INFORMATICALLY-INFORMED BIOCHEMICAL CHARACTERIZATION OF APTAMER-PROTEIN COMPLEXES

A Dissertation

presented to

the faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

PAIGE ROSE GRUENKE

Dr. Donald H. Burke-Agüero, Dissertation Supervisor

Dr. Margaret J. Lange-Osborn, Dissertation Co-Supervisor

MAY 2022

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

INFORMATICALLY-INFORMED BIOCHEMICAL CHARACTERIZATION OF APTAMER-PROTEIN COMPLEXES

Presented by Paige Rose Gruenke,

A candidate for the degree of

Doctor of Philosophy

And hereby certify that, in their opinion, it is worthy of acceptance.

_____

Dr. Donald H. Burke-Agüero

_____

Dr. Margaret J. Lange-Osborn

_____

Dr. Xiao Heng

_____

Dr. Thomas P. Quinn

_____

Dr. Marc C. Johnson

Dedication

In loving memory of my Dad

To my Mom, with love and endless appreciation

Acknowledgements

I would first like to thank Dr. Donald Burke-Agüero and Dr. Maggie Lange-Osborn for their guidance and mentorship over the years. They both have continuously challenged me to become a better scientist. I appreciate the supportive environment that you both have cultivated in your labs, and the training I have received will be invaluable in my future endeavors. I also want to thank the members of my doctoral committee, Dr. Xiao Heng, Dr. Thomas Quinn, and Dr. Marc Johnson. Thank you for the helpful feedback and insights you have provided to improve my research.

I want to thank the past and present members of the Burke lab through the years for their support, feedback, and good times. I am honored to have been able to work with some awesome people. I especially want to thank Dr. David Porciani for helping me experimentally and for his advice over the years, Dr. Khalid Alam for helping me get started in the lab and being a great collaborator since he left the lab, and Dr. Kwaku Tawiah for the infectious joy he brought to the lab. I also must give a special shout out to my fellow Burke lab graduate student and close friend Jordyn Lucas. Thank you for being there for me through the ups and downs of graduate school.

I also want to thank the past and present members of the Lange lab. It has been an honor being a part of the lab since its beginning, and I am truly proud of the community we have built. I especially want to thank Dr. Rachna Aneja and Dr. Sarah Welbourn, two fantastic colleagues that I have had the pleasure of working with and who taught me a lot on how to approach science.

I am thankful for the University of Missouri Molecular Life Sciences Fellowship and Dr. Mark Hannink and Debbie Allen for their support. I also am thankful for the Gehrke Graduate Fellowship in Biochemistry and the Wayne L. Ryan Graduate Fellowship (Ryan Foundation).

I am most grateful for my family and friends. Thank you all for your love and endless encouragement. To my sister Madison, thank you for always being there for me. I also must thank my dog Rosie for her unconditional love and solidarity.

I would not be here without my parents, Glenn and Tammy Gruenke. They have always been supportive of my academic journey. Dad, while you are no longer here with us, I still feel your love and support. Mom, thank you for being the rock in our family and my biggest cheerleader. Thanks for listening to me when I needed to vent and encouraging me to keep going.

Finally, but most certainly not least, I have to thank God for in Him all things are possible.

Table of Contents

List of Figures and Tables

Abstract

Nucleic acid aptamers are single-stranded oligonucleotides that fold into unique three-dimensional shapes that typically bind to targets of interest with high specificity and high affinity. However, a significant limiting step after performing an aptamer selection is the characterization of the resulting aptamers to identify those with the desired binding characteristics. By combining aptamer selections with high-throughput sequencing and bioinformatic analyses, we prioritized aptamers based on their enrichment profiles over the course of the selection for further characterization. The work described here uses biochemical techniques and bioinformatics to study aptamer-protein target interactions. We first describe the initial RNA aptamer selection against the assembled HIV-1 capsid lattice and the characterization of aptamer CA15-2. A capsid differentiation selection was then performed starting with the lattice round 15 library to identify subsets of aptamers capable of binding the assembled capsid lattice only or the capsid lattice and hexamer, and we utilized comparative sequence analyses and biochemical approaches to begin elucidating the sequence and structural requirements for aptamer binding to the capsid lattice. Finally, this work describes an aptamer reselection strategy to evaluate the effects of different 2′ modifications on a pre-enriched aptamer library using an aptamer library with affinity for HIV-1 reverse transcriptase (RT). While we identified RT aptamers capable of tolerating the three 2′ modifications used, we also observed that the presence of 2′-fluoro-modified pyrimidines caused all RNA sequences, including non-binding sequences, to bind and inhibit RT. The results described in this dissertation provide new strategies and insights for aptamer selections, bioinformatic analyses of aptamer populations from high-throughput sequencing, and biochemical characterization of aptamer-target interactions.

CHAPTER 1: INTRODUCTION

At its most basic level, life is a series of chemical reactions coordinated by molecular interactions among metabolites, proteins, nucleic acids, carbohydrates, and lipids. Investigating these molecular interactions can uncover how biological processes work, and altering molecular interactions could potentially have applications for human health, agriculture, and beyond. Nucleic acid aptamers are useful tools to study these molecular interactions due to their ability to bind specifically to their selected target. The research studies detailed within this dissertation focus on characterizing ribonucleic acid (RNA) aptamer-target interactions using biochemical methods and bioinformatic analyses. This chapter provides the background and the current questions and bottlenecks in the field that this research addressed. It is presented in three parts. First, an overview about the capsid protein in human immunodeficiency virus type 1 (HIV-1) will be provided, as it was the selection target for Chapters 2 and 3. Capsid's structure, its functions within the HIV-1 replication cycle, the ongoing questions about capsid biology, and the current tools that are used to study capsid will be summarized. The second part is an overview of nucleic acid aptamers, how they are generated, and their biological applications. Last, the remaining chapters in this dissertation will be introduced.

1.1 The HIV-1 Capsid

1.1.1. Human Immunodeficiency Virus

HIV-1 is the virus that is responsible for acquired immunodeficiency syndrome (AIDS). Belonging to the family *Retroviridae*, HIV-1 is a lentivirus that infects cells of the immune system, including $CD4^+$ T cells, macrophages, and dendritic cells. According

to the World Health Organization, there were approximately 37.7 million people living with HIV as of December 2020(1). Since the beginning of the HIV/AIDS global epidemic in the early 1980s, an estimated 36.3 million lives have been lost, with approximately 680,000 lives lost in 2020. Of those living with HIV, approximately 73% were receiving antiretroviral therapy, typically consisting of a combination of three antiretroviral drugs from two HIV drug classes. Antiretroviral therapy helps to suppress viral replication in these patients. However, some challenges with antiretroviral therapy are that it must be taken daily and that drug resistance can occur due to mutations that can occur in the HIV-1 genome. Therefore, while the ultimate goal in HIV research is to find a cure and to develop either an HIV-1 vaccine or a gene therapy approach to eradicate the virus, more research is needed to understand HIV biology and to develop new classes of antiretroviral drugs. In particular, there are several important questions regarding the roles of capsid during replication events that will be discussed below. Answering these questions could provide insights for capsid-targeting drugs.

1.1.2. Overview of the HIV-1 Replication Cycle

HIV-1 is an enveloped, single-stranded, positive-sense RNA virus. Schematics of a mature virion and the HIV-1 replication cycle are shown in Figure 1.1. To begin replication, a mature HIV-1 virion binds to the CD4 receptor and a co-receptor (CCR5 or CXCR4)(2,3). This initiates fusion of the viral membrane with the host plasma membrane and releases the capsid core, containing the HIV genome and viral enzymes, into the cytoplasm. From the release of the capsid core into the cytoplasm to integration of the provirus into the host's genome, a number of critical steps within the replication cycle must occur (reviewed in(4-6)). These include trafficking of the capsid core to the nucleus,

capsid uncoating, reverse transcription of the HIV RNA genome into DNA to form the pre-integration complex (PIC), and nuclear import. While there is much known about these steps, the exact spatiotemporal requirements for these steps are still being explored.

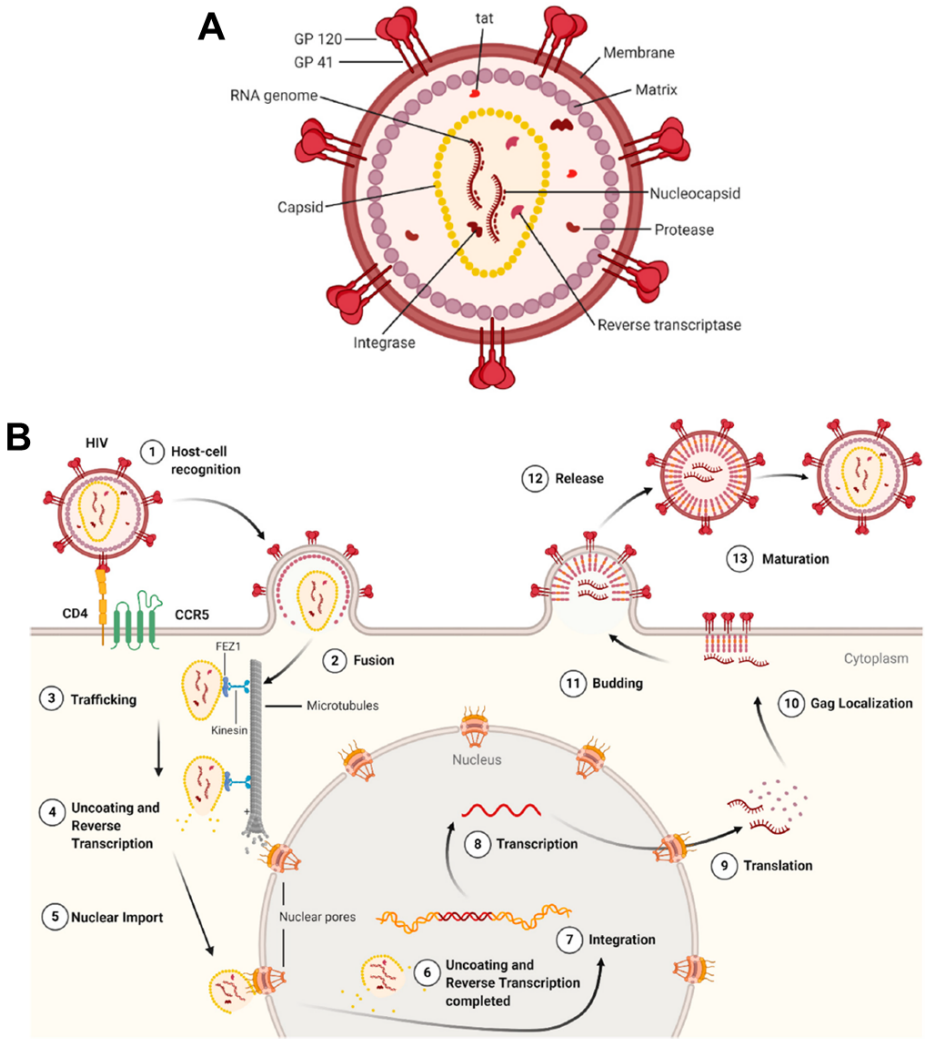

Figure 1.1: HIV-1 Replication Cycle. (A) Schematic of a mature HIV-1 virion. (B) Overview of the steps of HIV-1 replication cycle. Figure panels from (7). Licensed for use under the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

After integration, the provirus is transcribed using the host cellular machinery to generate various spliced and unspliced mRNA transcripts. Some of the mRNA transcripts are then translated into viral proteins. Two copies of the viral RNA genome and the translated viral proteins are translocated to the plasma membrane at viral assembly sites. The Gag polyprotein contains the major structural components of HIV-1 and is sufficient for assembly of virus particles (reviewed in (8-11)), which consist of the matrix protein (MA), capsid protein (CA), nucleocapsid protein (NC), p6 protein, and two spacer peptides (SP1 and SP2). Approximately 5% of the time due to a ribosomal frame shift, GagPol protein is translated(12). GagPol consists of the same components as Gag in addition to the protease (PR), reverse transcriptase (RT), and integrase (IN). The HIV-1 Gag polyprotein mediates the essential events in virion assembly, including binding the plasma membrane, forming an immature Gag lattice necessary to create spherical particles, concentrating the viral envelope (Env) protein, and packaging the genomic RNA via direct interactions with RNA packaging sequence (termed ψ). An immature HIV virion contains approximately 2,500 copies of the of Gag polyprotein(13). These Gag molecules assemble into an immature, spherical lattice due to interactions between the capsid domains with the N-terminal matrix domain directly interacts with the inner leaflet of the plasma membrane and the C-terminal pointing toward the center of the viral particle(14). The immature virion will bud from the producer cell by recruiting and utilizing some of the components present in the ESCRT (Endosomal Sorting Complex Required for Transport) machinery(11). The viral protease (PR) is then activated, and it cleaves the Gag polyprotein into its component parts. After this cleavage event, approximately 1,500 CA monomers assemble into the characteristic fullerene cone

structure, which contains the genomic RNA with associated NC, RT, IN, and viral accessory proteins. After this maturation event, this mature virion can infect a new target cell.

1.1.3. Roles of Capsid in the HIV-1 Replication Cycle

The mature HIV-1 capsid core is a protein shell composed of capsid protein (CA) monomers that houses the viral genome and required replication enzymes. Recent work has demonstrated that it also participates in and may mediate several critical replication events (reviewed in (4,6,15-17)), including uncoating (capsid core remodeling), reverse transcription, nuclear import, integration, evasion of host immune responses, assembly, and maturation. Furthermore, the capsid core interacts with host factors to coordinate these events (reviewed in (5,7,18)). Within these replication events, the exact roles of the different CA assembly states (CA monomers, CA hexamers, CA pentamers, or the assembled CA lattice have not been determined. For a productive infection to occur, a balance between capsid core stability and disassembly must be maintained, and mutations within CA can cause severe viral replication defects(19). Given the relationship between capsid stability and successful virus replication, the assembled capsid core remains a promising therapeutic target for HIV-1 antiretrovirals. CA is genetically fragile(20). Genetic fragility refers to the limited ability for a biological entity to preserve function due to sequence changes. In one study, 70% of 135 single amino acid substitution CA mutations, covering 102 of the 231 CA residues (44%), were found to be replication-defective, and the basis for this observation was that CA must be able to assemble into a mature core of optimal stability that is still capable to perform its critical functions(20). Most CA mutants identified have defects in CA assembly or cause an alteration in CA

stability(19). Therefore, unlike the HIV-1 enzymes which can be mutated to become resistant to inhibitors, CA has limited genetic flexibility to evade therapeutic intervention. However, it is important to note that drug-resistant CA variants have been identified that modulate virus-host interactions and are replication competent(19).

1.1.4. Capsid Structure and Assembly States

The mature HIV-1 capsid core has a fullerene cone structure and is made up of approximately 250 CA hexamers and exactly 12 pentamers (7 pentamers at the broad end and 5 pentamers at the narrow end) to facilitate closing of the cone(21). Not all copies of CA present in the virion are assembled into the mature capsid core. Structures of the CA monomer and the CA assembly states are shown in Figure 1.2. The fullerene type geometry of the capsid core was first proposed based on cryo-electron microscopy studies of *in vitro* assembled CA tubes and cones(21) and was supported in later studies using native viral cores(22). CA dimers were proposed to first organize into a trimer of CA dimers during mature capsid core assembly(23). Hexamer-hexamer spacing within the core is 9.6 nm(24). The capsid monomer consists of an N-terminal domain (NTD) and C-terminal domain (CTD), connected by a flexible linker(22,25-28). The NTD is an arrowhead-shaped domain made up of seven α-helices (numbered 1-7), and the CTD is a globular domain that is made up of a small 3_{10} helix, a highly conserved sheet-turn-helix motif termed the major homology region (MHR), and four α-helices (numbered 8-11)(9). There are three main types of CA-CA interactions important for hexamer and lattice formation. NTD-NTD interactions and NTD-CTD interactions are important for hexamer formation(29). Helices 1-3 of each NTD of a CA hexamer come together to form an 18-helix bundle. The NTD-CTD interface between neighboring CA molecules is formed

when the amino-terminal end of helix 4 of the NTD of one CA monomer is inserted into a groove, made up by helices 8 and 9 of the CTD of the other CA monomer. The dimerization interface between hexamers is made up of CTD-CTD interactions. Specifically, this interaction is formed from pair-wise packing of helix 9 from each CA monomer. The binding affinity of purified, isolated CTDs to each other was determined to be approximately 10-20 μM(30). Of note, alanine mutations at W184 and M185 in the middle of helix 9 decrease dimerization *in vitro* and infectivity in biological assays(31). Substantial water-mediated interactions were also found to help stabilize the capsid lattice(32,33).

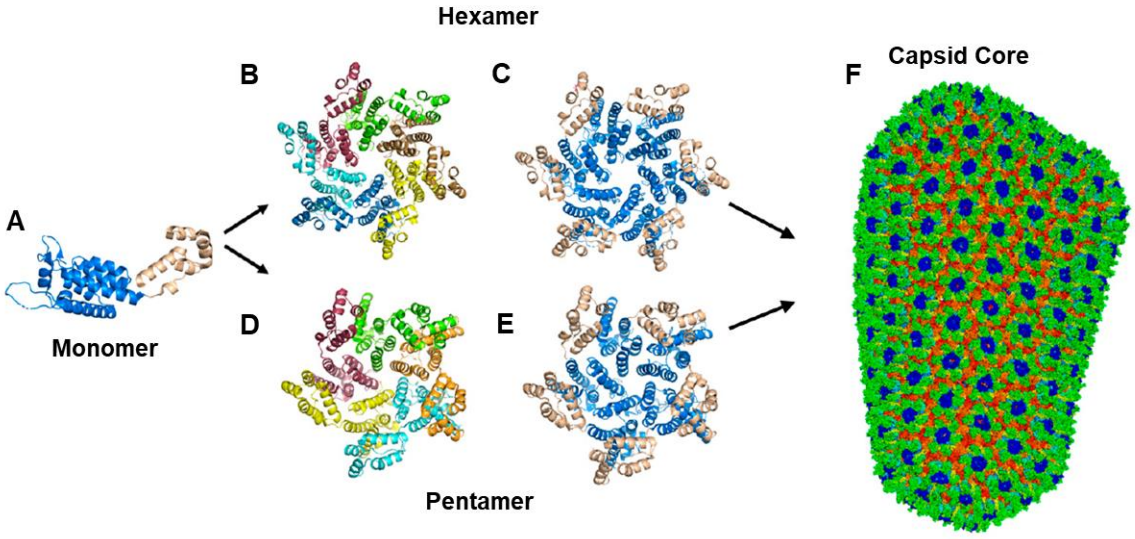

Figure 1.2: Structures of the capsid (CA) protein and its assembly states. (A) Structure of the CA monomer. (B & C) Structures of the CA hexamer. (D & E) Structure of the CA pentamer. For panels A, C, and E, the CA NTD is in blue, and the CA CTD is in tan. Panels B and D show each CA monomer in a different color within a hexamer (B) and pentamer (D). (F) Structure of a capsid core containing approximately 250 CA hexamers and 12 CA pentamers. Adapted from (7). Licensed for use under the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

CA pentamers form quasi-equivalent assemblies to CA hexamers. To form CA pentamers, there are subtle rearrangements of the CA-CA interactions. Similar to CA hexamers, the first three alpha-helices from each CA NTD pack together to form a 15-helix bundle, as compared to the 18-helix bundle in the hexamer, and the NTD-CTD interfaces were found to be very similar(34). Crystal-structures of disulfide-stabilized CA pentamers found that pentamer formation brings charged residues at the center of the ring into close proximity, causing both attractive and repulsive ionic interactions(34). This suggested that electrostatic forces control switching between the pentamer and hexamer, and that pentamers form in the capsid lattice only when required to reduce strain caused by local lattice curvature.

Curvature of the capsid lattice is required to form a closed fullerene cone structure. Therefore, there must be structural plasticity within individual CA monomers, such that they are quasi-equivalent. This is accomplished through flexibility of the NTD-CTD linker that causes changes in the relative orientations of the NTD and CTD(22,26,28,33), variability in the two-fold and three-fold CTD-CTD interactions(33,35), and changes in water-mediated contacts(32,34). Due to these structural variations, the angle between two CA hexamers at the CTD-CTD dimerization interface can range from 135° near hexamer-pentamer junctions to around 180°(9).

1.1.5. Questions in the Capsid Field

While there has been significant progress in understanding the structure and function of capsid and its assembly states, there are still some key questions that remain unanswered, each of which is developed in more detail below. The first is how the maturation of the capsid core occurs. The second is how, when, and where the capsid

core uncoating occurs. Lastly, more work is needed to understand the many interactions of capsid with host proteins and the roles of their interactions within HIV replication. Answering these questions is important not only for better understanding of HIV biology but also for understanding the targetable sites present on the capsid at different stages of virus replication.

1.1.5.1. Maturation and Capsid Assembly

One major question within the field is how the mature viral capsid forms from the immature core. As discussed above in the HIV Replication Cycle section, approximately 2,500 Gag molecules assemble together to form an immature lattice(13). Recently, structures of the immature lattice were determined using cryo-electron tomography and subtomagram averaging of CA-SP1 protein construct(36) and X-ray crystallography of a CA CTD-SP1 construct(37). Residues 356 to 370 of Gag, comprising the C terminus of CA and the first eight residues of SP1, assemble into a 6-helix bundle that is required for immature lattice assembly(36,37). The last eleven residues in CA that are part of the 6-helix bundle were found to be intrinsically disordered in the mature CA protein(25,33). The CA NTD in the immature lattice is rotated and displaced relative to its mature configuration and mediates dimerization instead of hexamerization(36). CA CTD dimerization interactions are important for extending the immature lattice(36,37). The minimal construct required for the assembly of immature virus-like particles was found not to include the CA NTD(38,39). Unlike the mature capsid lattice, the immature lattice contains no pentamers. Instead, it has irregular defects and gaps(40). A comparison of the immature and mature capsid lattice is shown in Figure 1.3. Formation of the CA-SP1 6-

helix bundle has been proposed to be the rate-limiting step of Gag assembly using an *in vitro* assembly reconstitution assay(41).

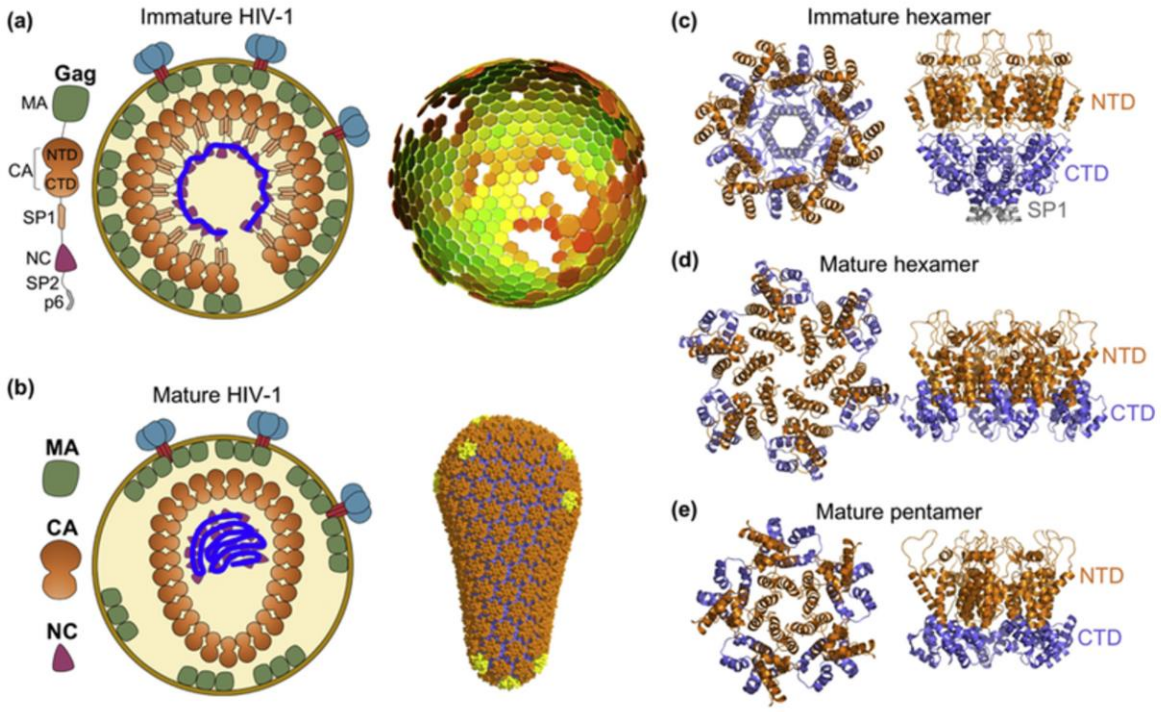

Figure 1.3: Comparison of the immature Gag lattice and the mature capsid lattice. (a) Schematic of the immature virion, which is organized by the Gag polyprotein and its domains. The immature lattice consists of assembled Gag hexamer. (b) Schematic of a mature virion, which is organized by Gag-derived structural proteins (MA, CA, and NC). The mature capsid core is composed of approximately 250 CA hexamers and 12 pentamers. (c) Structure of the immature hexamer, consisting of CA and SP1 (shown in gray). (d) Structure of the mature hexamer. (e) Structure of the mature pentamer. For panels b, c, d, and e, the CA NTD is shown in orange, and the CA CTD is shown in blue.

After the immature virion has budded, Gag is processed by the viral protease into its component parts, and the released CA monomers assemble into the mature capsid core (reviewed in (8-11,43)). How viral protease is activated to cleave the Gag/GagPol polyproteins is currently not known. However, once activated, protease cleaves Gag molecules through a series of five cleavage events that are kinetically regulated and that lead to an ordered release of the different Gag domains(44,45). The order of the cleavage events is SP1/NC, MA/CA and SP2/p6, NC/SP2, and CA/SP1. Similar to how mutations that alter stability of the mature capsid core negatively affect viral infectivity, mutations at the proteolytic cleavage sites lead to defects in core morphology and viral infectivity(46-48). Cleavage at the MA-CA site induces folding of the first twelve residues of the newly formed CA N terminus into a β hairpin, which is stabilized by a salt bridge between Pro1 and Asp51in helix 3(49,50). The final cleavage at the CA/SP1 site is proposed to be the maturation switch as is required for mature capsid assembly(48,51,52). The CA-SP1 cleavage site is located within the 6-helix bundle, and, therefore, the 6-helix bundle must unfold for it to be accessible to protease(37). The basic unit of immature Gag lattice assembly is thought to be a Gag dimer, and partial CA-SP1 helical bundles are present at the edges of the assembled immature lattice and may be where maturation is initiated(53). Another study used hydrogen/deuterium mass spectroscopy and found that the N-terminal β hairpin does not fold until after CA-SP1 cleavage (54), again suggesting that the CA/SP1 cleavage is the maturation switch.

It is currently poorly understood how the capsid monomers assemble into the mature capsid lattice. One model is that the immature CA lattice completely disassembles and reassembles into the mature lattice *de novo*(55). A second model is that is undergoes

a transition to the mature lattice without undergoing a complete disassembly through a displacive transition pathway(56-58). These two models may not be mutually exclusive. One study using a combination of biochemical data and computer simulations suggests that a combination of both displacive transition and disassembly-reassembly occurs(59). Three possible mechanisms for mature capsid assembly are shown in Figure 1.4.

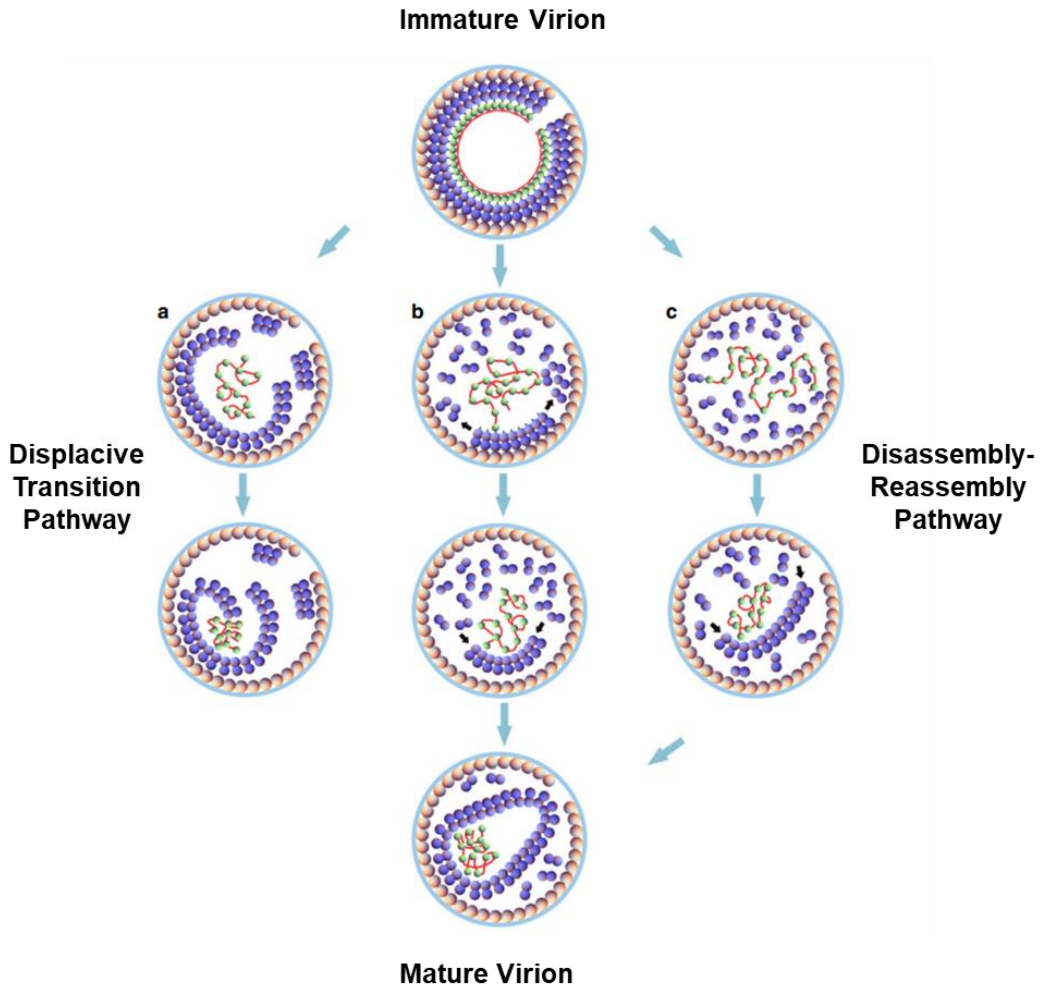

Figure 1.4: Possible models for capsid core maturation. Three models are shown for the maturation of the immature virion shown at the top into the mature virion at the bottom. Shown in the virions are the viral membrane (blue circle), MA proteins (orange circles), CA proteins (purple circles), NC proteins (green circles), and viral RNA genome (red line). Column (a) shows the displacive transition pathway. Column (b) shows a sequential combination of the displacive and *de novo* reassembly models. Column (c) shows the *de novo* disassembly-reassembly pathway. Adapted from (59). Licensed for use under the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

The naturally occurring, highly-negatively charged molecule inositol hexakisphosphate (IP6) recently has been determined to have critical roles in HIV-1 assembly and maturation(60,61). IP6 promotes assembly of CA-SP1 into immature virus-like particles(60), and unidentified electron density that matches the size of IP6 was observed in a cryo-electron microscopy structure of immature lattice between two rings of positively-charged residues (K290 at the bottom of the CA CTD hexamer and K359 at the top of the 6-helix bundle)(36). After proteolysis of the CA-SP1 cleavage site, maturation of the capsid core generates a new binding site for IP6 composed of a R18 ring in the central pore in a CA hexamer(60,61), suggesting that IP6 has a role in mature capsid core assembly and stabilization during post-entry events.

1.1.5.2. Capsid Uncoating

Major questions within the field are how, when, and where capsid uncoating occurs, and how capsid core disassembly is coordinated with the other critical events of HIV-1 infection, including trafficking to the nucleus, reverse transcription, nuclear import, integration, and evasion of host immune responses. Uncoating, also known as capsid core remodeling, is the process by which the capsid core undergoes structural rearrangements, opens, and/or disassembles. It is known that mutations that either increase or decrease capsid core stability reduce or abrogate viral infectivity(62-65). It was also found that mutations that alter the kinetics of uncoating (i.e., early or late uncoating) also cause a decrease in viral infectivity(66). Furthermore, the capsid protein was found to be the main determining factor for HIV's ability to infect non-dividing cells(67,68). Taken together, these observations suggest that capsid plays important roles in these post-entry events.

Models for capsid core uncoating that have been proposed include immediate uncoating upon viral entry, cytoplasmic uncoating, uncoating at the nuclear pore complex, and uncoating in the nucleus (Figure 1.5) (reviewed in (4,6,16,69)). Initially, it was thought that capsid uncoating occurs immediately upon viral entry, as the majority of the CA protein was not found to be associated with the HIV-1 RTCs (reverse transcription complexes)/PICs (pre-integration complexes) from infected cells after partial purification using fractionation (70,71). However, recent studies have pushed back against this model due to the discovery of host factors that bind to the assembled capsid lattice and capsid's role in evading host immune responses, which is discussed in more detail below. Another model proposes that the core begins to disassemble in the cytoplasm(72-75). In this model, some capsid core disassembly occurs in the cytoplasm, and some CA protein remains associated with the RTC to interact with host factors and mediate nuclear import. A third model proposes that the core dissociates at the nuclear pore complex(76,77). As this model suggests, the capsid core docks at the nuclear pore complex, undergoes uncoating, and nuclear import of the PIC follows with some CA associated. The fourth model proposed is that uncoating occurs in the nucleus(78-82). In this model, it is proposed that the intact core is transported through a nuclear pore complex (NPC) before uncoating of the core occurs in the nucleus. NPCs are made up of repeated assemblies of approximately 30 nucleoporins (Nups), and NPCs are present on the nuclear membrane and mediate nucleocytoplasmic transport(83). While small macromolecules less than 40 kDa can passively diffuse through the NPC, larger macromolecules must be actively transported through by nuclear transport factors(83). It was thought that the capsid core could not traverse through the nuclear pore complex due

to its size since the width of the broad end of the capsid core was approximately 60 nm(24), and the width of the NPC channel was measured to be approximately 40 nm(83). However, a recent study measured the nuclear pore complexes of uninfected and infected T cells *in situ* and found that the average NPC channel width was approximately 64 nm(81); therefore, it is possible for an intact capsid core to be transported through the NPC. Another study found that hyperstable capsid mutants could traverse through the NPC but were unable to dissociate from Nup153 on the inside of the NPC, suggesting that a capsid core remodeling event needs to occur to release the core from the NPC(84). Additional studies have observed capsid in the nucleus associated with RTCs/PICs using immunofluorescence microscopy(85,86), which is in agreement with the cytoplasmic, NPC, and nuclear uncoating models. A challenge of these studies is that the intactness of the capsid core or the presence of specific CA assembly states cannot be determined, and new tools are needed to answer these questions.

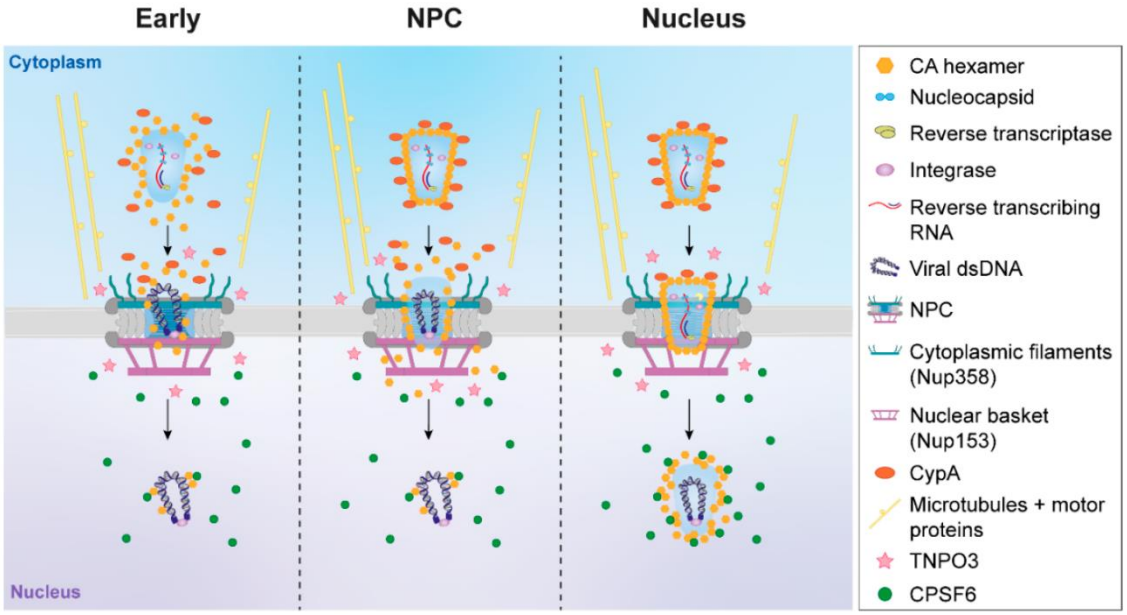

Figure 1.5: Models of capsid uncoating. Early uncoating, uncoating at the nuclear pore complex, and uncoating in the nucleus are depicted with some of the host factors involved in uncoating and nuclear entry. Reprinted from (16). Licensed for use under the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

There are multiple challenges to studying HIV-1 capsid uncoating to determine when and where uncoating occurs(4,16). First, as mentioned above, it is experimentally difficult to distinguish a completely intact core from a partially disassembled core. Second, it is difficult to directly tag the CA protein without affecting stability of the capsid core. Third, these studies utilize different cell lines in their assays, and not all cell lines used are natural targets of HIV-1 infection. Therefore, there may be discrepancies among the studies, and the results may not necessarily be biologically relevant. Fourth, although restriction factors or drugs that target capsid or the viral enzymes can be used in assays to look at how their addition affects viral infectivity or where the block in infection occurred, these studies rely on the assumptions based on the method of action of the drug or restriction factor used, which may not be fully known. Collectively, it is important to take into account these challenges when interpreting results from capsid uncoating studies.

Traditionally, it was thought that reverse transcription was completed in the cytoplasm after capsid uncoating but before nuclear import, and this was supported by a number of studies. From the above capsid uncoating studies, it is clear that capsid uncoating and reverse transcription are interconnected processes, and several studies had found that uncoating is dependent on reverse transcription(72,73,87-89). One study suggested that reverse transcription increases the pressure within the core and that this pressure triggers uncoating, as measured in *in vitro* studies using atomic force microscopy(89). A follow-up study proposed that the different stages of reverse transcription sequentially remodel the capsid until uncoating occurs(90). Christensen et al. were able to reconstitute reverse transcription and integration in a cell-free system

using capsid cores that were isolated from virions, and they were able to visualize mostly intact capsid cores that had lost parts of their lattice with viral cDNA looping out and back into the capsid core(91), suggesting that some remodeling of the capsid may occur during reverse transcription. It is important to note the lack of host factors in this *in vitro* assay. While some assays were done using cell lysates, there are likely host factors that may be cell-specific that contribute to efficient reverse transcription(91). Additionally, the capsid core may serve a container function by keeping the approximately 150 copies of RT in close proximity to the viral nucleic acids during reverse transcription, and, therefore, ensuring that the RT binds its template after dissociation events (13,15). Deoxyribonucleotide triphosphates (dNTPs) are required for reverse transcription, and dNTPs can pass through the electropositive R18 central pore of a CA hexamer(92-94). Molecular dynamics simulations further indicated that the competition for binding at the R18 pore by IP6 and ATP accelerates the rate of dNTP diffusion into the pore by accelerating the release of bound dNTPs(93,94), and a second ring of positively charged residues composed of K25 was also found in hexamer pore that can be coordinated by IP6(94,95). The presence of IP6 caused more efficient reverse transcription in atomic force microscopy and virological experiments(94). However, it has also been shown that the presence of too much IP6 deceases reverse transcription efficiency(91), suggesting the importance of capsid stability and flexibility for reverse transcription to occur.

Another important consideration for capsid uncoating relates to capsid's role in evading innate immune responses (reviewed in (96,97)). Recent studies propose that the ability of capsid to interact with host factors, such as cyclophilin A, cleavage and polyadenylation specificity factor 6 (CPSF6), and nucleoporin 358 (Nup358), prevents

detection by the innate immune system in macrophages and dendritic cells(98,99). This is done by regulating the timing of nuclear import and shielding the viral DNA from the cytosolic DNA sensor, cyclic GMP-AMP synthase (cGAS). Specifically, these studies found that altering CA-host factor interactions either by using CA mutants or changing host factor availability for CA binding (via protein knockdowns or adding a competitor, such as cyclosporin A for cyclophilin A) led to decreased viral infectivity and sensing of the viral DNA by cGAS. Another study found that Non-POU domain-containing octamer-binding protein (NONO) can bind to capsid and associate with cGAS to activate an innate immune response, although this was observed in cells infected with HIV-2 and not HIV-1 1 because NONO binds to the HIV-2 capsid with a higher affinity than the HIV-1 capsid(100). The stability of the capsid core was also found to be important for blocking cGAS sensing of viral DNA, as the infection by Gag cleavage-defective viruses or infection in the presence of a capsid destabilizing compound led to sensing by cGAS(101). Therefore, the integrity of the capsid core is important for evading the innate immune system, in addition to the other key early steps for a productive HIV-1 infection. As summarized in this section, how uncoating, trafficking to the nucleus, reverse transcription, nuclear import, integration, and evasion of host immune responses is spatiotemporally coordinated is not fully understood. New technologies to study these interconnected processes are needed to complement the current approaches.

1.1.5.3. Roles of Capsid Interactions with Host Factors

A third major question within the field is focused on determining the interacting partners with capsid and how their interactions with capsid affect viral infection. To date, at least 25 host factors have been identified as bona fide capsid interacting partners

(reviewed in (7,18)). Table 1.1 lists most of the capsid-interacting host factors that have been identified so far, the step(s) of replication affected by their interaction with capsid, and the binding sites on the capsid if known. These interactions need to be further investigated as it is possible that the virus could use different pathways to infect different cell types. Understanding these pathways and the binding sites on the capsid where these host factors interact could provide new strategies for therapeutic development (Figure 1.6). Furthermore, due to capsid's multiple functions within the replication cycle, there are likely more host factors that directly interact with capsid that have yet to be discovered. Specifically, there are likely factors that interact with the CA domain within the Gag polyprotein or interact with incoming capsid core. This section will discuss some of the capsid-binding host factors and the effects of these host factor interactions on viral replication.

Table 1.1: Capsid-Interacting Host Factors[1]

Name	Related Process	Binding Site on Capsid	Reference
BICD2 (Bicaudal D2)	Trafficking	ND[2]	(102,103)
CLASP2 (Cytoplasmic linker-associated protein 2)	Trafficking	ND	(104)
CLIP170 (Cytoplasmic linker protein 170)	Trafficking	EB1-like domain on pentamer interface	(105)
CPSF6 (Cleavage and polyadenylation specificity factor 6)	Nuclear Import/Integration Site Determination	NTD-CTD interface pocket	(106)
CypA (Cyclophilin A)	Uncoating	CypA binding loop, inter-hexamer interface	(107-109)
Daxx (Death domain-associated protein)	Uncoating inhibition	CA-CypA complex	(110)
DRFs (Diaphanous-related formins)	Uncoating/Trafficking	CA-NC	(111)
ERK2 (Extracellular signal-regulated kinase 2)	Uncoating	Phosphorylates Ser16 on CA	(112)
FEZ1 (Fasciculation and elongation protein zeta-1)	Trafficking	R18 pore	(113,114)
KIF5B (Kinesin-1 heavy chain)	Trafficking	ND	(74)
MAP1A/MAP1S (Microtubule-associated proteins 1A and 1S)	Trafficking	Monomeric CA interface	(115)
MELK (Maternal embryonic leucine zipper kinase)	Uncoating	Phosphorylates Ser149 on CA	(116)

MxB (Myxovirus resistance protein B)	Uncoating inhibition/ Nuclear import inhibition	Inter-hexamer interface	(117-120)
NONO (Non-POU domain-containing octamer-binding protein)	cGas immune response	CA NTD	(100)
Nup153 (Nucleoporin153)	Nuclear import	NTD-CTD interface pocket	(121)
Nup358 (Nucleoporin 358)	Nuclear import	CA NTD	(122-124)
PDZD8 (PDZ domain containing 8)	Uncoating	CA-NC and Gag	(125-127)
Pin1 (Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1)	Uncoating	Phosphorylated Ser16 on CA	(112,128)
REAF (RNA-associated early-stage antiviral factor)	Uncoating inhibitor/ Reverse transcription inhibitor	NTD-CTD interface pocket	(129,130)
Sec24C	Reverse Transcription/Nuclear import	NTD-CTD interface pocket	(131)
SUN1/SUN2	Nuclear import	CA-NC	(132)
TNPO1 (Transportin 1)	Nuclear import	CypA binding loop (G89 critical)	(133)
TNPO3 (Transportin 3)	Uncoating/Nuclear import/integration	CA-CPSF6 complex	(134,135)
TRIM5α (Tripartite motif-containing protein 5α)	Uncoating inhibition	Capsid lattice near CypA binding loop	(136)
TRIMCyp	Uncoating inhibition	CypA binding loop	(137)

[1]Adapted from (7,18)

[2]ND: Capsid binding site has not been determined.

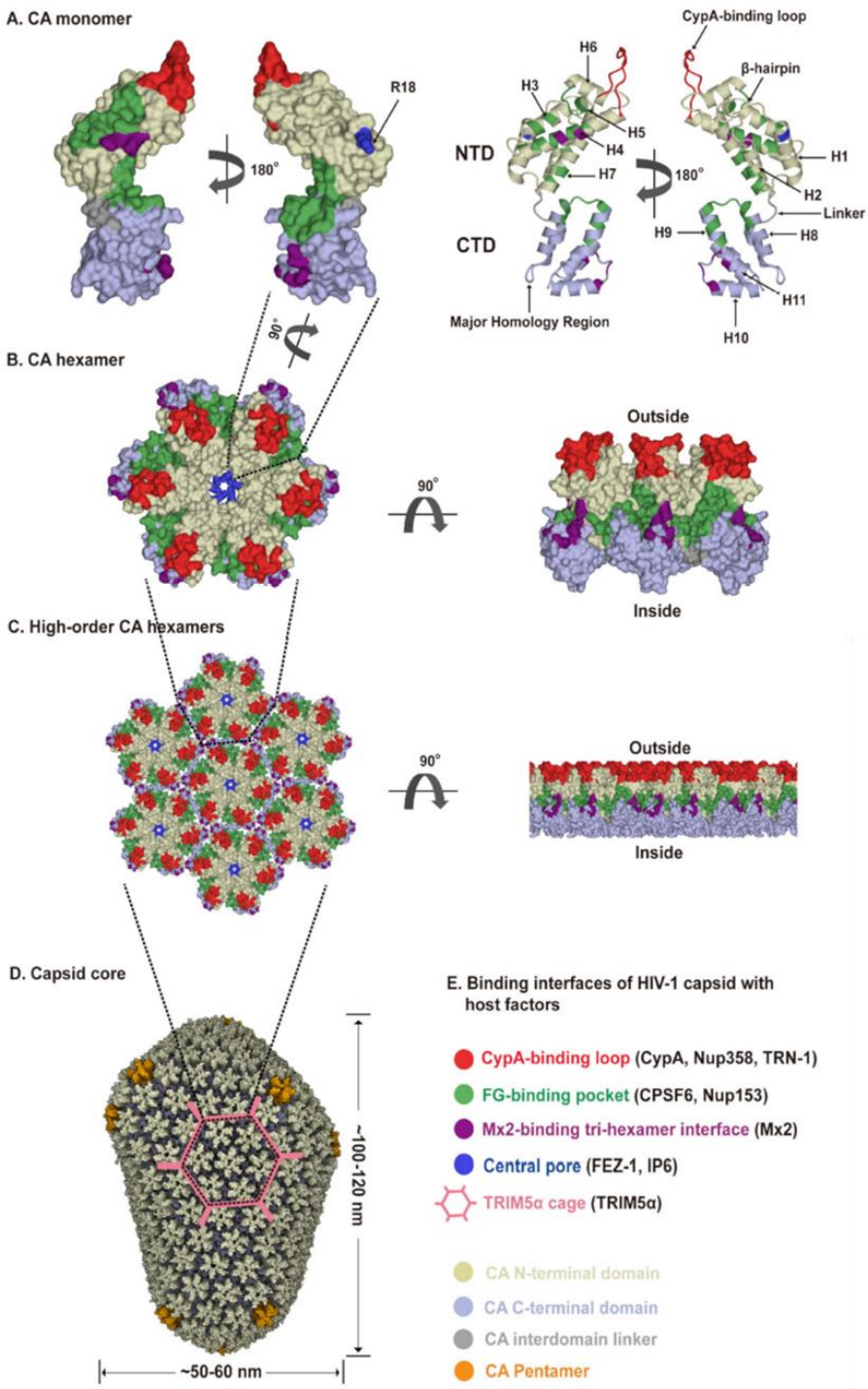

Figure 1.6: Known binding sites present on capsid assembly states. The binding interfaces of HIV-1 capsid with host factors is shown on the CA monomer (A), CA hexamer (B), high-order CA hexamers (C), and the capsid core (D) using the color scheme in (E). FG (phenylalanine-glycine motifs)-binding pocket is the NTD-CTD interface pocket. Figure reprinted from (138). Licensed for use under the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Cyclophilin A (CypA) is a peptidyl-prolyl cis-trans isomerase that directly binds to capsid to facilitate HIV-1 infection(107-109). CypA canonically binds to an exposed, proline-rich CypA binding loop (residues 85-93) in the CA NTD(139), and this loop is critical for CA stability(140). It has recently been shown that CypA can also non-canonically bind to capsid at a site between two adjacent hexamers(141). CypA was found to isomerize the G89-P90 bond in CA(142). While it is known that CypA is packaged into assembling virions, the CypA present in target cells was determined to be important for HIV-1 infection(143,144). The exact role of CypA in HIV-1 replication is unknown, and it has been shown to play roles both in promoting and restricting infection, depending on the cell type or virus strain used in the specific studies. CypA can promote CA core stability(145), and this stabilization delays capsid uncoating(146-148) and promotes reverse transcription(146). Conversely, it has been shown that CypA can promote CA core disassembly(149). Some studies have suggested that this observation is due to a dose-dependent biphasic effect of CypA on core stability, such that lower concentrations of CypA stabilize the core, and higher concentrations destabilize the core(141,147). It is possible that this concentration-dependent destabilization could explain why CypA promotes restriction by restriction factors TRIM5α(150) and MxB(117,151). An additional host factor shown to stabilize the capsid core is PDZ domain-containing protein 8 (PDZD8), which was found to directly bind with capsid and the Gag polyprotein(125-127), but PDZD8 was not required for HIV-1 infection(127).

Cytoplasmic trafficking and nuclear import of the pre-integration complex, whether in an intact or partially disassembled core, require host factor interactions with capsid. Capsid exploits the host cytoskeleton to travel through the cytoplasm to the

nucleus (reviewed in (7,17)). Microtubule-associated proteins MAP1A and MAP1S were shown to help tether the capsid to the microtubules(115). Fasciculation and elongation factor zeta 1 (FEZ1) is a kinesin-1 adaptor protein that directly binds to capsid at the central pore of a CA hexamer, and its interaction with capsid results in net retrograde movement toward the nucleus(113,114). Protein bicaudal D2 (BICD2) is a dynein adaptor protein that binds to capsid using its CC3 domain, and its interaction also contributes to retrograde movement(102,103). Furthermore, depletion of BICD2 led to increased expression of interferon stimulating genes, which may indicate that BICD2 helps HIV-1 evade host immune responses(102). Another study found that using the drug nocodazole to disrupt microtubules reduced capsid uncoating and that inhibiting dynein or kinesin-1 heavy chain (KIF5B) delayed uncoating, demonstrating that there is likely a connection between trafficking and uncoating(74). Capsid also binds some microtubule plus-end tracking proteins (+TIPs), which stabilize microtubules. Cytoplasmic linker-associated protein 2 (CLASP2) is a +TIP that capsid utilizes as an adaptor to traffic to the nucleus(104). Microtubule plus-end binding protein 1 (EB1) depletion was found to prevent HIV-1 from inducing stable microtubule structures(152). While capsid does not specifically bind EB1 itself, it is able to bind EB1-adaptor protein cytoplasmic linker protein-170 (CLIP170) through an EB1-like domain found within the MHR of CA(105). It is interesting to note that this domain is likely accessible for binding at pentamer interfaces due to the higher curvature present in these regions of the capsid core(105). Lastly, diaphanous-related formins (DRFs) Dia1 and Dia2 were shown to bind to *in vitro* CA-NC cores, and these interactions also lead to microtubule stabilization and help with trafficking and uncoating (111).

Phosphorylation of the capsid by host kinases has also been shown to be important for reverse transcription and uncoating. Extracellular Signal-Regulated Kinase 2 (ERK2) phosphorylates Ser16 of CA, and inhibiting this ERK2 activity impaired reverse transcription(112). Ser16 phosphorylation by ERK2 is also required for CA interactions with peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (Pin1) protein(112), which was shown to induce capsid uncoating and was required for successful infection(128). Maternal Embryonic Leucin Zipper Kinase (MELK) was found to phosphorylate Ser149 of CA, and phosphorylated Ser149 was required for reverse transcriptase and optimal uncoating(116).

Capsid cores associated with microtubules located near the NPC were recently visualized using cryo-electron tomography(81), demonstrating the connection between trafficking and nuclear import. Additionally, another study concluded that KIF5B induces a relocalization of Nup358 into the cytoplasm that is dependent on capsid and required for nuclear entry(122). Nup358 is an isomerase that targets the capsid core to the nuclear pore and docks the RTC/PIC to the cytoplasmic side of the NPC(123,124). Nup358 was found to bind to the CypA binding loop on capsid and was proposed to regulate the nuclear import machinery used(148), and its isomerase activity may also be important for coordinating nuclear entry with uncoating(123). Nup153 is the other important Nup for nuclear import of the RTC/PIC, and it directly interacts with capsid(121,153). It forms part of the nuclear basket of the NPC, and its phenylalanine-glycine motifs were found to bind with capsid at the hydrophobic pocket found between two adjacent CA monomers within a hexamer at the NTD-CTD interface(153). Additionally, overexpression of SUN1 and SUN2, members of the linker of nucleoskeleton and cytoskeleton complex, was

found to inhibit HIV-1 infection, and these two proteins were found to directly bind to *in vitro* CA-NC tubes(132). Therefore, these proteins could also play a role in nuclear import.

Cleavage and polyadenylation specificity factor 6 (CPSF6) facilitates nuclear import of the RTC/PIC(106,154,155) and has been implicated in integration site targeting to transcriptionally active chromatin(155,156). A C-terminal truncation of CPSF6 (CPSF6-358) lacking the nuclear localization signal was found to inhibit HIV-1 infection by blocking nuclear import(106). CPSF6 directly binds to the capsid using the same hydrophobic, interprotomer pocket as Nup153(157,158). While CPSF6 interactions with capsid are not required for infection to occur, it was proposed that this interaction confers a significant selective advantage for the virus(159). In this study, humanized mice were infected with HIV-1 containing a CA mutation (A77V) that significantly reduced binding to CPSF6, and it was found that the virus had reverted back to the wild-type CA sequence after 6-8 weeks post infection in 3 of the 4 mice infected(159).

Two transportin proteins have also been identified as capsid interacting host factors. Transportin-3 (TNPO3) is a member of the importin β superfamily of proteins required for importing proteins into the nucleus. TNPO3 was found to be required for infection, and this requirement was mapped to capsid(134,160). TNPO3was found to directly bind with capsid(161,162). However, recent studies suggest that TNPO3 is the nuclear import factor for CPSF6 and indirectly interacts with capsid through CPSF6(135,163,164). Finally, transportin-1 (TNPO1) was also found to bind CA at G89 on the CypA binding loop, and this interaction may also be important for nuclear import(133).

Myxovirus resistance protein B (MxB), also known as myxovirus resistance protein 2 (Mx2), is a restriction factor for HIV-1 infection induced by type I interferon by blocking capsid-dependent nuclear import(117-120). MxB was found to interact with the capsid lattice and not with the CA monomer, hexamer, or pentamer(165,166), and it specifically binds to the capsid lattice at tri-hexamer junctions(167). MxB oligomerization was found to be required for capsid binding(168,169). A recent study had shown that MxB blocks nuclear import through a Nup358-dependent pathway(170).

Tripartite motif-containing (TRIM) proteins TRIM5α and TRIMCyp from rhesus macaques and owl monkeys, respectively, bind directly to the assembled capsid core and restrict HIV-1 infection in non-human primate cells(137,171). These proteins are composed of RING, B-box 2, coiled-coil, and a capsid recognition domain. For TRIM5α, the SPRY (called PRYSPRY or B30.2) domain is the capsid recognition domain and binds the capsid lattice. While the exact nature of the interaction of the SPRY domain with the capsid lattice is still being determined, recent studies using engineered CA protein constructs, cryo-electron tomography, and molecular dynamics simulations suggests that it can bind to varied sites near the two-fold and three-fold symmetry axes between hexamers near the CypA binding loop(172-174). TRIMCyp binds capsid using its CypA domain. The RING domain is a RING E3 ubiquitin ligase, and as a dimer it interacts with a ubiquitin-conjugating E2 enzyme(175). Both TRIM proteins restrict HIV-1 infection by assembling a hexagonal lattice cage around the capsid core, and its binding to the core is enhanced by avidity effects(173,174,176-178). Higher order assembly of TRIM molecules facilitate the dimerization of RING domains and the arrangement of the capsid recognition domains within the assembled TRIM cages to match the arrangement

of the CA binding sites within the lattice. TRIM molecules dimerize through their coiled-coil domains, and higher-order assembly of TRIM dimers occurs through their B-box 2 domains(173,179). The precise mechanism of restriction by TRIM5α and TRIMCyp is not fully understood (reviewed in (180)), but it does not require core disruption and proteasomal degradation(181).

1.1.6. Tools to Study Capsid

As described above, there have been many advances in elucidating the roles of capsid and its assembly states within the HIV-1 replication. This is, in part, due to extensive structural biology studies, CA protein engineering, and the discovery of capsid-targeting small molecules and peptides. The structural biology studies using X-ray crystallography, nuclear magnetic resonance, cryo-electron microscopy/cryo-electron tomography, and molecular dynamics have been critical for understanding the interaction surfaces present within the CA assembly states and how capsid interacts with its binding partners.

Engineering stable forms of capsid assembly states have been important for *in vitro* studies for evaluating interactions between capsid and host factors or small molecules. The A14C/E45C capsid mutant assembles into hyperstable CA lattice tubes, due to the formation of disulfide bonds between adjacent CA monomers within a CA hexamer, under high salt conditions(182). Importantly, the assembled CA lattice tubes remain assembled after dialysis in buffer containing physiological salt conditions. Addition of the W184A/M185A mutations into this CA protein construct disrupts the CTD dimerization domain and abolishes higher order CA assembly. Therefore, the A14C/E45C/W184A/M185A capsid mutants can be assembled into soluble, disulfide-

stabilized CA hexamers(182). The assembled CA lattice and soluble CA hexamer protein constructs were critical for the selection of capsid lattice-binding aptamers described in Chapters 2 and 3. Recently, Summers et al. engineered different modular CA assemblies that can be used to look at specific CA recognition surfaces utilized by host factors or small molecules(183).

The discovery of capsid-targeting drugs has not only been an important advance for developing new classes of antiretroviral therapies, but capsid-targeting drugs also have become useful tools for studying capsid function. A recent review discussing the efforts in developing CA-targeting drugs is found in reference (184), and some of the CA-targeting drugs are summarized in Table 1.2. Like CA interactions with host factors, there have been a number of sites on capsid that have been targeted. PF-3450074 (PF74) was one of the first small molecules discovered to target the assembled capsid in 2010(185). PF74 is a small molecule HIV-1 inhibitor that targets the capsid protein and inhibits viral replication at both the early and lates stages(185). PF74 binds at the NTD-CTD interface and blocks binding of host factors, such as CPSF6 and NUP153(157,158). Unfortunately, PF74 has low metabolic stability and cannot be used in the clinic. Second-generation capsid-targeting small molecules, GS-CA1 and GS-6207 (lencapavir), bind in the same pocket as PF74, have increased metabolic stability, and were found to increase capsid stability and interfere with nuclear import of viral cDNA(186,187). GS-6207 is currently in phase II/III clinical trials. CAP-1 is a capsid-targeting small molecule that binds to a site on the CA NTD and was found to inhibit mature capsid assembly(188); however, due to its low affinity to CA, it was not further pursued. However, more potent small molecules that target this site could be developed. CAI (capsid assembly inhibitor)

is a peptide-based inhibitor that targets the hexamer dimerization interface(189). It was found to inhibit immature and mature HIV-1 assembly *in vitro* but was unable to inhibit assembly in biological assays(189). Finally, the central pore in a CA hexamer where IP6 binds could also be targeted by a drug(60,61). A previous study found that hexacarboxybenzene can block the central pores and inhibit reverse transcription(92). While this molecule is not a good candidate for therapeutics since it cannot cross the plasma membrane, this result suggests that therapeutics could be developed against this site. In summary, efforts in developing drugs targeted to capsid have identified several possible binding sites, including those present on the CA NTD of a CA monomer, either at the NTD-CTD interface or central pore of a CA hexamer, and at the hexamer dimerization interface found in the assembled CA lattice.

Table 1.2: Examples of CA-Targeting Drugs[a]

Compound	Binding Site on Capsid	Effect on CA Stability/Assembly[b]	HIV-1 Stage Targeted	Reference
PF74	NTD-CTD interface	Stabilize, Increase multimerization	Early/Late	(185)
BI-1 and BI-2	NTD-CTD interface	Stabilize, Increase Multimerization	Early/Late	(190)
Courmermycin-A1 (C-A1)	NTD-CTD interface	ND	Early	(191,192)
GS-CA1	NTD-CTD interface	Increase multimerization	Early/Late	(186)
GS-6207	NTD-CTD interface	Increase multimerization	Early/Late	(187)
CAP-1	CA NTD	Decrease multimerization	Late	(188)
CAI (capsid assembly inhibitor)	Hexamer dimerization interface	Decrease multimerization	Early/Late	(189)

[a]Adapted from (184)

[b]Determined using thermal shift assay or *in vitro* capsid assembly assay. ND: effects of compound on CA stability and assembly were not determined.

The advances described in this section have been important for studying capsid, its structure, and its functions. However, as described in section 1.1.5, there are still important questions regarding capsid that need to be answered. New tools are needed that can be used to differentiate capsid assembly states to assess their roles in the HIV-1 replication cycle. Chapters 2 and 3 will discuss the development of RNA aptamers selected to bind to capsid assembly states and their potential applications to elucidate the contributions of CA assembly states to HIV-1 replication.

1.2. Nucleic Acid Aptamers

Aptamers are single-stranded nucleic acid oligonucleotides that fold into unique, three-dimensional shapes capable of binding to a target with high specificity and usually high affinity. Due to their unique binding properties, aptamers can be used for a number of applications, including tools to study biological questions, therapeutics, diagnostics, delivery tools, *in vitro* and *in vivo* imaging, biomarker discovery and biosensors(193-202). Further, because they can fold into a variety of three-dimensional structures, they can be selected to discriminate against very similar targets, such as proteins with a single amino acid change(203,204) or different conformations of the same protein(205-209).

1.2.1. Aptamer Selections

Aptamers are generated through an iterative process termed Systematic Evolution of Ligands by EXponential enrichment (SELEX)(210,211). A schematic of SELEX is shown in Figure 1.7. A typical aptamer library designs consists of a random region, typically consisting of 30-70 random nucleotides, flanked by two constant regions for amplification purposes. A SELEX experiment starts with an aptamer library that consists of approximately 10^{14} to 10^{15} sequences. The aptamer library is incubated with the target

of interest, which can be a protein, metal ion, virus, or even a whole cell. Bound sequences are separated from unbound sequences during a partition step. The bound sequences are recovered, and for RNA libraries, they are reverse transcribed into cDNA, amplified using polymerase chain reaction (PCR), and the enriched library used in the next round is generated via *in vitro* transcription.

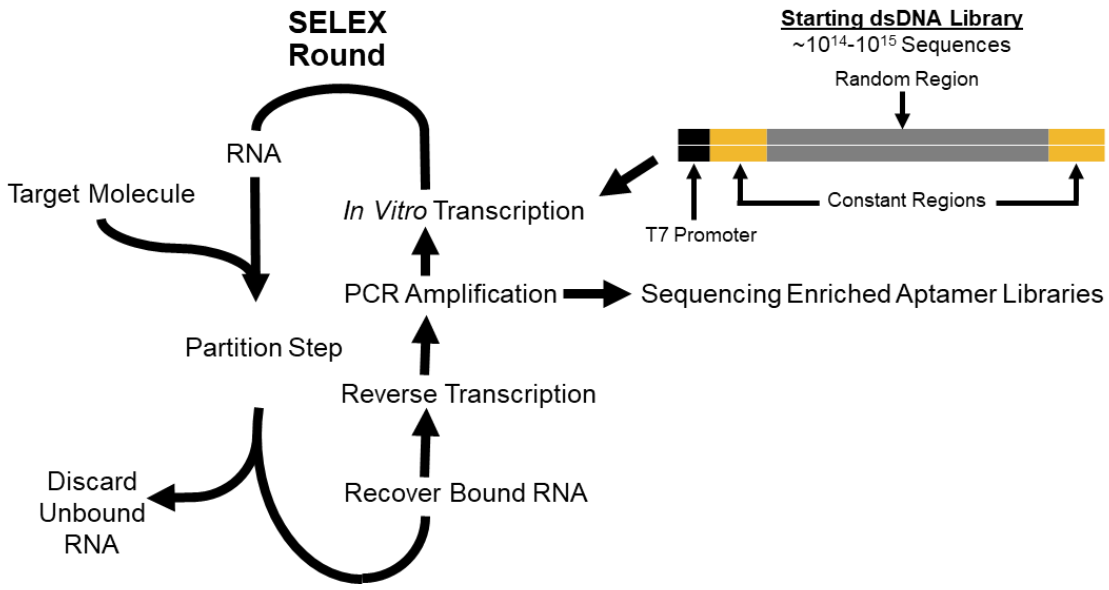

Figure 1.7: SELEX Schematic. For an RNA aptamer selection, a starting dsDNA library, composed of a random region flanked by two constant regions for amplification, can be transcribed into RNA by T7 RNA polymerase due to the presence of the T7 promoter. The RNA library is incubated with the target of interest, and a partition step occurs to separate the bound sequences from the unbound. The bound RNA is recovered, reverse transcribed into cDNA, and amplified via PCR to generate dsDNA templates that can be used in *in vitro* transcription reaction for the next round or for sequencing the aptamer libraries using Sanger sequencing or high-throughput sequencing.

The aptamer selection conditions should resemble the conditions that the aptamer will experience in future applications. Two important considerations are the buffer composition and temperature that the aptamer will experience when bound to a target in its native environment. Another consideration is the type of nucleic acid used in the selection, such as DNA or RNA with or without modifications. For example, using unmodified RNA aptamers is useful for intracellular expression of aptamers to cellular targets. However, if the aptamer will be used in environments that may have nucleases present, adding modifications to increase stability may be important. Aptamer modifications are useful not only for aptamer stability but also for increasing the types of interactions that can be made with the target(212,213). The types of modifications can occur on the sugar, nucleobase, and internucleotide linkages (Figure 1.8). There are two strategies commonly used to incorporate these modifications into aptamers(214,215). The first is to start the selection using the chemically-modified random libraries(212). The advantage of this *de novo* approach is that it evaluates the binding of a large number of modified sequences simultaneously during the selection. The challenges with this approach are that the modified nucleotides can be expensive or are challenging to synthesize in house, and that there must be polymerases that can recognize and/or incorporate the modified nucleotides of interest(216-218). The second strategy commonly used to generate modified aptamers is post-SELEX optimization(219). In this approach, selections are first done using unmodified libraries. Once an aptamer of interest is identified, modifications are systematically introduced into the aptamer sequences in various combinations, and the effects on aptamer binding and stability are evaluated. This process is repeated until the aptamer has the desired properties. A challenge of this

approach is that the introduction of modifications can alter the aptamer's structure and its ability to bind to its aptamer. Therefore, this process can become labor-intensive and expensive, and there is no guarantee that it will be successful. In Chapter 4, we present an alternative strategy where a pre-enriched aptamer library is re-transcribed with a modification of interest, and the modified library undergoes several rounds of selection to identify sequences that could tolerate or accommodate the modification.

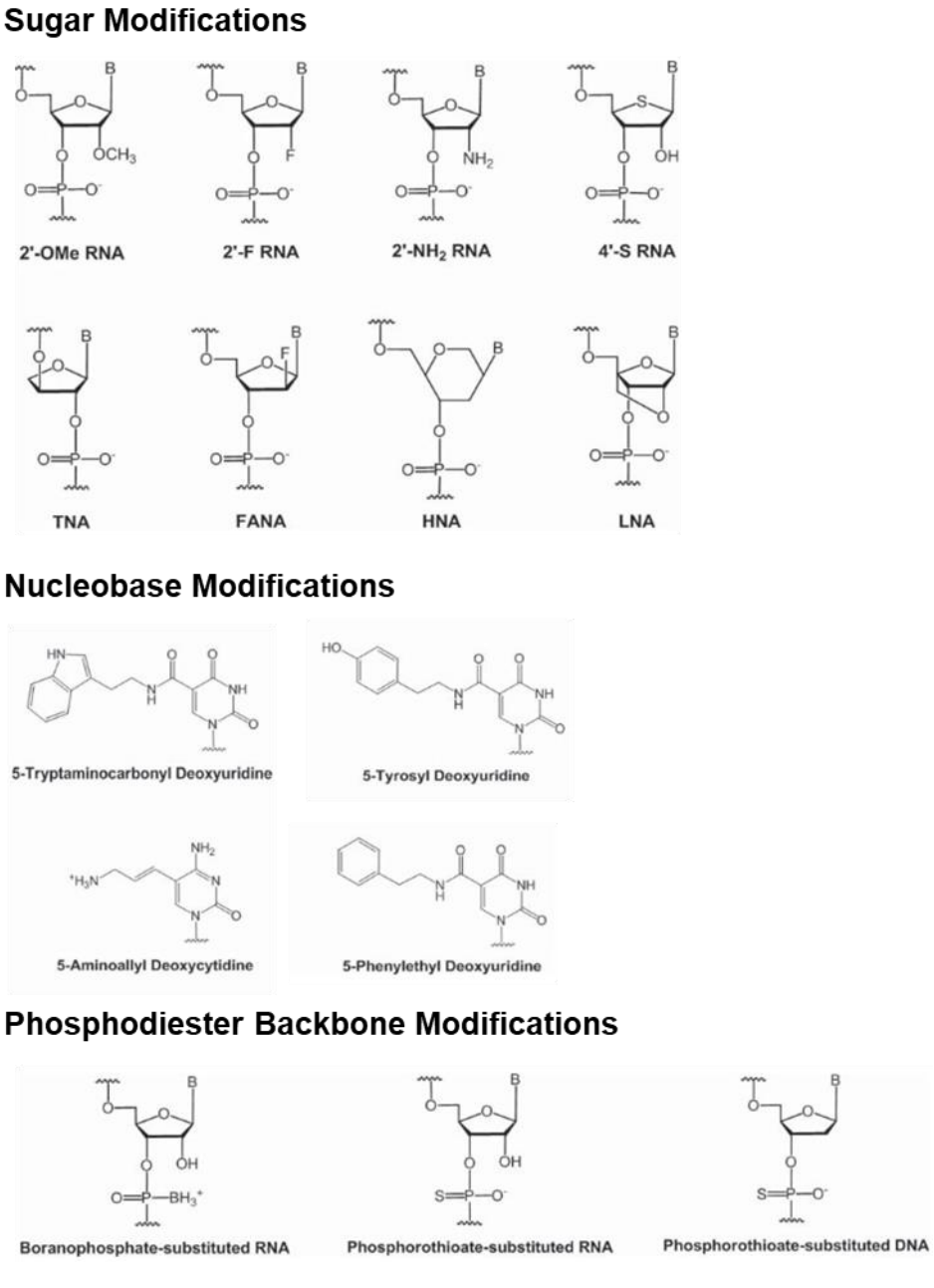

Figure 1.8: Examples of aptamer chemical modifications. Some of the common chemical sugar, nucleobase, and backbone modifications are shown. Adapted from (220).

1.2.3. HTS SELEX and Aptamer Bioinformatics

After a number of selection rounds, aptamer sequences are traditionally determined by cloning the aptamer libraries into a plasmid and using Sanger sequencing. These identified aptamer sequences are then tested for binding. This approach can provide a snapshot of the state of the aptamer library. For example, if one screens 20-50 clones, and there are few duplicate or related sequences, that suggests that the aptamer population has not converged. In contrast, if many of the clones screened are duplicate sequences or closely related sequences, that suggests that the population is converging. However, the amount of information one can get from Sanger sequencing is limited by the number of clones screened.

With high-throughput sequencing (HTS), millions of sequence reads from multiple rounds of selection can be determined, and this information can be used to identify aptamer candidates for further characterization (221-228). HTS has allowed for investigation of *in vitro* selection pools by revealing their sequence information and to assess how the populations evolve through the course of the experiment (229,230). These data can also provide information on the success of the selection(222,231), aptamer-target interactions(204,232-234), and the mutation landscape (221,235,236).

The development of bioinformatic toolkits to analyze the large datasets from HTS experiments overcame a major bottleneck for incorporating HTS analyses into combinatorial selection experiments(237). Common tasks of bioinformatics toolkits include counting the number of occurrences for each sequence(238,239), calculating enrichment of sequences between two or more rounds(239-241), filtering sequences based on the number of reads present in one or multiple rounds(242), clustering related

sequences(235,239,243,244), or analyzing predicted secondary structure motifs(235,245-249). FASTAptamer is one such toolkit and was developed in the Burke laboratory. FASTAptamer consists of five Perl scripts that can be used to analyze HTS data and address primary, sequence-level questions, such as counting sequence reads, clustering related sequences, and calculating enrichment values between two rounds.

A limiting step in aptamer research after completing a selection is the characterization of the resulting aptamers. Combining aptamer selections with HTS and bioinformatic analyses can help to speed up this process. However, one challenge of this approach is trying to limit false positives. Previous studies have shown that the most abundant sequences in the last round of selection are often not the best binders(224,231,241), likely due to amplification biases. Therefore, being able to track how individual aptamers and their cluster of related aptamers perform during the course of the selection can help to identify aptamers with the desired characteristics(227). Additionally, secondary structure motifs present within the aptamer sequences can be tracked if there is knowledge of the types of motifs present within the aptamer library(225). To increase the likelihood of success for identifying aptamers, it requires careful planning and execution of the selection, library preparation for HTS, and bioinformatic analyses.

1.3. Overview of Dissertation

This dissertation will study aptamer-protein interactions using a combination of bioinformatics and biochemical methods. Chapters 2 and 3 contain two studies using RNA aptamers that were selected to bind the assembled HIV-1 capsid lattice. Chapter 2 focuses on the initial aptamer selection and characterization of aptamer CA15-2. This

work represents that first successful aptamer selection against the assembled HIV-1 capsid, and it was published in *Nucleic Acids Research*(250). Aptamer CA15-2 specifically bound to the assembled capsid lattice but not to the soluble capsid hexamer or capsid protein monomer. Further, it was found that CA15-2 inhibited viral infectivity in producer cell assays.

Chapter 3 describes the selection and characterization aptamers that either bind only to the assembled capsid lattice or bind both the assembled capsid lattice and soluble capsid hexamer. This marks a critical step in identifying aptamer sequences capable of differentiating CA assembly states. Aptamers having these phenotypes were selected using a combination of positive and negative selection rounds. Aptamer populations from the differentiation selection trajectories were then subjected to high-throughput sequencing and bioinformatic analyses using FASTAptamer(239) and its R-based reimplementation FASTAptameR 2.0(251). Aptamer binding phenotypes (i.e., lattice-specific binders or lattice and hexamer binders) were predicted based on the enrichment values of the individual aptamers and their cluster of related sequences. The candidate aptamers were tested for binding to capsid lattice and hexamer, and 74% of the predicted binding phenotypes matched the actual binding phenotypes. Structural motifs present in the aptamer populations were studied using truncations, enzymatic probing, and covariation analyses to understand some of the sequence and/or structural requirements for aptamer binding to capsid, and current efforts are focused on the further elucidation of these requirements. Future experiments will explore the binding sites of lattice-specific and lattice and hexamer-binding aptamers.

