HIV-1 Gag Binding Specificity for Psi: Implications for Virus Assembly

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By
Shuohui Liu

Graduate Program in Chemistry

The Ohio State University

2017

Master's Examination Committee:

Dr. Karin Musier-Forsyth, Advisor

Dr. Thomas Magliery
Copyrighted by

Shuohui Liu

2017
Abstract

Human immunodeficiency virus type 1 (HIV-1) Gag is a polyprotein consisting of matrix (MA), capsid (CA), nucleocapsid (NC) and p6 domains, as well as two short spacer peptides, SP1 and SP2. During virus assembly, two copies of viral genomic RNA (gRNA) are selectively packaged into viral particles via specific interactions between Gag and gRNA packaging signal (Ψ), despite the vast abundance of cellular RNAs and spliced viral RNAs. However, the mechanism by which Gag specifically recognizes gRNA is not completely understood. In this study, several elements that are important for Gag’s binding specificity for Ψ were investigated using salt-titration binding assays. Salt-titration binding assays measure retroviral Gag’s ability to specifically recognize gRNA. Two parameters, $K_{d(1M)}$ and $Z_{\text{eff}}$, are obtained. $K_{d(1M)}$ measures the non-electrostatic component of binding and $Z_{\text{eff}}$ measures the effective charge of the binding, reflecting the electrostatic component.

In chapter 2, the role of MA domains from different retroviruses was tested. Even though MA does not directly interact with gRNA, it is still indispensable for maximum binding specificity in gRNA recognition. Chimeric proteins, wherein MA domains from HIV-1 and Rous sarcoma virus (RSV) were swapped, were made to test the specificity of MA. The results showed that MA affected RNA binding specificity of Gag in an indirect
manner. Additionally, individual clusters of exposed guanosine residues of HIV-1 Ψ were tested for the role in binding specificity.

In chapter 3, mutations of different Gag multimerization sites were tested. A mutation on the dimerization site in CA (WM mutation), reduced the binding specificity of Gag and changed the binding mode of Gag. This could be explained by different conformations of Gag when Gag interacts with Ψ RNA and non-Ψ RNA. Moreover, SP1 was demonstrated to contribute to the binding specificity for Ψ. A flexible linker, which disrupted the RNA binding specificity of Gag, displayed a role of SP1 as a transmitter. By combining the results from chapter 2 and chapter 3, a model describing the mechanism by which Gag specifically recognizes Ψ RNA and triggers virus assembly is proposed.

In conclusion, MA, CA, SP1, NC and Ψ RNA functions together to determine Gag’s binding specificity for Ψ and regulate HIV-1 assembly.
Acknowledgments

I would like to acknowledge my advisor, Dr. Karin Musier-Forsyth, for her conscientious support and patient guidance on this study. Without her help, I would never grow up and carry out my work so smoothly. I would also like to thank Dr. Thomas Magliery for being my examination committee.

I would like to thank Dr. Tiffiny Rye-McCurdy and Erik Olson for training me in this lab. Additionally, I would like to thank Christiana Binkley and Josh R. for working on the paper together. I also would like to thank Erik Olson and Dr. Ioulia Rouzina for any helpful discussions we had before. Moreover, I would like to thank the rest of smart people in my lab, for giving me any suggestions and letting me have a good time during the first year.

Finally, I would like thank The Ohio State University and the Department of Chemistry and Biochemistry for providing academic resources.
Vita

2015.................................................................B.S. Biology, Wuhan University, China

2015 to present ..............................................Graduate Teaching Associate, Department

of Chemistry, The Ohio State University

Publications


Fields of Study

Major Field: Chemistry
Table of Contents

Abstract ................................................................................................................................. ii

Acknowledgments ............................................................................................................... iv

Vita ......................................................................................................................................... v

List of Tables ......................................................................................................................... ix

List of Figures ......................................................................................................................... x

Chapter 1: Introduction ........................................................................................................ 1

1.1. AIDS ........................................................................................................................... 1

1.2. Replication cycle of HIV-1 ...................................................................................... 1

1.3. Role of the Gag polyprotein during HIV-1 replication cycle ...................................... 3

1.4. Genomic RNA and the importance of the 5'-UTR ...................................................... 7

1.5. Assembly of HIV-1 ....................................................................................................... 12

1.5.1. Overview of the assembly of HIV-1 ..................................................................... 12

1.5.2. In vitro assembly system ....................................................................................... 13

1.6. HIV-1 maturation ......................................................................................................... 15
1.7. Genome packaging selectivity ................................................................. 16

1.8. Current models for the selectivity of viral genome .................................... 18

1.8.1. 5’-UTR structural switch .................................................................... 18

1.8.2. Positive and negative regulatory elements in the viral genome .............. 19

1.9. Relation of thesis work to open questions in the retroviral field ............... 21

Chapter 2: Functional equivalence of retroviral MA domains in facilitating Psi RNA binding specificity by Gag ........................................................................... 22

2.1 Background ............................................................................................ 22

2.2 Materials and methods ............................................................................ 25

2.2.1 Plasmid and protein preparation ......................................................... 25

2.2.2 Preparation of RNA constructs .......................................................... 28

2.2.3 FA-based direct binding assay and salt titration binding assay ............... 29

2.3 Results ................................................................................................... 31

2.3.1 Optimization of the RNA concentrations used in the salt-titration binding assay ....................................................................................... 31

2.3.2 Functional equivalence of retroviral MA domain in discriminating Ψ RNA .... 33

2.3.3 Identifying HIV-1 Ψ residues required for Gag binding specificity........ 36

2.4 Discussion .............................................................................................. 39

Chapter 3: Contribution of CA and SP1 to Gag binding specificity for Psi ........ 45
3.1. Background .............................................................................................................. 45

3.2. Materials and methods ............................................................................................ 49
  3.2.1. Protein preparation ............................................................................................... 49
  3.2.2. Preparation of RNA constructs ........................................................................... 50
  3.2.3. FA-based direct binding assay and salt-titration binding assay ....................... 50

3.3. Results ...................................................................................................................... 50
  3.3.1. Mutation of the WM dimerization site converts the binding mode of Gag to NC-only ................................................................. 50
  3.3.2. Mutation on MA trimerization site does not affect the binding of GagΔp6 53
  3.3.3. SP1 alters the RNA binding specificity of Gag ................................................... 54

3.4. Discussion ................................................................................................................. 58
  3.4.1. WM mutation alters the RNA binding specificity of Gag ................................. 58
  3.4.2. Mutation of MA trimerization sites showes no effect on binding specificity of Gag........................................................................................................... 59
  3.4.3. SP1 contributes to the RNA binding specificity of Gag ............................. 60

3.5. A model describing Ψ RNA binding specificity of Gag ........................................ 61

3.6. Conclusion and future directions ............................................................................ 62

References ...................................................................................................................... 64
List of Tables

Table 1: Binding parameters determined by salt-titration binding assays with RSV GagΔPR and HIV-1 GagΔp6 at varying RNA concentrations. ................................................. 33

Table 2: Binding parameter determined by salt-titration binding assays of WT and chimeric Gag constructs. .................................................................................................................. 37

Table 3: Binding parameters determined from salt-titration binding assays of HIV-1 GagΔp6 with HIV-1 TARpolyA, Ψ-WT and Ψ RNA mutants ......................................................... 40

Table 4: Binding parameters determined from salt-titration binding assays using GagΔp6 and WM-GagΔp6. ........................................................................................................ 51

Table 5: Binding parameters determined from salt-titration binding assays using WT, L75E, L75G and A45E GagΔp6. .................................................................................................. 53

Table 6: Binding parameters determined from salt-titration binding assays using WT GagΔp6 and CA and SP1-deletion constructs. ................................................................. 56
List of Figures

Figure 1: Schematic of HIV-1 replication cycle................................................................. 3
Figure 2: Schematic illustration of HIV-1 Gag polyprotein.............................................. 4
Figure 3: Structural model of Gag protein......................................................................... 4
Figure 4: Organization of HIV-1 genome......................................................................... 8
Figure 5: Secondary structure prediction of 5'-UTR of HIV-1.......................................... 8
Figure 6: Comparison of HIV-1 immature virion and mature virion............................... 16
Figure 7: Two different conformations of viral genome proposed to function distinctly. 19
Figure 8: Model for regulation elements in the viral genome ......................................... 20
Figure 9: Model for genome selection................................................................................ 24
Figure 10: A phylogenetic tree showing the classification of retroviruses..................... 25
Figure 11: Protein constructs used in this work, with full-length RSV Gag and HIV-1 Gag shown at the top of each set for comparison................................................................. 26
Figure 12: End labeling with FTSC................................................................................... 29
Figure 13: Bar graphs showing $K_{d(1M)}$ values and $Z_{eff}$ values determined from salt titration binding assays using HIV-1 GagΔp6 with HIV-1 Ψ and HIV-1 TARpolyA, and RSV GagΔPR with RSV MΨ and RSV 167.................................................................................. 32
Figure 14: Bar graphs showing $K_{d(1M)}$ and $Z_{eff}$ values determined from salt-titration binding assays with RSV Gag∆PR, H132R, R155H and HIV-1 Gag∆p6 with RSV MΨ, RSV 167, HIV-1 Ψ and HIV-1 TARpolyA. ........................................ 36

Figure 15: HIV-1 RNA constructs used in this work. ........................................ 37

Figure 16: Bar graphs showing $K_{d(1M)}$ and $Z_{eff}$ values determined from salt-titration binding assays with HIV-1 Gag∆p6 and HIV-1 TARpolyA, Ψ RNA and Ψ RNA mutants. .............................................................. 39

Figure 17: The crystal structure showing trimeric HIV-1 MA. ................................ 46

Figure 18: Helical wheel showing that α-helical SP1 is amphipathic. ...................... 47

Figure 19: Model describing SP1-triggered HIV-1 virus assembly. ......................... 48

Figure 20: Protein constructs used in this work .................................................. 49

Figure 21: Bar graphs showing $K_{d(1M)}$ and $Z_{eff}$ determined from salt-titration binding assays using Gag∆p6 and WM-Gag∆p6 binding to HIV-1 Ψ, HIV-1 TARpolyA, SIV Ψ and SIV TAR. ................................................................. 52

Figure 22: Bar graphs showing $K_{d(1M)}$ and $Z_{eff}$ determined from salt-titration binding assays using WT, L75E, L75G, A45E Gag∆p6 binding to HIV-1 Ψ and HIV-1 TARpolyA. ................................................................. 54

Figure 23: Bar graphs showing $K_{d(1M)}$ and $Z_{eff}$ determined from salt-titration binding assays using Gag∆p6 and CA and SP1-deletion constructs binding to HIV-1 Ψ and HIV-1 TARpolyA. ................................................................. 57

Figure 24: Model for describing Ψ RNA binding specificity of Gag. ....................... 62
Chapter 1: Introduction

1.1. AIDS

Human immunodeficiency virus (HIV), which is the pathogen of acquired immunodeficiency syndrome (AIDS), threatens the life of humans throughout the world. So far there are no effective vaccines so precaution against the spreading of the disease is of great significance. There are two types of viruses that can cause AIDS: HIV-1 and HIV-2. HIV-1 is the main causative agent of AIDS all over the world; in contrast, HIV-2 infection happens mainly in Africa and leads to less severe consequences [1]. The primary focus of this work is HIV-1. HIV-1 can establish a state of latent infection without showing obvious symptoms and the presence of the latent reservoir is the major obstacle to curing HIV infection [2]. Alternatively, HIV-1 infection depletes the CD4+ T cells in the human body as well as the immune system [3]. Although the major way to prevent disease is by changing people’s behavior, significant effort has been put into developing therapies [4, 5]. Nevertheless, new strategies for therapies and effective drugs are needed and remain to be discovered.

1.2. Replication cycle of HIV-1

Because of the severe implications of HIV-1 infection, studying the HIV-1 replication cycle is of great importance. Figure 1 shows a schematic of HIV-1 replication
cycle. In general, the replication cycle starts when an infectious mature viral particle encounters a target cell, e.g. a CD4+ T cell. Envelope (Env) glycoprotein gp120 binds to the CD4 receptor with the help of co-receptors CXCR4/CCR5, facilitating fusion of both the viral and cellular membranes [6]. After entry, the viral core including the full-length viral RNA is released into the cytoplasm of the cell. Meanwhile, viral RNA is reverse transcribed into double-stranded cDNA, primed by human tRNA\(^{\text{Lys,3}}\) which is specifically packaged into virions together with tRNA\(^{\text{Lys1,2}}\). Following reverse transcription, a pre-integration complex (PIC) forms [7], which is transported into the nucleus [8]. In the nucleus, the viral cDNA is integrated into the host genome by viral integrase (IN) [9].

