian, S. and J. Luban, *TRIM5alpha selectively binds a restriction-sensitive retroviral capsid.* Retrovirology, 2005. **2**: p. 40.
71. Zhao, G., et al., *Rhesus TRIM5alpha disrupts the HIV-1 capsid at the inter-hexamer interfaces.* PLoS Pathog, 2011. **7**(3): p. e1002009.
72. Nakayama, E.E. and T. Shioda, *Role of Human TRIM5alpha in Intrinsic Immunity.* Front Microbiol, 2012. **3**: p. 97.
73. Luban, J., *Cyclophilin A, TRIM5, and resistance to human immunodeficiency virus type 1 infection.* J Virol, 2007. **81**(3): p. 1054-61.

74. Sayah, D.M., et al., *Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1.* Nature, 2004. **430**(6999): p. 569-73.
75. Brennan, G., Y. Kozyrev, and S.L. Hu, *TRIMCyp expression in Old World primates Macaca nemestrina and Macaca fascicularis.* Proc Natl Acad Sci U S A, 2008. **105**(9): p. 3569-74.
76. Newman, R.M., et al., *Evolution of a TRIM5-CypA splice isoform in old world monkeys.* PLoS Pathog, 2008. **4**(2): p. e1000003.
77. Nisole, S., et al., *A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1.* Proc Natl Acad Sci U S A, 2004. **101**(36): p. 13324-8.
78. Virgen, C.A., et al., *Independent genesis of chimeric TRIM5-cyclophilin proteins in two primate species.* Proc Natl Acad Sci U S A, 2008. **105**(9): p. 3563-8.
79. Caines, M.E., et al., *Diverse HIV viruses are targeted by a conformationally dynamic antiviral.* Nat Struct Mol Biol, 2012. **19**(4): p. 411-6.
80. Price, A.J., et al., *Active site remodeling switches HIV specificity of antiretroviral TRIMCyp.* Nat Struct Mol Biol, 2009. **16**(10): p. 1036-42.
81. Ylinen, L.M., et al., *Conformational adaptation of Asian macaque TRIMCyp directs lineage specific antiviral activity.* PLoS Pathog, 2010. **6**(8): p. e1001062.
82. Sokolskaja, E., D.M. Sayah, and J. Luban, *Target cell cyclophilin A modulates human immunodeficiency virus type 1 infectivity.* J Virol, 2004. **78**(23): p. 12800-8.
83. Hatziioannou, T., et al., *Cyclophilin interactions with incoming human immunodeficiency virus type 1 capsids with opposing effects on infectivity in human cells.* J Virol, 2005. **79**(1): p. 176-83.
84. Franke, E.K., H.E. Yuan, and J. Luban, *Specific incorporation of cyclophilin A into HIV-1 virions.* Nature, 1994. **372**(6504): p. 359-62.
85. Yoo, S., et al., *Molecular recognition in the HIV-1 capsid/cyclophilin A complex.* J Mol Biol, 1997. **269**(5): p. 780-95.
86. Goujon, C., et al., *Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection.* Nature, 2013. **502**(7472): p. 559-62.
87. Kane, M., et al., *MX2 is an interferon-induced inhibitor of HIV-1 infection.* Nature, 2013. **502**(7472): p. 563-6.
88. Liu, Z., et al., *The interferon-inducible MxB protein inhibits HIV-1 infection.* Cell Host Microbe, 2013. **14**(4): p. 398-410.
89. Haller, O., et al., *Mx GTPases: dynamin-like antiviral machines of innate immunity.* Trends Microbiol, 2015. **23**(3): p. 154-63.
90. Busnadiego, I., et al., *Host and viral determinants of mx2 antiretroviral activity.* J Virol, 2014. **88**(14): p. 7738-52.
91. Goujon, C., et al., *Transfer of the amino-terminal nuclear envelope targeting domain of human MX2 converts MX1 into an HIV-1 resistance factor.* J Virol, 2014.
92. Fricke, T., et al., *MxB binds to the HIV-1 core and prevents the uncoating process of HIV-1.* Retrovirology, 2014. **11**: p. 68.