Chapter 4 explores the unexpected finding that the presence of 2′-fluoro-modified pyrimidines in RNA oligonucleotides causes non-specific RNAs to be able to bind HIV-1 reverse transcriptase (RT). This observation highlighted the importance of confirming that the modification of interest does not cause non-specific binding to the target. This work was published in *RNA*(252). In this study, pre-enriched aptamer libraries to RT were re-transcribed with 2′-modified pyrimidines (2′-fluoro, 2′-O-methyl, or 2′-amino) and underwent three additional rounds of selection. Representative aptamers from several structural families identified by high-throughput sequencing were tested for RT inhibition, and for each modification, there were aptamers that were able to inhibit RT when modified. Therefore, this reselection approach provided an alternative strategy for identifying aptamers that tolerate chemical modifications. 2′-fluoro-pyrimidine (2′-FY) aptamers were the strongest RT inhibitors, as compared to aptamers modified with the 2′-O-methyl or 2′-amino pyrimidines. Unexpectedly, a diverse panel of retroviral RTs was strongly inhibited by all 2′-FY-modified transcripts, including sequences that do not inhibit those RTs as unmodified RNA. The magnitude of RT inhibition by 2′-FY RNAs was proportional to mole fraction 2′-FY present in the RNA transcript. 2′-FY RNA binding to RT was more sensitive to salt concentration than binding by unmodified RNAs, suggesting that the interaction with retroviral RTs is more ionic in character for 2′-FY RNA than for unmodified 2′-OH RNA.

In Chapter 5, the overall conclusions and future perspectives from this work will be discussed.

1.4. References

1. World Health Organization. (2021). HIV/AIDS. https://www.who.int/news-room/fact-sheets/detail/hiv-aids
2. Douek, D.C., Brenchley, J.M., Betts, M.R., Ambrozak, D.R., Hill, B.J., Okamoto, Y., Casazza, J.P., Kuruppu, J., Kunstman, K., Wolinsky, S. *et al.* (2002) HIV preferentially infects HIV-specific CD4+ T cells. *Nature*, **417**, 95-98.
3. Chen, B. (2019) Molecular Mechanism of HIV-1 Entry. *Trends Microbiol.*, **27**, 878-891.
4. Campbell, E.M. and Hope, T.J. (2015) HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat. Rev. Microbiol.*, **13**, 471-483.
5. Novikova, M., Zhang, Y., Freed, E.O. and Peng, K. (2019) Multiple roles of HIV-1 capsid during the virus replication cycle. *Virol. Sin.*, **34**, 119-134.
6. Ambrose, Z. and Aiken, C. (2014) HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. *Virology*, **454-455**, 371-379.
7. Rossi, E., Meuser, M.E., Cunanan, C.J. and Cocklin, S. (2021) Structure, function, and interactions of the HIV-1 capsid protein. *Life*, **11,** 100.
8. Briggs, J.A. and Krausslich, H.G. (2011) The molecular architecture of HIV. *J. Mol. Biol.*, **410**, 491-500.
9. Ganser-Pornillos, B.K., Yeager, M. and Pornillos, O. (2012) Assembly and architecture of HIV. *Adv. Exp. Med. Biol.*, **726**, 441-465.
10. Sundquist, W.I. and Krausslich, H.G. (2012) HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.*, **2**, a006924.

11. Freed, E.O. (2015) HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.*, **13**, 484-496.

12. Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J. and Varmus, H.E. (1988) Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature*, **331**, 280-283.

13. Carlson, L.-A., Briggs, J.A.G., Glass, B., Riches, J.D., Simon, M.N., Johnson, M.C., Müller, B., Grünewald, K. and Kräusslich, H.-G. (2008) Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. *Cell Host Microbe*, **4**, 592-599.

14. Wilk, T., Gross, I., Gowen, B.E., Rutten, T., de Haas, F., Welker, R., Kräusslich, H.-G., Boulanger, P. and Fuller, S.D. (2001) Organization of immature human immunodeficiency virus type 1. *J. Virol.*, **75**, 759-771.

15. Aiken, C. and Rousso, I. (2021) The HIV-1 capsid and reverse transcription. *Retrovirology*, **18**, 29.

16. Guedán, A., Caroe, E.R., Barr, G.C.R. and Bishop, K.N. (2021) The role of capsid in HIV-1 nuclear entry. *Viruses*, **13,** 1425.

17. Naghavi, M.H. (2021) HIV-1 capsid exploitation of the host microtubule cytoskeleton during early infection. *Retrovirology*, **18**, 19.

18. Yamashita, M. and Engelman, A.N. (2017) Capsid-dependent host factors in HIV-1 infection. *Trends Microbiol.*, **25**, 741-755.

19. Saito, A. and Yamashita, M. (2021) HIV-1 capsid variability: viral exploitation and evasion of capsid-binding molecules. *Retrovirology*, **18**, 32.

20. Rihn, S.J., Wilson, S.J., Loman, N.J., Alim, M., Bakker, S.E., Bhella, D., Gifford, R.J., Rixon, F.J. and Bieniasz, P.D. (2013) Extreme genetic fragility of the HIV-1 capsid. *PLoS Pathog.*, **9**, e1003461.

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

22. Mattei, S., Glass, B., Hagen, W.J.H., Kräusslich, H.-G. and Briggs, J.A.G. (2016) The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science*, **354**, 1434-1437.

23. Tsiang, M., Niedziela-Majka, A., Hung, M., Jin, D., Hu, E., Yant, S., Samuel, D., Liu, X. and Sakowicz, R. (2012) A trimer of dimers Is the basic building block for human immunodeficiency virus-1 capsid assembly. *Biochemistry*, **51**, 4416-4428.

24. Briggs, J.A., Wilk, T., Welker, R., Kräusslich, H.G. and Fuller, S.D. (2003) Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J.*, **22,** 1701-1715.

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

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

27. Worthylake, D.K., Wang, H., Yoo, S., Sundquist, W.I. and Hill, C.P. (1999) Structures of the HIV-1 capsid protein dimerization domain at 2.6 A resolution. *Acta Crystallogr.D*, **55**, 85-92.

28. Bayro, M.J., Chen, B., Yau, W.-M. and Tycko, R. (2014) Site-specific structural variations accompanying tubular assembly of the HIV-1 capsid protein. *J. Mol. Biol.*, **426**, 1109-1127.

29. Yufenyuy, E.L. and Aiken, C. (2013) The NTD-CTD intersubunit interface plays a critical role in assembly and stabilization of the HIV-1 capsid. *Retrovirology*, **10**, 29.

30. Gamble, T.R., Yoo, S., Vajdos, F.F., von Schwedler, U.K., Worthylake, D.K., Wang, H., McCutcheon, J.P., Sundquist, W.I. and Hill, C.P. (1997) Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science*, **278**, 849-853.

31. Yu, X., Wang, Q., Yang, J.-C., Buch, I., Tsai, C.-J., Ma, B., Cheng, S.Z.D., Nussinov, R. and Zheng, J. (2009) Mutational analysis and allosteric effects in the HIV-1 capsid protein carboxyl-terminal dimerization domain. *Biomacromolecules*, **10**, 390-399.

32. Gres, A.T., Kirby, K.A., KewalRamani, V.N., Tanner, J.J., Pornillos, O. and Sarafianos, S.G. (2015) X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science*, **349**, 99-103.

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

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

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

36. Schur, F.K.M., Obr, M., Hagen, W.J.H., Wan, W., Jakobi, A.J., Kirkpatrick, J.M., Sachse, C., Kräusslich, H.-G. and Briggs, J.A.G. (2016) An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science*, **353**, 506-508.

37. Wagner, J.M., Zadrozny, K.K., Chrustowicz, J., Purdy, M.D., Yeager, M., Ganser-Pornillos, B.K. and Pornillos, O. (2016) Crystal structure of an HIV assembly and maturation switch. *eLife*, **5**, e17063.

38. Borsetti, A., Ohagen, A. and Göttlinger, H.G. (1998) The C-terminal half of the human immunodeficiency virus type 1 Gag precursor is sufficient for efficient particle assembly. *J. Virol.*, **72**, 9313-9317.

39. Accola, M.A., Strack, B. and Göttlinger, H.G. (2000) Efficient particle production by minimal Gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain. *J. Virol.*, **74**, 5395-5402.

40. Briggs, J.A.G., Riches, J.D., Glass, B., Bartonova, V., Zanetti, G. and Kräusslich, H.-G. (2009) Structure and assembly of immature HIV. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 11090-11095.

41. Kucharska, I., Ding, P., Zadrozny, K.K., Dick, R.A., Summers, M.F., Ganser-Pornillos, B.K. and Pornillos, O. (2020) Biochemical reconstitution of HIV-1 assembly and maturation. *J. Virol.*, **94**, e01844-19.

42. Pornillos, O. and Ganser-Pornillos, B.K. (2019) Maturation of retroviruses. *Curr. Opin. Virol.*, **36**, 47-55.

43. Mattei, S., Schur, F.K.M. and Briggs, J.A.G. (2016) Retrovirus maturation—an extraordinary structural transformation. *Curr. Opin. Virol.*, **18**, 27-35.

44. Pettit, S.C., Moody, M.D., Wehbie, R.S., Kaplan, A.H., Nantermet, P.V., Klein, C.A. and Swanstrom, R. (1994) The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J. Virol.*, **68**, 8017-8027.

45. Wiegers, K., Rutter, G., Kottler, H., Tessmer, U., Hohenberg, H. and Krausslich, H.G. (1998) Sequential steps in human immunodeficiency virus particle maturation revealed by alterations of individual Gag polyprotein cleavage sites. *J. Virol.*, **72,** 2846-2854.

46. Müller, B., Anders, M., Akiyama, H., Welsch, S., Glass, B., Nikovics, K., Clavel, F., Tervo, H.-M., Keppler, O.T. and Kräusslich, H.-G. (2009) HIV-1 Gag processing intermediates trans-dominantly interfere with HIV-1 infectivity *J. Biol. Chem.*, **284**, 29692-29703.

47. Fontana, J., Keller, P.W., Urano, E., Ablan, S.D., Steven, A.C. and Freed, E.O. (2015) Identification of an HIV-1 mutation in spacer peptide 1 that stabilizes the immature CA-SP1 lattice. *J. Virol.*, **90**, 972-978.

48. Datta, S.A., Temeselew, L.G., Crist, R.M., Soheilian, F., Kamata, A., Mirro, J., Harvin, D., Nagashima, K., Cachau, R.E. and Rein, A. (2011) On the role of the SP1 domain in HIV-1 particle assembly: a molecular switch? *J. Virol.*, **85**, 4111-4121.

49. Gitti, R.K., Lee, B.M., Walker, J., Summers, M.F., Yoo, S. and Sundquist, W.I. (1996) Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science*, **273**, 231-235.

50. von Schwedler, U.K., Stemmler, T.L., Klishko, V.Y., Li, S., Albertine, K.H., Davis, D.R. and Sundquist, W.I. (1998) Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *EMBO J.*, **17**, 1555-1568.

51. Kräusslich, H.G., Fäcke, M., Heuser, A.M., Konvalinka, J. and Zentgraf, H. (1995) The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *J. Virol.*, **69**, 3407-3419.

52. Gross, I., Hohenberg, H., Wilk, T., Wiegers, K., Grättinger, M., Müller, B., Fuller, S. and Kräusslich, H.-G. (2000) A conformational switch controlling HIV-1 morphogenesis. *EMBO J.*, **19**, 103-113.

53. Tan, A., Pak, A.J., Morado, D.R., Voth, G.A. and Briggs, J.A.G. (2021) Immature HIV-1 assembles from Gag dimers leaving partial hexamers at lattice edges as potential substrates for proteolytic maturation. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2020054118.

54. Monroe, E.B., Kang, S., Kyere, S.K., Li, R. and Prevelige, P.E. (2010) Hydrogen/deuterium exchange analysis of HIV-1 capsid assembly and maturation. *Structure*, **18**, 1483-1491.

55. Keller, P.W., Huang, R.K., England, M.R., Waki, K., Cheng, N., Heymann, J.B., Craven, R.C., Freed, E.O. and Steven, A.C. (2013) A two-pronged structural analysis of retroviral maturation indicates that core formation proceeds by a disassembly-reassembly pathway rather than a displacive transition. *J. Virol.*, **87**, 13655-13664.

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

57. Frank, G.A., Narayan, K., Bess, J.W., Del Prete, G.Q., Wu, X., Moran, A., Hartnell, L.M., Earl, L.A., Lifson, J.D. and Subramaniam, S. (2015) Maturation of the HIV-1 core by a non-diffusional phase transition. *Nat. Commun.*, **6**, 5854.

58. Woodward, C.L., Cheng, S.N. and Jensen, G.J. (2015) Electron cryotomography studies of maturing HIV-1 particles reveal the assembly pathway of the viral core. *J. Virol.*, **89**, 1267-1277.

59. Ning, J., Erdemci-Tandogan, G., Yufenyuy, E.L., Wagner, J., Himes, B.A., Zhao, G., Aiken, C., Zandi, R. and Zhang, P. (2016) In vitro protease cleavage and computer simulations reveal the HIV-1 capsid maturation pathway. *Nat. Commun.*, **7**, 13689.

60. Dick, R.A., Zadrozny, K.K., Xu, C., Schur, F.K.M., Lyddon, T.D., Ricana, C.L., Wagner, J.M., Perilla, J.R., Ganser-Pornillos, B.K., Johnson, M.C. *et al.* (2018) Inositol phosphates are assembly co-factors for HIV-1. *Nature*, **560**, 509-512.

61. Mallery, D.L., Márquez, C.L., McEwan, W.A., Dickson, C.F., Jacques, D.A., Anandapadamanaban, M., Bichel, K., Towers, G.J., Saiardi, A., Böcking, T. *et al.* (2018) IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *eLife*, **7**, e35335.

62. Forshey, B.M., von Schwedler, U., Sundquist, W.I. and Aiken, C. (2002) Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.*, **76,** 5667-5677.

63. Hulme, A.E., Kelley, Z., Okocha, E.A. and Hope, T.J. (2015) Identification of capsid mutations that alter the rate of HIV-1 uncoating in infected cells. *J. Virol.*, **89**, 643-651.

64. Wacharapornin, P., Lauhakirti, D. and Auewarakul, P. (2007) The effect of capsid mutations on HIV-1 uncoating. *Virology*, **358,** 48-54.

65. Manocheewa, S., Swain, J.V., Lanxon-Cookson, E., Rolland, M. and Mullins, J.I. (2013) Fitness costs of mutations at the HIV-1 capsid hexamerization interface. *PLoS ONE*, **8**, e66065.

66. Dismuke, D.J. and Aiken, C. (2006) Evidence for a functional link between uncoating of the human immunodeficiency virus type 1 core and nuclear import of the viral preintegration complex. *J. Virol.*, **80,** 3712-3720.

67. Yamashita, M. and Emerman, M. (2004) Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. *J. Virol.*, **78,** 5670-5678.

68. Yamashita, M., Perez, O., Hope, T.J. and Emerman, M. (2007) Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells. *PLoS Pathog.*, **3,** 1502-1510.

69. Fassati, A. (2012) Multiple roles of the capsid protein in the early steps of HIV-1 infection. *Virus Res.*, **170**, 15-24.

70. Fassati, A. and Goff, S.P. (2001) Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. *J. Virol.*, **75,** 3626-3635.

71. Miller, M.D., Farnet, C.M. and Bushman, F.D. (1997) Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.*, **71,** 5382-5390.

72. Hulme, A.E., Perez, O. and Hope, T.J. (2011) Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 9975-9980.

73. Mamede, J.I., Cianci, G.C., Anderson, M.R. and Hope, T.J. (2017) Early cytoplasmic uncoating is associated with infectivity of HIV-1. *Proc. Natl. Acad. Sci. U.S.A.,* **114,** E7169-E7178.

74. Lukic, Z., Dharan, A., Fricke, T., Diaz-Griffero, F. and Campbell, E.M. (2014) HIV-1 uncoatingis facilitated by dynein and kinesin 1. *J. Virol.*, **88**, 13613-13625.

75. Xu, H., Franks, T., Gibson, G., Huber, K., Rahm, N., De Castillia, C.S., Luban, J., Aiken, C., Watkins, S., Sluis-Cremer, N. *et al.* (2013) Evidence for biphasic uncoating during HIV-1 infection from a novel imaging assay. *Retrovirology*, **10**, 70.

76. Francis, A.C. and Melikyan, G.B. (2018) Single HIV-1 imaging reveals progression of infection through CA-dependent steps of docking at the nuclear pore, uncoating, and nuclear transport. *Cell Host Microbe*, **23**, 536-548.

77. Burdick, R.C., Delviks-Frankenberry, K.A., Chen, J., Janaka, S.K., Sastri, J., Hu, W.-S. and Pathak, V.K. (2017) Dynamics and regulation of nuclear import and nuclear movements of HIV-1 complexes. *PLoS Pathog.*, **13**, e1006570.

78. Burdick, R.C., Li, C., Munshi, M., Rawson, J.M.O., Nagashima, K., Hu, W.-S. and Pathak, V.K. (2020) HIV-1 uncoats in the nucleus near sites of integration. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 5486-5493.

79. Li, C., Burdick, R.C., Nagashima, K., Hu, W.S. and Pathak, V.K. (2021) HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. *Proc Natl Acad Sci U S A*, **118,** e2019467118.

80. Dharan, A., Bachmann, N., Talley, S., Zwikelmaier, V. and Campbell, E.M. (2020) Nuclear pore blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus. *Nat. Microbiol.*, **5**, 1088-1095.

81. Zila, V., Margiotta, E., Turoňová, B., Müller, T.G., Zimmerli, C.E., Mattei, S., Allegretti, M., Börner, K., Rada, J., Müller, B. *et al.* (2021) Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell*, **184**, 1032-1046.e1018.

82. Selyutina, A., Persaud, M., Lee, K., KewalRamani, V. and Diaz-Griffero, F. (2020) Nuclear import of the HIV-1 core precedes reverse transcription and uncoating. *Cell Rep.*, **32**, 108201.

83. Lin, D.H. and Hoelz, A. (2019) The structure of the nuclear pore complex (an update). *Annu. Rev. Biochem.*, **88**, 725-783.

84. Guedán, A., Donaldson, C.D., Caroe, E.R., Cosnefroy, O., Taylor, I.A. and Bishop, K.N. (2021) HIV-1 requires capsid remodelling at the nuclear pore for nuclear entry and integration. *PLoS Pathog.*, **17**, e1009484.

85. Hulme, A.E., Kelley, Z., Foley, D. and Hope, T.J. (2015) Complementary assays reveal a low level of CA associated withviral complexes in the nuclei of HIV-1-infected cells. *J. Virol.*, **89**, 5350-5361.

86. Peng, K., Muranyi, W., Glass, B., Laketa, V., Yant, S.R., Tsai, L., Cihlar, T., Müller, B. and Kräusslich, H.-G. (2014) Quantitative microscopy of functional HIV post-entry complexes reveals association of replication with the viral capsid. *eLife*, **3**, e04114.

87. Yang, Y., Fricke, T. and Diaz-Griffero, F. (2013) Inhibition of reverse transcriptase activity increases stability of the HIV-1 core. *J. Virol.*, **87**, 683-687.

88. Cosnefroy, O., Murray, P.J. and Bishop, K.N. (2016) HIV-1 capsid uncoating initiates after the first strand transfer of reverse transcription. *Retrovirology*, **13**, 58.

89. Rankovic, S., Varadarajan, J., Ramalho, R., Aiken, C. and Rousso, I. (2017) Reverse transcription mechanically initiates HIV-1 capsid disassembly. *J. Virol.*, **91,** e00289-17.

90. Rankovic, S., Deshpande, A., Harel, S., Aiken, C., Rousso, I. and Kirchhoff, F. (2021) HIV-1 uncoating cccurs via a series of rapid biomechanical changes in the core related to individual stages of reverse transcription. *J. Virol.*, **95**, e00166-21.

91. Christensen, D.E., Ganser-Pornillos, B.K., Johnson, J.S., Pornillos, O. and Sundquist, W.I. (2020) Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro. *Science*, **370**, eabc8420.

92. Jacques, D.A., McEwan, W.A., Hilditch, L., Price, A.J., Towers, G.J. and James, L.C. (2016) HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. *Nature*, **536**, 349-353.

93. Song, G. (2020) Uncovering the release mechanism of nucleotide import by HIV-1 capsid. *Phys. Biol.*, **18**, 016004.

94. Xu, C., Fischer, D.K., Rankovic, S., Li, W., Dick, R.A., Runge, B., Zadorozhnyi, R., Ahn, J., Aiken, C., Polenova, T. *et al.* (2020) Permeability of the HIV-1 capsid to metabolites modulates viral DNA synthesis. *PLoS Biol.*, **18**, e3001015.

95. Renner, N., Mallery, D.L., Faysal, K.M.R., Peng, W., Jacques, D.A., Böcking, T. and James, L.C. (2021) A lysine ring in HIV capsid pores coordinates IP6 to drive mature capsid assembly. *PLoS Pathog.*, **17**, e1009164.

96. Le Sage, V., Mouland, A.J. and Valiente-Echeverría, F. (2014) Roles of HIV-1 capsid in viral replication and immune evasion. *Virus Res.*, **193**, 116-129.

97. Maelfait, J., Seiradake, E. and Rehwinkel, J. (2014) Keeping your armour intact: how HIV-1 evades detection by the innate immune system. *Bioessays*, **36**, 649-657.

98. Lahaye, X., Satoh, T., Gentili, M., Cerboni, S., Conrad, C., Hurbain, I., El Marjou, A., Lacabaratz, C., Lelièvre, J.D. and Manel, N. (2013) The capsids of HIV-1 and HIV-2 determine immune detection of the viral cDNA by the innate sensor cGAS in dendritic cells. *Immunity*, **39**, 1132-1142.

99. Rasaiyaah, J., Tan, C.P., Fletcher, A.J., Price, A.J., Blondeau, C., Hilditch, L., Jacques, D.A., Selwood, D.L., James, L.C., Noursadeghi, M. *et al.* (2013) HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature*, **503**, 402-405.

100. Lahaye, X., Gentili, M., Silvin, A., Conrad, C., Picard, L., Jouve, M., Zueva, E., Maurin, M., Nadalin, F., Knott, G.J. *et al.* (2018) NONO detects the nuclear HIV capsid to promote cGAS-mediated innate immune activation. *Cell*, **175**, 488-501.e22.

101. Sumner, R.P., Harrison, L., Touizer, E., Peacock, T.P., Spencer, M., Zuliani-Alvarez, L. and Towers, G.J. (2020) Disrupting HIV-1 capsid formation causes cGAS sensing of viral DNA. *EMBO J.*, **39**, e103958.

102. Dharan, A., Opp, S., Abdel-Rahim, O., Keceli, S.K., Imam, S., Diaz-Griffero, F. and Campbell, E.M. (2017) Bicaudal D2 facilitates the cytoplasmic trafficking and nuclear import of HIV-1 genomes during infection. *Proc. Natl. Acad. Sci. U.S.A.* **114,** E10707-10716.

103. Carnes, S.K., Zhou, J., Aiken, C. and Sundquist, W.I. (2018) HIV-1 engages a dynein-dynactin-BICD2 Complex for infection and transport to the nucleus. *J. Virol.*, **92**, e00358-18.

104. Mitra, S., Shanmugapriya, S., Silva, E.S.d., Naghavi, M.H. and Kirchhoff, F. (2020) HIV-1 exploits CLASP2 to induce microtubule stabilization and facilitate virus trafficking to the nucleus. *J. Virol.*, **94**, e00404-20.

105. Santos da Silva, E., Shanmugapriya, S., Malikov, V., Gu, F., Delaney, M.K. and Naghavi, M.H. (2020) HIV-1 capsids mimic a microtubule regulator to coordinate early stages of infection. *EMBO J.*, **39**, e104870.

106. Lee, K., Ambrose, Z., Martin, T.D., Oztop, I., Mulky, A., Julias, J.G., Vandegraaff, N., Baumann, J.G., Wang, R., Yuen, W. *et al.* (2010) Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe*, **7,** 221-233.

107. Luban, J., Bossolt, K.L., Franke, E.K., Kalpana, G.V. and Goff, S.P. (1993) Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell*, **73**, 1067-1078.

108. Franke, E.K., Yuan, H.E. and Luban, J. (1994) Specific incorporation of cyclophilin A into HIV-1 virions. *Nature*, **372**, 359-362.

109. Thali, M., Bukovsky, A., Kondo, E., Rosenwirth, B., Walsh, C.T., Sodroski, J. and Göttlinger, H.G. (1994) Functional association of cyclophilin A with HIV-1 virions. *Nature*, **372**, 363-365.

110. Maillet, S., Fernandez, J., Decourcelle, M., El Koulali, K., Blanchet, F.P., Arhel, N.J., Maarifi, G. and Nisole, S. (2020) Daxx inhibits HIV-1 reverse transcription and uncoating in a SUMO-dependent manner. *Viruses*, **12**, 636.

111. Delaney, M.K., Malikov, V., Chai, Q., Zhao, G. and Naghavi, M.H. (2017) Distinct functions of diaphanous-related formins regulate HIV-1 uncoating and transport. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E6932-E6941.

112. Dochi, T., Nakano, T., Inoue, M., Takamune, N., Shoji, S., Sano, K. and Misumi, S. (2014) Phosphorylation of human immunodeficiency virus type 1 capsid protein at serine 16, required for peptidyl-prolyl isomerase-dependent uncoating,

is mediated by virion-incorporated extracellular signal-regulated kinase 2. *J. Gen. Virol.*, **95**, 1156-1166.

113. Malikov, V., da Silva, E.S., Jovasevic, V., Bennett, G., de Souza Aranha Vieira, D.A., Schulte, B., Diaz-Griffero, F., Walsh, D. and Naghavi, M.H. (2015) HIV-1 capsids bind and exploit the kinesin-1 adaptor FEZ1 for inward movement to the nucleus. *Nat. Commun.*, **6**, 6660.

114. Huang, P.T., Summers, B.J., Xu, C., Perilla, J.R., Malikov, V., Naghavi, M.H. and Xiong, Y. (2019) FEZ1 Is recruited to a conserved cofactor site on capsid to promote HIV-1 trafficking. *Cell Rep.*, **28**, 2373-2385.e7.

115. Fernandez, J., Portilho, D.M., Danckaert, A., Munier, S., Becker, A., Roux, P., Zambo, A., Shorte, S., Jacob, Y., Vidalain, P.-O. *et al.* (2015) Microtubule-associated proteins 1 (MAP1) promote human immunodeficiency virus type I (HIV-1) intracytoplasmic routing to the ucleus. J. *Biol. Chem.*, **290**, 4631-4646.

116. Takeuchi, H., Saito, H., Noda, T., Miyamoto, T., Yoshinaga, T., Terahara, K., Ishii, H., Tsunetsugu-Yokota, Y. and Yamaoka, S. (2017) Phosphorylation of the HIV-1 capsid by MELK triggers uncoating to promote viral cDNA synthesis. *PLoS Pathog.*, **13**, e1006441.

117. Liu, Z., Pan, Q., Ding, S., Qian, J., Xu, F., Zhou, J., Cen, S., Guo, F. and Liang, C. (2013) The interferon-inducible MxB protein inhibits HIV-1 infection. *Cell Host Microbe*, **14**, 398-410.

118. Kane, M., Yadav, S.S., Bitzegeio, J., Kutluay, S.B., Zang, T., Wilson, S.J., Schoggins, J.W., Rice, C.M., Yamashita, M., Hatziioannou, T. *et al.* (2013) MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature*, **502**, 563-566.

119. Goujon, C., Moncorgé, O., Bauby, H., Doyle, T., Ward, C.C., Schaller, T., Hué, S., Barclay, W.S., Schulz, R. and Malim, M.H. (2013) Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature*, **502**, 559-562.

120. Fricke, T., White, T.E., Schulte, B., de Souza Aranha Vieira, D.A., Dharan, A., Campbell, E.M., Brandariz-Nuñez, A. and Diaz-Griffero, F. (2014) MxB binds to the HIV-1 core and prevents the uncoating process of HIV-1. *Retrovirology*, **11**, 68.

121. Matreyek, K.A. and Engelman, A. (2011) The requirement for nucleoporin NUP153 during human immunodeficiency virus type 1 infection is determined by the viral capsid. *J. Virol.*, **85**, 7818-7827.

122. Dharan, A., Talley, S., Tripathi, A., Mamede, J.I., Majetschak, M., Hope, T.J. and Campbell, E.M. (2016) KIF5B and Nup358 cooperatively mediate the nuclear import of HIV-1 during infection. *PLoS Pathog.*, **12**, e1005700.

123. Bichel, K., Price, A.J., Schaller, T., Towers, G.J., Freund, S.M. and James, L.C. (2013) HIV-1 capsid undergoes coupled binding and isomerization by the nuclear pore protein NUP358. *Retrovirology*, **10**, 81.

124. Di Nunzio, F., Danckaert, A., Fricke, T., Perez, P., Fernandez, J., Perret, E., Roux, P., Shorte, S., Charneau, P., Diaz-Griffero, F. *et al.* (2012) Human Nucleoporins Promote HIV-1 Docking at the Nuclear Pore, Nuclear Import and Integration. *PLoS ONE*, **7**, e46037.

125. Henning, M.S., Morham, S.G., Goff, S.P. and Naghavi, M.H. (2010) PDZD8 is a novel Gag-interacting factor that promotes retroviral infection. *J. Virol.*, **84**, 8990-8995.

126. Guth, C.A. and Sodroski, J. (2014) Contribution of PDZD8 to stabilization of the human immunodeficiency virus type 1 capsid. *J. Virol.*, **88**, 4612-4623.

127. Zhang, S. and Sodroski, J. (2015) Efficient human immunodeficiency virus (HIV-1) infection of cells lacking PDZD8. *Virology*, **481**, 73-78.

128. Misumi, S., Inoue, M., Dochi, T., Kishimoto, N., Hasegawa, N., Takamune, N. and Shoji, S. (2010) Uncoating of human immunodeficiency virus type 1 requires prolyl isomerase Pin1. *J. Biol. Chem.*, **285,** 25185-15195.

129. Marno, K.M., Ogunkolade, B.W., Pade, C., Oliveira, N.M.M., O'Sullivan, E. and McKnight, Á. (2014) Novel restriction factor RNA-associated early-stage anti-viral factor (REAF) inhibits human and simian immunodeficiency viruses. *Retrovirology*, **11**, 3.

130. Marno, K.M., O'Sullivan, E., Jones, C.E., Díaz-Delfín, J., Pardieu, C., Sloan, R.D. and McKnight, Á. (2017) RNA-associated early-stage antiviral factor is a major component of Lv2 restriction. *J. Virol.*, **91**, e01228-16.

131. Rebensburg, S.V., Wei, G., Larue, R.C., Lindenberger, J., Francis, A.C., Annamalai, A.S., Morrison, J., Shkriabai, N., Huang, S.-W., KewalRamani, V. *et al.* (2021) Sec24C is an HIV-1 host dependency factor crucial for virus replication. *Nat. Microbiol.*, **6**, 435-444.

132. Schaller, T., Bulli, L., Pollpeter, D., Betancor, G., Kutzner, J., Apolonia, L., Herold, N., Burk, R. and Malim, M.H. (2017) Effects of inner nuclear membrane proteins SUN1/UNC-84A and SUN2/UNC-84B on the early steps of HIV-1 infection. *J. Virol.*, **91,** e00483-17.

133. Fernandez, J., Machado, A.K., Lyonnais, S., Chamontin, C., Gärtner, K., Léger, T., Henriquet, C., Garcia, C., Portilho, D.M., Pugnière, M. *et al.* (2019) Transportin-1 binds to the HIV-1 capsid via a nuclear localization signal and triggers uncoating. *Nat. Microbiol.*, **4**, 1840-1850.

134. Krishnan, L., Matreyek, K.A., Oztop, I., Lee, K., Tipper, C.H., Li, X., Dar, M.J., Kewalramani, V.N. and Engelman, A. (2010) The requirement for cellular transportin 3 (TNPO3 or TRN-SR2) during infection maps to human immunodeficiency virus type 1 capsid and not integrase. *J. Virol.*, **84,** 397-406.

135. De Iaco, A., Santoni, F., Vannier, A., Guipponi, M., Antonarakis, S. and Luban, J. (2013) TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm. *Retrovirology*, **10**, 20.

136. Stremlau, M., Owens, C.M., Perron, M.J., Kiessling, M., Autissier, P. and Sodroski, J. (2004) The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature*, **427**, 848-853.

137. Nisole, S., Lynch, C., Stoye, J.P. and Yap, M.W. (2004) A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1. *Proc. Natl. Acad. Sci.U.S.A.*, **101**, 13324-13328.

138. Zhuang, S. and Torbett, B.E. (2021) Interactions of HIV-1 capsid with host factors and their implications for developing novel therapeutics. *Viruses*, **13**, 417.

139. Gamble, T.R., Vajdos, F.F., Yoo, S., Worthylake, D.K., Houseweart, M., Sundquist, W.I. and Hill, C.P. (1996) Crystal structure of humancyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell*, **87**, 1285-1294.

140. Cortines, J.R., Lima, L.M.T.R., Mohana-Borges, R., Millen, T.d.A., Gaspar, L.P., Lanman, J.K., Prevelige, P.E. and Silva, J.L. (2015) Structural insights into the stabilization of the human immunodeficiency virus type 1 capsid protein by the cyclophilin-binding domain and implications on the virus cycle. *Biochim. Biophys. Acta*, **1854**, 341-348.

141. Liu, C., Perilla, J.R., Ning, J., Lu, M., Hou, G., Ramalho, R., Himes, B.A., Zhao, G., Bedwell, G.J., Byeon, I.-J. *et al.* (2016) Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. *Nat. Commun.*, **7**, 10714.

142. Bosco, D.A., Eisenmesser, E.Z., Pochapsky, S., Sundquist, W.I. and Kern, D. (2002) Catalysis of cis/trans isomerization in native HIV-1 capsid by human cyclophilin A. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 5247-5252.

143. Hatziioannou, T., Perez-Caballero, D., Cowan, S. and Bieniasz, P.D. (2005) Cyclophilin interactions with incoming human immunodeficiency virus type 1 capsids with opposing effects on infectivity in human cells. *J. Virol.*, **79**, 176-183.

144. Sokolskaja, E., Sayah, D.M. and Luban, J. (2004) Target cell cyclophilin A modulates human immunodeficiency virus type 1 infectivity. *J. Virol.*, **78**, 12800-12808.

145. Shah, V.B., Shi, J., Hout, D.R., Oztop, I., Krishnan, L., Ahn, J., Shotwell, M.S., Engelman, A. and Aiken, C. (2013) The host proteins transportin SR2/TNPO3 and cyclophilin A exert opposing effects on HIV-1 uncoating. *J. Virol.*, **87**, 422-432.

146. De Iaco, A. and Luban, J. (2014) Cyclophilin A promotes HIV-1 reverse transcription but its effect on transduction correlates best with its effect on nuclear entry of viral cDNA. *Retrovirology*, **11**, 11.

147. Li, Y., Kar, A.K. and Sodroski, J. (2009) Target cell type-dependent modulation of human immunodeficiency virus type 1 capsid disassembly by cyclophilin A. *J. Virol.*, **83,** 10951-10962.

148. Schaller, T., Ocwieja, K.E., Rasaiyaah, J., Price, A.J., Brady, T.L., Roth, S.L., Hué, S., Fletcher, A.J., Lee, K., KewalRamani, V.N. *et al.* (2011) HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathog.*, **7**, e1002439.

149. Fricke, T., Brandariz-Nuñez, A., Wang, X., Smith, A.B., 3rd and Diaz-Griffero, F. (2013) Human cytosolic extracts stabilize the HIV-1 core. *J. Virol.*, **87**, 10587-10597.

150. Burse, M., Shi, J. and Aiken, C. (2017) Cyclophilin A potentiates TRIM5α inhibition of HIV-1 nuclear import without promoting TRIM5α binding to the viral capsid. *PLoS ONE*, **12**, e0182298.

151. Liu, Z., Pan, Q., Liang, Z., Qiao, W., Cen, S. and Liang, C. (2015) The highly polymorphic cyclophilin A-binding loop in HIV-1 capsid modulates viral resistance to MxB. *Retrovirology*, **12**, 1.

152. Sabo, Y., Walsh, D., Barry, D.S., Tinaztepe, S., de Los Santos, K., Goff, S.P., Gundersen, G.G. and Naghavi, M.H. (2013) HIV-1 induces the formation of stable microtubules to enhance early infection. *Cell Host Microbe*, **14**, 535-546.

153. Matreyek, K.A., Yücel, S.S., Li, X. and Engelman, A. (2013) Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. *PLoS Pathog.*, **9**, e1003693.

154. Chin, Christopher R., Perreira, Jill M., Savidis, G., Portmann, Jocelyn M., Aker, Aaron M., Feeley, Eric M., Smith, Miles C. and Brass, Abraham L. (2015) Direct visualization of HIV-1 replication intermediates shows that capsid and CPSF6 modulate HIV-1 intra-nuclear invasion and integration. *Cell Rep.*, **13**, 1717-1731.

155. Buffone, C., Martinez-Lopez, A., Fricke, T., Opp, S., Severgnini, M., Cifola, I., Petiti, L., Frabetti, S., Skorupka, K., Zadrozny, K.K. *et al.* (2018) Nup153 unlocks the nuclear pore complex for HIV-1 nuclear translocation in nondividing cells. *J. Virol.*, **92,** e00648-18 .

156. Sowd, G.A., Serrao, E., Wang, H., Wang, W., Fadel, H.J., Poeschla, E.M. and Engelman, A.N. (2016) A critical role for alternative polyadenylation factor CPSF6 in targeting HIV-1 integration to transcriptionally active chromatin. *Proc. Natl.l Acad. Sci. U.S.A.*, **113**, E1054-E1063.

157. Price, A.J., Jacques, D.A., McEwan, W.A., Fletcher, A.J., Essig, S., Chin, J.W., Halambage, U.D., Aiken, C. and James, L.C. (2014) Host cofactors and pharmacologic ligands share an essential interface in HIV-1 capsid that is lost upon disassembly. *PLoS Pathog.*, **10**, e1004459.

158. Bhattacharya, A., Alam, S.L., Fricke, T., Zadrozny, K., Sedzicki, J., Taylor, A.B., Demeler, B., Pornillos, O., Ganser-Pornillos, B.K., Diaz-Griffero, F. *et al.* (2014)

Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. *Proc. Natl. Acad. Sci. U.S.A.,* **111**, 18625-18630.

159. Saito, A., Henning, M.S., Serrao, E., Dubose, B.N., Teng, S., Huang, J., Li, X., Saito, N., Roy, S.P., Siddiqui, M.A. *et al.* (2016) Capsid-CPSF6 interaction is dispensable for HIV-1 replication in primary cells but is selected during virus passage in vivo. *J. Virol.*, **90**, 6918-6935.

160. De Iaco, A. and Luban, J. (2011) Inhibition of HIV-1 infection by TNPO3 depletion is determined by capsid and detectable after viral cDNA enters the nucleus. *Retrovirology*, **8**, 98.

161. Valle-Casuso, J.C., Di Nunzio, F., Yang, Y., Reszka, N., Lienlaf, M., Arhel, N., Perez, P., Brass, A.L. and Diaz-Griffero, F. (2012) TNPO3 is required for HIV-1 replication after nuclear import but prior to integration and binds the HIV-1 core. *J. Virol.*, **86**, 5931-5936.

162. Zhou, L., Sokolskaja, E., Jolly, C., James, W., Cowley, S.A. and Fassati, A. (2011) Transportin 3 promotes a nuclear maturation step required for efficient HIV-1 integration. *PLoS Pathog.*, **7**, e1002194.

163. Fricke, T., Valle-Casuso, J.C., White, T.E., Brandariz-Nuñez, A., Bosche, W.J., Reszka, N., Gorelick, R. and Diaz-Griffero, F. (2013) The ability of TNPO3-depleted cells to inhibit HIV-1 infection requires CPSF6. *Retrovirology*, **10**, 46.

164. Maertens, G.N., Cook, N.J., Wang, W., Hare, S., Gupta, S.S., Öztop, I., Lee, K., Pye, V.E., Cosnefroy, O., Snijders, A.P. *et al.* (2014) Structural basis for nuclear import of splicing factors by human Transportin 3. *Proc. Natl. Acad. Sci. U.S.A.*,, **111**, 2728-2733.

165. Fribourgh, J.L., Nguyen, H.C., Matreyek, K.A., Alvarez, F.J.D., Summers, B.J., Dewdney, T.G., Aiken, C., Zhang, P., Engelman, A. and Xiong, Y. (2014) Structural insight into HIV-1 restriction by MxB. *Cell Host Microbe*, **16**, 627-638.

166. Kong, J., Xu, B., Wei, W., Wang, X., Xie, W. and Yu, X.-F. (2014) Characterization of the amino-terminal domain of Mx2/MxB-dependent interaction with the HIV-1 capsid. *Protein Cell*, **5**, 954-957.

167. Smaga, S.S., Xu, C., Summers, B.J., Digianantonio, K.M., Perilla, J.R. and Xiong, Y. (2019) MxB restricts HIV-1 by targeting the tri-hexamer interface of the viral capsid. *Structure*, **27**, 1234-1245.e5.

168. Buffone, C., Schulte, B., Opp, S. and Diaz-Griffero, F. (2015) Contribution of MxB oligomerization to HIV-1 capsid binding and restriction. *J. Virol.*, **89**, 3285-3294.

169. Dicks, M.D.J., Goujon, C., Pollpeter, D., Betancor, G., Apolonia, L., Bergeron, J.R.C. and Malim, M.H. (2016) Oligomerization requirements for MX2-mediated suppression of HIV-1 infection. *J. Virol.*, **90**, 22-32.

170. Xie, L., Chen, L., Zhong, C., Yu, T., Ju, Z., Wang, M., Xiong, H., Zeng, Y., Wang, J., Hu, H. *et al.* (2020) MxB impedes the NUP358-mediated HIV-1 pre-integration complex nuclear import and viral replication cooperatively with CPSF6. *Retrovirology*, **17**, 16.

171. Stremlau, M., Owens, C.M., Perron, M.J., Kiessling, M., Autissier, P. and Sodroski, J. (2004) The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature*, **427,** 848-853.

172. Morger, D., Zosel, F., Bühlmann, M., Züger, S., Mittelviefhaus, M., Schuler, B., Luban, J., Grütter, M.G. and Kirchhoff, F. (2018) The three-fold axis of the HIV-1 capsid lattice is the species-specific binding interface for TRIM5α. *J. Virol.*, **92**, e01541-17.

173. Skorupka, K.A., Roganowicz, M.D., Christensen, D.E., Wan, Y., Pornillos, O. and Ganser-Pornillos, B.K. (2019) Hierarchical assembly governs TRIM5alpha recognition of HIV-1 and retroviral capsids. *Sci. Adv.*, **5**, eaaw3631.

174. Yu, A., Skorupka, K.A., Pak, A.J., Ganser-Pornillos, B.K., Pornillos, O. and Voth, G.A. (2020) TRIM5α self-assembly and compartmentalization of the HIV-1 viral capsid. *Nat. Commun.*, **11**, 1307.

175. Yudina, Z., Roa, A., Johnson, R., Biris, N., de Souza Aranha Vieira, D.A., Tsiperson, V., Reszka, N., Taylor, A.B., Hart, P.J., Demeler, B. *et al.* (2015) RING dimerization links higher-order assembly of TRIM5α to synthesis of K63-linked polyubiquitin. *Cell Rep.*, **12**, 788-797.

176. Ganser-Pornillos, B.K., Chandrasekaran, V., Pornillos, O., Sodroski, J.G., Sundquist, W.I. and Yeager, M. (2011) Hexagonal assembly of a restricting TRIM5α protein. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 534-539.

177. Li, Y.-L., Chandrasekaran, V., Carter, S.D., Woodward, C.L., Christensen, D.E., Dryden, K.A., Pornillos, O., Yeager, M., Ganser-Pornillos, B.K., Jensen, G.J. *et al.* (2016) Primate TRIM5 proteins form hexagonal nets on HIV-1 capsids. *eLife*, **5**, e16269.

178. Wagner, J.M., Christensen, D.E., Bhattacharya, A., Dawidziak, D.M., Roganowicz, M.D., Wan, Y., Pumroy, R.A., Demeler, B., Ivanov, D.N., Ganser-

Pornillos, B.K. *et al.* (2018) General model for retroviral capsid pattern recognition by TRIM5 proteins. *J. Virol.*, **92,** e01563-17.

179. Roganowicz, M.D., Komurlu, S., Mukherjee, S., Plewka, J., Alam, S.L., Skorupka, K.A., Wan, Y., Dawidowski, D., Cafiso, D.S., Ganser-Pornillos, B.K. *et al.* (2017) TRIM5alpha SPRY/coiled-coil interactions optimize avid retroviral capsid recognition. *PLoS Pathog.*, **13**, e1006686.

180. Ganser-Pornillos, B.K. and Pornillos, O. (2019) Restriction of HIV-1 and other retroviruses by TRIM5. *Nat. Rev. Microbiol.*, **17**, 546-556.

181. Kutluay, S.B., Perez-Caballero, D. and Bieniasz, P.D. (2013) Fates of retroviral core components during unrestricted and TRIM5-restricted infection. *PLoS Pathog.*, **9**, e1003214.

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

183. Summers, B.J., Digianantonio, K.M., Smaga, S.S., Huang, P.T., Zhou, K., Gerber, E.E., Wang, W. and Xiong, Y. (2019) Modular HIV-1 capsid assemblies reveal diverse host-capsid recognition mechanisms. *Cell Host Microbe*, **26**, 203-216.e6.

184. McFadden, W.M., Snyder, A.A., Kirby, K.A., Tedbury, P.R., Raj, M., Wang, Z. and Sarafianos, S.G. (2021) Rotten to the core: antivirals targeting the HIV-1 capsid core. *Retrovirology*, **18**, 41.

185. Blair, W.S., Pickford, C., Irving, S.L., Brown, D.G., Anderson, M., Bazin, R., Cao, J., Ciaramella, G., Isaacson, J., Jackson, L. *et al.* (2010) HIV capsid is a

tractable target for small molecule therapeutic intervention. *PLoS Pathog.*, **6**, e1001220.

186. Yant, S.R., Mulato, A., Hansen, D., Tse, W.C., Niedziela-Majka, A., Zhang, J.R., Stepan, G.J., Jin, D., Wong, M.H., Perreira, J.M. *et al.* (2019) A highly potent long-acting small-molecule HIV-1 capsid inhibitor with efficacy in a humanized mouse model. *Nat. Med.*, **25**, 1377-1384.

187. Link, J.O., Rhee, M.S., Tse, W.C., Zheng, J., Somoza, J.R., Rowe, W., Begley, R., Chiu, A., Mulato, A., Hansen, D. *et al.* (2020) Clinical targeting of HIV capsid protein with a long-acting small molecule. *Nature*, **584**, 614-618.

188. Tang, C., Loeliger, E., Kinde, I., Kyere, S., Mayo, K., Barklis, E., Sun, Y., Huang, M. and Summers, M.F. (2003) Antiviral inhibition of the HIV-1 capsid protein. *J. Mol. Biol.*, **327**, 1013-1020.

189. Sticht, J., Humbert, M., Findlow, S., Bodem, J., Müller, B., Dietrich, U., Werner, J. and Kräusslich, H.-G. (2005) A peptide inhibitor of HIV-1 assembly in vitro. *Nat. Struct. Mol. Biol.*, **12**, 671-677.

190. Lamorte, L., Titolo, S., Lemke, C.T., Goudreau, N., Mercier, J.F., Wardrop, E., Shah, V.B., von Schwedler, U.K., Langelier, C., Banik, S.S. *et al.* (2013) Discovery of novel small-molecule HIV-1 replication inhibitors that stabilize capsid complexes. *Antimicrob. Agents Chemother.*, **57**, 4622-4631.

191. Vozzolo, L., Loh, B., Gane, P.J., Tribak, M., Zhou, L., Anderson, I., Nyakatura, E., Jenner, R.G., Selwood, D. and Fassati, A. (2010) Gyrase B inhibitor impairs HIV-1 replication by targeting Hsp90 and the capsid protein. *J. Biol. Chem.*, **285**, 39314-39328.

192. Chen, N.Y., Zhou, L., Gane, P.J., Opp, S., Ball, N.J., Nicastro, G., Zufferey, M., Buffone, C., Luban, J., Selwood, D. *et al.* (2016) HIV-1 capsid is involved in post-nuclear entry steps. *Retrovirology*, **13**, 28.

193. Keefe, A.D., Pai, S. and Ellington, A. (2010) Aptamers as therapeutics. *Nat. Rev. Drug Discov.*, **9**, 537-550.

194. Burnett, John C. and Rossi, John J. (2012) RNA-based therapeutics: current progress and future prospects. *Chem. Biol.*, **19**, 60-71.

195. Shigdar, S., Macdonald, J., Connor, M., Wang, T., Xiang, D., Al.Shamaileh, H., Qiao, L., Wei, M., Zhou, S.-F., Zhu, Y. *et al.* (2013) Aptamers as theranostic agents: modifications, serum stability and functionalisation. *Sensors*, **13**, 13624-13637.

196. Zhou, W., Huang, P.J., Ding, J. and Liu, J. (2014) Aptamer-based biosensors for biomedical diagnostics. *Analyst*, **139**, 2627-2640.

197. Zhou, J. and Rossi, J. (2016) Aptamers as targeted therapeutics: current potential and challenges. *Nat. Rev. Drug Discov.*, **16,** 430.

198. Tawiah, K.D., Porciani, D. and Burke, D.H. (2017) Toward theselection of cell targeting aptamers with extended biological functionalities to facilitate endosomal escape of cargoes. *Biomedicines*, **5,** 51.

199. Ning, Y., Hu, J. and Lu, F. (2020) Aptamers used for biosensors and targeted therapy. *Biomed. Pharmacother.*, **132**, 110902.

200. Huang, J., Chen, X., Fu, X., Li, Z., Huang, Y. and Liang, C. (2021) Advances in aptamer-based biomarker discovery. *Front. Cell Dev. Biol.*, **9,** 659760.

201. González, V.M., Martín, M.E., Fernández, G. and García-Sacristán, A. (2016) Use of aptamers as diagnostics tools and antiviral agents for human viruses. *Pharmaceuticals*, **9**, 78.

202. Bouvier-Müller, A. and Ducongé, F. (2018) Application of aptamers for in vivo molecular imaging and theranostics. *Adv. Drug Deliv. Rev.*, **134**, 94-106.

203. Chen, L., Rashid, F., Shah, A., Awan, H.M., Wu, M., Liu, A., Wang, J., Zhu, T., Luo, Z. and Shan, G. (2015) The isolation of an RNA aptamer targeting to p53 protein with single amino acid mutation. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 10002-10007.

204. Alam, K.K., Chang, J.L., Lange, M.J., Nguyen, P.D.M., Sawyer, A.W. and Burke, D.H. (2018) Poly-target selection identifies broad-spectrum RNA aptamers. *Mol. Ther. Nucleic Acids*, **13**, 605-619.

205. Hermann, T. and Patel, D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820-825.

206. Zichel, R., Chearwae, W., Pandey, G.S., Golding, B. and Sauna, Z.E. (2012) Aptamers as a sensitive tool to detect subtle modifications in therapeutic proteins. *PLoS ONE*, **7**, e31948.

207. Moore, M.D., Bobay, B.G., Mertens, B., Jaykus, L.-A. and Coyne, C.B. (2016) Human norovirus aptamer exhibits high degree of target conformation-dependent-binding similar to that of receptors and discriminates particle functionality. *mSphere*, **1**, e00298-16.

208. Kahsai, A.W., Wisler, J.W., Lee, J., Ahn, S., Cahill Iii, T.J., Dennison, S.M., Staus, D.P., Thomsen, A.R., Anasti, K.M., Pani, B. *et al.* (2016) Conformationally

selective RNA aptamers allosterically modulate the β2-adrenoceptor. *Nat. Chem. Biol.*, **12**, 709-716.

209. Weber, A.M., Kaiser, J., Ziegler, T., Pilsl, S., Renzl, C., Sixt, L., Pietruschka, G., Moniot, S., Kakoti, A., Juraschitz, M. *et al.* (2019) A blue light receptor that mediates RNA binding and translational regulation. *Nat. Chem. Biol.*, **15**, 1085-1092.

210. Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818-822.

211. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.

212. Meek, K.N., Rangel, A.E. and Heemstra, J.M. (2016) Enhancing aptamer function and stability via in vitro selection using modified nucleic acids. *Methods*, **106**, 29-36.

213. Dellafiore, M.A., Montserrat, J.M. and Iribarren, A.M. (2016) Modified nucleoside triphosphates for in-vitro selection techniques. *Front Chem*, **4**, 18.

214. Darmostuk, M., Rimpelova, S., Gbelcova, H. and Ruml, T. (2015) Current approaches in SELEX: an update to aptamer selection technology. *Biotechnol. Adv.*, **33**, 1141-1161.

215. Keefe, A.D. and Cload, S.T. (2008) SELEX with modified nucleotides. *Curr. Opin. Chem. Biol.*, **12**, 448-456.

216. Kimoto, M., Meyer, A.J., Hirao, I. and Ellington, A.D. (2017) Genetic alphabet expansion transcription generating functional RNA molecules containing a five-

letter alphabet including modified unnatural and natural base nucleotides by thermostable T7 RNA polymerase variants. *ChemComm*, **53**, 12309-12312.

217. Lauridsen, L.H., Rothnagel, J.A. and Veedu, R.N. (2012) Enzymatic recognition of 2'-modified ribonucleoside 5'-triphosphates: towards the evolution of versatile aptamers. *ChemBioChem*, **13**, 19-25.

218. Pinheiro, V.B., Taylor, A.I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J.C., Wengel, J., Peak-Chew, S.-Y., McLaughlin, S.H. *et al.* (2012) Synthetic genetic polymers capable of heredity and evolution. *Science*, **336**, 341-344.

219. Eaton, B.E., Gold, L., Hicke, B.J., Janjié, N., Jucker, F.M., Sebesta, D.P., Tarasow, T.M., Willis, M.C. and Zichi, D.A. (1997) Post-SELEX combinatorial optimization of aptamers. *Bioorg. Med. Chem.*, **5**, 1087-1096.

220. Lipi, F., Chen, S., Chakravarthy, M., Rakesh, S. and Veedu, R.N. (2016) In vitro evolution of chemically-modified nucleic acid aptamers: pros and cons, and comprehensive selection strategies. *RNA Biol.*, **13**, 1232-1245.

221. Hamada, M. (2018) In silico approaches to RNA aptamer design. *Biochimie*, **145**, 8-14.

222. Takahashi, M., Wu, X., Ho, M., Chomchan, P., Rossi, J.J., Burnett, J.C. and Zhou, J. (2016) High throughput sequencing analysis of RNA libraries reveals the influences of initial library and PCR methods on SELEX efficiency. *Sci. Rep.*, **6**, 33697.

223. Gotrik, M.R., Feagin, T.A., Csordas, A.T., Nakamoto, M.A. and Soh, H.T. (2016) Advancements in aptamer discovery technologies. *Acc. Chem. Res.*, **49**, 1903-1910.

224. Berezhnoy, A., Stewart, C.A., McNamara Ii, J.O., Thiel, W., Giangrande, P., Trinchieri, G. and Gilboa, E. (2012) Isolation and optimization of murine IL-10 receptor blocking oligonucleotide aptamers using high-throughput sequencing. *Mol. Ther.*, **20**, 1242-1250.

225. Thiel, W.H., Bair, T., Peek, A.S., Liu, X., Dassie, J., Stockdale, K.R., Behlke, M.A., Miller, F.J., Jr. and Giangrande, P.H. (2012) Rapid identification of cell-specific, internalizing RNA aptamers with bioinformatics analyses of a cell-based aptamer selection. *PLoS ONE*, **7**, e43836.

226. Valenzano, S., De Girolamo, A., DeRosa, M.C., McKeague, M., Schena, R., Catucci, L. and Pascale, M. (2016) Screening and identification of DNA aptamers to tyramine using in vitro selection and high-throughput sequencing. *ACS Comb. Sci.*, **18**, 302-313.

227. Blind, M. and Blank, M. (2015) Aptamer selection technology and recent advances. *Mol. Ther. Nucleic Acids*, **4**, e223.

228. Komarova, N., Barkova, D. and Kuznetsov, A. (2020) Implementation of high-throughput sequencing (HTS) in aptamer selection technology. *Int. J. Mol. Sci.*, **21**, 8774.

229. Jijakli, K., Khraiwesh, B., Fu, W., Luo, L., Alzahmi, A., Koussa, J., Chaiboonchoe, A., Kirmizialtin, S., Yen, L. and Salehi-Ashtiani, K. (2016) The in vitro selection world. *Methods*, **106**, 3-13.

230. Kinghorn, A.B., Fraser, L.A., Lang, S., Shiu, S.C.-C. and Tanner, J.A. (2017) Aptamer bioinformatics. *Int. J. Mol. Sci.*, **18**, 2516.

231. Zimmermann, B., Gesell, T., Chen, D., Lorenz, C. and Schroeder, R. (2010) Monitoring genomic sequences during SELEX using high-throughput sequencing: neutral SELEX. *PLoS ONE*, **5**, e9169.

232. Ditzler, M.A., Lange, M.J., Bose, D., Bottoms, C.A., Virkler, K.F., Sawyer, A.W., Whatley, A.S., Spollen, W., Givan, S.A. and Burke, D.H. (2013) High-throughput sequence analysis reveals structural diversity and improved potency among RNA inhibitors of HIV reverse transcriptase. *Nucleic Acids Res.*, **41**, 1873-1884.

233. Dupont, D.M., Larsen, N., Jensen, J.K., Andreasen, P.A. and Kjems, J. (2015) Characterisation of aptamer–target interactions by branched selection and high-throughput sequencing of SELEX pools. *Nucleic Acids Res.*, **43**, e139.

234. Spiga, F.M., Maietta, P. and Guiducci, C. (2015) More DNA–aptamers for small drugs: a capture–SELEX coupled with surface plasmon resonance and high-throughput sequencing. *ACS Comb. Sci.*, **17**, 326-333.

235. Nguyen Quang, N., Bouvier, C., Henriques, A., Lelandais, B. and Ducongé, F. (2018) Time-lapse imaging of molecular evolution by high-throughput sequencing. *Nucleic Acids Res.*, **46**, 7480-7494.

236. Levay, A., Brenneman, R., Hoinka, J., Sant, D., Cardone, M., Trinchieri, G., Przytycka, T.M. and Berezhnoy, A. (2015) Identifying high-affinity aptamer ligands with defined cross-reactivity using high-throughput guided systematic evolution of ligands by exponential enrichment. *Nucleic Acids Res.*, **43**,e82.

237. Nguyen Quang, N., Perret, G. and Ducongé, F. (2016) Applications of high-throughput sequencing for in vitro selection and characterization of aptamers. *Pharmaceuticals*, **9,** 76 .

238. Thiel, W.H. and Giangrande, P.H. (2016) Analyzing HT-SELEX data with the Galaxy Project tools--a web based bioinformatics platform for biomedical research. *Methods*, **97**, 3-10.

239. Alam, K.K., Chang, J.L. and Burke, D.H. (2015) FASTAptamer: a bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Mol. Ther. Nucleic Acids*, **4**, e230.

240. Cho, M., Xiao, Y., Nie, J., Stewart, R., Csordas, A.T., Oh, S.S., Thomson, J.A. and Soh, H.T. (2010) Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 15373-15378.

241. Schütze, T., Wilhelm, B., Greiner, N., Braun, H., Peter, F., Mörl, M., Erdmann, V.A., Lehrach, H., Konthur, Z., Menger, M. *et al.* (2011) Probing the SELEX process with next-generation sequencing. *PLoS ONE*, **6**, e29604.

242. Thiel, W.H. (2016) Galaxy workflows for web-based bioinformatics analysis of aptamer high-throughput sequencing data. *Mol. Ther. Nucleic Acids*, **5**, e345.

243. Hoinka, J., Berezhnoy, A., Sauna, Z.E., Gilboa, E. and Przytycka, T.M. (2014) AptaCluster - a method to cluster HT-SELEX aptamer pools and lessons from its application. *Res. Comput. Mol. Biol.*, **8394**, 115-128.

244. Kato, S., Ono, T., Minagawa, H., Horii, K., Shiratori, I., Waga, I., Ito, K. and Aoki, T. (2020) FSBC: fast string-based clustering for HT-SELEX data. *BMC Bioinform.*, **21**, 263-263.

245. Hoinka, J., Zotenko, E., Friedman, A., Sauna, Z.E. and Przytycka, T.M. (2012) Identification of sequence-structure RNA binding motifs for SELEX-derived aptamers. *Bioinformatics*, **28**, i215-i223.

246. Hoinka, J., Berezhnoy, A., Dao, P., Sauna, Z.E., Gilboa, E. and Przytycka, T.M. (2015) Large scale analysis of the mutational landscape in HT-SELEX improves aptamer discovery. *Nucleic Acids Res.*, **43**, 5699-5707.

247. Dao, P., Hoinka, J., Takahashi, M., Zhou, J., Ho, M., Wang, Y., Costa, F., Rossi, J.J., Backofen, R., Burnett, J. *et al.* (2016) AptaTRACE elucidates RNA sequence-structure motifs from selection trends in HT-SELEX experiments. *Cell Syst.*, **3**, 62-70.

248. Caroli, J., Taccioli, C., De La Fuente, A., Serafini, P. and Bicciato, S. (2015) APTANI: a computational tool to select aptamers through sequence-structure motif analysis of HT-SELEX data. *Bioinformatics*, **32**, 161-164.

249. Shieh, K.R., Kratschmer, C., Maier, K.E., Greally, J.M., Levy, M. and Golden, A. (2020) AptCompare: optimized de novo motif discovery of RNA aptamers via HTS-SELEX. *Bioinformatics*, **36**, 2905-2906.

250. Gruenke, P.R., Aneja, R., Welbourn, S., Ukah, O.B., Sarafianos, S.G., Burke, D.H. and Lange, M.J. (2022) Selection and identification of an RNA aptamer that specifically binds the HIV-1 capsid lattice and inhibits viral replication. *Nucleic Acids Res.*, **50**, 1701-1717.

251. Kramer, S.T., Gruenke, P.R., Alam, K.K., Xu D. and Burke, D.H. (2022) FASTAptameR 2.0: a web tool for combinatorial sequence selections. bioRxiv doi: https://doi.org/10.1101/2022.04.27.489774, 29 April 2022, pre-print: not peer-reviewed.

252. Gruenke, P.R., Alam, K.K., Singh, K. and Burke, D.H. (2020) 2′-fluoro-modified pyrimidines enhance affinity of RNA oligonucleotides to HIV-1 reverse transcriptase. *RNA*, **26**, 1667-1679.

CHAPTER 2: SELECTION AND IDENTIFICATION OF AN RNA APTAMER THAT SPECIFICALLY BINDS THE HIV-1 CAPSID LATTICE AND INHIBITS VIRAL REPLICATION

As found in: Paige R. Gruenke*, Rachna Aneja*, Sarah Welbourn, Obiaara B. Ukah, Stefan G. Sarafianos, Donald H. Burke, and Margaret J. Lange. Selection and identification of an RNA aptamer that specifically binds the HIV-1 capsid lattice and inhibits viral replication. *Nucleic Acids Res.* 2022; 50: 1701-1717. (*These authors contributed equally to the work.)

2.1. Abstract

The HIV-1 capsid core participates in several replication processes. The mature capsid core is a lattice composed of capsid (CA) monomers thought to assemble first into CA dimers, then into ~250 CA hexamers and 12 CA pentamers. CA assembly requires conformational flexibility of each unit, resulting in the presence of unique, solvent-accessible surfaces. Significant advances have improved our understanding of the roles of the capsid core in replication; however, the contributions of individual CA assembly forms remain unclear and there are limited tools available to evaluate these forms *in vivo*. Here, we have selected aptamers that bind CA lattice tubes. We describe aptamer CA15-2, which selectively binds CA lattice, but not CA monomer or CA hexamer, suggesting that it targets an interface present and accessible only on CA lattice. CA15-2 does not compete with PF74 for binding, indicating that it likely binds a non-overlapping site. Furthermore, CA15-2 inhibits HIV-1 replication when expressed in virus producer cells, but not target cells, suggesting that it binds a biologically-relevant site during virus production that is either not accessible during post-entry replication steps or is accessible but unaltered by aptamer binding. Importantly, CA15-2 represents the first aptamer that specifically recognizes the HIV-1 CA lattice.

2.2. Introduction

The HIV-1 capsid (CA) protein plays an indispensable role in both early and late stages of viral replication, performing multifaceted functions that extend beyond housing the viral RNA and proteins required for replication (1-8). The capsid core facilitates reverse transcription and formation of the pre-integration complex (2,3,9-11), enables transport through the cytosol to the nucleus via interactions with a variety of host factors (3,12-15), facilitates nuclear import (1,3,8,16-20), delivers the pre-integration complex for integration into the host chromosome (18-25), and enables HIV-1 to evade host innate immune surveillance during replication (6,26-28). CA also plays a significant role in viral assembly and maturation (8,11,29,30). As disruption of any of these events can significantly impact viral replication, the viral capsid core is an important target both for better understanding virus biology and developing new therapeutics to combat the ongoing AIDS pandemic.

Prior work has highlighted the dynamic nature of the viral capsid core. The mature capsid lattice is composed of ~1500 CA monomers thought to assemble first into CA dimers, then into ~250 CA hexamers and 12 pentamers that facilitate the curvature of the mature capsid core (31). Each CA monomer contains independently folded N-terminal (CA_{NTD}) and C-terminal (CA_{CTD}) domains that are connected by a flexible linker (32-35). Neighboring subunits in each hexamer and pentamer are connected by intermolecular CA_{NTD}-CA_{NTD} and CA_{NTD}-CA_{CTD} interactions, while pairs of subunits between hexamers are connected by dimeric and trimeric CA_{CTD}-CA_{CTD} contacts (33-39). Interactions between the CA_{NTD} and CA_{CTD} regulate and stabilize the lattice in conjunction with a hydration layer composed of thousands of water molecules (40,41).

Assembly of CA into these independent states requires conformational flexibility of each CA protein (38,40), supporting the presence of unique, solvent-accessible binding surfaces on each assembly state. Replication events and thus viral infectivity can be severely compromised by mutations that impact lattice stability, alter interactions between or among CA assembly states, or disrupt the hydration layer (1,5,31,36,40,42-50). While destabilization must eventually occur for successful replication, a delicate balance between stability and dissociation must be maintained, demonstrating an intricate link between the capsid core and viral infectivity (3,5,6,9,46,51).

Substantial developments in protein engineering, structural biology, and molecular modeling have led to an improved understanding of the molecular architecture and interaction surfaces of various CA assembly forms (12,38,40,52-56). Indeed, several critical binding interactions between CA and host factors have been defined. These host factors include restriction factors such as TRIM5alpha, TRIMCyp and MxB (27,57-67), and infectivity factors such as Cyclophilin A (CypA), cleavage and polyadenylation specific factor 6 (CPSF6), nucleoporin 153 (Nup153), and recently identified Sec24C (11,15,68-73). Interestingly, several of these host factors share overlapping binding sites at CA hexamer junctions that are lost upon capsid core disassembly, illustrating the importance of assembly state-specific interactions (74). Small molecules also bind sites associated with specific assembly forms. The recent discovery of the stabilizing role of inositol hexakisphosphate (IP6) (75-80) has facilitated important developments such as the recent demonstration of efficient reverse transcription and integration by capsid cores in a cell-free system (2). The identification of small molecules that bind CA has also contributed to our understanding of the role of CA in HIV-1 replication. PF-3450074

(PF74), which binds to a pocket found at the CA_{NTD}-CA_{CTD} interface of assembled CA hexamers, has been shown to alter the stability of the CA lattice (71,81,82). PF74 seems to interfere with the interaction of the capsid core and host proteins, namely CPSF6 and NUP153 (69,71,74,82,83). Gilead compounds, GS-CA1 and GS-6207, more recently described CA-targeting small molecules that bind in the same pocket as PF74, interfere with both nuclear import of viral DNA and CA assembly (14,82,84). GS-6207 is currently in clinical trials. Importantly, the ability of known restriction factors and small molecules to specifically recognize different CA assembly forms supports the presence of unique, biologically relevant binding sites present on these CA assembly forms and underscores the importance of identifying novel targetable sites.

Despite these advances, there are many unresolved questions regarding CA structural dynamics and their impact on viral biology. Recent work supports the presence of partial Gag hexamers at the edges of the immature Gag hexamer lattice, which could serve as substrates for proteolytic maturation and contribute to assembly of the immature lattice (53). However, further study is required to fully understand the order in which immature lattice assembly occurs and how it leads to activation of the viral protease, as well as the assembly mechanisms underlying remodeling of proteolytically cleaved CA into the mature capsid core. In addition, the mature capsid core may also undergo some degree of remodeling to facilitate reverse transcription, nuclear entry and integration (2). Indeed, dissociation is required for access of the proviral DNA to the host chromosome during integration, although the degree of dissociation required and the location at which dissociation begins remain controversial (18,19,21-25,51,85-88). Furthermore, we still do not fully understand the broad spectrum of CA interactions with host proteins and their

implications for virus replication, and previously undescribed interactions and targetable surfaces likely exist. New tools are needed to support studies of the contributions of CA assembly forms to diverse replication events, to identify novel interaction sites on different CA assembly forms, or to compete with various host factors for binding to CA assembly forms.

To address these gaps, we sought to develop aptamers that specifically target accessible binding sites present on the assembled CA lattice. Aptamers are structured oligonucleotides generated through an enrichment process termed SELEX (Systematic Evolution of Ligands by EXponential enrichment) to bind molecular targets. Through adoption of highly structured 3D conformations, aptamers have been shown to discriminate among very similar proteins (e.g. a single amino acid change (89-91) or different conformations of the same protein (92-97)). Prior work has identified aptamers targeted to various HIV-1 proteins, including nucleocapsid (98), Gag polyprotein (99,100), integrase (101), reverse transcriptase (89,102-105), and others (106-108). Aptamers targeting the Gag polyprotein were demonstrated to bind the Gag polyprotein as well as the matrix (MA) or nucleocapsid (NC) components of Gag, likely due to the presence of highly basic regions in these proteins. Notably, some of the Gag aptamers were able to bind Gag and the CANC portion of Gag, but not NC alone. However, they were not reported to bind recombinant CA alone (99, 100) and failed to bind the assembled CA lattice utilized here (Figure 2.1). Thus, aptamers that specifically target HIV-1 CA or its assembly forms have not been previously described. In this study, we report the identification and characterization of aptamer CA15-2, which binds specifically the HIV-1 CA lattice, but not the CA monomer or CA hexamer. Binding was

not dependent upon mutations introduced into the CA lattice construct, as CA15-2 was also able to bind assembled native CA lattice tubes. We also demonstrate that aptamer CA15-2 significantly inhibits HIV-1 replication in producer cells, but not in target cells.

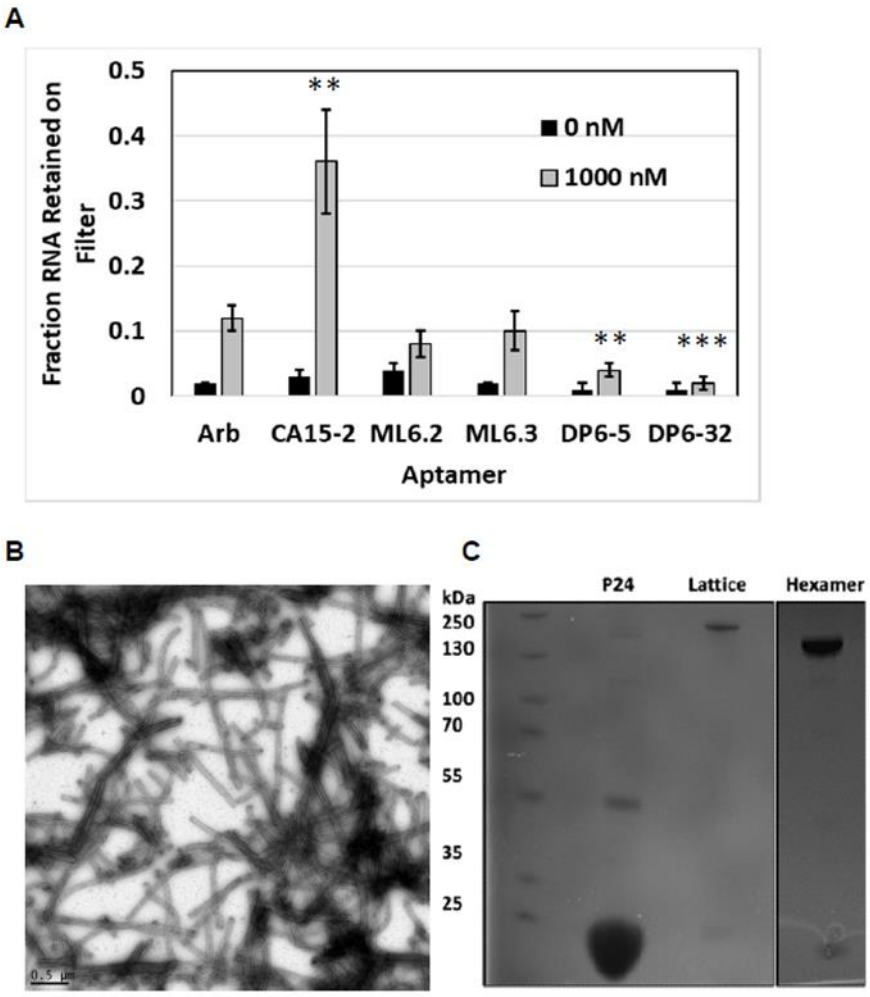

Figure 2.1: Previous anti-HIV Gag aptamers do not bind CA lattice and validation of CA protein assemblies used in this study. (A) Binding of previously identified aptamers selected against the HIV-1 Gag polyprotein to the assembled CA lattice. (n=3). Values are the mean ± SD. ** ($P < 0.01$), *** ($P < 0.001$). (B) Transmission electron microscopy image of assembled CA lattice tubes. (C) Non-reducing SDS-PAGE analysis of the CA monomer (p24), CA lattice, and CA hexamer. For panel B, the grid was prepared by PRG and SW, and the image was taken by the MU Electron Microscopy Core. Panel C was generated by RA.

2.3. Results and Discussion

2.3.1. Selection of RNA aptamers with affinity to the assembled HIV-1 CA lattice

The viral capsid core is a complex target that contains a variety of potential binding sites for aptamers and other novel ligands. Some of these sites are uniquely present in viral capsid cores because they lie at the junctions between hexamer or pentamer units; others are present within individual CA hexamers, pentamers, or monomers. However, intact viral capsid cores are difficult to isolate with the purity and quantity required for aptamer selection. Therefore, we utilize an *in vitro* assembled, stabilized HIV-1 CA lattice (referred to here as CA lattice) as a surrogate for the lattice present within the viral capsid core (37). The A14C/E45C CA double mutant forms disulfide bonds between adjacent CA monomers within a hexamer, and these cross-linked hexamers assemble into hyperstabilized CA lattice tubes at high salt concentrations (Figure 2.1B). Importantly, due to the cross-linking, these stabilized tubes can be removed from high salt buffers without disrupting the lattice and have been shown to maintain interactions with lattice-binding host factors (65). In addition, the CA lattice tubes can be pelleted by centrifugation, enabling partitioning of bound versus unbound RNAs during the aptamer selection process. Aptamer-targetable sites on the CA lattice include those present on the CA monomer, CA hexamer, and/or CA lattice (Figure 2.2A).