After integration, the late phase of the viral replication cycle begins. The integrated viral genome, which is referred as proviral DNA, is transcribed by the host cell machinery to generate full-length, unspliced mRNA, as well as spliced mRNAs [10]. Full-length RNA serves as both the viral genome specifically incorporated into the viral particles and the mRNA encoding Gag and Gag-pol polyproteins [11]. Gag, which will be discussed in more details in the next section, is the main structural protein of HIV-1 and drives virion assembly and release [12]. Gag-pol is produced by a ribosomal frameshift and contains coding regions for viral enzymes protease (PR), reverse transcriptase (RT) and IN [11]. Gag and Gag-pol are then transported to the plasma membrane (PM) where the immature viral particles form. Following assembly, immature viral particles bud and are released from the PM with the help of endosomal sorting complexes required for transport (ESCRT) [13, 14]. After release, immature viral particles undergo changes to form mature particles, triggered by PR-mediated cleavage of
Gag and Gag-pol [12]. At this moment, the mature viral particle is ready for a new round of infection.

**Figure 1:** Schematic of HIV-1 replication cycle, taken from [15].

### 1.3. Role of the Gag polyprotein during HIV-1 replication cycle

The Gag polyprotein consists of four major domains, matrix (MA), capsid (CA), nucleocapsid (NC) and p6, as well as two spacer peptides SP1 and SP2 which are located between CA and NC and between NC and p6, respectively (Figure 2 and Figure 3) [11].
Figure 2: Schematic illustration of HIV-1 Gag polyprotein, with arrows indicating the cleavage sites by PR, taken from [12].

Figure 3: Structural model of Gag protein, created by combination of different structures (2GOL, 1BAJ, 1MFS). Arrows indicate cleavage sites, taken from [12].

Matrix (MA)

The MA domain is a membrane-binding domain, comprising an N-terminal myristate moiety, as well as a highly basic region (HBR) [16]. The N-terminal myristate moiety is covalently attached to Gly-2 post-transcriptionally and Gly-2 is critical for
budding [17]. In addition, mutations in the HBR lead to a mistargeting of Gag to the late endosome [18]. Strikingly, MA can also interact with RNAs and it contributes to the binding specificity of Gag to the viral genome [19]. The PM has a specific phospholipid, phosphatidylinositol 4,5-bisphosphate (PI(4,5)P2), which specifically binds to MA, facilitating the specific targeting of Gag to the PM [20]. Depletion of PI(4,5)P2 causes a defect that is similar to the mutations in the HBR [21]. The myristate moiety has been shown to have two distinct conformations: One where the myristate is buried inside the hydrophobic core of MA and another where the myristate is exposed to the solvent. NMR studies have shown that PI(4,5)P2 triggers myristate moiety exposure, leading to the insertion and anchoring of Gag in the membrane [22]. Alternatively, in the absence of PI(4,5)P2, the myristate moiety is sequestered in the hydrophobic core of MA [23]. Structures show that MA can oligomerize to form trimers [24]. MA is also responsible for the incorporation of Env protein into viral particles during assembly [25]. Indeed, the trimer formation is relevant to the encapsidation of Env [26].

**Capsid (CA)**

The CA domain promotes Gag multimerization during assembly by providing multiple intermolecular interaction sites [12, 27]. CA folds into two domains, the N-terminal domain (NTD) and the C-terminal domain (CTD). The latter one is mainly responsible for the multimerization and formation of immature viral particles [28]. A proline-rich loop in the middle of the NTD and CTD is vital for the correct assembly of immature viral particles [11]. In helix 14 of the CTD, there are two amino acids “WM”
that are critical for dimerization of Gag. Mutation of “WM” leads to a defect of inter-Gag interaction [29]. After cleavage, in the mature viral particles, CA is responsible for the formation of the conical capsid core [30]. Structures of immature HIV-1 CA in viral particles have been determined and the oligomerization interfaces have also been defined [31].

**Nucleocapsid (NC)**

Overall, NC has two diverse functions. First, the NC domain of Gag mediates the specific recognition of genomic RNA (gRNA), resulting in the selective encapsidation of gRNA via the function of basic residues, as well as two zinc finger (ZF) motifs [19, 32]. A chimera in which NC is replaced by leucine zipper motifs that trigger multimerization shows similar efficiency of assembly as WT Gag, indicating a role for NC in bringing Gag molecules together [33]. This effect will be further discussed in section 1.5.2. Secondly, mature NC can also serve as a nucleic acid chaperone and function in several processes during the replication cycle, e.g., strand transfer during reverse transcription [34]. Compared to Gag, which also has a chaperone activity, NC has a better destabilization ability due to a higher dissociation rate [35]. NC’s chaperone activity is also mediated by the basic residues and ZFs [36, 37]. It has been shown that for genome packaging, the second ZF motif is more important while for chaperone activity, the first ZF motif is more critical [19, 38].
P6 domain

The P6 domain, which is also called the “L-domain” (or late domain), is indispensable during virus budding and release. It is also the domain that is in charge of recruitment of ESCRT machinery, composed of four protein complexes (ESCRT-0, I, II and III) [13, 14]. P6 contains two significant motifs: A Pro-Thr-Ala-Pro motif (PTAP) that binds to ESCRT-I, and a Tyr-Pro-Xn-Leu motif (YPXnL, where n is 1-4 amino acids) which binds to the ESCRT-associated factor Alix. Both motifs are suggested to be indispensable for the replication cycle [39, 40].

Spacer peptide 1 (SP1)

SP1 regulates the process of assembly. It can convert from a random coil to an α-helix, triggering the start of virus assembly [41]. The role of SP1 will be further discussed in Chapter 3. The SP1 region is also involved in the regulation of maturation and this will be discussed in section 1.6.

1.4. Genomic RNA and the importance of the 5’-UTR

All of the HIV-1 viral genes are encoded by the genomic RNA (gRNA) (Figure 4). The full-length gRNA encodes Gag and Gag-Pol, while the spliced RNAs encode Env and accessory proteins like Vif, Nef, Vpu, Vpr, Rev and Tat. Since unspliced RNA and incompletely spliced RNA are unable to be transported to the cytoplasm by a regular transportation pathway, they are exported from the nucleus to cytosol with the help of the Rev protein [12]. Like other retroviruses, genome packaging selectivity is achieved due
to several key elements located in 5'-untranslated region (5'-UTR). The predicted secondary structure of HIV-1 5'-UTR is shown in Figure 5. From the secondary structure, several stem-loops (SL) are predicted, including the trans-activation region (TAR), 5'-polyadenylation signal (PolyA), primer binding site (PBS), dimerization initiation site (DIS or SL1), major splice donor (SD or SL2) and Psi hairpin (Ψ or SL3). The function of the 5'-UTR will be discussed below and the mechanism of genome selection will be discussed in section 1.7 and 1.8.

Figure 4: Organization of HIV-1 genome. Adapted from [42].

Figure 5: Secondary structure prediction of 5'-UTR of HIV-1. Two alternative conformations are indicated. Taken from [43].
Trans-activation region (TAR)

The TAR hairpin is essential for Tat-mediated transcriptional activation [44]. Multiple roles of TAR in the HIV-1 replication cycle have been indicated, including strand transfer during reverse transcription [45]. The TAR hairpin contains a conserved U-rich region, which is the binding site of Tat protein [46]. The TAR hairpin has been identified as a second dimerization initiation site (DIS) in the absence of the primary DIS [47], although the significance of the dimerization ability still remains to be determined.

PolyA hairpin

The second hairpin in the HIV-1 5'-UTR is the PolyA hairpin, which contains the AAUAAA polyadenylation signal in the apical loop [48]. Indeed, little is known about the function of the PolyA hairpin. Previously, it was shown that mutations that tend to disrupt the secondary structure of the hairpin have a drastic effect on the packaging of gRNA [49, 50]. The PolyA hairpin is also suggested to be involved in a long-distance interaction (LDI). It is proposed to form base pairs with the DIS region [51] or may form a pseudoknot by base pairing with the MA coding region of the genome [52].

Primer binding site (PBS)

The PBS region is of great significance because it is the site that human tRNA^[Lys,3] anneals to and initiates reverse transcription [53, 54]. Recently, a small-angle X-ray scattering (SAXS)-derived ab initio envelope revealed the tertiary structure of the HIV-1 5'-UTR. The PBS region mimics the overall fold of tRNA [55]. This so-called tRNA-like
element (TLE) additionally binds to LysRS with a high affinity, thereby facilitating recruitment of the primer to the PBS for reverse transcription [56].

**Importance of dimerization and the DIS, SL1**

The HIV-1 genome is packaged as dimer, which renders the genome resistance for potential cleavage by nucleases [57, 58]. Since infected cells can have multiple integrated proviruses, there is a high possibility for heterozygous gRNAs to form dimers and be packaged, giving rise to the observed high rate of recombination [59]. HIV-1 genomes initially form weak and non-covalently linked dimers and the dimeric structure is further stabilized during maturation [60]. The cellular location of dimerization is a topic of intense investigation. Rous sarcoma virus (RSV) and Moloney murine leukemia virus (MoMuLV) genomes are likely to dimerize inside the nucleus [61]. Dimerization of MoMuLV in the nucleus is further suggested by the fact that the high probability of packaged homodimeric genomes, with a much smaller probability of recombination [62]. On the contrary, in the case of HIV-1, the proportion of homodimer and heterodimer correspond to random association, indicating that dimerization happens outside the nucleus [63]. However, whether dimerization happens in the cytoplasm or on the PM remains controversial. In 2016, one study using TIRF microscopy showed that dimerization of gRNA happens on the PM whereas in the same year another study using super-resolution and fluctuation microscopy concluded that the initial dimerization event happened in the cytoplasm [64, 65].
Because of the GC-rich palindromic loop, the DIS initiates the formation of a kissing loop complex between two copies of gRNA [66]. Mutations designed to disrupt the structure of DIS are detrimental to both genome packaging and infectivity of HIV-1 [67].

**Splice donor (SD), SL2**

During HIV-1 replication cycle, splicing is critical for translating multiple viral accessory proteins [42]. The SD loop is involved in the splicing events and it is the major splice donor, i.e., both the completely spliced viral RNAs and incompletely spliced viral RNAs contain regions upstream the SD loop [68]. SL2 is able to bind to NC with high affinity. Because of the exposed G-rich loop in the hairpin, the NC ZFs preferentially bind to single-stranded G-rich regions in SL2 [69]. Despite this fact, mutations designed to disrupt the secondary structure of SL2 didn’t disrupt genome packaging, indicating that this stem-loop is not critical for gRNA selection [70].

**Ψ, SL3**

Early studies suggested that SL3 is the stem-loop that is responsible for genome packaging. Deletions and mutations designed to disrupt the structure of SL3 showed large reduction in gRNA packaging [71]. Similar to SL2, SL3 also contains a G-rich loop and is capable of binding to NC with high affinity. However, more recent studies have shown that SL3 is less important in terms of genome packaging [72], due to the fact that early
mutations of SL3 may have affected the overall folding of the RNA. Indeed, SL1 was recently shown to be indispensable for the high affinity binding of full-length Gag [73].

1.5. Assembly of HIV-1

1.5.1. Overview of the assembly of HIV-1

The assembled virions contain all of the components that are critical for infectivity, including two copies of full-length, unspliced gRNA, cellular tRNA\textsuperscript{1lys}, Env, Gag, PR, RT, IN, as well as other regulatory viral proteins Vif, Vpr and Vpu [12]. The main component in a virion is the Gag polyprotein, constituting roughly half of the mass of the whole virion [74]. Since HIV-1 Gag is translated in polysomes, it needs to be trafficked to the PM to initiate the assembly. It has been shown that a certain detergent-resistant microdomain in the PM is the location of HIV-1 assembly [75]. Assembly is largely dependent on the cholesterol and sphingolipid content and the microdomain where assembly occurs is called a “lipid raft” [76]. A mass spectrometry study has indicated that the constitutive difference between the viral membrane and the PM, showing that in the “raft”, sphingomyelin, cholesterol and phosphatidyl ethanol (PE) and saturated fatty acids, are enriched [77]. Consistent with the idea that PI(4,5)P\textsubscript{2} plays a role in targeting Gag to the membrane, the viral membrane also contains a high content of PI(4,5)P\textsubscript{2} [78].

Gag tends not to form polymers in the cytoplasm but it forms higher-order oligomers on the PM [79]. Correspondingly, monomeric Gag proteins in the cytoplasm tend to adopt a compact conformation but will adopt an optimized conformation for
interactions between MA-membrane, NC-RNA and Gag-Gag after targeting to the PM [80, 81]. Recently, assembly of HIV-1 has been directly visualized by using live-cell microscopy and this indicates that the mobility of gRNA on the PM is reduced when gRNA is co-expressed with Gag [82]. From the same study, the authors determined that the approximate time for assembly is ~ 10 min [82]. Another live-cell imaging study has determined that assembly is nucleated mainly by Gag molecules that are directly from the pool of cytosolic Gag [83].

The corresponding product of assembly is called the immature virion. In the immature virion, Gag molecules are radial, with the N-terminus of MA facing the exterior and associating with the viral membrane and the C-terminal p6 domain facing the interior near the center of the virion [84]. By using cryo-electron tomography, reconstruction of three-dimensional structures of authentic HIV-1 virion cores was achieved [85]. Cryo-electron tomography showed the arrangement of Gag in the immature virion as an incomplete hexameric lattice containing holes [86]. Recently, a crystallography study showed that CA-CTD and SP1 fragments form a goblet-shaped hexamer [87]. Moreover, another cryo-electron tomography study showed a similar hexameric CA-SP1 structure within the immature HIV-1 particles in high resolution, indicating a regulatory role of this region in assembly and maturation [88].