93. Opp, S., et al., *MxB Is Not Responsible for the Blocking of HIV-1 Infection Observed in Alpha Interferon-Treated Cells.* J Virol, 2016. **90**(6): p. 3056-64.
94. Fribourgh, J.L., et al., *Structural insight into HIV-1 restriction by MxB.* Cell Host Microbe, 2014. **16**(5): p. 627-38.
95. Matreyek, K.A., et al., *Host and viral determinants for MxB restriction of HIV-1 infection.* Retrovirology, 2014. **11**: p. 90.
96. Alvarez, F.J.D., et al., *CryoEM structure of MxB reveals a novel oligomerization interface critical for HIV restriction.* Sci Adv, 2017. **3**(9): p. e1701264.
97. Buffone, C., et al., *Contribution of MxB oligomerization to HIV-1 capsid binding and restriction.* J Virol, 2015. **89**(6): p. 3285-94.
98. Dicks, M.D., et al., *Oligomerization Requirements for MX2-Mediated Suppression of HIV-1 Infection.* J Virol, 2016. **90**(1): p. 22-32.
99. Schulte, B., et al., *Restriction of HIV-1 Requires the N-Terminal Region of MxB as a Capsid-Binding Motif but Not as a Nuclear Localization Signal.* J Virol, 2015. **89**(16): p. 8599-610.
100. Goujon, C., et al., *A triple-arginine motif in the amino-terminal domain and oligomerization are required for HIV-1 inhibition by human MX2.* J Virol, 2015. **89**(8): p. 4676-80.
101. Kong, J., et al., *Characterization of the amino-terminal domain of Mx2/MxB-dependent interaction with the HIV-1 capsid.* Protein Cell, 2014. **5**(12): p. 954-7.
102. Schaller, T., et al., *HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency.* PLoS Pathog, 2011. **7**(12): p. e1002439.
103. Gindhart, J.G., et al., *The kinesin-associated protein UNC-76 is required for axonal transport in the Drosophila nervous system.* Mol Biol Cell, 2003. **14**(8): p. 3356-65.
104. Alborghetti, M.R., et al., *Human FEZ1 protein forms a disulfide bond mediated dimer: implications for cargo transport.* J Proteome Res, 2010. **9**(9): p. 4595-603.
105. Gamble, T.R., et al., *Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein.* Science, 1997. **278**(5339): p. 849-53.
106. 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.
107. 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.
108. Grime, J.M. and G.A. Voth, *Early stages of the HIV-1 capsid protein lattice formation.* Biophys J, 2012. **103**(8): p. 1774-83.
109. Chen, B. and R. Tycko, *Simulated self-assembly of the HIV-1 capsid: protein shape and native contacts are sufficient for two-dimensional lattice formation.* Biophys J, 2011. **100**(12): p. 3035-44.

110. Tsiang, M., et al., *A trimer of dimers is the basic building block for human immunodeficiency virus-1 capsid assembly.* Biochemistry, 2012. **51**(22): p. 4416-28.
111. Kirchdoerfer, R.N., et al., *Pre-fusion structure of a human coronavirus spike protein.* Nature, 2016. **531**(7592): p. 118-21.
112. Yang, X., et al., *Highly stable trimers formed by human immunodeficiency virus type 1 envelope glycoproteins fused with the trimeric motif of T4 bacteriophage fibritin.* J Virol, 2002. **76**(9): p. 4634-42.
113. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
114. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Methods Enzymol, 1997. **276**: p. 307-26.
115. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.
116. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 235-42.
117. Vagin, A., Steiner, RS, Lebedev, AA, Potterton, L, McNicholas, S, Long, F and Murshudov, GN, *REFMAC5 dictionary: organisation of prior chemical knowledge and guidelines for its use.* Acta Cryst, 2004(D60).
118. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
119. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
120. Schrodinger, LLC, *The PyMOL Molecular Graphics System, Version 1.8*. 2015.
121. Scheres, S.H., *RELION: implementation of a Bayesian approach to cryo-EM structure determination.* J Struct Biol, 2012. **180**(3): p. 519-30.
122. Zakeri, B., et al., *Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin.* Proc Natl Acad Sci U S A, 2012. **109**(12): p. E690-7.
123. Dietrich, E.A., et al., *Variable prevalence and functional diversity of the antiretroviral restriction factor TRIMCyp in Macaca fascicularis.* J Virol, 2011. **85**(19): p. 9956-63.
124. Wagner, J.M., et al., *Mechanism of B-box 2 domain-mediated higher-order assembly of the retroviral restriction factor TRIM5alpha.* Elife, 2016. **5**.
125. Keown, J.R., et al., *Characterisation of assembly and ubiquitylation by the RBCC motif of Trim5alpha.* Sci Rep, 2016. **6**: p. 26837.
126. Keown, J.R. and D.C. Goldstone, *Crystal structure of the Trim5alpha Bbox2 domain from rhesus macaques describes a plastic oligomerisation interface.* J Struct Biol, 2016. **195**(3): p. 282-285.
127. Yudina, Z., et al., *RING Dimerization Links Higher-Order Assembly of TRIM5alpha to Synthesis of K63-Linked Polyubiquitin.* Cell Rep, 2015. **12**(5): p. 788-97.
128. Altschul, S.F., et al., *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.

129. Agarwal, S., et al., *Isolation, characterization, and genetic complementation of a cellular mutant resistant to retroviral infection.* Proc Natl Acad Sci U S A, 2006. **103**(43): p. 15933-8.
130. Slavoff, S.A., et al., *A human short open reading frame (sORF)-encoded polypeptide that stimulates DNA end joining.* J Biol Chem, 2014. **289**(16): p. 10950-7.
131. Price, A.J., et al., *CPSF6 defines a conserved capsid interface that modulates HIV-1 replication.* PLoS Pathog, 2012. **8**(8): p. e1002896.
132. Shi, J., et al., *Small-molecule inhibition of human immunodeficiency virus type 1 infection by virus capsid destabilization.* J Virol, 2011. **85**(1): p. 542-9.
133. Mallery, D.L., et al., *IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis.* Elife, 2018. **7**.
134. Dick, R.A., et al., *Inositol phosphates are assembly co-factors for HIV-1.* Nature, 2018. **560**(7719): p. 509-512.
135. Jacques, D.A., et al., *HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis.* Nature, 2016. **536**(7616): p. 349-53.
136. Morger, D., et al., *The Three-Fold Axis of the HIV-1 Capsid Lattice Is the Species-Specific Binding Interface for TRIM5alpha.* J Virol, 2018. **92**(5).