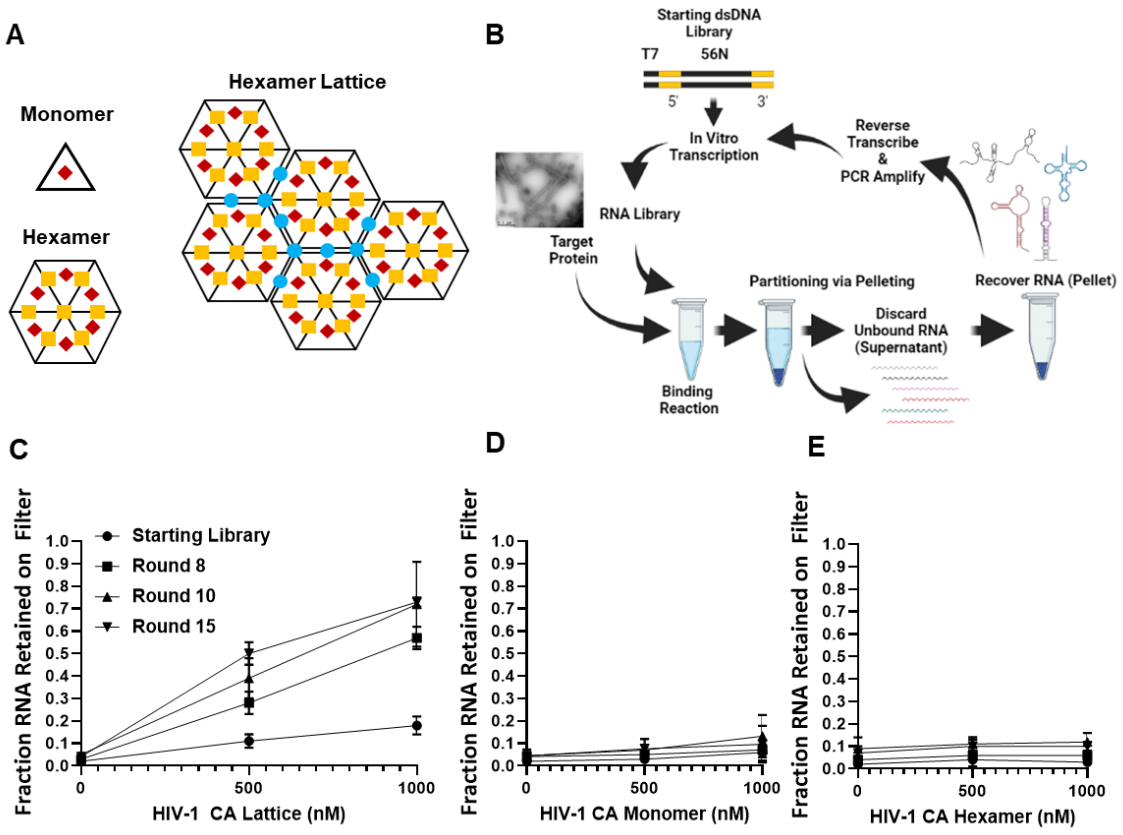

Figure 2.2: CA lattice aptamer selection. (A) Potential binding interfaces for CA-binding aptamers that are associated with the CA monomer (diamonds), CA hexamer (squares), and CA lattice (circles). (B) Schematic representation of the HIV-1 CA lattice aptamer selection created using BioRender.com. The starting dsDNA library containing the T7 promoter and a 56-nucleotide random region flanked by 5′ and 3′ constant regions was transcribed using T7 RNA polymerase. The RNA library was incubated with assembled HIV-1 CA lattice tubes (transmission electron microscopy image of the lattice tubes is shown) in a binding reaction. Library partitioning via pelleting was performed to separate the bound RNA (pellet) from the unbound RNA (supernatant). Bound RNA was recovered, reverse transcribed into cDNA, PCR-amplified, and transcribed to generate the RNA library to be used in the next round of selection. (C, D, E) Binding of aptamer libraries from rounds 8, 10, and 15 to the HIV-1 capsid lattice (C), CA monomer (D) and CA hexamer (E) was evaluated using nitrocellulose filter binding assays (n=3). Values are the mean ± SD. Binding for each library set was compared to the starting library for determination of statistical significance; ns ($P > 0.05$), ** ($P<0.01$), *** ($P<0.001$). Statistical significance for each round (500 nM and 1000 nM CA lattice) is as follows: Round 8 (**, ***), Round 10 (**, **) and Round 15 (***, ***). Throughout this study, CA concentrations for all assembly forms are given in terms of total monomer. Panels A and B were generated by MJL.

To identify CA lattice-binding aptamers, the refolded 56N RNA library was incubated with CA lattice, followed by pelleting of CA lattice to partition bound versus unbound RNA sequences. Unbound sequences were discarded with the supernatant, while bound sequences were recovered from the pellet after washing. Recovered RNA was reverse transcribed and PCR amplified to restore the T7 RNA polymerase promoter (see Table 2.1). The resulting dsDNA was then used for *in vitro* transcription to generate the RNA library for the next round of selection (Figure 2.2B). A total of 15 rounds of selection were performed.

Table 2.1: Aptamer and primer sequences used in this study[a]

Name	Sequence
56N Library Top Strand[b]	GCCTAATACGACTCACTATAGGAAGAAGAGAATCATACACAAGA-$(N)_{56}$-GGGCATAAGGTAGGTAAGTCCATA
56N Forward Primer	GCCTAATACGACTCACTATAGGAAGAAGAGAATCATACACAAGA
56N Reverse Primer	TATGGACTTACCTACCTTATGCCC
56N InFusion Forward Primer	CCTTGTATCTGAATTCGAAGAGAATCATACACAAGA
56N InFusion Reverse Primer	CACAATCCGTGACAGGGCCCTATGGACTTACCTACCTTATGCCC
Arb Forward Primer	GCCTAATACGACTCACTATAGGGAAAAGCGAATCATACACAAGA
Arb Reverse Primer	TATGGAATTAAATACCTTATGCCC
56N 5′ Anti-Leader	TCTTGTGTATGATTCTCTTCTTCC
CA15-2	ggaagaagagaaucauacacaagaUCGACGUACCUCAGGGUGGUGUAUGACUGAGGUGAAGACUGUGAACCAUGGCAUGCgggcauaagguagguaaguccaua
CA8-4	ggaagaagagaaucauacacaagaCCAGCGCGCGUCCAGUAACAGCCGGCCCAUGCCGCCUACUGUCUUGCAGAUUCCACgggcauaagguagguaaguccaua
CA8-6	ggaagaagagaaucauacacaagaAACAAGUGAUUAAGGAAGGUGUAUAGGAGGCGAAUCACUGUAAGGUGUUGGAAUAgggcauaagguagguaaguccaua
CA10-2	ggaagaagagaaucauacacaagaCCAUCCCCCAACUUGUCAAGAGGGAAGGAGUGGGAAAGGAAGACAAGGACAUGUGgggcauaagguagguaaguccaua
CA10-5	ggaagaagagaaucauacacaagaGGCGAAUUACACGACAAUCUGGACGACCUACACUAUGAGAGGAGGGUAGAGUGGAAgggcauaagguagguaaguccaua
CA15-1	ggaagaagagaaucauacacaagaCCUACCUGUAUGAAAGAGACGAACCUAACUACAAUGCUAGCAAUGGUGUAUGGUGCgggcauaagguagguaaguccaua
CA15-8	ggaagaagagaaucauacacaagaCCGUGCCCAGGGUCUUGCUGUAUGUAUGGUGCUGAUCUACUGUAGCCAUGUGUAUGgggcauaagguagguaaguccaua

Scrambled CA15-2	ggaagaagagaaucauacacaagaGGAGGUGCAGGUGCAUAAUGCGUGAAGGCUCGGCCUGUACACUCGAAGUACUUAGUgggcauaagguagguaaguccaua
Arbitrary RNA	gggaaaagcgaaucauacacaagaAAUUUGGACUUUCCGCCCUUCUUGGCCUUUAUGAGGAUCUCUCUGAUUUUUCUUGCGUCGAGUUUUCCGGgggcauaagguauuuaauuccaua
CA15-2 1-80 Stem Disrupt (1-80*)	ggaagaagagaaucauacacaagaUCGACGUAGAGCAGGGUGGUGUAUGACUGAGGUGAAGACUGUGAACCAUGGCAUGC
CA15-2 1-80 Stem Rescue (1-80**)	ggaagaagagaaucauacacaagaUCGACGUAGAGCAGGGUGGUGUAUGACUGCUCUGAAGACUGUGAACCAUGGCAUGC
CA15-2 25-104[c]	GGUCGACGUACCUCAGGGUGGUGUAUGACUGAGGUGAAGACUGUGAACCAUGGCAUGCgggcauaagguagguaaguccaua
CA15-2 25-80	GGUCGACGUAGAGCAGGGUGGUGUAUGACUGCUCUGAAGACUGUGAACCAUGGCAUGC
CA15-2 25-60	GGUCGACGUAGAGCAGGGUGGUGUAUGACUGCUCUGAA
CA15-2 30-60	GUACCUCAGGGUGGUGUAUGACUGAGGUGAA

[a]The constant regions for the RNAs are in lowercase. [b]The T7 promoter is underlined. [c]For CA15-2 transcripts beginning at position 25, two guanine nucleotides (underlined) were added to the 5' end to aid in *in vitro* transcription. The addition of these nucleotides did not alter the predicted secondary structures represented in Figure 2.10.

We determined the binding profiles of the RNA libraries from rounds 8, 10, and 15 to different HIV-1 CA assembly forms using nitrocellulose filter binding assays, including CA lattice (selection target), soluble CA monomer, and soluble CA hexamer (Figure 2.1C). The CA monomer expression construct contains no mutations, while the soluble CA hexamer construct contains the same A14C/E45C mutations as the CA lattice, along with two additional alanine mutations (W184A/M185A) that abolish higher order lattice assembly (37). As expected, the three aptamer libraries exhibited dose-dependent binding to CA lattice as compared to the starting library (Figure 2.2C). In contrast, there was no significant change in dose-dependent binding of the aptamer libraries to either the CA monomer (Figure 2.2D) or CA hexamer (Figure 2.2E), indicating that most of the RNA species in these libraries recognize epitopes that are uniquely present on CA lattice.

Sanger sequencing of plasmids generated from cloning of Rounds 8, 10, and 15 (Table 2.1) did not reveal obvious convergence. However, in preliminary screens to evaluate binding of a subset of the individual clones to CA lattice (Figure 2.3), aptamer CA15-2 consistently exhibited strong, dose-dependent binding to CA lattice, and we chose to focus on aptamer CA15-2 for further characterization.

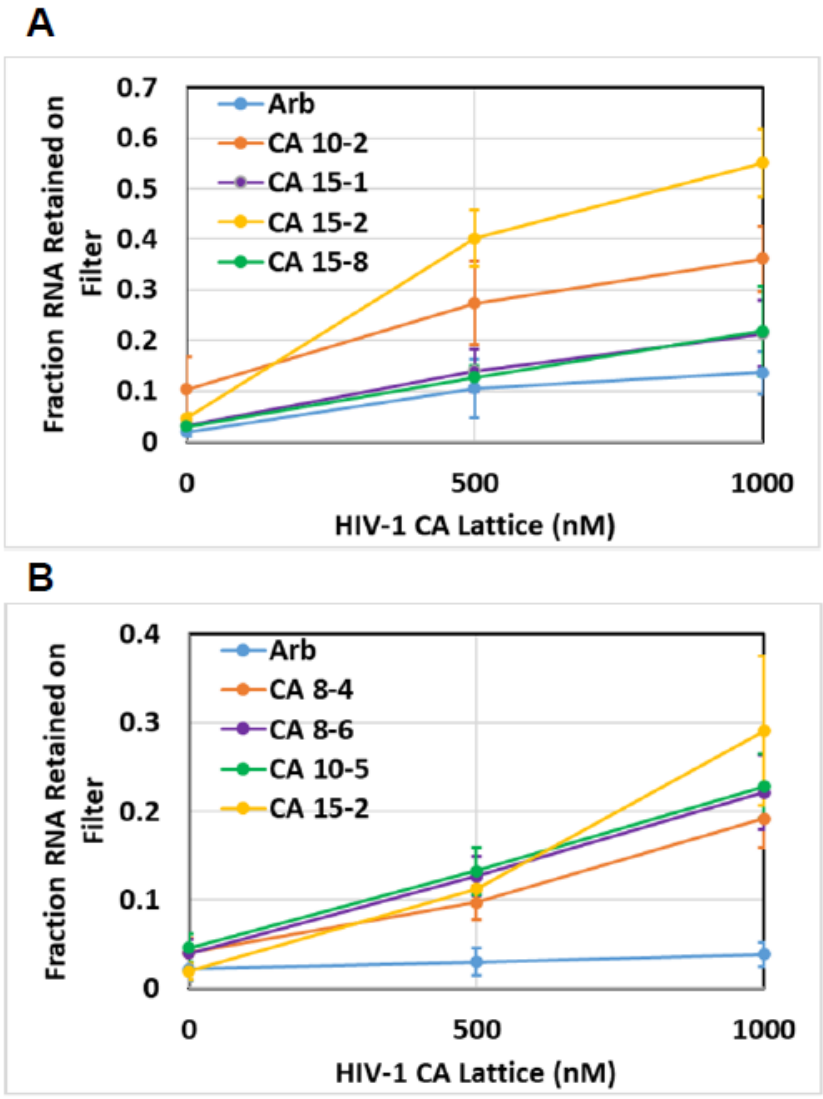

Figure 2.3: Screening individual aptamers for binding to the CA lattice. In panels A and B, two different nitrocellulose filter binding screens were performed to determine the ability of a panel of aptamer clones to bind the CA lattice including aptamer CA15-2, other identified aptamer clones, and an arbitrary RNA control. Values are the mean ± SD for at least three experiments.

2.3.2. Aptamer CA15-2 binds to the assembled CA lattice but not the soluble CA hexamer or CA monomer

Binding of aptamer CA15-2 to different CA protein assembly forms was evaluated using nitrocellulose filter binding assays, electrophoretic mobility shift assays (EMSA), microscale thermophoresis (MST), and size exclusion chromatography (SEC). The purity of each protein is represented in Figure 2.1. Aptamer CA15-2 exhibited a significant mobility shift in the presence of CA lattice via both EMSA (Figure 2.4A) and SEC (Figure 2.5A). It also exhibited dose-dependent binding to CA lattice in nitrocellulose filter binding assays (Figure 2.4B and 2.4C) and MST assays (Figure 2.4D). In contrast, the arbitrary control RNA showed little or no interaction with CA lattice (Figure 2.4B, 2.4C, 2.4D and 2.5A) (109). Furthermore, aptamer CA15-2 showed no evidence of binding to CA monomer (Figure 2.4E and 2.5B) or to CA hexamer (Figure 2.4F and 2.5C) in contrast to positive control, PF74, which bound both the CA monomer and CA hexamer (data not shown). Collectively, these data demonstrate specificity of aptamer CA15-2 for CA lattice.

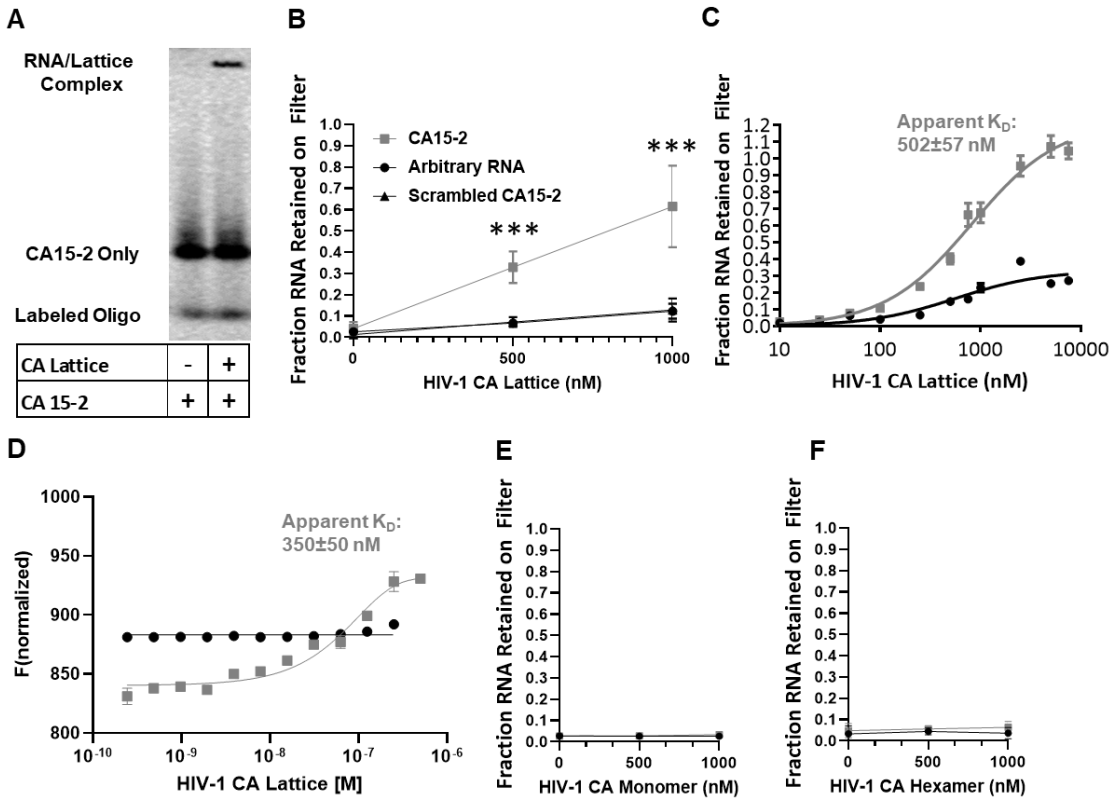

Figure 2.4: Specific binding of aptamer CA15-2 to the HIV-1 CA lattice. (A) An electrophoretic mobility shift assay demonstrates formation of a complex between Cy5-labeled aptamer CA15-2 and CA lattice. (B, C, E, and F) Nitrocellulose filter binding assays were used to evaluate binding of radiolabeled aptamer CA15-2 or arbitrary RNA to the CA lattice (B and C; n=8), hexamer (E; n=6), and monomer (F; n=3). Values are the mean ± SD. Statistical comparisons were made between CA15-2 and the arbitrary control. *** ($P<0.001$). (D) Microscale thermophoresis assays were used to confirm aptamer CA15-2 binding to the CA lattice using Cy5-labeled CA15-2 and the indicated concentrations of CA lattice. Panel D was generated by RA.

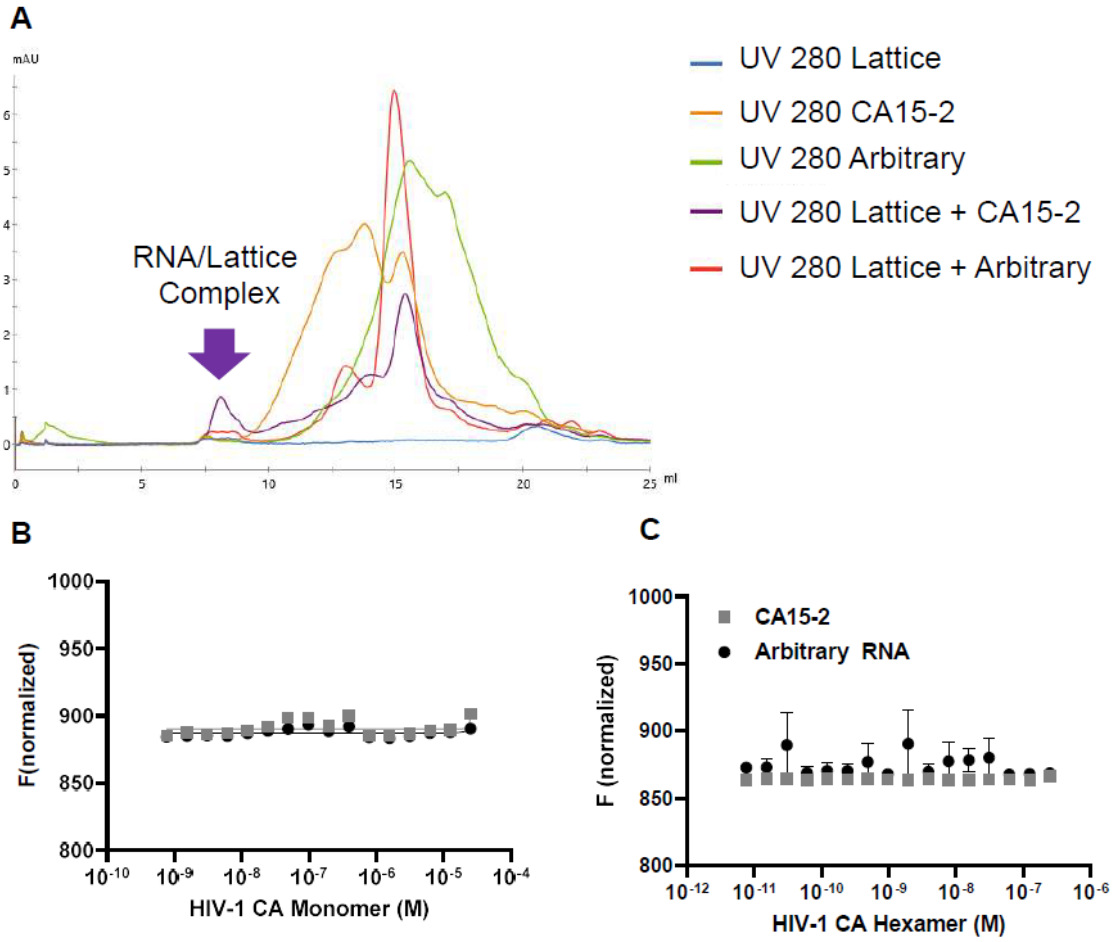

Figure 2.5: Assessing binding of aptamer CA15-2 to CA assemblies using size exclusion chromatography and microscale thermophoresis. (A) Size exclusion chromatography demonstrating formation of an aptamer CA15-2 and CA lattice complex. (B and C) Microscale thermophoresis analysis of aptamer CA15-2 and the arbitrary control to the CA monomer (B) and CA hexamer (C). Values are the mean ± SD for at least three experiments. Figure was generated by RA.

When binding of aptamer CA15-2 to the CA lattice was evaluated using nitrocellulose filter binding assays at a wider range of CA protein concentrations (10-7500 nM), an apparent dissociation constant (K_{Dapp}) for aptamer CA15-2 was determined to be 502 ± 57 nM, while the apparent kD for the arbitrary control was greater than 7.5 μM (Figure 2.4C). In a similar experiment performed using MST at a CA protein concentration range of 0.1 nM to 1000 nM, aptamer CA15-2 was observed to bind the HIV-1 CA lattice with an equilibrium dissociation constant of 350±50 nM, while the arbitrary control demonstrated no binding (Figure 2.4D). Notably, the apparent kD values obtained here may not be representative of the true CA-binding strength of aptamer CA15-2. Due to the heterogeneity of the CA lattice tubes, which likely contain varying numbers of binding sites depending on tube length and diameter, CA concentrations were calculated based on the total number of CA monomer (p24) present. Finally, to confirm aptamer CA15-2 binding to CA lattice, we examined the aptamer-CA lattice complex using SEC, where we observed the retention volume of the complex to be ~7 mL as compared to ~15 mL for aptamer CA15-2 alone and ~20 mL for CA lattice alone (Figure 2.5A). Collectively, these results demonstrate that aptamer CA15-2 specifically binds to CA lattice, likely at protein interfaces that are uniquely present on CA lattice but that are not present on the CA hexamer or CA monomer (Figure 2.2A).

2.3.3. Unlabeled CA15-2, but not other unlabeled nucleic acids, competes with labeled CA15-2 for binding to CA lattice

To confirm the binding specificity of aptamer CA15-2 to CA lattice, we performed competition assays using aptamer CA15-2 annealed to Cy5-labeled antisense

oligonucleotide (EMSA) or radiolabeled aptamer CA15-2 (nitrocellulose filter binding assays) paired with a panel of unlabeled competitors. Unlabeled competitor was first incubated with CA lattice prior to addition of Cy5-labeled aptamer CA15-2. For the EMSA, 0-200 nM unlabeled competitor was used, along with 50 nM aptamer CA15-2 and 2 μM CA lattice. For the nitrocellulose filter binding assays, 0-160 nM unlabeled competitor was used with 10 nM aptamer CA15-2 and 250 nM CA lattice. The fraction of Cy5-labeled aptamer CA15-2 that bound in the absence and presence of increasing concentrations of unlabeled competitor was then compared to determine the impact of the competitor on the complex between aptamer CA15-2 and CA lattice. As expected, the relative amount of Cy5-labeled CA15-2 bound to CA lattice decreased as the concentration of unlabeled aptamer CA15-2 increased (Figure 2.6 and 2.7). Notably, the relative amount of labeled aptamer CA15-2 bound to CA lattice when unlabeled aptamer CA15-2 was present was significantly different from the amount present when unlabeled nonspecific RNAs were present in the binding reaction, even at the highest concentrations of unlabeled competitor. Unlabeled nonspecific RNAs used included Arbitrary RNA (Figure 2.6, 2.7 and 2.8A), a 3′ truncation of Arbitrary RNA (Arbitrary 1-80nt; Figure 2.6 and 2.7), Scrambled CA15-2 (Figure 2.6, 2.7 and 2.8B), and yeast tRNA (Figure 2.8C). These results between specific and non-specific competitors are consistent with a model in which unlabeled aptamer CA15-2 blocks binding by labeled aptamer CA15-2 via saturation of specific interaction sites on CA lattice. Minimal competition was observed with all other non-specific RNAs in the nitrocellulose filter binding assay (Figure 2.8). The experiment utilized a low concentration of radiolabeled aptamer CA15-2, where binding would not be saturated. Thus, specific binding in this case would be

lower than maximal binding. This is particularly relevant due to the possibility that binding affinity may be underrepresented by our apparent Kd determinations in Figure 2.4, as discussed above. Notably, defining the minimal binding structure of aptamer CA15-2 may improve the observed apparent binding affinity.

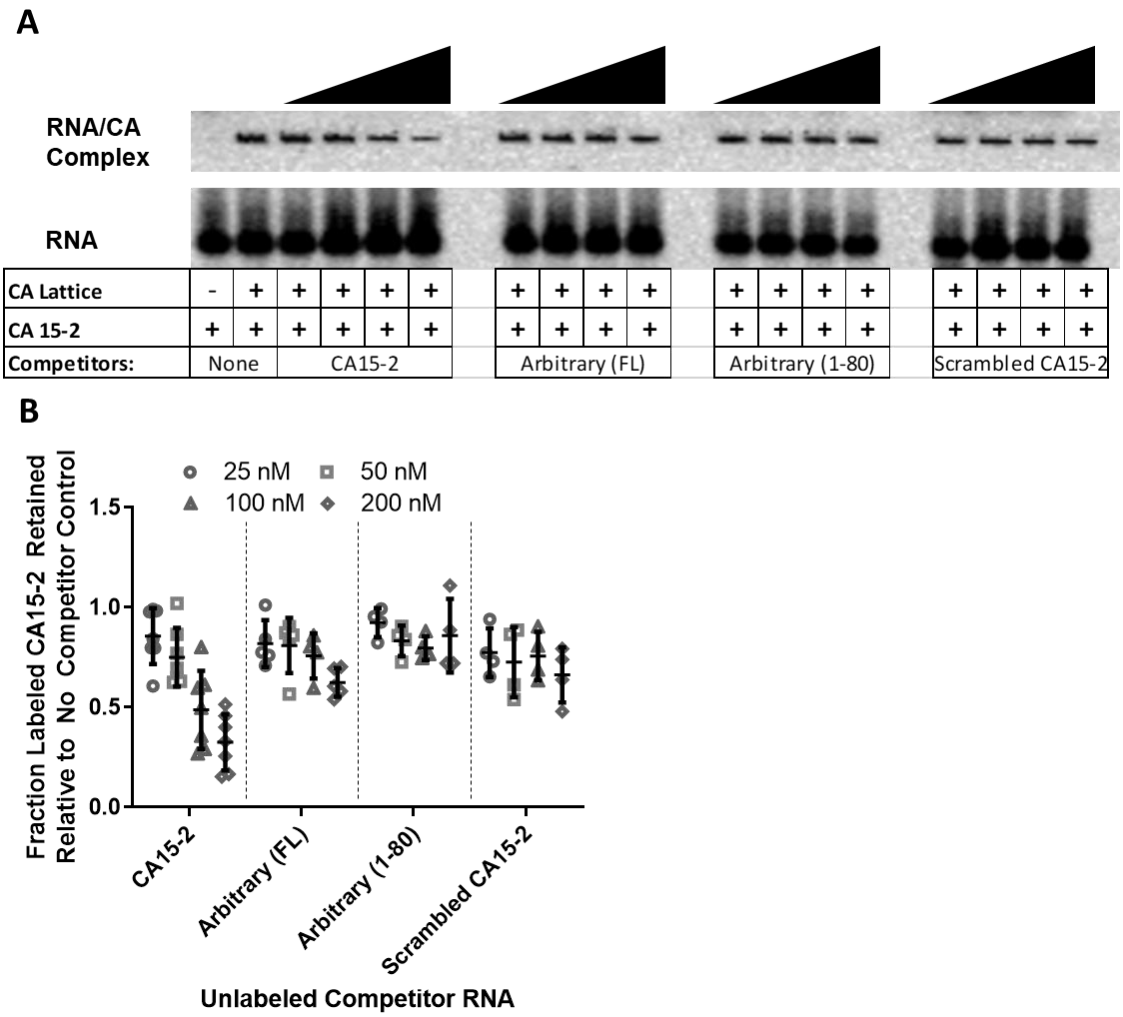

Figure 2.6. **Competition of unlabeled competitors with Cy5-labeled aptamer CA15-2 for binding to CA lattice.** (A) Electrophoretic mobility shift assays were used to determine the ability of unlabeled competitors to compete with Cy5-labeled aptamer CA15-2 for binding to the assembled CA lattice. Unlabeled full length CA15-2 was able to compete for binding to a greater degree than non-binding control RNAs (Arbitrary (FL), Arbitrary (1-80), and Scrambled CA15-2). (B) The fraction CA15-2 retained in wells relative to No Competitor was quantified using Multi Gauge software (Fujifilm) using the following equation: (intensity of signal retained in well – intensity of signal in well of RNA only lane)/(intensity of signal retained in well of No Competitor lane – intensity of signal in well of RNA only lane). Values are the mean ± SD for at least four experiments.

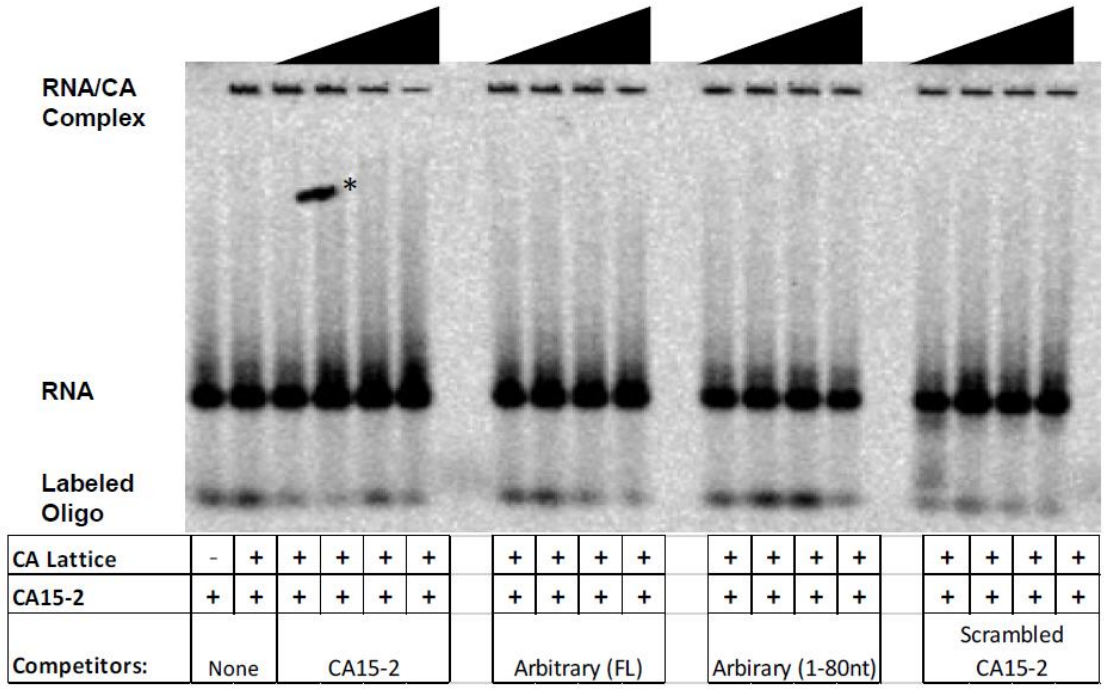

Figure 2.7: Full gel depiction of Figure 2.6. Competition of unlabeled competitors with Cy5-labeled aptamer CA15-2 for binding to the assembled CA lattice. Electrophoretic mobility shift assays were used to determine the ability of unlabeled competitors to compete with Cy5-labeled aptamer CA15-2 for binding to the CA lattice. Unlabeled full length CA15-2 was able to compete for binding to a greater degree than non-binding control RNAS (Arbitrary, Arbitrary 1-80, and Scrambled CA15-2). The (*) denotes a stray image defect present in the gel.

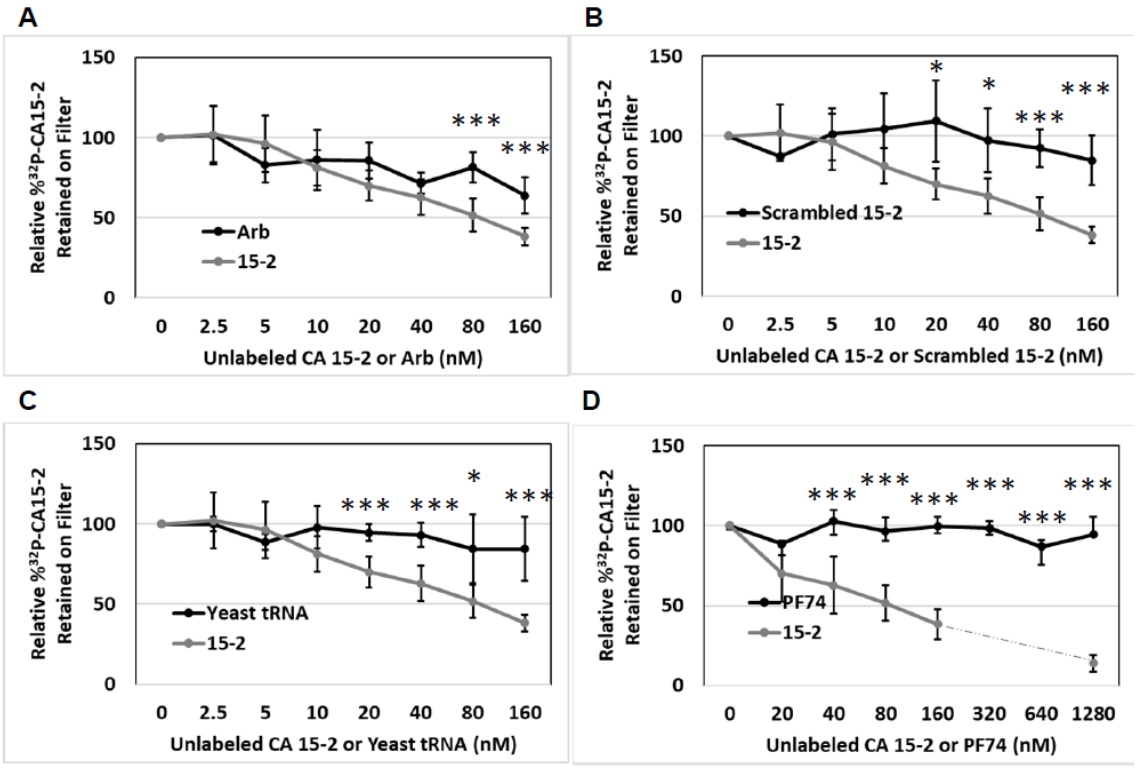

Figure 2.8: Competition of unlabeled competitors and radiolabeled aptamer CA15-2 for binding to the CA lattice. The relative percent of radiolabeled aptamer CA15-2 retained on the nitrocellulose filter was determined using full length unlabeled CA15-2 (n=5) and arbitrary control (A), scrambled 15-2 (B), yeast tRNA (C), and PF74 (D) as competitors (n=3). Values are the mean ± SD. * ($P < 0.05$), *** ($P < 0.001$).

We also evaluated competition between aptamer CA15-2 and PF74, a small molecule that inhibits HIV-1 replication (71,74,81) and that binds CA in a pocket located at the CA_{NTD}-CA_{CTD} interface of assembled hexamers with a Kd of 176 ± 78 nM (71). Radiolabeled aptamer CA15-2 did not compete with PF74 for binding to CA lattice at any of the concentrations tested (20 nM to 1280 nM; Figure 2.8D), suggesting that aptamer CA15-2 does not share an overlapping binding site with PF74. This interpretation is further supported by the lack of binding observed for aptamer CA15-2 to CA hexamer (Figure 2.4E, and 2.5B).

2.3.4. Involvement of the 3′ and 5′ constant regions of aptamer CA15-2 in binding to the CA lattice

The 56N library design included constant regions at the 5′ and 3′ ends for amplification during the selection process. To determine whether either constant region is required for aptamer CA15-2 binding to CA lattice, antisense oligonucleotides were separately annealed to each end of aptamer CA15-2 as represented in Figure 2.9A. The predicted CA15-2 structure was generated using NUPACK(110). After oligonucleotide annealing, binding to CA lattice was evaluated using nitrocellulose filter binding assays. Annealing of an antisense oligonucleotide to the 3' constant region did not significantly alter binding of aptamer CA15-2 to CA lattice as compared to the no oligonucleotide control (Figure 2.9B), indicating that sequestration via formation of a double-stranded region at the 3′ end does not disrupt the proper folding of aptamer CA15-2 into a CA lattice-binding structure. In contrast, when an antisense oligonucleotide was annealed to the 5′ constant region, binding to CA lattice was significantly decreased as compared to

the no oligonucleotide control (Figure 2.9B). This indicates that sequestration of the 5′ constant region via formation of a double stranded region interferes with the proper folding of aptamer CA15-2, ultimately blocking binding to CA lattice.

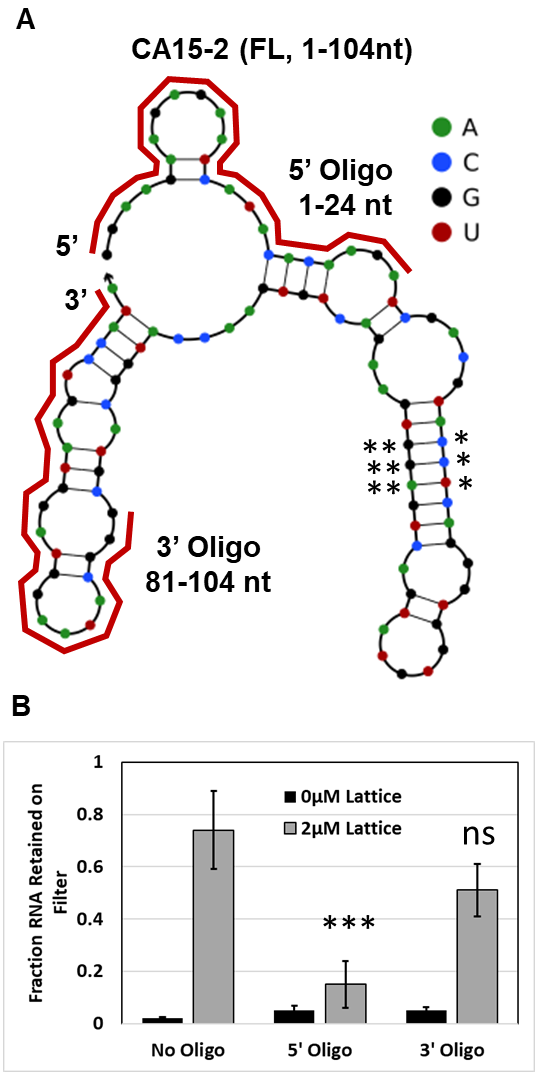

Figure 2.9. **Contributions of the 5´ and 3´ constant regions to aptamer CA15-2 binding to the CA lattice via antisense oligonucleotide annealing.** (A) The secondary structure of aptamer CA15-2 was predicted using the NUPACK web server (110). Red lines indicate binding sites for anti-sense oligonucleotides used to sequester the 5´ and 3´ constant regions. Asterisks indicate the location of stem disruption or stem restoration mutations evaluated in Figure 2.11. (B) Antisense oligonucleotides were annealed to either the 5´ or 3´ constant regions of aptamer CA15-2, followed by binding analysis using nitrocellulose filter assays. Statistical comparisons were made against the binding values for the arbitrary control. Values are the mean ± SD for three experiments. Panel A was generated by MJL.

To further define the sequence requirements for aptamer CA15-2, we tested the ability of truncated aptamer variants to bind CA lattice. Structures for each variant were predicted using NUPACK (Figure 2.10A-K) and the variants are named according to the nucleotides remaining after truncation. As shown in Figure 2.11A, full-length CA15-2 (1-104) and truncation 1-80 demonstrated significant binding to CA lattice in nitrocellulose assays ($p<0.001$), consistent with the oligonucleotide annealing data above. Interestingly, truncation 25-104 also demonstrated significant binding, suggesting that the 5′ constant region is not required for binding. Thus, the disruption of binding upon annealing of the 5′ antisense oligonucleotide observed above (Figure 2.9B) may indicate that annealing introduced steric hinderance that disrupted the binding structure. Notably, binding was abolished in truncation 25-80, which removes both the 5′ and 3′ constant regions. This suggests that while there is not a specific requirement for either constant region, each constant region can contribute to the functional binding structure in the absence of the other.

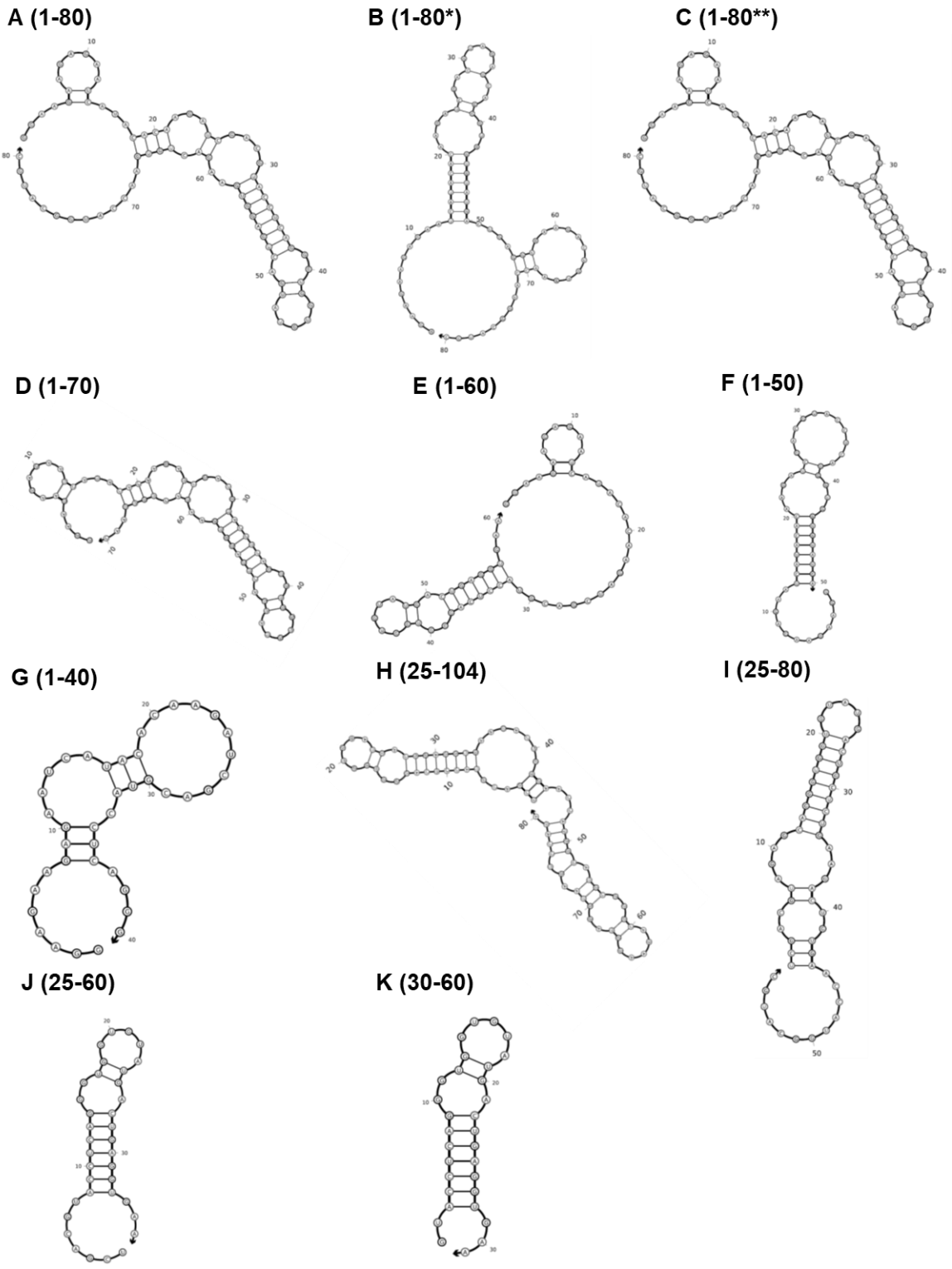

Figure 2.10 Structural predictions for the aptamer CA15-2 3´ truncations evaluated in Figure 4 and 5. Secondary structure predictions were generated using the NUPACK web server (110).

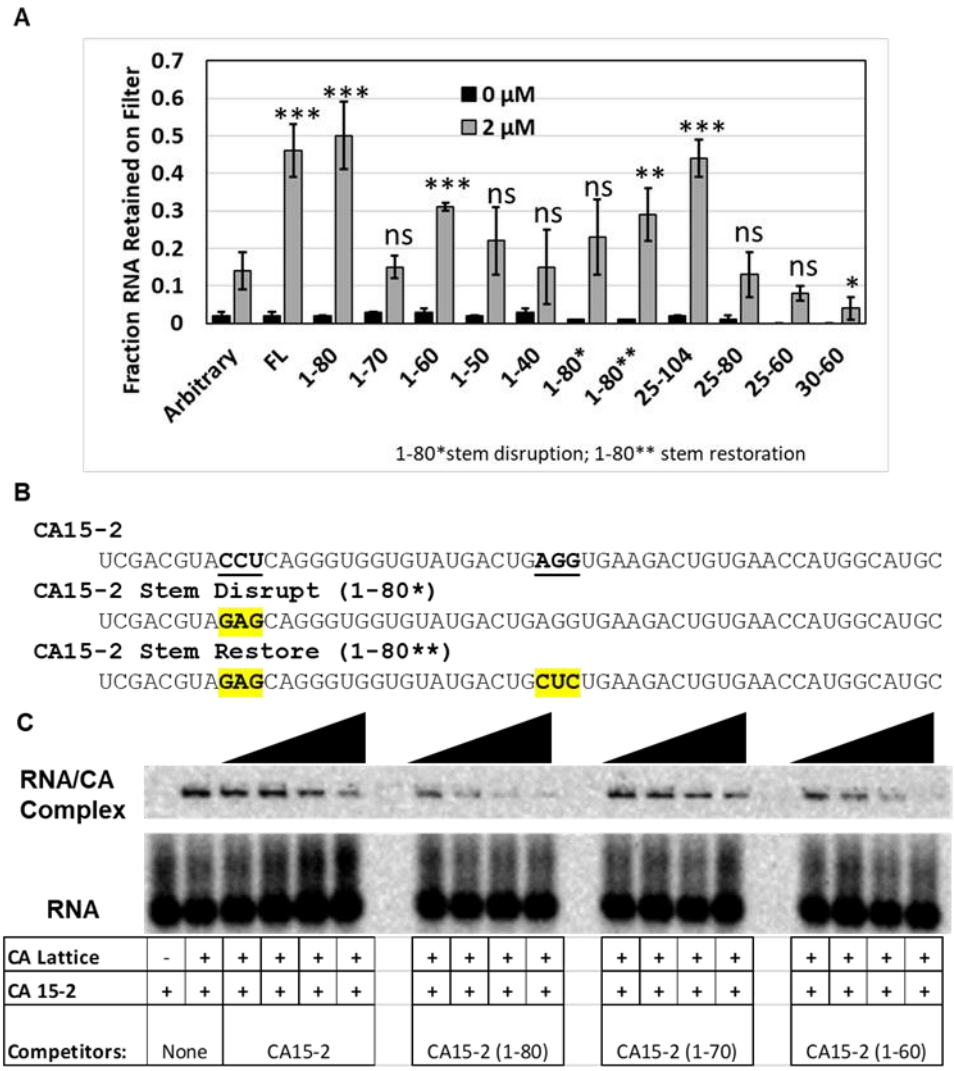

Figure 2.11. **Contributions of the 5´ and 3´ constant regions to aptamer CA15-2 binding to CA lattice via truncations.** Binding of aptamer CA15-2 and various truncations of aptamer CA15-2 was evaluated using nitrocellulose filter binding assays (A) and electrophoretic mobility shift assays (B). (A) Aptamers or control RNAs were radiolabeled, incubated with CA lattice, and evaluated using nitrocellulose filter binding assays. Values are the mean ± SD for three experiments. Statistical comparisons were made for each truncation versus the level of binding for the arbitrary control. ns ($P > 0.05$), *** ($P < 0.001$). (B) Mutations introduced for disruption or restoration of a predicted stem structure in CA15-2 are shown in yellow. (C) Electrophoretic mobility shift assays were used to determine the ability of unlabeled truncated aptamers to compete with Cy5-labeled aptamer CA15-2 for binding to CA lattice. The fraction of Cy5-labeled, full-length aptamer CA15-2 retained in wells relative to No Competitor was quantified using Multi Gauge software (Fujifilm) using the following equation: (intensity of signal retained in well – intensity of signal in well of RNA only lane)/(intensity of signal retained in well of No Competitor lane – intensity of signal in well of RNA only lane). Experiments were repeated at least three times.

Interestingly, truncation 1-60 demonstrated a greater degree of binding as compared to truncation 1-70 (Figure 2.11A; $p<0.001$). These results were corroborated by EMSA-based competition assays, wherein the full-length aptamer (1-104) and truncations 1-80 and 1-60 competed with full-length CA15-2 for binding, as shown by the decrease in mobility shift with increasing competitor concentration (Figure 2.11C and 2.12). It is possible that 1-60 generates a stabilized structure capable of binding CA lattice, albeit to a lesser degree than 1-104 or 1-80, which may not be formed by 1-70. Truncations 1-50, 1-40, 25-80, 25-60, and 30-60 did not bind significantly over the arbitrary control (Figure 2.11A).

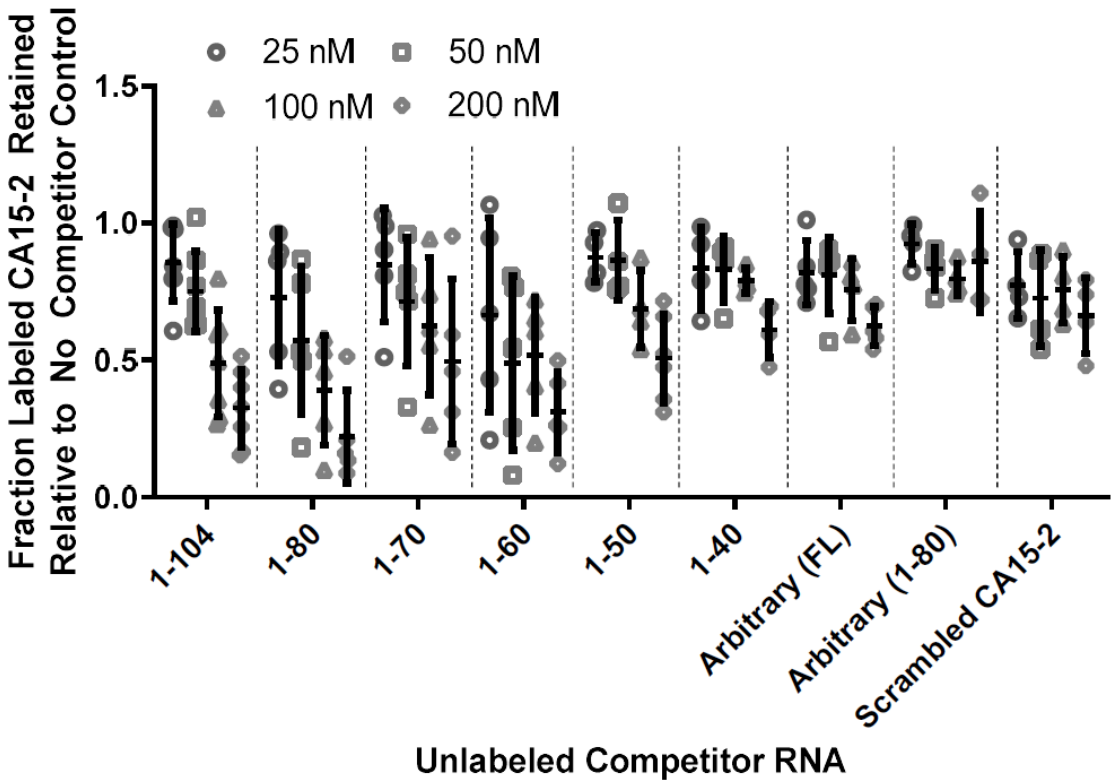

Figure 2.12 Quantification of Figure 2.11C demonstrating the ability of aptamer CA15-2 3′ truncations to compete or not compete with full length CA15-2 for binding to the CA lattice by electrophoretic mobility shift assay. The fraction of Cy5-labeled, full-length aptamer CA15-2 retained in wells relative to No Competitor was quantified using Multi Gauge software (Fujifilm) using the following equation: (intensity of signal retained in well – intensity of signal in well of RNA only lane)/(intensity of signal retained in well of No Competitor lane – intensity of signal in well of RNA only lane). Values are the mean ± SD of at least three experiments.

The predicted secondary structures of all active truncations of aptamer CA15-2 include a stem-loop structure from approximately position 25-62. We tested the importance of this stem by either disrupting the stem (1-80*) or restoring the stem (1-80**) (Figure 2.9A (asterisks), 2.11A, and 2.11B (yellow highlight)). Disruption of the stem resulted in binding that was not statistically different from the arbitrary control, whereas restoration of the stem demonstrated significant binding ($p<0.01$) as compared to the arbitrary control (Figure 2.11A). However, there was not a significant difference in binding between the disrupted stem and the restored stem, and restoration of the stem did not restore binding to the level of 1-80. Thus, aptamer CA15-2 likely adopts a different structure than predicted by NUPACK or has specific nucleotide requirements. Collectively, these data suggest that the nucleotides required for aptamer binding reside within nucleotides 1-60, as the truncated aptamer (1-60) retained binding activity, albeit to a lower extent. The reduction of target affinity displayed by 1-60 is consistent with previous reports (111). Notably, our results also suggest that additional residues located within nucleotides 80-104 also play an important role, either by stabilizing the active fold or through additional contacts with the target molecule.

2.3.5. Aptamer CA15-2 binds native CA lattice tubes

Aptamer CA15-2 was selected against CA lattice containing mutations A14C/E45C. To determine whether these mutations were required for binding, we sought to evaluate binding to native CA lattice tubes. Common methods for assembly of native tubes rely on inclusion of IP6 or high concentrations of salt; however, these conditions abolished aptamer CA15-2 binding (Figure 2.13). PF74 analog, compound 15, provides

stabilization to allow for assembly of native CA tubes under physiological salt conditions (Figure 2.14A). As we previously demonstrated that PF74 is unable to compete with aptamer CA15-2 for binding (Figure 2.8D), we performed similar competition experiments with compound 15, followed by binding assays using native CA lattice tubes assembled in the presence of compound 15. As shown in Figure 2.14, compound 15 is unable to compete with aptamer CA15-2 for binding to the assembled CA lattice (Figure 2.14B) and aptamer CA15-2 is able to bind assembled native CA lattice tubes significantly over the arbitrary control, similarly to the assembled CA lattice used in the selection (Figure 2.14C). These results demonstrate that binding of aptamer CA15-2 is not dependent upon the presence of the A14C/E45C mutations.

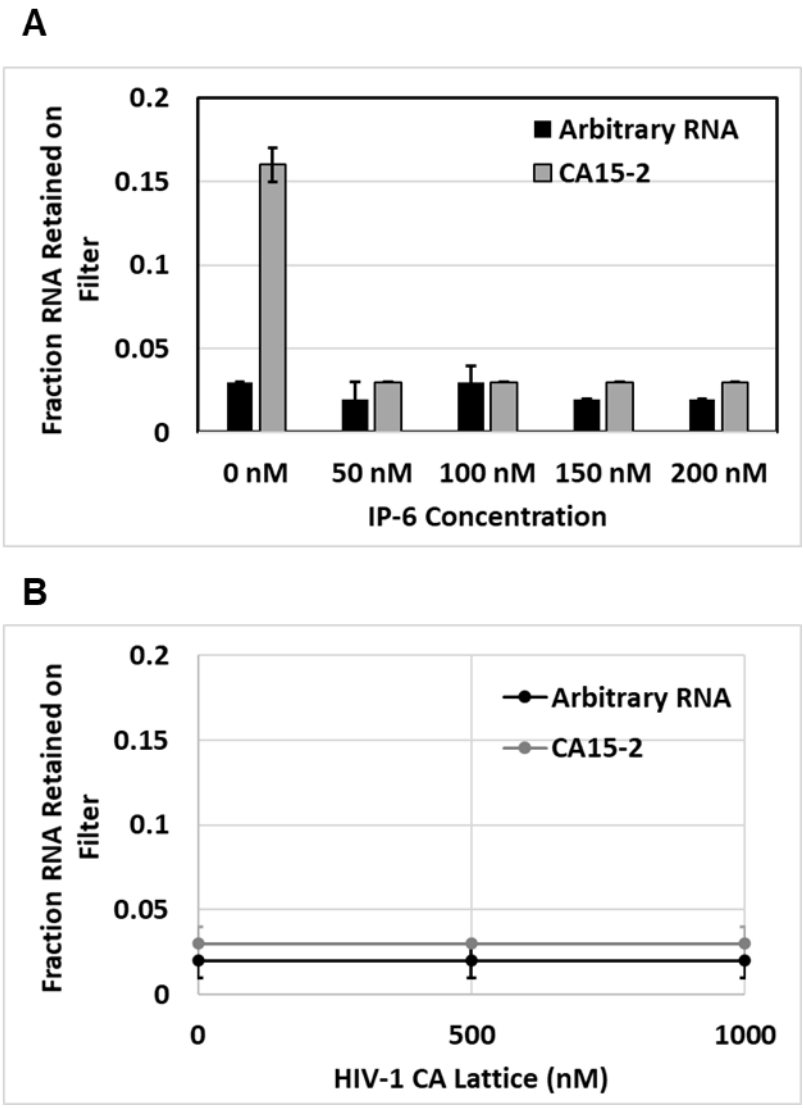

Figure 2.13: Aptamer CA15-2 binding is abolished in the presence of IP6 or high salt. (A) Nitrocellulose filter binding assays were performed to evaluate the ability of aptamer CA15-2 to bind to the CA lattice in the presence of increasing concentrations of IP6. (B) Nitrocellulose filter binding assays were performed to evaluate the ability of aptamer CA15-2 to bind to the CA lattice in the presence of 1M NaCl. Values are the mean ± SD for three experiments.

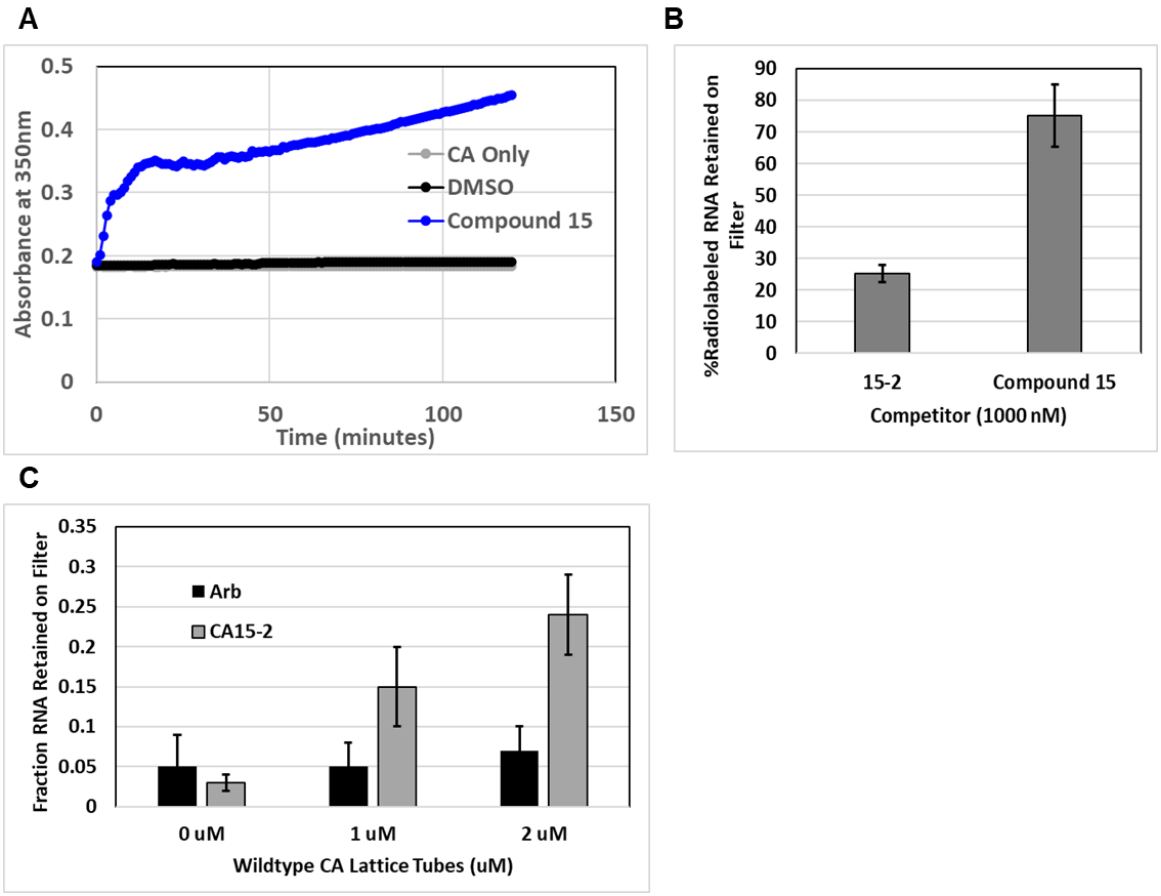

Figure 2.14: Aptamer CA15-2 binds to native assembled CA lattice tubes. (A) Compound 15, a PF74 analog, facilitates assembly of native CA tubes at physiological salt concentrations. Experiment was performed three times, and representative data is shown. (B) Compound 15 does not compete with aptamer CA15-2 for binding to the assembled CA lattice. Values are the mean % radiolabeled CA15-2 retained on filter ± SD of three experiments. (C) Aptamer CA15-2 binds to native CA lattice tubes assembled in the presence of Compound 15. Values are the mean ± SD of four experiments.

2.3.6. Aptamer CA15-2 inhibits HIV-1 infectivity at the producer cell stage but not at the target cell stage.

To investigate whether intracellular expression of aptamer CA15-2 affects HIV-1 replication, we performed single-cycle infectivity assays using VSV-G-pseudotyped HIV-1 as previously described (109). In principle, such effects could occur during the efferent phase of the viral replication cycle (virus assembly and release) or the afferent phase (entry and infection). To examine these stages separately, we first determined the effect of aptamer CA15-2 during virus production by producing VSV-G-pseudotyped HIV-1 particles in cells that expressed either aptamer CA15-2 or the arbitrary RNA control (Figure 2.15A). Aptamer or arbitrary control RNA-expressing plasmids were co-transfected into 293FT cells along with the proviral plasmid, pNL4-3-CMV-EGFP and pMD-G which expresses VSV-G. Viruses were harvested 48 hours post infection and were used to infect fresh, non-aptamer-expressing TZM-GFP target cells. TZM-GFP cells produce GFP upon infection, increasing the intensity of the GFP signal. Total virus production, as measured by p24 ELISA, was slightly elevated for aptamer-expressing cells, but the difference was not statistically significant (Figure 2.15B). In contrast, relative infectivity, as measured by the percentage of EGFP-positive cells normalized to p24 and the arbitrary control, was more than five-fold lower for aptamer CA15-2 compared to the arbitrary control (Figure 2.15C). Collectively, these results demonstrate that aptamer CA15-2 inhibits virus replication as compared to the arbitrary control when present during virus production and that these effects are not due to a decrease in viral particle release.

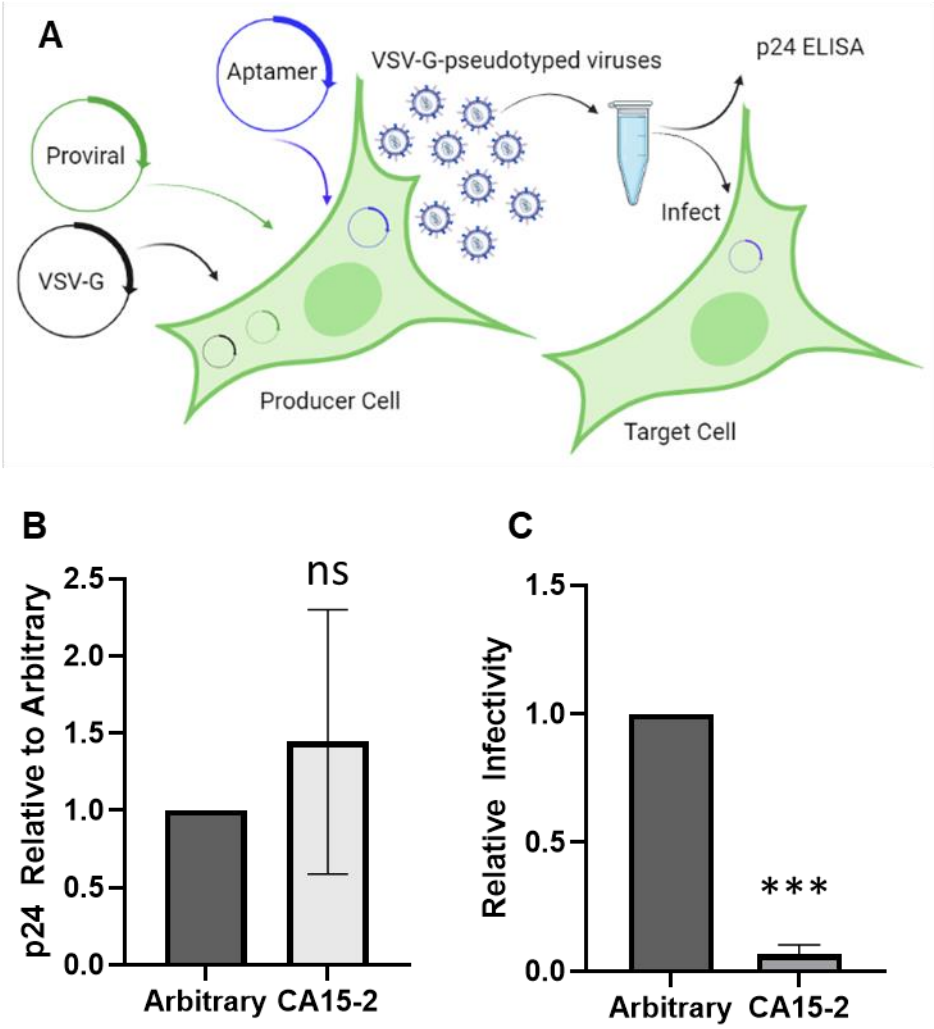

Figure 2.15. **Aptamer CA15-2 effects on HIV-1 replication in producer cells.** (A) Schematic depiction of the producer cell assay to evaluate aptamer effects when present during virus production created using BioRender.com. (B) p24 amounts were determined using ELISA for each sample and normalized to the arbitrary control for comparison. (C) Viral supernatants generated in aptamer-expressing cells were used to infect fresh TZM-GFP cells. Relative infectivity was determined by normalizing the percentage of EGFP positive cells to the amount of p24 present in each sample, followed by normalization to the arbitrary control. Values are the mean ± SD for five experiments. Panel A was generated by MJL, and panels B and C were generated by RA.

In contrast with the producer cell assay, we did not observe any reduction in infectivity when 293FT cells were first transfected with plasmids that expressed aptamer CA15-2 or the arbitrary control RNA prior to challenge with VSV-G-pseudotyped virus produced in the absence of aptamer (Figure 2.16). These results suggest that aptamer CA15-2 is unable to protect cells from incoming virus infection at the dose of aptamer used here. It is possible that aptamer CA15-2 can interact with the incoming viral capsid core, but that these interactions do not significantly impact post-entry steps of replication or capsid core stability. Alternatively, aptamer CA15-2 may not be present in a cellular location that enables interaction in target cells or may interact with a binding interface that is either not present or not accessible on the incoming viral capsid core but that is present and accessible during or after virus assembly in the producer cell.

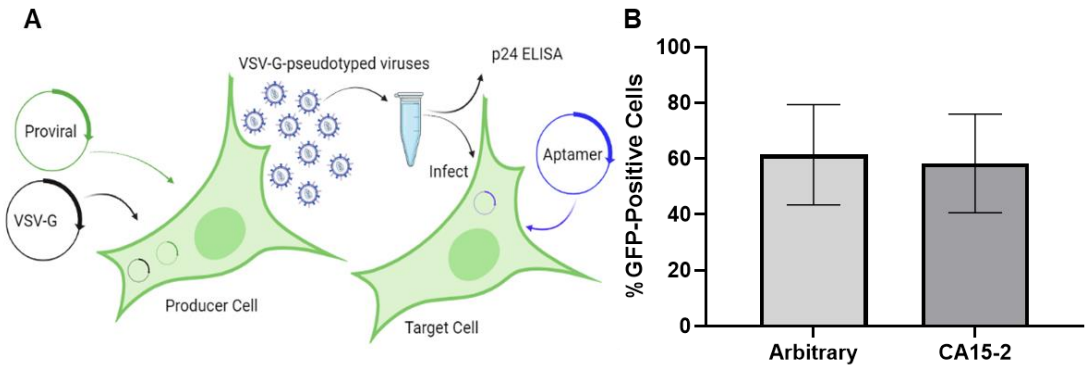

Figure 2.16. **Aptamer CA15-2 effects on HIV-1 replication in target cells.** (A) Schematic depiction of the target cell assay to evaluate aptamer effects on incoming virus created using BioRender.com. (B) Viral supernatants generated in the absence of aptamer were used to infect 293FT cells transfected for 48 hrs with plasmids for expression of the arbitrary control or CA15-2. Infectivity was determined by flow cytometry analysis of the number of EGFP-positive cells in each sample. Values are the mean ± SD for three experiments. Panel A was generated by MJL, and panel B was generated by RA.

There are many possibilities for the mechanism underlying the antiviral activity of CA15-2, on which we speculate here. As the level of p24 in the viral supernatant was similar in aptamer CA15-2 and the control samples, the defect does not appear to manifest in altered production of viral particles. It is possible that aptamer CA15-2 may impact genome packaging, virus assembly, encapsidation of a necessary host factor, virus maturation, or capsid stability, all of which are important for virus infectivity and would impact the ability of the virus to perform downstream replication steps. Recent work has suggested that the immature lattice is formed via recruitment of Gag dimers to the site of assembly to form hexamers, which then assemble into the immature lattice (53). We have demonstrated here that aptamer CA15-2 binds specifically to CA lattice but not to CA monomer or CA hexamer assembly forms. If the assembling immature lattice is available and accessible within the cell during virus production, it is possible that aptamer CA15-2 interacts with the assembling immature lattice. This interaction could drive specific incorporation of CA15-2 into viral particles, where CA15-2 could redirect the assembly process, resulting in defective viral particles. Alternatively, aptamer CA15-2 could be non-specifically incorporated into particles, providing an opportunity to interact with the lattice during remodeling that occurs as part of viral maturation to redirect the assembly process, leading to the replication defect. Indeed, we have previously demonstrated that, at the dosage of aptamer expression plasmid utilized in these experiments, aptamers, including non-binding controls, can be non-specifically encapsidated into viral particles (90). Further experiments are needed to elucidate these potential mechanisms.

2.3.7. Conclusions on CA lattice-targeting aptamers as new tools for interrogating the role of CA in HIV-1 replication

Aptamers offer several advantages over small molecules and antibodies in that they can fold into a variety of different structures, can be selected to display increased target affinity and specialized binding characteristics, and be easily modified for a variety of applications. In the case of RNA aptamers, specifically, they can also be expressed in cells as RNA transcripts. These advantages allow aptamers to engage in a variety of binding mechanisms, to outcompete other binding interactions, to achieve proper cellular localization, and to aid in visualization of cellular processes. Recent developments in our understanding of HIV-1 assembly and CA structure, as well as the availability of stabilized CA lattice tubes, provided a unique opportunity for the identification of aptamers capable of selectively binding the CA lattice. The work presented here details the successful selection of aptamers that bind CA lattice. We further characterized aptamer CA15-2, which binds CA lattice, but not CA monomer or CA hexamer, suggesting that aptamer CA15-2 binds an interface that is present on CA lattice, but not on CA monomer or CA hexamer. We found that aptamer CA15-2 was able to inhibit viral replication when expressed in producer cells, but not when expressed in target cells, opening the door to further examination of the mechanisms underlying viral inhibition by aptamer CA15-2. Aptamer CA15-2 and other aptamer clones from the same aptamer population represent new tools for discrimination of CA lattice from other CA assembly forms. Importantly, aptamers that recognize accessible sites specific to CA lattice or other CA assembly forms provide an opportunity to investigate impact of these interactions on diverse steps of HIV-1 replication, to further examine CA interactions with host factors,

and to identify potentially novel interaction surfaces present on CA assembly forms, all of which could increase our understanding of viral biology and inform development of therapeutics.

2.4. Materials and Methods

2.4.1. Reagents

Unless otherwise noted, all chemicals were purchased from Sigma-Aldrich (St. Louis, MO). The 56N library for aptamer selection, DNA templates for aptamer truncations and all primers were ordered from Integrated DNA Technologies (Coralville, IA).

Plasmid constructs for expression and purification of soluble HIV-1 CA hexamer (pET11a-CA (A14C, E45C, W184A and M185A)), CA hexamer lattice (pET11a-CA (A14C and E45C)), and CA monomer (pET11a-CA) were kindly provided by Dr. Owen Pornillos and Dr. Barbie Ganser-Pornillos at the University of Virginia. For the purposes of the work described here, the CA hexamer lattice will be referred to as the CA lattice to avoid confusion with the soluble CA hexamer. The CMV-driven plasmid encoding aptamer or control RNA sequences flanked by stabilizing structures for expression in mammalian cells was developed previously (109). To insert anti-HIV-1 CA aptamers or controls into the expression plasmid, the aptamer expression plasmid was digested using *Eco*RI and *Apa*I (New England Biolabs, Ipswitch, MA) and purified using agarose gel extraction. DNA encoding aptamers or control RNAs was amplified by PCR using appropriate primers to enable InFusion Cloning into Stellar competent cells (Takara Bio USA, Inc., Mountain View, CA). Sequences of aptamers, control RNAs and

amplification primers are listed in Table 1. All plasmids were confirmed by DNA sequencing. The HIV-1_{NL3-4}-derived CMV-EGFP plasmid (pNL4-3-CMV-EGFP) used in single-cycle infectivity assays was kindly provided by Vineet Kewal Ramani (National Cancer Institute [NCI]-Fredrick) and has been previously described (109). The vesicular stomatitis virus glycoprotein (VSV-G)-expressing plasmid, pMD-G, was obtained from Invitrogen (Carlsbad, CA).

2.4.2. Biological Resources

The human cell lines 293FT (Invitrogen, Carlsbad, CA) and TZM-GFP (112) were maintained in standard culture media containing Dulbecco's Minimum Essential Medium (Sigma, St. Louis, MO), 10% FBS (Sigma, St. Louis, MO), 2 mM L-glutamine (Gibco, Life Technologies, Grand Island, NY), 1 mM non-essential amino acids (Gibco, Life Technologies, Grand Island, NY), and 1 mM sodium pyruvate (Gibco, Life Technologies, Grand Island, NY). The 293FT cells were cultured in 0.5 mg/mL G418 (Sigma, St. Louis, MO). All cell lines were maintained at 37°C in 5% carbon dioxide with splitting approximately three times per week.

2.4.3. Expression and purification of HIV-1 CA proteins and crosslinking of CA hexamers and CA lattice tubes

Purified, assembled HIV-1 CA lattice for the aptamer selection was kindly provided by Dr. Owen Pornillos and Dr. Barbie Ganser-Pornillos at the University of Virginia. Post-selection experiments were performed using newly purified proteins, which were expressed and purified as previously described (37,113). Briefly, the HIV-1

CA proteins were expressed by IPTG induction in *Escherichia coli* BL21(DE3) cells for 6–12 hr at 25°C. The monomeric p24 CA protein was purified using a Hi-Trap Q-Sepharose column after ammonium sulfate precipitation. Homogenously purified protein fractions were concentrated and reconstituted in 25 mM Tris-HCl (pH 8.1) and 40 mM NaCl, flash frozen and stored at -80°C.