1.5.2. In vitro assembly system

To simplify the study of viral assembly, in vitro assembly systems have been developed. In an in vitro assembly system, Gag is induced to assemble into particles that
resemble real viral particles, which are called “virus-like particles (VLP)” [89]. Aqueous solutions of Gag can form VLPs only after addition of nucleic acid. Surprisingly, even short nucleic acids can trigger this *in vitro* assembly [89].

It is believed that an increase of the local concentration of Gag is a prerequisite for assembly. Three factors were proposed to explain this phenomenon. First, Gag binding to nucleic acids increases the local concentration of Gag near the nucleic acid [90]. Second, oligomerization of Gag molecules, especially the dimerization of Gag via CA-CTD has similar effect. Third, PM binding is another possible way to increase local concentration of Gag, considering that PM binding constricts the space that Gag molecules can occupy [90]. *In vitro* assembly experiments are consistent with these three factors contributing to the assembly of HIV-1 [91]. It has been shown that when only one of these factors is defective, Gag still forms viral particles that have roughly the correct shape with little defect [91]. However, any combination of two leads to a complete incapability of forming particles with the correct shape and curvature, indicating that the factors work together to trigger assembly [91]. Remarkably, the successful assembly of VLPs requires inositol hexakisphosphate (IP6). In the absence of IP6, Gag assembles into 30 nm diameter particles whereas in the presence of IP6, Gag assembles into ~100-150 nm diameter particles, which correspond to the correct size for virions [92]. This difference in particle size is suggested to be relevant to the different conformations of Gag and this will be discussed in detail in Chapter 3.
1.6. **HIV-1 maturation**

Viral maturation begins at the same time as budding or slightly later than budding (budding is reviewed elsewhere [12, 93, 94] and will not be discussed here). This process is mediated by cleavage of PR, generating protein segments MA, CA, NC, p6, PR, RT and IN. Other changes take place as well, such as activation of viral Env protein [95, 96], stabilization of gRNA dimer and rearrangement of tRNA\textsuperscript{Lys,3} primer-genome complex [34, 60]. Varied rates of different cleavage sites result in an ordered maturation process. The relative rate of Gag cleavage sites is summarized as SP1/NC >> SP2/p6 ~ MA/CA >> NC/SP2 ~ CA/SP1 [97]. A putative maturation switch spans the SP1 region [98]. This maturation switch, as well as the detailed structural information of CA/SP1 junction, has been determined by cryo-EM and crystallization [87, 88].

Drastic structural changes happen during maturation (Figure 6). In the mature virions, MA remains associated with the viral membrane, whereas CA forms a conical shell of the core particle. Within the core particle, NC is surrounded by CA and is associated with two copies of the viral genome [12]. Unlike immature viral particles, HIV-1 mature viral particles contain continuous hexagonal units with a few pentagonal defects [99]. In order to maintain the structure of mature lattice, CA is involved in multiple inter-molecular interactions, including a 6-fold axes interaction via CA-NTD contacts and 2, 3-fold axes interactions via CA-CTD contacts [100, 101].
**Figure 6**: Comparison of HIV-1 immature virion (A) and mature virion (B). Adapted from [12].

1.7. **Genome packaging selectivity**

During assembly, two copies of the full-length, unspliced gRNA are specifically encapsidated into newly assembled virus particles, despite the vast excess of cellular RNA and spliced viral RNA [102, 103]. How the genome is selectively packaged is still not completely understood. Genome packaging is mediated by Gag protein. To be more specific, selection of the viral genome is mediated via the interaction between NC and the genome packaging signal, termed Psi (Ψ), which is located in the 5'-UTR of the gRNA [32]. However, high-affinity binding is not enough to explain the selective packaging because most RNAs have similar dissociation constants upon binding to Gag under physiological condition [104]. The genome plays a significant role in assembly, despite the fact that the role of gRNA can be replaced by the function of cellular mRNAs during assembly [105]. What is interesting is that certain types of cellular RNA, like 7SL and
tRNA<sup>Lys</sup>, are also relatively selectively packaged into virions, although to a smaller extent [106]. For other retroviruses like RSV and MoMuLV, the smallest elements that are sufficient for directing RNA packaging have been identified [107-109], but in the case of HIV-1, the minimal packaging signal is hard to identify. Moreover, it seems that the whole 5'-UTR and part of the gag coding region are both required for efficient packaging [110-112]. However, recently, a 159-residue RNA packaging signal within HIV-1 5'-UTR showing similar packaging efficiency and NC binding properties as the intact 5'-UTR was identified [113].

**Role of NC in the genome selection**

The NC domain is regarded as indispensable during the selection of the HIV-1 viral RNA genome. Substitution of HIV-1 NC by MoMuLV NC or substitution of MoMuLV NC by HIV-1 NC lead to preferential packaging of MoMuLV and HIV-1 genomes, respectively [114, 115]. However, swapping of NC domains of HIV-1 and mouse mammary tumor viruses leads to the original packaging preference, indicating that other elements must be involved in the genome selection [116]. HIV-1 NC proteins contain two copies of the CCHC zinc knuckle motifs that can bind to zinc ion with high affinity [117, 118]. It has been suggested that zinc knuckles play an central role in genome selection [119].
Role of MA in the genome selection

MA plays an essential role in membrane targeting. Therefore, it is important to understand whether MA can also contribute to selectivity or not [18]. Because MA is highly basic, it is able to bind to nucleic acids with high affinity [81]. Recently it has been shown that liposomes containing PI(4,5)P_2 can compete with nucleic acids for MA binding but those that lack PI(4,5)P_2 cannot compete, suggesting the role of RNA in membrane targeting [120-122]. Furthermore, although MA itself seems not distinguish between the viral genome and other RNAs, MA does contribute to the selectivity of viral genome binding [19]. It has also been shown that tRNA binding to MA regulates its functions [123, 124]. Despite the recent discoveries, the exact role of MA in genome selection still remains to be clarified.

1.8. Current models for the selectivity of viral genome

1.8.1. 5'-UTR structural switch

Full-length gRNA can serve as either mRNA used for translation of Gag and Gag-Pol or gRNA that is selectively packaged into viral particles. A possible mechanism of functional regulation was recently suggested (Figure 7) [43]. It has been proposed that there are two conformers of the 5'-UTR: In the “AUG start codon exposed” conformer, the U5 region base pairs with the DIS, thus exposing the AUG start codon and sequestering DIS. This conformation favors translation but the dimerization of gRNA is not favorable. Considering the fact that the dimerized genome is preferentially packaged, this conformation doesn’t promote packaging. In contrast, another conformer is “DIS
exposed” where U5 base pairs with the AUG region. In this conformation, AUG is sequestered, which means that translation is not favorable. As a result, this conformation promotes dimerization and packaging. Recently, it has been shown that proviral DNA of HIV-1 can have multiple transcriptional start sites (TSS), causing transcripts with different numbers of guanosines at the beginning [125]. It has been proposed that 1G RNA is dimeric which promotes packaging, whereas 2G and 3G RNAs are monomeric which promote translation, corresponding to those conformations [126].

![Diagram of two conformational states of viral genome](image)

**Figure 7:** Two different conformations of viral genome proposed to function distinctly. Taken from [43].

### 1.8.2. Positive and negative regulatory elements in the viral genome

How Gag discriminates and selectively package full-length, unspliced gRNA against spliced RNA is an intriguing question. As discussed in section 1.4, previously it was believed that SL3 is the major structural element that is responsible for the selective packaging of gRNA. Since SL3 is downstream of the splice donor, only unspliced gRNA contains SL3, providing a possible mechanism for Gag to distinguish between spliced
and unspliced gRNA. However, more recent studies have indicated that SL1 is the Psi element that is more important [72, 127, 128]. In one model (Figure 8), it is proposed that RNA upstream of the packaging signal serves as an inhibitory element because RNA constructs that contain the whole packaging signal, as well as upstream elements (TAR, PolyA and PBS), show less efficient binding to Gag [73]. Moreover, the inclusion of a downstream element restores the high binding affinity of RNA to Gag, indicating that a downstream region serves as an enhancer element. In addition, it has also been proposed that the downstream positive regulatory element suppresses the inhibitory element rather than directly contributing to the high affinity. This provides new insights into the mechanism of selectivity: Only unspliced gRNA contains the positive regulatory element, counteracting the effect of the negative regulatory element, while spliced gRNAs lack the positive regulatory element, leading to the selective packaging of full-length gRNA.

Figure 8: Model for regulation elements in the viral genome. Taken from [73]
1.9. *Relation of thesis work to open questions in the retroviral field*

Despite numerous studies in the retroviral field as shown before in this chapter, there are still a lot of open questions remained to answer. Several most important open questions nowadays are: By which mechanism is $\Psi$ specifically recognized by HIV-1 Gag and packaged into HIV-1 virions? Despite the MA and NC domains, are other domains also involved in determining the binding specificity of Gag? How does MA determine the RNA binding specificity of Gag? How exactly does $\Psi$ trigger the HIV-1 assembly?

This work dedicates to determine the key elements that are important for the RNA binding specificity of HIV-1 Gag and virus assembly. In chapter 2, MA was shown to contribute to the RNA binding specificity of Gag indirectly. Several exposed guanosine residues in great proximity were also shown to be involved in Gag-gRNA interaction. In chapter 3, the critical role of SP1 and CA was demonstrated by investigating several mutants using salt-titration binding assays. Finally, a model is proposed to describe how recognition of gRNA by Gag triggers the HIV-1 assembly.
Chapter 2: Functional equivalence of retroviral MA domains in facilitating Psi RNA binding specificity by Gag

2.1 Background

Despite the vast abundance of cellular RNA and spliced vRNA, a pair of HIV-1 full-length, unspliced dimeric gRNA molecules is specifically incorporated into immature virions. However, the mechanism behind this packaging specificity is still not completely understood. Although the minimal element within the HIV-1 genome for selective packaging is not determined, a “core encapsidation signal” (ΨCES) has been proposed [113]. It is obvious that Ψ (packaging signal) provides an advantage to be packaged because only in the absence of Ψ can the cellular RNA be incorporated randomly [129]. Exposed G residues within Ψ were shown to be high-affinity binding sites for the NC domain. This is supported by different experimental approaches such as selective 2’-hydroxyl acylation analyzed by primer extension (SHAPE) [130], in vitro RNA binding assays [73] and crosslinking-immunoprecipitation-sequencing (CLIP-seq) studies [123].

Salt-titration binding assays have also been used to measure retroviral Gag’s ability to specifically recognize gRNA [19, 127]. Two important parameters, $K_d(1\text{M})$ and $Z_{\text{eff}}$, can be obtained by this technique. $K_d(1\text{M})$ measures the non-electrostatic component of binding and is obtained from extrapolation by theoretically increasing the salt concentration to 1 M, where electrostatic interactions are ruled out completely [131]. $Z_{\text{eff}}$ measures the effective charge of the binding, reflecting the process in which Na$^+$ gets
displaced by electrostatic contacts [131]. More detailed information about this technique will be given in Section 2.2.3.

The NC domain alone is able to differentiate Ψ RNA versus non-Ψ RNA [19], as demonstrated previously in our lab using the fluorescence anisotropy (FA)-based salt-titration binding assay. However, Gag binds to Ψ RNA with an even greater non-electrostatic binding component and a low effective charge. In contrast, for non-Ψ RNA, the binding is characterized by a smaller non-electrostatic binding component and a higher effective charge [19]. For either Ψ RNA or non-Ψ RNA binding, another construct CANC (or ∆MA) shows similar binding behavior to NC alone. This work suggested that the difference in the effective charge of Gag binding to Ψ RNA versus non-Ψ RNA is due to the contribution of MA [19]. Thus, a model was proposed herein (Figure 9). Ψ RNA binding corresponds to an NC-only binding mode, whereas non-Ψ RNA binding corresponds to an NC & MA binding mode [19]. In the case of Ψ RNA binding, MA is free to interact with the plasma membrane. Therefore, this binding mode is more favorable for assembly. Indeed, an important implication is that different conformations of Gag may contribute to the different RNA binding modes and this idea will be discussed in more detail in Chapter 3.

Whether other retroviruses have similar mechanisms for Ψ RNA recognition by Gag is unknown. A phylogenetic tree describing the classification of different types of retroviruses is shown in Figure 10. Rous sarcoma virus (RSV) is a typical example of an Alpha-retrovirus. It is a cancer-causing retrovirus, which causes sarcoma in chicken. Similar to HIV-1, the 5'-UTR of RSV plays an important role in selective genome
packaging. Moreover, the minimal element that directs specific packaging of heterologous RNAs with a similar efficiency as the full-length genome of RSV has been identified. Initially this element was called MΨ and includes nucleotides (nt) 156-315 in the 5'-UTR [108, 132]. More recently, it was shown that the minimal region is nt 156-237 and this region is designated μΨ [107, 133]. Similar to HIV-1, the interaction between RSV NC and the Ψ region of RSV is believed to be indispensable for selective packaging of the RSV genome [134-136]. Similar to HIV-1, The 5'-UTR of RSV also has a complex predicted secondary structure where several stem loops have been implicated to be responsible for specific genome recognition by NC [137-139].