Soluble CA hexamers (pET11a-CA (A14C/E45C/W184A/M185A)) and CA lattice tubes (pET11a-CA (A14C/E45C)) were assembled as previously described (34,37). Briefly, assembly was performed *in vitro* by sequential overnight dialysis. Pooled and concentrated fractions were dialyzed into assembly buffer (50 mM Tris, pH 8.1, 1 M NaCl, and 200 mM βME) followed by dialysis into storage buffer (20 mM Tris, pH 8.1, and 40 mM NaCl). The integrity of the assembled CA lattice tubes was verified using transmission electron microscopy (TEM) with assistance from the University of Missouri Electron Microscopy Core (Figure 2.1B). Diluted CA lattice was spotted onto grids and stained with uranyl acetate (34,37). Tubes were visualized on the JEOL JEM-1400 120kV TEM. The purity and integrity of monomeric p24, assembled soluble CA hexamers and CA lattice were confirmed by non-reducing SDS-PAGE (Figure 2.1C). Importantly, for the purposes of this study, all CA protein concentrations are reported in terms of the total CA monomer.

2.4.4. Assembly of Native Capsid Tubes

To determine the ability of compound 15 (114,115) to facilitate assembly of native CA at physiological salt concentrations, assembly of 200 μM native CA monomer into assembled native CA lattice tubes was initiated in 1X binding buffer by adding

compound 15 to a final concentration of 20 μM from a 100 μM stock containing 10% DMSO. Absorbance at 350 nm was monitored at 1 min intervals using a BioTek Synergy HT plate reader (BioTek Instruments, Winooski, VT) for 120 min. In addition, the absorbance at 350 nm for negative control samples containing only the capsid protein or the capsid protein and DMSO were also monitored.

To prepare native CA lattice tubes for nitrocellulose filter binding assays, 100 μM working stocks of compound 15 were prepared using a final concentration of 10% DMSO. Native CA lattice tubes were assembled by incubating 200 μM CA monomer and 20 μM compound 15 in 1X binding buffer at 37°C for 1 hour. Assembled tubes were pelleted by centrifugation at 16000 x g for 5 min. The supernatant was removed, and the pelleted tubes were washed with 1X binding buffer containing 20 μM compound 15. The centrifugation step was repeated, and the supernatant was removed. The native capsid tubes were resuspended in 1X binding buffer containing 20 μM compound 15 for subsequent binding assays.

2.4.5. CA Lattice Aptamer Selection

Prior to the first round of selection, the double-stranded 56N library transcription template was generated from the bottom strand of the 56N starting library in a 10 mL PCR. The top strand of the 56N starting library transcription template and all relevant primers are listed in Table 2.1. The total transcript length is 104 nucleotides. The selection was initiated using approximately 10^{15} unique sequences in a 1 mL transcription.

Run-off transcription reactions were performed using the Y639F mutant T7 RNA polymerase (116), *in vitro* transcription buffer (50 mM Tris-HCl pH 7.5, 15 mM $MgCl_2$, 5 mM DTT, and 2 mM spermidine), and 2 mM each NTP. Reactions were incubated at 37°C for a minimum of 4 hr and halted by adding an equal volume of denaturing gel loading buffer (90% formamide and 50 mM EDTA with trace amounts of xylene cyanol and bromophenol blue). RNAs were purified by denaturing polyacrylamide gel electrophoresis (6% TBE-PAGE, 8 M urea). Bands corresponding to the expected product sizes were visualized by UV shadow, excised, and eluted overnight in 300 mM sodium acetate (pH 5.4) at room temperature. Eluates were ethanol precipitated, resuspended in nuclease-free water and stored at -20°C until further use. RNA concentrations were determined using a NanoDropOne spectrophotometer (Thermo Fisher Scientific, Waltham, MA). For the first round of selection, 5 nmol of RNA (~3.0 x 10^{15} molecules) was used, followed by 1.5 nmol RNA (~9.0 x 10^{14} molecules) in subsequent rounds. Prior to each selection round, the RNA was refolded by heating to 90°C for 90 seconds followed by cooling to room temperature for 5 minutes. CA lattice solution was then added, followed by addition of 10X binding buffer to a final concentration of 1X (50 mM Tris-HCl [pH 7.5], 100 mM KCl, 50 mM NaCl, 1 mM $MgCl_2$). For the first round of selection, 1 nmol CA lattice was added to the refolded RNA libraries in a final volume of 2 mL (2500 nM RNA to 500 nM lattice). In subsequent rounds, 300 pmol CA lattice was added in a final volume of 1 mL (1500 nM RNA to 300 nM lattice). Binding reactions were incubated at 37°C for 15 min followed by partitioning of the bound versus unbound RNA species from the CA lattice by centrifugation at 16,000 x g for 10 min at 4°C. Following centrifugation, the supernatant

containing unbound RNA species was removed and the CA lattice pellet was washed with 1X binding buffer. The sample was then centrifuged again at 16,000 x g for 10 min at 4°C. The wash step was performed a total of two times. CA lattice-bound RNA was recovered using phenol/chloroform extraction and ethanol precipitation. The recovered RNA was reverse transcribed using ImProm-II Reverse Transcriptase (Promega, Madison, WI) and PCR amplified using Pfu DNA polymerase to generate the transcription template for the next round of selection.

2.4.6. Library Cloning into the Aptamer Expression Plasmid

The double-stranded DNA templates for round 8, 10, and 15 libraries were cloned into an aptamer expression cassette as described above (see Plasmids). The resulting plasmids were purified from overnight cultures grown from single colonies picked from ampicillin-containing agar plates. Sanger sequencing was performed at the University of Missouri Genomics Technology Core to determine individual aptamer sequences for each clone. Aptamers were named according to the round and clone number. For example, aptamer 15-2 was the sequence present in the second colony picked from the round 15 plate.

2.4.7. DNA Templates and RNA Transcription

RNA aptamers (Table 2.1) were generated for each clone described above for use in biochemical assays. Transcription templates were generated by PCR amplification of the aptamer expressing plasmids described above using Pfu DNA polymerase, the 56N forward primer to append the T7 promoter, and the reverse primer complimentary to the

3′ constant region. Amplicon size was confirmed by agarose gel electrophoresis. Amplicons were transcribed *in vitro* and purified as described above.

2.4.8. Nitrocellulose Filter Binding Assays

In vitro transcribed RNA was treated with Antarctic phosphatase (Fermentas, Waltham, MA) to remove the 5' terminal phosphate and then labeled by T4 polynucleotide kinase (New England Biolabs, Ipswitch, MA) in the presence of γ-^{32}P-labeled ATP (PerkinElmer, Waltham, MA). Radiolabeled RNAs were gel-purified using denaturing PAGE as described above. To evaluate dose-dependent binding of aptamer libraries or individual aptamers to CA proteins, 50 nM 5′-radiolabeled and refolded RNA was incubated with varying concentrations of CA lattice, CA hexamer, or CA monomer in 1X binding buffer at 37°C for 15 min. A separate aliquot of the same RNA was incubated with 0 nM CA protein to determine non-specific nucleic acid retention. RNA:protein complexes were partitioned from unbound RNA by filtering samples through a pre-wet, alkaline-treated nitrocellulose filter under vacuum and immediately washing with 1 mL binding buffer. Radioactivity retained on the filter was counted by placing filters into scintillation vials, adding 4 mL Emulsifier-safe liquid scintillation fluid (Perkin Elmer, Waltham, MA), and counting using a liquid scintillation counter. An unfiltered 'No Wash' sample was counted to determine the total amount of radioactivity present in each binding reaction. The fraction of RNA retained on the filter was calculated by dividing the radioactivity retained on the filter by the radioactivity present in the 'No Wash' sample. At least three replicates were performed for each binding assay. To decrease non-specific nucleic acid retention, nitrocellulose filters were pre-incubated

in 0.5 M KOH for 20 min, washed extensively with MilliQ water, and then incubated in 1X binding buffer for at least 45 min as previously described (117).

To determine the apparent dissociation constant (K_{Dapp}) using nitrocellulose filter binding assays, approximately 20,000 counts-per-minute (cpm) of 5′-radiolabeled and refolded RNA was incubated with varying concentrations of CA lattice (0 nM CA lattice to determine non-specific binding, or 10 to 7500 nM CA lattice) in 1X binding buffer at 37°C for 15 min. The nitrocellulose filter binding assay was then performed as described above. The values for the fraction of RNA retained on filter were fit to a 1:1 binding curve model ($Y = Bmax*X/[K_D+X]$) using GraphPad Prism 6.2. In the equation, Bmax is the maximum specific binding, K_D is the dissociation constant, X is the CA concentration (total monomer), and Y is the fraction of RNA retained on filter. Binding assays were performed in triplicate.

To determine whether aptamer CA15-2 retained binding to CA lattice when an antisense oligonucleotide was annealed to either its 5′ or 3′ constant region, 50 nM 5′-radiolabeled CA15-2 RNA and 62.5 nM antisense oligonucleotide were mixed on ice in 1X binding buffer. For thermal renaturation and annealing of the antisense oligonucleotide to the aptamer, the samples were transferred into a preheated aluminum insert within a dry heat block set to 90°C for 1 min, then the aluminum insert was removed from the heat block and placed on the benchtop and allowed to cool to 37°C. 2 μM CA lattice (or 0 nM CA lattice to determine non-specific binding) was then added to each reaction, and the reaction was incubated at 37°C for 15 min. The nitrocellulose filter binding assay was then performed as described above to determine fraction RNA retained on filter. Experiments were performed at least three times.

To evaluate the effect of IP6 on CA15-2 binding to CA lattice, 50 nM 5'-radiolabeled and refolded CA15-2 or arbitrary RNA was incubated with 1 μM CA lattice at varying concentrations of IP6 (0 to 200 nM IP6) in 1X binding buffer at 37°C for 15 min. The nitrocellulose filter binding assay was then performed as described above to determine the fraction RNA retained on filter. These experiments were performed in triplicate.

To evaluate the effect of high salt concentration on CA15-2 binding to CA lattice, 50 nM 5'-radiolabeled and refolded CA15-2 or arbitrary RNA was incubated in varying concentration of assembled CA lattice (0 to 1000 nM CA lattice) in high salt binding buffer (50mM Tris-HCl [pH 7.5], 100 mM KCl, 1 M NaCl, 1 mM $MgCl_2$) at 37°C for 15 min. Nitrocellulose filter binding assays were performed as described above in triplicate.

To evaluate the ability of CA15-2 to bind assembled native CA lattice tubes, nitrocellulose filter binding assays were performed in the presence of 20 μM compound 15 for stabilization of the native CA lattice tubes.

2.4.9. Size Exclusion Chromatography for the complex of the CA Lattice with Aptamers

To determine whether CA15-2 forms stable complex with CA lattice, 100 nM refolded RNA (CA15-2 or Arbitrary) was combined with 1000 nM of CA lattice and incubated in 1X binding buffer for 30 min. Size exclusion chromatography (SEC) analysis was then performed using a Superdex 200 Increase 10/300GL column. RNA alone and CA lattice alone were also analyzed to define the peaks corresponding to the

complex of RNA and the CA lattice. Experiments were performed in triplicate and representative curves of one experiment are shown in Figure 2.5A.

2.4.10. Dye conjugation to antisense oligos

The 56N reverse primer with a 5'-amino group attached to a C-6 alkyl chain (5' Amino Modifier C6) was purchased from Integrated DNA Technologies (Coralville, IA) and used in a dye conjugation reaction with Cy5-NHS ester (Lumiprobe, Hunt Valley, MD). The conjugation reaction, subsequent reverse-phase HPLC purification, and concentration/buffer exchange of the Cy5-labeled 56N reverse primer were performed as described previously (118).

2.4.11. Electrophoretic Mobility Shift Assays (EMSA) Competition Assays

2 μM CA lattice was preloaded with increasing concentrations of refolded, unlabeled competitor RNAs (25 to 200 nM) in 1X binding buffer at 37°C for 10 min. A separate sample with 0 nM competitor was prepared in parallel to determine the maximum binding of aptamer CA15-2. A 10X annealing reaction containing 500 nM CA15-2 and 450 nM Cy5-labeled 56N reverse primer in 1X binding buffer underwent the same thermal renaturation and annealing described. The annealed CA15-2:Cy5-labeled 56N reverse primer complex was then added to the binding reaction to a 1X final concentration (50 nM CA15-2 and 45 nM Cy5-labeled 56N reverse primer) and incubated at 37°C for 15 min. 4X native loading dye (2X tris-borate [TB], 50% glycerol) was added to each reaction to a final concentration of 1X, and the samples were run on a 2% 0.5X TB-agarose gel at 8V/cm. Gels were scanned for Cy5 fluorescence using the

Typhoon FLA 9000 phosphoimager (GE Healthcare, Chicago, IL). The fraction of CA15-2 annealed to Cy5-labeled oligo retained in wells relative to No Competitor was quantified using Multi Gauge software (Fujifilm) using the following equation: (intensity of signal retained in well – intensity of signal in well of RNA only lane)/(intensity of signal retained in well of the No Competitor lane – intensity of signal in well of RNA only lane). Experiments were performed in triplicate.

2.4.12. Nitrocellulose Filter Binding Competition Assays

To determine whether an unlabeled competitor was able to compete with 5′-radiolabeled CA15-2 for binding to the CA lattice, varying concentrations of unlabeled competitor (2.5 to 160 nM) were incubated with 250 nM CA lattice in 1X binding buffer at 37°C for 10 min. A separate sample with 0 nM competitor was prepared in parallel to determine the maximum binding of aptamer CA15-2. Unlabeled competitor RNAs (CA15-2, Arbitrary, Scrambled CA15-2, or yeast tRNA [Invitrogen, Carlsbad, CA]) were first refolded as described above. PF74 (Sigma Aldrich, St. Louis, MO) working stocks were prepared using a final concentration of 10% DMSO. 10 nM 5′-radiolabeled and refolded CA15-2 was then added to the binding reaction, which was incubated at 37°C for 15 min. Nitrocellulose filter binding assays were performed and the fraction of radiolabeled CA15-2 retained on the filter was calculated as described above. To calculate the relative % radiolabeled CA15-2 retained on filter, the fraction of radiolabeled CA15-2 retained on filter at each concentration of competitor was divided by the fraction radiolabeled CA15-2 retained on filter when no competitor was present (0 nM) and multiplied by 100%. The assays were done in triplicate.

To determine whether compound 15 was able to compete with 5′-radiolabeled aptamer CA15-2 prior to performing binding assays using compound 15-stabilized, assembled native CA lattice tubes, 10 μM working stocks of compound 15 was prepared using a final concentration of 10% DMSO. 1 μM unlabeled CA15-2 RNA or compound 15 was incubated with 1 μM assembled CA lattice in 1X binding buffer for 10 min at 37°C. A separate sample with no competitor added was prepared in parallel to determine the maximum binding of aptamer CA15-2 to CA lattice. 50 nM 5′-radiolabeled and refolded CA15-2 was then added to the binding reaction, which was incubated at 37°C for 15 min. Nitrocellulose filter binding assays were performed as described and the fraction of radiolabeled CA15-2 retained on the filter was calculated as detailed above.

2.4.13. **Microscale Thermophoresis (MST)**

The binding affinities of aptamer CA15-2 or the arbitrary RNA control for CA lattice, CA hexamer, and CA monomer were determined by measuring thermophoresis of Cy5-labeled RNAs in the presence of increasing concentrations of each individual capsid protein. Cy5-labeleing of the RNAs was performed using the Label IT Nucleic Acid Labeling Kit according to the manufacturer's instructions (Mirus Bio, Madison, WI). Unreacted dye was removed using a Sephadex G-25 spin column provided with the kit. The ratio of RNA to dye was determined by measuring absorption at 260 nm for the RNA and 650 nm for the dye (molar absorbance: 250,000 $M^{-1}cm^{-1}$). An average of 5 dye molecules were incorporated per RNA. Prior to performing MST experiments, nitrocellulose filter binding assays were performed as described above using the Cy5-labeled RNAs to ensure that Cy5 labeling did not alter the observed binding phenotypes.

Reaction mixtures containing 100 nM Cy5-labeled aptamer CA15-2 or arbitrary RNA and increasing concentrations of the individual CA proteins (1–1000 nM) were prepared in binding buffer (50 mM Tris-HCl [pH 7.5], 100 mM KCl, 50 mM NaCl, 1 mM $MgCl_2$) containing 0.1% Pluoronic acid. Reaction mixtures were incubated for 30 min at room temperature and then loaded into capillaries. Thermophoresis was monitored on a Monolith NT.115 MST instrument (NanoTemper Technologies GmbH, Munich, Germany) at 20% LED power and high MST power with 20 sec MST-on time, and data were analyzed with MO.Affinity software (version 2.3) using a 1:1 binding model (Nano Temper Technologies, CA). The final data were plotted using GraphPad Prism (Version 6.0) (GraphPad Inc., La Jolla, CA). Experiments were performed at least three times.

2.4.14. **Producer Cell Assay**

For experiments evaluating the effects of aptamer CA15-2 on a single cycle of HIV-1 replication at the producer cell stage (90,109), 293FT cells were co-transfected with pNL4-3-CMV-GFP (250 ng), pMD-G (100 ng) and aptamer- or non-binding RNA control-expressing plasmids at the indicated dosages using polyethylenimine (PEI; 1 μL of 1 mg/mL PEI per μg DNA). After 48 hr, supernatants containing VSV-G-pseudotyped viruses produced in the presence of aptamers or controls were harvested. Cellular debris was removed by centrifugation and viral supernatants were frozen at -80˚C. Infectivity was determined by adding equal volumes of viral supernatant per sample to TZM-GFP cells such that the arbitrary RNA control infectivity levels were ~10%. Infected cells were collected 48 hr post-infection and fixed with 2% paraformaldehyde. The number of infected cells was determined by measuring the number of GFP-positive cells using an

Accuri C6 Flow Cytometer (BD Biosciences, Franklin Lakes, NJ). The number of GFP-positive cells was normalized to p24 ELISA values and to the arbitrary RNA control to determine the relative infectivity for each sample. Biological replicates for these experiments were performed at least three times and each experiment included two technical replicates.

2.4.15. Target Cell Assay

For experiments evaluating the ability of aptamer CA15-2 or the arbitrary RNA control to protect cells from infection by pseudotyped HIV-1 particles (90), viruses were first produced in the absence of aptamer or controls by co-transfecting 293FT cells in 10 cm dishes with pNL4-3-CMV-GFP (250 ng) and pMD-G (100 ng) complexed with PEI. Virus was harvested 48 hr post-transfection as described above. Separately, aptamer- or arbitrary RNA control-expressing plasmids (1000 ng) were transfected into fresh target 293FT cells using PEI. At 48 hr post-transfection, the aptamer- or control-expressing cells were challenged with virus produced in the absence of aptamer. Infected cells were harvested 48 hr later, fixed with 2% paraformaldehyde, and analyzed by flow cytometry for expression of EGFP as described above. Experiments were performed at least three times.

2.4.16. Data Availability/Sequence Data Resources

All aptamer sequences utilized in this work are listed in Table 2.1. The full sequence of the pcDNA3.1-based aptamer expression plasmid is available upon request.

2.4.17. Statistical Analysis

To determine whether there was a statistically significant difference between two sets of samples, P values were calculated using an unpaired t test computed by GraphPad Prism.

2.4.18. Web Sites/Data Base Referencing

Structure prediction throughout this work was performed using the NUPACK web server (www.nupack.org) (110).

2.5. References

1. Campbell, E.M. and Hope, T.J. (2015) HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat. Rev. Microbiol.*, **13**, 471-483.
2. Christensen, D.E., Ganser-Pornillos, B.K., Johnson, J.S., Pornillos, O. and Sundquist, W.I. (2020) Reconstitution and visualization of HIV-1 capsid-dependent replication and integration in vitro. *Science*, **370**, eabc8420.
3. Fassati, A. (2012) Multiple roles of the capsid protein in the early steps of HIV-1 infection. *Virus Res.*, **170**, 15-24.
4. Sundquist, W.I. and Krausslich, H.G. (2012) HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspec. Med.*, **2**, a006924.
5. Rihn, S.J., Wilson, S.J., Loman, N.J., Alim, M., Bakker, S.E., Bhella, D., Gifford, R.J., Rixon, F.J. and Bieniasz, P.D. (2013) Extreme genetic fragility of the HIV-1 capsid. *PLoS Pathog.*, **9**, e1003461.
6. Le Sage, V., Mouland, A.J. and Valiente-Echeverria, F. (2014) Roles of HIV-1 capsid in viral replication and immune evasion. *Virus Res.*, **193**, 116-129.
7. James, L.C. (2019) The HIV-1 capsid: more than just a delivery package. *Adv. Exp. Med. Biol.*, **1215**, 69-83.
8. Novikova, M., Zhang, Y., Freed, E.O. and Peng, K. (2019) Multiple roles of HIV-1 capsid during the virus replication cycle. *Virol. Sin.*, **34**, 119-134.
9. Forshey, B.M., von Schwedler, U., Sundquist, W.I. and Aiken, C. (2002) Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.*, **76**, 5667-5677.

10. Eschbach, J.E., Elliott, J.L., Li, W., Zadrozny, K.K., Davis, K., Mohammed, S.J., Lawson, D.Q., Pornillos, O., Engelman, A.N. and Kutluay, S.B. (2020) Capsid lattice destabilization leads to premature loss of the viral genome and integrase enzyme during HIV-1 infection. *J. Virol.*, **95**, e00984-20.
11. Rossi, E., Meuser, M.E., Cunanan, C.J. and Cocklin, S. (2021) Structure, function, and interactions of the HIV-1 capsid protein. *Life (Basel)*, **11**, 100.
12. Summers, B.J., Digianantonio, K.M., Smaga, S.S., Huang, P.T., Zhou, K., Gerber, E.E., Wang, W. and Xiong, Y. (2019) Modular HIV-1 capsid assemblies reveal diverse host-capsid recognition mechanisms. *Cell Host Microbe*, **26**, 203-216.
13. Naghavi, M.H. (2021) HIV-1 capsid exploitation of the host microtubule cytoskeleton during early infection. *Retrovirology*, **18**, 19.
14. Zhuang, S. and Torbett, B.E. (2021) Interactions of HIV-1 capsid with host factors and their implications for developing novel therapeutics. *Viruses*, **13**, 417.
15. Mascarenhas, A.P. and Musier-Forsyth, K. (2009) The capsid protein of human immunodeficiency virus: interactions of HIV-1 capsid with host protein factors. *FEBS J.*, **276**, 6118-6127.
16. Burdick, R.C., Delviks-Frankenberry, K.A., Chen, J., Janaka, S.K., Sastri, J., Hu, W.S. and Pathak, V.K. (2017) Dynamics and regulation of nuclear import and nuclear movements of HIV-1 complexes. *PLoS Pathog.*, **13**, e1006570.
17. Achuthan, V., Perreira, J.M., Ahn, J.J., Brass, A.L. and Engelman, A.N. (2019) Capsid-CPSF6 interaction: master regulator of nuclear HIV-1 positioning and integration. *J. Life Sci. (Westlake Village)*, **1**, 39-45.

18. Francis, A.C., Marin, M., Prellberg, M.J., Palermino-Rowland, K. and Melikyan, G.B. (2020) HIV-1 uncoating and nuclear import precede the completion of reverse transcription in cell lines and in primary macrophages. *Viruses*, **12**, 1234.

19. Selyutina, A., Persaud, M., Lee, K., KewalRamani, V. and Diaz-Griffero, F. (2020) Nuclear import of the HIV-1 core precedes reverse transcription and uncoating. *Cell Rep.*, **32**, 108201.

20. Blanco-Rodriguez, G. and Di Nunzio, F. (2021) The viral capsid: a master key to access the host nucleus. *Viruses*, **13**, 1178.

21. Dharan, A., Bachmann, N., Talley, S., Zwikelmaier, V. and Campbell, E.M. (2020) Nuclear pore blockade reveals that HIV-1 completes reverse transcription and uncoating in the nucleus. *Nat. Microbiol.*, **5**, 1088-1095.

22. Burdick, R.C., Li, C., Munshi, M., Rawson, J.M.O., Nagashima, K., Hu, W.S. and Pathak, V.K. (2020) HIV-1 uncoats in the nucleus near sites of integration. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 5486-5493.

23. Muller, T.G., Zila, V., Peters, K., Schifferdecker, S., Stanic, M., Lucic, B., Laketa, V., Lusic, M., Muller, B. and Krausslich, H.G. (2021) HIV-1 uncoating by release of viral cDNA from capsid-like structures in the nucleus of infected cells. *Elife*, **10**, e64776.

24. Li, C., Burdick, R.C., Nagashima, K., Hu, W.S. and Pathak, V.K. (2021) HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2019467118.

25. Zila, V., Margiotta, E., Turonova, B., Muller, T.G., Zimmerli, C.E., Mattei, S., Allegretti, M., Borner, K., Rada, J., Muller, B. *et al.* (2021) Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell*, **184**, 1032-1046.

26. Takemura, T. and Murakami, T. (2012) Functional constraints on HIV-1 capsid: their impacts on the viral immune escape potency. *Front. Microbiol.*, **3**, 369.

27. Grutter, M.G. and Luban, J. (2012) TRIM5 structure, HIV-1 capsid recognition, and innate immune signaling. *Curr. Opin. Virol.*, **2**, 142-150.

28. Sultana, T., Mamede, J.I., Saito, A., Ode, H., Nohata, K., Cohen, R., Nakayama, E.E., Iwatani, Y., Yamashita, M., Hope, T.J. *et al.* (2019) Multiple pathways to avoid beta interferon sensitivity of HIV-1 by mutations in capsid. *J. Virol.*, **93**, e00986-19.

29. Pornillos, O. and Ganser-Pornillos, B.K. (2019) Maturation of retroviruses. *Curr. Opin. Virol.*, **36**, 47-55.

30. Kucharska, I., Ding, P., Zadrozny, K.K., Dick, R.A., Summers, M.F., Ganser-Pornillos, B.K. and Pornillos, O. (2020) Biochemical reconstitution of HIV-1 assembly and maturation. *J. Virol.*, **94**, e01844-19.

31. Tsiang, M., Niedziela-Majka, A., Hung, M., Jin, D., Hu, E., Yant, S., Samuel, D., Liu, X. and Sakowicz, R. (2012) A trimer of dimers is the basic building block for human immunodeficiency virus-1 capsid assembly. *Biochemistry*, **51**, 4416-4428.

32. Worthylake, D.K., Wang, H., Yoo, S., Sundquist, W.I. and Hill, C.P. (1999) Structures of the HIV-1 capsid protein dimerization domain at 2.6 A resolution. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 85-92.

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

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

35. Byeon, I.J., Meng, X., Jung, J., Zhao, G., Yang, R., Ahn, J., Shi, J., Concel, J., Aiken, C., Zhang, P. *et al.* (2009) Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell*, **139**, 780-790.

36. von Schwedler, U.K., Stray, K.M., Garrus, J.E. and Sundquist, W.I. (2003) Functional surfaces of the human immunodeficiency virus type 1 capsid protein. *J. Virol.*, **77**, 5439-5450.

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

38. Meng, X., Zhao, G. and Zhang, P. (2011) Structure of HIV-1 capsid assemblies by cryo-electron microscopy and iterative helical real-space reconstruction. *J. Vis. Exp.*, **54**, 3041.

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

40. Gres, A.T., Kirby, K.A., KewalRamani, V.N., Tanner, J.J., Pornillos, O. and Sarafianos, S.G. (2015) X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science*, **349**, 99-103.

41. Cortines, J.R., Lima, L.M., Mohana-Borges, R., Millen Tde, A., Gaspar, L.P., Lanman, J.K., Prevelige, P.E., Jr. and Silva, J.L. (2015) Structural insights into the stabilization of the human immunodeficiency virus type 1 capsid protein by the cyclophilin-binding domain and implications on the virus cycle. *Biochim. Biophys. Acta*, **1854**, 341-348.

42. Wacharapornin, P., Lauhakirti, D. and Auewarakul, P. (2007) The effect of capsid mutations on HIV-1 uncoating. *Virology*, **358**, 48-54.

43. Yu, X., Wang, Q., Yang, J.C., Buch, I., Tsai, C.J., Ma, B., Cheng, S.Z., Nussinov, R. and Zheng, J. (2009) Mutational analysis and allosteric effects in the HIV-1 capsid protein carboxyl-terminal dimerization domain. *Biomacromolecules*, **10**, 390-399.

44. Manocheewa, S., Swain, J.V., Lanxon-Cookson, E., Rolland, M. and Mullins, J.I. (2013) Fitness costs of mutations at the HIV-1 capsid hexamerization interface. *PLoS One*, **8**, e66065.

45. Yufenyuy, E.L. and Aiken, C. (2013) The NTD-CTD intersubunit interface plays a critical role in assembly and stabilization of the HIV-1 capsid. *Retrovirology*, **10**, 29.

46. Hulme, A.E., Kelley, Z., Okocha, E.A. and Hope, T.J. (2015) Identification of capsid mutations that alter the rate of HIV-1 uncoating in infected cells. *J. Virol.*, **89**, 643-651.

47. Bayro, M.J., Chen, B., Yau, W.M. and Tycko, R. (2014) Site-specific structural variations accompanying tubular assembly of the HIV-1 capsid protein. *J. Mol. Biol.*, **426**, 1109-1127.

48. Lampel, A., Bram, Y., Ezer, A., Shaltiel-Kario, R., Saad, J.S., Bacharach, E. and Gazit, E. (2015) Targeting the early step of building block organization in viral capsid assembly. *ACS Chem. Biol.*, **10**, 1785-1790.

49. Lampel, A., Varenik, M., Regev, O. and Gazit, E. (2015) Hierarchical multi-step organization during viral capsid assembly. *Colloids Surf. B Biointerfaces*, **136**, 674-677.

50. Hulme, A.E., Kelley, Z., Foley, D. and Hope, T.J. (2015) Complementary assays reveal a low level of CA associated with viral complexes in the nuclei of HIV-1-infected cells. *J. Virol.*, **89**, 5350-5361.

51. Dismuke, D.J. and Aiken, C. (2006) Evidence for a functional link between uncoating of the human immunodeficiency virus type 1 core and nuclear import of the viral preintegration complex. *J. Virol.*, **80**, 3712-3720.

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

53. Tan, A., Pak, A.J., Morado, D.R., Voth, G.A. and Briggs, J.A.G. (2021) Immature HIV-1 assembles from Gag dimers leaving partial hexamers at lattice edges as potential substrates for proteolytic maturation. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2020054118.

54. Mattei, S., Glass, B., Hagen, W.J., Krausslich, H.G. and Briggs, J.A. (2016) The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science*, **354**, 1434-1437.

55. Perilla, J.R. and Gronenborn, A.M. (2016) Molecular architecture of the retroviral capsid. *Trends. Biochem. Sci.*, **41**, 410-420.

56. Mendonca, L., Sun, D., Ning, J., Liu, J., Kotecha, A., Olek, M., Frosio, T., Fu, X., Himes, B.A., Kleinpeter, A.B. *et al.* (2021) CryoET structures of immature HIV Gag reveal six-helix bundle. *Commun. Biol.*, **4**, 481.

57. Kong, J., Xu, B., Wei, W., Wang, X., Xie, W. and Yu, X.F. (2014) Characterization of the amino-terminal domain of Mx2/MxB-dependent interaction with the HIV-1 capsid. *Protein Cell*, **5**, 954-957.

58. Buffone, C., Schulte, B., Opp, S. and Diaz-Griffero, F. (2015) Contribution of MxB oligomerization to HIV-1 capsid binding and restriction. *J. Virol.*, **89**, 3285-3294.

59. Smaga, S.S., Xu, C., Summers, B.J., Digianantonio, K.M., Perilla, J.R. and Xiong, Y. (2019) MxB restricts HIV-1 by targeting the tri-hexamer interface of the viral capsid. *Structure*, **27**, 1234-1245.

60. Diaz-Griffero, F., Vandegraaff, N., Li, Y., McGee-Estrada, K., Stremlau, M., Welikala, S., Si, Z., Engelman, A. and Sodroski, J. (2006) Requirements for capsid-binding and an effector function in TRIMCyp-mediated restriction of HIV-1. *Virology*, **351**, 404-419.

61. Zhao, G., Ke, D., Vu, T., Ahn, J., Shah, V.B., Yang, R., Aiken, C., Charlton, L.M., Gronenborn, A.M. and Zhang, P. (2011) Rhesus TRIM5alpha disrupts the HIV-1 capsid at the inter-hexamer interfaces. *PLoS Pathog.*, **7**, e1002009.

62. Yang, H., Ji, X., Zhao, G., Ning, J., Zhao, Q., Aiken, C., Gronenborn, A.M., Zhang, P. and Xiong, Y. (2012) Structural insight into HIV-1 capsid recognition by rhesus TRIM5alpha. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 18372-18377.

63. Kovalskyy, D.B. and Ivanov, D.N. (2014) Recognition of the HIV capsid by the TRIM5alpha restriction factor is mediated by a subset of pre-existing conformations of the TRIM5alpha SPRY domain. *Biochemistry*, **53**, 1466-1476.

64. Li, Y.L., Chandrasekaran, V., Carter, S.D., Woodward, C.L., Christensen, D.E., Dryden, K.A., Pornillos, O., Yeager, M., Ganser-Pornillos, B.K., Jensen, G.J. *et al.* (2016) Primate TRIM5 proteins form hexagonal nets on HIV-1 capsids. *Elife*, **5**, e16269.

65. Selyutina, A., Bulnes-Ramos, A. and Diaz-Griffero, F. (2018) Binding of host factors to stabilized HIV-1 capsid tubes. *Virology*, **523**, 1-5.

66. Morger, D., Zosel, F., Buhlmann, M., Zuger, S., Mittelviefhaus, M., Schuler, B., Luban, J. and Grutter, M.G. (2018) The three-fold axis of the HIV-1 capsid lattice is the species-specific binding interface for TRIM5alpha. *J. Virol.*, **92**, e01541-17.

67. Ganser-Pornillos, B.K. and Pornillos, O. (2019) Restriction of HIV-1 and other retroviruses by TRIM5. *Nat. Rev. Microbiol.*, **17**, 546-556.

68. Price, A.J., Fletcher, A.J., Schaller, T., Elliott, T., Lee, K., KewalRamani, V.N., Chin, J.W., Towers, G.J. and James, L.C. (2012) CPSF6 defines a conserved capsid interface that modulates HIV-1 replication. *PLoS Pathog.*, **8**, e1002896.

69. Matreyek, K.A., Yucel, S.S., Li, X. and Engelman, A. (2013) Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. *PLoS Pathog.*, **9**, e1003693.

70. De Iaco, A., Santoni, F., Vannier, A., Guipponi, M., Antonarakis, S. and Luban, J. (2013) TNPO3 protects HIV-1 replication from CPSF6-mediated capsid stabilization in the host cell cytoplasm. *Retrovirology*, **10**, 20.

71. Bhattacharya, A., Alam, S.L., Fricke, T., Zadrozny, K., Sedzicki, J., Taylor, A.B., Demeler, B., Pornillos, O., Ganser-Pornillos, B.K., Diaz-Griffero, F. *et al.* (2014) Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 18625-18630.

72. Yamashita, M. and Engelman, A.N. (2017) Capsid-dependent host factors in HIV-1 infection. *Trends Microbiol.*, **25**, 741-755.

73. Rebensburg, S.V., Wei, G., Larue, R.C., Lindenberger, J., Francis, A.C., Annamalai, A.S., Morrison, J., Shkriabai, N., Huang, S.W., KewalRamani, V. *et al.* (2021) Sec24C is an HIV-1 host dependency factor crucial for virus replication. *Nat. Microbiol.*, **6**, 435-444.

74. Price, A.J., Jacques, D.A., McEwan, W.A., Fletcher, A.J., Essig, S., Chin, J.W., Halambage, U.D., Aiken, C. and James, L.C. (2014) Host cofactors and pharmacologic ligands share an essential interface in HIV-1 capsid that is lost upon disassembly. *PLoS Pathog.*, **10**, e1004459.

75. Dick, R.A., Zadrozny, K.K., Xu, C., Schur, F.K.M., Lyddon, T.D., Ricana, C.L., Wagner, J.M., Perilla, J.R., Ganser-Pornillos, B.K., Johnson, M.C. *et al.* (2018) Inositol phosphates are assembly co-factors for HIV-1. *Nature*, **560**, 509-512.

76. Mallery, D.L., Marquez, C.L., McEwan, W.A., Dickson, C.F., Jacques, D.A., Anandapadamanaban, M., Bichel, K., Towers, G.J., Saiardi, A., Bocking, T. *et al.*

(2018) IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *Elife*, **7**, e35335.

77. Jennings, J., Shi, J., Varadarajan, J., Jamieson, P.J. and Aiken, C. (2020) The host cell metabolite inositol hexakisphosphate promotes efficient endogenous HIV-1 reverse transcription by stabilizing the viral capsid. *mBio*, **11**, e02820-20.

78. Yu, A., Lee, E.M.Y., Jin, J. and Voth, G.A. (2020) Atomic-scale characterization of mature HIV-1 capsid stabilization by inositol hexakisphosphate (IP6). *Sci Adv*, **6**, eabc6465.

79. Dostalkova, A., Kaufman, F., Krizova, I., Vokata, B., Ruml, T. and Rumlova, M. (2020) In vitro quantification of the effects of IP6 and other small polyanions on immature HIV-1 particle assembly and core stability. *J. Virol.*, **94**, e00991-20.

80. Mallery, D.L., Kleinpeter, A.B., Renner, N., Faysal, K.M.R., Novikova, M., Kiss, L., Wilson, M.S.C., Ahsan, B., Ke, Z., Briggs, J.A.G. *et al.* (2021) A stable immature lattice packages IP6 for HIV capsid maturation. *Sci. Adv.*, **7**, eabe4716.

81. Rankovic, S., Ramalho, R., Aiken, C. and Rousso, I. (2018) PF74 reinforces the HIV-1 capsid to impair reverse transcription-induced uncoating. *J. Virol.*, **92**, e00845-18.

82. Singh, K., Gallazzi, F., Hill, K.J., Burke, D.H., Lange, M.J., Quinn, T.P., Neogi, U. and Sonnerborg, A. (2019) GS-CA compounds: first-in-class HIV-1 capsid inhibitors covering multiplegrounds. *Front. Microbiol.*, **10**, 1227.

83. Zhou, J., Price, A.J., Halambage, U.D., James, L.C. and Aiken, C. (2015) HIV-1 resistance to the capsid-targeting inhibitor PF74 results in altered dependence on host factors required for virus nuclear entry. *J Virol.*, **89**, 9068-9079.

84. Carnes, S.K., Sheehan, J.H. and Aiken, C. (2018) Inhibitors of the HIV-1 capsid, a target of opportunity. *Curr. Opin. HIV AIDS*, **13**, 359-365.

85. Cosnefroy, O., Murray, P.J. and Bishop, K.N. (2016) HIV-1 capsid uncoating initiates after the first strand transfer of reverse transcription. *Retrovirology*, **13**, 58.

86. Rankovic, S., Varadarajan, J., Ramalho, R., Aiken, C. and Rousso, I. (2017) Reverse transcription mechanically initiates HIV-1 capsid disassembly. *J. Virol.*, **91**, e00289-17.

87. Rankovic, S., Deshpande, A., Harel, S., Aiken, C. and Rousso, I. (2021) HIV-1 uncoating occurs via a series of rapid biomechanical changes in the core related to individual stages of reverse transcription. *J Virol.*, **95**, e00166-21.

88. Francis, A.C. and Melikyan, G.B. (2018) Single HIV-1 imaging reveals progression of infection through CA-dependent steps of docking at the nuclear pore, uncoating, and nuclear transport. *Cell Host Microbe*, **23**, 536-548 e536.

89. Alam, K.K., Chang, J.L., Lange, M.J., Nguyen, P.D.M., Sawyer, A.W. and Burke, D.H. (2018) Poly-target selection identifies broad-spectrum RNA aptamers. *Mol. Ther. Nucleic Acids*, **13**, 605-619.

90. Lange, M.J., Nguyen, P.D.M., Callaway, M.K., Johnson, M.C. and Burke, D.H. (2017) RNA-protein interactions govern antiviral specificity and encapsidation of broad spectrum anti-HIV reverse transcriptase aptamers. *Nucleic Acids Res.*, **45**, 6087-6097.

91. Chen, L., Rashid, F., Shah, A., Awan, H.M., Wu, M., Liu, A., Wang, J., Zhu, T., Luo, Z. and Shan, G. (2015) The isolation of an RNA aptamer targeting to p53

protein with single amino acid mutation. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 10002-10007.

92. Hermann, T. and Patel, D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820-825.

93. Zichel, R., Chearwae, W., Pandey, G.S., Golding, B. and Sauna, Z.E. (2012) Aptamers as a sensitive tool to detect subtle modifications in therapeutic proteins. *PLoS One*, **7**, e31948.

94. Thirunavukarasu, D. and Shi, H. (2015) An RNA aptamer specific to Hsp70-ATP conformation inhibits its ATPase activity independent of Hsp40. *Nucleic Acid Ther.*, **25**, 103-112.

95. Moore, M.D., Bobay, B.G., Mertens, B. and Jaykus, L.A. (2016) Human norovirus aptamer exhibits high degree of target conformation-dependent binding similar to that of receptors and discriminates particle functionality. *mSphere*, **1**, e00298-16.

96. Kahsai, A.W., Wisler, J.W., Lee, J., Ahn, S., Cahill Iii, T.J., Dennison, S.M., Staus, D.P., Thomsen, A.R., Anasti, K.M., Pani, B. *et al.* (2016) Conformationally selective RNA aptamers allosterically modulate the beta2-adrenoceptor. *Nat. Chem. Biol.*, **12**, 709-716.

97. Weber, A.M., Kaiser, J., Ziegler, T., Pilsl, S., Renzl, C., Sixt, L., Pietruschka, G., Moniot, S., Kakoti, A., Juraschitz, M. *et al.* (2019) A blue light receptor that mediates RNA binding and translational regulation. *Nat. Chem. Biol.*, **15**, 1085-1092.

98. Kim, S.J., Kim, M.Y., Lee, J.H., You, J.C. and Jeong, S. (2002) Selection and stabilization of the RNA aptamers against the human immunodeficiency virus type-1 nucleocapsid protein. *Biochem. Biophys. Res. Commun.*, **291**, 925-931.

99. Ramalingam, D., Duclair, S., Datta, S.A., Ellington, A., Rein, A. and Prasad, V.R. (2011) RNA aptamers directed to human immunodeficiency virus type 1 Gag polyprotein bind to the matrix and nucleocapsid domains and inhibit virus production. *J. Virol.*, **85**, 305-314.

100. Lochrie, M.A., Waugh, S., Pratt, D.G., Jr., Clever, J., Parslow, T.G. and Polisky, B. (1997) In vitro selection of RNAs that bind to the human immunodeficiency virus type-1 gag polyprotein. *Nucleic Acids Res.*, **25**, 2902-2910.

101. Rose, K.M., Alves Ferreira-Bravo, I., Li, M., Craigie, R., Ditzler, M.A., Holliger, P. and DeStefano, J.J. (2019) Selection of 2'-deoxy-2'-fluoroarabino nucleic acid (FANA) aptamers that bind HIV-1 integrase with picomolar affinity. *ACS Chem. Biol.*, **14**, 2166-2175.

102. Li, N., Wang, Y., Pothukuchy, A., Syrett, A., Husain, N., Gopalakrisha, S., Kosaraju, P. and Ellington, A.D. (2008) Aptamers that recognize drug-resistant HIV-1 reverse transcriptase. *Nucleic Acids Res.*, **36**, 6739-6751.

103. Lai, Y.T. and DeStefano, J.J. (2012) DNA aptamers to human immunodeficiency virus reverse transcriptase selected by a primer-free SELEX method: characterization and comparison with other aptamers. *Nucleic Acid Ther.*, **22**, 162-176.

104. Tuerk, C., MacDougal, S. and Gold, L. (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 6988-6992.

105. Burke, D.H., Scates, L., Andrews, K. and Gold, L. (1996) Bent pseudoknots and novel RNA inhibitors of type 1 human immunodeficiency virus (HIV-1) reverse transcriptase. *J. Mol. Biol.*, **264**, 650-666.

106. Konopka, K., Lee, N.S., Rossi, J. and Duzgunes, N. (2000) Rev-binding aptamer and CMV promoter act as decoys to inhibit HIV replication. *Gene*, **255**, 235-244.

107. Yamamoto, R., Katahira, M., Nishikawa, S., Baba, T., Taira, K. and Kumar, P.K. (2000) A novel RNA motif that binds efficiently and specifically to the Tat protein of HIV and inhibits the trans-activation by Tat of transcription in vitro and in vivo. *Genes Cells*, **5**, 371-388.

108. Sayer, N., Ibrahim, J., Turner, K., Tahiri-Alaoui, A. and James, W. (2002) Structural characterization of a 2'F-RNA aptamer that binds a HIV-1 SU glycoprotein, gp120. *Biochem. Biophys. Res. Commun.*, **293**, 924-931.

109. Lange, M.J., Sharma, T.K., Whatley, A.S., Landon, L.A., Tempesta, M.A., Johnson, M.C. and Burke, D.H. (2012) Robust suppression of HIV replication by intracellularly expressed reverse transcriptase aptamers is independent of ribozyme processing. *Mol. Ther.*, **20**, 2304-2314.

110. Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M. and Pierce, N.A. (2011) NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.*, **32**, 170-173.

111. Porciani, D., Signore, G., Marchetti, L., Mereghetti, P., Nifosi, R. and Beltram, F. (2014) Two interconvertible folds modulate the activity of a DNA aptamer against transferrin receptor. *Mol. Ther. Nucleic Acids*, **3**, e144.

112. Rosa, A., Chande, A., Ziglio, S., De Sanctis, V., Bertorelli, R., Goh, S.L., McCauley, S.M., Nowosielska, A., Antonarakis, S.E., Luban, J. *et al.* (2015) HIV-1 Nef promotes infection by excluding SERINC5 from virion incorporation. *Nature*, **526**, 212-217.

113. Yoo, S., Myszka, D.G., Yeh, C., McMurray, M., Hill, C.P. and Sundquist, W.I. (1997) Molecular recognition in the HIV-1 capsid/cyclophilin A complex. *J. Mol. Biol.*, **269**, 780-795.

114. Wang, L., Casey, M.C., Vernekar, S.K.V., Sahani, R.L., Kirby, K.A., Du, H., Zhang, H., Tedbury, P.R., Xie, J., Sarafianos, S.G. *et al.* (2021) Novel PF74-like small molecules targeting the HIV-1 capsid protein: balance of potency and metabolic stability. *Acta. Pharm. Sin. B*, **11**, 810-822.

115. Wang, L., Casey, M.C., Vernekar, S.K.V., Sahani, R.L., Kankanala, J., Kirby, K.A., Du, H., Hachiya, A., Zhang, H., Tedbury, P.R. *et al.* (2020) Novel HIV-1 capsid-targeting small molecules of the PF74 binding site. *Eur. J. Med. Chem.*, **204**, 112626.

116. Sousa, R. and Padilla, R. (1995) A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J.*, **14**, 4609-4621.

117. McEntee, K., Weinstock, G.M. and Lehman, I.R. (1980) recA protein-catalyzed strand assimilation: stimulation by Escherichia coli single-stranded DNA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.*, **77**, 857-861.

118. Porciani, D., Cardwell, L.N., Tawiah, K.D., Alam, K.K., Lange, M.J., Daniels, M.A. and Burke, D.H. (2018) Modular cell-internalizing aptamer nanostructure enables targeted delivery of large functional RNAs in cancer cell lines. *Nat. Commun.*, **9**, 2283.

CHAPTER 3: HIGH-THROUGHPUT SEQUENCE ANALYSIS REVEALS RNA APTAMERS WITH DIFFERENT SPECIFICITIES FOR HIV-1 CAPSID ASSEMBLY STATES

This work will constitute most of the material that will be submitted for publication after further developments outlined at the end of the chapter. Anticipated co-authors include Rachna Aneja, Gilmarie Rodriguez Martinez, William Schulze, Donald Burke, and Margaret Lange.

3.1. Abstract

The mature HIV-1 capsid core has critical roles in several replication steps. It has a fullerene cone structure composed of approximately 250 capsid protein (CA) hexamers and 12 CA pentamers. Mutations that alter the relative stability of CA assembly states (i.e., hexamers or assembled lattice) cause severe infectivity defects. While there has been progress in understanding capsid's roles in HIV infectivity, there are limited tools available to differentiate among CA assembly states to evaluate their role in replication events. Starting with a pre-enriched aptamer library with affinity for the assembled CA lattice, we performed a differentiation selection to select for a subset of aptamers capable of either binding both lattice and hexamer or that are specific for lattice. The resultant aptamer libraries were subjected to high-throughput sequencing and bioinformatic analyses. Candidate aptamers were identified, and their binding specificities were predicted based on their enrichment profiles within the selection trajectories. The aptamers were screened for binding to CA lattice and hexamer, and 17 of the 23 aptamers (~74%) had binding specificities that matched their predictions. Importantly, we

identified aptamers that were lattice-specific or lattice and hexamer binders. Current work is focused on understanding the structural requirements for aptamer binding to CA lattice, and future work will determine the aptamer binding sites on the CA lattice.

3.2. Introduction

The HIV-1 capsid protein (CA) is involved in multiple steps within the viral replication cycle that lead to a successful infection, including reverse transcription, cytoplasmic trafficking, nuclear import, capsid uncoating, and integration(1-7). During viral assembly, as part of the Gag polyprotein, CA is responsible for assembling Gag into an immature lattice(8-11). Upon activation of the viral protease, Gag is cleaved into its component parts, and the CA proteins assemble into a mature capsid core, composed of approximately 250 CA hexamers and 12 CA pentamers(12). Therefore, multiple CA assembly states are present during the viral replication cycle, such as CA monomers, CA hexamers, CA pentamers, and the assembled CA lattice. To coordinate the steps within the replication cycle, the capsid core interacts with host factors(1,13), helps the virus evade host immune responses(14-16), and must maintain structural stability until capsid uncoating needs to occur(17). Mutations that alter capsid's structural stability or host factor interactions lead to disruptions in viral infection(17-19). It is unknown how the various CA assembly states contribute to capsid function in viral replication. Therefore, understanding the roles of capsid and its assembly states is important for HIV biology but also for the discovery of novel accessible sites on the capsid for targeting therapeutics.

Limited tools exist for differentiating CA assembly states to evaluate the roles of the different CA assembly states within the HIV-1 replication cycle. Therefore, we previously described an RNA aptamer selection against the assembled HIV-1 CA

lattice(20). Aptamers are useful reagents for probing the roles of similar proteins due to their unique binding properties, including proteins that differ by a single amino acid mutation(21-23) or protein conformation(24-29). They are selected through an *in vitro* evolution process known as SELEX (Systematic Evolution of Ligands by EXponential enrichment)(30,31). Therefore, we reasoned aptamers could be selected to differentiate among the available accessible surfaces present on the CA assembly states. Fifteen rounds of selection were performed against the *in vitro* assembled, stabilized CA lattice, that was used as surrogate for the CA lattice present within a viral core(32). We had cloned selection rounds 8, 10, and 15 into an aptamer-expression plasmid and had identified some aptamer clones using Sanger sequencing. One of the aptamers identified was CA15-2, which was found to specifically bind to the assembled CA lattice and not to the soluble CA hexamer or CA monomer. Furthermore, we observed that CA15-2 inhibited viral infectivity in producer cell assays but not target cell assays.

To determine whether we could identify subsets of aptamers within our pre-enriched aptamer population to the assembled CA lattice that either specifically bound the assembled CA lattice (i.e., CA15-2) or bound both lattice and hexamers, we performed a CA differentiation selection starting with the lattice round 15 library with three selection trajectories. In the first trajectory, we selected for sequences that bound both lattice and hexamer. In the second trajectory, we selected for sequences that were lattice-specific by performing negative selection rounds with CA hexamer. Because nitrocellulose filter retention was used as the partition method for the hexamer trajectories, we additionally did two rounds of selection against nitrocellulose to ensure that sequences that enriched in the lattice and hexamer trajectory were not nitrocellulose

binders. Aptamers populations from the original CA lattice selection and the three trajectories underwent high-throughput sequencing (HTS) and were analyzed using FASTAptamer(33) and a pre-release version of FASTAptameR 2.0(34) to identify candidate aptamers for further characterization. The combination of HTS and bioinformatic analyses allows for tracking how individual aptamer sequences performed during the selection trajectories. Based on their enrichment and/or depletion across the different trajectories, the binding specificities for lattice and hexamer were predicted and subsequently biochemically evaluated. The success rate between the predicted binding phenotype and experimentally determined binding phenotype was approximately 74%. This suggested that the strategy utilized here of a differentiation selection followed by HTS and bioinformatic analyses looking at enrichment profiles across different selection trajectories could be useful for future selections against other targets. While most aptamers were lattice-specific binders, we identified two aptamers capable of binding both lattice and hexamer. Next, we explored the sequence and/or structural requirements that are important for capsid binding for four representative aptamers using comparative sequence analyses and biochemical assays, and these studies provided initial insights on the structural features of these aptamers. Future work will determine the binding sites of the lattice-specific and lattice and hexamer binders.

3.3. Results and Discussion

3.3.1. CA Differentiation Selection

To identify a subset of aptamers from the CA lattice round 15 library that either specifically bind CA lattice or that bind both CA lattice and CA hexamer, we performed a differentiation selection using the pre-enriched CA lattice round 15 library as the starting

point. The protein targets for this differentiation selection were the assembled CA lattice tubes used in the original selection and the soluble CA hexamer. The assembled CA lattice is composed of cross-linked A14C/E45C CA hexamers, while the soluble CA hexamer proteins are composed of cross-linked A14C/E45C/W184A/M185A CA hexamers that are incapable of forming higher-order CA assemblies(32). The CA lattice and hexamer proteins were previously validated using transmission electron microscopy and SDS-PAGE(20)(see Figure 2.1). We previously had shown that the lattice round 15 aptamer library exhibited dose-dependent binding to CA lattice, but there was no significant dose-dependent binding to CA hexamer, suggesting that most of the aptamer population binds to epitopes uniquely present on the CA lattice(20).

The capsid differentiation selection consisted of three selection trajectories (Figure 3.1). The first trajectory enriched for sequences capable of binding both CA lattice and CA hexamer. For four rounds, positive selection rounds against CA hexamer and CA lattice were toggled, followed by five additional positive selection rounds against CA hexamer. Positive selection rounds using CA lattice used CA lattice pelleting via centrifugation as the partition method, as previously described(20). Nitrocellulose filter retention was used as the partition method for the positive selection rounds against the CA hexamer. In positive selection rounds against the CA hexamer, sequences capable of binding CA hexamer were kept while sequences not able to bind CA hexamer were discarded. The goal of the second trajectory was to deplete sequences capable of binding the CA hexamer from the aptamer population. Negative selection rounds against CA hexamer and positive selection rounds against CA lattice were toggled for five rounds. Therefore, a total of three CA hexamer depletion rounds were done, where CA hexamer

protein was in excess of the RNA aptamer library. As with the positive selection rounds against CA hexamer, nitrocellulose was used in the partitioning step. However, instead of recovering sequences retained on the nitrocellulose filter (i.e., hexamer-binding sequences), the non-binding sequences in the flow-through were recovered and amplified to be used in the next round. Because nitrocellulose was used for the partitioning steps for the first two trajectories, the third trajectory consisted of two rounds against nitrocellulose to decrease the likelihood of identifying false positive hits for CA hexamer binders. The binding reactions for this trajectory were prepared as before; however, no protein target was added the RNA aptamer library. The sample was applied to the nitrocellulose filter under vacuum, and sequences that were retained on the filter (i.e., nitrocellulose-binding sequences) were recovered and amplified for the next round.

Aptamer libraries were named according to the number of partition steps against the different selection targets and the type of selection step performed. For example, the lattice 17 + hexamer 7 (L17+H7) library underwent 17 total positive rounds against CA lattice and seven positive selection rounds against CA hexamer, while lattice 17 – hexamer 3 (L17-H3) library underwent 17 total positive selection rounds against CA lattice and three negative selection rounds against CA hexamer.

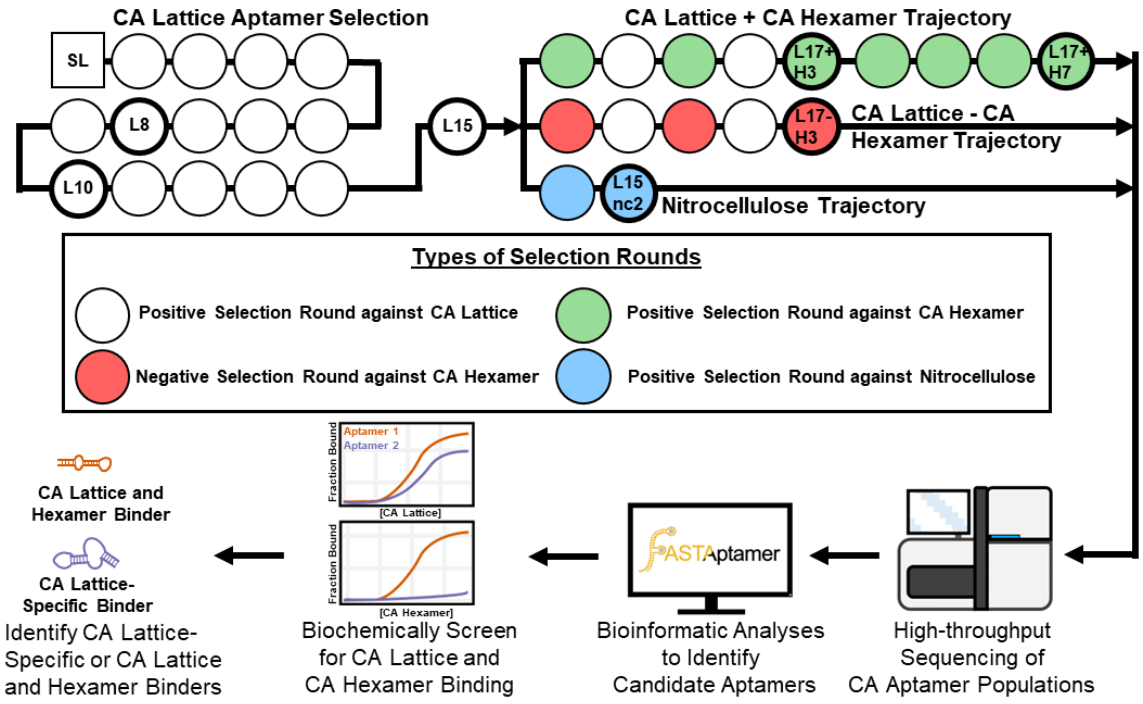

Figure 3.1: Schematic of Capsid Differentiation Selection. Previously, 15 rounds of selection against assembled CA lattice were performed(20). Starting from the lattice round 15 library, three selection trajectories were performed: the CA lattice and hexamer trajectory, the hexamer depletion trajectory, and the nitrocellulose trajectory. The seven libraries labeled (bold, thick outlines) were submitted for high-throughput sequencing, and candidate aptamers were identified for further characterization based on how they enriched and/or depleted in the selection trajectories. The candidate aptamers were screened for binding to CA lattice and hexamer to identify aptamers that were CA lattice-specific or CA lattice and hexamer binders.

3.3.2. HTS and Bioinformatic Analyses of CA Aptamer Libraries

Seven CA aptamer libraries were submitted for HTS: lattice rounds 8, 10, and 15; lattice 15 nitrocellulose 2 (L15nc2), lattice 17 + hexamer 3 (L17+H3), L17+H7, and L17-H3 (Figure 3.1). HTS data was preprocessed to merge paired end reads, trim off constant regions, and remove low quality or ambiguous reads. After data preprocessing, more than 34 million high-quality reads were generated from the seven aptamer libraries. The libraries were then analyzed using the FASTAptamer toolkit(33) and its R-based reimplementation FASTAptameR 2.0(34). First, the libraries were analyzed using FASTAptameR-Count to determine the rank, total read count, and the normalized sequence frequency in reads per million (RPM) for each unique sequence present in each library. The read counts and number of unique aptamer sequences for each library are shown in Table 3.1. Plotting the total number of reads against the rank of each aptamer within a library yields a line plot with a steep decline to a relatively horizontal line for all libraries (Figure 3.2), suggesting that the CA aptamer libraries have converged. In contrast, if a population that had not converged yet, a horizontal line would result. Comparing the number of unique sequences to the total reads sequenced for each population to calculate the percentage of unique sequences within a population (Table 3.1), the lattice round 15 population and the subsequent rounds performed in the CA differentiation selection are more converged than lattice rounds 8 and 10. This is further visualized by plotting the binned sequence abundance of each aptamer library (Figure 3.3). Based on the steepness present in lattice round 8 (Figure 3.2), this would suggest that the library had fully converged. However, almost half of the total reads in lattice round 8 come from unique sequences sampled once (i.e., singletons) (Figure 3.3). As the

rounds progressed from lattice round 8 to round 15, the percentage of reads that are from singletons decreased to less than 2%. Therefore, when determining if a library has converged, it is important to not only look at whether there are sequences that are abundant in the library but also at the overall sequence diversity present in the library. Finally, most aptamer reads had a sequence length of 56 nucleotides as expected (Figure 3.4), since the random region of the starting library for the CA lattice aptamer selection was 56 nucleotides. Although preprocessing allowed for sequences with lengths of 56 ± 6 nt, the actual distributions were much tighter. Therefore, minimal insertions or deletions occurred during the amplification steps (i.e., PCR and *in vitro* transcription) over the course of the selection.

Table 3.1: Read Counts from Capsid Aptamer HTS Libraries

Selection Round	Total Reads	Total Unique Sequences	% Unique Sequences
Lattice Round 8	6,118,235	2,916,568	47.7
Lattice Round 10	4,678,445	725,600	15.5
Lattice Round 15	5,449,375	150,347	2.8
Lattice 15 Nitrocellulose 2	3,802,822	99,969	2.6
Lattice 17 + Hexamer 3	4,730,617	119,760	2.5
Lattice 17 + Hexamer 7	3,756,752	64,662	1.7
Lattice 17 - Hexamer3	6,124,384	221,093	3.6

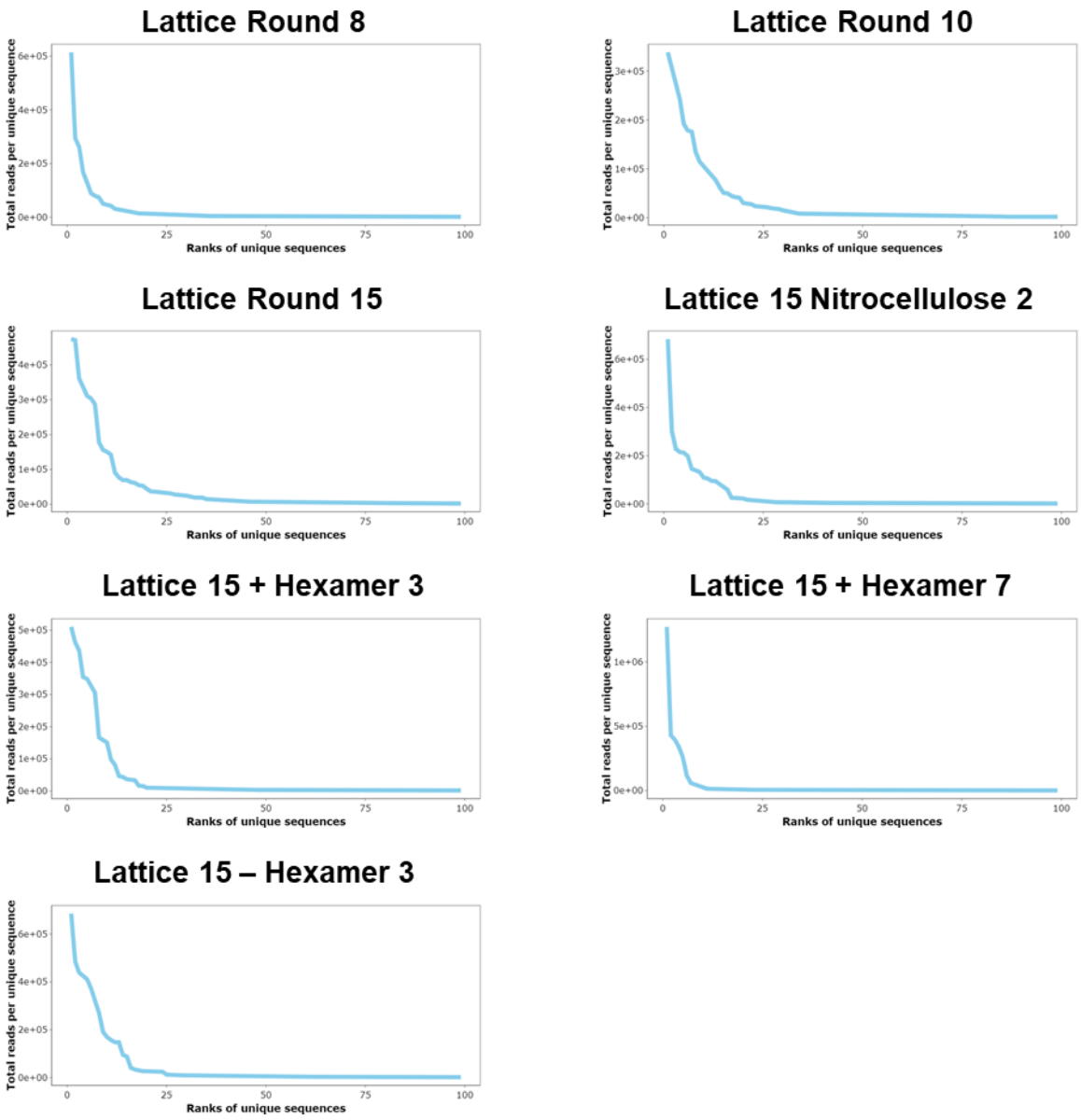

Figure 3.2: Reads per ranks of unique sequences within the CA aptamer libraries. The total read counts for the top 100 most abundant sequences within each specified CA aptamer library were plotted.

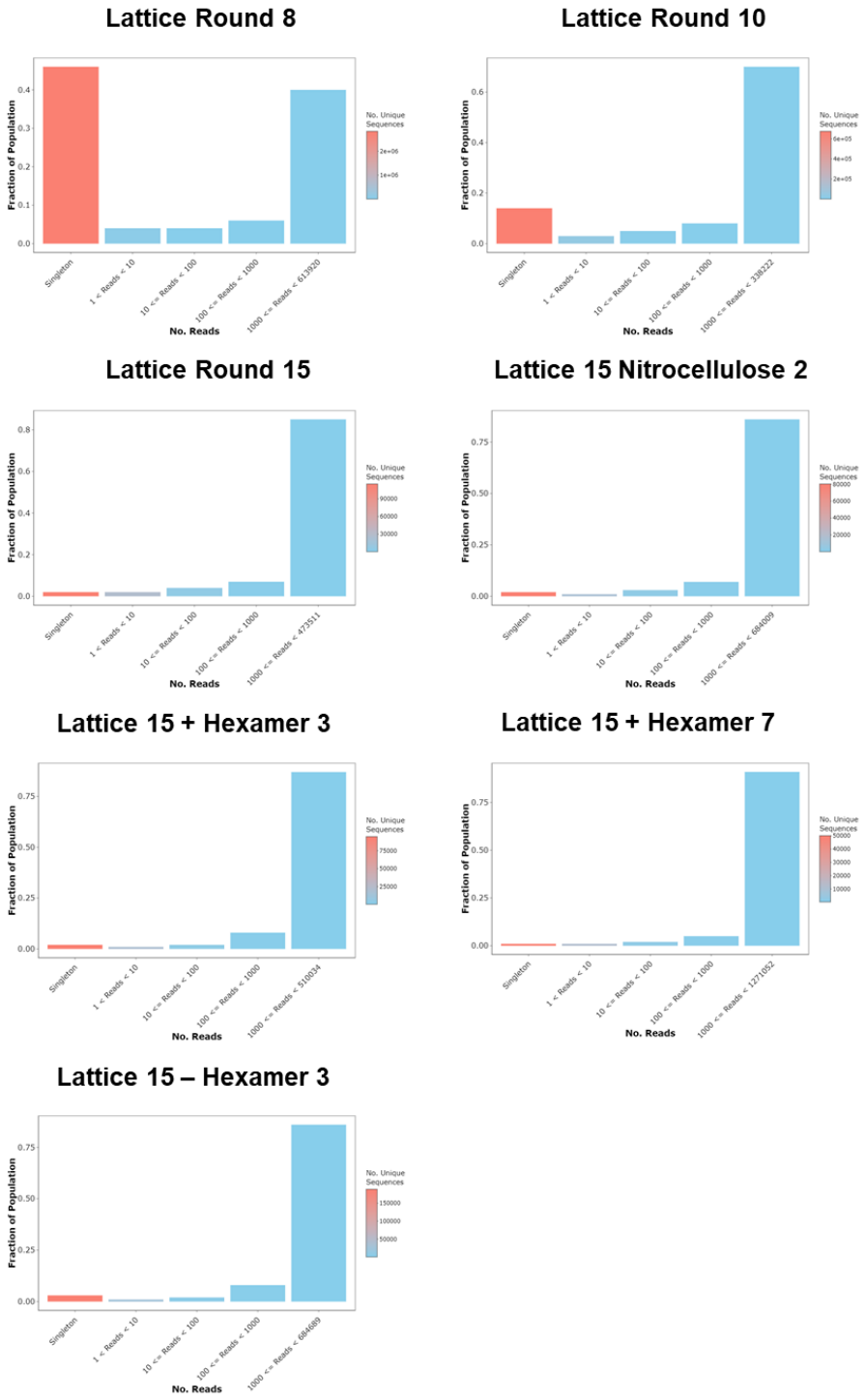

Figure 3.3: Binned sequence abundance plots for CA aptamer libraries. Unique sequences were binned according to their read counts for each library specified above. The color of the bar represents the number of sequences within each bin, and the height of the bar represents the fraction of the total read count of the library that contains the binned unique sequences.

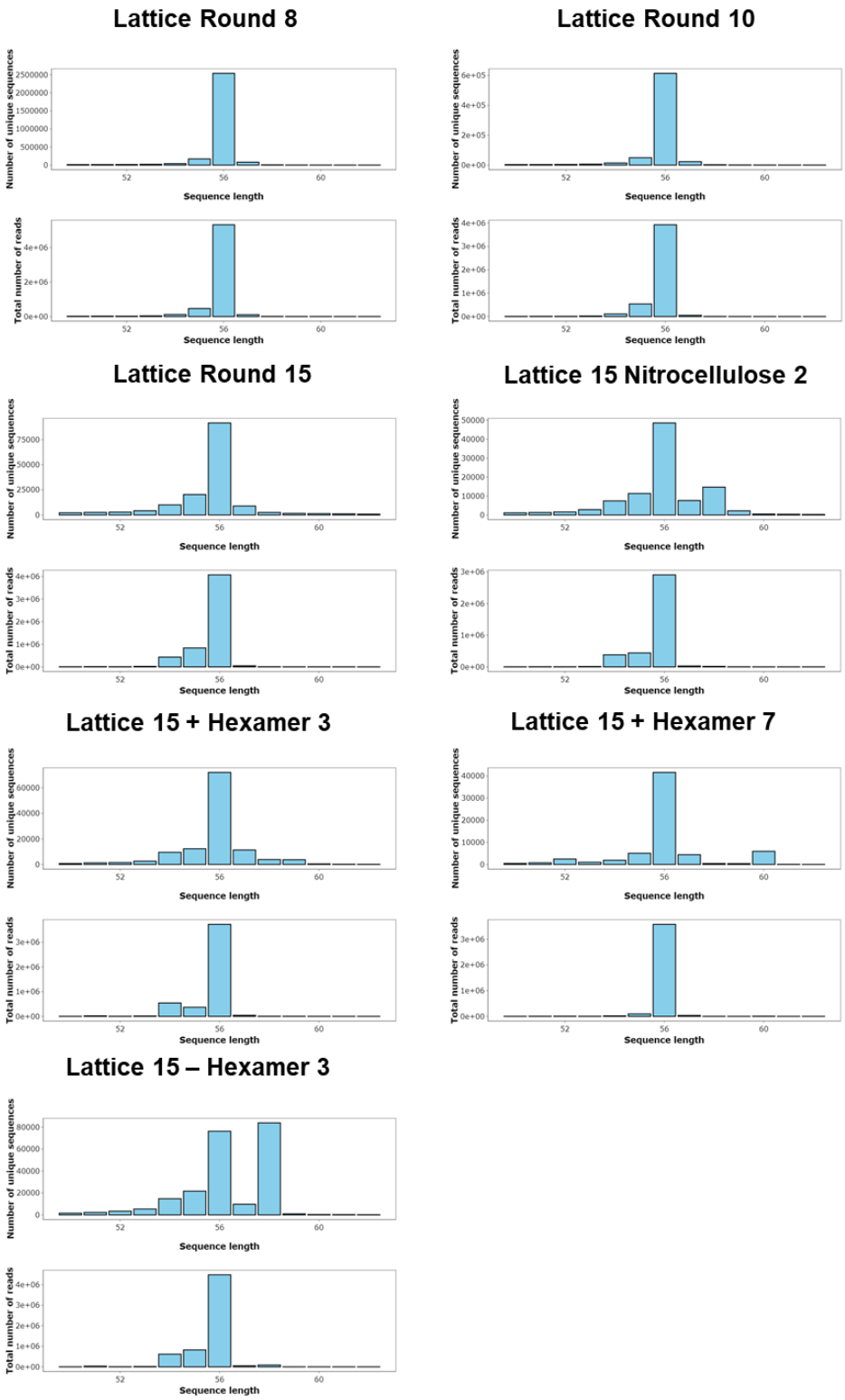

Figure 3.4: Sequence length histograms of the CA aptamer library. For each library specified above, the number of unique sequences and the total number or reads were plotted at each sequence length.

Sequences were next clustered using FASTAptameR-Cluster based on their Levenshtein edit distance (LED). The LED represents the number of insertions, deletions, and substitution required to convert one sequence into another. Sequences within a small LED from each other likely evolved from a parent sequence. Therefore, grouping related sequences simplifies the analysis, and we can determine how the cluster of related sequences performed during the selection trajectories. We utilized an edit distance of 6 for clustering, as most sequences had small edit distances relative to the cluster seed sequence (i.e., most abundant sequence within a cluster). This was visualized using the distance plot from FASTAptameR-Distance where the edit distance for each sequence read or unique sequence within the lattice round 15 library was calculated relative to the most abundant sequence within the library, aptamer L15.1.1 (Figure 3.5). These plots show a group of reads and sequences within a small edit distance of L15.1.1, suggesting that those sequences belong to its cluster, and a second group with edit distances between approximately 24 to 36, belonging to the remaining sequences that do not belong to its cluster.

The top one thousand clusters were generated for each of the aptamer libraries, except for lattice round 8 where the top one hundred clusters were generated due to the large number of unique sequences present in the library. The clustered sequence files were further analyzed with FASTAptameR-Cluster Diversity. For the top 100 clusters within each aptamer library, the number of unique sequences, number of reads, and the average LED of each sequence within a cluster were plotted (Figure 3.6). The total number of sequences per cluster and total reads per cluster plots demonstrate that the most abundant clusters tend to have the most unique sequences and the most reads, which

is expected since the seed sequences for these clusters are the most abundant sequences within the libraries. Clusters with more unique sequences also tend to have larger average LED relative to the seed sequence. However, this is more variable for clusters that have few unique sequences, which can be observed in the rapid drop-off in the L17+H7 library. This was likely due to the depletion of less abundant aptamers that were not able to bind to the CA hexamer during the five consecutive positive rounds against CA hexamer. Finally, the top ten clusters from each library were visualized using a k-mer-based principal component analysis (PCA) plots using a k-mer size of three from FASTAptameR-Cluster Diversity (Figure 3.7). These plots show that the individual sequences within clusters do cluster together, and furthermore, it suggests that some of the clusters may share sequence similarities as some of the clusters overlap.

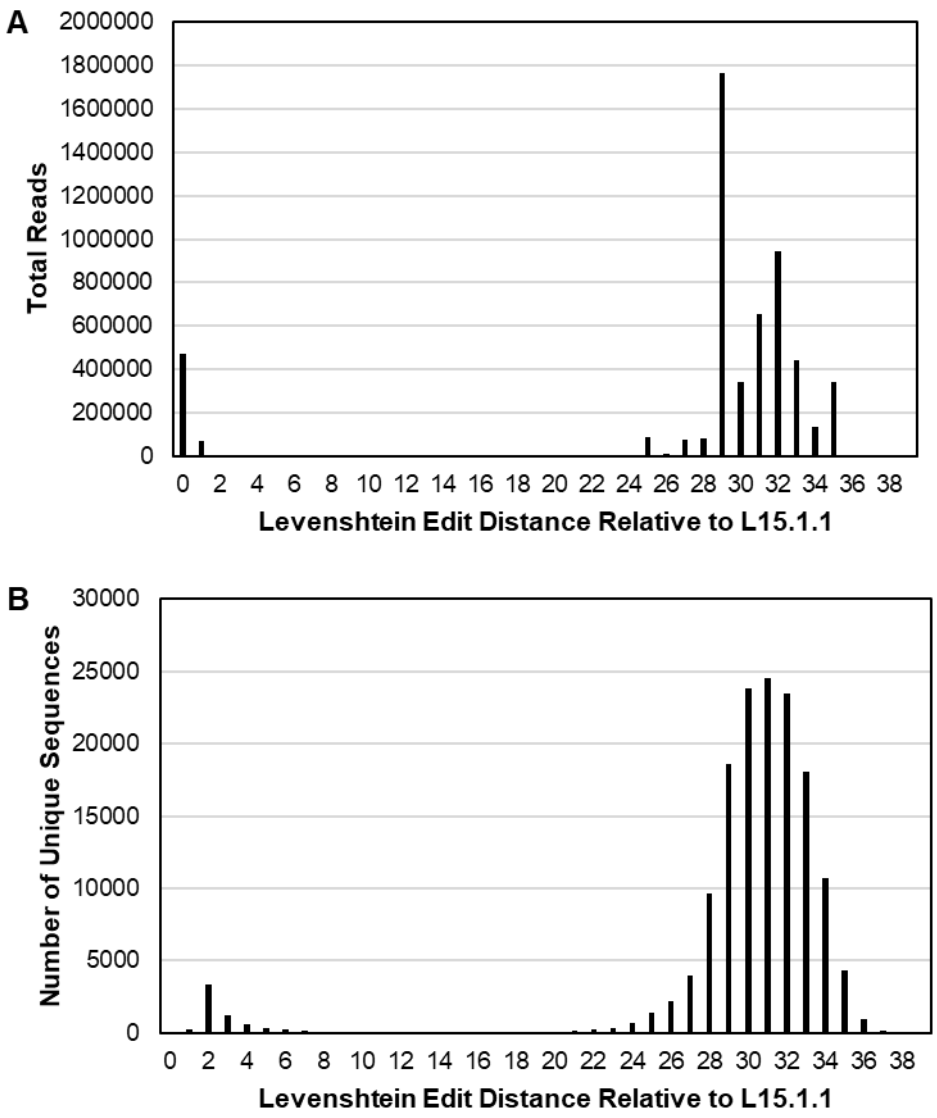

Figure 3.5: Levenshtein edit distance of lattice round 15 aptamer sequences to L15.1.1. (A) The total sequence reads and (B) the number of unique sequences were plotted according to their Levenshtein edit distance relative to aptamer L15.1.1, the most abundant aptamer within the library.

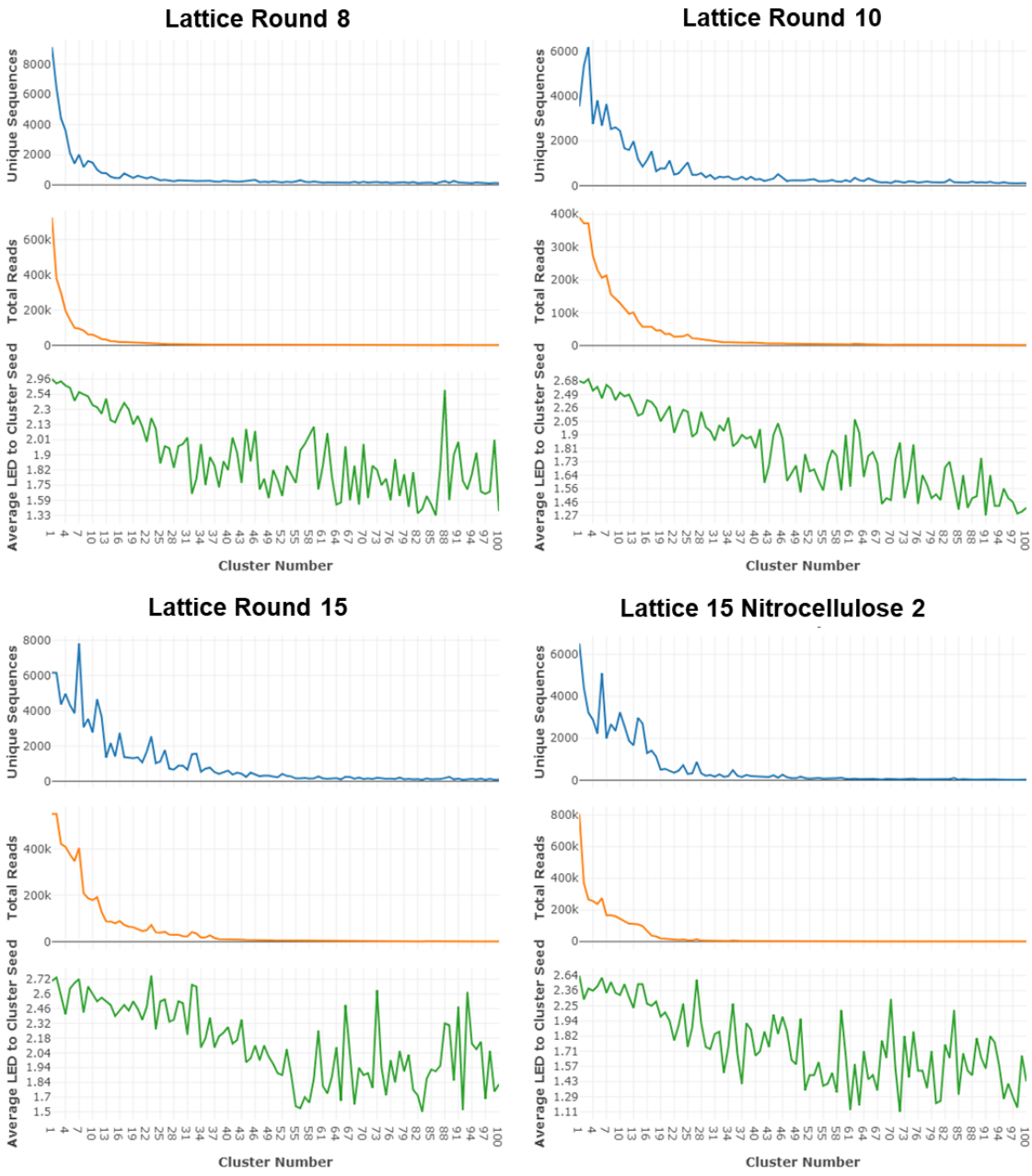

Figure 3.6 continues to next page.

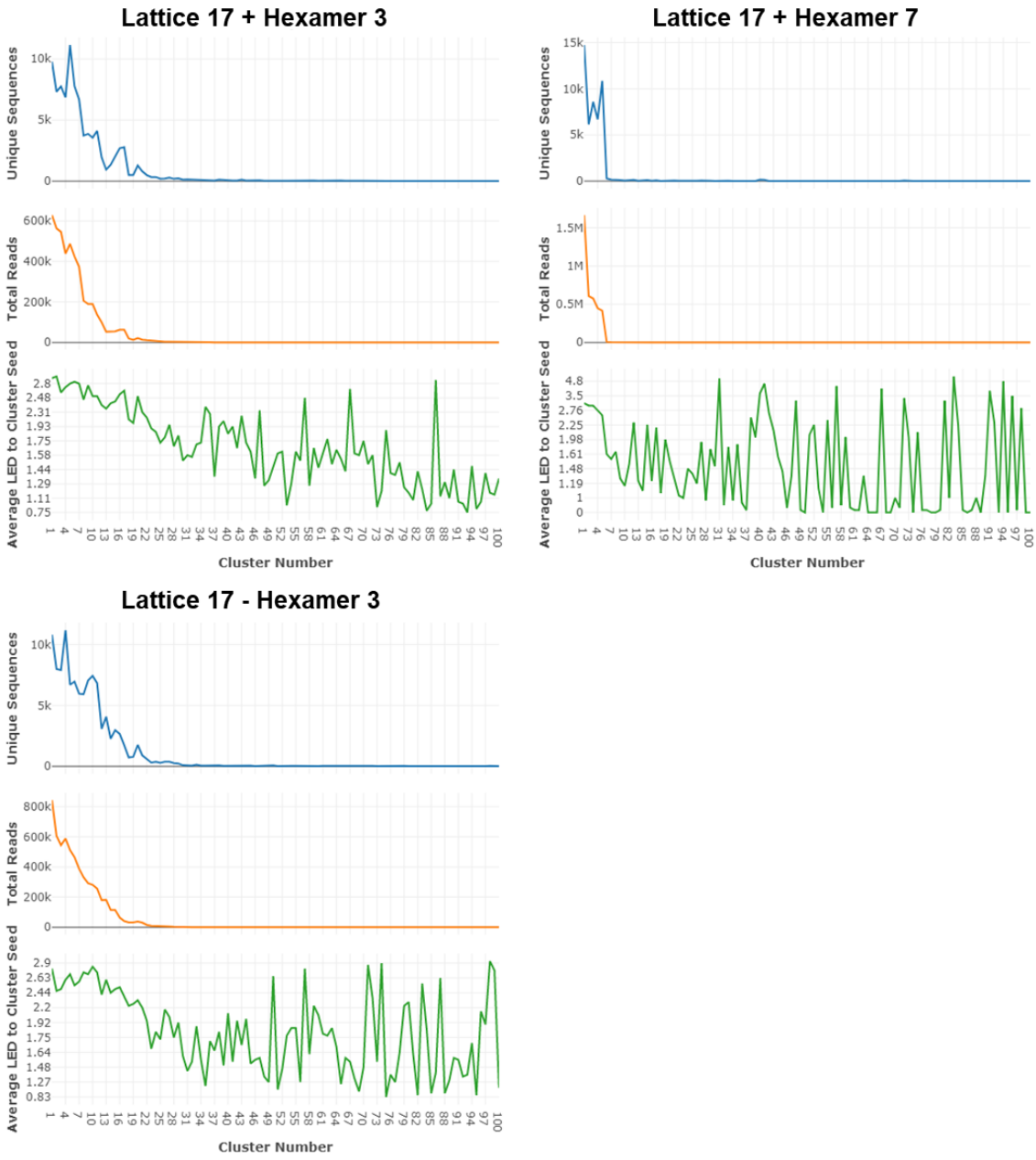

Figure 3.6: Cluster data metaplots for CA aptamer libraries. For each library specified above the plots, the number of unique sequences per cluster (top plot), the total reads per cluster (middle plot), and the average Levenshtein edit distance required to transform sequence into cluster seed sequence for the top 100 clusters (bottom plot) are shown.

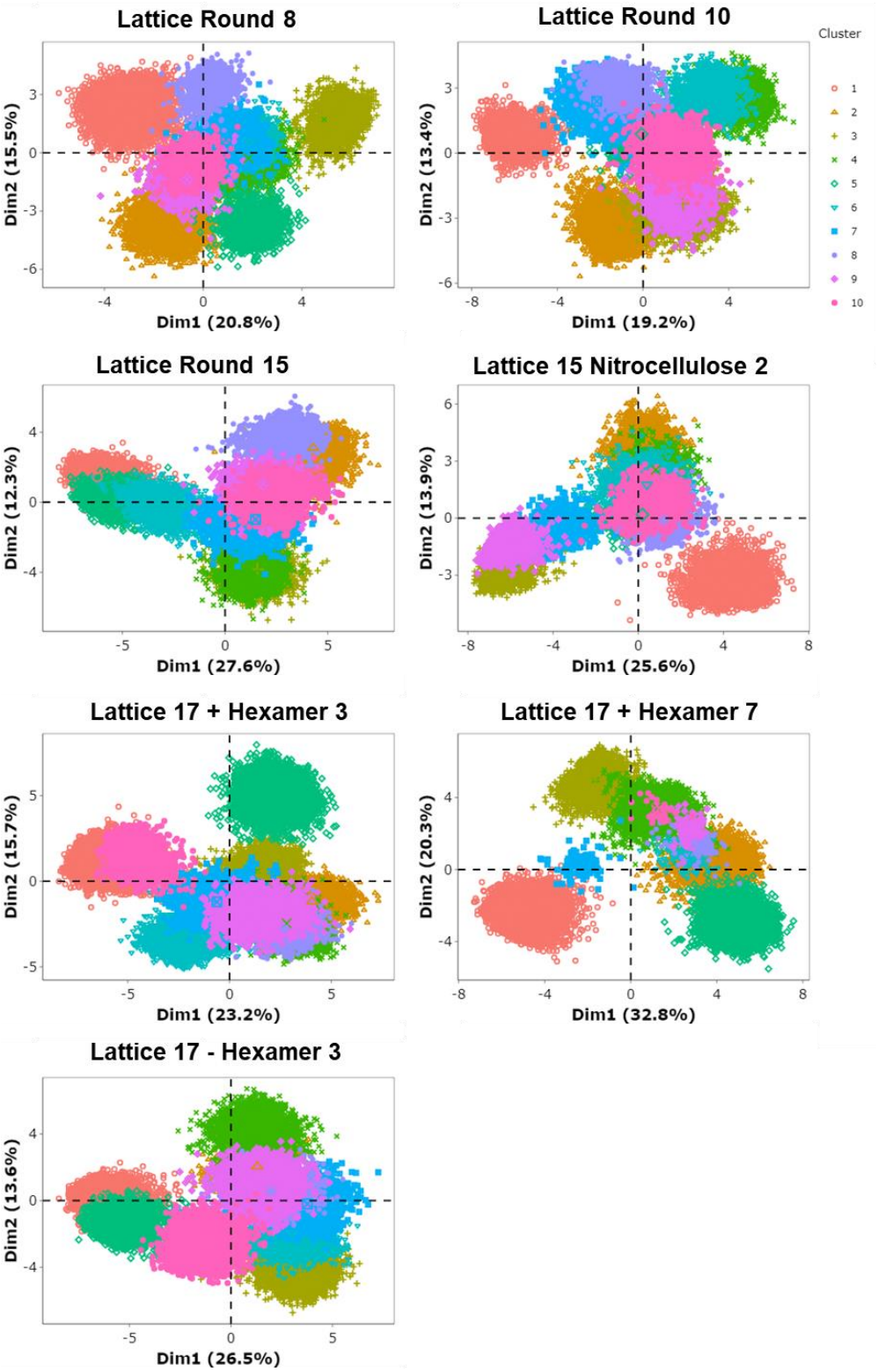

Figure 3.7: Cluster k-mer PCA plots for CA aptamer libraries. A PCA was done for the top 10 clusters for each aptamer library the plot using k = 3.

3.3.3. Enrichment Profiles of Aptamer Libraries

Next, we looked at how the individual sequences and clusters enriched or depleted within the different trajectories. Determining how a sequence or cluster enriched or depleted can inform how the sequence or cluster performed during the selection and possibly predict its binding affinity to its target of interest. Pairwise comparisons between aptamer libraries within trajectories were done to determine the enrichment values of individual sequences using FASTAptameR-Sequence Enrichment. Enrichment values for a given sequence are equal to the ratio of (RPM in round y [RPM_y]) to (RPM in round x [RPM_x]). Enrichment values can only be calculated for sequences that are present in both libraries being compared. Most sequences within the two libraries being compared were only present in one or the other (Figure 3.8). For the sequences that were present in both populations, we plotted the sequences based on their normalized frequency (RPM) in both populations (Figure 3.9). Sequences whose frequencies stayed approximately the same had points that fell on or near the $y = x$ line. In contrast, sequences that enriched or depleted were found to be above or below the $y = x$ line, respectively. Additionally, we binned the shared unique sequences based on their $\log_2$(Enrichment) (Figure3.10). Sequences that neither enriched nor depleted have an enrichment value of approximately 1, and their $\log_2$(Enrichment) is approximately 0. Therefore, enriched sequences are in the histograms to the right of the $\log_2$(Enrichment) = 0 histogram, and the depleted sequences are to the left of this histogram.

During the original lattice aptamer selection, we observed a shift in the number of sequences that were highly enriched as the selection progressed (Figure 3.9 A-C, Figure 3.10 A-C), suggesting that the composition of the lattice round 15 library contained more

sequences with higher fitness than rounds 8 or 10. For the nitrocellulose trajectory, we observed that the aptamer library generally had not changed from the lattice round 15 to round L15nc2, as most sequences neither enriched nor depleted (Figure 3.9 D, Figure 3.10 D). This suggests that most sequences do not bind nitrocellulose. For the positive selection trajectory against the CA hexamer, we observed a bimodal distribution of sequences, where some sequences had enriched (i.e., potential CA hexamer-binders) and others had depleted (i.e., potential CA lattice-specific binders) (Figure 3.9 E-G, Figure 3.10 E-G). It is interesting to compare how the aptamer library changed from lattice round 15 to round L17+H3 (Figure 3.9 E, Figure 3.10E) and from round L17+H3 to round L17+H7 (Figure 3.9F, Figure 3.10F). A clear decrease in the number of enriched sequences and increase in depleted sequences was observed in the L17+H7 vs L17+H3 comparison, compared to the L17+H3 vs lattice round 15 comparison. This suggests that the lattice-specific binders in the L17+H3 library depleted during the remaining four positive rounds against the CA hexamer. It is possible that the toggle between positive rounds against CA hexamer and CA lattice at the beginning of this selection trajectory allowed CA-lattice specific binders to initially persist in the library. For the CA hexamer depletion trajectory, we again saw a bimodal distribution of enriched and depleted sequences (Figure 3.9H, Figure 3.10H). For this trajectory, we expected any CA hexamer-binders to deplete, while CA lattice-specific binders should have sequence frequencies that enriched or stayed approximately the same.

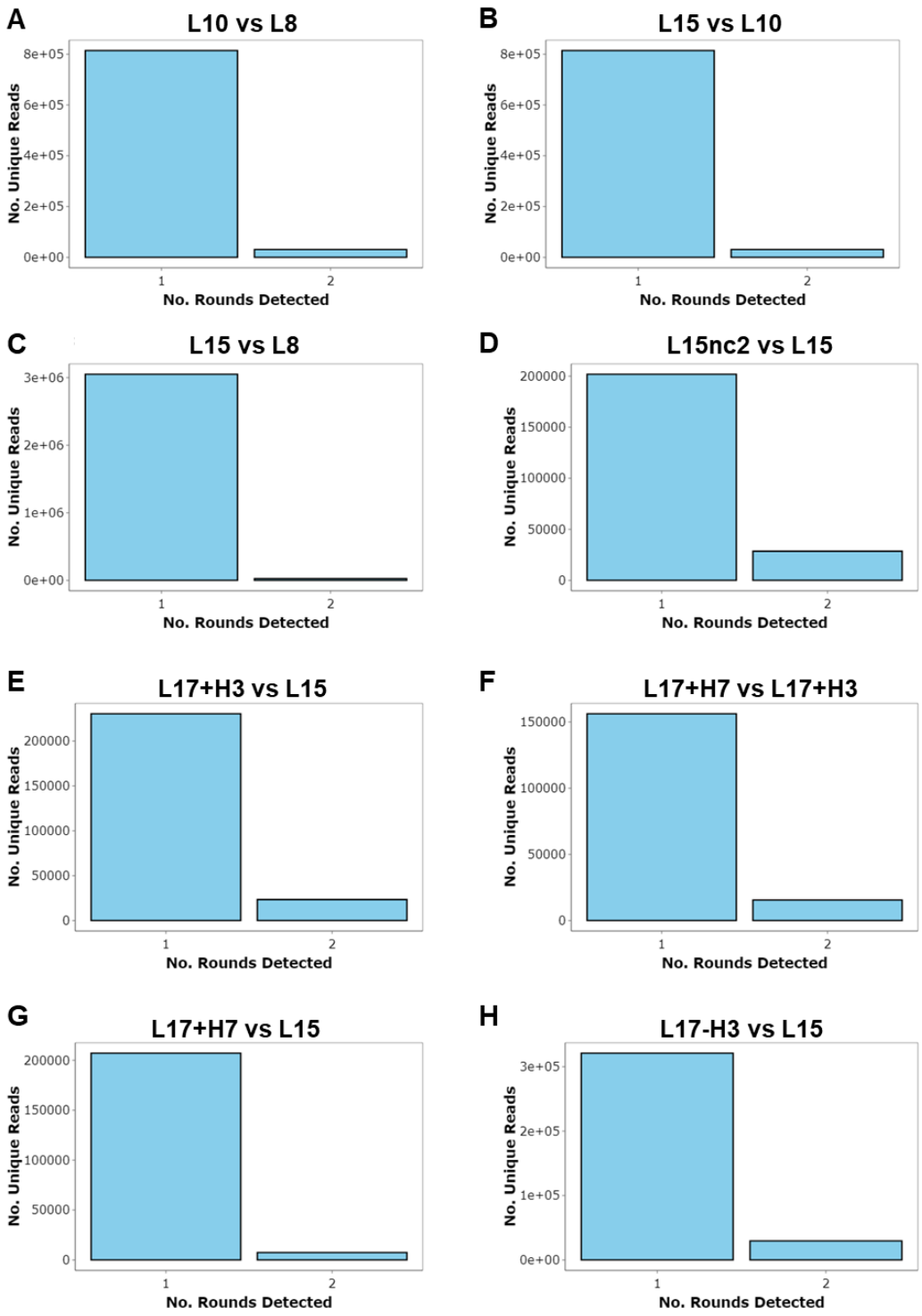

Figure 3.8: Sequence persistence between CA aptamer libraries. The number of unique sequences present in either one or both aptamer libraries were plotted.

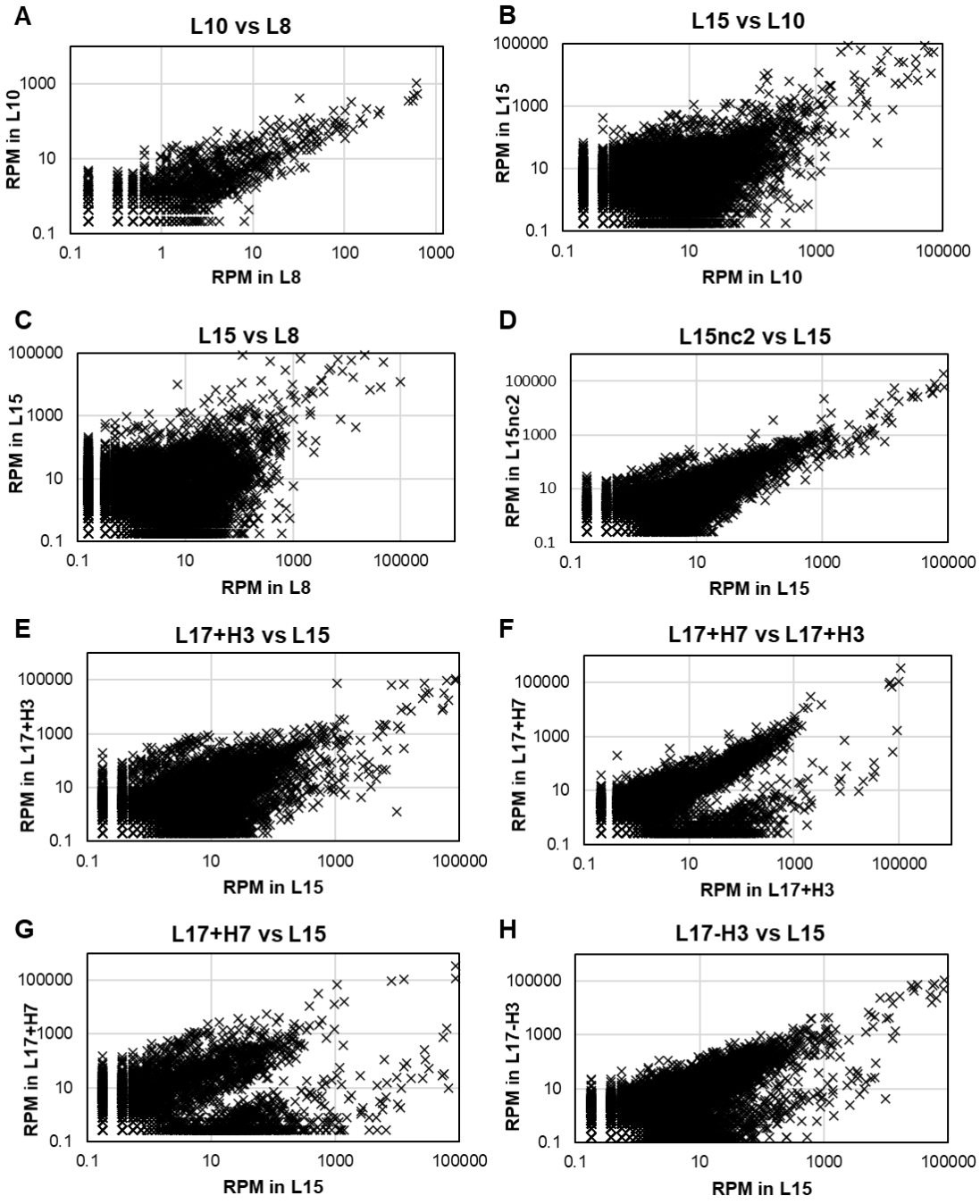

Figure 3.9: Comparison of sequence read frequencies among CA aptamer libraries. Sequences were plotted based on their read frequencies in both aptamer libraries in reads per million (RPM). Only sequences that were present in both libraries were plotted. Both axes are plotted on a logarithmic scale.

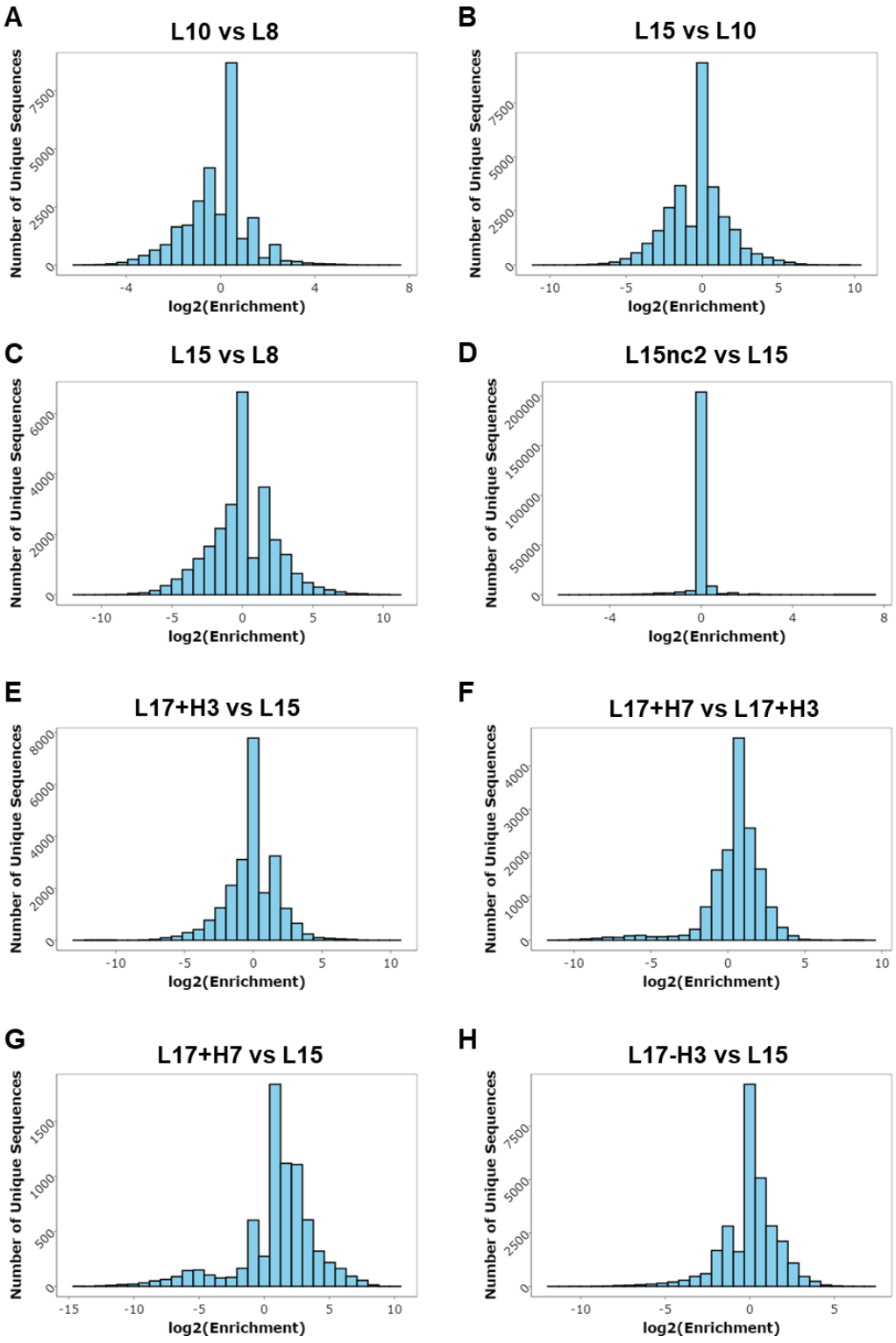

Figure 3.10: Distribution of enrichment values among shared sequences between two CA aptamer libraries. Unique sequences were binned according to the $\log_2$ of their enrichment value. Only sequences that were present in both libraries had calculated enrichment values.

Aptamer sequences were identified as candidates for a binding screen against CA lattice and CA hexamer based on how they and their cluster of related sequences enriched and/or depleted in the different selection trajectories. Because the pre-enriched lattice round 15 library was used as the branch point, we expected sequences identified as candidates to be capable of binding to the CA lattice. Sequences that depleted in the lattice and hexamer trajectory but did not significantly deplete in the hexamer depletion (CA lattice – CA hexamer) trajectory were predicted to be CA lattice-specific binders. In contrast, sequences that enriched in the lattice and hexamer trajectory and significantly depleted in the hexamer depletion trajectory were predicted to be aptamers that are capable of binding both CA lattice and hexamer. Because nitrocellulose filters were used as the partition method for the rounds using CA hexamer, the nitrocellulose trajectory was performed to limit the likelihood of identifying nitrocellulose-binding aptamers, instead of CA hexamer-binding aptamers. Aptamers predicted to be capable of binding CA hexamer were chosen if they did not also enrich significantly in the nitrocellulose trajectory. Aptamers were named according to their cluster number and rank in cluster in lattice round 15. If the aptamer was not present in lattice round 15, it was named according to its cluster number and rank in round L17+H7. For example, L15.6.1 was the seed sequence for cluster 6 within lattice round 15, while H7.10.1 was the seed sequence for cluster 10 in round L17+H7. Clusters were also named according to their cluster number in lattice round 15. If the cluster was not one of the top 1000 clusters in lattice round 15, then it was named according to its cluster number in round L17+H7.

The initial CA lattice selection consisted of fifteen rounds of selection against the assembled capsid lattice using pelleting as the partition method, as previously

described(20). Table 3.2 shows the ranks, RPM values, and enrichment values for the top 20 most abundant sequences in round 15. All these sequences are the seed sequences for their respective clusters. It is important to note that aptamer CA15-2 is aptamer L15.6.1. The variability in how the frequencies of these sequences enriched (>2), depleted (<0.5), or stayed approximately the same (0.5-2.0, 2-fold change) as the rounds progressed from round 8 to round 15 can be observed. Some sequences enriched earlier in the selection and began to deplete, while others enriched in the later rounds. A decrease in enrichment during this trajectory does not mean that the aptamer is a weaker binder to CA lattice. Rather, it is likely a consequence due to the amplification of other sequences with weaker binding within the population(35,36). We also determined the enrichment values for the aptamer clusters. For the first 20 aptamer clusters in round 15, the enrichment values for most clusters were either neutral (between 0.5 and 2, 2-fold change) or positive (>2) (Table 3.3). However, clusters 16 and 20 depleted from round 10 to round 15. Due to the time intensity of the cluster algorithm, only the top 100 clusters were determined for round 8, and the top 1000 clusters were determined for the other aptamer populations. Therefore, some cluster data were not determined for some of the aptamer clusters. It was expected that aptamer sequences with multiple reads in these libraries should bind the assembled lattice. We focused the candidate aptamers on sequences that were the seed sequences of their respective clusters with a sequence RPM of at least 10, a cluster RPM of at least 50, and both the sequence and cluster had not depleted more than 2-fold (i.e., enrichment value greater than 0.5 from L8 to L10, L10 to L15, or L8 to L15). Alternatively, if a sequence or cluster was not previously sampled in lattice rounds 8 or

10, such that an enrichment value could not be calculated, it was also considered. In total, 143 aptamers met these criteria and were on the short-list for the binding screen.

Table 3.2: Enrichment Values of the Top 20 Most Abundant Sequences in Lattice Round 15 Aptamer Library

Sequence	Rank (L8)	RPM (L8)	Rank (L10)	RPM (L10)	Rank (L15)	RPM (L15)	Enrichment (L10/L8)	Enrichment (L15/L10)	Enrichment (L15/L8)
GCCGATGCAGTCCCAATGTCGTATGAGATGTGA CTACCTACCACTGTTTAGTGTAT	247	114.1	30	3241.3	1	86619.7	28.4	26.7	759.4
CCACCCATGCCAAGGAAGGAGGTCGGAGGAGG TCACAGGCATTGGTGTGCTAGGAT	5	21001.1	4	51877.3	2	86094.7	2.5	1.7	4.1
CGCGCAAGCTTTCACGAGGGAGGTTGCGAGGGC GGATGTGGTTTGCGACGTCGTA	27	1356.4	23	5037.7	3	65992.2	3.7	13.1	48.7
CGTGCTCATACCGCCGAATACGAGGGAGGAGAG GGTGGCAAAACTGCGTGGGTT	11	6895.5	14	13413.6	4	61824.5	1.9	4.6	9.0
CCTACCTGTATGAAAGAGACGAACCTAACTACA ATGCTAGCAATGGTGTATGGTGC	8	11939.6	1	72124.4	5	56884.7	6.0	0.8	4.8
TCGACGTACCTCAGGGTGGTGTATGACTGAGGTG AAGACTGTGAACCATGGCATGC	79	368.6	32	2491.8	6	55461.3	6.8	22.3	150.5
CCAACCAGAGACGCACCAGAGTCCTCGAGGGA AAGGAGTGGGTGGTACGGGACATG	2	47829.1	3	59303.0	7	52362.8	1.2	0.9	1.1
CCATCCCCCAACTTGTCAAGAGGGAAGGAGTGG GAAAGGAAGACAAGGACATGTG	12	5077.6	6	38084.4	8	32558.7	7.5	0.9	6.4
CCCGGCCCCGCCAGTGGAAGTGGGAGGCATGA GGGAAAGGTTAAACACTGTCGGCT	36	718.4	16	10663.6	9	28572.3	14.8	2.7	39.8
GGACAGAACAACCCTTGAACGAATGAGGGACA GGTGCGAGGGAAGGATTCGTTGGA	13	4669.4	11	20524.0	10	27554.9	4.4	1.3	5.9
ACGCGAAAGCACGATCGGCACCACCTAATGTCA TGAGGGACGGAGTGAGGGAGGCA	9	8164.0	9	24472.1	11	26033.8	3.0	1.1	3.2
CAACACTAAAGTTCGACAACAACGGATATTGCC TACCAAGAGGATGGGTGTGTGGA	7	12974.3	7	37547.2	12	16679.8	2.9	0.4	1.3
TACCGCGACCTACCCCAGTGCAGGCTAGTGAGG GAGGAGGGTGGCATCTAGACACT	16	3301.7	18	9003.9	13	14128.1	2.7	1.6	4.3
ATCGGAACAAAACGCGACCCGGGACTAGACCTT ACCTAAGTGGATGGGTGTGTGGA	108	260.0	50	1123.8	14	12703.8	4.3	11.3	48.9
CGCCAAATTGCTCAGTCCAGTGCGGGAGGTGAG GGCGGAACAACATGACGGACCA	15	3451.7	15	10892.2	15	12692.8	3.2	1.2	3.7
GGCGAATTACACGACAATCTGGACGACCTACAC TATGAGAGGAGGGTAGAGTGGAA	1	99761.6	2	65497.3	16	11698.8	0.7	0.2	0.1
TGCGCGCGATGCCCGTACCATGCTATGTGGGTG GCGAGGGAGGACACTGCATGTTG	253	111.6	222	172.7	17	11208.0	1.5	64.9	100.4
CCCTGGAGACCACCAACAGAGGGAGGTGTGAG GGAAAGGATGATAGGTGCATGTCA	32	952.9	28	4048.5	18	10039.8	4.2	2.5	10.5
CCGTGCCCAGGGTCTTGCTGTATGTATGGTGCTG ATCTACTGTAGCCATGTGTATG	3152	7.2	243	152.9	19	9758.4	21.4	63.8	1364.8
GGGTCGCATCGGTCTCTATGGATAAATCTGCCTT ATTGGAGGATGGGAAATGTGGA	3	42712.5	8	28621.4	20	8157.8	0.7	0.3	0.2

Table 3.3: Enrichment Values of the Top 20 Clusters in Lattice Round 15 Aptamer Library

Seed Sequence	Cluster (L8)	Cluster Sequences (L8)	Cluster RPM (L8)	Cluster (L10)	Cluster Sequences (L10)	Cluster RPM (L10)	Cluster (L15)	Cluster Sequences (L15)	Cluster RPM (L15)	Enrichment (L10/L8)	Enrichment (L15/L10)	Enrichment (L15/L8)
GCCGATGCAGTCCCAATGTCGTATG AGATGTGACTACCTACCACTGTTTAG TGTAT				29	374	3646.3	1	6164	101176.9		27.7	
CCACCCATGCCAAGGAAGGAGGTC GGAGGAGGTCACAGGCATTGGTGTG CTAGGAT	5	2104	23564.8	4	2754	58433.7	2	6155	101148.1	2.5	1.7	4.3
CGCGCAAGCTTTCACGAGGGAGGTT GCGAGGGCGGATGTGGTTTGCGACG TCGTA	25	307	1554.4	22	495	5745.1	3	4359	77250.5	3.7	13.4	49.7
CGTGCTCATACCGCCGAATACGAGG GAGGAGAGGGTGGCAAAACTGCGTG GGTT	11	1016	8060.0	14	1192	15763.8	4	4978	75180.7	2.0	4.8	9.3
CCTACCTGTATGAAAGAGACGAACC TAACTACAATGCTAGCAATGGTGTAT GGTGC	8	1193	13736.5	1	3536	83366.2	5	4326	69025.0	6.1	0.8	5.0
TCGACGTACCTCAGGGTGGTGTATG ACTGAGGTGAAGACTGTGAACCATG GCATGC	69	135	418.3	31	298	2784.3	6	3869	63749.5	6.7	22.9	152.4
CCAACCAGAGACGCACCAGAGTCC TCGAGGGAAAGGAGTGGGTGGTACG GGACATG	2	6422	61861.0	3	6179	79508.9	7	7817	74123.5	1.3	0.9	1.2
CCATCCCCCAACTTGTCAAGAGGGA AGGAGTGGGAAAGGAAGACAAGGA CATGTG	12	794	5892.1	6	2684	44108.5	8	3080	38458.9	7.5	0.9	6.5
CCCGGCCCCGCCAGTGGAAGTGGG AGGCATGAGGGAAAGGTTAAACACT GTCGGCT	33	263	836.5	16	1138	12301.7	9	3539	34448.5	14.7	2.8	41.2
GGACAGAACAACCCTTGAACGAATG AGGGACAGGTGCGAGGGAAGGATTC GTTGGA	13	781	5478.5	11	1661	24048.4	10	2770	32834.6	4.4	1.4	6.0
ACGCGAAAGCACGATCGGCACCAC CTAATGTCATGAGGGACGGAGTGAG GGAGGCA	9	1586	10119.9	9	2602	30489.2	11	4654	35166.6	3.0	1.2	3.5
CAACACTAAAGTTCGACAACAACGG ATATTGCCTACCAAGAGGATGGGTG TGTGGA	7	2011	15460.2	7	3632	45506.4	12	3669	23599.6	2.9	0.5	1.5
TACCGCGACCTACCCCAGTGCAGGC TAGTGAGGGAGGAGGGTGGCATCTA GACACT	15	459	3636.7	18	647	9874.0	13	1350	16102.9	2.7	1.6	4.4
ATCGGAACAAAACGCGACCCGGGA CTAGACCTTACCTAAGTGGATGGGT GTGTGGA	92	167	320.8	46	344	1378.4	14	2158	15962.4	4.3	11.6	49.8
CGCCAAATTGCTCAGTCCAGTGCGG GAGGTGAGGGCGGAACAACATGAC GGACCA	14	548	3901.3	15	851	12294.0	15	1419	14555.2	3.2	1.2	3.7
GGCGAATTACACGACAATCTGGACG ACCTACACTATGAGAGGAGGGTAGA GTGGAA	1	9120	118265.6	2	5339	79531.1	16	2737	16349.2	0.7	0.2	0.1
TGCGCGCGATGCCCGTACCATGCTA TGTGGGTGGCGAGGGAGGACACTGC ATGTTG				146	60	194.5	17	1377	13336.9		68.6	
CCCTGGAGACCACCAACAGAGGGA GGTGTGAGGGAAAGGATGATAGGTG CATGTCA	30	285	1095.4	27	482	4616.7	18	1357	11872.7	4.2	2.6	10.8
CCGTGCCCAGGGTCTTGCTGTATGTA TGGTGCTGATCTACTGTAGCCATGTG TATG				156	66	173.6	19	1312	11467.4		66.1	
GGGTCGCATCGGTCTCTATGGATAA ATCTGCCTTATTGGAGGATGGGAAAT GTGGA	3	4409	48581.8	8	2528	33209.8	20	1360	9833.8	0.7	0.3	0.2

In the nitrocellulose trajectory, the majority of the top 20 most abundant sequences in lattice round 15 had enrichment values less than 2 and did not enrich (Table 3.4). Sequences L15.2.1 and L15.14.1 had enrichment values of 2.1 and 2.2, respectively, and sequence L15.13.1 had the highest enrichment value among the top 20 sequences in round 15 of approximately four-fold. Of the 1,286 sequences that had enriched more than four-fold, only 38 sequences had greater than 50 reads in lattice round 15 (data not shown). Of the top 20 most enriched clusters in round L15nc2, the two-highest were approximately 20-fold enriched, and the rest dropped quickly off (Table 3.5). These data suggest that most of lattice-binding aptamers do not have a strong affinity to alkaline-treated nitrocellulose. Candidates were removed from the short list if the sequence and its respective cluster had enriched greater than 5-fold in the nitrocellulose trajectory. Three sequences met these criteria and were excluded from the short list (L15.45.1, L15.82.1, and L15.112.1).

Table 3.4: Enrichment Values of Top 20 Most Abundant Sequences in Lattice Round 15 in the Nitrocellulose Trajectory

Sequence	Rank (L15)	RPM (L15)	Rank (L15nc2)	RPM (L15nc2)	Enrichment (L15nc2/L15)
GCCGATGCAGTCCCAATGTCGTATGAGATGTGACTACCTACCACTGTTTAGTGTAT	1	86619.7	3	59898.6	0.7
CCACCCATGCCAAGGAAGGAGGTCGGAGGAGGTCACAGGCATTGGTGTGCTAGGAT	2	86094.7	1	179529.6	2.1
CGCGCAAGCTTTCACGAGGGAGGTTGCGAGGGCGGATGTGGTTTGCGACGTCGTA	3	65992.2	4	56525.9	0.9
CGTGCTCATACCGCCGAATACGAGGGAGGAGAGGGTGGCAAAACTGCGTGGGTT	4	61824.5	2	79758.4	1.3
CCTACCTGTATGAAAGAGACGAACCTAACTACAATGCTAGCAATGGTGTATGGTGC	5	56884.7	9	34566.1	0.6
TCGACGTACCTCAGGGTGGTGTATGACTGAGGTGAAGACTGTGAACCATGGCATGC	6	55461.3	7	38176.8	0.7
CCAACCAGAGACGCACCAGAGTCCTCGAGGGAAAGGAGTGGGTGGTACGGGACATG	7	52362.8	6	52078.9	1.0
CCATCCCCCAACTTGTCAAGAGGGAAGGAGTGGGAAAGGAAGACAAGGACATGTG	8	32558.7	13	24763.7	0.8
CCCGGCCCCGCCAGTGGAAGTGGGAGGCATGAGGGAAAGGTTAAACACTGTCGGCT	9	28572.3	8	36717.0	1.3
GGACAGAACAACCCTTGAACGAATGAGGGACAGGTGCGAGGGAAGGATTCGTTGGA	10	27554.9	12	24966.6	0.9
ACGCGAAAGCACGATCGGCACCACCTAATGTCATGAGGGACGGAGTGAGGGAGGCA	11	26033.8	10	28582.1	1.1
CAACACTAAAGTTCGACAACAACGGATATTGCCTACCAAGAGGATGGGTGTGTGGA	12	16679.8	15	17942.5	1.1
TACCGCGACCTACCCCAGTGCAGGCTAGTGAGGGAGGAGGGTGGCATCTAGACACT	13	14128.1	5	55769.2	3.9
ATCGGAACAAAACGCGACCCGGGACTAGACCTTACCTAAGTGGATGGGTGTGTGGA	14	12703.8	11	27559.3	2.2
CGCCAAATTGCTCAGTCCAGTGCGGGAGGTGAGGGCGGAACAACATGACGGACCA	15	12692.8	25	2836.7	0.2
GGCGAATTACACGACAATCTGGACGACCTACACTATGAGAGGAGGGTAGAGTGGAA	16	11698.8	17	6672.7	0.6
TGCGCGCGATGCCCGTACCATGCTATGTGGGTGGCGAGGGAGGACACTGCATGTTG	17	11208.0	21	4428.6	0.4
CCCTGGAGACCACCAACAGAGGGAGGTGTGAGGGAAAGGATGATAGGTGCATGTCA	18	10039.8	24	3148.8	0.3
CCGTGCCCAGGGTCTTGCTGTATGTATGGTGCTGATCTACTGTAGCCATGTGTATG	19	9758.4	59	580.1	0.1
GGGTCGCATCGGTCTCTATGGATAAATCTGCCTTATTGGAGGATGGGAAATGTGGA	20	8157.8	16	15326.5	1.9

Table 3.5: Top 20 Enriched Aptamer Clusters in the Nitrocellulose Trajectory

Seed Sequence	Cluster (L15)	Cluster Sequences (L15)	Cluster RPM (L15)	Cluster (L15nc2)	Cluster Sequences (L15nc2)	Cluster RPM (L15nc2)	Enrichment (L15nc2/ L15)
CCTACTGAAAGAGCGTGCTACGTACCGGTGGAG GGAGAGTGAGGAGGGTGTGCGTA	82	122	218.4	20	543	4533.5	20.8
AGAGAATAGGCAAATACACCAATCTACCATAAT GCTGGCGGAGGGAGAATGAGGAG	45	499	1429.3	14	2969	28562.7	20.0
CGCGCAACGTGCATTTACACCAATCTACCATAA TGCTGGCGGAGGGAGAATGAGGAG	989	4	2.2	110	24	31.0	14.1
CGGCCACGTCGACTACCACAAGTATACTACCTA CCATAGTGGAGGGTAAATGTGGA	112	108	123.1	32	278	1066.3	8.7
GGCCACCGCGGTACACCAATCTACCATAATGCT GGCGGAGGGAGAATGAGGAG	780	3	2.8	146	20	20.2	7.4
ACAACGAATCGGCCTAAACTCACCGTGGAGGGT GTGCGGGGGGATGGTGATGTGCA	950	5	2.4	171	10	12.6	5.3
CCTGCAAAGTGATGATCGCCATCGCGCCCAGAG GGTTAGGAGCGAGGGACAGGGTT	733	11	4.6	141	19	21.3	4.6
TACCGCGACCTACCCCAGTGCAGGCTAGTGAGG GAGGAGGGTGGCATCTAGACACT	13	1349	16033.9	5	2221	62287.4	3.9
CACCGATAAACCGTCGTGATGCGGTGGGCGATG AGGAGGGTTCACGTACGGTGGAA	320	16	11.9	98	17	45.2	3.8
GGACACCAGACCGAATTACATCCTTATGCTTAG GTAATCGAGGGAGGAGGGTGGTA	747	10	4.2	175	11	13.9	3.3
ACCACCAGGCCAATGCGAAAGGGACATGGAGG GAGGAGGGAGGCTTTGGATCGCAG	96	147	191.9	44	233	564.1	2.9
CTGCTCATGCCGAACGTGTGCCCGTGTGGCAAG GGAGGAGCGAGGGCGGTACCAGC	776	6	3.5	228	7	8.2	2.3
ACCACCAGGCCAATGCGAAAGGGACATGGAGG GAGGAGGGTGGCATCTAGACACT	788	5	3.1	272	6	7.1	2.3
ATCGGAACAAAACGCGACCCGGGACTAGACCTT ACCTAAGTGGATGGGTGTGTGGA	14	2160	15933.6	11	2576	34642.7	2.2
CCACCCATGCCAAGGAAGGAGGTCGGAGGAGG TCACAGGCATTGGTGTGCTAGGAT	2	6155	101109.2	1	6503	212150.6	2.1
TAACAGTACGCCCCAATTCTGAGCAAGAGGAGG AGAGGATAGGGTGACAGATGTGC	882	6	2.9	282	5	5.5	1.9
GGGTCGCATCGGTCTCTATGGATAAATCTGCCTT ATTGGAGGATGGGAAATGTGGA	20	1360	9833.8	16	1292	18087.1	1.8
ATGGTAACACTCGTCTCGGACGCGGAGGAGCGG GAGGAGGAGTAAACGTCCACGGC	425	9	7.2	186	10	11.8	1.7
CACACGGCAACGCAGCGACATGAGACTGCCAC ATGCATAGGAAAGAGGGTCAGGGA	792	13	5.9	465	19	9.2	1.6
CGTGCTCATACCGCCGAATACGAGGGAGGAGAG GGTGGCATCTAGACACT	125	56	72.1	68	35	112.3	1.6

In the lattice and hexamer trajectory, most of the top 20 most abundant sequences in lattice round 15 significantly depleted during this trajectory (Table 3.6). This suggests that most of these aptamers are likely specific for CA lattice. However, sequences L15.1.1, L15.14.1, and L15.20.1 enriched during this trajectory, and sequence L15.2.1 remained approximately the same. Of the most enriched clusters in the L17+H7 library, only nine clusters had enrichment values greater than 1, and three of those clusters were not highly abundant as their cluster RPM values were less than 50 (Table 3.7). Specifically, clusters 45, 20, 14, and 1 enriched at least 4.4-fold. Clusters 2 and 82 had enrichments of 1.6 and 1.1, respectively, and the remaining clusters in this trajectory had depleted. Because the selection target initially toggled between CA hexamer and CA lattice, it was also important to compare how sequences had enriched and/or depleted from the L17+H3 to L17+H7 rounds. It is possible that sequences capable of binding CA lattice specifically were able to persist in the aptamer populations due to the presence of the positive rounds against CA lattice. However, since only positive rounds against the CA hexamer were performed from rounds L17+H3 to L17+H7, any CA lattice-specific binders should have depleted. Sequence H7.10.1 and its cluster strongly enriched from rounds L17+H3 to L17+H7 (Tables 3.8 & 3.9). Of the top twenty sequences that had enriched from L17+H3 to L17+H7, ten of them were from L15 cluster 1, which was also the largest cluster in L17+H3 and L17+H7 (Table 3.8). Only 18 of the top 20 clusters had enriched from L17+H3 to L17+H7, including clusters 1, 2, 14, 19, 20, and 45 from lattice round 15 (Table 3.9). Overall, most individual sequences and sequence clusters strongly depleted in the lattice and hexamer trajectory, suggesting that most sequences in lattice round 15 do not bind the soluble CA hexamer. This is in agreement with the previous

binding result where lattice round 15 library was unable to bind to the CA hexamer(20). There were four aptamers (L15.1.1, L15.2.1, L15.14.1, and L15.20.1) from the lattice 15 library short list of candidates that had enrichment values greater than 1 for both the individual sequence and its cluster from round L15 to L17+H7 and from round L17+H3 to L17+H7, suggesting that they could be capable of binding lattice and hexamer. Alternatively, for sequences and their respective clusters not present in the lattice round 15 library but were present in the L17+H3 and L17+H7 libraries, sequences were considered if they had a sequence RPM greater than 10, a cluster RPM greater than 50, and had enriched as a sequence and cluster greater than 5-fold from round L17+H3 to L17+H7. Aptamers H7.10.1 and H7.12.1 met these criteria. Finally, for sequences and their respective clusters only present in the L17+H7 library, sequences were considered if they had a sequence RPM greater than 10 and a cluster RPM greater than 50. Only aptamer H7.20.1 met these criteria.

Table 3.6: Top 20 Sequences from Lattice Round 15 Library in the Lattice and Hexamer Trajectory

Sequence	Rank (L15)	RPM (L15)	Rank (L17+H3)	RPM (L17+H3)	Rank (L17+H7)	RPM (L17+H7)	Enrichment (L17+H3/ L15)	Enrichment (L17+H7/ L17+H3)	Enrichment (L17+H7/ L15)
GCCGATGCAGTCCCAATGTCGTATGAGATGTGACTACCTACCACTGTTTAGTGTAT	1	86619.7	1	107674.6	1	338110.4	1.2	3.1	3.9
CCACCCATGCCAAGGAAGGAGGTCGGAGGAGGTCACAGGCATTGGTGTGCTAGGAT	2	86094.7	2	97633.4	2	114338.1	1.1	1.2	1.3
CGCGCAAGCTTTCACGAGGGAGGTTGCGAGGGCGGATGTGGTTTGCGACGTCGTA	3	65992.2	12	17242.0	1351	9.6	0.3	0.001	0.0001
CGTGCTCATACCGCCGAATACGAGGGAGGAGAGGGTGGCAAAACTGCGTGGGTT	4	61824.5	3	92244.4	25	1637.6	1.5	0.02	0.03
CCTACCTGTATGAAAGAGACGAACCTAACTACAATGCTAGCAATGGTGTATGGTGC	5	56884.7	10	31853.3	693	31.9	0.6	0.001	0.001
TCGACGTACCTCAGGGTGGTGTATGACTGAGGTGAAGACTGTGAACCATGGCATGC	6	55461.3	14	9263.2	60	741.1	0.2	0.1	0.01
CCAACCAGAGACGCACCAGAGTCCTCGAGGGAAAGGAGTGGGTGGTACGGGACATG	7	52362.8	15	7718.5	764	27.7	0.1	0.004	0.001
CCATCCCCCAACTTGTCAAGAGGGAAGGAGTGGGAAAGGAAGACAAGGACATGTG	8	32558.7	8	35182.3	498	50.5	1.1	0.001	0.002
CCCGGCCCCGCCAGTGGAAGTGGGAGGCATGAGGGAAAGGTTAAACACTGTCGGCT	9	28572.3	9	33132.0	299	110.7	1.2	0.003	0.004
GGACAGAACAACCCTTGAACGAATGAGGGACAGGTGCGAGGGAAGGATTCGTTGGA	10	27554.9	4	74860.3	162	260.2	2.7	0.003	0.01
ACGCGAAAGCACGATCGGCACCACCTAATGTCATGAGGGACGGAGTGAGGGAGGCA	11	26033.8	11	20794.2	889	21.8	0.8	0.001	0.001
CAACACTAAAGTTCGACAACAACGGATATTGCCTACCAAGAGGATGGGTGTGTGGA	12	16679.8	17	7121.0	891	21.8	0.4	0.003	0.001
TACCGCGACCTACCCCAGTGCAGGCTAGTGAGGGAGGAGGGTGGCATCTAGACACT	13	14128.1	13	9960.3	363	81.7	0.7	0.008	0.01
ATCGGAACAAAACGCGACCCGGGACTAGACCTTACCTAAGTGGATGGGTGTGTGGA	14	12703.8	6	68300.2	3	105397.0	5.4	1.5	8.3
CGCCAAATTGCTCAGTCCAGTGCGGGAGGTGAGGGCGGAACAACATGACGGACCA	15	12692.8	153	286.1	435	60.7	0.0	0.2	0.005
GGCGAATTACACGACAATCTGGACGACCTACACTATGAGAGGAGGGTAGAGTGGAA	16	11698.8	16	7553.8	1357	9.6	0.6	0.001	0.001
TGCGCGCGATGCCCGTACCATGCTATGTGGGTGGCGAGGGAGGACACTGCATGTTG	17	11208.0	29	1576.8	4250	1.9	0.1	0.001	0.0002
CCCTGGAGACCACCAACAGAGGGAGGTGTGAGGGAAAGGATGATAGGTGCATGTCA	18	10039.8	24	1913.5	3358	2.4	0.2	0.001	0.0002
CCGTGCCCAGGGTCTTGCTGTATGTATGGTGCTGATCTACTGTAGCCATGTGTATG	19	9758.4	7493	1.3	837	23.9	0.0	18.9	0.002
GGGTCGCATCGGTCTCTATGGATAAATCTGCCTTATTGGAGGATGGGAAATGTGGA	20	8157.8	7	64566.5	4	90909.7	7.9	1.4	11.1

Table 3.7: Top 20 Enriched Aptamer Clusters in the Lattice and Hexamer Trajectory

Seed Sequence	Cluster (L15)	Cluster Sequences (L15)	Cluster RPM (L15)	Cluster (L17+H3)	Cluster Sequences (L17+H3)	Cluster RPM (L17+H3)	Cluster (L17+H7)	Cluster Sequences (L17+H7)	Cluster RPM (L17+H7)	Enrichment (L17+H3/ L15)	Enrichment (L17+H7/ L17+H3)	Enrichment (L17+H7/ L15)
AGAGAATAGGCAAATACACCAATCTACCATAATGCTGGCGGAGGGAGAATGAGGAG	45	499	1429.3	5	11141	102973.2	5	10864	111017.2	72.0	1.1	77.7
GGGTCGCATCGGTCTCTATGGATAAATCTGCCTTATTGGAGGATGGGAAATGTGGA	20	1360	9833.8	7	6681	79058.2	4	6683	119447.1	8.0	1.5	12.1
ATCGGAACAAAACGCGACCCGGGACTAGACCTTACCTAAGTGGATGGGTGTGTGGA	14	2160	15933.6	6	7773	89891.7	3	8576	153189.8	5.6	1.7	9.6
GGCCACCGCGGTACACCAATCTACCATAATGCTGGCGGAGGGAGAATGAGGAG	780	3	2.8	67	39	30.0	31	32	21.8	10.9	0.7	7.9
GCCGATGCAGTCCCAATGTCGTATGAGATGTGACTACCTACCACTGTTTAGTGTAT	1	6165	101202.1	1	9771	133009.5	1	14735	443891.2	1.3	3.3	4.4
CGCGCAACGTGCATTTACACCAATCTACCATAATGCTGGCGGAGGGAGAATGAGGAG	989	4	2.2	94	15	9.7	44	12	7.5	4.4	0.8	3.4
CCACCCATGCCAAGGAAGGAGGTCGGAGGAGGTCACAGGCATTGGTGTGCTAGGAT	2	6155	101109.2	2	7324	119126.3	2	6153	161906.6	1.2	1.4	1.6
GGGTCGCATCGGTCTCTATGGATTAACCTGCCTGCAGTGGATGGGAAGCAGAGGAG	418	24	15.4	57	50	55.0	33	25	22.9	3.6	0.4	1.5
CCTACTGAAAGAGCGTGCTACGTACCGGTGGAGGGAGAGTGAGGAGGGGTGTGCGTA	82	122	218.4	28	192	559.1	11	79	234.2	2.6	0.4	1.1
GGGTCGCATCGGCACCACCTAATGTCATGAGGGACGGAGTGAGGGAGGCA	417	9	7.5	86	14	11.2	60	3	2.4	1.5	0.2	0.3
ACTGTCGACCTGGTTGCGAGGGATAGGAGTGAGGGTGGACGACCATGTACGACAAT	842	4	2.6				832	1	0.3			0.1
GGGTCGCATCGGTCTCTATGGATAAATCTGCCTTAGACGAGGAGGAACGTGGGACA	443	16	9.0	70	24	22.4	193	1	0.5	2.5	0.02	0.1
CACTATGGACAATCTGGACGACCTACACTATGAGAGGAGGGTAGAGTGGAA	44	254	1364.7	18	506	4227.1	19	40	76.9	3.1	0.02	0.1
ACCTACAGACCTGCGATATTGAGAGGGAGGTTGCGAGGGAAAGGAACAGTATAGTG	497	8	6.1				792	1	0.3			0.04
CCCGAACCACCTAAATCCGGTCAAAGTGAGGGGGGGAAAGGAGCGGAATTTGACCA	315	18	12.7				112	1	0.5			0.04
CGTGCTCATACCGCCGAATACGAGGGAGGAGAGGGTGGCAAAACTGCGTGGGTT	4	4965	75082.9	3	7748	115305.3	6	297	1972.7	1.5	0.02	0.03
CCAACTGATGACTCTGAGGGAGGAATGTGAGGGTGGAAACGAGTCTACGGTGAGAT	24	1031	7392.4	19	507	2689.1	13	38	122.2	0.4	0.05	0.02
TCGACGTACCTCAGGGTGGTGTATGACTGAGGTGAAGACTGTGAACCATGGCATGC	6	3866	63679.6	14	1340	11148.7	7	156	824.4	0.2	0.1	0.01
GGACAGAACAACCCTTGAACGAATGAGGGACAGGTGCGAGGGAAGGATTCGTTGGA	10	2772	32879.7	4	6854	92634.0	9	107	324.2	2.8	0.003	0.01

Table 3.8: Top 20 Most Enriched Sequences from Round L17+H3 to Round L17+H7

Sequence	Rank (L15)	RPM (L15)	Cluster (L15)	Cluster Rank (L15)	Rank (L17+H3)	RPM (L17+H3)	Cluster (L17+H3)	Cluster Rank (L17+H3)	Rank (L17+H7)	RPM (L17+H7)	Cluster (L17+H7)	Cluster Rank (L17+H7)	Enrichment (L17+H3/ L15)	Enrichment (L17+H7/ L17+H3)	Enrichment (L17+H7/ L15)
TGGCGGGTAAACTAAGTCCTTGG GAGGGAGCGATGGTTCTTAGCCT CGAGGGATGG					25391	0.4	687	1	202	199.0	10	1		473.7	
GCCGATGCAGTCCCCATGTCGT ATGAGATGTGACTACCTACCACT GTTTGGTGTAT	21065	0.6	1	800	97455	0.2	1	7960	645	36.2	1	192	0.4	172.3	65.8
TGCAGTCCCAATGTCGTATGAGA TGTGACTACCTACCACTGTTTAG TGTAT	5682	3.3	1	216	3461	4.2	1	246	116	357.5	1	41	1.3	84.7	108.7
CACTGCTGCCTAGCCCCATAGT ATCCGCACCACGCCAGACTCAT ATGTTAAACAGT					56964	0.2			1084	15.2	31	1		72.2	
GCCGTTGCAGTCCCAATGTCGTA TGAGATGTGACTACCTACCACTG TTTGGTGTAT	33237	0.4	1	1466	99164	0.2	1	9305	1090	14.9	1	293	0.6	71.0	40.3
GCCGATGCAGTCCCGATGTCGT ATGAGATGTGACTACCTACCACT GTTGGTGTAT					97517	0.2	1	8019	1089	14.9	1	292		71.0	
GCCGATGCAGTCCCAATGTCGT ATGTGATGTGACTACCTACCACT GTTAGTGTAT	21053	0.6	1	788	96872	0.2	1	7405	1103	14.6	1	296	0.4	69.7	26.6
TGGACGAGGTGTATTTGGGCAGA CCTGACCTAAGTGGATGGGTGTG TGGA					120460	0.2	268	4	1164	13.3	21	2		63.3	
GCCGATGCAGTCCCAATGTCGT ATGAGATGTGACTACCTAACACT GTTTGGTGTAT	127111	0.2	1	3919	95820	0.2	1	6372	1175	13.0	1	312	1.2	62.0	72.4
GCCGATGCAGTCCCAATGTCGT ATGAGCTGTGACTACCTACCACT GTTTGGTGTAT	127749	0.2	1	4523	22737	0.4	1	1954	812	25.5	1	225	2.3	60.8	141.9
GGGTCGCACCGGTCTCTATGGAT AAATCTGCCTTATTGGAGGATGG GTAATGTGGA					15163	0.6	7	626	649	35.7	4	100		56.6	
CCCGATGCAGTCCCAATGTCGT ATGAGATGTGACTACCTACCACT GTTTGGTGTAT	82909	0.2	1	1699	69038	0.2	1	2595	1280	10.6	1	336	1.2	50.7	59.1
ACCAAACTCACCACCGAATGAG TTCGCCGACAACGGAGTGCCTT GCTTTCCCAGGTT					26852	0.2			1287	10.4	35	1		49.4	
GCCGATGCAGTCTCAATGTCGTA TGAGATGTGACTACCTACCACTG TTTGGTGTAT	128641	0.2	1	5336	14569	0.6	1	1309	780	26.9	1	217	3.5	42.7	149.3
GCCCAAACGTCTGGCTAAGTCG CAGGATTCATCCTGATCCTTCAC ATCTCCGCACA					92021	0.2			1427	8.8	39	1		41.8	
GGGTCGCACCGGTATCTATGGAT AAATCTGCCTTATTGGAGGATGG GAAATGTGGA					110732	0.2	7	2540	1469	8.5	4	210		40.5	
GGGTCGCACCGGTGTCTATGGAT AAATCTGCCTTATTGGAGGATGG GAAATGTGGA					110777	0.2	7	2585	1473	8.5	4	214		40.5	
GCCGATGCAGTCCCAATGTCGT ATGAGATGTGACTTCCTACCACT GTTTGGTGTAT	13522	0.9	1	440	22700	0.4	1	1917	1050	16.5	1	284	0.5	39.3	18.1
GCCGATGCAGTCCCAATGTCGT ATAAGATGTGACTACCTACCACT GTTTGGTGTAT	20963	0.6	1	700	7694	1.3	1	596	516	48.2	1	170	2.3	37.9	87.5
GGGTCGCATCGGTCTCTATGGAT AACCCTGCCCGGAAGTGGAGGA ATGTGGA					3332	4.4	44	2	232	163.9	12	1		37.0	

Table 3.9: Top 20 Most Enriched Clusters from Round L17+H3 to Round L17+H7

Seed Sequence	Cluster (L15)	Cluster Sequences (L15)	Cluster RPM (L15)	Cluster (L17+H3)	Cluster Sequences (L17+H3)	Cluster RPM (L17+H3)	Cluster (L17+H7)	Cluster Sequences (L17+H7)	Cluster RPM (L17+H7)	Enrichment (L17+H3/ L15)	Enrichment (L17+H7/ L17+H3)	Enrichment (L17+H7/ L15)
TGGCGGGTAAACTAAGTCCTTGGGAGGGAGCGATG GTTCTTAGCCTCGAGGGATGG				651	1	0.4	10	51	224.7		531.4	
ATGGTAACCTACCCATGCTCACGTAGCTACAATGGT AGAGCAGCTTGACTTGATAT				418	1	0.4	37	2	8.5		20.1	
CCGTGCCCAGGGTCTTGCTGTATGTATGGTGCTGAT CTACTGTAGCCATGTGTATG	19	1312	11467.4	217	10	3.8	25	11	27.7	0.0003	7.3	0.002
GACTGGAACGCGACCCGGGACTAGACCTTACCTAA GTGGATGGGTGTGTGGA				191	52	14.8	41	142	60.7		4.1	
GGGTCGCATCGGTCTCTATGTCGTATGAGATGTGAC TACCTACCACTGTTTAGTGTAT				204	2	1.7	45	8	6.1		3.6	
GCCGATGCAGTCCCAATGTCGTATGAGATGTGACTA CCTACCACTGTTTAGTGTAT	1	6165	101202.1	1	9771	133009.5	1	14735	443891.2	1.3	3.3	4.4
CCTACACGCATCGACGATCTACCATAATGCTGGCG GAGGGAGAATGAGGAG				218	3	1.9	53	7	5.9		3.1	
CCGACGACGTGCGACCCGGGACTAGACCTTACCTA AGTGGATGGGTGTGTGGA				335	6	1.7	57	10	4.8		2.8	
GTGTATTGCCGCTGTGATCGTGAGCACCTATCTTAC AGTGGAGGGTTAGCAGTGGA				126	10	5.7	39	13	12.2		2.1	
ATCGGAACAAAACGCGACCCGGGACTAGACCTTAC CTAAGTGGATGGGTGTGTGGA	14	2160	15933.6	6	7773	89891.7	3	8576	153189.8	5.6	1.7	9.6
AGAGAATCATACACAAGACGCGACCCGGGACTAGA CCTTACCTAAGTGGATGGGTGTGTGGA				306	2	0.8	76	2	1.3		1.6	
GGGTCGCATCGGTCTCTATGGATAAATCTGCCTTAT TGGAGGATGGGAAATGTGGA	20	1360	9833.8	7	6681	79058.2	4	6683	119447.1	8.0	1.5	12.1
GGGTCGCATCGGTCTCTATGGATATTGCCTACCAAG AGGATGGGTGTGTGGA				36	103	165.5	15	119	238.8		1.4	
CCACCCATGCCAAGGAAGGAGGTCGGAGGAGGTCA CAGGCATTGGTGTGCTAGGAT	2	6155	101109.2	2	7324	119126.3	2	6153	161906.6	1.2	1.4	1.6
GGGTCGCATCGGTCTCTATGGATAAATCTACCATAA TGCTGGCGGAGGGAGAATGAGGAG				39	84	103.4	17	91	138.9		1.3	
ATCGGAACAAAACGCGACCCGGGACTAGACCTTAT TGGAGGATGGGAAATGTGGA				416	2	0.6	104	2	0.8		1.3	
GTCGGACAGGACGCACACCAATCTACCATAATGCT GGCGGAGGGAGAATGAGGAG				89	9	9.9	38	15	11.2		1.1	
AGAGAATAGGCAAATACACCAATCTACCATAATGC TGGCGGAGGGAGAATGAGGAG	45	499	1429.3	5	11141	102973.2	5	10864	111017.2	72.0	1.1	77.7
GCGAGGCAGCCTGCGACTGTTGGCACTACCATAAT GCTGGCGGAGGGAGAATGAGGAG				248	1	1.1	186	3	1.1		1.0	
GCTAGTGAGCGGACACCAATCTACCATAATGCTGG CGGAGGGAGAATGAGGAG				365	3	1.1	188	3	1.1		1.0	

In the hexamer depletion trajectory, of the top twenty most abundant sequences in lattice round 15, most sequences had enrichment values between 0.3 and 3 (Table 3.10). However, L15.19.1 was strongly depleted, suggesting that it could potentially be a hexamer binder. On a cluster level, cluster 19 was also the most strongly depleted (Table 3.11). While it was expected that sequences that significantly depleted in the hexamer depletion should have enriched in the lattice and hexamer trajectory, this prediction was only observed once for L15.19.1, where it had enriched from round L17+H3 to L17+H7. This trajectory was not very informative when it came to identifying candidate aptamers for further characterization. This, in part, could have occurred because the lattice round 15 likely did not contain many aptamers capable of binding to the CA hexamer. While sequences that were capable of binding hexamer were likely filtered out, sequences not capable of binding hexamer were also depleted. Additionally, the toggling between hexamer depletion steps and positive selection rounds against CA lattice likely allowed sequences to recover to an extent between the hexamer depletion steps, especially sequences from more abundant aptamer clusters. Therefore, it was harder to observe sequences that were strongly depleting.

Table 3.10: Top 20 Most Abundant Sequences in Lattice Round 15 in the Hexamer Depletion Trajectory

Sequence	Rank (L15)	RPM (L15)	Rank (L17-H3)	RPM (L17-H3)	Cluster (L17-H3)	Enrichment (L17-H3/ L15)
GCCGATGCAGTCCCAATGTCGTATGAGATGTG ACTACCTACCACTGTTTAGTGTAT	1	86619.7	1	111559.1	1	1.3
CCACCCATGCCAAGGAAGGAGGTCGGAGGAG GTCACAGGCATTGGTGTGCTAGGAT	2	86094.7	7	51643.6	7	0.6
CGCGCAAGCTTTCACGAGGGAGGTTGCGAGGG CGGATGTGGTTTGCGACGTCGTA	3	65992.2	12	24155.6	12	0.4
CGTGCTCATACCGCCGAATACGAGGGAGGAGA GGGTGGCAAAACTGCGTGGGTT	4	61824.5	2	78917.4	2	1.3
CCTACCTGTATGAAAGAGACGAACCTAACTAC AATGCTAGCAATGGTGTATGGTGC	5	56884.7	5	66912.0	5	1.2
TCGACGTACCTCAGGGTGGTGTATGACTGAGG TGAAGACTGTGAACCATGGCATGC	6	55461.3	14	15585.8	14	0.3
CCAACCAGAGACGCACCAGAGTCCTCGAGGG AAAGGAGTGGGTGGTACGGGACATG	7	52362.8	9	31272.4	9	0.6
CCATCCCCCAACTTGTCAAGAGGGAAGGAGTG GGAAAGGAAGACAAGGACATGTG	8	32558.7	3	71489.1	3	2.2
CCCGGCCCCGCCAGTGGAAGTGGGAGGCATG AGGGAAAGGTTAAACACTGTCGGCT	9	28572.3	8	44167.5	8	1.5
GGACAGAACAACCCTTGAACGAATGAGGGAC AGGTGCGAGGGAAGGATTCGTTGGA	10	27554.9	6	60765.4	6	2.2
ACGCGAAAGCACGATCGGCACCACCTAATGTC ATGAGGGACGGAGTGAGGGAGGCA	11	26033.8	4	68958.7	4	2.6
CAACACTAAAGTTCGACAACAACGGATATTGC CTACCAAGAGGATGGGTGTGTGGA	12	16679.8	10	27655.8	10	1.7
TACCGCGACCTACCCCAGTGCAGGCTAGTGAG GGAGGAGGGTGGCATCTAGACACT	13	14128.1	35	1352.2	23	0.1
ATCGGAACAAAACGCGACCCGGGACTAGACC TTACCTAAGTGGATGGGTGTGTGGA	14	12703.8	15	14383.33	15	1.1
CGCCAAATTGCTCAGTCCAGTGCGGGAGGTGA GGGCGGAACAACATGACGGACCA	15	12692.8	51	753.24	25	0.1
GGCGAATTACACGACAATCTGGACGACCTACA CTATGAGAGGAGGGTAGAGTGGAA	16	11698.8	11	25630.18	11	2.2
TGCGCGCGATGCCCGTACCATGCTATGTGGGT GGCGAGGGAGGACACTGCATGTTG	17	11208.0	20	4414.7	19	0.4
CCCTGGAGACCACCAACAGAGGGAGGTGTGA GGGAAAGGATGATAGGTGCATGTCA	18	10039.8	23	4023.66	21	0.4
CCGTGCCCAGGGTCTTGCTGTATGTATGGTGCT GATCTACTGTAGCCATGTGTATG	19	9758.4	3722	4.4	115	0.0005
GGGTCGCATCGGTCTCTATGGATAAATCTGCCT TATTGGAGGATGGGAAATGTGGA	20	8157.8	13	24037.5	13	2.9

Table 3.11: Top 20 Enriched Clusters in the Hexamer Depletion Trajectory

Seed Sequence	Cluster (L15)	Cluster Sequences (L15)	Cluster RPM (L15)	Cluster (L17-H3)	Cluster Sequences (L17-H3)	Cluster RPM (L17-H3)	Enrichment (L17-H3/ L15)
CCGTGCCCAGGGTCTTGCTGTATGTATGGTGCTGA TCTACTGTAGCCATGTGTATG	19	1312	11467.4	101	12	6.5	0.0006
TGGATCCTGGTGTGTATGGCAGAATCCCAGTTGTA TGATATTGGACGCCTATCGAT	34	548	3269.2	141	4	3.3	0.001
CGAACGCTGCACCGACACTTATGTACCCCCTGAG GGCGGTGTGAGGGATAGGAATG	39	516	1951.0	233	7	2.0	0.001
CTCGACCAGTGCGAGGCGAGGGAGGGTGGTACAA TGGTCGAAGTCGCCACTGATCA	55	162	510.3	305	1	0.7	0.001
CCCACACTGGTTGTTTTTCGGTGGAGAGGGTGGAG GAGCGAGAATTACAATCAGG	52	421	1138.8	295	6	1.5	0.001
ATAACAACAACCCCTGCCTAGCAAGTGGGAGGAA TGAGGGAAAGGCTGCTGAAGCA	101	91	144.1	421	1	0.3	0.002
CAACAATCCCGTCTTGCGCTCACGTGGGAGGAATT GCGAGGGTGGATGTGAGCAAC	63	142	341.9	253	1	0.8	0.002
CCCATGGCCAATCCACCGCGTGGTGAGGGAGGGA GGCTGAATGTGGTTGGTCTAAA	29	885	5532.4	61	25	16.0	0.003
CACACTTGGCCTACACAAGTAGTGTCCGCCCCGT GGGAGGTATGCGTGGGAGGATG	43	430	1540.5	122	10	4.6	0.003
CCCGCGACTCCCAACAGTTGTGTAGGCACTACCA AACCATGCTAGGTGTATGGGCA	73	208	326.5	264	2	1.0	0.003
CCCCCCCCAAAGATACTGGCGCATGTGAGGGAGG CGAGGGAGGTACGCGTCCTTT	23	2532	13261.0	37	44	44.4	0.003
CGATAATGGACTAGTAATGTGTAGGACTACTGCCT TTTTACCTAGTGTGTGTGATA	27	715	5573.1	46	21	23.5	0.004
AGCGCAACATGGCCCGTGTGCGTGAGTGTATCGA CTGTACCTACGCGCTAGTGTAT	104	89	149.4	285	1	0.7	0.004
CCGCCACCCCCGCCAGCAGATGAGGGAGGTCAGA GGGTAAGGATAAAGCAGCTGTC	25	1134	7152.9	40	43	34.9	0.005
CCGGAGATCCGAGAGGGAGGACTGCGCGGGAGGA TATGGATCTTGCGGCACAGCCT	131	54	66.6	472	1	0.3	0.005
CAACCTGTTCACACGTACTGGTGAGGGTCAGGAG CGGGAGGAACACAGACGTGTGA	86	132	198.7	226	1	1.0	0.005
CCGCCGCCTTACCTGTCGTTGAGGGTTAGGAGAGG GAGGTAATCGATTGGTAGGTC	41	385	1614.3	78	11	8.3	0.005
CCAGTGACAGCCCAAGCACGTGAGGGAGGCCAGA GGGAGGTTATGTGCTTGATGTG	47	293	1106.7	95	7	5.9	0.005
CATCGCACTACCACCCATGATGCTAGTCATGAGG GAGGCTTGAGGGAGGATTGACA	109	65	91.8	438	2	0.5	0.005
CGCGAGCTGGTCCGGATTATACGACGTGGGCGGA GCGAGGGATAGGATGTCGAACC	48	328	1117.6	102	10	6.2	0.006

3.3.4. Candidate Aptamer Identification and Screening of CA Lattice and Hexamer Binding

The criteria discussed above using the enrichment and depletion patterns above from the lattice, nitrocellulose, and lattice and hexamer trajectories were considered for identifying candidate aptamers for further testing. In summary, the sequence should have first performed well in the lattice selection, using the criteria described in the lattice selection section. Second, the sequence and its cluster should not have high enrichment for alkaline-treated nitrocellulose. Third, in the lattice and hexamer trajectory, sequences that bind to hexamer should have individual and cluster enrichment values that either enrich or stay relatively constant if the sequence/cluster was highly abundant. In contrast, sequences that do not bind hexamer should deplete, especially from rounds L17+H3 to L17+H7, since there were only positive selection rounds against the soluble hexamer. Additionally, sequences that were not present in the lattice round 15 library but were present in rounds L17+H3 and/or L17+H7 were considered if the met the criteria discussed in the lattice and hexamer trajectory section. Because aptamer L15.45.1 and its cluster had enriched the lattice and hexamer trajectory and the nitrocellulose trajectory, it was also included in the screen to test whether enrichment in the nitrocellulose trajectory was predictive of an aptamer incapable of binding CA hexamer. Table 3.12 lists the aptamers that were screened for CA lattice and hexamer binding, their enrichment values from the different selection trajectories, and their predicted binding phenotype based on how they enriched and/or depleted.