Figure 9: Model for genome selection. There are two distinct Gag conformations and binding modes for Ψ RNA (right) and non-Ψ RNA (left). Taken from [19].
In this work, the role of HIV-1 and RSV MA in gRNA recognition was investigated. Preliminary studies suggested that MA was dispensable for maximum specificity, even though MA does not directly interact with the gRNA. Chimeric proteins were made to test the specificity of MA. Furthermore, individual exposed guanosines of HIV-1 Ψ were investigated for their potential role in binding specificity. Overall, this work identified domains of Gag and viral gRNA elements that influence the binding specificity of retroviral Gag.

2.2 Materials and methods

2.2.1 Plasmid and protein preparation

The protein constructs that were used in this work are shown in Figure 11.
Figure 11: Protein constructs used in this work, with full-length RSV Gag and HIV-1 Gag shown at the top of each set for comparison. In the case of chimeras, the residues at the junctions are explicitly shown. H132R contains the 132-residue HIV-1 MA in place of RSV MA in the context of RSV Gag, and R155H contains the 155-residue RSV MA in place of HIV-1 MA in the context of HIV-1 Gag.

The plasmid encoding HIV-1 GagΔp6 was a gift from Dr. Alan Rein (HIV dynamics and Replication Program, Center for Cancer Research, National Cancer Institute, Frederick, MD, USA). HIV-1 GagΔp6 was expressed in E. coli and purified by an established protocol [140]. Briefly, when OD$_{600}$ of the culture was in the range of 0.6 - 0.8, protein expression was induced by 0.1 mg/ml isopropyl β-D-1-thiogalactopyranoside (IPTG, Roche Life Sciences) for 4 h at 37 °C. The pellets were resuspended in Lysis Buffer (20 mM Tris HCl, pH 7.4, 10 mM 2-mercaptoethanol (βME), 1 μM ZnCl$_2$, 750 mM NaCl, 10% glycerol, 0.05% Trition X-100 and 1 cOmplete™ Mini ethylenediaminetetraacetic acid (EDTA)-free Protease Inhibitor Cocktail Tablet (Roche Life Sciences)). After resuspension, polyethylenimine (PEI, Sigma-Aldrich) precipitation, which removes nucleic acids, was performed by adding PEI to a final concentration of 0.15%. This was followed by 40% ammonium sulfate precipitation, which removes impurities and Gag aggregates. The fractions that were recovered were incubated with...
pre-equilibrated cellulose phosphate (PC resin) for 30 min, followed by stepwise elution with NaCl concentration gradient (100 mM, 200 mM, 500 mM, 1000 mM). Eluted protein fractions were monitored by SDS-PAGE and the desired fractions were collected and precipitated overnight by 0.375 g/ml ammonium sulfate. The next day, the precipitated proteins were recovered by Dialysis Buffer (20 mM Tris-HCl, pH 7.4, 500 mM NaCl, 10% glycerol, 10 mM βME, 1 µM ZnCl₂) to remove remaining cellulose phosphate particles. The recovered samples were purified on a HiLoad 16/600 Superdex 200 prep grade column (GE Healthcare) run in Dialysis Buffer.

RSV GagΔPR plasmid pET28(-His).RSVGagΔPR was derived from the Prague C strain of RSV [141]. Purification of RSV GagΔPR was similar to HIV-1 GagΔp6, except for the contents of the Lysis Buffer (50 mM Tris-HCl, pH 7.4, 700 mM NaCl, 0.1 mM βME, 10% glycerol, 0.1% Triton X-100 and 1 Protease Inhibitor Cocktail Tablet) and Dialysis Buffer (10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.4, 500 mM NaCl, 0.1 mM ZnCl₂, 0.1 mM EDTA and 1 mM βME) [142]. The plasmid encoding the R155H chimera was prepared by amplifying RSV MA from pET28(-His).RSVGagΔPR by PCR, followed by NdeI digestion and ligation. The method for purifying R155H was the same as the preparation of RSV GagΔPR. H132R, which contains a His₆ tag, was purified by affinity chromatography using Ni-NTA resin as described [142]. Concentrations of all of the proteins were determined by measuring the absorbance at 280 nm, and using the following molar extinction coefficients: 63,090 M⁻¹ cm⁻¹ (HIV-1 GagΔp6), 63,348 M⁻¹ cm⁻¹ (RSV GagΔPR), 67,295 M⁻¹ cm⁻¹ (R155H) and 59,275 M⁻¹ cm⁻¹ (H132R).
2.2.2 Preparation of RNA constructs

RNA constructs were in vitro transcribed from linearized plasmids by T7 RNA polymerase using published methods [143]. DNA sequences encoding the RSV MΨ and RSV 167 were synthesized and cloned into pIDTSMART-AMP by Integrated DNA Technologies (IDT). RSV 167 was derived from nt 1249-1409, whereas RSV MΨ was derived from nt 156-315 of the RSV Prague C strain genome. Each RNA was designed with two additional non-genomic guanosines added at the 5’-end to increase the efficiency of transcription. HindIII was used to linearize the plasmid for transcription. HIV-1 Ψ and TARpolyA were derived from nt 229-333 and nt 1-104 of the HIV-1 NL4-3 genome [144]. Plasmids were linearized by FokI digestion to generate the correct 3’-end of the templates. Two additional guanosines at the 5’-end were also added to promote the transcription. HIV-1 Ψ mutants were obtained by site-directed ligase-independent mutagenesis (SLIM) [145] and were designated as follows: Ψ-Mut1 (G241, 242A), Ψ-Mut2 (G274, 275A), Ψ-Mut3 (G292A, U293A, G294A), Ψ-Mut4 (G311A), Ψ-Mut5 (G319A, G321A), Ψ-Mut6 (G239, 330A), Ψ-Mut7 (G235A) and Ψ-Mut8 (G267, 268A). RNAs were labeled with fluorescein-5-thiosemicarbazide (FTSC) at the 3’-end using an established method (Figure 12) [146]. Concentrations of all of the labeled RNA constructs and labeling efficiency were determined by measuring the absorbance at 260 nm and 495 nm and using the following molar extinction coefficients: \( \varepsilon_{260nm} = 9.3 \times 10^5 \text{ M}^{-1} \cdot \text{cm}^{-1} \) (HIV-1 Ψ and HIV-1 Ψ mutants 1-8), \( \varepsilon_{260nm} = 9.7 \times 10^5 \text{ M}^{-1} \cdot \text{cm}^{-1} \) (HIV-1 TARpolyA).
Figure 12: End labeling with FTSC. The 3'-end of an RNA is first oxidized, followed by the attachment of the FTSC dye. Taken from [147].

2.2.3 **FA-based direct binding assay and salt titration binding assay**

Direct and salt-titration binding assays were performed according to established methods [19, 131]. A FA-based direct binding assay was first performed to obtain a preliminary assessment of binding affinity and to determine the protein concentrations used in the salt-titration binding assay. The assays were performed at 20 mM HEPES, pH 7.5, 1 mM MgCl₂, 50 mM NaCl and 1.5 nM RNA (for RSV RNAs) or 20.5 nM RNA (for HIV-1 RNAs). Briefly, RNAs were folded in 50 mM HEPES, pH 7.4 by heating at 80 °C for 2 min, cooling at 60 °C for 2 min, adding 0.1 M MgCl₂ to a final concentration of 10 mM and cooling on ice for at least 30 min before mixing with proteins of different concentrations and the final reaction buffer. The final reaction mixtures were incubated at room temperature in the dark for at least 30 min to reach equilibrium. The FA and
fluorescence intensity were measured using a SpectraMax M5 plate reader (Molecular Devices). FA data were fitted to equation 1:

\[ A(P) = A_{\text{min}} + \theta(P) \cdot (A_{\text{max}} - A_{\text{min}}) \] (Equation 1),

where \( A_{\text{min}} \) and \( A_{\text{max}} \) are the FA values of the unbound and saturated bound RNA, respectively. Fraction of the RNA that is bound, \( \theta(P) \), is determined by the binding isotherm (equation 2):

\[ \theta(P) = \frac{(P+K_d+nR-\sqrt{(P+K_d+nR)^2-4nP})}{2nP} \] (Equation 2),

where \( P \) is the total concentration of the protein, \( K_d \) is the dissociation constant of the binding, \( n \) is the number of protein-binding sites per RNA molecule (usually set to 1) and \( R \) is the total concentration of RNA.

Salt-titration binding assays were performed with 1.5-20.5 nM RNA in 20 mM HEPES, pH 7.5, 1 mM MgCl\(_2\), 400 nM HIV-1 Gag\(\Delta p6\), 400 nM R155H, 300 nM H132R and 300-500 nM RSV Gag\(\Delta PR\). In the final reaction mixture, NaCl concentrations varied (from 50 mM to 1 M). To correct for the FA changes due to the increasing NaCl concentration, FA values in the absence of protein were subtracted from the FA values for protein-RNA mixtures at the same salt concentrations. The data were analyzed as described [19, 131]. Briefly, the dissociation constant, \( K_d \), as a function of Na\(^+\) concentrations is described by equation 3:

\[ K_d = K_{d(1M)} \times [Na^+]^{Z_{\text{eff}}} \] (Equation 3),

where \( K_{d(1M)} \) is the extrapolated dissociation constant of the RNA-protein interaction at 1 M NaCl concentration when all electrostatic charges are screened out. This value reflects the strength of the non-electrostatic binding component. \( Z_{\text{eff}} \) reflects the number of
electrostatic contacts in the RNA-protein interaction. Equation 3 is substituted into Equation 2, leading to the determination of the two parameters $K_{d(1M)}$ and $Z_{eff}$.

2.3 Results

2.3.1 Optimization of the RNA concentrations used in the salt-titration binding assay

Surprisingly, we found that the Gag binding specificity for $\Psi$ might be different under different RNA concentrations (data not shown). In order to determine the optimal concentrations, a range of $\Psi$ RNA and non-$\Psi$ RNA concentrations (1.5 nM – 20.5 nM) was tested using salt-titration binding assays (Figure 13 and Table 1). Interestingly, the specificity for cognate $\Psi$ RNAs varied largely under different RNA concentrations. For HIV-1, at 20.5 nM RNA concentration (1:20 RNA:protein ratio), Gag$\Delta$p6 showed the highest specificity for HIV-1 $\Psi$ ($K_{d(1M)} \sim 3.6 \times 10^{-5}$ M). In contrast, at 1.5 nM RNA concentration, it showed the lowest specificity for HIV-1 $\Psi$ ($K_{d(1M)} \sim 5.4 \times 10^{-3}$ M). Indeed, the specificity for HIV-1 $\Psi$ increased as the RNA concentration increased (Figure 13A, Table 1). Meanwhile, the $K_{d(1M)}$ measured for HIV-1 Gag$\Delta$p6 binding to HIV-1 TARpolyA remained largely unchanged. RSV$\Delta$PR, however, displayed the opposite trend, showing the greatest specificity ($K_{d(1M)} \sim 7.1 \times 10^{-5}$ M) for RSV M$\Psi$ at 1.5 nM RNA (1:200 RNA:protein ratio). At 20.5 nM RNA concentration, RSV Gag$\Delta$PR showed the lowest specificity for RSV M$\Psi$ ($K_{d(1M)} \sim 1.6 \times 10^{-3}$ M). Although there was a significant fluctuation in the $K_{d(1M)}$ values measured for cognate non-$\Psi$ RNA, RSV Gag$\Delta$PR bound to RSV 167 non-specifically at all RNA concentration tested. Taken together, both HIV-1 Gag$\Delta$p6 and RSV Gag$\Delta$PR showed different specificity for cognate
RNAs at different RNA concentrations. For the remaining studies, we used the RNA concentration resulting in the greatest binding specificity for cognate \( \Psi \) RNA, i.e., 20.5 nM for HIV-1 \( \Psi \) and HIV-1 TARpolyA (1:20 RNA:protein ratio) and 1.5 nM for RSV M\( \Psi \) and RSV 167 (1:200 RNA:protein ratio).

**Figure 13:** Bar graphs showing \( K_{d(1M)} \) values and \( Z_{\text{eff}} \) values determined from salt titration binding assays using HIV-1 Gag\( \Delta p6 \) with HIV-1 \( \Psi \) and HIV-1 TARpolyA (A, C), and RSV Gag\( \Delta \text{PR} \) with RSV M\( \Psi \) and RSV 167 (B, D). RNA concentrations range from 1.5 nM to 20.5 nM. Values of three trials performed in each case are shown with the height of the bar indicating the mean value.
Table 1: Binding parameters determined by salt-titration binding assays with RSV GagΔPR and HIV-1 GagΔp6 at varying RNA concentrations.

<table>
<thead>
<tr>
<th>RNA-Concentration (nM)</th>
<th>RSV GagΔPR</th>
<th>HIV-1 GagΔp6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_d(1M)$ (M)</td>
<td>$(7.1 \pm 5) \times 10^{-5}$</td>
</tr>
<tr>
<td>Cognate Ψ RNA-1.5$^a$</td>
<td>$Z_{eff}$</td>
<td>$3.7 \pm 0.5$</td>
</tr>
<tr>
<td></td>
<td>$K_d(1M)$ (M)</td>
<td>$(1.3 \pm 2) \times 10^{-1}$</td>
</tr>
<tr>
<td>Cognate Non-Ψ RNA-1.5$^b$</td>
<td>$Z_{eff}$</td>
<td>$7.0 \pm 1.2$</td>
</tr>
<tr>
<td></td>
<td>$K_d(1M)$ (M)</td>
<td>$(3.0 \pm 1) \times 10^{-4}$</td>
</tr>
<tr>
<td>Cognate Ψ RNA-4$^a$</td>
<td>$Z_{eff}$</td>
<td>$4.2 \pm 0.4$</td>
</tr>
<tr>
<td></td>
<td>$K_d(1M)$ (M)</td>
<td>$(1.2 \pm 0.9) \times 10^{2}$</td>
</tr>
<tr>
<td>Cognate Non-Ψ RNA-4$^b$</td>
<td>$Z_{eff}$</td>
<td>$11.6 \pm 0.4$</td>
</tr>
<tr>
<td></td>
<td>$K_d(1M)$ (M)</td>
<td>$(9.3 \pm 7) \times 10^{-4}$</td>
</tr>
<tr>
<td>Cognate Ψ RNA-8$^a$</td>
<td>$Z_{eff}$</td>
<td>$4.7 \pm 1.0$</td>
</tr>
<tr>
<td></td>
<td>$K_d(1M)$ (M)</td>
<td>$(4.0 \pm 4) \times 10^{1}$</td>
</tr>
<tr>
<td>Cognate Non-Ψ RNA-8$^b$</td>
<td>$Z_{eff}$</td>
<td>$10.1 \pm 1.2$</td>
</tr>
<tr>
<td></td>
<td>$K_d(1M)$ (M)</td>
<td>$(1.6 \pm 1) \times 10^{-3}$</td>
</tr>
<tr>
<td>Cognate Ψ RNA-20.5$^a$</td>
<td>$Z_{eff}$</td>
<td>$5.4 \pm 0.6$</td>
</tr>
<tr>
<td></td>
<td>$K_d(1M)$ (M)</td>
<td>$(2.0 \pm 1) \times 10^{-1}$</td>
</tr>
<tr>
<td>Cognate Non-Ψ RNA-20.5$^b$</td>
<td>$Z_{eff}$</td>
<td>$7.5 \pm 0.6$</td>
</tr>
</tbody>
</table>

$^a$ Cognate Ψ RNA is HIV-1 Ψ for HIV-1 GagΔp6 and RSV MΨ for RSV GagΔPR.

$^b$ Cognate non-Ψ RNA is TARpolyA for HIV-1 GagΔp6 and RSV 167 for RSV GagΔPR.

2.3.2 Functional equivalence of retroviral MA domain in discriminating Ψ RNA

Previously, it was shown that HIV-1 GagΔp6 binds to Ψ and TARpolyA (non-Ψ RNA control) with similar affinity ($K_d \sim 50$ nM) [19], indicating that affinity itself cannot explain the specificity difference between Ψ and non-Ψ. Moreover, it was shown that HIV-1 MA contributes to the binding specificity [19]. Similarly, in the case of RSV, direct binding assays were also performed for RSV GagΔPR binding to RSV MΨ or RSV 167. These RNAs showed similar Gag binding affinity ($K_d \sim 15-18$ nM) [127].
Furthermore, by performing salt-titration binding assays, RSV GagΔPR binding to RSV MΨ was shown to have a ~1900-fold stronger non-electrostatic binding component and 3 fewer electrostatic contacts than binding to RSV 167 [127]. More strikingly, RSV CANC bound to RSV MΨ and RSV 167 with a similar non-electrostatic binding strength and similar electrostatic contacts [127]. This indicates that in the case of RSV, like HIV-1, MA also contributes to Gag’s ability to distinguish between Ψ and non-Ψ RNA.

The exact mechanism by which MA contributes to the binding specificity of Gag is unknown. In order to investigate this question, two protein chimeras were prepared wherein MA domains from HIV-1 and RSV were swapped. In R155H, HIV-1 MA was replaced by RSV MA in the context of HIV-1 GagΔp6, whereas in H132R, RSV MA was replaced by HIV-1 MA in the context of RSV GagΔPR (Figure 11). Using salt-titration binding assays, we found that H132R bound to RSV MΨ with high specificity (K₁M(1M) ~ 7.2 × 10⁻⁵ M), which was similar to RSV GagΔPR (K₁M(1M) ~ 7.1 × 10⁻⁵ M). The Z eff (~ 5) is slightly higher than that measured for RSV GagΔPR (~ 4). Moreover, H132R bound to RSV 167 non-specifically with a K₁M(1M) (~ 1.9 M) and Z eff (~ 9) that are similar to RSV GagΔPR (Figure 14 and Table 2). This indicates that RSV Gag still can distinguish cognate Ψ from non-Ψ RNA in the presence of a heterologous HIV-1 MA domain. In the reciprocal experiment, R155H interacting with HIV-1 Ψ showed a specific K₁M(1M) (~ 3.1 × 10⁻⁵ M) and a small Z eff (~ 4), whereas these values are ~ 2.6 M and ~ 9, respectively, for HIV-1 TARpolyA, similar to the parameters obtained for HIV-1 GagΔp6, showing a non-specific binding (Figure 14 and Table 2). Therefore, HIV-1 Gag can also differentiate cognate Ψ from non-Ψ RNA in the presence of the RSV MA domain. Taken
together, MA was shown to be indispensable for discrimination of Ψ but functioned in a non-specific way as it can be substituted by a MA domain from a different retrovirus.

To further confirm that the MA domain does not confer the ability to recognize non-cognate Ψ RNA, salt-titration binding assays were performed to measure WT and chimeric Gag binding to non-cognate RNAs. $K_{d(1M)}$ values of $\sim 3.2 \times 10^{-2}$ M and $\sim 4.0 \times 10^{-2}$ M were measured for RSV GagΔPR binding to HIV-1 Ψ and HIV-1 TARpolyA, respectively, and a $Z_{\text{eff}} \sim 6$ for both, indicating that RSV GagΔPR is unable to specifically recognize HIV-1 Ψ (Figure 14 and Table 2). Strikingly, H132R showed similar binding behavior ($K_{d(1M)} \sim 1.9 \times 10^{-2}$ and $Z_{\text{eff}} \sim 7$ for HIV-1 Ψ, $K_{d(1M)} \sim 2.9 \times 10^{-1}$ and $Z_{\text{eff}} \sim 8$ for HIV-1 TARpolyA) as RSV GagΔPR (Figure 14 and Table 2). In contrast, HIV-1 GagΔp6 showed specific binding to RSV MΨ with a $K_{d(1M)} \sim 5.5 \times 10^{-5}$ M and a $Z_{\text{eff}} \sim 5$, whereas showed non-specific binding to RSV 167 with a $K_{d(1M)} \sim 1.0 \times 10^{-2}$ M and a $Z_{\text{eff}} \sim 8$ (Figure 14 and Table 2). Similarly, R155H interacted with RSV MΨ with a specific $K_{d(1M)}$ ($\sim 6.7 \times 10^{-5}$ M) and a small $Z_{\text{eff}}$ ($\sim 5$), and interacted with RSV 167 with a non-specific $K_{d(1M)}$ ($\sim 2.3 \times 10^{-1}$ M) and a large $Z_{\text{eff}}$ ($\sim 10$) (Figure 14 and Table 2). Thus, R155H is still able to distinguish between RSV MΨ and RSV 167. In summary, recognition by the chimeric constructs follows the same pattern as the corresponding WT Gag constructs, supporting the conclusion that the NC domain determines the RNA binding specificity of Gag, while MA appears to assist in Ψ recognition in a non-specific manner.
2.3.3 Identifying HIV-1 Ψ residues required for Gag binding specificity

In agreement with previous work [19], HIV-1 GagΔp6 exhibits ~15,000-fold greater specificity for WT Ψ RNA ($K_{d(1M)} \sim 3.6 \times 10^{-5}$ M) and less electrostatic interaction ($Z_{eff} \sim 5$) than TARpolyA ($K_{d(1M)} \sim 5.6 \times 10^{-1}$ M and $Z_{eff} \sim 11$) (Figure 16 and Table 3). To identify the guanosine residues critical for this binding specificity, we made several mutations on the HIV-1 Ψ region (Figure 15).

Figure 14: Bar graphs showing $K_{d(1M)}$ (A) and $Z_{eff}$ values (B) determined from salt-titration binding assays with RSV GagΔPR, H132R, R155H and HIV-1 GagΔp6 with RSV MΨ, RSV 167, HIV-1 Ψ and HIV-1 TARpolyA. Values of at three or four trials performed in each case are shown with the height of the bar indicating the mean value.
Table 2: Binding parameter determined by salt-titration binding assays of WT and chimeric Gag constructs.

<table>
<thead>
<tr>
<th>RNA</th>
<th>RSV Gag∆PR</th>
<th>H132R</th>
<th>HIV-1 Gag∆p6</th>
<th>R155H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_{d(1M)}$ (M)</td>
<td>$Z_{eff}$</td>
<td>$K_{d(1M)}$ (M)</td>
<td>$Z_{eff}$</td>
</tr>
<tr>
<td>RSV MΨ</td>
<td>$(7.1 \pm 0.2) \times 10^{-5}$</td>
<td>$3.7 \pm 0.5$</td>
<td>$(1.3 \pm 2) \times 10^{-1}$</td>
<td>$7.0 \pm 1.2$</td>
</tr>
<tr>
<td></td>
<td>$(7.2 \pm 3) \times 10^{-5}$</td>
<td>$4.7 \pm 0.5$</td>
<td>$(1.9 \pm 3) \times 10^{-2}$</td>
<td>$8.9 \pm 2.3$</td>
</tr>
<tr>
<td>HIV-1 Ψ</td>
<td>$(5.5 \pm 2) \times 10^{-5}$</td>
<td>$4.8 \pm 0.2$</td>
<td>$(1.0 \pm 1) \times 10^{-2}$</td>
<td>$8.4 \pm 1.0$</td>
</tr>
<tr>
<td></td>
<td>$(6.7 \pm 3) \times 10^{-5}$</td>
<td>$4.5 \pm 0.4$</td>
<td>$(2.3 \pm 2) \times 10^{-1}$</td>
<td>$9.9 \pm 1.3$</td>
</tr>
<tr>
<td>TARpolyA</td>
<td>$(5.5 \pm 2) \times 10^{-5}$</td>
<td>$4.8 \pm 0.2$</td>
<td>$(3.6 \pm 2) \times 10^{-5}$</td>
<td>$5.4 \pm 0.5$</td>
</tr>
<tr>
<td></td>
<td>$(6.7 \pm 3) \times 10^{-5}$</td>
<td>$4.5 \pm 0.4$</td>
<td>$(3.1 \pm 1) \times 10^{-5}$</td>
<td>$4.1 \pm 0.4$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RNA</th>
<th>RSV Gag∆PR</th>
<th>H132R</th>
<th>HIV-1 Gag∆p6</th>
<th>R155H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_{d(1M)}$ (M)</td>
<td>$Z_{eff}$</td>
<td>$K_{d(1M)}$ (M)</td>
<td>$Z_{eff}$</td>
</tr>
<tr>
<td>RSV MΨ</td>
<td>$(3.2 \pm 5) \times 10^{-2}$</td>
<td>$5.9 \pm 1.1$</td>
<td>$(1.0 \pm 1) \times 10^{-2}$</td>
<td>$7.1 \pm 2.2$</td>
</tr>
<tr>
<td></td>
<td>$(2.9 \pm 2) \times 10^{-1}$</td>
<td>$7.1 \pm 2.2$</td>
<td>$(3.6 \pm 2) \times 10^{-1}$</td>
<td>$5.4 \pm 0.5$</td>
</tr>
<tr>
<td>HIV-1 Ψ</td>
<td>$(3.6 \pm 2) \times 10^{-5}$</td>
<td>$5.4 \pm 0.5$</td>
<td>$(3.1 \pm 1) \times 10^{-5}$</td>
<td>$4.1 \pm 0.4$</td>
</tr>
<tr>
<td></td>
<td>$(3.1 \pm 1) \times 10^{-5}$</td>
<td>$4.1 \pm 0.4$</td>
<td>$(2.3 \pm 2) \times 10^{-1}$</td>
<td>$9.9 \pm 1.3$</td>
</tr>
<tr>
<td>TARpolyA</td>
<td>$(4.0 \pm 2.7) \times 10^{-2}$</td>
<td>$6.2 \pm 0.3$</td>
<td>$(5.6 \pm 6) \times 10^{-1}$</td>
<td>$10.5 \pm 1.0$</td>
</tr>
<tr>
<td></td>
<td>$(2.9 \pm 2) \times 10^{-1}$</td>
<td>$7.9 \pm 1.2$</td>
<td>$(2.6 \pm 3) \times 10^{0}$</td>
<td>$9.2 \pm 0.7$</td>
</tr>
</tbody>
</table>

* RNA concentrations were 1.5 nM for RSV Gag∆PR and H132R and 20.5 nM for HIV-1 Gag∆p6 and R155H.