Table 3.12: Candidate Aptamers for CA Binding Screen[a]

Sequence	Lattice	Hexamer	Nitrocellulose	Predicted Binding Phenotype	Actual Binding Phenotype
L15.1.1	759.4 (27.7 L15/L10)	3.9 (4.9)	0.7 (0.7)	Lattice and Hexamer	Lattice only
L15.2.1	4.1 (4.3)	1.3 (1.6)	2.1 (2.1)	Lattice and Hexamer	Lattice and Hexamer
L15.3.1	48.7 (49.7)	0.0001 (0.0002)	0.9 (0.9)	Lattice only	Lattice only
L15.4.1	9.0 (9.3)	0.03 (0.03)	1.3 (1.3)	Lattice only	Lattice only
L15.5.1	4.8 (5.0)	0.0006 (0.0006)	0.6 (0.6)	Lattice only	Lattice only
L15.6.1	150.5 (152.4)	0.01 (0.01)	0.7 (0.7)	Lattice only	Lattice only
L15.7.1	1.1 (1.2)	0.0006 (0.0007)	1.0 (1.0)	Lattice only	Lattice only
L15.8.1	6.4 (6.5)	0.002 (0.002)	0.8 (0.8)	Lattice only	Lattice only
L15.9.1	39.8 (41.2)	0.004 (0.004)	1.3 (1.3)	Lattice only	Lattice only
L15.10.1	5.9 (6.0)	0.009 (0.01)	0.9 (0.9)	Lattice only	Lattice only
L15.13.1	4.3 (4.4)	0.006 (0.006)	3.9 (3.9)	Lattice only	Lattice only
L15.14.1	48.9 (49.8)	8.3 (9.6)	2.2 (2.2)	Lattice and Hexamer	Lattice only
L15.15.1	3.7 (3.7)	0.005 (0.004)	0.2 (0.2)	Lattice only	Lattice only
L15.16.1	0.1 (0.1)	0.0008 (0.001)	0.6 (0.6)	Lattice only	Lattice only
L15.17.1	100.4 (68.6 L15/L10)	0.0002 (0.0002)	0.4 (0.6)	Lattice only	Lattice only
L15.19.1	1364.8 (66.1 L15/L10)	0.002 (0.002); enriched18.9 (7.3)from L17+H3 to L15+H7; strong depletion in hexamer depletion trajectory [0.0005 (0.0006)]	0.06 (0.06)	Lattice and Hexamer	Lattice only
L15.20.1	0.2 (0.2)	11.1 (12.1)	1.9 (1.8)	Lattice and Hexamer	Lattice only
L15.21.1	99.2 (43.8 L15/L10)	0.002 (0.004)	0.1 (0.1)	Lattice only	Lattice only
L15.23.1	178.5	0.00004 (.006 L10/L8)	0.3 (0.3)	Lattice only	Lattice only
L15.45.1	2.4 (0.5)	65.6 (77.7)	20.4 (20.0)	Lattice only	Lattice only
H7.10.1	Not present	Not present in L15, 474.1 L17+H7/L17+H3 (531.4 L17+H7/L17+H3)	Not present	Lattice and Hexamer	Lattice and Hexamer
H7.12.1	Not present	Not present in L15, 36.9 L17+H7/L17+H3 (5.8 L17+H7/L17+H3)	Not present	Lattice and Hexamer	Lattice only
H7.20.1	Not present	Only present in L15+H7, Rank 489; Cluster consists of 27 sequences and a RPM of 76.1	Not present	Lattice and Hexamer	Lattice only

[a]For the lattice, hexamer, and nitrocellulose columns, the enrichment value for the sequence is shown, and if calculated, its cluster's enrichment value is shown in parentheses. For the lattice column, the enrichment values shown are for L15/L8, unless indicated otherwise. For the hexamer column, the enrichment value is for L17+H7/L15, unless indicated otherwise. For the nitrocellulose column, the enrichment value is L15nc2/L15.

Aptamer binding to CA lattice and hexamer was screened using nitrocellulose filter binding assays. All candidate aptamers exhibited dose-dependent binding to CA lattice, as expected (Figure 3.11A). Two aptamers exhibited binding to CA hexamer, H7.10.1 and L15.2.1 (Figure 3.11B). Aptamer H7.10.1 exhibited dose-dependent binding to CA hexamer that was similar to its binding to CA lattice. In contrast, L15.2.1 exhibited more binding to CA hexamer than Arb. Aptamer L15.2.1 had higher non-specific binding when no protein was present, but its binding at 1000 nM CA hexamer was significantly different than its binding when no hexamer was present. It is likely that L15.2.1 is a weaker binder to CA hexamer than aptamer H7.10.1.

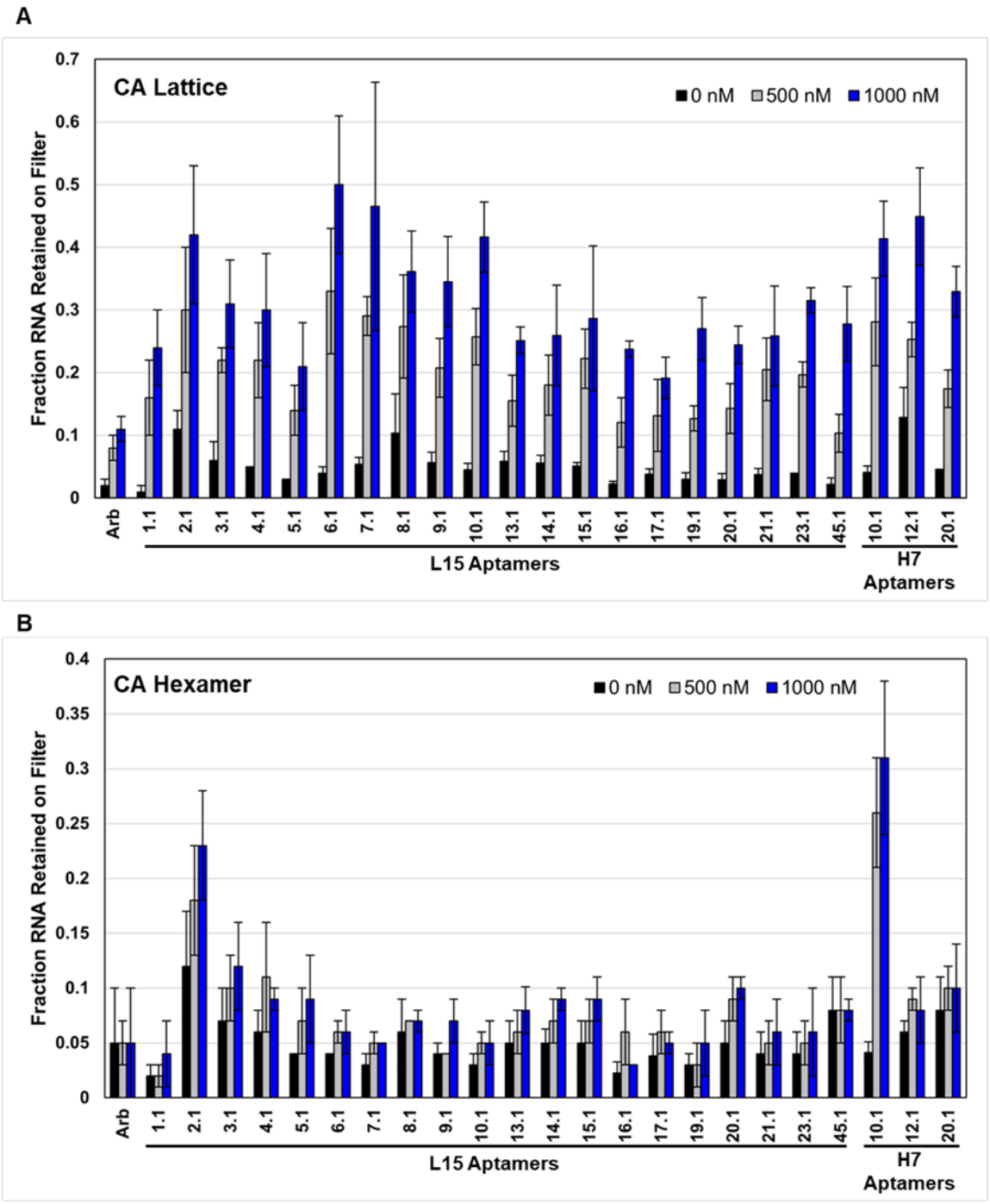

Figure 3.11: CA binding by candidate aptamers. Binding of candidate aptamers to CA lattice (A) and CA hexamer (B) were assessed using nitrocellulose filter binding assays. Experiments were performed at least in triplicate.

The observed binding phenotypes for the candidate aptamers are summarized in Table 3.12. Comparing the predicted binding phenotypes to the experimentally-determined binding phenotypes, 17 of the 23 binding predictions (74%) were correct. All aptamers that were predicted to be lattice-specific binders had specificity for the assembled CA lattice, suggesting that depletion in the lattice and hexamer trajectory was a good predictor for aptamers that were unable to bind CA hexamer. Of note, aptamer L15.45.1 and its cluster had enriched in both the lattice and hexamer and nitrocellulose trajectories, and because of its enrichment in the nitrocellulose trajectory, it was predicted and was found to be a lattice-specific binder. Therefore, having a control trajectory for the partition method was helpful for predicting binding specificities. However, only two of the eight aptamers predicted to bind both CA lattice and hexamer could bind both CA assembly states. Of these sequences that were present in the nitrocellulose trajectory, most had enrichment values above 1. Only two rounds against nitrocellulose were performed. It would be interesting to see what the enrichment value for the aptamers would have been if the same number of rounds against nitrocellulose were done as the number of positive rounds against CA hexamer using nitrocellulose as the partition method. Although it must be taken into account that some sequences, like L15.2.1, may have some affinity for nitrocellulose but also be able to bind the CA hexamer. Therefore, careful planning, execution, and bioinformatic analyses of the different selection trajectories can lead to quicker identification of aptamers with the desired binding specificities. However, all predictions must be experimentally validated.

3.3.5. Determination of Structural Motifs Present Using MEME Suite

We next wanted to determine what sequence motifs were present in the aptamer libraries, as the presence of recurring sequence motifs within the aptamer libraries may provide insights into the sequence requirements for capsid binding. Therefore, we analyzed the seed sequences for the top 1000 cluster in the lattice round 15 library using MEME Suite software(37) to identify recurring sequence motifs. Table 3.13 shows the top three most abundant sequence motifs, the number of sequences in which they were present, and the candidate aptamers from the prioritized list in Table 3.12 that contain the motifs. The sequence logos for these three motifs are shown in Figure 3.12. The most abundant motif was found in 725 of the top 1000 seed sequences, and this motif was highly G-rich. The other two motifs were found in 75 sequences each. Motif 2 contains a highly conserved GUGUAU motif, and motif 3 contains a conserved ACCUCC motif followed by a G-rich region. While we do not know if every sequence that contains one of these motifs has the motif in the same structural context, this analysis helped to narrow down the focus on a few candidate aptamers. Of note, the lattice and hexamer binding aptamer H7.10.1 did not contain any of the motifs identified from the lattice round 15 library top 1000 cluster seed sequences, nor did it contain any of the twenty most abundant sequence motifs identified when the top 1000 cluster seed sequences from round L17+H7 was used, suggesting that it contains unique sequence and/or structural requirements to be able to bind to CA lattice and hexamer.

Table 3.13: Top Three Most Abundant Sequence Motifs Identified in the Lattice Round 15 Library

Motif	Number of Sequences	Candidate Aptamers
NNGDGGGRGRKGHGAGGGAGGRDAD	725	L15.3.1, L15.4.1, L15.7.1, L15.8.1, L15.9.1, L15.10.1, L15.13.1, L15.15.1, L15.17.1, L15.23.1
NCHMMCNHNDBSYWRGUGUAU	75	L15.1.1, L15.5.1, L15.6.1, L15.14.1, L15.19.1, L15.21.1
HHWACCUACCWDADDRGGGGRGGRR	75	L15.16.1, L15.20.1, L15.45.1

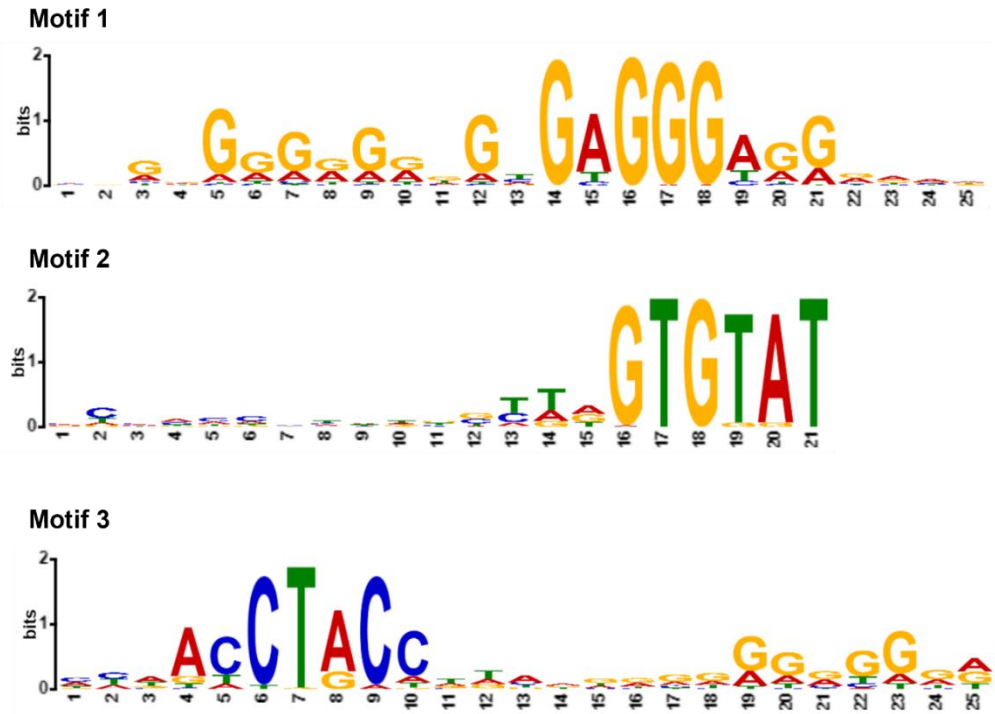

Figure 3.12: Sequence logos for the top three most abundant sequence motifs in the lattice round 15 library. Conserved sequence motifs were searched among the top 1000 cluster seed sequences from the lattice round 15 library using the MEME Suite software(37).

3.3.6. Binding of Aptamer Candidates to CA Lattice when an Antisense Oligonucleotide was Annealed to Either Constant Region

To determine whether the aptamers retained binding to the CA lattice when an antisense oligonucleotide was annealed to either the 5′ or 3′ constant region, we performed a nitrocellulose filter binding assay screen. This provided initial insights about which aptamers may require their constant regions for CA binding and determine which aptamers to use in the competition assay described in the next section that could tolerate an antisense oligonucleotide annealed to it 3′ constant region. We observed that most but not all aptamers retained binding to CA lattice when an oligonucleotide was annealed to its 3′ constant region but not to 5′ constant region in this binding screen (Figure 3.13). Since this screen was performed only once, it is possible that false positives could have been identified for use during the competition assay. However, because the binding of the aptamer annealed to an antisense oligonucleotide to the CA lattice is repeatedly measured in the competition assay, the aptamer's ability to bind to the CA lattice when an antisense oligonucleotide was annealed to its 3′ constant region could be confirmed.

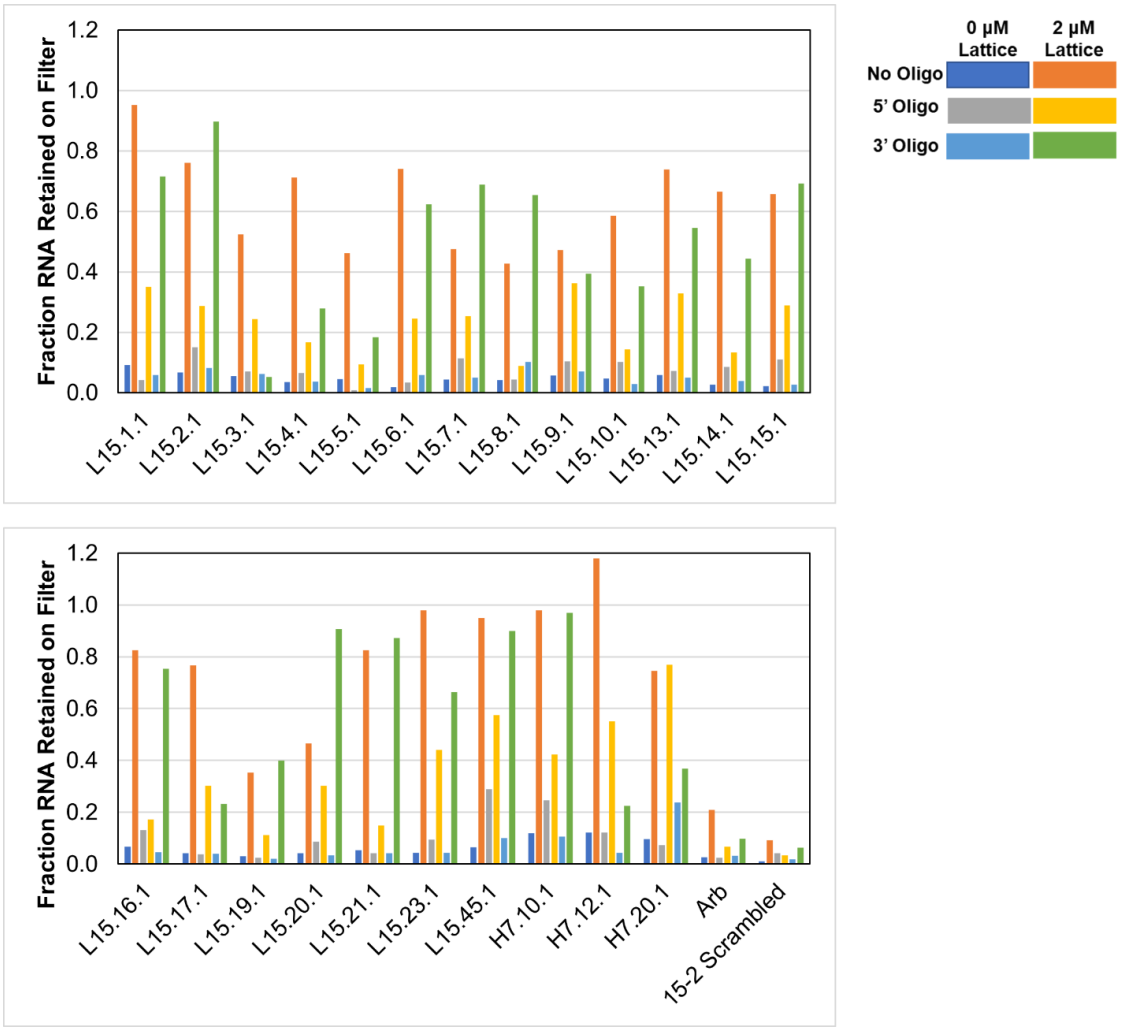

Figure 3.13: Candidate aptamer binding to CA lattice when either a 5′ or 3′ antisense oligonucleotide was annealed to its constant regions. Binding to CA lattice by aptamers when an oligonucleotide was annealed to one of constant was screened using a nitrocellulose filter binding assay. Data shown is the result of one experiment.

3.3.7. Aptamer Competition for Binding to CA Lattice

To determine whether different CA binding aptamers competed against each other for binding to the CA lattice, we utilized an electrophoretic mobility shift assay (EMSA). The EMSA allows for the visualization of competition by comparing how much labeled aptamer is retained in the wells (i.e., bound to CA lattice) and a higher throughput of samples. This assay was previously used to determine if truncations of CA15-2 and non-specific RNAs competed against labeled CA15-2 for binding to CA lattice(20). In this assay, 200 nM unlabeled aptamer or arbitrary RNA control was first incubated with the 2 μM CA lattice present in the reaction to block binding sites. Then 50 nM Cy5-labeled aptamer was added to the binding reaction, and the amount of labeled aptamer bound to the CA lattice was compared in the presence and absence of competitor. The aptamer is labeled by annealing a Cy5-modified antisense oligonucleotide to its 3′ constant region. Representative aptamers were chosen for this assay based on whether they could tolerate the presence of the antisense oligonucleotide (Figure 3.13) and the binding results in Figure 3.11. We chose to use L15.7.1 as a representative of motif 1, L15.20.1 as a representative of motif 3, and H7.10.1 as the lattice and hexamer binder. Based on previous results with L15.6.1 (CA15-2)(20), it was chosen as the representative of motif 2. Aptamers L15.1.1 and L15.2.1 were also included as part of the aptamer competition panel, since they were the two most abundant aptamers present in the lattice round 15 library.

We expected that aptamers with different specificities for CA assembly states (capable of binding lattice only or lattice and hexamer) likely bind to different sites on the CA lattice. Therefore, aptamers that bind to the same site should compete, while

aptamers that bind to different sites should not compete. However, we observed that most aptamers were able to compete for binding to CA lattice, regardless of aptamer binding specificities to CA assembly states (Figure 3.14). Generally, competition was observed between aptamers L15.6.1, L15.7.1, L15.20.1 and H7.10.1 for lattice binding. The lattice and hexamer binder H7.10.1 directly competed with lattice specific binders. Aptamers L15.1.1 and L15.2.1 were not good competitors for the other aptamers tested or even against themselves, suggesting that even though these two aptamers were the most abundant in the lattice round 15 library, they do not have the strongest binding affinities to the assembled CA lattice. These competition data also suggest that the binding of one aptamer blocks nearby binding sites on the CA lattice that can no longer be bound by other aptamers. This result is similar to a previous aptamer study that found that two aptamers that bind to different sites on VEGF-165 were unable to both bind at the same time to the VEGF-165 homodimer, suggesting that the orientation of one of the aptamer bound to the VEGF-165 homodimer spatially clashes with the second aptamer(38). Therefore, direct competition experiments using the assembled capsid lattice was not an informative method for determining whether different aptamers bind to the same site on the lattice.

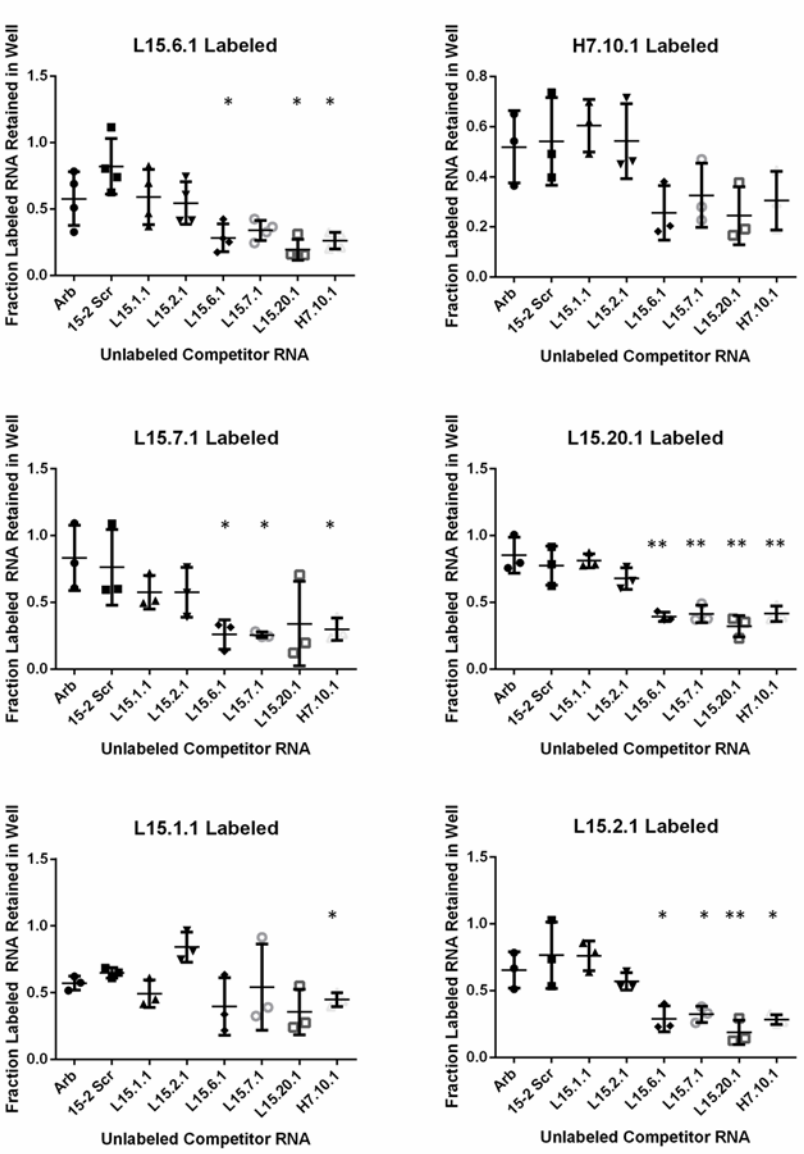

Figure 3.14: Aptamer competition for CA lattice binding. Electrophoretic mobility shift assays were used to determine the ability of unlabeled competitor aptamers to compete with a Cy5-labeled aptamer for binding to the assembled CA lattice. The fraction labeled RNA retained in well was quantified relative to when no competitor was added using Multi Gauge software (Fujifilm) using the following equation: (intensity of signal retained in well – intensity of signal in well of RNA only lane)/(intensity of signal retained in well of No Competitor lane – intensity of signal in well of RNA only lane). Values are the mean ± SD. Assays were done at least in triplicate. * (P value <0.05), ** (P value <0.01).

Because aptamer H7.10.1 was found to bind both CA lattice and hexamer, we wanted to determine whether it competed with small molecule PF74 that inhibits HIV-1 replication (39-41) and that binds CA in a pocket located at the CA NTD (N-terminal domain)-CA CTD (C-terminal domain) interface of two CA monomers within assembled hexamers(41). A nitrocellulose filter competition assay was performed by first allowing 250-1000 nM unlabeled competitor (H7.10.1 or PF74) to bind to 1 μM CA lattice, followed by the addition of 50 nM radiolabeled H7.10.1. As expected, aptamer H7.10.1 competed with itself for binding to CA lattice, but PF74 did not compete with H7.10.1 for binding to lattice (Figure 3.15). This suggests that H7.10.1 may not bind CA in the same pocket as PF74 or that the binding of PF74 does not block the binding of H7.10.1.

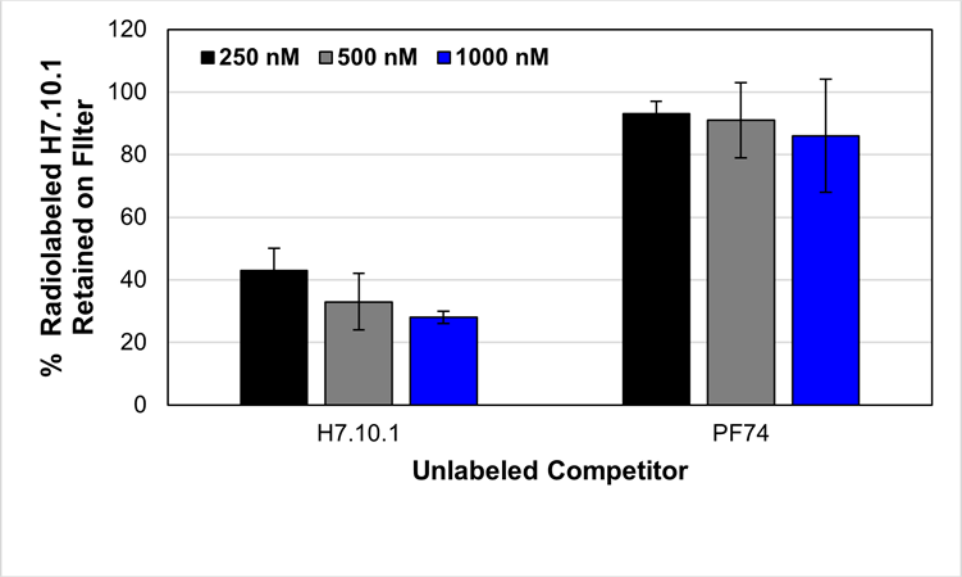

Figure 3.15: Competition of unlabeled aptamer H7.10.1 or PF74 and radiolabeled aptamer H7.10.1 for binding to CA lattice. The relative percent of radiolabeled aptamer H7.10.1 retained on the nitrocellulose filter was determined using full length unlabeled H7.10.1 and PF74 as competitors (n = 3). Values are the mean ± SD.

3.3.8. Elucidation of the Sequence and/or Structural Requirements for Aptamer Binding to CA Lattice

To understand more about the interaction between the aptamers and the assembled CA lattice, we focused on uncovering the sequence and structural requirements for aptamer binding to CA lattice for aptamers L15.7.1, L15.6.1, and L15.20.1 as representatives of the three most common sequence motifs identified within the lattice round 15 library and aptamer H7.10.1 as a lattice and hexamer binder. We utilized multiple approaches to look at the aptamer requirements for binding, including generating covariance models (CMs) from the HTS data, testing the binding of aptamer truncations from the 3′ end, and enzymatic probing using structure-sensitive nucleases.

Aptamer L15.7.1 was the representative sequence used for the most common motif that was highly G-rich. First, CMs for cluster 7 and the other top ten clusters from the lattice round 15 library that contained this motif were generated using Infernal(42). Briefly, we aligned the top 300 most abundant sequences from the cluster of interest and generated a predicted secondary structure from the alignment using RNAalifold(43). The sequence alignment and predicted structure were used to create a CM in Infernal. The CM was then calibrated and searched against the top 1000 cluster seed sequences from the lattice round 15 library. Sequences that matched with the CM were then aligned against the CM, and the CM was further refined. This process repeated until either no more sequences matched against the CM or ten iterations of this process were done. Not surprisingly, when generating CMs from clusters known to contain this G-rich motif, hundreds of cluster seed sequences matched to the final CM. Unexpectedly, the CMs generated from clusters containing this G-rich motif no longer had any conserved

secondary structure after multiple refinements of the CMs in Infernal. Rather, sequence conservation of the G-rich motif among the CMs from the clusters containing this motif was high (Figure 3.16), suggesting the importance of the G-rich motif for aptamer binding to CA lattice. Due to the presence of four sets of G nucleotides, it is possible that this aptamer may form a G-quadruplex. This may explain why there was no conserved secondary structure present in the CMs, as G-quadruplexes do not form traditional Watson-Crick base pairs. Next, a series of truncations from the 3′ end of aptamer L15.7.1 were generated and tested for CA lattice binding. None of the truncated forms of L15.7.1 was able to bind to the CA lattice (Figure 3.17A). Since the aptamer no longer binds without its 3′ constant region but can bind lattice when an antisense oligonucleotide is annealed to its 3′ constant region, this suggests that the aptamer may require a structural element located 3′ of the G-rich sequence motif. Finally, we probed the structure of L15.7.1 using T1 RNase (cleaves after accessible G residues), S1 nuclease (cleaves single-stranded regions), and V1 RNase (cleaves double-stranded regions) under native conditions. The S1 reactions were done in the absence and presence of 1 mM $ZnSO_4$ because the S1 nuclease has optimal activity in the presence of 1 mM Zn^{2+}(44). Probing the structure of L15.7.1 using structure-sensitive nucleases generated a distinctive banding pattern (Figure 3.17B). Specifically, we observed strong S1 nuclease cleavage in the region containing the G-rich motif (~ nucleotides 51-78). Because L15.7.1 may contain a G-quadruplex, we generated a predicted secondary structure where the G-rich motif was forced to be single-stranded and not form base pairs with other parts of the aptamer using mfold(45), and the resulting secondary structure was visualized using NUPACK(46). When mapping the enzymatic cleavage patterns onto a predicted

secondary structure of L15.7.1 where the G-rich motif was forced to be single-stranded (Figure 3.17C), there was agreement between the digestion results and the predicted secondary structure. This suggests that the secondary structure of L15.7.1 may be similar to what was proposed in Figure 3.17C, but additional work is needed to determine the secondary structure of aptamer L15.7.1.

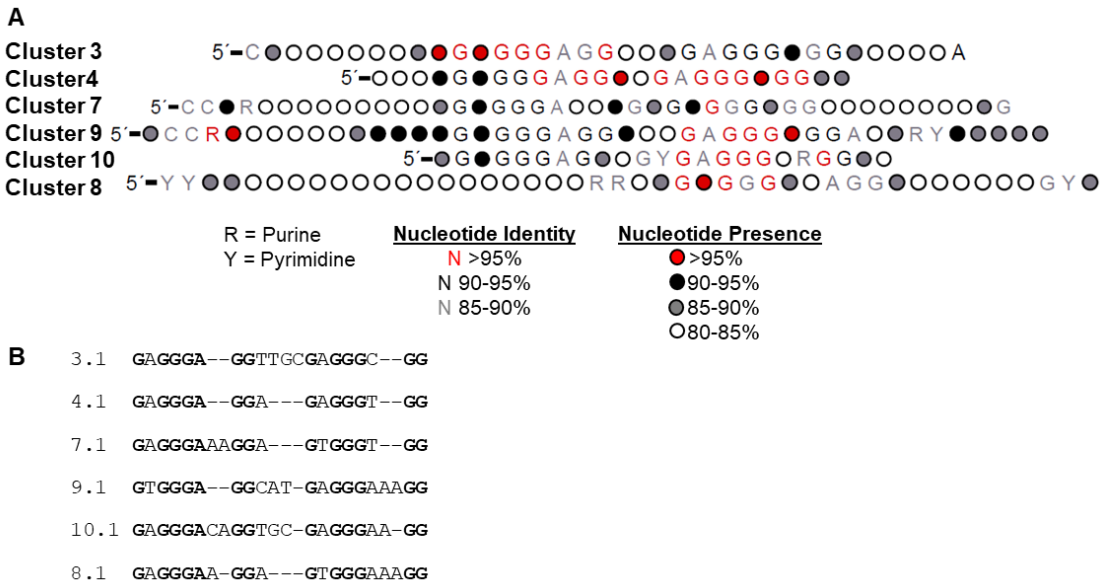

Figure 3.16: Sequence conservation among covariance models containing Motif 1. (A) Covariance models generated using Infernal for the top clusters that contain motif 1 are aligned. (B) Alignment of the G-rich motif 1 in the cluster seed sequences of the clusters that were used to generate the covariance models shown in (A).

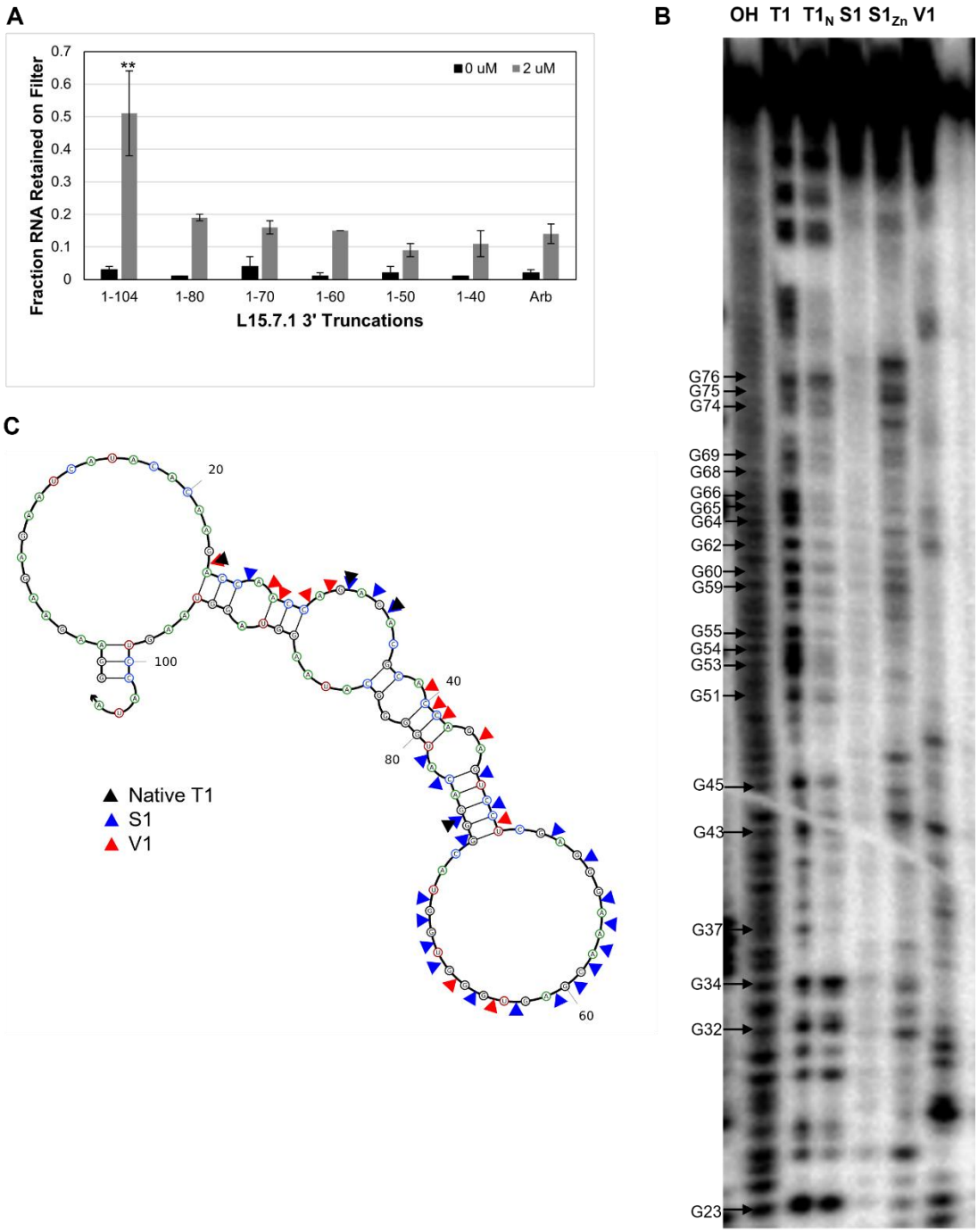

Figure 3.17: Exploration of the secondary structure of aptamer L15.7.1. (A) Binding of 3′ truncations to CA lattice assessed using nitrocellulose filter binding assays (n=3). Values are mean ± SD. ** (P <0.01 vs Arb). (B) Enzymatic probing of L15.7.1 using structure-sensitive nucleases. (C) Mapping of digestion results onto predicted secondary structure of L15.7.1.

For aptamer L15.6.1 (CA15-2), we previously determined that the 1-80 and 25-104 truncations were able to bind the CA lattice, while the 25-80 truncation was unable to bind the CA lattice(20). L15.6.1 was predicted to form a stem-loop structure consisting of nucleotides 30 through 60, and the GUGUAU motif (nucleotides 43-48) were predicted to be located within the loop. We observed strong S1 nuclease cleavage at these positions, and the G residues in this motif had strong T1 cleavage under native conditions (Figure 3.18A). When the cleavage patterns from the nuclease digestions were mapped onto the predicted secondary structure for aptamer L15.6.1 (Figure 3.18B), they were mostly in agreement with each other. Additionally, the CM generated from cluster 6 also had the GUGUAU motif in a loop (Figure 3.18C). Taken together, these results suggest that the structure of L15.6.1 is similar to the structure predicted by the structure prediction software. For the top five other GUGUAU-containing aptamers, the motif is also predicted to be a part of a hairpin loop for L15.1.1, L15.5.1, L15.19.1, and L15.21.1 and is a part of an internal loop in L15.14.1 (Figure 3.19). These predicted structures of GUGUAU-containing aptamers suggest that the motif is likely present in the structural context determined by the enzymatic probing and CM of L15.6.1.

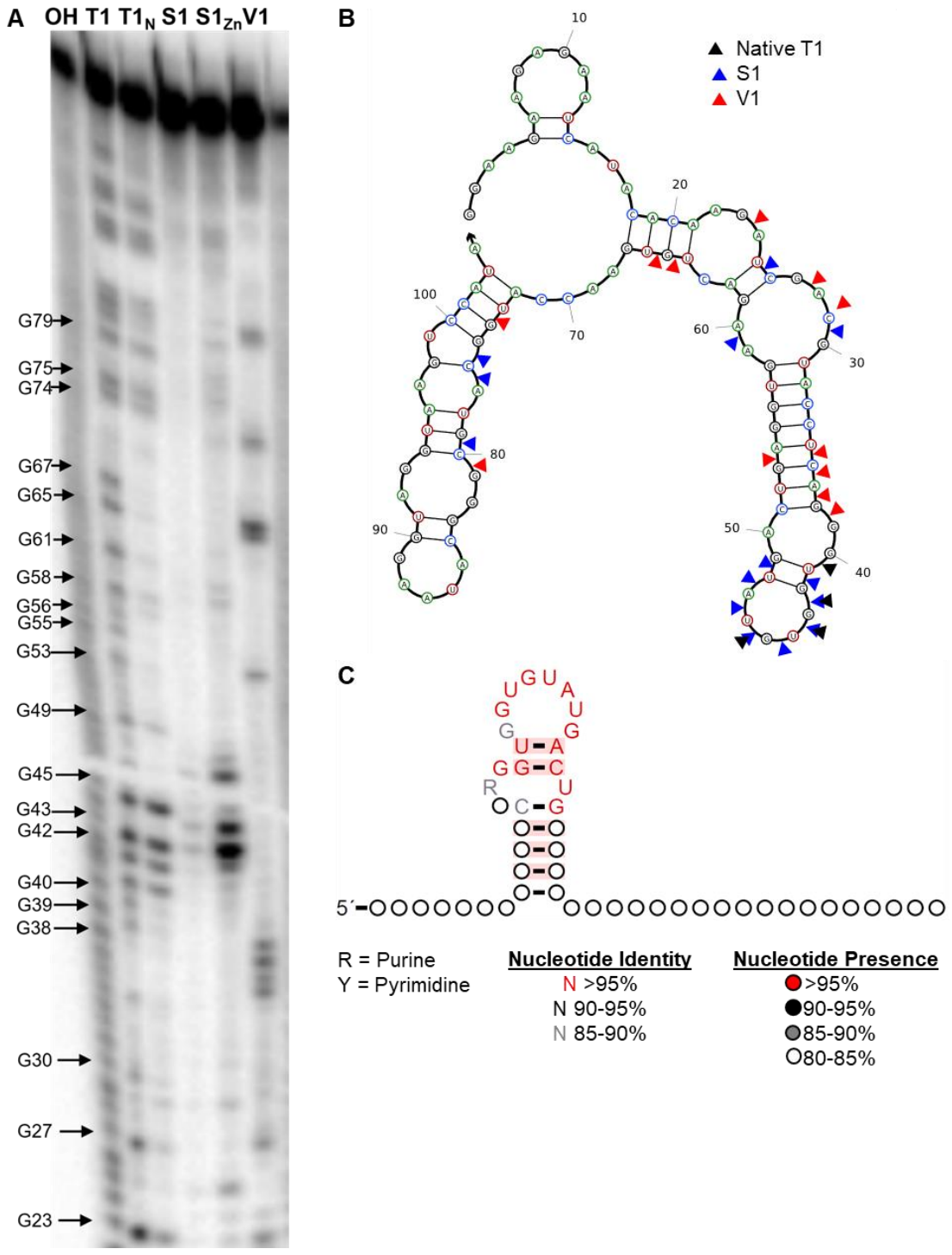

Figure 3.18: Exploration of the secondary structure of aptamer L15.6.1. (A) Enzymatic probing of L15.6.1 using structure-sensitive nucleases. (B) Mapping the cleavage products onto the predicted secondary structure of L15.6.1. (C) Covariance model generated starting from the cluster of L15.6.1. Red shading of base pairs means that there was no variation in the base pair when present.

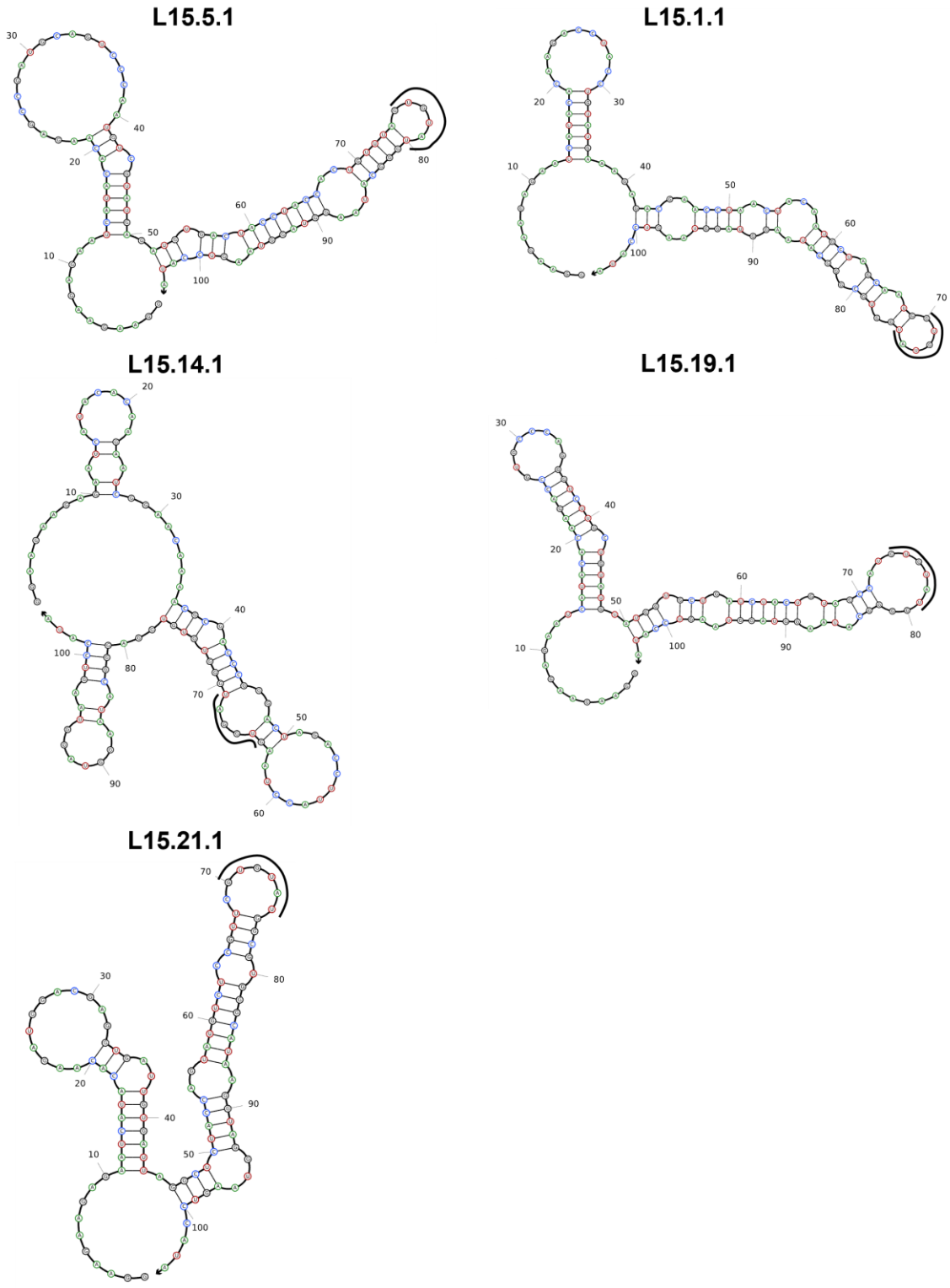

Figure 3.19: Predicted secondary structures of other GUGUAU-containing aptamers. For the aptamers shown above, the predicted secondary structures generated using NUPACK(46) are shown, and the GUGUAU motif is marked with a black line.

Aptamer L15.20.1 was used as the representative for motif 3, the other G-rich sequence motif. None of the 3′ truncations for aptamer L15.20.1 was able to bind CA lattice (Figure 3.20A). Like aptamer L15.7.1, it is possible that this aptamer requires a structural element located 3′ of the predicted sequence motif. The CM generated from its cluster predicted a conserved stem-loop structure and a highly-conserved, single-stranded G-rich sequence motif (Figure 3.20B). When the structure of this aptamer was probed using structure-sensitive nucleases, the cleavage patterns for the most part did not match with the predicted secondary structure of L15.20.1 when the structure from the CM was used as constraints in the secondary structure prediction software (Figure 3.20 C & D). This suggests that aptamer L15.20.1 likely folds into a different structure than the one that is predicted, and a preliminary 1D NMR experiment performed by the Heng lab suggests that this aptamer may have multiple conformations (data not shown). Due to the presence of four sets of G residues within its sequence, it is possible that this aptamer may fold into a G-quadruplex.

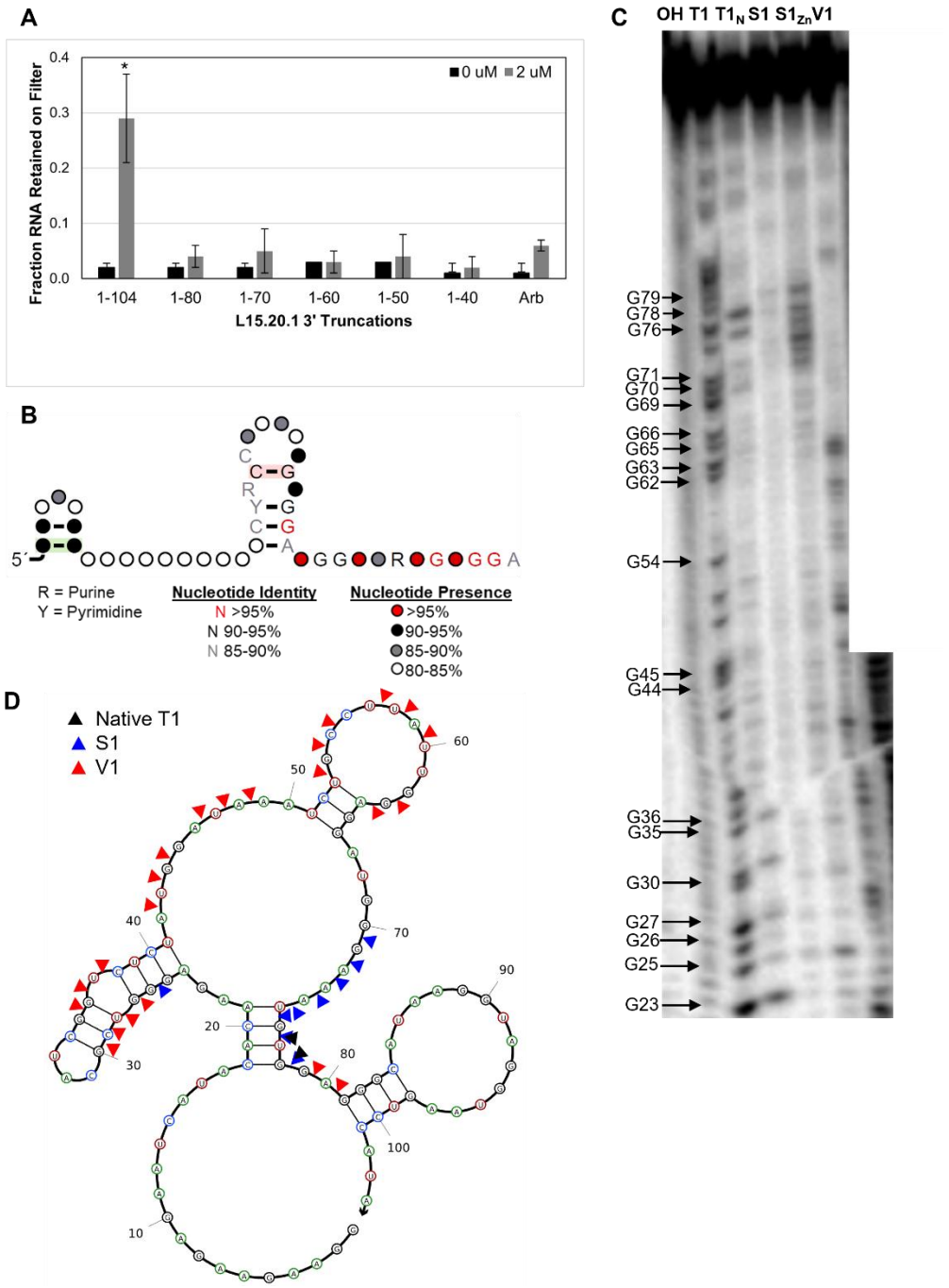

Figure 3.20: Exploration of the secondary structure of aptamer L15.20.1. (A) 3′ truncations of L15.20.1 were assessed for lattice binding using nitrocellulose filter binding assays (n = 3). Values are mean ± SD.* ($P < 0.05$). (B) Covariance model generated from L15.20.1 and its cluster. Red shading of base means that there was no variation in the base pair. (C) Enzymatic probing of L15.20.1 using structure-sensitive nucleases. (D) Mapping the cleavage products onto the secondary structure prediction for L15.20.1 based on the covariance model.

Aptamer H7.10.1 can bind both CA lattice and hexamer. Considering that this aptamer can bind both CA lattice and hexamer, and most aptamers in the libraries cannot bind CA hexamer, it was not surprising that it was not identified to contain a common sequence motif found in the top 1000 cluster seed sequences from the lattice round 15 library or the top 1000 cluster seed sequences from the L17+H7 library during the sequence motif search using MEME Suite(37). Therefore, other approaches were necessary to uncover the requirements for this aptamer to bind CA lattice and hexamer. The 1-80 truncation was still able to bind to the CA lattice to the same degree as the full-length aptamer (Figure 3.21A). Therefore, this aptamer does not require its 3′ constant region for binding. In addition, the 1-70 and 1-60 truncations exhibited significantly more binding to CA lattice than Arb but less than the full-length aptamer. A CM was generated from the alignment of the 51 sequences from H7.10.1's cluster and the predicted consensus structure (Figure 3.21B). When this CM was searched against the top 1000 cluster seed sequences from the lattice round 15 library or the L17+H7 library using Infernal, no other cluster seed sequences matched. The secondary structure of H7.10.1 was probed by structure-sensitive nucleases (Figure 3.21C). The cleavage patterns from the enzymatic digestions mapped well onto the predicted secondary structure of the 1-80 truncation of H7.10.1 (Figure 3.21D). Taken together, these results suggest that aptamer H7.10.1 may form a stem-loop structure.

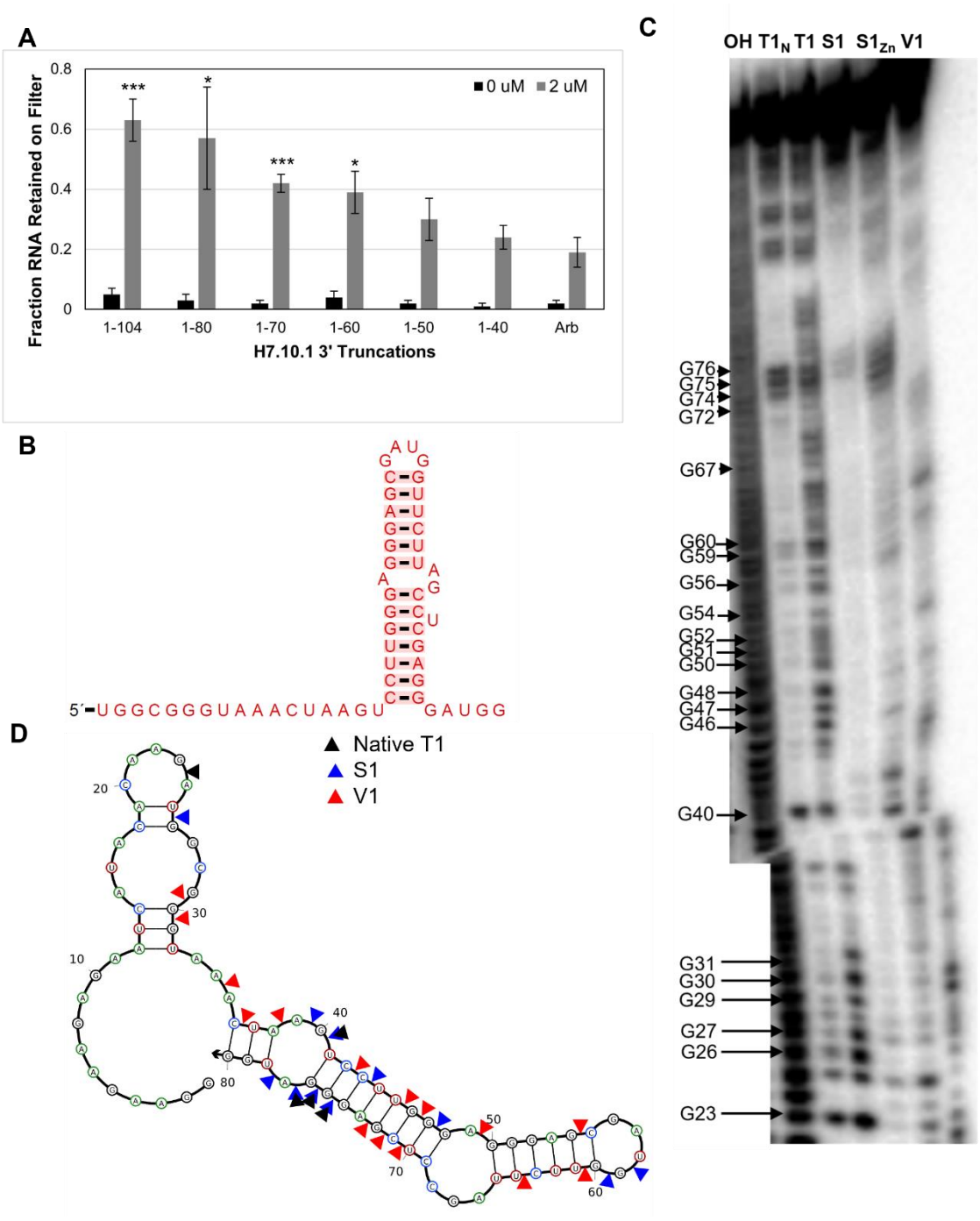

Figure 3.21: Exploration of the secondary structure of aptamer H7.10.1. (A) 3′ truncations of H7.10.1 were assessed for binding to CA lattice using nitrocellulose filter binding assays (n = 3). Values are mean ± SD. * (P value <0.05), *** (P <0.001). (B) Covariance model generated from the sequences from H7.10.1's cluster. The nucleotide identity for all positions was greater than 95% (red letters). Red shading of base pairs refers to no variation within the base pairs (C) Probing of H7.10.1 structure using structure-sensitive nucleases. (D) Mapping the cleavage patterns onto the predicted secondary structure of the 1-80 truncation of H7.10.1.

3.3.9. Assessing the Potential Presence of G-Quadruplexes in Motifs 1 and 3

Since the sequence motifs that L15.7.1 and L15.20.1 are highly G-rich and potentially could form G-quadruplexes, we wanted to determine whether binding to CA lattice by these aptamers was sensitive to the binding buffer conditions. It is known that potassium ions stabilize G-quadruplexes more than sodium ions(47). Therefore, we replaced the KCl in the binding buffer with NaCl and performed binding assays to CA lattice with the two binding buffers. We observed that aptamers L15.6.1 and L15.20.1 were not sensitive to buffer conditions, as they bound CA lattice in both binding buffers to the same degree (Figure 3.22). In contrast, aptamer L15.7.1 did not bind the CA lattice significantly more than the arbitrary control in the buffer where the potassium ions were replaced with sodium ions, suggesting that this aptamer could be a G-quadruplex. Additional studies will be needed to confirm these observations.

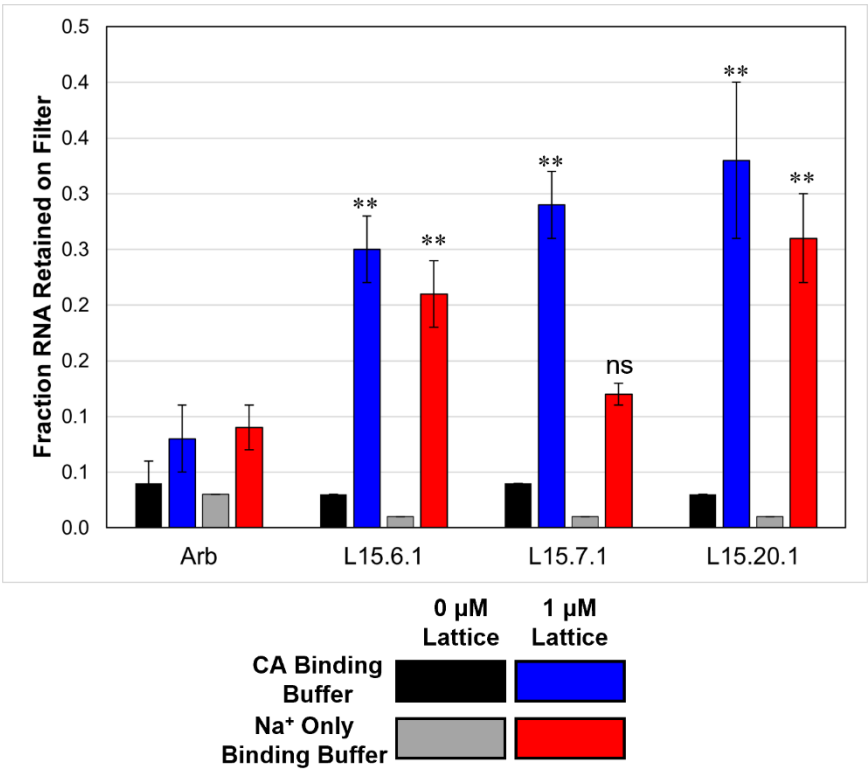

Figure 3.22: Assessing binding of aptamers to CA lattice in different binding buffers. Nitrocellulose filter binding assays were performed to determine whether aptamer binding was affected when the potassium ions in the capsid binding buffer were replaced with sodium ions (n = 3). Values are mean ± SD. Statistical comparisons were made against the binding values of the arbitrary control in the same buffer. ** ($P<0.01$), ns ($P>0.05$).

3.4. Conclusions and Future Studies

The initial goal of this work was to identify aptamers that were either lattice-specific binders or capable of binding both CA lattice and hexamer that can be used to probe the roles of CA assembly states in HIV-1 replication. To accomplish this, we performed a CA differentiation selection starting from the lattice round 15 library. This selection was composed of three selection trajectories. Two trajectories were focused on the CA hexamer in which one trajectory selected for sequences that bound both lattice and hexamer and a hexamer depletion trajectory that removed CA hexamer binders from the population. The third trajectory was against nitrocellulose, as nitrocellulose filters were used as the partition method for CA hexamer rounds. Seven aptamer libraries were submitted for high-throughput sequencing. Candidate aptamers were identified based on their enrichment profiles among the different trajectories and their binding specificities to CA lattice and hexamer were predicted. Seventeen of the 23 aptamers screened for binding to CA lattice and hexamer had their predicted binding specificities match their actual binding specificities to CA lattice and hexamer. Importantly, we were able to identify aptamers capable of binding both CA lattice and hexamer and more aptamers that were lattice specific, such as L15.6.1 (CA15-2), suggesting that aptamers can be used to differentiate among CA assembly states.

This result of identifying aptamers that are capable of binding CA lattice and hexamer or lattice only highlighted the utility of this selection approach for identifying aptamers with desired binding specificities. We were able to determine how sequences performed during the different selection trajectories by performing HTS and bioinformatic analyses, and the enrichment profiles of sequences and their respective

clusters could be used to predict binding specificities to CA lattice and hexamer. Typically, for an aptamer selection against a protein target starting with a random library, the success rate for identifying aptamer was found to be less than 30%(48). This approach successfully predicted the binding specificities for 74% of the aptamer candidates, which suggests that this approach can be used for other aptamer selections against different targets. These results also provide insights for how to improve the selection. First, it may be better to use a pre-enriched library that was less converged. For the lattice and hexamer trajectory, aptamers that had strongly depleted generally were not able to bind CA hexamer. In contrast, aptamers that had enrichment values approximately 1 or greater within this trajectory were less likely to actually be lattice and hexamer binders. This is in part due to the lattice round 15 library likely not containing many lattice and hexamer binders, and, therefore, it is difficult to select for aptamers that are not present within the library. A pre-enriched library that is less converged will have higher sequence diversity and less sequences that are dominant in the population. The position selection trajectory against nitrocellulose filters which was used in the partition step for the selection rounds against CA hexamer was useful to limit the likelihood of identifying false positive aptamers. It may be better to do the same number of rounds against nitrocellulose as the positive rounds against CA hexamer so that the enrichment values for sequences present in both libraries can be directly compared.

Next, we began to characterize the sequence and structural requirements for aptamer binding to CA assemblies. We learned that nearly three-quarters of the top 1000 aptamer clusters from the lattice round 15 population contained a highly conserved G-rich motif, and we focused our studies on representative aptamers that contained the three

most abundant sequence motifs (which collectively account for approximately 88% of the top 1000 clusters in the lattice round 15 library) and the lattice and hexamer-binding aptamer H7.10.1. Using a combination of testing aptamer truncations, enzymatic probing using structure-sensitive nucleases, and generating covariance models from aptamer cluster sequence data, we were able to begin uncovering the important sequence and structural features of these aptamers for CA binding. In particular, we found that aptamer H7.10.1 contains sequence and structural features that are not present in the top 1000 cluster seed sequences from lattice round 15 or the other top 1000 cluster seed sequence from the L17+H7 library. This suggests that aptamers capable of binding to different CA assemblies likely have different sequence and structural requirements.

Future work that will be done prior to manuscript submission is outlined below. We will continue to elucidate the sequence and sequence requirements for aptamer binding to capsid proteins. First, we will test the covariance models by testing aptamer mutants for lattice binding. For aptamer L15.7.1, the mutants to be tested include variants with the G residues mutated to potentially disrupt the formation of a G-quadruplex and mutations that transform L15.7.1 sequence to match the consensus G-rich sequence motif. For aptamer L15.20.1, the mutants to be tested include variants that disrupt and rescue the predicted stem-loop structure and variants that mutate the conserved sequence motifs, including the G-rich motif. For aptamer L15.6.1, a variant that mutates the GUGUAU motif was generated, in addition to two additional truncated forms. For aptamer H7.10.1, variants that disrupt and rescue the predicted stem-loop structure were made. Additional studies will explore the presence of G quadruplexes within some of the predicted motifs, including circular dichroism spectroscopy studies. Circular dichroism

spectroscopy is a useful technique for monitoring changes in nucleic acid secondary structure due to changes in buffer composition(49). For G quadruplexes, the spectra are predicted to change when buffers containing K^+, Na^+, or Li^+ are used. We also want to determine whether these aptamers inhibit HIV infectivity in producer or target cell assays. It will be interesting to learn whether aptamers containing the same structural motifs have similar inhibition patterns and whether aptamers containing different structural motifs utilize different mechanisms for inhibition. Previously, we had shown that CA15-2 (L15.6.1) had inhibited HIV infectivity in producer cell assays but not target cell assays(20). The aptamers in this study were cloned into the aptamer expression plasmid, and these assays are in progress. Potentially, to determine where on the capsid the aptamers bind, we will perform footprinting assays and cryo-electron microscopy. Future work that is beyond the scope of this manuscript will be discussed in Chapter 5.

3.5. Materials and Methods

3.5.1. Reagents

Unless otherwise noted, all chemicals were purchased from Sigma-Aldrich (St. Louis, MO). DNA templates for aptamers and all primers were ordered from Integrated DNA Technologies (Coralville, IA). The sequences of the RNAs and DNA primers used in this study are in Table 3.14.

Plasmid constructs for expression and purification of soluble HIV-1 CA hexamer (pET11a-CA (A14C, E45C, W184A and M185A)), CA hexamer lattice (pET11a-CA (A14C and E45C)), and CA monomer (pET11a-CA) were kindly provided by Dr. Owen Pornillos and Dr. Barbie Ganser-Pornillos (University of Virginia).

Table 3.14: RNA and primer sequences used in this study[a]

Name	Sequence
L15.1.1	ggaagaagagaaucauacacaagaGCCGATGCAGTCCCAATGTCGTATGAGATGTGACTACCTACCACTGTTTAGTGTATgggcauaagguagguaaguccaua
L15.2.1	ggaagaagagaaucauacacaagaCCACCCATGCCAAGGAAGGAGGTCGGAGGAGGTCACAGGCATTGGTGTGCTAGGATgggcauaagguagguaaguccaua
L15.3.1	ggaagaagagaaucauacacaagaCGCGCAAGCTTTCACGAGGGAGGTTGCGAGGGCGGATGTGGTTTGCGACGTCGTAgggcauaagguagguaaguccaua
L15.4.1	ggaagaagagaaucauacacaagaCGTGCTCATACCGCCGAATACGAGGGAGGAGAGGGTGGCAAAACTGCGTGGGTTgggcauaagguagguaaguccaua
L15.5.1	ggaagaagagaaucauacacaagaCCTACCTGTATGAAAGAGACGAACCTAACTACAATGCTAGCAATGGTGTATGGTGCgggcauaagguagguaaguccaua
L15.6.1	ggaagaagagaaucauacacaagaTCGACGTACCTCAGGGTGGTGTATGACTGAGGTGAAGACTGTGAACCATGGCATGCgggcauaagguagguaaguccaua
L15.7.1	ggaagaagagaaucauacacaagaCCAACCAGAGACGCACCAGAGTCCTCGAGGGAAAGGAGTGGGTGGTACGGGACATGgggcauaagguagguaaguccaua
L15.8.1	ggaagaagagaaucauacacaagaCCATCCCCCAACTTGTCAAGAGGGAAGGAGTGGGAAAGGAAGACAAGGACATGTGgggcauaagguagguaaguccaua
L15.9.1	ggaagaagagaaucauacacaagaCCCGGCCCCGCCAGTGGAAGTGGGAGGCATGAGGGAAAGGTTAAACACTGTCGGCTgggcauaagguagguaaguccaua
L15.10.1	ggaagaagagaaucauacacaagaGGACAGAACAACCCTTGAACGAATGAGGGACAGGTGCGAGGGAAGGATTCGTTGGAgggcauaagguagguaaguccaua
L15.13.1	ggaagaagagaaucauacacaagaTACCGCGACCTACCCCAGTGCAGGCTAGTGAGGGAGGAGGGTGGCATCTAGACACTgggcauaagguagguaaguccaua
L15.14.1	ggaagaagagaaucauacacaagaATCGGAACAAAACGCGACCCGGGACTAGACCTTACCTAAGTGGATGGGTGTGTGGAgggcauaagguagguaaguccaua
L15.15.1	ggaagaagagaaucauacacaagaCGCCAAATTGCTCAGTCCAGTGCGGGAGGTGAGGGCGGAACAACATGACGGACCAgggcauaagguagguaaguccaua
L15.16.1	ggaagaagagaaucauacacaagaGGCGAATTACACGACAATCTGGACGACCTACACTATGAGAGGAGGGTAGAGTGGAAgggcauaagguagguaaguccaua
L15.17.1	ggaagaagagaaucauacacaagaTGCGCGCGATGCCCGTACCATGCTATGTGGGTGGCGAGGGAGGACACTGCATGTTGgggcauaagguagguaaguccaua
L15.19.1	ggaagaagagaaucauacacaagaCCGTGCCCAGGGTCTTGCTGTATGTATGGTGCTGATCTACTGTAGCCATGTGTATGgggcauaagguagguaaguccaua
L15.20.1	ggaagaagagaaucauacacaagaGGGTCGCATCGGTCTCTATGGATAAATCTGCCTTATTGGAGGATGGGAAATGTGGAgggcauaagguagguaaguccaua
L15.21.1	ggaagaagagaaucauacacaagaTGGACGAGGTGATTGTGATTAGGCTCTACCAGTATGTCTCCGTTCGTGTATGGCGTgggcauaagguagguaaguccaua
L15.23.1	ggaagaagagaaucauacacaagaCCCCCCCCAAAGATACTGGCGCATGTGAGGGAGGCGAGGGAGGTACGCGTCCTTTgggcauaagguagguaaguccaua
L15.45.1	ggaagaagagaaucauacacaagaAGAGAATAGGCAAATACACCAATCTACCATAATGCTGGCGGAGGGAGAATGAGGAGgggcauaagguagguaaguccaua
H7.10.1	ggaagaagagaaucauacacaagaTGGCGGGTAAACTAAGTCCTTGGGAGGGAGCGATGGTTCTTAGCCTCGAGGGATGGgggcauaagguagguaaguccaua

H7.12.1	ggaagaagagaaucauacacaagaGGGTCGCATCGGTCTCTATGGATAACCCTGCCCGGA AGTGGAGGAATGTGGAgggcauaagguagguaaguccaua
H7.20.1	ggaagaagagaaucauacacaagaTGGACGAGGTGTATTTGGGCAGACCTTACCTAAGTG GATGGGTGTGTGGAgggcauaagguagguaaguccaua
Arbitrary RNA (Arb)	gggaaaagcgaaucauacacaagaAAUUUGGACUUUCCGCCCUUCUUGGCCUUUAUGAG GAUCUCUCUGAUUUUUCUUGCGUCGAGUUUUCCGGgggcauaagguauuuaauuccaua
15-2 Scrambled	ggaagaagagaaucauacacaagaGGAGGUGCAGGUGCAUAAUGCGUGAAGGCUCGGCC UGUACACUCGAAGUACUUAGUgggcauaagguagguaaguccaua
56N Forward Primer	GCCTAATACGACTCACTATAGGAAGAAGAGAATCATACACAAGA
56N Reverse Primer	TATGGACTTACCTACCTTATGCCC
Arb Fowrard Primer	GCCTAATACGACTCACTATAGGGAAAAGCGAATCATACACAAGA
Arb Reverse Primer	TATGGAATTAAATACCTTATGCCC
56N 5′ Anti-leader	TCTTGTGTATGATTCTCTTCTTCC

[a]For the RNA sequences, the constant regions are shown in lowercase.