Figure 15: HIV-1 RNA constructs used in this work. Predicted secondary structures of HIV-1 Ψ (left) and HIV-1 TARpolyA (right) are shown. Eight mutant Ψ RNAs are indicated by the boxed nt and arrows (Mut1-Mut8).
HIV-1 GagΔp6 exhibited only a minor loss in specificity upon mutation of residues in the loops of SL2 and SL3 of HIV-1 Ψ (Mut 3 and 5; $K_{d(1M)} \sim 6.8 \times 10^{-5}$ M and relative specificity of 0.47-0.61 of Ψ-WT), and only a modest increase in the number of electrostatic charges mediating the interaction ($Z_{\text{eff}} \sim 5-6$) (Figure 16 and Table 3). Based on the partial overlap of the parameter values measured for those mutants with those for Ψ-WT, these ~2-fold differences were not significant. A mutation on the single-stranded region between SL2 and SL3 (Mut 4) led to a greater loss in specificity ($K_{d(1M)} \sim 3 \times 10^{-4}$ M; relative specificity of 0.14) (Figure 16 and Table 3). However, because of the variability of the parameter values, we concluded it was also not significantly different from Ψ-WT. In contrast, a significant loss in specificity was observed upon mutation of residues in the bulge regions of SL1 (Mut 1 and 2) and in the region downstream of SL3 (Mut 6) ($K_{d(1M)} \sim 3-9.5 \times 10^{-4}$ M, relative specificity of 0.04-0.11) (Figure 16 and Table 3). Mut 7 and Mut 8 interacted with HIV-1 GagΔp6 with similar $K_{d(1M)}$ values (~ 2-4 × 10^{-5} M, relative specificity of 0.97-1.7) and with similar electrostatic contacts ($Z_{\text{eff}} \sim 4-5$) as Ψ-WT (Figure 16 and Table 3). Importantly, none of the single mutants reduced RNA binding specificity completely to the level of non-Ψ RNA, even though we observed different effects of the HIV-1 Ψ mutants on the binding specificity of Gag. Moreover, we also performed native-PAGE to determine if Ψ mutations disrupted the global RNA fold (data not shown). It turned out that majority of RNAs migrated as a single band corresponding to the RNA dimers, indicating that changes to the global RNA fold and dimerization state cannot explain the altered binding specificity for the mutants.
Figure 16: Bar graphs showing $K_{d(1M)}$ and $Z_{eff}$ values determined from salt-titration binding assays with HIV-1 GagΔp6 and HIV-1 TARpolyA, Ψ RNA and Ψ RNA mutants. Values of three or four trials performed in each case are shown with the height of the bar indicating the mean value.

2.4 Discussion

In this work, we observed an RNA concentration-dependence in the capability of HIV-1 Gag to specifically interact with its cognate Ψ RNA element, with optimal discrimination between Ψ and non-Ψ RNA occurring at 20.5 nM RNA (1:20 RNA:protein ratio) (Figure 13 and Table 1). We also performed native-PAGE analysis to investigate whether this concentration-dependence is due to the change of the
Table 3: Binding parameters determined from salt-titration binding assays of HIV-1 GagΔp6 with HIV-1 TARpolyA, Ψ-WT and Ψ RNA mutants.

<table>
<thead>
<tr>
<th>RNA</th>
<th>$K_{d(1\text{M})}$ (M)</th>
<th>$Z_{\text{eff}}$</th>
<th>Relative Specificity$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TARpolyA</td>
<td>$(5.6 \pm 6) \times 10^{-1}$</td>
<td>$10.5 \pm 1.0$</td>
<td>$6.4 \times 10^{-5}$</td>
</tr>
<tr>
<td>Ψ-WT</td>
<td>$(3.6 \pm 2) \times 10^{-5}$</td>
<td>$5.4 \pm 0.5$</td>
<td>$1$</td>
</tr>
<tr>
<td>Ψ-Mut1</td>
<td>$(3.4 \pm 5) \times 10^{-4}$</td>
<td>$6.7 \pm 1.4$</td>
<td>$0.11$</td>
</tr>
<tr>
<td>Ψ-Mut2</td>
<td>$(9.5 \pm 16) \times 10^{-4}$</td>
<td>$7.1 \pm 1.4$</td>
<td>$0.038$</td>
</tr>
<tr>
<td>Ψ-Mut3</td>
<td>$(5.9 \pm 4) \times 10^{-5}$</td>
<td>$5.5 \pm 0.9$</td>
<td>$0.61$</td>
</tr>
<tr>
<td>Ψ-Mut4</td>
<td>$(2.6 \pm 3) \times 10^{-4}$</td>
<td>$6.7 \pm 2.1$</td>
<td>$0.14$</td>
</tr>
<tr>
<td>Ψ-Mut5</td>
<td>$(7.6 \pm 6) \times 10^{-5}$</td>
<td>$5.9 \pm 0.6$</td>
<td>$0.47$</td>
</tr>
<tr>
<td>Ψ-Mut6</td>
<td>$(4.6 \pm 3) \times 10^{-4}$</td>
<td>$7.8 \pm 0.8$</td>
<td>$0.078$</td>
</tr>
<tr>
<td>Ψ-Mut7</td>
<td>$(3.7 \pm 2) \times 10^{-5}$</td>
<td>$5.2 \pm 0.7$</td>
<td>$0.97$</td>
</tr>
<tr>
<td>Ψ-Mut8</td>
<td>$(2.1 \pm 2) \times 10^{-5}$</td>
<td>$4.4 \pm 1.2$</td>
<td>$1.7$</td>
</tr>
</tbody>
</table>

$^a$ Specificity of the WT Ψ RNAs was set to 1.0 and the relative specificity of the non-Ψ or mutant Ψ RNAs was calculated as $K_{d(1\text{M})}(\Psi-\text{WT})/K_{d(1\text{M})}(\Psi-\text{variant})$.

oligomerization state of the RNA. The results showed that at both the lowest (1.5 nM) and the highest (20.5 nM) RNA concentration tested, HIV-1 Ψ RNA was exclusively dimeric (data not shown). In contrast, RSV GagΔPR shows the greatest MΨ binding specificity at 1.5 nM RNA (1:200 RNA:protein ratio) and remains selective at high RNA concentration (Figure 13 and Table 1). Interestingly, RSV Gag is hypothesized to interact with gRNA in the nucleus initially [148]. Although the concentration-dependence of RNA on the binding specificity of HIV-1 GagΔp6 could be an artifact caused by in vitro assays, it is also possible that the local RNA fold is changed, while the change could not be detected by native-PAGE. Alternatively, the changes in RNA:protein ration might influence Gag’s binding specificity in a biologically relevant manner. Importantly, earlier
studies using salt-titration binding assays showed dramatic effects of NC zinc finger mutations on RNA binding specificity [19], consistent with results based on cell-based assays [119], supporting the robustness of the salt-titration binding assay. We then assume that the highly specific HIV-1 Gag binding at 20.5 nM Ψ RNA concentration (1:20 RNA:protein ratio) may reflect the possibility that HIV-1 Gag initially recognizes gRNA in the cytoplasm where the Gag concentration is low but not on the PM where Gag has a higher local concentration [79, 103, 123, 149]. More work is needed to test this idea.

In this work, we underscored the role of the MA domain in recognizing Ψ RNA in the case of both HIV-1 and RSV. Previously CLIP-seq studies showed that there is a lack of HIV-1 MA-gRNA interaction in vivo [123]. However, MA is detected to interact with tRNA [123] and may serve other functions including PM targeting [121, 122]. Our results provide further biochemical support for the CLIP-seq data showing a lack of MA-Ψ interactions, which would leave MA free to interact with other RNAs and facilitate virus assembly at the PM. Overall, we showed that both RSV and HIV-1 MA played an important role in gRNA recognition by Gag.

We also found that both HIV-1 and RSV Gag were able to accommodate a heterologous MA domain while still retaining the same RNA specificity (Figure 14 and Table 2). Those results are intriguing because HIV-1 MA and RSV MA have extremely different nucleic acid binding affinity [81, 142, 150] and membrane binding properties [120, 150-153]. HIV-1 MA is myristoylated and requires PI(4,5)P₂ for targeting to the PM [150, 151, 154], whereas RSV MA is not myristoylated and still requires PI(4,5)P₂ for PM localization in cells [120, 152]. To our knowledge, cross-packaging between
HIV-1 and RSV Gag proteins and gRNAs has never been investigated in cells. Based on our results, we predict that HIV-1 Gag may be capable of packaging RSV gRNA, whereas the converse would not be true. Taken together, the role of MA domains in genome recognition of retroviruses is significant but the exact mechanism still remains to be determined.

Previous salt-titration binding assays investigated a HIV-1 Ψ variant (Ψ-12M) wherein 12 single-stranded guanosine residues in loops and bulges were simultaneously mutated, showing an ~ 25-fold weaker non-electrostatic binding component relative to WT Ψ RNA [19]. Those guanosine residues were proposed to be high-affinity NC binding sites based on a SHAPE footprinting study [130]. In this work, we investigated individual mutations at six regions (Figure 15). Mutations of two G-rich bulges in SL1 (Mut 1 and 2) and single-stranded G residues downstream of SL3 (Mut 6) showed the greatest loss in specificity (Figure 16 and Table 3), consistent with another study showing that the upper bulge in SL1 is the critical site for HIV-1 Gag interaction [73]. The mutation in the upper single-stranded bulge of SL1 (Mut 2) has a similar effect as previously reported for Ψ-12M [19]. Moreover, consistent with the results, the two bulges (Mut 1 and 2) were shown to be the most significant binding sites for NC in the SHAPE study [130]. According to our results, Ψ Mut 3 is less critical for binding specificity, consistent with the previous studies in which SL2 was believed to be less essential during genome recognition [32, 73, 123]. Ψ Mut 5 showed only a minor effect on the binding specificity of HIV-1 Gag. Surprisingly, numerous studies support the idea that SL3 is important for genome packaging [32, 71]. Importantly, single stranded G residues
mutated in Ψ mut 7 and 8, which are not identified as important residues in genome recognition, showed no effect on the binding specificity in our in vitro binding assays. Taken together, SL1 is the most critical determinant for gRNA selection.

Interestingly, after mapping residues that diminish the binding specificity the most onto the tertiary structural model of HIV-1 Ψ RNA generated by SAXS [55], these residues cluster in the central part of the RNA. Because no sites were shown to completely reduce the specific Ψ RNA recognition by Gag to the level of non-Ψ RNA, we hypothesized that the clustering of several binding sites with high affinity and high specificity in close proximity is a critical prerequisite for genome recognition. Indeed, Gag forms a hexamer due to inter-Gag interactions [87]. Thus, dimerization of gRNA might bring the high-affinity NC binding sites even closer, facilitating the interactions with oligomers of Gag, ensuring the selective packaging of dimeric gRNA.

In summary, in this work, salt-titration binding assays were used to test the binding of RSV and HIV-1 Gag constructs. We found an intriguing concentration-dependence binding specificity of Gag from different retroviruses. In addition, MA was found to be indispensable for specific Ψ recognition by Gag and surprisingly, MA domains can be swapped between different retroviruses without loss in cognate Ψ binding specificity in vitro. Finally, the contribution of clusters of individual G bases in HIV-1 Ψ for Gag binding specificity was investigated and based on these results, a mechanism for genome packaging of dimeric HIV-1 gRNA is proposed.
Acknowledgement

In this work, data collection and analysis are performed together with Dr. Tiffny Rye-McCurdy, Erik Olson, Christiana Binkley and Joshua-Paolo Reyes.
Chapter 3: Contribution of CA and SP1 to Gag binding specificity for Psi

3.1. Background

Recently, a Förster resonance energy transfer (FRET) study suggested that HIV-1 Gag has two conformations: a compact conformation and an extended conformation [92]. Only Gag in its extended conformation can form VLPs which have a similar size as authentic viral particles [92]. In addition, as described in chapter 2, Gag has two distinct binding modes, corresponding to Ψ RNA binding and non-Ψ RNA binding [19]. Moreover, the FRET study also showed that the compact conformation of Gag is highly stable in solution [92]. Those studies lead to a hypothesis: The two conformations of Gag correspond to the two binding modes. For Ψ RNA binding, Gag adopts the extended conformation and this is favorable for virus assembly. In contrast, for non-Ψ RNA binding, Gag adopts the compact conformation, which is stabilized by an intra-Gag interaction. Testing of this hypothesis will be described in this chapter.