3.5.2. Expression and purification of HIV-1 CA proteins and crosslinking of CA hexamers and CA lattice tubes

HIV-1 CA protein assemblies were expressed and purified as previously described (32,50). Briefly, the mutant HIV-1 CA proteins were expressed by IPTG induction in *Escherichia coli* BL21(DE3) cells for 6–12 hr at 25°C. The monomeric p24 CA protein was precipitated using 25% ammonium sulfate, and the pellet was resuspended in 20 mM Tris pH 8.0, 40 mM NaCl, and 60 mM βME, followed by purification using a Hi-Trap Q-Sepharose column. The purified protein fractions were concentrated and reconstituted in 25 mM Tris-HCl (pH 8.1) and 40 mM NaCl, flash frozen and stored at -80°C.

Soluble CA hexamers (pET11a-CA (A14C/E45C/W184A/M185A)) and CA lattice tubes (pET11a-CA (A14C/E45C)) were assembled as previously described (32,51). Briefly, assembly was performed *in vitro* by sequential overnight dialysis. Pooled and concentrated fractions were dialyzed into assembly buffer (50 mM Tris, pH 8.1, 1.0 M NaCl, 200 mM βME) followed by dialysis into storage buffer (20 mM Tris, pH 8.1, 40 mM NaCl). The integrity of the assembled CA lattice tubes was verified using transmission electron microscopy (TEM) with assistance from the University of Missouri Electron Microscopy Core. The purity and integrity of monomeric p24, assembled soluble CA hexamers and CA lattice were confirmed by non-reducing SDS-PAGE. Importantly, for the purposes of this study, all CA protein concentrations are reported in terms of the total CA monomer.

3.5.3. CA Differentiation Selection

The starting library for this work was the HIV-1 assembled capsid lattice round 15 libraries described previously(20). To initiate the differentiation selection, double-stranded DNAs from this library were first PCR-amplified using Taq DNA polymerase for 25 cycles and then transcribed using the Y639F mutant T7 RNA polymerase purified in house (52), *in vitro* transcription buffer (50 mM Tris-HCl pH 7.5, 15 mM $MgCl_2$, 5 mM DTT, 2 mM spermidine), and 2 mM each NTP (ATP, CTP, GTP, and UTP). Reactions were incubated at 37°C for a minimum of 4 hr and halted by adding an equal volume of denaturing gel loading buffer (90% formamide and 50 mM EDTA with trace amounts of xylene cyanol and bromophenol blue). RNAs were purified by denaturing polyacrylamide gel electrophoresis (6% TBE-PAGE, 8 M urea). Bands corresponding to the expected product sizes were visualized by UV shadow, excised, and eluted overnight in 300 mM sodium acetate (pH 5.4) at room temperature. Eluates were ethanol precipitated, resuspended in nuclease-free water and stored at -20°C until further use. RNA concentrations were determined using a NanoDropOne spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

For each selection trajectory, the order and types of selection rounds performed are shown in Figure 3.1. Samples from each selection round were named according to the selection targets used and the number of selection rounds against each target. For example, the Lattice 17 + Hexamer 7 population had gone through 17 rounds of positive selection against the CA lattice and 7 rounds of positive selection against the CA hexamer, while the Lattice 17 – Hexamer 3 had gone through 17 rounds of positive selection against CA lattice and three rounds of negative selection against CA hexamer.

For each selection round, 0.5 nmol RNA (~3.0 x 10^{14} molecules) was used and was refolded by heating to 90°C for 90 sec followed by cooling to room temperature for 5 min, followed by addition of 10X binding buffer to a final concentration of 1X (50 mM Tris-HCl [pH 7.5], 100 mM KCl, 50 mM NaCl, 1 mM $MgCl_2$). For positive selection rounds against CA lattice or CA hexamer, 100 pmol of the respective protein was added (500 nM RNA to 100 nM CA protein). For negative selection rounds against hexamer, 2.5 nmol CA hexamer was added (500 nM RNA to 2500 nM CA hexamer). For the nitrocellulose trajectory, no protein was added to the binding reaction. The final volume for the binding reactions was 1 mL.

Binding reactions were incubated at 37°C for 15 min followed by partitioning of the bound versus unbound RNA species. For positive selection rounds against CA lattice, partitioning was done as described previously using centrifugation(20). Briefly, the binding reaction was centrifuged at 16,000 x g for 10 min at 4°C. Following centrifugation, the supernatant was removed, and the CA lattice pellet was washed with 1X binding buffer. The sample was then centrifuged again at 16,000 x g for 10 min at 4°C. The wash step was performed a total of two times. CA lattice-bound RNA was recovered using phenol/chloroform extraction and ethanol precipitation. For the rounds against hexamer or nitrocellulose, the partitioning step was done using alkaline-treated nitrocellulose filters to decrease non-specific nucleic acid retention on the filter. Alkaline-treatment of the nitrocellulose filters involves pre-incubating the filters in 0.5 M KOH for 20 min, washing them extensively with MilliQ water, and then incubating them in 1X binding buffer for at least 45 min as previously described (53). While the binding reactions were incubating, the treated filters were placed on a sampling manifold and pre-

wet with 1 mL 1X binding buffer. Immediately before applying the binding reactions to the filter, the nitrocellulose filter was washed again with 1 mL 1X binding buffer under vacuum. The reaction was then applied to the filter and immediately wash with 1 mL 1X binding buffer. The filters remained under vacuum for an additional 2-3 min. For positive rounds against hexamer and rounds against nitrocellulose, the filter was incubated in 400 μL extraction buffer (8 M urea, 10 mM EDTA, 50 mM NaCl) and 300 μL phenol/chloroform for 10 min at room temperature. RNA was recovered by phenol/chloroform extraction and ethanol precipitation. For negative rounds against CA hexamer, the RNA in the flow-through (i.e., non-hexamer binders) was recovered using ethanol precipitation. The recovered RNA was reverse transcribed using ImProm-II Reverse Transcriptase (Promega, Madison, WI) and PCR amplified using Taq DNA polymerase for 10 cycles to generate the transcription template for the next round of selection.

3.5.4. Illumina Sequencing

The libraries specified in Figure 3.1 were prepared for sequencing using a series of PCR steps to append the Illumina adapters and sequencing indices for the multiplexing of the libraries as previously described(54). The primers used to append the Illumina adapters and sequencing indices can be found in Table 3.15. Sequencing was performed on an Illumina NextSeq 500 (University of Missouri Genomics Technology Core) to generate 75 bp paired-end reads. The populations were demultiplexed, and the paired-end reads were assembled into sequences using PANDASEQ(55) with a minimum overlap of 45 nt and a quality threshold of 0.6 to remove low-quality assemblies. Data preprocessing

was performed using cutadapt(56) to trim the 5′ and 3′ constant regions from sequences and to discard any uncut sequences or sequences with lengths not within ± 6 nt of the expected size (56 nt) after trimming. These populations were then analyzed using the FASTAptamer software(33) and its R-based re-implementation FASTAptameR 2.0(34) to count and normalize reads (FASTAptameR-Count), calculate fold enrichment and compare populations (FASTAptameR-Sequence Enrichment), group related sequences together (FASTAptameR-Cluster), analyze cluster composition (FASTAptameR-Cluster Diversity), and calculate fold enrichment of clusters (FASTAptameR-Cluster Enrichment). Aptamers were named according to the cluster in which they were found in lattice round 15 and their rank in terms of abundance with the cluster. If the aptamer was not present in lattice round 15, the aptamer was named according to the cluster in which it was found in the lattice 17 + hexamer 7 population and its rank within the cluster in terms of abundance. For example, L15.6.1 was the most abundant aptamer within cluster 6 of lattice round 15, while H7.10.1 was the most abundant aptamer within cluster 10 of the lattice 17 + hexamer 7 population.

Table 3.15: Primer Sequences to Append Illumina Adapters and Sequencing Indices

Library	PCR 1 Forward	PCR 1 Reverse	PCR 2 Forward	PCR 2 Reverse	PCR 3 Forward	PCR 3 Reverse[a]
Lattice Round 8	TAATACGACTCACTATAGGACGCAGGGCTTCATGCACAAGA	TATGGACTTACCTACCTTATGCCC	TTTCCCTACACGACGCTCTTCCGATCTCGCAGGGCTTCATGCACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATGGACTTACCTACCTTA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	CAAGCAGAAGACGGCATACGAGAT**CGAGTAAT**GTGACTGGAGTTCAGACGTG
Lattice Round 10	TAATACGACTCACTATAGGAAGCATCTATTCATGCACAAGA	TATGGACTTACCTACCTTATGCCC	TTTCCCTACACGACGCTCTTCCGATCTGCATCTATTCATGCACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATGGACTTACCTACCTTA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	CAAGCAGAAGACGGCATACGAGAT**TCTCCGGA**GTGACTGGAGTTCAGACGTG
Lattice Round 15	TAATACGACTCACTATAGGAAGTAGCTACTCATGCACAAGA	TATGGACTTACCTACCTTATGCCC	TTTCCCTACACGACGCTCTTCCGATCTTAGCTACTCATGCACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATGGACTTACCTACCTTA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	CAAGCAGAAGACGGCATACGAGAT**AATGAGCG**GTGACTGGAGTTCAGACGTG
Lattice 15 Nitrocellulose 2 (L15nc2)	TAATACGACTCACTATAGGAAGAATTGACTCATACACAAGA	TATGGACTTACCTACCTTATGCCC	TTTCCCTACACGACGCTCTTCCGATCTATTGACTCATACACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATGGACTTACCTACCTTA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	CAAGCAGAAGACGGCATACGAGAT**GGAATCTC**GTGACTGGAGTTCAGACGTG
Lattice 17 + Hexamer 3 (L17+H3)	TAATACGACTCACTATAGGACGCAGGGCTTCATGCACAAGA	TATGGACTTACCTACCTTATGCCC	TTTCCCTACACGACGCTCTTCCGATCTCGCAGGGCTTCATGCACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATGGACTTACCTACCTTA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	CAAGCAGAAGACGGCATACGAGAT**TTCTGAAT**GTGACTGGAGTTCAGACGTG
Lattice 17 + Hexamer 7 (L17+H7)	TAATACGACTCACTATAGGAAGCATCTATTCATGCACAAGA	TATGGACTTACCTACCTTATGCCC	TTTCCCTACACGACGCTCTTCCGATCTGCATCTATTCATGCACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATGGACTTACCTACCTTA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	CAAGCAGAAGACGGCATACGAGAT**ACGAATTC**GTGACTGGAGTTCAGACGTG
Lattice 17 – Hexamer 3 (L17-H3)	TAATACGACTCACTATAGGAAGTAGCTACTCATGCACAAGA	TATGGACTTACCTACCTTATGCCC	TTTCCCTACACGACGCTCTTCCGATCTTAGCTACTCATGCACAAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTATGGACTTACCTACCTTA	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT	CAAGCAGAAGACGGCATACGAGAT**AGCTTCAG**GTGACTGGAGTTCAGACGTG

[a]The index sequence is bolded in the PCR 3 reverse primer.

3.5.5. DNA Templates and RNA Transcription

For each aptamer to be transcribed, DNA oligonucleotides were purchased as 'left' and 'right' halves from Integrated DNA Technologies (Coralville, IA). Oligonucleotides corresponding to the 'right' half were 5′-phosphorylated using T4 polynucleotide kinase (New England Biolabs, Ipswitch, MA). Equimolar amounts of 'right' strand, 'left' strand, and a complimentary bridge oligo were annealed and ligated using T4 DNA ligase (New England Biolabs, Ipswitch, MA). Ligated templates were PCR amplified using Pfu DNA polymerase, a forward primer to append the T7 promoter, and a reverse primer complimentary to the 3′ constant region. Amplified products were verified for size using agarose gel electrophoresis. The double-stranded DNA templates were then transcribed *in vitro* and purified as above.

3.5.6. Nitrocellulose Filter Binding Assays

In vitro transcribed RNA was treated with Antarctic phosphatase (Fermentas, Waltham, MA) to remove the 5' terminal phosphate and then labeled by T4 polynucleotide kinase (New England Biolabs, Ipswitch, MA) in the presence of γ-^{32}P-labeled ATP (PerkinElmer, Waltham, MA). Radiolabeled RNAs were gel-purified using denaturing PAGE as described above. To evaluate dose-dependent binding of aptamer libraries or individual aptamers to CA proteins, 50 nM 5′-radiolabeled and refolded RNA was incubated with varying concentrations of CA lattice, CA hexamer, or CA monomer in 1X binding buffer at 37°C for 15 min. RNA:protein complexes were partitioned from unbound RNA by filtering samples through a pre-wet, alkaline-treated nitrocellulose filter under vacuum and immediately washing with 1 mL binding buffer. Radioactivity

retained on the filter was counted by placing filters into scintillation vials, adding 4 mL Emulsifier-safe liquid scintillation fluid (Perkin Elmer, Waltham, MA), and counting using a liquid scintillation counter. A separate aliquot of the same RNA was incubated without CA protein and applied to filters under vacuum to determine non-specific nucleic acid retention. An unfiltered 'No Wash' sample was counted to determine the total amount of radioactivity present in each binding reaction. The fraction of RNA retained on the filter was calculated by dividing the radioactivity retained on the filter by the radioactivity present in the 'No Wash' sample. At least three replicates were performed for each binding assay.

To determine the effect of buffer composition on aptamer binding, binding assay were performed as above using either the standard 1X binding buffer or 1X binding buffer where the KCl was replaced with NaCl (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$). Binding assays were done in triplicate.

To determine whether an aptamer retained binding to CA lattice when an antisense oligonucleotide was annealed to either its 5′ or 3′ constant region, 50 nM 5′-radiolabeled aptamer RNA (~10,000-20,000 cpm) and 62.5 nM antisense oligonucleotide were mixed on ice in 1X binding buffer. For thermal renaturation and annealing of the antisense oligonucleotide to the aptamer, the samples were transferred into a preheated aluminum insert within a dry heat block set to 90°C for 1 min, then the aluminum insert was removed from the heat block and placed on the benchtop and allowed to cool to 37°C. 2 μM CA lattice (or 0 nM CA lattice to determine non-specific binding) was then added to each reaction, and the reaction was incubated at 37°C for 15 min. The

nitrocellulose filter binding assay was then performed as described above to determine fraction RNA retained on filter.

3.5.7. Structural motif search using MEME Suite

The top 1000 cluster seed sequences were searched for conserved structural motifs using the MEME discovery tool as part of MEME Suite(37). The software was asked to find 20 motifs with a maximum width of 30 nucleotides, and the sequence motifs were prioritized based on the number of sequences that contained each motif.

3.5.8. Dye conjugation to antisense oligos

The 56N reverse primer with a 5'-amino group attached to a C-6 alkyl chain (5' Amino Modifier C6) was purchased from Integrated DNA Technologies (Coralville, IA) and used in a dye conjugation reaction with Cy5-NHS ester (Lumiprobe, Hunt Valley, MD). The conjugation reaction, subsequent reverse-phase HPLC purification, and concentration/buffer exchange of the Cy5-labeled 56N reverse primer were performed as described previously (57).

3.5.9. Electrophoretic Mobility Shift Assays (EMSA) Competition Assays

2 μM CA lattice was preloaded with increasing concentrations of refolded, unlabeled competitor RNAs (25 to 200 nM) in 1X binding buffer at 37°C for 10 min. A separate sample with 0 nM competitor was prepared in parallel to determine the maximum binding of the labeled aptamer (i.e., aptamer annealed to Cy5-labeled antisense oligonucleotide). A 10X annealing reaction containing 500 nM aptamer RNA and 450

nM Cy5-labeled 56N reverse primer in 1X binding buffer underwent the same thermal renaturation and annealing described above. The annealed aptamer:Cy5-labeled 56N reverse primer complex was then added to the binding reaction to a 1X final concentration (50 nM aptamer and 45 nM Cy5-labeled 56N reverse primer) and incubated at 37°C for 15 min. 4X native loading dye (2X tris-borate [TB, 2X = 180 mM Tris, 180 mM boric acid], 50% glycerol) was added to each reaction to a final concentration of 1X, and the samples were run on a 2% 0.5X TB-agarose gel at 8V/cm. Gels were scanned for Cy5 fluorescence using the Typhoon FLA 9000 phosphoimager (GE Healthcare, Chicago, IL). The fraction of aptamer annealed to Cy5-labeled oligo retained in wells relative to No Competitor was quantified using Multi Gauge software (Fujifilm) using the following equation: [(intensity of signal retained in well) – (intensity of signal in well of RNA only lane)]/[(intensity of signal retained in well of the No Competitor lane) – (intensity of signal in well of RNA only lane)]. Experiments were performed in triplicate.

3.5.10. Nitrocellulose Filter Binding Competition Assays

To determine whether an unlabeled competitor was able to compete with 5′-radiolabeled H7.10.1 for binding to the CA lattice, varying concentrations of unlabeled competitor (0.25-1 μM) were incubated with 1 μM CA lattice in 1X binding buffer at 37°C for 10 min. A separate sample with 0 nM competitor was prepared in parallel to determine the maximum binding of aptamer H7.10.1. Unlabeled competitor RNA was first refolded as described above. PF74 (Sigma Aldrich, St. Louis, MO) working stocks were prepared using a final concentration of 10% DMSO. 50 nM 5′-radiolabeled and

refolded H7.10.1 was then added to the binding reaction, which was incubated at 37°C for 15 min. Nitrocellulose filter binding assays were performed and the fraction of radiolabeled H7.10.1 retained on the filter was calculated as described above. To calculate the relative % radiolabeled H7.10.1 retained on filter, the fraction of radiolabeled H7.10.1 retained on filter at each concentration of competitor was divided by the fraction radiolabeled H7.10.1 retained on filter when no competitor was present (0 nM) and multiplied by 100%. The assays were done in triplicate.

3.5.11. Enzymatic Probing of Aptamer Structures

The 5′ radiolabeled aptamer structures were probed using enzymatic digestions(44). To generate the OH ladder, 1 μL 5′ radiolabeled RNA (50,000-150,000 cpm/μL) was incubated in 50 mM sodium carbonate pH 9.0 buffer at 90°C for 10 min. The alkaline digestion reaction was stop by adding an equal volume of colorless loading dye (10 M urea, 15 mM EDTA) and snap cooling in a bath of dry ice and ethanol. A denaturing T1 digestion was performed to cut after each G residue by incubating1 μL 5′ radiolabeled RNA (50,000-150,000 cpm/μL) in 25 mM sodium citrate pH 5.0 buffer containing 7 M urea and 10.5 mM EDTA with 1 U RNase T1 (Thermo Fisher Scientific, Waltham, MA) at 55°C for 5 min. The reaction was stopped by adding an equal volume of colorless loading dye and snap cooling. For the native RNase T1, S1 nuclease, and VI RNase digestions, the 5′ radiolabeled RNA (50,000-150,000 cpm per reaction) was first heated in water at 90°C for 5 min and then allowed to cool at 0.1°C/s in a thermocycler. When the temperature reached approximately 65°C, 10X capsid binding buffer was added to the RNA to a concentration of approximately 1.1X. Once room temperature was

reached, the RNAs were placed on ice. Prior to adding an enzyme, the reaction was equilibrated to 37°C for 2 min. For the native T1 digestion, 0.1 U was added to the reaction and incubated at 37°C for 2 min. For the S1 nuclease (Thermo Fisher Scientific, Waltham, MA), the reactions were done with and without the addition of 1 mM $ZnSO_4$ prior to equilibration at 37°C. 20 U of S1 nuclease was added to the reaction and incubated at 37°C for 5 min. For RNase VI (Ambion, Austin, TX), 0.005 U were added to the reaction and incubated at 37°C for 5 min. The enzymatic digestions were stopped by adding an equal volume of colorless loading dye and snap cooling. The samples were run on a 0.4 mm thick, 8% denaturing polyacrylamide gel (TBE-8 M urea). The gel was exposed to a phosphoimager screen overnight at -20°C, and the phosphoimager screen was scanned using the Typhoon FLA 9000 phosphoimager (GE Healthcare, Chicago, IL). The structural probing experiments were done in triplicate to establish the reproducibility of the experiment. A representative gel of the results is shown in the figures.

3.5.12. Generation of Covariance Models using Infernal

Covariance models were generated following a similar protocol described previously for HIV-1 RT aptamers(54). For the top 10 clusters and cluster 20 from the lattice round 15 library, the 300 most abundant sequences within each cluster were aligned using MAFFT(58), and these alignments were used to predict a secondary structure using RNAalifold (weight of the covariance term = 0.6, penalty for non-compatible sequences = 6)(43). Both the MAFFT alignment and RNAalifold prediction were done within the Jalview Desktop application(59). The results from RNAalifold were exported as a Stockholm file within Jalview. To use the Stockholm file in Infernal, the

centroid structure was renamed SS_cons, and an extra '-' was removed from the end of the centroid structure line. The input Stockholm file containing the alignment and centroid structure predicted by RNAalifold were then used to generate a covariance model (CM) using the Infernal program(42). Infernal was then used to calibrate and search the covariance models against the population. CMs were searched against the top 1000 cluster seed sequences from the lattice round 15 library. For all CM searches, seed sequences matched with the CM if the expectation score was below an expectation score cut-off of 0.1. The default value is 0.01, which means about one false positive would be expected every 100 queries on average. To increase the number of hits, an expectation cut-off score of 0.1 was used, as done previously(54). Groups of clusters that were identified from the CM search were then aligned to the CM and used to generate a new CM. This process was repeated until either no more sequences matched the CM or up to 10 rounds total. The results from the final CM search were then visualized using R2R(60) to see sequence and/or structure conservation.

3.5.13. Statistical Analysis

To determine whether there was a statistically significant difference between two sets of samples, P values were calculated using an unpaired t test computed by GraphPad Prism.

3.5.14. Secondary Structure Prediction Software

Structure prediction throughout this work was performed using mfold(45) and the NUPACK web server (www.nupack.org) (46).

3.6. References

1. Rossi, E., Meuser, M.E., Cunanan, C.J. and Cocklin, S. (2021) Structure, function, and interactions of the HIV-1 capsid protein. *Life*, **11,** 100.
2. Campbell, E.M. and Hope, T.J. (2015) HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat. Rev. Microbiol.*, **13**, 471-483.
3. Novikova, M., Zhang, Y., Freed, E.O. and Peng, K. (2019) Multiple roles of HIV-1 capsid during the virus replication cycle. *Virol. Sin.*, **34**, 119-134.
4. Aiken, C. and Rousso, I. (2021) The HIV-1 capsid and reverse transcription. *Retrovirology*, **18**, 29.
5. Ambrose, Z. and Aiken, C. (2014) HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. *Virology*, **454-455**, 371-379.
6. Guedán, A., Caroe, E.R., Barr, G.C.R. and Bishop, K.N. (2021) The role of capsid in HIV-1 nuclear entry. *Viruses*, **13,** 1425.
7. Fassati, A. (2012) Multiple roles of the capsid protein in the early steps of HIV-1 infection. *Virus Res.*, **170**, 15-24.
8. Ganser-Pornillos, B.K., Yeager, M. and Pornillos, O. (2012) Assembly and architecture of HIV. *Adv. Exp. Med. Biol.*, **726**, 441-465.
9. Freed, E.O. (2015) HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.*, **13**, 484-496.
10. Briggs, J.A. and Krausslich, H.G. (2011) The molecular architecture of HIV. *J. Mol. Biol.*, **410**, 491-500.
11. Sundquist, W.I. and Krausslich, H.G. (2012) HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.*, **2**, a006924.

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

13. Yamashita, M. and Engelman, A.N. (2017) Capsid-dependent host factors in HIV-1 infection. *Trends Microbiol.*, **25**, 741-755.

14. Le Sage, V., Mouland, A.J. and Valiente-Echeverría, F. (2014) Roles of HIV-1 capsid in viral replication and immune evasion. *Virus Res.*, **193**, 116-129.

15. Lahaye, X., Satoh, T., Gentili, M., Cerboni, S., Conrad, C., Hurbain, I., El Marjou, A., Lacabaratz, C., Lelièvre, J.D. and Manel, N. (2013) The capsids of HIV-1 and HIV-2 determine immune detection of the viral cDNA by the innate sensor cGAS in dendritic cells. *Immunity*, **39**, 1132-1142.

16. Rasaiyaah, J., Tan, C.P., Fletcher, A.J., Price, A.J., Blondeau, C., Hilditch, L., Jacques, D.A., Selwood, D.L., James, L.C., Noursadeghi, M. *et al.* (2013) HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature*, **503**, 402-405.

17. Forshey, B.M., von Schwedler, U., Sundquist, W.I. and Aiken, C. (2002) Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J. Virol.*, **76,** 5667-5677.

18. Wacharapornin, P., Lauhakirti, D. and Auewarakul, P. (2007) The effect of capsid mutations on HIV-1 uncoating. *Virology*, **358,** 48-54.

19. Manocheewa, S., Swain, J.V., Lanxon-Cookson, E., Rolland, M. and Mullins, J.I. (2013) Fitness costs of mutations at the HIV-1 capsid hexamerization interface. *PLoS ONE*, **8**, e66065.

20. Gruenke, P.R., Aneja, R., Welbourn, S., Ukah, O.B., Sarafianos, S.G., Burke, D.H. and Lange, M.J. (2022) Selection and identification of an RNA aptamer that specifically binds the HIV-1 capsid lattice and inhibits viral replication. *Nucleic Acids Res.*, **50**, 1701-1717.

21. Chen, L., Rashid, F., Shah, A., Awan, H.M., Wu, M., Liu, A., Wang, J., Zhu, T., Luo, Z. and Shan, G. (2015) The isolation of an RNA aptamer targeting to p53 protein with single amino acid mutation. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 10002-10007.

22. Alam, K.K., Chang, J.L., Lange, M.J., Nguyen, P.D.M., Sawyer, A.W. and Burke, D.H. (2018) Poly-target selection identifies broad-spectrum RNA aptamers. *Mol. Ther. Nucleic Acids*, **13**, 605-619.

23. Lange, M.J., Nguyen, P.D.M., Callaway, M.K., Johnson, M.C. and Burke, D.H. (2017) RNA-protein interactions govern antiviral specificity and encapsidation of broad spectrum anti-HIV reverse transcriptase aptamers. *Nucleic Acids Res.*, **45**, 6087-6097.

24. Hermann, T. and Patel, D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820-825.

25. Zichel, R., Chearwae, W., Pandey, G.S., Golding, B. and Sauna, Z.E. (2012) Aptamers as a sensitive tool to detect subtle modifications in therapeutic proteins. *PLoS ONE*, **7**, e31948.

26. Thirunavukarasu, D. and Shi, H. (2015) An RNA Aptamer specific to Hsp70-ATP conformation inhibits its ATPase activity independent of Hsp40. *Nucleic Acid Ther.*, **25**, 103-112.

27. Moore, M.D., Bobay, B.G., Mertens, B., Jaykus, L.-A. and Coyne, C.B. (2016) Human norovirus aptamer exhibits high degree of target conformation-dependent binding similar to that of receptors and discriminates particle functionality. *mSphere*, **1**, e00298-00216.

28. Kahsai, A.W., Wisler, J.W., Lee, J., Ahn, S., Cahill Iii, T.J., Dennison, S.M., Staus, D.P., Thomsen, A.R., Anasti, K.M., Pani, B. *et al.* (2016) Conformationally selective RNA aptamers allosterically modulate the β2-adrenoceptor. *Nature Chem. Biol.*, **12**, 709-716.

29. Weber, A.M., Kaiser, J., Ziegler, T., Pilsl, S., Renzl, C., Sixt, L., Pietruschka, G., Moniot, S., Kakoti, A., Juraschitz, M. *et al.* (2019) A blue light receptor that mediates RNA binding and translational regulation. *Nature Chem. Biol.*, **15**, 1085-1092.

30. Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818-822.

31. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.

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

33. Alam, K.K., Chang, J.L. and Burke, D.H. (2015) FASTAptamer: a bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Mol. Ther. Nucleic Acids*, **4**, e230.

34. Kramer, S.T., Gruenke, P.R., Alam, K.K., Xu D. and Burke, D.H. (2022) FASTAptameR 2.0: a web tool for combinatorial sequence selections. bioRxiv doi: https://doi.org/10.1101/2022.04.27.489774, 29 April 2022, pre-print: not peer-reviewed.

35. Schütze, T., Wilhelm, B., Greiner, N., Braun, H., Peter, F., Mörl, M., Erdmann, V.A., Lehrach, H., Konthur, Z., Menger, M. *et al.* (2011) Probing the SELEX process with next-generation sequencing. *PLoS One*, **6**, e29604.

36. Zimmermann, B., Gesell, T., Chen, D., Lorenz, C. and Schroeder, R. (2010) Monitoring genomic sequences during SELEX using high-throughput sequencing: neutral SELEX. *PLoS ONE*, **5**, e9169.

37. Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28-36.

38. Manochehry, S., McConnell, E.M. and Li, Y. (2019) Unraveling determinants of affinity enhancement in dimeric aptamers for a dimeric protein. *Sci. Rep.*, **9**, 17824.

39. Price, A.J., Jacques, D.A., McEwan, W.A., Fletcher, A.J., Essig, S., Chin, J.W., Halambage, U.D., Aiken, C. and James, L.C. (2014) Host cofactors and pharmacologic ligands share an essential interface in HIV-1 capsid that is lost upon disassembly. *PLoS Pathog.*, **10**, e1004459.

40. Rankovic, S., Ramalho, R., Aiken, C. and Rousso, I. (2018) PF74 reinforces the HIV-1 capsid to impair reverse transcription-induced uncoating. *J. Virol.*, **92,** e00845-18.

41. Bhattacharya, A., Alam, S.L., Fricke, T., Zadrozny, K., Sedzicki, J., Taylor, A.B., Demeler, B., Pornillos, O., Ganser-Pornillos, B.K., Diaz-Griffero, F. *et al.* (2014) Structural basis of HIV-1 capsid recognition by PF74 and CPSF6. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 18625-18630.

42. Nawrocki, E.P. and Eddy, S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**, 2933-2935.

43. Bernhart, S.H., Hofacker, I.L., Will, S., Gruber, A.R. and Stadler, P.F. (2008) RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinform*, **9**, 474.

44. Biondi, E. and Burke, D.H. (2014) RNA structural analysis by enzymatic digestion. *Methods Mol. Biol.*, **1086**, 41-52.

45. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406-3415.

46. Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M. and Pierce, N.A. (2011) NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.*, **32**, 170-173.

47. Bhattacharyya, D., Mirihana Arachchilage, G. and Basu, S. (2016) Metal cations in G-quadruplex folding and stability. *Front. Chem.*, **4,** 38.

48. Rohloff, J.C., Gelinas, A.D., Jarvis, T.C., Ochsner, U.A., Schneider, D.J., Gold, L. and Janjic, N. (2014) Nucleic acid ligands with protein-like side chains:

modified aptamers and their use as diagnostic and therapeutic agents. *Mol. Ther. Nucleic Acids*, **3**, e201.

49. Del Villar-Guerra, R., Trent, J.O. and Chaires, J.B. (2018) G-quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angew. Chem. Int. Ed.*, **57**, 7171-7175.

50. Yoo, S., Myszka, D.G., Yeh, C., McMurray, M., Hill, C.P. and Sundquist, W.I. (1997) Molecular recognition in the HIV-1 capsid/cyclophilin A complex. *J. Mol. Biol.*, **269**, 780-795.

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

52. Sousa, R. and Padilla, R. (1995) A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J.*, **14**, 4609-4621.

53. McEntee, K., Weinstock, G.M. and Lehman, I.R. (1980) recA protein-catalyzed strand assimilation: stimulation by Escherichia coli single-stranded DNA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.*, **77**, 857-861.

54. Ditzler, M.A., Lange, M.J., Bose, D., Bottoms, C.A., Virkler, K.F., Sawyer, A.W., Whatley, A.S., Spollen, W., Givan, S.A. and Burke, D.H. (2013) High-throughput sequence analysis reveals structural diversity and improved potency among RNA inhibitors of HIV reverse transcriptase. *Nucleic Acids Res.*, **41**, 1873-1884.

55. Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G. and Neufeld, J.D. (2012) PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinform.*, **13**, 31.

56. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.*, **17**, 3.

57. Porciani, D., Cardwell, L.N., Tawiah, K.D., Alam, K.K., Lange, M.J., Daniels, M.A. and Burke, D.H. (2018) Modular cell-internalizing aptamer nanostructure enables targeted delivery of large functional RNAs in cancer cell lines. *Nat. Commun.*, **9**, 2283.

58. Katoh, K., Misawa, K., Kuma, K.i. and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, **30**, 3059-3066.

59. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189-1191.

60. Weinberg, Z. and Breaker, R.R. (2011) R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinform.*, **12**, 3.

CHAPTER 4: 2′-FLUORO-MODIFIED PYRIMIDINES ENHANCE AFFINITY OF RNA OLIGONUCLEOTIDES TO HIV-1 REVERSE TRANSCRIPTASE

As found in: Paige R. Gruenke, Khalid K. Alam, Kamal Singh, and Donald H. Burke. 2′-Fluoro-Modified Pyrimidines Enhance Affinity of RNA Oligonucleotides to HIV-1 Reverse Transcriptase. *RNA*. 2020; 26(11):1667-1679.

4.1. Abstract

Nucleic acid aptamers can be chemically modified to enhance function, but modifying previously selected aptamers can have non-trivial structural and functional consequences. We present a reselection strategy to evaluate the impact of several modifications on pre-existing aptamer pools. RNA aptamer libraries with affinity to HIV-1 reverse transcriptase (RT) were re-transcribed with 2′-F, 2′-OMe, or 2′-NH_2 pyrimidines and subjected to three additional selection cycles. RT inhibition was observed for representative aptamers from several structural families identified by high-throughput sequencing when transcribed with their corresponding modifications. Thus, reselection identified specialized subsets of aptamers that tolerated chemical modifications from unmodified pre-enriched libraries. Inhibition was the strongest with the 2′-F-pyrimidine (2′-FY) RNAs, as compared to inhibition by the 2′-OMeY and 2′-NH_2Y RNAs. Unexpectedly, a diverse panel of retroviral RTs was strongly inhibited by all 2′-FY-modified transcripts, including sequences that do not inhibit those RTs as unmodified RNA. The magnitude of promiscuous RT inhibition was proportional to mole fraction 2′-FY in the transcript. RT binding by 2′-FY transcripts was more sensitive to salt concentration than binding by unmodified transcripts, indicating that interaction with retroviral RTs is more ionic in character for 2′-FY RNA than for unmodified 2′-OH

RNA. These surprising features of 2′-FY-modified RNA may have general implications for applied aptamer technologies.

4.2. Introduction

Aptamers are structured nucleic acids that bind to molecular targets, often with high affinity and specificity. They are generated through an iterative selection process termed systematic evolution of ligands by exponential enrichment (SELEX) (1,2). In some cases, chemically modified nucleotide building blocks within the aptamers can improve the binding affinity of an aptamer to its target of interest. For example, very high affinity aptamers against the human immunodeficiency virus type 1 reverse transcriptase (HIV-1 RT) and HIV-1 integrase were recently selected using the 2′-deoxy-2′-fluoroarabinonucleotide (FANA) modification (3,4) ,various classes of modifications can improve the biostability of aptamers and enable reaction chemistries that are inaccessible to unmodified nucleic acids (5-7). To stabilize RNA aptamers against rapid degradation by serum endonucleases for therapeutic and diagnostic applications, one strategy is to modify the 2′ position on the ribose ring of pyrimidines (8-10). Common 2′-modifications such as 2′-fluoro (-F), 2′-amino (-NH_2), and 2′-O-methyl (-OMe) significantly increase RNA serum half-life, but they also can have structural and functional consequences for the aptamer by perturbing the binding interface, binding affinity, three-dimensional structure, and overall pharmacokinetic and pharmacodynamic properties (11). Among the above-mentioned modifications, the 2′-F-pyrimidine (2′-FY) modification has been extensively used in selecting aptamers for use *in vivo* or in serum, such as cell-targeted or cell-internalizing aptamers (12,13). Modifications at the 2′-position of the pentose sugar in nucleic acids can affect sugar puckering and the types of hydrogen bond patterns (14)

and alters electronic distribution throughout the nucleotide (15). The high electronegativity of the fluorine makes it a weak hydrogen bond acceptor relative to 2′-OH, and enhances the C3′-endo character of the sugar pucker relative to unmodified ribose to reinforce the canonical A-form helix (15-17). In contrast, the 2′-NH_2 substitution more strongly prefers the C2′-endo sugar pucker found in DNA-like B-form helices (15). 2′-OMe and unmodified 2′-OH nucleotides are intermediate between the two extremes (2′-F and 2′-NH_2), favoring the C3′-endo sugar pucker to approximately the same degree (15). Many other nucleotide analogs are available with modifications on the sugar, nucleobase, or internucleotide linkages, each associated with their own potential for modulating nucleic acid function.

Two approaches are conventionally used to generate chemically-modified aptamers. In the first approach *de novo* selection is performed with chemically modified random libraries from the beginning. An advantage of this in-SELEX strategy is that it leverages the full sequence diversity of the random library to investigate a large number of modified sequences simultaneously (18). Many types of modified nucleotides can be incorporated into selections to enhance aptamer functionality and stability, such as modifications within the nucleobases, phosphate backbone, or other sugar modifications (10). However, the modified nucleotides must be recognized and/or incorporated by relevant polymerases (19), and many modified nucleotides are expensive or require specialized synthetic expertise to produce them in-house (20). Polymerase engineering has yielded many new polymerases that incorporate diverse modified nucleotides efficiently and greatly expanded the alphabet of accessible nucleotide analogs for *in vitro* selection (21-23). For example, the Y639F mutation in T7 RNA polymerase allows

efficient incorporation of 2′-F-, 2′-NH_2-, and 2′-H nucleotides into RNA during *in vitro* transcription (24,25), and the Y639F/H784A double mutant T7 RNA polymerase incorporates bulkier 2′-modified nucleotides such as 2′-OMe and 2′-azido nucleotide triphosphates (NTPs) (26). Another example is the DNA polymerase I from *Geobacillus stearthermophilus* (Bst), as it is able to reverse transcribe RNA and xeno nucleic acids (XNAs) into DNA (27), including FANA and threose nucleic acid (TNA), and has been used in *in vitro* selections (28-30). In the second approach chemical modifications are incorporated through post-SELEX chemical optimization of aptamers that were originally selected as native DNA, RNA, or other polymers by systematically introducing modified nucleotides into pre-existing aptamers in various combinations and determining their impact on binding affinity, serum stability, and function through systematic screens (18,20,31,32). This process is repeated until the modified aptamer has the desired properties, which typically requires screening a large number of variants (6), which can be laborious, expensive, and hit-or-miss.

We developed a reselection strategy that identifies aptamers that can either accommodate or adapt to nucleotide chemical modifications and that reduces many of the obstacles commonly faced with the *de novo* modified SELEX and post-SELEX optimization approaches. We began with a pre-enriched library of RNA aptamers (2′-OH) that had previously undergone 14 rounds of affinity selection against HIV-1 RT (33). This library was re-transcribed with either 2′-F, 2′-OMe, or 2′-NH_2 pyrimidines, and each new set was subjected to three additional rounds of selection followed by high-throughput sequencing and bioinformatics analyses. From the resulting enrichment data, we chose candidate aptamers to further characterize their ability to bind and inhibit HIV-

1 RT enzymatic activity in primer extension assays. This approach identified 2′-modified RNA aptamers that maintained their originally selected function of binding to HIV-1 RT despite the 2′ modification. Unexpectedly, all transcripts carrying 2′-FY modifications – including non-binding control sequences – were found to bind and inhibit several retroviral RTs, but either did not inhibit or only very weakly inhibited bacterial and human DNA polymerases. The enhanced inhibition by the 2′-FY-modified transcripts is proportional to the mole fraction of 2′-FY and is due, in part, to increased ionic interactions of the 2′-FY transcripts with HIV-1 RT in comparison with 2′-OH transcripts. These observed properties of 2′-FY-modified RNA may provide new insights on the binding interactions of 2′-FY aptamers to their respective targets.

4.3. Results

4.3.1. Reselection Strategy to Identify Aptamers that Can Tolerate Chemical Modifications

To rapidly identify aptamers that can productively accommodate or adapt to alternative nucleotide alphabets, a pre-enriched population of RNA aptamers that emerged from 14 rounds of selection for binding to HIV-1 RT (33) was subjected to three additional rounds of selection for binding to HIV-1 RT in which the populations carried either the original (2′-OH) or 2′ modified pyrimidines (2′-FY, 2′-OMeY, or 2′-NH_2Y) (Figure 4.1A & B). Each reselection trajectory was done in duplicate. Output libraries for each round were screened for RT inhibition, defined as a reduction in the ‘normalized fraction extended’ value from 1.0 in the absence of aptamer to less than 0.6 (Figure 4.1C) (34). The results shown in Figure 4.1C demonstrate that the 2′-FY and 2′-OMeY pre-enriched libraries strongly inhibited RT. In contrast, the pre-enriched library was non-

inhibitory upon substitution to 2′-NH_2Y but regained some RT inhibition after several rounds of reselection with 2′-NH_2Y.

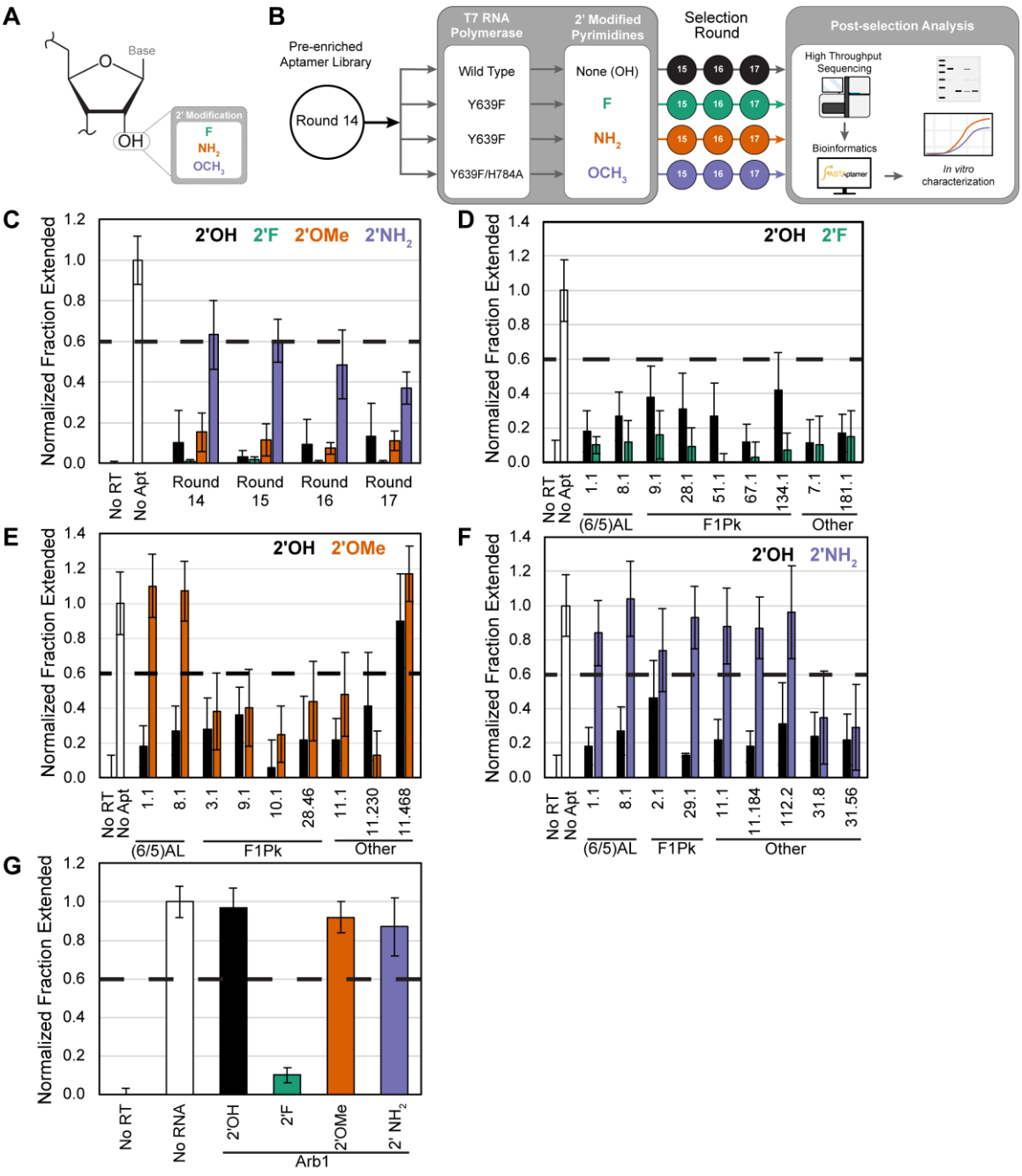

Figure 4.1: Effect of 2′-pyrimdine modifications on RT inhibition by candidate aptamers. (A) Structure of ribose in RNA highlighting the 2′ position and modifications tested in this study. (B) Schematic of the reselection strategy. A pre-enriched aptamer library that had previously undergone 14 rounds of selection for affinity to HIV-1 RT was transcribed with either 2′-OH, 2′-F, 2′-OMe or 2′-NH_2 pyrimidines and reselected for three additional rounds, with each trajectory performed in duplicate. (C) Quantification of primer extension assays showing fraction of primer converted into full-length product in control reactions in the absence of RT (No RT) or aptamer (No Apt) and in reactions containing selected aptamer libraries from the 2′-OH (black), 2′-F (green), 2′-OMe (orange) or 2′-NH_2

(purple) trajectory after each reselection round. (D-F) Quantification of primer extension assays showing fraction of primer converted into full-length product in control reactions in the absence of RT (No RT) or aptamer (No Apt) or in reactions containing aptamers either transcribed with 2′-OH or (D) 2′-F, (E) 2′-OMe, or (F) 2′-NH_2 pyrimidines. Aptamers are grouped together by structural class: (6/5) asymmetric loop family [(6/5)AL] or family 1 pseudoknot (F1Pk). Aptamers that did not contain the consensus sequence features of any of the characterized structural motifs were grouped as Other. (G) Evaluation of primer extension assays in the presence of 2′-modified Arb1. Plotted values and vertical error bars represent the means and standard deviations of fraction primer extended to full-length product normalized to the no aptamer (or no RNA) control (set to 1) and to the no RT control (set to 0) of four independent replicates (n = 4). RNAs are considered inhibitory if the normalized fraction extended value is below 0.6, which is marked by a dashed horizontal line.

High-throughput sequencing data for aptamer populations from rounds 14 and 17 were analyzed using FASTAptamer, a bioinformatics toolkit that counts the number of times an individual sequence is present in each population, clusters highly similar sequences within a population, and calculates enrichment of sequences between populations (35). The total reads and number of unique sequences for each aptamer population is given in Table 4.1. Aptamers that co-enriched at least two-fold in each of the duplicate trajectories for a given modification were prioritized for further characterization (Tables 4.2-4.3). Most of these belonged to the characterized family 1 pseudoknot (F1Pk) motif (33,36,37) or (6/5) asymmetric loop ((6/5)AL) motif (37) aptamer families (Figure 4.2A). (6/5)AL aptamers are broad-spectrum inhibitors of reverse transcriptases across multiple strains of HIV-1, while F1Pk aptamers are potent inhibitors of HIV-1 reverse transcriptases from Group M:Subtype B, which is the target against which they were originally selected. A single point mutation R277K confers resistance to inhibition by F1Pk aptamers. For comparison, additional sequences were chosen that co-depleted at least two-fold in each of the duplicate trajectories. Aptamers from the different structural families demonstrated different sensitivities to the three modifications tested, and similar results were observed with a subset of previously identified HIV-1 RT aptamers (Figure 4.2). 2′-FY versions of aptamers from the 2′-FY trajectories retained ability to inhibit RT to at least the same degree as the unmodified (2′-OH) versions of these same aptamers (Figure 4.1D). 2′-OMeY versions of aptamers from the 2′-OMe trajectory exhibited different effects on RT inhibition depending on the structural class of the aptamer (Figure 4.1E). The two (6/5)AL aptamers no longer inhibited RT as 2′-OMeY transcripts, while the F1Pk aptamers were only moderately

affected (see Section 4.3.1.2 and Figure 4.3 for more analysis on the effects of the 2′-OMeY modification on these structural classes). 2′-NH_2Y versions of aptamers from the 2′-NH_2 trajectory were generally unable to inhibit HIV-1 RT (Figure 4.1F), although each remained inhibitory as 2′-OH transcripts. Only two aptamers, both from cluster 31, inhibited RT to the same degree as their unmodified RNA versions when generated as 2′-NH_2Y transcripts (Figure 4.1F; also see Sections 4.3.1.2 and 4.3.1.3 and Figure 4.4 for initial structural evaluation).

Table 4.1: Read counts (total, unique, reads per sequence) and number of unique sequence clusters from high throughput sequences of aptamer populations

Library ID	Total Reads	Unique Sequences	Average Reads per Sequence[a]	Sequence Clusters[b]
70N Round 14	2,160,216	72,921	30	1,250
70N 2′-OH (17A)	2,028,611	44,811	45	604
70N 2′-OH (17B)	2,760,126	42,998	64	
70N 2′-FY (17A)	3,542,268	129,540	27	2,360
70N 2′-FY (17B)	2,774,057	115,242	24	
70N 2′-OMeY (17A)	2,564,164	73,009	35	188
70N 2′-OMeY (17B)	1,862,002	62,405	30	
70N 2′-NH_2Y (17A)	977,733	25,481	38	818
70N 2′-NH_2Y (17B)	2,913,710	73,550	40	

[a]Average reads per sequences was rounded to the nearest whole number.

[b]The number of unique sequence clusters was combined for the two replicates of the same trajectory.

Table 4.2: RNA sequences used in this study[a]

RNA	Sequence
70N 1.1	ACGUUGUCGAAAGCCUAUGCAAAUUAAGGACUGUCGACGAAACCUUGCGUAGACUCGCCACGCUUGGUGU
70N 2.1	CAUAGCGACUGUCCACGAAUCCGAAGCCUAACGGGACAAAAGGCAAGAGCGCGAUACCAAUGCUGGACUG
70N 3.1	AACCGCAAGCAACACCCAGCAAGAAACAUCCGACGCACGACGGGAGAAAGUGCAUUACCACGAUGUCGAU
70N 7.1	GCGAACCAAACCCAGAUUACUAACCGUGGGCCUGAAACACGGGACAAAACAGGCAUCAAUGGAGUGGUAC
70N 8.1	ACGUUGUGCACGGAUGCCCACGGUCGCACGAAACCUUGUGUGGGAUAGCGCGAAUACUGACGAGUGUGCC
70N 9.1	ACCAAAUCCCGAACUACAAAUCCGAACGCUAACGGGACAAUUGCGAAAUGGAACAUACGGGCCUGGUUGA
70N 10.1	GUGCGCUACCACAUGAUCCGAGGCAAAACGGGAAAAGAUAGCAUCGAUUACGGAACCGGCCACGCACA
70N 11.1	ACCAAGAUAAACCGAGGUGUAAACGGGAAAACACAUGGAAUGACGUGCUGUAGUAGGUGUGAUCACUGC
70N 11.184	ACCAAGAUAAACCGAGGUGUAAAACGGGAAAACACAUGGAAUGACGUGCUGUAGUAGGUGUGAUCACUGC
70N 11.230	ACCAAGAUAAACCGAGGUGUAAACGGGAUAACACAUGGAAUGACGUGCUGUAGUAGGUGUGAUCACUGC
70N 11.468	ACCAAGAUAAACCGAGGUGUAAACGGGAAAUCACAUGGAAUGACGUGCUGUAGUAGGUGUGAUCACUGC
70N 28.1	GGCGCACACUACCAAUUGAAUCCGAAAGACACGGGACAAGUCUAAAUGCGAUGAAAUACCGAAUCGAGA
70N 28.46	GGCACACUACCAAUUGAAUCCGAAAGACACGGGACAAGUCUAAAUGCGAUGAAAUACCGAAUCGAGA
70N 29.1	GCGCGAUUGCCAUAGAAUCCGAGGGUCAACGGGAAAAGGACCAACCAACGCGUCGAGCAGCCACAUGCAG
70N 31.8	UAACCGCACAUCCGAAAACGAUCCCACGGCAAGGGGAUAAUGCCAGCGAUUAAUGGAGCCUACGUCAACA
70N 31.56	UAACCGCACAUCCGAGAACGAUCCCAUGGCAAGGGGAUAAUGCCAGCGAUUAAUGGAGCCUACGUCAACA
70N 51.1	ACAUUGGCAAACCACAGAAUCCGAGGCGACACGGGAUAACGCAACCUUAAGGUCGACAUGUUUCGCACAA
70N 67.1	AAGACUAACACAUGACCCAAUACACGAAUCCGAAGGCAAACGGGACAACUGCCGAUACCUGUGUACCGAGCUAC

70N 112.2	CUAAUCCUGGAGAAGGAACAGGGACAAACACCUUCAACAUCGAACCAAAUCCCGAACUGCACCGCGCCAU
70N 134.1	ACAAUCCGAGGCAUAACGGGAGAAGAUGCAAAAAUUGAGUGGAAAACACAAAAGUGCCGUCGUAACGUAC
70N 181.1	ACUGAAGGGGAAAGAGUGTGUAUCCCUUCCACACAACAUGGCGCGCAAUAUGAUUGGUGAUGCUUGGC
Arb1	AAUUUGGACUUUCCGCCCUUCUUGGCCUUUAUGAGGAUCUCUCUGAUUUUUCUUGCGUCGAGUUUUCCGG

[a]The full RNA sequence consists of the 5′ constant region (gggaaaagcgaaucauacacaaga), the ~70 nucleotide sequence displayed above, and the 3′ constant region (gggcauaagguauuuaauuccaua).

Table 4.3: Enrichment Values for Candidate Aptamers[a]

RNA	2'OH		2'F		2'OMe		2'NH_2	
	A	B	A	B	A	B	A	B
70N 1.1	0.042	0.0036	0.053	0.088	0.0027	0.00018	0.000026	0.027
70N 2.1	1.11	1.06	-	-	-	-	1.28	0.78
70N 3.1	0.63	0.27	-	-	0.99	0.81	-	-
70N 7.1	0.0017	0.00024	0.011	0.050	-	-	-	-
70N 8.1	0.032	0.0018	0.14	0.48	0.00041	0.00032	0.00015	0.064
70N 9.1	0.70	0.51	48.13	47.09	8.64	5.53	-	-
70N 10.1	0.72	0.10	-	-	5.23	2.32	-	-
70N 11.1	390.18	486.84	-	-	364.20	384.43	47.2	374.01
70N 11.184	246.48	422.15	-	-	-	-	733.74	3039.61
70N 11.230	1996.43	1961.97	-	-	2534.93	3151.13	-	-
70N 11.468	38.59	41.74	-	-	4268.70	5927.48	-	-
70N 28.1	16.47	25.2	673.67	447.92	-	-	-	-
70N 28.46	NP	NP	-	-	22.89	656.15	-	-
70N 29.1	1.74	0.34	-	-	-	-	19.15	4.24
70N 31.8	44.33	6.26	-	-	-	-	408.37	35.31
70N 31.56	52.50	NP	-	-	-	-	647.02	56.70
70N 51.1	9.02	0.75	2671.60	673.52	-	-	-	-

70N 67.1	0.37	0.11	1084.92	67.45	-	-	-	-
70N 112.2	NP	NP	-	-	-	-	NP	218.76
70N 134.1	244.80	256.63	1288.5	401.54	-	-	-	-
70N 181.1	12.26	NP	225.08	72.68	-	-	-	-

[a]Enrichment values are listed for the modification trajectories that the RNA was modified with and tested for RT inhibition. NP = No enrichment value was present for the RNA in a particular trajectory.

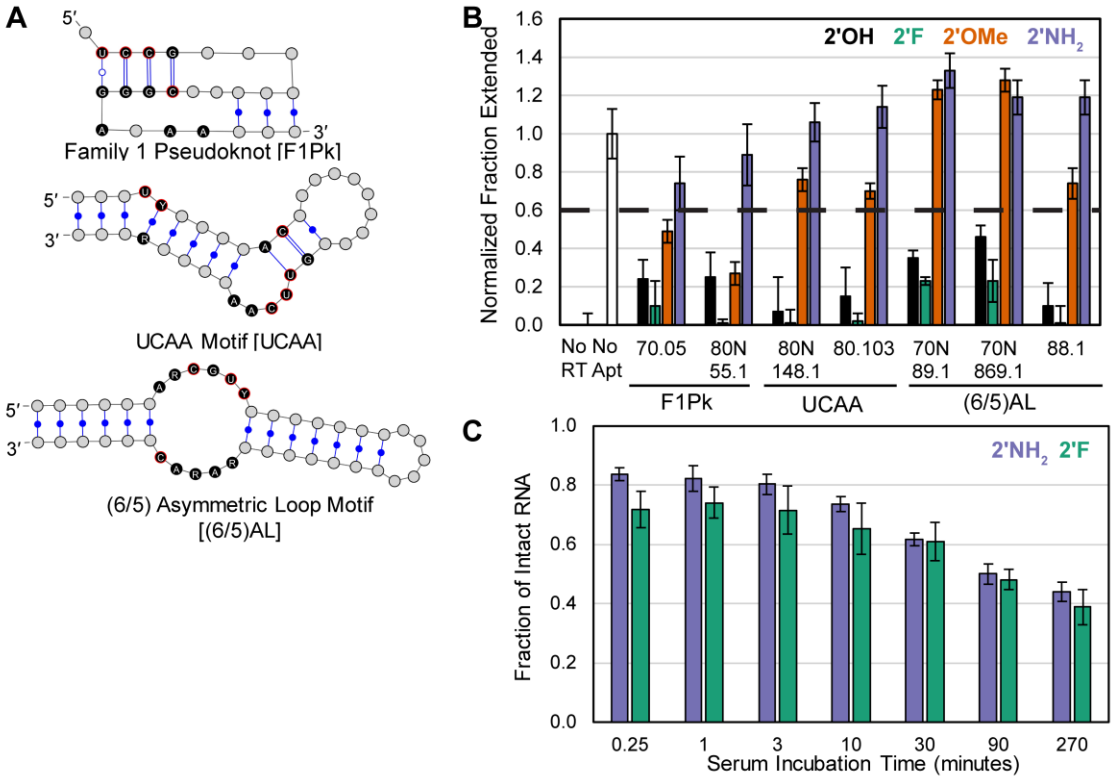

Figure 4.2: Effect of 2′ modifications on inhibition of HIV-1 RT by anti-HIV-1 RT RNA aptamers. (A) Representations of the characteristic secondary structural features of these three classes of RT aptamers: (6/5) asymmetric loop motif [(6/5)AL], UCAA motif (UCAA) and family 1 pseudoknot (F1Pk). Highly conserved nucleotides are shown in black. Positions marked with R or Y are purines or pyrimidines, respectfully. (B) Quantification of primer extension assays showing fraction of primer converted into full-length in control reactions in the absence of RT (No RT) or aptamer (No Apt) and in reactions containing various full-length RNA aptamers belonging to the F1Pk, UCAA and (6/5)AL structural motifs. Aptamers were transcribed using either 2′-OH (black), 2′-F (green), 2′-OMe (orange) or 2′-NH_2 (purple) pyrimidines (n = 4). (C) Quantification of serum nuclease assays performed against an RNA aptamer (70N 1.1) transcribed with 2′-F or 2′-NH_2 pyrimidines and incubated in DMEM media + 10% fetal bovine serum for the indicated period of time (n = 3).

4.3.1.1. Impact of 2′ Modifications on Previously Identified Anti-HIV-1 RT Aptamers

Previously identified RNA aptamers that bind and inhibit HIV-1 RT were transcribed using 2′-OH, 2′-F, 2′-NH_2, or 2′-OMe pyrimidines and were evaluated for RT inhibition. The aptamers tested belonged to the family 1 pseudoknot (F1Pk) motif (33,36,37), UCAA motif (38), or (6/5) asymmetric loop(AL) motif (37) aptamer families (Figure 4.2A). Each of the three 2′ modifications had a different impact on the abilities of these aptamers to inhibit RT (Figure 4.2B), similar to the trends observed in Figure 4.1. When modified with 2′-F pyrimidines, aptamers from all three structural classes inhibited RT at least as well as the unmodified versions, and many seemed to inhibit to a greater extent compared to the unmodified aptamers. When 2′-OMe pyrimidines were incorporated, two F1Pk aptamers remained moderately inhibitory, but aptamers from the UCAA and (6/5)AL family aptamers were no longer able to inhibit RT. When these aptamers were modified with 2′-NH_2 pyrimidines, RT inhibition was no longer observed regardless of structural motif. These results were in agreement with the trends observed in Figure 4.1. As expected, the 2′ modifications increase the half-life of the aptamer 70N 1.1 in the presence of 10% fetal bovine serum from <<15sec (not shown) to 2-3 hours (Figure 4.2C).

4.3.1.2. Additional Analyses of Results from 2′-OMeY and 2′-NH_2Y Trajectories

As shown in Figure 4.1E, F1Pk aptamers were still able to inhibit RT when modified with 2′-OMe pyrimidines, while (6/5)AL aptamers lost ability to inhibit RT. The moderate reduction in inhibition by 2′-OMeY-modified F1Pk aptamers is consistent with a previous study that found that 2′-OMe substitution of C13, which is the first

nucleotide of the 5′-CGGG-3′ element within stem 1, reduced affinity to HIV-1 RT by approximately 5-fold. For that same aptamer, an approximate 10-fold reduction in affinity was observed when all positions were modified with 2′-OMe pyrimidines (39). Therefore, it is likely that the incorporation of 2′-OMe pyrimidines into F1Pk aptamers studied here decreased, but did not abolish, their affinity to HIV-1 RT. However, any decrease in affinity did not greatly affect HIV RT inhibition by these aptamers under the conditions of this assay. For (6/5)AL aptamers, at least one cytosine and at least one uracil are required to be 2′-OH, as singly-substituted 2′-OMeY transcripts were non-inhibitory (Figure 4.3).

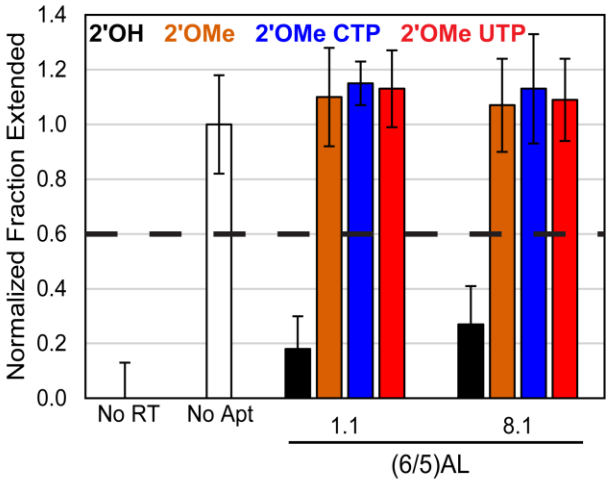

Figure 4.3: Effect of Single 2′-OMe-Pyrimidine Modification on RT inhibition by (6/5)AL Aptamers. Two (6/5)AL aptamers (70N 1.1 and 70N 8.1) were transcribed in reactions in which only 2′-OMe-modified CTP (blue) or only 2′-OMe-modified UTP (red) replaced the normal NTP, and these RNAs were evaluated for RT inhibition in primer extension assays (n = 4). Primer extension assay results using aptamers transcribed with both pyrimidines supplied as 2′-OH (black) or as 2′-OMe (orange) were plotted for comparison. Loss of RT inhibition with either substitution indicates that at least one C and at least one U must retain 2′-OH in order for these aptamers to inhibit RT.

Four aptamer variants from cluster 11 (11.1, 11.184, 11.230, and 11.468) were highly enriched in the 2′-OMeY and 2′-NH_2Y trajectories. Their sequences do not contain recognizable motifs for F1Pk, UCAA, or (6/5)AL families, and their functional structure is not known. Each of these aptamers was transcribed with various NTP combinations and tested for RT inhibition. All four aptamers moderately inhibited HIV-1 RT as unmodified transcripts (Figure 4.1E and 4.1F). When modified with 2′-OMe pyrimidines, RT inhibition by aptamer 11.1 was slightly weaker (not significant) while inhibition by aptamer 11.230 was about the same as for the 2′-OH version. In contrast, aptamer 11.468 lost all ability to inhibit HIV-1 RT after 2′-OMe modification. Neither aptamer 11.1 nor 11.184 inhibited RT when modified with 2′-NH_2 pyrimidines. While sequences from cluster 11 were highly enriched, it seems that these sequences were artefacts from the selection. Therefore, studies further exploring the structural class of cluster 11 aptamers were not pursued.

4.3.1.3. Secondary Structure Insights for Cluster 31 Aptamers from the 2′-NH_2Y Trajectory

As shown in Figure 4.1F, only the two aptamers from cluster 31 were able to inhibit RT when modified with 2′-NH_2 pyrimidines. The cluster 31 aptamers do not belong to any of the previously defined structural families of HIV-1 RT RNA aptamers, and the required functional structure is not known. A series of variants of aptamer 31.56 were evaluated for RT inhibition to gain insight into the functional requirements of this novel aptamer, using 2′-OH transcripts to simplify the analysis. Removing ten nucleotides at a time from either the 5′ or 3′ end abolished RT inhibition (Figure 4.4A), indicating that nucleotides near each end are required for the aptamer to remain fully

functional. Among the 1596 unique sequences present in this cluster, 1334 of these contained two or more nucleotide differences relative to the cluster seed sequence, aptamer 70N 31.1 (Figure 4.4B), and were therefore capable of covariation. The 300 most abundant of these sequences were used to generate two covariation models using the RNAalifold web server (40) by applying two different parameter settings (Figure 4.4C). Model 1 predicted that portions of the 5′ and 3′ constant regions base paired with each other to form a 4 base pair stem. Model 2, in contrast, predicted that three nucleotides from the 5′ constant region base paired with three nucleotides from the random region and five nucleotides from the 3′ constant region base paired with five nucleotides from the random region. Both models are consistent with initial 5′ and 3′ truncation data, but invariance within the constant regions precludes using covariation data to discriminate between these models. Therefore, sequence variants were generated that preferentially disrupted, rescued, or stabilized one or the other of the two structures, and these variants were tested for RT inhibition. Specifically, each predicted stem was mutated in one strand to disrupt its formation, and the original predicted pairing pattern was restored by introducing compensatory mutations in the other strand. For Model 1, the stem disruption remained fully inhibitory, while introducing compensatory mutations to restore pairing, or extending the stem to reinforce base pairing between the 5′ and 3′ constant regions, eliminated RT inhibition (Figure 4.4D). These data rule out Model 1. Sequence variants that disrupted or rescued the predicted stems in Model 2 all inhibited RT to the same degree as the full-length unmodified or 2′-NH_2 versions of aptamer 31.56, while an internal deletion of 19 nucleotides abolished RT inhibition. Model 2 predicted structures at the 5′ and 3′ ends, while the data do not provide evidence that these

structures are functionally required. Further, non-inhibition by the internal deletion could mean that a portion of the deleted sequence may be involved in making the active structure or induces a misfold of the aptamer. Finally, when aptamer 70N 31.56 was transcribed with 2′-FY in place of 2′-NH_2Y, the 2′-FY version strongly inhibited RT (Figure 4.4E). These data were in agreement with the other primer extension assays using 2′-FY-modified aptamers in Figure 4.1 and suggested that the 2′-FY modification was either well tolerated by the RT aptamers tested or enhanced non-specific binding of the RNA to RT.

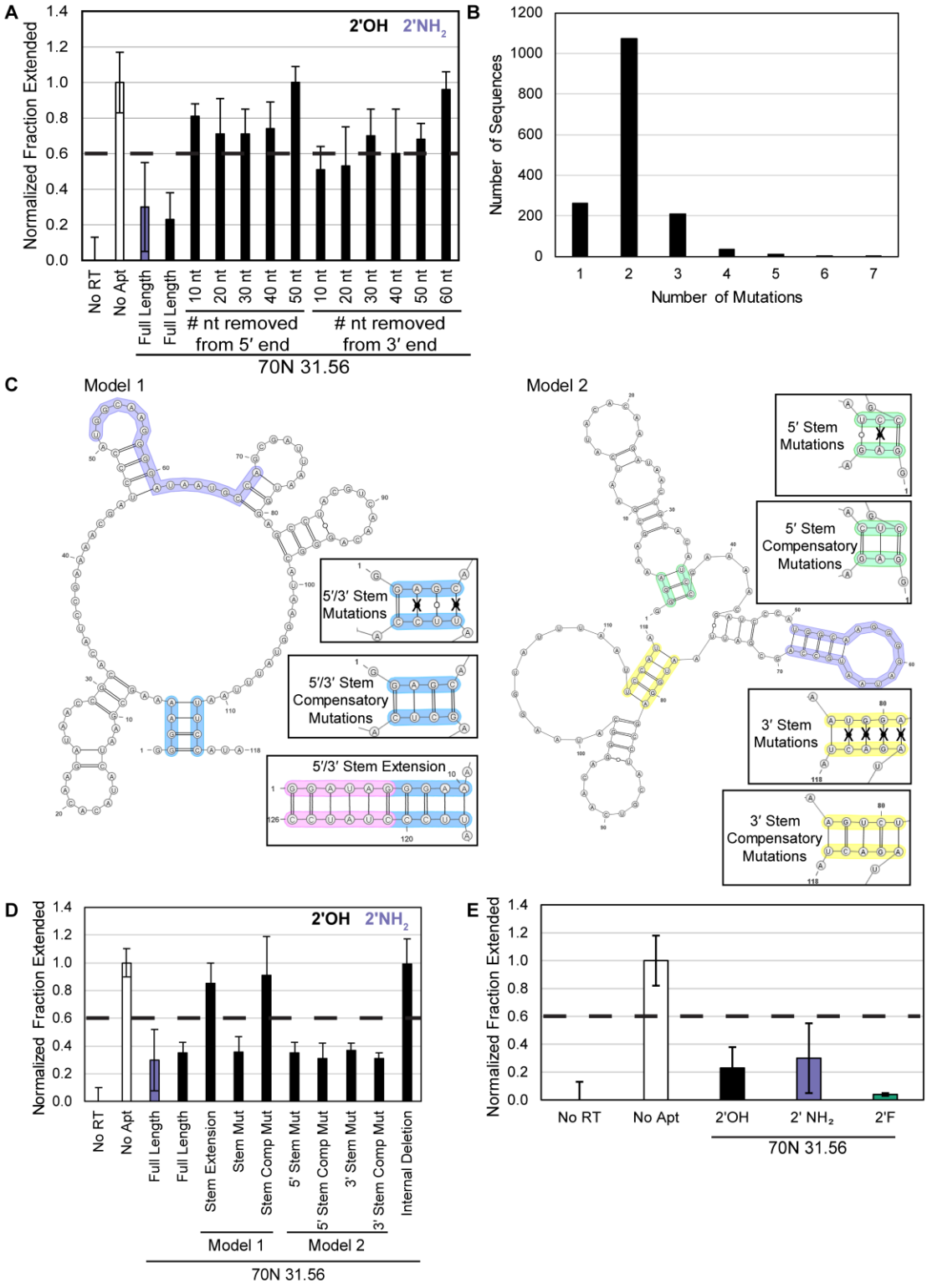

Figure 4.4: Investigating the Secondary Structure of Cluster 31 Aptamers. (A) Aptamer 70N 31.56 was systematically truncated by removing a block of 10 nucleotides from either the 5′ or 3′ end, and these truncations were used in primer extension assays. Full-length 70N 31.56 transcribed with either 2′-OH (black) or 2′-NH_2 (purple) pyrimidines

was included for comparison. 2′-OH pyrimidines were used in these assays because the unmodified and 2′-NH_2Y versions of this RNA showed similar inhibition profiles. (B) Distribution of cluster 31 sequences showing number of sequences with the indicated number of mutations relative the seed sequence. The 300 most abundant sequences with at least two mutations were used to the generate the covariation models in (C). (C) A covariation model for the secondary structure of cluster 31 aptamers ("Model 1") determined using RNAalifold with parameters of weight of covariance of 0.6 and penalty for noncompatible sequences of 6. The 5′/3′ stem is highlighted in blue. To test this model, three variants were transcribed that mutate or extend the 5′/3′ stem and used in primer extension assays shown in (D). The changed regions of the variants are shown in the boxes. A second covariation model for the secondary structure of cluster 31 aptamers ("Model 2") determined using RNAalifold with parameters of weight of covariance of 1 and penalty for noncompatible sequences of 1. The 5′ and 3′ stems are highlighted in green and yellow, respectfully. To test this model, variants were transcribed that mutate the 5′ or 3′ stem and used in primer extension assays as shown in (D). The changed regions of the variants are shown in the boxes. The 19-nucleotide region highlighted in purple in both models was removed for the internal deletion variant tested in (D). The sequences used to generate the covariation models were first aligned using MAFFT(41), and the output from MAFFT was used as the input for RNAalifold. (D) The variants of the two cluster 31 covariation models were used in primer extension assays to test the base-pairing hypotheses of the two models. Full-length 70N 31.56 transcribed with either 2′-OH (black) or 2′-NH_2 (purple) pyrimidines was included for comparison. (E) To test whether cluster 31 aptamers from the 2′-NH_2-pyrimidine selection inhibit HIV-1 RT when transcribed with 2′-F pyrimidines, 70N 31.56 was transcribed using 2′-F pyrimidines (green) and used in primer extension assays (n = 4). Full-length 70N 31.56 transcribed with either 2′-OH (black) or 2′-NH_2 (purple) pyrimidines was included for comparison.

Over the course of the selections, relative enrichment of the F1Pk and (6/5)AL structural families as a whole followed the trends predicted from screens of individual exemplars of these families. The F1Pk structural motif did not enrich or deplete more than two-fold in any of the trajectories (Table 4.4). In contrast, the (6/5)AL motif was much more strongly depleted in the 2′-OMeY and 2′-NH_2Y trajectories (Table 4.5), consistent with its sensitivity to these modifications. These results demonstrate the feasibility of the reselection strategy for identifying the subset of aptamers in a pre-enriched library that can function with modified nucleotides, and they highlight the variable degree to which 2′-modified pyrimidines impact various structural families of RNA aptamers with affinity for HIV-1 RT.

Table 4.4: Structural Motif Enrichment Data for Family 1 Pseudoknot Aptamers

Library ID	Number Sequences Matched	Total Motif Reads	Motif RPM (reads per million)	Enrichment Value[a]
70N Round 14	40,066	1,166,477	539,981.65	
70N 2′-OH (17A)	19,157	712,192	351,073.71	0.65
70N 2′-OH (17B)	16,875	759,350	275,114.25	0.51
70N 2′-FY (17A)	103,009	3,235,054	913,271.95	1.69
70N 2′-FY (17B)	86,412	2,388,807	861,123.98	1.60
70N 2′-OMeY (17A)	30,908	969,288	378,013.26	0.70
70N 2′-OMeY (17B)	25,039	600,101	322,288.05	0.60
70N 2′-NH_2Y (17A)	17,139	753,248	770,402.55	1.43
70N 2′-NH_2Y (17B)	39,599	1,200,166	411,903.04	0.76

[a]Enrichment value is calculated by taking the motif RPM at round 17 divided by the motif RPM at round 14.

Table 4.5: Structural Motif Enrichment Data for (6/5) Asymmetric Loop Aptamers

Library ID	Number Sequences Matched	Total Motif Reads	Motif RPM (reads per million)	Enrichment Value
70N Round 14	16,086	818,995	379,126.44	
70N 2′-OH (17A)	2,516	41,014	20,217.77	0.053
70N 2′-OH (17B)	834	6,019	2,180.70	0.0058
70N 2′-FY (17A)	9,470	150,785	42,563.76	0.11
70N 2′-FY (17B)	9,952	186,152	67,104.61	0.18
70N 2′-OMeY (17A)	413	3,105	1,210.92	0.0032
70N 2′-OMeY (17B)	71	192	103.11	0.00027
70N 2′-NH_2Y (17A)	57	280	286.38	0.00076
70N 2′-NH_2Y (17B)	1,970	34,168	11,726.63	0.031

4.3.2. 2′-FY Transcripts as General Inhibitors of Retroviral RTs

In many of the RT enzymatic inhibition assays above, 2′-FY aptamer transcripts were more potent than 2′-OH or other forms in inhibiting HIV-1 RT. We therefore evaluated the impact of 2′-FY modifications on RT inhibition for Arb1, an arbitrary control RNA that contains 70 nucleotides from the luciferase gene mRNA, flanked by the same constant regions as the aptamers. This sequence has previously been shown not to bind RT *in vitro* or to inhibit virus replication in cells (42,43). As expected, a 2′-OH version of Arb1 did not inhibit HIV-1 RT. 2′-OMeY and 2′-NH_2Y versions were similarly non-inhibitory. In contrast, the 2′-FY version of Arb1 strongly inhibited HIV-1 RT (Figure 4.1G). In other words, substituting the 2′-FY in place of 2′-OH converted a non-inhibitory RNA transcript into an inhibitor for HIV-1 RT.