Multimerization of Gag is critical for virus assembly [12]. W316, M317 in the CA domain are residues that are important for the dimerization of Gag. Replacement of these two residues by alanines leads to a defect of inter-Gag interaction [29]. The WM mutation causes an increase of the dissociation constant of dimerization from 5.5 μM to 0.53 mM [155]. Moreover, MA was also shown to be involved in multimerization. A crystallography study suggested a trimeric arrangement of MA (Figure 17) [156]. Another low-resolution EM study showed MA forms a hexamer of trimers on the plasma
membrane [24], compatible with a hexamer-of-trimers arrangement of CA-NTD [31].

MA trimerization plays a role in Env incorporation and influences infectivity of HIV-1 [26].

![Figure 17](image)

**Figure 17:** The crystal structure showing trimeric HIV-1 MA. Taken from [156].

SP1, which is located between CA and NC, has been shown to be important for HIV-1 assembly. Several reports showed that mutations on SP1 significantly disrupt the assembly of HIV-1 [41, 157, 158]. A very important characteristic of SP1 is that it can form an \( \alpha \)-helix together with residues at the C-terminal end of CA. At low concentration in solution, the CA-SP1 junction is largely unstructured [159]. However, when the dielectric constant of the medium is reduced by adding 30% TFE, CA-SP1 adopts an \( \alpha \)-helical structure [160]. Similarly, a CD study also showed that the CA-SP1 junction forms an \( \alpha \)-helical structure when the concentration is raised to \( \sim 5 \mu\text{M} \) [41]. Interestingly, this \( \alpha \)-helix has a hydrophobic face and a hydrophilic face, indicating that the hydrophobic faces from different SP1 molecules are able to be buried together and
trigger the oligomerization of SP1 (Figure 18) [41]. Recently, a study on a chimeric protein SP1-Zip (residues at the C-terminal end of CA and residues at SP1 are fused to a leucine zipper) showed that the SP1 region adopts a helical conformation even at low concentration [161]. Thus, the leucine zipper, which promotes dimerization, has a similar effect on increasing the local concentration of SP1 as Gag-RNA binding [161]. Overall, these studies on SP1 suggested a mechanism by which SP1 could trigger virus assembly (Figure 19) [90]. Briefly, binding of Gag to RNA increases local Gag concentration. The increase of local Gag concentration triggers a conformational switch of SP1, forming an $\alpha$-helical structure. This $\alpha$-helical SP1 facilitates the formation of new oligomerization interfaces in CA and initiates virus assembly [90].

Figure 18: Helical wheel showing that $\alpha$-helical SP1 is amphipathic. Taken from [41].
Figure 19: Model describing SP1-triggered HIV-1 virus assembly. Taken from [90].

In this work, by using salt-titration binding assays, the Ψ RNA binding specificity of several Gag mutants was tested. The roles that the WM dimerization site and the MA trimerization site play in RNA binding specificity of Gag were investigated. Mutation of the WM dimerization site changed Gag binding specificity significantly, whereas mutation of the MA trimerization sites showed no effect. In addition, mutants related to SP1 were tested, showing distinct effects on the RNA binding specificity of Gag. Overall, this work indicates that the WM dimerization site and SP1 contribute to the binding specificity of Gag.
3.2. Materials and methods

3.2.1. Protein preparation

Protein constructs that were used in this work are shown in Figure 20. In addition, three point mutations (L75E, L75G, A45E) in the context of GagΔp6 were also made in the MA trimerization site.

![Protein constructs](image)

**Figure 20**: Protein constructs used in this work. Full-length HIV-1 Gag and HIV-1 GagΔp6 are shown at the top for comparison. WM-GagΔp6 contains two mutations W316A, M317A in the dimerization site of CA. MA-linker-SP1-NC represents a construct wherein CA is removed and replaced by a 16-amino acid flexible linker (SGGGGSGG)$_2$. MA-linker-NC is a construct wherein both CA and SP1 is deleted and replaced by the same flexible linker. GagΔp6ΔSP1 is a SP1-deleted mutant in the context of GagΔp6. In linker-GagΔp6ΔSP1, SP1 is replaced by the flexible linker (SGGGGSGG)$_2$.

All the mutations were made by site-directed ligase-independent mutagenesis (SLIM) [145]. The methods for protein preparation are the same as described in Chapter 2. Concentrations of all the proteins were determined by measuring absorbance at 280 nm and using the following molar extinction coefficients: 63,223 M$^{-1}$·cm$^{-1}$ (GagΔp6,
GagΔSP1, linker-GagΔSP1, L75E GagΔp6, L75G GagΔp6 and A45E GagΔp6), 57,723 M⁻¹·cm⁻¹ (WM-GagΔp6), 29,700 M⁻¹·cm⁻¹ (MA-linker-SP1-NC, MA-linker-NC).

3.2.2. Preparation of RNA constructs

All RNA constructs were in vitro transcribed from linearized plasmids using T7 RNA polymerase and labeled with fluorescein-5-thiosemicarbazide (FTSC) at the 3’ end, as described in Chapter 2. The concentration and labeling efficiency were determined by measuring absorbance at 260 nm and 495 nm and using the following molar extinction coefficients: \( \varepsilon_{260\text{nm}} = 9.7 \times 10^5 \text{ M}^{-1} \cdot \text{cm}^{-1} \) (HIV-1 Ψ), \( \varepsilon_{260\text{nm}} = 9.3 \times 10^5 \text{ M}^{-1} \cdot \text{cm}^{-1} \) (HIV-1 TARpolyA), \( \varepsilon_{260\text{nm}} = 1.1 \times 10^6 \text{ M}^{-1} \cdot \text{cm}^{-1} \) (SIV TAR), \( \varepsilon_{260\text{nm}} = 1.3 \times 10^6 \text{ M}^{-1} \cdot \text{cm}^{-1} \) (SIV Ψ).

3.2.3. FA-based direct binding assay and salt-titration binding assay

Direct and salt-titration binding assays, as well as the data analysis, were performed as described in Chapter 2. Salt-titration binding assays were performed with 20.5 nM RNA concentration and 400 nM protein concentration.

3.3. Results

3.3.1. Mutation of the WM dimerization site converts the binding mode of Gag to NC-only

The effect of WM dimerization site mutation was investigated. WM-GagΔp6 interacted with HIV-1 Ψ with a specific \( K_d(1\text{M}) \) (\( \sim 1.6 \times 10^{-4} \text{ M} \)), which was similar to
GagΔp6 (~ 3.6 × 10⁻⁵ M), whereas Z_{eff} was slightly smaller (~ 4) compared to GagΔp6 (~ 5) (Figure 21 and Table 4). However, WM-GagΔp6 interacted with HIV-1 TARpolyA, which is the non-Ψ RNA, with a more specific \( K_{d(1M)} \) (~ 3.8 × 10⁻³ M) and a much smaller \( Z_{eff} \) (~ 6), relative to the binding of GagΔp6 (\( K_{d(1M)} \) ~ 5.6 × 10⁻¹ M and \( Z_{eff} \) ~ 10.5) (Figure 21 and Table 4). The small \( Z_{eff} \) measured for WM-GagΔp6 binding to HIV-1 TARpolyA is consistent with an NC-only binding mode [19]. By comparison, binding of CANC construct (or ∆MA) to HIV-1 Ψ and HIV-1 TARpolyA showed values similar to those measured for WM-GagΔp6 (\( K_{d(1M)} \) ~ 2.5 × 10⁻⁴ M and \( Z_{eff} \) ~ 4 for HIV-1 Ψ, \( K_{d(1M)} \) ~ 3.9 × 10⁻³ M and \( Z_{eff} \) ~ 5 for HIV-1 TARpolyA) [19], supporting an NC-only binding mode for both variants.

Table 4: Binding parameters determined from salt-titration binding assays using GagΔp6 and WM-GagΔp6.

<table>
<thead>
<tr>
<th>RNA</th>
<th>GagΔp6a</th>
<th>WM-GagΔp6</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV-1 Ψ</td>
<td>( K_{d(1M)} ) (M) (3.6 ± 1.6) × 10⁻⁵</td>
<td>(1.6 ± 2.4) × 10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>( Z_{eff} ) 5.4 ± 0.5</td>
<td>3.9 ± 1.5</td>
</tr>
<tr>
<td>HIV-1 TARpolyA</td>
<td>( K_{d(1M)} ) (M) (5.6 ± 6.4) × 10⁻¹</td>
<td>(3.8 ± 1.7) × 10⁻³</td>
</tr>
<tr>
<td></td>
<td>( Z_{eff} ) 10.5 ± 1.0</td>
<td>5.6 ± 0.3</td>
</tr>
<tr>
<td>SIV Ψ</td>
<td>( K_{d(1M)} ) (M) (4.7 ±4.0) × 10⁻²</td>
<td>(1.1 ± 0.7) × 10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>( Z_{eff} ) 8.8 ± 1.0</td>
<td>4.3 ± 0.9</td>
</tr>
<tr>
<td>SIV TAR</td>
<td>( K_{d(1M)} ) (M) (1.2 ± 1.9) × 10⁻¹</td>
<td>(1.9 ± 1.0) × 10⁻⁴</td>
</tr>
<tr>
<td></td>
<td>( Z_{eff} ) 8.0 ± 1.4</td>
<td>4.2 ± 0.6</td>
</tr>
</tbody>
</table>

*Parameters of GagΔp6 binding with SIV Ψ and SIV TAR are from [162].
Figure 21: Bar graphs showing $K_{d(1M)}$ (A) and $Z_{eff}$ (B) determined from salt-titration binding assays using GagΔp6 and WM-GagΔp6 binding to HIV-1 Ψ, HIV-1 TARpolyA, SIV Ψ and SIV TAR. Values of three trials performed in each case are shown with the height of the bar indicating the mean value. Parameters of GagΔp6 binding to SIV Ψ and SIV TAR are from [162].

To further test whether WM-GagΔp6 results in an NC-only binding mode for other non-specific RNAs, simian immunodeficiency virus (SIV) Ψ and SIV TAR were used. A recent study using salt-titration binding assays showed that HIV-1 GagΔp6 interacted with both SIV Ψ and SIV TAR in a non-specific manner ($K_{d(1M)} \approx 4.7 \times 10^{-2} \text{ M}$).
and \(Z_{\text{eff}} \sim 9\) for SIV Ψ, \(K_{d(1M)} \sim 1.2 \times 10^{-1}\) M and \(Z_{\text{eff}} \sim 8\) for SIV TAR) [162]. WM-GagΔp6, however, interacted with both SIV Ψ and SIV TAR with a less specific \(K_{d(1M)}\) (~ \(1.1 \times 10^{-4}\) M for SIV Ψ and ~ \(1.1 \times 10^{-4}\) M for SIV TAR) and a smaller \(Z_{\text{eff}}\) (~ 4 for both RNAs) (Figure 21 and Table 4). This supports the conclusion that WM mutation converts the binding of HIV-1 GagΔp6 to an NC-only binding mode.

3.3.2. Mutation of MA trimerization site does not affect the binding of GagΔp6

To investigate whether mutation of another Gag multimerization site has a similar effect on the binding of Gag as mutation on the WM site, several point mutations (L75E, L75G and A45E) in the context of GagΔp6 were made in the MA trimerization site. These mutations were previously shown to disrupt the trimerization of MA significantly as determined by a chemical cross-linking-based experiment [26]. Salt-titration binding assays showed that all the mutants interacted with HIV-1 Ψ and HIV-1 TARpolyA in a similar manner as GagΔp6 (\(K_{d(1M)} \sim 10^{-5}\) M and \(Z_{\text{eff}} \sim 6\)) (Figure 22 and Table 5), indicating that the specificity for Ψ RNA binding is not affected by mutation of the MA trimerization interface.

<table>
<thead>
<tr>
<th>RNA</th>
<th>GagΔp6</th>
<th>L75E</th>
<th>L75G</th>
<th>A45E</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV-1 Ψ</td>
<td>(K_{d(1M)}) (M)</td>
<td>(3.6 ± 1.6) × 10^{-5}</td>
<td>(6.0 ± 1.1) × 10^{-5}</td>
<td>(1.7 ± 1.6) × 10^{-4}</td>
</tr>
<tr>
<td></td>
<td>(Z_{\text{eff}})</td>
<td>5.4 ± 0.5</td>
<td>5.4 ± 0.2</td>
<td>6.4 ± 1.6</td>
</tr>
<tr>
<td>HIV-1 TARpolyA</td>
<td>(K_{d(1M)}) (M)</td>
<td>(5.6 ± 6.4) × 10^{-1}</td>
<td>(2.2 ± 3.6) × 10^{-1}</td>
<td>(7.0 ± 11.7) × 10^{-1}</td>
</tr>
<tr>
<td></td>
<td>(Z_{\text{eff}})</td>
<td>10.5 ± 1.0</td>
<td>10.2 ± 1.6</td>
<td>9.6 ± 1.8</td>
</tr>
</tbody>
</table>

Table 5: Binding parameters determined from salt-titration binding assays using WT, L75E, L75G and A45E GagΔp6.
Figure 22: Bar graphs showing $K_d(1M)$ (A) and $Z_{eff}$ (B) determined from salt-titration binding assays using WT, L75E, L75G and A45E GagΔp6 binding to HIV-1 Ψ and HIV-1 TARpolyA. Values of three trials performed in each case are shown with the height of the bar indicating the mean value.

3.3.3. *SP1 alters the RNA binding specificity of Gag*

Considering that the WM mutation located in CA had a significant effect on the RNA binding specificity of Gag, we next wanted to determine the effect of CA domain deletion. A MA-linker-SP1-NC construct wherein CA was replaced by a flexible linker
(SGGGGSGGG)_2 was prepared and tested. This mutant maintains structural flexibility, but avoids the direct connection of the MA and NC domains. Salt-titration binding assay showed that this mutant interacted with HIV-1 Ψ in a less specific manner (K_d(1M) ~ 2.0 × 10^{-3} M) and with a slightly greater Z_{eff} (~ 6) compared to GagΔp6. However, this mutant interacted with HIV-1 TARpolyA in a similar manner (K_d(1M) ~ 6.3 × 10^{-1} M) as GagΔp6, although with a slightly smaller Z_{eff} (~ 9) (Figure 23 and Table 6). Thus, MA-linker-SP1-NC has a diminished binding specificity for Ψ but still discriminates between Ψ and non-Ψ RNA. Since SP1 remains intact in this construct, we next wanted to determine whether SP1 contributes to the binding specificity. To test this, another construct MA-linker-NC, which lacks both CA and SP1, was tested. MA-linker-NC interacted with HIV-1 Ψ more specifically (K_d(1M) ~ 1.1 × 10^{-4} M) and with fewer electrostatic contacts (Z_{eff} ~ 6) than the linker construct with SP1. For interaction with HIV-1 TARpolyA, a relatively specific K_d(1M) (~ 5.9 × 10^{-3} M) and a small Z_{eff} (~ 6) were observed (Figure 23 and Table 6). Interestingly, MA-linker-NC binding to HIV-1 Ψ and HIV-1 TARpolyA was nearly the same as the binding of WM-GagΔp6 and CANC, indicating an NC-only binding mode. Therefore, SP1 contributes to the binding specificity in the context of a ΔCA variant of GagΔp6.

To test the role of SP1 in the context of full-length GagΔp6, another construct GagΔp6ΔSP1 was tested. Unexpectedly, GagΔp6ΔSP1 maintains a similar RNA binding specificity as GagΔp6, with a specific K_d(1M) (~ 3.9 × 10^{-5} M) and a small Z_{eff} (~ 6) for HIV-1 Ψ binding and a non-specific K_d(1M) (~ 5.5 × 10^{-1} M) and a large Z_{eff} (~ 11) for HIV-1 TARpolyA binding (Figure 23 and Table 6). Therefore, in the presence of CA, the
removal of SP1 does not impact the RNA binding specificity of Gag. Considering SP1 is proposed to serve as a transmitter, triggering oligomerization of CA when NC binds to an RNA [90], we next determined whether replacement of SP1 with a flexible linker disrupts the Ψ binding specificity of Gag. In the construct linker-GagΔp6ΔSP1, SP1 is replaced by the same flexible linker (SGGGGSGG)$_2$. Its interaction with HIV-1 Ψ was highly specific ($K_{d(1M)} \approx 2.4 \times 10^{-5}$ M), similar to GagΔp6. Moreover, this interaction had similar electrostatic contacts ($Z_{eff} \approx 5$). However, this construct interacted with HIV-1 TARpolyA with a more specific $K_{d(1M)} \approx 7.2 \times 10^{-3}$ M and a smaller $Z_{eff} \approx 8$ compared to GagΔp6 (Figure 23 and Table 6). Therefore, a flexible linker appears to disrupt the communication between CA and NC so that the binding mode is closer to an NC-only one. Taken together, the data suggest that SP1 contributes to RNA binding specificity of Gag.

**Table 6:** Binding parameters determined from salt-titration binding assays using WT GagΔp6 and CA and SP1-deletion constructs.

<table>
<thead>
<tr>
<th>Protein</th>
<th>HIV-1 Ψ</th>
<th>HIV-1 TARpolyA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_{d(1M)}$ (M)</td>
<td>$Z_{eff}$</td>
</tr>
<tr>
<td>GagΔp6</td>
<td>$3.6 \pm 1.6 \times 10^{-5}$</td>
<td>$5.4 \pm 0.5$</td>
</tr>
<tr>
<td>MA-linker-SP1-NC</td>
<td>$2.0 \pm 2.4 \times 10^{-3}$</td>
<td>$6.2 \pm 0.8$</td>
</tr>
<tr>
<td>MA-linker-NC</td>
<td>$1.1 \pm 0.4 \times 10^{-4}$</td>
<td>$4.9 \pm 0.3$</td>
</tr>
<tr>
<td>GagΔp6ΔSP1</td>
<td>$3.9 \pm 1.1 \times 10^{-5}$</td>
<td>$6.0 \pm 0.2$</td>
</tr>
<tr>
<td>linker-GagΔp6ΔSP1</td>
<td>$2.4 \pm 1.4 \times 10^{-5}$</td>
<td>$5.1 \pm 0.7$</td>
</tr>
</tbody>
</table>
Figure 23: Bar graphs showing $K_d(1\text{M})$ (A) and $Z_{\text{eff}}$ (B) determined from salt-titration binding assays using WT GagΔp6 and CA and SP1-deletion constructs binding to HIV-1 Ψ and HIV-1 TARpolyA. Values of three or four trials performed in each case are shown with the height of the bar indicating the mean value.
3.4. Discussion

3.4.1. WM mutation alters the RNA binding specificity of Gag

Studying the 5'-UTR of HIV-1 gRNA in cell-based assays is challenging because of multiple effects on the life cycle [104]. Therefore, using in vitro assays to investigate the role of specific Gag-RNA interactions is necessary.

It has been proposed that Gag adopts an NC-only binding mode upon binding to Ψ RNA and a MA & NC binding mode when bound to non-Ψ RNA [19]. Hypothetically, those two binding modes correspond to two different conformations—an extended conformation and a compact conformation, respectively. Only Ψ RNA is able to trigger the extended conformation of Gag so that in this case MA is able to interact with the PM and initiate the virus assembly [19]. This is especially important because Gag is dominantly in a compact conformation in solution at low concentration [92].

An intra-Gag interaction may play a role in stabilizing the compact conformation of Gag. In this work, the effect of WM mutation in the dimerization site of CA on specific RNA binding was demonstrated. WM-Gag∆p6 bound to all RNAs tested with a rather specific $K_d(1M)$ and a small $Z_{eff}$, suggesting an NC-only binding mode. Since WM-Gag∆p6 doesn’t interact with non-Ψ RNA in a MA & NC binding mode, it appears that the compact conformation is not stable in WM-Gag∆p6. Therefore, we hypothesize that the WM site is involved in an intra-Gag interaction to stabilize the compact conformation of Gag. However, the possibility that the altered specificity for RNA binding is due to the weakening of the oligomeric state of Gag cannot be excluded. Considering the protein concentration (400 nM) used in salt-titration binding assays and the dissociation constant
for dimerization of Gag (~ 5.5 μM) [155], the predominant form of GagΔp6 in the assays is monomer. However, it is still possible that the oligomeric state of Gag is affected upon RNA binding. Moreover, salt-titration binding assays are not sensitive to the changes in protein concentration [19]. CANC, which is capable of dimerizing, also shows similar $K_{d(1M)}$ and $Z_{\text{eff}}$ values for $\Psi$ RNA and non-$\Psi$ RNA [19]. All of the evidence suggests that WM mutation influences the conformation adopted by Gag rather than simply influencing the oligomeric state of Gag. Unfortunately, it is very difficult to directly study the effect of WM mutation on gRNA incorporation in cells because WM mutation largely inhibits the formation of viral particles [91]. Overall, WM mutation alters the RNA binding specificity of Gag and could possibly change the conformation adopted by Gag.

3.4.2. Mutation of MA trimerization site shows no effect on binding specificity of Gag

HIV-1 MA has diverse functions throughout the HIV-1 replication cycle and studying MA is of great interest [11]. Crystallographic structures reveal a MA trimer [156], whereas NMR structures only reveal a monomer [163]. However, so far the structure of MA is still not available in the context of authentic virions [31]. Recently, MA trimerization has been suggested to correlate with Env incorporation [26, 164]. Therefore, it is of interest to establish whether MA trimers have additional functions. In this work, the RNA binding specificity was determined for three trimerization mutants of MA in the context of GagΔp6. No altered RNA binding specificity was observed for these point mutants, although the possibility that other residues in the MA trimerization
site may have an effect cannot be excluded. The MA trimer is not a highly stable structure and it is proposed to be maintained by relatively weak hydrophobic interactions [26]. Because of the hydrophobic character, the MA trimerization site may be involved in the proposed intra-Gag interaction, as discussed in the previous section. However, MA trimerization does not impact formation of VLPs [26], and based on the results of salt-titration binding assays, the MA trimerization site is not important for Gag-RNA interaction.

3.4.3. *SP1 contributes to the RNA binding specificity of Gag*

Even subtle changes in the SP1 region significantly impact virus assembly [41, 157, 158, 161, 165-168]. In this work, using salt-titration binding assays, the role of SP1 in determining the RNA binding specificity of Gag was revealed. MA-linker-SP1-NC showed diminished RNA binding specificity, whereas MA-linker-NC almost completely lost RNA binding specificity for Ψ (Figure 23 and Table 6). This indicates that SP1 contributes to the RNA binding specificity of Gag for Ψ in the absence of CA. Interestingly, in the presence of CA, GagΔp6ΔSP1 showed the same RNA binding specificity as GagΔp6 (Figure 23 and Table 6), but RNA binding specificity can be significantly impacted by substituting SP1 with a flexible linker, indicating that the communication between NC and CA is blocked in this case (Figure 23 and Table 6).

Considering the role of SP1 as a transmitter, a flexible linker near that region may provide too much flexibility for Gag and influence the ability of Gag to distinguish Ψ RNA versus non-Ψ RNA. This is consistent with the diminished binding specificity for Ψ
in the context of MA-linker-NC and linker-GagΔp6ΔSP1. In these cases, NC alone determines the specificity, and an NC-only binding mode is observed. It should be emphasized that SP1 forms a helical structure together with residues from the C-terminal end of CA [169]. Therefore, the remaining residues of CA could be the explanation for the reason why GagΔp6ΔSP1 still distinguishes Ψ RNA versus non-Ψ RNA. It would be interesting to see whether additional deletion of residues at the C-terminal end of CA together with ΔSP1 impacts the binding more drastically. For MA-linker-SP1-NC, the binding is largely different from any other mutants, showing a less specific $K_d(1M)$ for Ψ RNA binding, relative to an NC-only binding. In this construct, it is conceivable that SP1-NC determines the specificity. In fact, SP1-NC shows a different RNA binding specificity from NC (data not shown). Thus, the unique binding patterns of MA-linker-SP1-NC may be explained by the difference in binding specificity of NC and SP1-NC. Overall, the data suggest that the SP1 region contributes to the RNA binding specificity of Gag.

3.5. A model describing Ψ RNA binding specificity of Gag

Based on the results described above and several previous models [19, 161], a model for the Ψ RNA binding specificity of Gag is proposed (Figure 24). In the case of Ψ RNA, Gag adopts an NC-only binding mode. Multiple high-affinity and high-specificity binding sites for Gag that are proximal to each other (described in Chapter 2) increase the local concentration of Gag, triggering conformational transition of SP1 to an α-helix. The α-helical SP1 then triggers the oligomerization of Gag and initiates virus assembly. In the
case of non-Ψ RNA, Gag adopts a MA & NC binding mode. Correspondingly, the local concentration of extended Gag is too low to trigger the conformation transition of SP1. Therefore, slow, non-optimal virus assembly occurs.

**Figure 24**: Model for describing Ψ RNA binding specificity of Gag. The asterisks on Ψ RNA represent high-affinity NC binding sites.

### 3.6. Conclusion and future directions

This study suggests that a mutation of two residues in the dimerization site of CA can significantly change the binding behavior of Gag. Conversion of the binding mode of Gag could be explained by a conformational change of Gag triggered by specific NC-RNA binding. Moreover, this study also sheds light on the critical role of SP1 in gRNA
recognition. In summary, several key elements that are important for gRNA recognition and virus assembly were demonstrated.

Future work should focus on obtaining direct evidence for the intra-Gag interaction. Using salt-titration binding assays alone is not sufficient to support the hypothesis that the WM dimerization site is altering the conformation of Gag and is involved in the intra-Gag interaction. FRET is a technique that can probe the proposed mechanism. In addition, more work needs to be done to determine the role of SP1. For instance, it is necessary to additionally delete the residues at the C-terminal end of CA. Moreover, in order to avoid influencing the global structure of Gag, point mutations should also be considered. It is also necessary to figure out the effect of SP1 on the RNA binding specificity of NC. In conclusion, more work still needs to be carried out before a complete understanding of the mechanism of gRNA recognition by Gag is obtained.

Acknowledgement
Salt-titration binding assays for MA-trimerization mutants were performed together with Songyu Dong (Department of Chemistry and Biochemistry, The Ohio State University)
References


70