To establish the generalizability of this unexpected result, unmodified and 2′-FY modified transcripts were generated for a broad-spectrum (6/5)AL aptamer (70N 1.1) and for an F1Pk aptamer (70N 51.1), as well as for control RNAs Arb1 described above and Arb2 (an RNA aptamer selected to bind an unrelated protein). These eight transcripts (4 sequences, each ± 2′-FY modifications) were screened for their abilities to inhibit a panel of eleven DNA polymerases that included RTs from phylogenetically diverse HIV-1, RTs from other retroviruses, and non-viral DNA polymerases. In unmodified form, both anti-RT aptamers inhibited the HIV-1 RT from strain HIV-1 HXB2 (Group M:Subtype B), but only the (6/5)AL aptamer inhibited the R277K point mutant of HIV-1 HXB2 RT and the RT from HIV-1 strain 94CY (Group M:Subtype A), consistent with previously observed specificities for these structural families (37,44,45). Neither Arb1 nor Arb2 inhibited these three RTs as 2′-OH transcripts. In contrast, the 2′-FY versions of all four

RNAs strongly inhibited these Group M RTs (Figure 4.5A). The RT from the chimpanzee strain of simian immunodeficiency virus (SIV_{cpz}), TAN1B, was not inhibited by either anti-RT aptamer as 2′-OH transcripts, consistent with previous observations (44,45), while the 2′-FY-modified versions of these aptamers were moderately better inhibitors of TAN1B RT. Arb1 and Arb2 were both non-inhibitory as 2′-OH transcripts but strongly inhibited TAN1B RT as 2′-FY transcripts (Figure 4.5B, left panel).

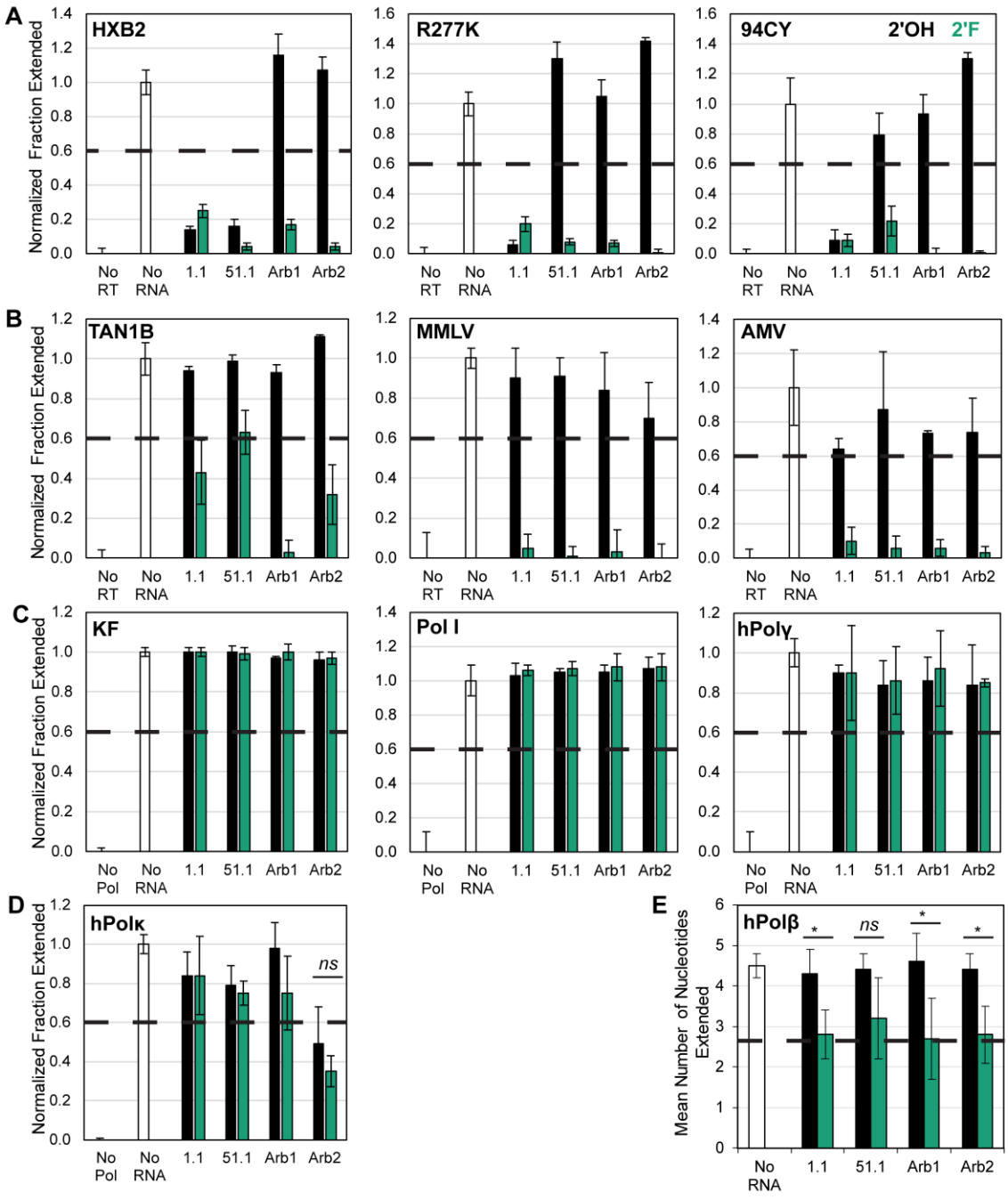

Figure 4.5: 2′-FY RNA Sequences Inhibit Retroviral Reverse Transcriptases. Primer extension assays (n = 4) were used to monitor polymerase inhibition by unmodified RNA (black) and 2′-FY RNA (green) transcripts using (A) RT from HIV-1 Group M strains and an F1Pk-resistant point mutant; (B) RT from SIVcpz and other retroviruses; (C) Klenow Fragment and holoenzyme of DNA Polymerase I and human DNA polymerase γ; (D) human DNA polymerase κ; and (E) human DNA polymerase β. In (E), the dashed horizontal line marks 60% of the No RNA value. Inhibition is observed if the mean number of nucleotides incorporated is below this value. *ns* ($P > 0.05$), * ($P < 0.05$).

RTs from two other retroviruses were also screened. Moloney murine leukemia virus (MMLV) is a gammaretrovirus for which the 75 kDa RT is monomeric in crystal structures (46). Avian myeloblastosis virus (AMV) is an alpharetrovirus with a heterodimeric RT consisting of a 63 kDa α subunit and a 95 kDa β subunit (47). The 2′-FY RNAs strongly inhibited both retroviral RTs, but none of the unmodified RNAs inhibited either MMLV or AMV RT (Figure 4.5B, middle and right panels). Finally, non-viral DNA-dependent DNA polymerases were screened. The Klenow Fragment of *Escherichia coli* DNA polymerase I (KF), *E. coli* DNA polymerase I holoenzyme, and human DNA polymerase γ (hPolγ) were not inhibited by 2′-OH nor 2′-FY RNAs (Figure 4.5C). Human DNA polymerase κ (hPolκ), a Y-family DNA polymerase (48), was not inhibited by the 2′-OH nor the 2′-FY versions of the RT aptamers and Arb1 (Figure 4.5D). Although it was modestly inhibited to equivalent extents by both the 2′-OH and 2′-FY versions of Arb2, there is no clear 2′-FY-mediated enhancement of this inhibition. Human DNA polymerase β (hPolβ) is an X-family DNA polymerase that is typically involved in the base excision repair pathway (49). hPolβ is typically processive for short gaps (1-6 nucleotides) (49), and its low processivity was observed here, compared to the other polymerases tested in this panel. Therefore, we quantified the mean number of nucleotides extended from the 18 nucleotide primer that is annealed to the 31 nucleotide template (maximum of 13 nucleotide extension). A representative gel from this primer extension assay is shown in Figure 4.6. The mean number of nucleotides incorporated in the absence of aptamer or when 2′-OH RNAs were present was approximately 4, while the mean number of nucleotides incorporated when 2′-FY RNAs was approximately 3 (Figure 4.5E). Therefore, while there is a difference between most unmodified and 2′-FY

RNAs, the mean number of nucleotides incorporated was not reduced below 60% of the No RNA value. We conclude that 2′-FY RNAs bind and inhibit RTs from HIV-1 and other retroviruses with little or no specificity, but that this generalized inhibition does not extend to the non-reverse transcribing DNA polymerases tested here.

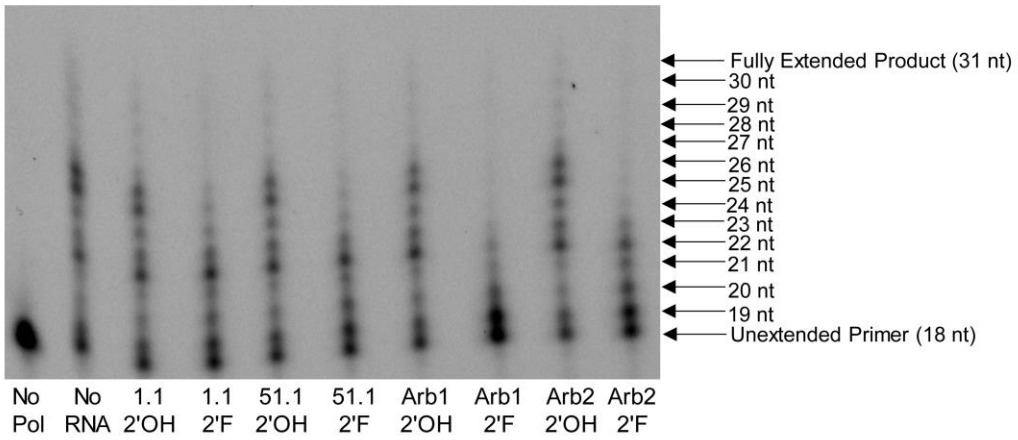

Figure 4.6: Representative Primer Extension Gel for Human Polymerase Beta. Primer extension by hPolβ in the presence of the 2′-OH or 2′-FY RNA sequences labeled below each lane was quantified to determine the mean number of nucleotides incorporated (n=4). The bands corresponding to the unextended primer (18 nt), fully extended product (31 nt), and intermediate products are annotated on the gel.

4.3.3. Sequence and Compositional Determinants of the 2′-FY Effect for RT Inhibition

In principle, RNA structure, length, and composition could all contribute to the strong RT inhibition observed for 2′-FY transcripts that were otherwise non-inhibitory as unmodified 2′-OH transcripts. The Arb1 control RNA is predicted by mfold (50) to be fairly unstructured, with 30-50% of the nucleotides unpaired. When the full-length 2′-FY version of Arb1 was annealed with its unmodified reverse complement, Arb1 alone and the 2′-FY/2′-OH duplex, but not the unmodified reverse complement strand, strongly inhibited HXB2 RT (Figure 4.7). Next, Arb1 RNA was truncated from the 3′ end to generate 100, 80, 64, and 45 nucleotide versions, each of which had similar ratios of pyrimidines to purines as the full-length version. All truncations strongly inhibited HXB2 RT (Figure 4.7). The 2′-fluoro moieties must be part of an oligonucleotide to observe RT inhibition, as adding monomeric 2′-FY NTPs to the extension reactions did not inhibit HXB2 RT (Figure 4.8). To determine how extensively the pyrimidines need to be 2′-F-modified in order to observe strong inhibition of HIV-1 RT, four RNAs were transcribed *in vitro* with different percentages of 2′-FY NTP substitutions, ranging from 0% to 100%, and each RNA was tested for inhibition of RTs that they would not normally inhibit. The F1Pk aptamer was tested against the R277K point mutant of HXB2 RT since this mutation confers resistance to inhibition by F1Pk aptamers, as noted above. The other three RNAs (Arb1, Arb2, and a random-sequence library with 56 random positions) were evaluated against both HXB2 RT and the R277K point mutant. For all seven RT-RNA combinations, no inhibition was observed for fully-unmodified transcripts (0% 2′-FY) and strong inhibition was observed for fully-modified transcripts (100% 2′-FY), with

half-maximal inhibition being observed at approximately 40-60% 2′-FY for most RNAs and slightly higher for the F1Pk aptamer against the R277K point mutant (Figure 4.9). The progressive increase in RT inhibition with increasing mole fraction 2′-FY, observed across multiple unrelated transcripts with difference sequences and structures, suggests that the RT binding mode of 2′-FY RNAs is likely due to direct interactions of the 2′-F groups with RT.

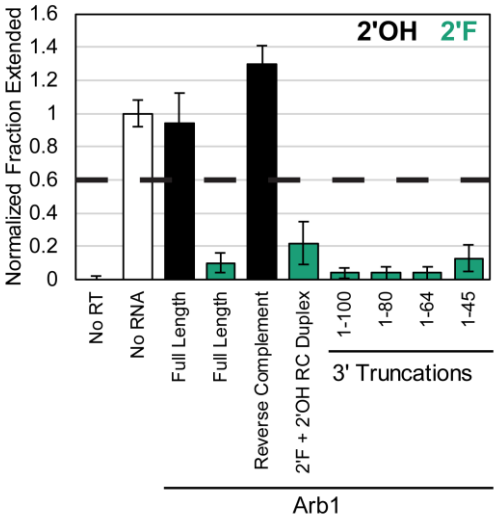

Figure 4.7: RT inhibition by 2′-FY Versions of Alternative forms of Arb1. Arbitrary RNA sequence 1 was converted into a duplex consisting of the 2′FY RNA and its unmodified reverse complement. Additionally, 3′ truncations of the 2′-FY version were generated that contained the indicated nucleotide segments and had similar ratios of pyrimidines to purines of 0.54, 0.55, 0.56, 0.50, and 0.47 for full-length, 100 nt, 80 nt, 64 nt, and 45 nt variants, respectively, and these variants were tested for RT inhibition using a primer extension assay (n = 4).

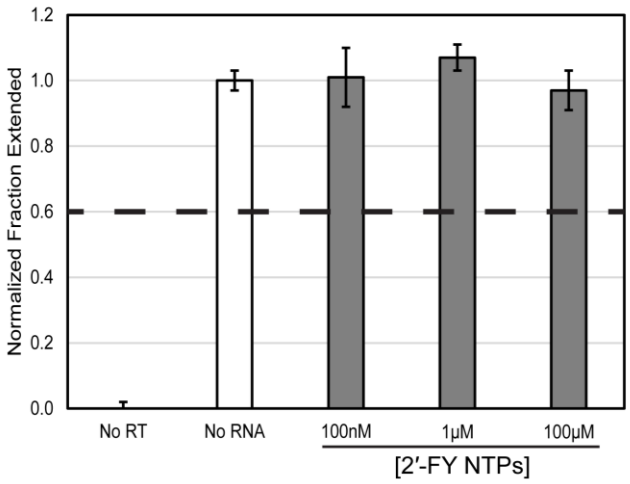

Figure 4.8: No Effect of 2′-FY Nucleotide Triphosphates on RT Inhibition. Quantification of primer extension assays showing fraction of primer converted into full-length product in control reactions in the absence of RT (No RT) or in the absence of RNA inhibitor (No RNA) and in reactions containing varying concentrations of 2′-FY nucleotide triphosphates (NTPs) (n = 4).

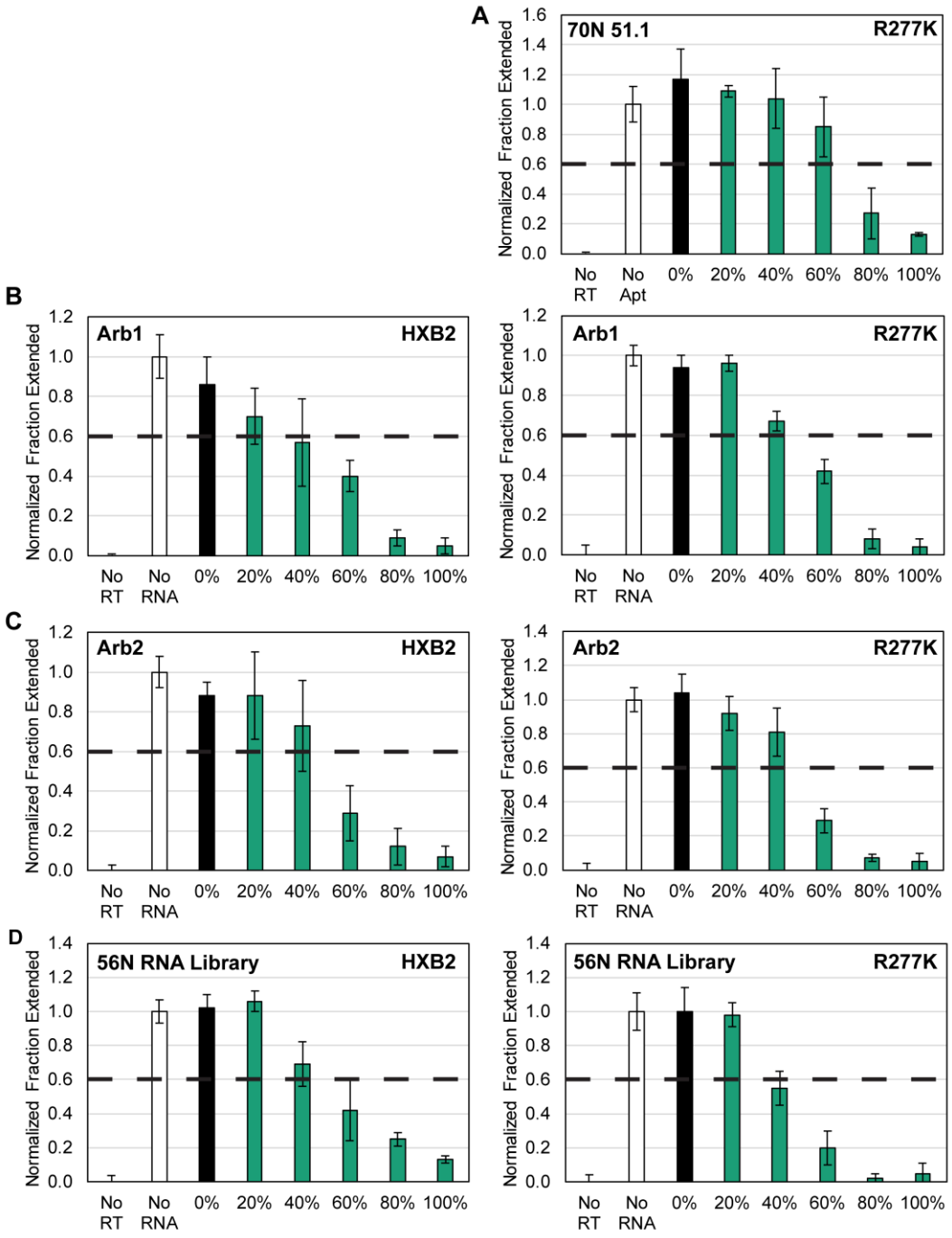

Figure 4.9: Effect of 2′-FY Composition on RT Inhibition. RNA sequences indicated in the upper left of each panel were transcribed with the indicated percentages of 2′-FY NTPs present in the transcription reactions and were then tested for RT inhibition (n = 4) of the RT indicated in the upper right of each panel. Data are plotted for (A) aptamer 70N 51.1, (B) Arb1 control, (C) Arb2 control, and (D) 56N random RNA library.

4.3.4. The 2′-FY Modification Increases the Ionic Character of the Interaction between RNA and HIV-1 RT

The enzymatic inhibition assays above indirectly monitor binding of RT by the RNA transcripts because the transcripts compete with primer-template for access to RT. Therefore, nitrocellulose filter binding assays were performed to measure RT binding directly and to compare apparent binding affinities of the modified and unmodified transcripts. When generated as unmodified transcripts, the (6/5)AL aptamer 70N 1.1 and the F1Pk aptamer 70N 51.1 had apparent dissociation constants (K_D) in the low nanomolar range, as expected, while arbitrary sequences Arb1 and Arb2 did not strongly bind RT (Table 4.6 and Figure 4.10). As 2′-FY RNA transcripts, aptamers 70N 1.1 and 70N 51.1 bound RT with a similar or slightly stronger affinity as the unmodified version. For the 2′-FY versions of arbitrary control RNAs, Arb1 bound RT to a similar degree as 70N 1.1 while Arb2 bound approximately 2-fold less strongly.

Table 4.6: Apparent Binding Affinity (K_D) of Unmodified and 2′-FY RNAs to HXB2 RT

RNA	2′OH K_D (nM)	2′F K_D (nM)
70N 1.1	38 ± 5	43 ± 8
70N 51.1	29 ± 5	17 ± 6
Arb1	>800	44 ± 8
Arb2	>800	87 ± 16

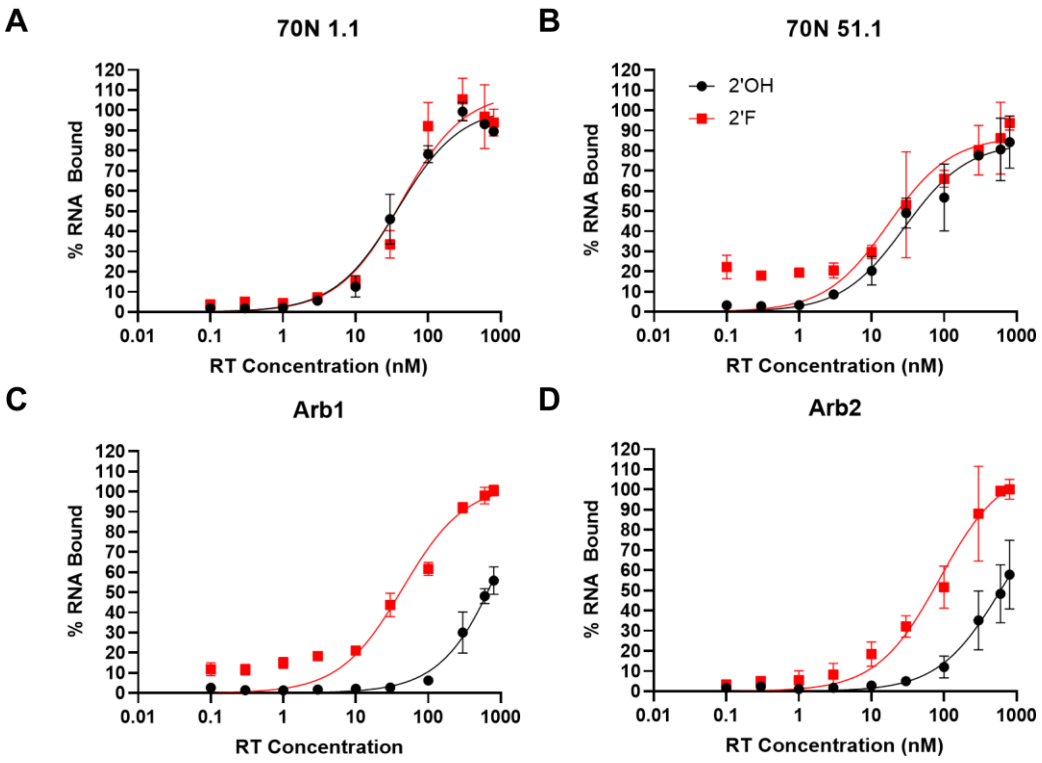

Figure 4.10: Binding Curves of RNA Sequences to HXB2 RT. Binding of (A) 70N 1.1, (B) 70N 51.1, (C) Arb1 control, and (D) Arb2 control to HXB2 RT was measured using a nitrocellulose filter binding assay (n = 3).

To evaluate the ionic character of the binding interaction between RT and 2′-FY RNA, we investigated how salt concentration affected the binding of RT to the 2′-FY and unmodified versions of 70N 1.1 and Arb1. Nitrocellulose filter binding assays were done using binding buffers in which the potassium chloride (KCl) concentration ranged from 75 mM to 500 mM. In all cases, affinity decreased with increasing KCl, as expected for a protein-nucleic acid complex. However, the effect appears to be stronger for the 2′-FY transcripts (Figure 4.11A & B). By plotting the percent RNA bound from each of the fitted curves for the four RNAs when RT concentration is equal to 100 nM versus the KCl concentration present in the binding buffer, there was a nearly linear decrease in binding as salt concentration increased for the four RNAs (Figure 4.11C). However, the binding of 2′-FY RNAs to RT decreased more rapidly as salt concentrations increased, in comparison to the slower drop-off observed for their corresponding unmodified versions, indicating that the 2′-FY modification increases the ionic character of the interaction between the modified RNAs and HIV-1 RT.

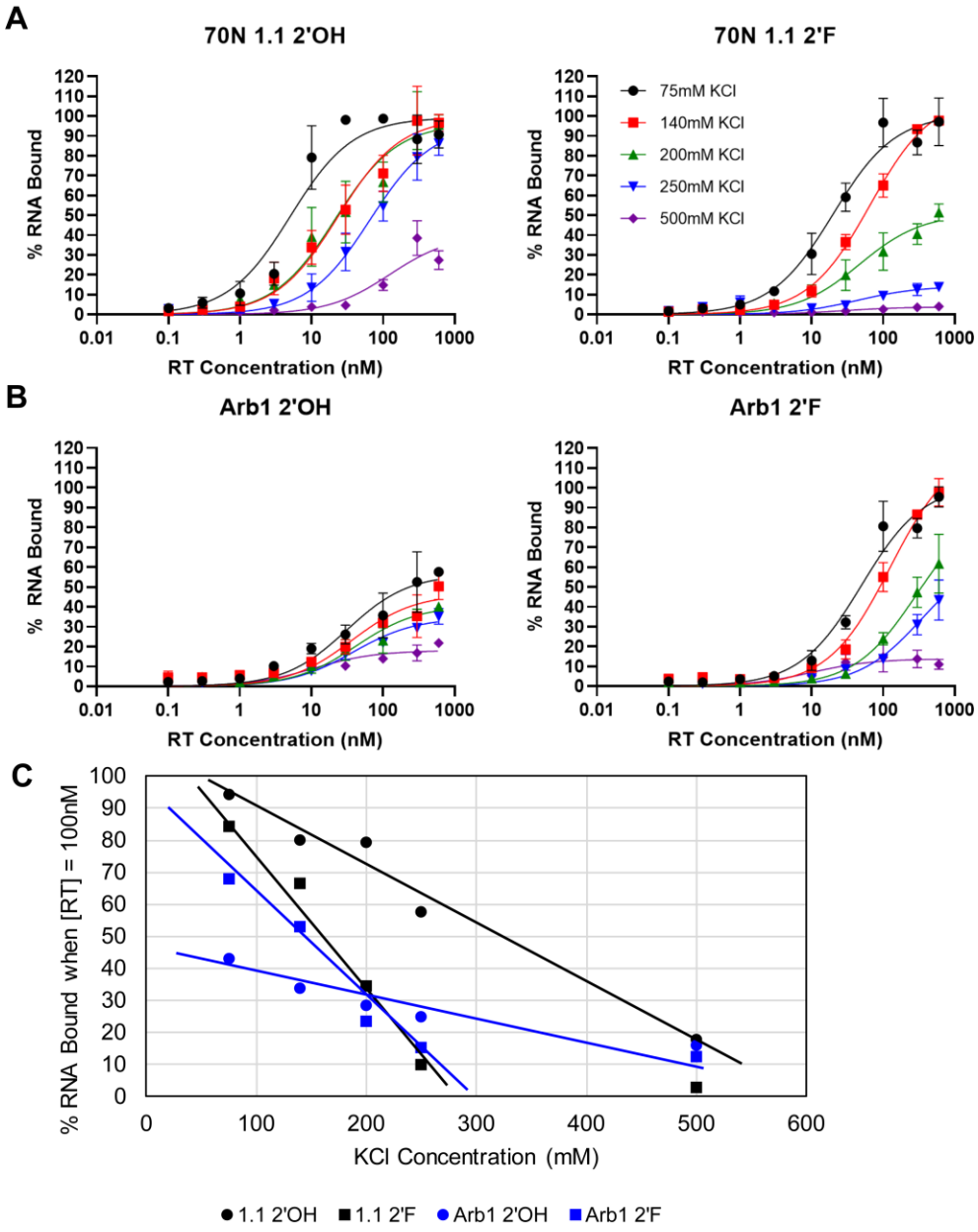

Figure 4.11: Effect of Increasing Salt Concentrations on Binding of RNA Sequences to HXB2 RT. Binding of unmodified and 2′-FY modified (A) 70N 1.1 and (B) Arb1 RNAs to HXB2 RT at increasing salt concentrations. Binding was measuring using a nitrocellulose filter binding assay with alkaline-treated filters (n = 3). (C) Plot of the percent RNA bound when RT concentration is equal to 100 nM from each of the fitted curves for the four data sets in (A and B). A line-of-best-fit is plotted for the first four data points of each set to illustrate the general trend of decreasing binding as salt concentration increases for each sequence.

4.4. Discussion

This study demonstrated a reselection strategy for the identification of chemically-modified aptamers and provided new insights on the impact of 2′-FY modifications on target recognition. A pre-enriched RNA library with affinity for HIV-1 RT was re-transcribed with 2′ modifications, and three additional rounds of selection with the modified libraries were performed. High-throughput sequencing and bioinformatics identified aptamer sequences that productively accommodated 2′ pyrimidine modifications and that still bound HIV-1 RT *in vitro*, and co-enrichment in the parallel trajectories helped identify which sequences to test. Structurally distinct aptamer families exhibited different inhibition profiles to the 2′ modifications. All aptamers tested strongly inhibited RT when 2′-F pyrimidines were incorporated. None of the HIV-1 RT aptamers from previously identified structural families inhibited RT when 2′-NH_2 pyrimidines were incorporated, likely due in part to the fact that the 2′-NH_2 modifications favor the C2′-endo sugar pucker and destabilize RNA duplexes (51). The reselection identified two new aptamers from a single cluster (cluster 31) that were able to inhibit HIV-1 RT when transcribed with 2′-NH_2 pyrimidines. The impact of 2′-OMe pyrimidines was structure-specific, as F1Pk aptamers still inhibited RT, but (6/5)AL and UCAA motif aptamers no longer inhibited RT.

The reselection strategy described here is conceptually similar to the recently described “Poly-Target” selection approach, in which the pre-enriched RNA aptamer library with affinity for RT from the BH10 strain of HIV-1 was subjected to additional cycles of selection for affinity to RTs from phylogenetically diverse lentiviruses (44). Both approaches sought to find specialized aptamer subsets from within pre-enriched

libraries. While the present study looked to find aptamers that tolerate or productively accommodate various 2′ pyrimidine modifications, the goal of the Poly-Target selection approach was to identify aptamers with broad-spectrum target recognition (44). An analogous approach could be used to find aptamer subsets that are compatible with alternative alphabets from among pre-enriched aptamer libraries with other initial compositions and with affinity for other targets. In addition, while the reselection phase of the strategy described here introduced 2′-pyrimidine modifications, the strategy can easily be generalized to other chemical modifications, so long as there are reverse transcriptases and/or polymerases that recognize and incorporate the modified nucleotides. The 'pre-enrichment + reselection' approach requires that there be overlap in the fitness landscapes in which the same sequence (or a close relative) is active for target recognition when generated from two different monomer alphabets. Given appropriately matched monomer sets for the two phases of the selection, this approach could enable exploration of nucleotide chemistries that are too expensive or technically challenging to carry out with the modified nucleotides from the beginning.

The promiscuous recognition of viral reverse transcriptases by generic 2′-FY-modified RNAs has not been described previously and was unexpected. The 2′-F moiety is relatively small and favors the same sugar pucker as unmodified RNA. Previous studies suggested that 2′-F-modified aptamers often have higher affinities than unmodified aptamers and attributed this to formation of more thermodynamically stable secondary structures (51-53). Additional studies determined that 2′-F-modified RNA duplexes are less hydrated than unmodified RNA duplexes, resulting in stronger Watson-Crick hydrogen bonding and π-π stacking interactions between nucleobases (16,54). A

combination of these effects may partially explain our observation that RT inhibition in primer extension assays by *bona fide* RT aptamers was equivalent or more potent for 2′-FY transcripts than for the corresponding 2′-OH versions. Nevertheless, it was surprising that the negative control 2′-FY arbitrary RNA sequences also strongly inhibited RT (Figure 4.1G and Figure 4.5) and that the 2′-FY RNA sequences inhibited diverse retroviral RTs but generally did not inhibit non-viral DNA polymerases (Figure 4.5). For multiple unrelated transcripts, inhibition by 2′-FY RNAs was proportional to the mole fraction of 2′-F pyrimidines present in the transcript (Figure 4.9), thereby establishing that no specific position or subset of positions can be responsible for the observed effects, and further suggesting that the 2′-F moieties may bind directly with the RTs. We also observed that the binding of 2′-FY RNAs to HIV-1 RT was more sensitive to increasing salt concentrations than were their unmodified versions (Figure 4.11), suggesting that the 2′-F modification may increase the ionic character of the interaction between the RNA transcripts and reverse transcriptases. The precise physico-chemical basis for the increased ionic character is not immediately apparent. An important contributing factor may be due to loss of interaction with the proton of the 2′-hydroxyl. Specifically, in unmodified RNA, the 2′-hydroxyl proton can partially neutralize the negative charge of the phosphate backbone, but when replaced with the more electronegative fluorine group, this partial neutralization no longer occurs, and the phosphate retains its full negative charge.

2′-FY modifications are commonly used in aptamer selections because 2′-FY incorporation confers serum stability and because the nucleotide triphosphates are commercially available and are efficiently incorporated by modified T7 RNA

polymerases. Our findings raise the possibility that fluoro modifications could contribute directly to the affinities observed for other aptamer-target pairs by various mechanisms, including via contributions to ionic interactions. 2′-FY RNA aptamers with affinity for HIV-1 RT have not been previously selected directly from random libraries; however, Alves Ferreira-Bravo et al recently reported selection of FANA aptamers with enhanced affinity against HIV-1 RT (3). FANA is structurally distinct from 2′-F NTPs, as the fluoro substituent is in a β conformation for FANA, as compared to an α conformation for 2′-F ribose, and the FANA sugar pucker is in a C2′/O4′-endo conformation, as compared to C3′-endo in 2′-F NTPs (55,56). Those authors found that some of the selected FANA aptamers, such as 'FA1,' appear to bind RT with K_D values in the single-digit pM range, while the FANA random pools bound HIV-1, MMLV, and AMV RTs at low nanomolar affinities. Although they observed that FA1 affinity for RT decreased with increasing ionic strength, it is not yet known whether the salt sensitivity of FANA transcripts is greater than that of RNA or DNA, or whether an increased ionic character may account for some of the enhanced affinity observed for FANA aptamers to HIV-1 RT.

Increased ionic character may be an important feature of other 2′-FY aptamers that are in development for use as theranostics and biosensors. In addition, our observation that the 2′-FY modifications can convert non-inhibitory RNA transcripts into RT inhibitors raises the possibility that 2′-FY aptamers for other targets may be more prone to off-target binding than aptamers with other compositions when presented with highly-basic alternative targets, especially under low salt conditions. In general, *in vitro* assessment of aptamer function should carefully control for ionic composition,

particularly in comparing results from one lab or experimental setting to another. Finally, the inferences made in this study regarding potential contributions of fluoro moieties to target recognition were only made possible by observing a *gain* of function when unmodified aptamer or arbitrary sequence were re-transcribed with 2′-FY nucleotides. Many aptamers that were initially selected with 2′-F pyrimidines have been shown to *lose* affinity for their targets when transcribed with unmodified pyrimidines. The precise basis for this loss of affinity is rarely delineated with precision but could simply reflect structural or other constraints associated with having evolved with the 2′-F modification. It remains to be seen whether fluoro moieties in 2′-FY or FANA aptamers contribute directly to binding affinities of other fluoro-substituted aptamers to their respective targets and/or to unwanted interactions with alternative molecular targets.

4.5. Materials and Methods

4.5.1. RT Expression and Purification. Enzymatically active p66/p51 heterodimer RTs from different viral strains of HIV were purified and validated as described previously (44). Protein concentration was calculated from UV absorbance at 280 nm using an estimated extinction coefficient of 260,120 $M^{-1}cm^{-1}$ based on the amino acid sequence. RT preps were validated for purity and size by SDS-PAGE and for activity by comparison of primer extension assays against previous preps. Proteins were stored at -80° C after addition of glycerol to 50% (v/v).

4.5.2. Modified Aptamer Reselection, High-Throughput Sequencing, and Bioinformatics. The starting library for this work was the round 14 70N library described previously (33). Double-stranded DNA from this library was transcribed *in vitro* to generate unmodified (2′-OH) and 2′-pyrimidine-modified input RNA libraries for

the reselection. 2′-modified pyrimidines were purchased from TriLink Biotechnologies (San Diego, CA). Run-off transcription reactions of 2′-OH, 2′-FY and 2′-NH_2Y transcripts were performed using the Y639F mutant T7 RNA polymerase (25), *in vitro* transcription buffer (50 mM Tris-HCl pH 7.5, 15 mM $MgCl_2$, 5 mM DTT, and 2 mM spermidine), and 2 mM each of ATP, GTP, and the corresponding 2′-modified CTP and UTP. Transcription reactions for 2′-OMe RNAs utilized the same components as above, except that a Y639F/H784 double mutant T7 RNA polymerase (26) was used, and 2 mM each 2′-OMe-CTP and 2′-OMe-UTP, 6 mM $MnCl_2$, 2 mM GMP, 10% PEG 6000, and 1U 10U/μL inorganic pyrophosphatase (Thermo Fisher Scientific) were added to the reaction. Reactions were incubated at 37°C for a minimum of 4 hours and halted with the addition of denaturing gel loading buffer (90% formamide, 50 mM EDTA and trace amounts of xylene cyanol and bromophenol blue). RNAs were purified by denaturing polyacrylamide gel electrophoresis (6% TBE-PAGE, 8 M urea), and bands corresponding to the expected product sizes were excised from the gel and eluted while tumbling overnight in 300 mM sodium acetate pH 5.4. Eluates were ethanol precipitated, resuspended in buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA), and stored at -20°C until further use. RNA concentrations were determined on a NanoDropOne spectrophotometer (Thermo Fisher Scientific). For each trajectory and round of selection, 200 pmol of transcribed libraries (~1.2×10^{14} molecules) were resuspended in 100 μL binding buffer (50mM Tris-HCl [pH 7.5], 150 mM KCl, and 10 mM $MgCl_2$) and renatured by heating to 65°C and cooling on ice. 40 pmol HXB2 RT was then added to a final concentration of 400 nM, and the mixture was incubated on ice for an additional 20 min. The bound RNA species were partitioned from the unbound species using nitrocellulose filters and

recovered as previously described (44). The recovered RNA was reverse-transcribed using ImProm-II Reverse Transcriptase (Promega) and PCR-amplified for the next round of selection or for HTS. Round 14 and Round 17 libraries were prepared for sequencing using a series of PCR steps to add Illumina adapters and sequencing indices for multiplexing of the 70N libraries as previously described (37). The primers used to append the Illumina adapters and sequencing indices can be found in Table 4.7. Sequencing was performed on an Illumina HiSeq2000 (University of Missouri DNA Core Facility). Populations were demultiplexed to identify and parse the 5′ and 3′ constant regions. Data preprocessing was performed using cutadapt (57) to trim 5′ and 3′ constant regions from sequences and to discard any uncut sequences or sequences not within ±3 nt of the expected size (70 nt) after trimming. Trimmed sequences were then filtered for high-quality reads using FASTQ quality filter from the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx-toolkit/). Quality filtering eliminated a sequence if a single position had a Phred quality score of less than 20. Trimmed and quality filtered sequences were then processed using the FASTAptamer toolkit (35) to count and normalize sequence reads (FASTAptamer-Count), calculate fold enrichment from round 14 to round 17 (FASTAptamer-Enrich), and group related sequences into clusters (FASTAptamer-Cluster). Aptamers were named according to the cluster in which they were found in round 14 and their rank in terms of abundance within the cluster. For example, aptamer 70N 31.56 was part of cluster 31 in round 14 and was the 56th most abundant sequence within that cluster. To determine motif enrichment values, FASTAptamer-Search was used to identify sequences from the libraries that matched sequence patterns for the F1Pk [TCCG and GCCC] and (6/5)AL [RCGTY and RARAC]

structural motifs, and a custom Perl script was used to count the total number of motif reads, normalize motif reads, and calculate motif enrichment values.

Table 4.7: Primer Sequences to Append Illumina Adapters and Sequencing Indices

Library	PCR 1 Forward	PCR 1 Reverse	PCR 2 Forward	PCR 2 Reverse	PCR 3 Forward	PCR 3 Reverse[a]
70N Round 14	TAATAC GACTCA CTATAG GGAGAA TCGAAT CATACA CAAGA	TATGGA ATTAAA TACCTT ATGCCC	TTTCCCTAC ACGACGCTC TTCCGATCT GAATCGAAT CATACACAA GA	GTGACTGGA GTTCAGACG TGTGCTCTT CCGATCTTA TGGAATTAA ATACCTTAT	AATGATACG GCGACCACC GAGATCTAC ACTCTTTCC CTACACGAC GCTCTT	CAAGCAGAA GACGGCATA CGAGAT**TCG TAAGCCGTC** GTGACTGGA GTTCAGACG TG
70N 2′OH Round 17 A/B	TAATAC GACTCA CTATAG GGAAAA GCGAAT GATACA CAAGA	TATGGA ATTAAA TACCTT ATGCCC	TTTCCCTAC ACGACGCTC TTCCGATCT GCGAATGAT ACACAAGA	GTGACTGGA GTTCAGACG TGTGCTCTT CCGATCTTA TGGAATTAA ATACCTTAT	AATGATACG GCGACCACC GAGATCTAC ACTCTTTCC CTACACGAC GCTCTT	CAAGCAGAA GACGGCATA CGAGAT**GCT ACTGGTATG** GTGACTGGA GTTCAGACG TG
70N 2′FY Round 17 A/B	TAATAC GACTCA CTATAG GGAAAA GCCAAT CATACA CAAGA	TATGGA ATTAAA TACCTT ATGCCC	TTTCCCTAC ACGACGCTC TTCCGATCT AGCCAATCA TACACAAGA	GTGACTGGA GTTCAGACG TGTGCTCTT CCGATCTTA TGGAATTAA ATACCTTAT	AATGATACG GCGACCACC GAGATCTAC ACTCTTTCC CTACACGAC GCTCTT	CAAGCAGAA GACGGCATA CGAGAT**CAG CGTTTAGCC** GTGACTGGA GTTCAGACG TG
70N 2′OMeY Round 17 A/B	TAATAC GACTCA CTATAG GGAGAA TCGAAT CATACA CAAGA	TATGGA ATTAAA TACCTT ATGCCC	TTTCCCTAC ACGACGCTC TTCCGATCT GAATCGAAT CATACACAA GA	GTGACTGGA GTTCAGACG TGTGCTCTT CCGATCTTA TGGAATTAA ATACCTTAT	AATGATACG GCGACCACC GAGATCTAC ACTCTTTCC CTACACGAC GCTCTT	CAAGCAGAA GACGGCATA CGAGAT**GCC AAATCGCTC** GTGACTGGA GTTCAGACG TG
70N 2′NH_2Y Round 17 A/B	TAATAC GACTCA CTATAG GGAACA GCGAAT CATACA CAAGA	TATGGA ATTAAA TACCTT ATGCCC	TTTCCCTAC ACGACGCTC TTCCGATCT CAGCGAATC ATACACAAG A	GTGACTGGA GTTCAGACG TGTGCTCTT CCGATCTTA TGGAATTAA ATACCTTAT	AATGATACG GCGACCACC GAGATCTAC ACTCTTTCC CTACACGAC GCTCTT	CAAGCAGAA GACGGCATA CGAGAT**GCC ATAGTGTGT** GTGACTGGA GTTCAGACG TG

[a]The index sequence is bolded in the PCR 3 reverse primer.

4.5.3. DNA Templates and RNA Transcription. For each aptamer to be transcribed, DNA oligonucleotides were purchased as 'left' and 'right' halves from Integrated DNA Technologies (Coralville, IA). Oligonucleotides corresponding to the 'right' half were 5′-phosphorylated using T4 polynucleotide kinase (New England Biolabs). Equimolar amounts of 'right' strand, 'left' strand, and a complimentary bridge oligo were annealed and ligated using T4 DNA ligase (New England Biolabs). Ligated templates were PCR amplified using Pfu DNA polymerase, a forward primer to append the T7 promoter, and a reverse primer complimentary to the 3′ constant region. Amplified products were verified for size using agarose gel electrophoresis. The double-stranded DNA templates were then transcribed *in vitro* and purified as above. The RNA sequences used in this study are in Table 4.2. The Arb2 sequence was kindly provided by Dr. Margaret Lange (University of Missouri, Columbia, MO). Two guanine nucleotides were added to the 5′ end of the reverse complement of Arb1 to aid transcription, resulting in a two-nucleotide 5′ overhang in the annealed duplex. For transcriptions to generate mixed 2′-OH/2′-FY polymers, NTPs were adjusted to maintain a constant total NTP concentration. For example, transcription of an RNA containing 40% 2′-FY NTPs utilized 0.8 mM each of 2′-F-CTP and 2′-F-UTP and 1.2 mM each of CTP and UTP.

4.5.4. Primer Extension Assays. Inhibition of the DNA-dependent DNA polymerase activities of HIV-1 RT and other DNA polymerases was monitored using the primer extension assay as previously described (37). Briefly, 5′-Cy3-end-labeled 18 nt DNA primer corresponding to the 3′ end of $tRNA_{Lys3}$ was annealed with a 31 nt DNA template. 20 nM RT was pre-assembled with excess (100 nM) RNA in extension buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$) for 10 min. Note that the ionic conditions

of the extension buffer were different than those of the binding buffer used in the binding studies. The lower monovalent and higher divalent ion concentrations in the extension buffer, relative to the binding buffer, may affect binding affinity of 2′-FY RNAs to RT. Polymerization reactions were initiated by the addition of pre-annealed 10 nM primer and 20 nM template and 100 μM of each dNTP (final concentrations). After 10 min at 37°C, reactions were stopped by adding an equal volume of denaturing gel loading, then heated to 90°C for 2 min immediately prior to loading onto a 10% polyacrylamide, 8M urea denaturing gel. Gels were scanned for Cy3 fluorescence using a Typhoon FLA 9000 phosphoimager (GE Healthcare Life Sciences). The fraction of primer extended to full-length product was quantified by measuring the intensity of the bands using Multiguage software (Fujifilm) and normalized by setting the fraction of primer extended when no RNA was present to 1. Extension reactions using the Klenow fragment of *E. coli* DNA polymerase I (New England Biolabs) were done using 20 nM enzyme per reaction for 5 min in extension buffer containing 10 mM Tris-HCl [pH 7.9], 50 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT. MMLV RT (Promega) extension reactions were done using 20 nM enzyme per reaction for 10 min in extension buffer containing 50 mM Tris-HCl [pH 8.3], 75 mM KCl, 3 mM $MgCl_2$, and 10 mM DTT. Human DNA polymerase γ was gift from Dr. Whitney Yin (University of Texas Medical Center, Galveston, TX), human DNA polymerase κ was a gift from Dr. Robert Eoff (University of Arkansas Medical Center, Little Rock, AR), and human DNA polymerase β, *E. coli* DNA polymerase I holoenzyme, and AMV RT were gifts from Dr. Mukund J. Modak (Rutgers University, Newark, NJ). Extension time for AMV RT and human DNA polymerase γ was 60 min, and the final concentration for these two proteins was 100 nM. The AMV RT reactions

were done in buffer containing 50 mM Tris-HCl [pH 8.3], 50 mM KCl, 10 mM $MgCl_2$, and 10 mM DTT. The human polymerase γ reactions were done in buffer containing 25 mM Tris-HCl [pH 8], 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, and 1 mg/mL BSA. The final concentration of the DNA polymerase I holoenzyme was 20 nM, and the extension time was 15 min. The buffer used for this enzyme was 50 mM Tris-HCL [pH 7.5], 100 mM NaCl, 5 mM $MgCl_2$, and 1 mM DTT. For human DNA polymerase κ, the final concentration was 100 nM, and extension time was 30 min. The buffer used for hPolκ was 40 mM Tris-HCl [pH 8], 60 mM KCl, 5 mM $MgCl_2$, and 10 mM DTT. For human DNA polymerase β, the final concentration was 200 nM, and the extension time was 60 min. The buffer used for hPolβ was the same used for the DNA polymerase I holoenzyme. To quantify the largely distributive extension observed due to the low processivity by hPolβ, the mean number of nucleotides extended was determined. First, the intensity of each individual band within a given lane (up to 14 bands total, 0-13 nucleotides extended onto the primer) was quantified. Values were weighted by multiplying each band's intensity by the number of nucleotides that were extended to produce that DNA product. The sum of the weighted intensities was then divided by the sum of the unweighted intensities within the lane to calculate the mean number of nucleotides extended. P values were calculated using an unpaired t-test computed by GraphPad Prism.

4.5.5. Binding Affinity Determination. Dissociation constant (K_D) values were determined using a nitrocellulose filter binding assay. Approximately 20,000 counts-per-minute of 5′-radiolabeled and refolded RNA was incubated with varying concentrations of RT (0.1 to 800 nM or without RT to determine background binding) in binding buffer

(50 mM Tris-HCl [pH 7.5], 140 mM KCl, 1 mM $MgCl_2$, and 0.1 μg/mL BSA, unless otherwise noted) and allowed to equilibrate at room temperature for 15 min. RNA:RT complexes were then partitioned from unbound RNA by passing samples through a pre-wet nitrocellulose filter under vacuum, as described above for the partitioning step of the selection, and immediately washing with 500 μL binding buffer. Radioactivity retained on the filter was counted by placing filters into scintillation vials, adding 4 mL liquid scintillation fluid, and counting using a liquid scintillation counter. An unfiltered 'No Wash' sample was counted to determine the total amount of radioactivity ('100%') that was added to each binding reaction. The percent RNA bound was calculated by dividing the radioactivity retained on the filter by the radioactivity present in the 'No Wash' sample multiplied by 100%. The percent RNA bound values were fit to a one-site, specific binding curve ($Y = Bmax*X/[K_D+X]$) using GraphPad Prism 6.2. In the equation, Bmax is the maximum specific binding, K_D is the dissociation constant, X is the RT concentration, and Y is the percent RNA bound. Binding assays were done in triplicate. To decrease background binding to nitrocellulose filters for Arb2 RNA and during salt titrations, nitrocellulose filters were incubated in 0.5 M KOH for 20 min, washed extensively with MilliQ water, and then incubated in the appropriate binding buffer for at least 45 min. Such alkaline-treatment has been previously shown to reduce non-specific binding to nitrocellulose filters (58).

4.5.6. Serum Nuclease Assays. 5′ radiolabeled RNA was incubated in DMEM medium containing 10% fetal bovine serum at 37°C. At specified time points, an aliquot of the reaction was removed, added to an equal volume of denaturing gel loading buffer, and placed in an ethanol and dry ice bath to halt the reaction. Samples were run on a

denaturing polyacrylamide gel (6% TBE-PAGE, 8 M urea). The gel was imaged using a Typhoon FLA 9000 phosphoimager (GE Healthcare Life Sciences). The fraction of full-length RNA present in the sample was quantified by measuring the intensity of the band corresponding to the full-length RNA and dividing it by the intensity of all bands present in the lane of the gel using Multiguage software (Fujifilm).

4.6. References

1. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.
2. Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818-822.
3. Alves Ferreira-Bravo, I., Cozens, C., Holliger, P. and DeStefano, J.J. (2015) Selection of 2'-deoxy-2'-fluoroarabinonucleotide (FANA) aptamers that bind HIV-1 reverse transcriptase with picomolar affinity. *Nucleic Acids Res.*, **43**, 9587-9599.
4. Rose, K.M., Alves Ferreira-Bravo, I., Li, M., Craigie, R., Ditzler, M.A., Holliger, P. and DeStefano, J.J. (2019) Selection of 2′-deoxy-2′-fluoroarabino nucleic acid (FANA) aptamers that bind HIV-1 integrase with picomolar affinity. *ACS Chem. Biol.*, **14**, 2166-2175.
5. Kuwahara, M. and Sugimoto, N. (2010) Molecular evolution of functional nucleic acids with chemical modifications. *Molecules*, **15**, 5423-5444.
6. Lapa, S.A., Chudinov, A.V. and Timofeev, E.N. (2016) The toolbox for modified aptamers. *Mol. Biotechnol.*, **58**, 79-92.
7. Stovall, G.M., Bedenbaugh, R.S., Singh, S., Meyer, A.J., Hatala, P.J., Ellington, A.D. and Hall, B. (2014) In vitro selection using modified or unnatural nucleotides. *Curr. Protoc. Nucleic Acid Chem.*, **56**, 9.6.1-9.6.33.
8. Keefe, A.D., Pai, S. and Ellington, A. (2010) Aptamers as therapeutics. *Nat. Rev. Drug Discov.*, **9**, 537-550.

9. Adler, A., Forster, N., Homann, M. and Goringer, H.U. (2008) Post-SELEX chemical optimization of a Trypanosome-specific RNA aptamer. *Com. Chem. Highthroughput Screen.*, **11**, 16-23.

10. Dellafiore, M.A., Montserrat, J.M. and Iribarren, A.M. (2016) Modified nucleoside triphosphates for in-vitro selection techniques. *Front. Chem.*, **4**, 18.

11. Burnett, J.C. and Rossi, J.J. (2012) RNA-based therapeutics: current progress and future prospects. *Chem. Biol.*, **19**, 60-71.

12. Zhou, J. and Rossi, J. (2016) Aptamers as targeted therapeutics: current potential and challenges. *Nat. Rev. Drug Discov.*, **16**, 181-202.

13. Tawiah, K.D., Porciani, D. and Burke, D.H. (2017) Toward the selection of cell targeting aptamers with extended biological functionalities to facilitate endosomal escape of cargoes. *Biomedicines*, **5**, 51.

14. Shigdar, S., Macdonald, J., Connor, M., Wang, T., Xiang, D., Al.Shamaileh, H., Qiao, L., Wei, M., Zhou, S.-F., Zhu, Y. *et al.* (2013) Aptamers as theranostic agents: modifications, serum stability and functionalisation. *Sensors*, **13**, 13624-13637.

15. Guschlbauer, W. and Jankowski, K. (1980) Nucleoside conformation is determined by the electronegativity of the sugar substituent. *Nucleic Acids Res.*, **8**, 1421-1433.

16. Patra, A., Paolillo, M., Charisse, K., Manoharan, M., Rozners, E. and Egli, M. (2012) 2'-fluoro RNA shows increased Watson-Crick H-bonding strength and stacking relative to RNA: evidence from NMR and thermodynamic data. *Angew. Chem. Int. Ed.*, **51**, 11863-11866.

17. Zhou, P., Zou, J., Tian, F. and Shang, Z. (2009) Fluorine bonding — how does it work in protein−ligand interactions? *J. Chem. Inf. Model.*, **49**, 2344-2355.

18. Meek, K.N., Rangel, A.E. and Heemstra, J.M. (2016) Enhancing aptamer function and stability via in vitro selection using modified nucleic acids. *Methods*, **106**, 29-36.

19. Lauridsen, L.H., Rothnagel, J.A. and Veedu, R.N. (2012) Enzymatic recognition of 2'-modified ribonucleoside 5'-triphosphates: towards the evolution of versatile aptamers. *ChemBioChem.*, **13**, 19-25.

20. Darmostuk, M., Rimpelova, S., Gbelcova, H. and Ruml, T. (2015) Current approaches in SELEX: an update to aptamer selection technology. *Biotechnol. Adv.*, **33**, 1141-1161.

21. Kimoto, M., Meyer, A.J., Hirao, I. and Ellington, A.D. (2017) Genetic alphabet expansion transcription generating functional RNA molecules containing a five-letter alphabet including modified unnatural and natural base nucleotides by thermostable T7 RNA polymerase variants. *Chem. Commun.*, **53**, 12309-12312.

22. Larsen, A.C., Dunn, M.R., Hatch, A., Sau, S.P., Youngbull, C. and Chaput, J.C. (2016) A general strategy for expanding polymerase function by droplet microfluidics. *Nat. Commun.*, **7**, 11235.

23. Pinheiro, V.B., Taylor, A.I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J.C., Wengel, J., Peak-Chew, S.-Y., McLaughlin, S.H. *et al.* (2012) Synthetic genetic polymers capable of heredity and evolution. *Science*, **336**, 341-344.

24. Huang, Y., Eckstein, F., Padilla, R. and Sousa, R. (1997) Mechanism of ribose 2′-group discrimination by an RNA polymerase. *Biochemistry*, **36**, 8231-8242.

25. Sousa, R. and Padilla, R. (1995) A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J.*, **14**, 4609-4621.

26. Padilla, R. and Sousa, R. (2002) A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Res.*, **30**, e138.

27. Jackson, L.N., Chim, N., Shi, C. and Chaput, J.C. (2019) Crystal structures of a natural DNA polymerase that functions as an XNA reverse transcriptase. *Nucleic Acids Res.*, **47**, 6973-6983.

28. Wang, Y., Ngor, A.K., Nikoomanzar, A. and Chaput, J.C. (2018) Evolution of a general RNA-cleaving FANA enzyme. *Nat. Commun.*, **9**, 5067.

29. Mei, H., Shi, C., Jimenez, R.M., Wang, Y., Kardouh, M. and Chaput, J.C. (2017) Synthesis and polymerase activity of a fluorescent cytidine TNA triphosphate analogue. *Nucleic Acids Res.*, **45**, 5629-5638.

30. Mei, H., Liao, J.Y., Jimenez, R.M., Wang, Y., Bala, S., McCloskey, C., Switzer, C. and Chaput, J.C. (2018) Synthesis and evolution of a threose nucleic acid aptamer bearing 7-deaza-7-substituted guanosine residues. *J. Am. Chem. Soc.*, **140**, 5706-5713.

31. Eaton, B.E., Gold, L., Hicke, B.J., Janjié, N., Jucker, F.M., Sebesta, D.P., Tarasow, T.M., Willis, M.C. and Zichi, D.A. (1997) Post-SELEX combinatorial optimization of aptamers. *Bioorg. Med. Chem.*, **5**, 1087-1096.

32. Aaldering, L.J., Tayeb, H., Krishnan, S., Fletcher, S., Wilton, S.D. and Veedu, R.N. (2015) Smart functional nucleic acid chimeras: enabling tissue specific RNA targeting therapy. *RNA Biol.*, **12**, 412-425.

33. Burke, D.H., Scates, L., Andrews, K. and Gold, L. (1996) Bent pseudoknots and novel RNAiInhibitors of type 1 human immunodeficiency virus (HIV-1) reverse transcriptase. *J. Mol. Biol.*, **264**, 650-666.

34. Nguyen, P.D.M., Zheng, J., Gremminger, T.J., Qiu, L., Zhang, D., Tuske, S., Lange, M.J., Griffin, P.R., Arnold, E., Chen, S.-J. *et al.* (2020) Binding interface and impact on protease cleavage for an RNA aptamer to HIV-1 reverse transcriptase. *Nucleic Acids Res.*, **48**, 2709-2722.

35. Alam, K.K., Chang, J.L. and Burke, D.H. (2015) FASTAptamer: a bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Mol. Ther. Nucleic Acids*, **4**, e230.

36. Tuerk, C., MacDougal, S. and Gold, L. (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 6988-6992.

37. Ditzler, M.A., Lange, M.J., Bose, D., Bottoms, C.A., Virkler, K.F., Sawyer, A.W., Whatley, A.S., Spollen, W., Givan, S.A. and Burke, D.H. (2013) High-throughput sequence analysis reveals structural diversity and improved potency among RNA inhibitors of HIV reverse transcriptase. *Nucleic Acids Res.*, **41**, 1873-1884.

38. Whatley, A.S., Ditzler, M.A., Lange, M.J., Biondi, E., Sawyer, A.W., Chang, J.L., Franken, J.D. and Burke, D.H. (2013) Potent inhibition of HIV-1 reverse

transcriptase and replication by nonpseudoknot, "UCAA-motif" RNA aptamers. *Mol. Ther. Nucleic Acids*, **2**, e71.

39. Green, L., Waugh, S., Binkley, J.P., Hostomska, Z., Hostomsky, Z. and Tuerk, C. (1995) Comprehensive chemical modification interference and nucleotide substitution analysis of an RNA pseudoknot inhibitor to HIV-1 reverse transcriptase. *J. Mol. Biol.*, **247**, 60-68.
40. Bernhart, S.H., Hofacker, I.L., Will, S., Gruber, A.R. and Stadler, P.F. (2008) RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinform.*, **9**, 474.
41. Katoh, K., Misawa, K., Kuma, K.I. and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*, **30**, 3059-3066.
42. Lange, M.J., Nguyen, P.D.M., Callaway, M.K., Johnson, M.C. and Burke, D.H. (2017) RNA-protein interactions govern antiviral specificity and encapsidation of broad spectrum anti-HIV reverse transcriptase aptamers. *Nucleic Acids Res.*, **45**, 6087-6097.
43. Lange, M.J., Sharma, T.K., Whatley, A.S., Landon, L.A., Tempesta, M.A., Johnson, M.C. and Burke, D.H. (2012) Robust suppression of HIV replication by intracellularly expressed reverse transcriptase aptamers is independent of ribozyme processing. *Mol. Ther.*, **20**, 2304-2314.
44. Alam, K.K., Chang, J.L., Lange, M.J., Nguyen, P.D.M., Sawyer, A.W. and Burke, D.H. (2018) Poly-target selection identifies broad-spectrum RNA aptamers. *Mol. Ther. Nucleic Acids*, **13**, 605-619.

45. Held, D.M., Kissel, J.D., Thacker, S.J., Michalowski, D., Saran, D., Ji, J., Hardy, R.W., Rossi, J.J. and Burke, D.H. (2007) Cross-clade inhibition of recombinant human immunodeficiency virus type 1 (HIV-1), HIV-2, and simian immunodeficiency virus SIVcpz reverse transcriptases by RNA pseudoknot aptamers. *J. Virol.*, **81**, 5375-5384.

46. Das, D. and Georgiadis, M.M. (2004) The crystal structure of the monomeric reverse transcriptase from Moloney murine leukemia virus. *Structure*, **12**, 819-829.

47. Konishi, A., Nemoto, D., Yasukawa, K. and Inouye, K. (2011) Comparison of the thermal stabilities of the αβ heterodimer and the α subunit of avian myeloblastosis virus reverse transcriptase. *Biosci. Biotech. Bioch.*, **75**, 1618-1620.

48. Sale, J.E., Lehmann, A.R. and Woodgate, R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.*, **13**, 141-152.

49. Yamtich, J. and Sweasy, J.B. (2010) DNA polymerase family X: function, structure, and cellular roles. *Biochim. Biophys. Acta*, **1804**, 1136-1150.

50. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406-3415.

51. Aurup, H., Tuschl, T., Benseler, F., Ludwig, J. and Eckstein, F. (1994) Oligonucleotide duplexes containing 2'-amino-2'-deoxycytidines: thermal stability and chemical reactivity. *Nucleic Acids Res.*, **22**, 20-24.

52. Pagratis, N.C., Bell, C., Chang, Y.-F., Jennings, S., Fitzwater, T., Jellinek, D. and Dang, C. (1997) Potent 2′-amino-, and 2′-fluoro-2′- deoxyribonucleotide RNA inhibitors of keratinocyte growth factor. *Nat. Biotech.*, **15**, 68-73.

53. Cummins, L.L., Owens, S.R., Risen, L.M., Lesnik, E.A., Freier, S.M., McGee, D., Guinosso, C.J. and Cook, P.D. (1995) Characterization of fully 2'-modified oligoribonucleotide hetero- and homoduplex hybridization and nuclease sensitivity. *Nucleic Acids Res.*, **23**, 2019-2024.

54. Pallan, P.S., Greene, E.M., Jicman, P.A., Pandey, R.K., Manoharan, M., Rozners, E. and Egli, M. (2011) Unexpected origins of the enhanced pairing affinity of 2'-fluoro-modified RNA. *Nucleic Acids Res.*, **39**, 3482-3495.

55. Berger, I., Tereshko, V., Ikeda, H., Marquez, V.E. and Egli, M. (1998) Crystal structures of B-DNA with incorporated 2'-deoxy-2'-fluoro-arabino-furanosyl thymines: implications of conformational preorganization for duplex stability. *Nucleic Acids Res.* **26**, 2473-2480.

56. Trempe, J.F., Wilds, C.J., Denisov, A.Y., Pon, R.T., Damha, M.J. and Gehring, K. (2001) NMR solution structure of an oligonucleotide hairpin with a 2'F-ANA/RNA stem: implications for RNase H specificity toward DNA/RNA hybrid duplexes. *J. Am. Chem. Soc.*, **123**, 4896-4903.

57. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 3.

58. McEntee, K., Weinstock, G.M. and Lehman, I.R. (1980) recA protein-catalyzed strand assimilation: stimulation by Escherichia coli single-stranded DNA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.*, **77**, 857-861.

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

Due to their unique binding properties, aptamers can potentially be useful in a variety of applications. However, a limiting step after performing an aptamer selection is the identification of aptamer sequences with the desired properties. The results described in this dissertation provide new strategies and insights for aptamer selections, bioinformatic analyses of aptamer populations from high-throughput sequencing, and biochemical characterization of aptamer-target interactions that may help to increase the likelihood of identifying aptamers with desired binding specificities in future selections.

5.1. Summary of Results

Chapter 2 described the first successful aptamer selection against the assembled capsid lattice and the identification of aptamer CA15-2(1), which was found to bind specifically to the assembled capsid lattice and not to the capsid hexamer or capsid monomer. We demonstrated that the binding of CA15-2 to the assembled capsid lattice is specific since unlabeled CA15-2, but not other unlabeled nucleic acids, could compete for binding to the lattice against labeled CA15-2. Furthermore, CA15-2 was found to inhibit HIV infectivity at the producer cell stage but not the target cell stage, indicating that CA15-2 may impact virus assembly or maturation.

Chapter 3 described the follow-up study using the assembled HIV-1 capsid lattice aptamers. In this study, we utilized a differentiation selection approach using a combination of positive and negative selection steps against the assembled capsid lattice or capsid hexamer. The resulting aptamer populations were analyzed using high-throughput sequencing and bioinformatics using FASTAptamer(2) and a pre-release

version of FASTAptameR 2.0(3) to identify aptamer candidates for further characterization. Based on enrichment values of individual aptamer sequences and their related clusters of aptamer sequences, we predicted the binding phenotypes for the aptamer candidates and biochemically tested these aptamers for binding against the assembled capsid lattice and hexamer. We found that 17 of the 23 candidate aptamers tested had binding phenotypes that matched the predictions (~74% success rate). After the initial binding characterization, we then wanted to determine what sequence or structural motifs were present in the aptamer populations. A G-rich motif was identified in 725 of the top 1000 cluster seed sequences from the lattice round 15 library. In addition, we also identified the GUGUAU motif that is present in L15.6.1 (CA15-2), and another G-rich motif that was present in 75 cluster seed sequences each. Focusing on representative aptamers that contain one of the top three sequence motifs identified and the lattice and hexamer binder H7.10.1, we began to explore the structural features of these aptamers by testing truncations, enzymatic probing using structure-sensitive nucleases, and the generation of covariance models starting from a sequence alignment and predicted consensus secondary structures of the top sequences within a cluster and searching against the top 1000 cluster seed sequences from the lattice round 15 library. Using this combination of approaches, the sequence and structural requirements for these aptamers to bind to the assembled capsid lattice have begun to be uncovered.

Chapter 4 described an unexpected binding phenotype for RNA oligonucleotides modified with 2′-fluoro pyrimidines and HIV-1 reverse transcriptase(4). This study had started with developing a new strategy for selecting aptamers with chemical modifications by starting with aptamer populations that were pre-enriched for binding to

the target of interest, transcribing them with the modification of interest, and performing additional selection rounds. After performing high-throughput sequencing on the aptamer populations and using FASTAptamer to analyze the results, we found that family 1 pseudoknot aptamers retained inhibition of HIV-1 reverse transcriptase when modified with 2′-O-methyl-modified pyrimidines, and we identified a cluster of previously unidentified reverse transcriptase aptamers that could still inhibit RT when modified with 2′-amino-modified pyrimidines. However, we observed that all 2′-fluoro-pyrimidine (2′-FY) aptamers inhibited RT. This RT inhibition was also observed with a 2′-FY-modified arbitrary RNA, while the unmodified version and the versions modified with 2′-O-methyl- or 2′-amino-modified did not inhibit RT. We found that the 2′-FY-modified RNAs inhibited a diverse panel of retroviral RTs, including sequences that did not inhibit these RTs as unmodified RNAs. We also found that the magnitude of the RT inhibition by 2′-FY RNAs was proportional to the mole fraction of 2′-FY present in the RNA transcript. Finally, we observed that binding of 2′-FY RNAs to HIV-1 RT was more sensitive to increasing salt concentration than unmodified RNAs, suggesting that the interaction of 2′-FY RNAs to RT is more ionic in character. This unexpected binding phenotype with 2′-FY RNAs and RT is a clear example of why it is important to check that the modification being incorporated into an aptamer library does not cause all sequences to bind to the target prior to starting a selection or reselecting with modifications.

5.2. Considerations for Future Aptamer Selections Combined with High-Throughput Sequencing and Bioinformatic Analyses

Combining aptamer selections with HTS and subsequent bioinformatic analyses can increase the likelihood of identifying aptamers with desired phenotypes. The capsid differentiation selection and the RT aptamer re-selection with 2′-modified pyrimidines both demonstrate the usefulness of using HTS to analyze aptamer libraries, as well as highlight other considerations that should be taken into account for future studies. To get the most useful information out of the HTS data, it is important to sequence multiple libraries from a selection trajectory. Ideally, the rounds sequenced should include early rounds, final rounds, and any branch points within a differentiation selection. Sequencing only the final round of an aptamer selection can determine the relative frequencies of the aptamers present within the population. However, no comparisons can be made, and often the most abundant aptamer present in a population is not the best binder.

The results from the capsid differentiation selection and the RT aptamer re-selection with the 2′-modified pyrimidines highlight the benefits of looking at individual sequence-level enrichment and population-level enrichment. Similar enrichment values calculated from the normalized reads per million (RPM) from two rounds of a selection can occur for two aptamers with very different frequencies within two populations being compared. For example, two aptamers have RPM values of 10 and 1,000, respectively, in round X and 1,000 and 100,000, respectively, in round Y will both have enriched 100-fold. By looking at how clusters of related aptamers enrich or deplete using the same two rounds of selection for comparison, it can increase the accuracy of the prediction since it takes into account more data.

However, both these selections also highlight the challenge of using *in vitro* selection to identify sequences with specific characteristics. In particular, it can be difficult to identify aptamers from libraries that had converged. For the capsid differentiation selection, approximately 67% of the HTS reads in lattice round 15 were the top twenty most abundant sequences. Although this was not known until after high-throughput sequencing was performed, this population had converged. Most of these top sequences were not predicted to be hexamer binders due to their strong depletion from lattice round 15 to round lattice 17 + hexamer 7, and all aptamers that were predicted to be lattice specific and not bind to the hexamer were indeed lattice-specific binders. For a selection against multiple targets, strong depletion against a target seems to be a good indicator that the aptamer likely does not bind the target. In contrast, binding predictions were less successful for aptamer sequences and their related clusters that had enrichment values near 1 (i.e., population did not change) or greater and were, therefore, predicted to be lattice and hexamer binders. This could be due to the fact that there were not many lattice and hexamer binders in the lattice round 15 library that was used as the branch point. It is also likely that the sequences that had enrichment values near 1 or greater in the lattice and hexamer trajectory had some affinity for the nitrocellulose filter that was used in the partition step.

For the RT aptamer re-selection using the different 2′-modifications, greater than 95% of the RT aptamer round 14 library had previously been found to be family 1 pseudoknot (F1Pk) aptamers. Therefore, it would be difficult to remove F1Pk aptamers from the population during the re-selection, even though the 2′-amino-modified versions of F1Pk aptamers did not inhibit RT in primer extension assays, suggesting that this

modification abolished binding to RT. It was found that F1Pk aptamers had motif enrichment values of 1.43 and 0.76 for the two trajectories using the 2′-amino-modified pyrimidines (Table 4.4). In contrast, for the two modifications not tolerated by the (6/5) asymmetric loop [(6/5)AL] aptamers, 2′-O-methyl- and 2′-amino-modified pyrimidines, the (6/5)AL aptamers had strongly depleted enrichment values less than 0.031 for the two trajectories per modification (Table 4.5).

From these observations using converged libraries, it may be better to start a differentiation selection using pre-enriched but less converged libraries. Depending on the selection, aptamer libraries have been found to typically converge between rounds five and fifteen(5). Therefore, to determine where to start the differentiation selection, high-throughput sequencing can be performed on multiple rounds from the aptamer selection that had demonstrated the desired phenotype in binding assays. Based on the sequencing data, the branch point for the differentiation selection would occur at a round where some aptamer enrichment was observed but had not completely converged. The branch point should be an aptamer library that contain a few highly abundant sequences with RPM values between 10,000 and 100,000 (i.e., 1-10% of the total sequence reads each) and high sequence diversity. However, it is difficult to suggest specific parameters since sequence distributions are dependent on the selection and will vary across aptamer libraries. For example, the lattice round 15 library had seven sequences with RPM values greater than 50,000 and made up nearly 50% of the total reads from this library. In contrast, the lattice round 8 library had one sequence with an RPM value greater than 50,000, and the ten most abundant sequences in this library made up nearly 30% of the

total reads. Based on this information, it may have been better to use the lattice round 8 library as the branch point instead of the lattice round 15 library.

One step that improved the ability to interpret the HTS data from the capsid differentiation selection was the incorporation of a positive selection trajectory using the nitrocellulose filters. By performing two complete rounds of selection against nitrocellulose, every RNA sequence from the lattice round 15 library had the opportunity to pass through or bind to the nitrocellulose filter. Therefore, enrichment values from the nitrocellulose trajectory and from the lattice and hexamer trajectory could be compared to determine if enrichment in the lattice and hexamer trajectory was due to hexamer binding or nitrocellulose binding. While most sequences did not show much enrichment in the nitrocellulose trajectory, L15.45.1 had enriched approximately 20-fold in the nitrocellulose trajectory and 66-fold in the positive hexamer trajectory. Because of its enrichment within the nitrocellulose trajectory, L15.45.1 was predicted to likely not be a hexamer binder, and it was found experimentally not to bind to the hexamer. To improve the usefulness of this trajectory, it may be useful to do the number of rounds with nitrocellulose equal to the number of rounds where nitrocellulose is used in the positive hexamer rounds. Therefore, if sequences do not bind hexamer but do bind nitrocellulose, then the enrichment values for these sequences in these two trajectories should be comparable.

5.3. Future Studies for HIV-1 Capsid-Binding Aptamers

In addition to the insights gained for future aptamer selections and bioinformatic analyses of HTS data, the two HIV-1 capsid aptamer studies lay the foundation for a very exciting research program for using capsid-binding aptamers to study the roles of capsid

and its assembly states in HIV-1 replication. As mentioned in Chapter 3, current efforts are focused on determining the secondary structures of some of the aptamers and where the aptamers bind on the capsid lattice. By understanding how the aptamer binds to the lattice, capsid-binding aptamers can be engineered to be leveraged to answer interesting questions. For example, these aptamers could become the molecular recognition element of an aptamer-based biosensor for capsid for use *in vitro* or in biological assays(6). Additionally, multivalent CA aptamer assemblies can be engineered to mimic the TRIM5α cage around the incoming capsid core in target cells. Currently, none of the aptamers tested inhibit infection in a target cell, although that does not necessarily mean that they do not bind to the incoming capsid core. The multimerization of aptamers that cooperatively bind to the capsid lattice could possibly alter capsid stability or host factor interactions and, therefore, inhibit viral replication. Because some aptamers have been shown to tolerate having an antisense oligonucleotide on their 3′ constant region, one design could include an oligonucleotide containing repeats of the antisense sequence of the 3′ constant region separated by linker regions in between to accommodate the space between binding sites on the lattice and space for the aptamer to properly fold. Therefore, the multivalent aptamer assembly can form through base pairing. Other designs could utilize a DNA origami approach(7,8), where the aptamers are placed in specific locations within a larger DNA-based nanostructure to match the geometry of the binding sites of the capsid lattice. Additionally, recent studies have utilized aptamers to recruit E3 ubiquitin ligases to target a protein of interest for proteasomal degradation(9,10). Therefore, it could be possible to mimic TRIM5α's ability to ubiquitinate the capsid core and target it for proteasomal degradation. This potentially could be done by attaching a

capsid lattice-binding aptamer to an E3 ubiquitin ligase ligand to generate a proteolysis targeting chimera(10).

Capsid-binding aptamers can also be used as probes to further understand HIV biology. Nucleic acids can be functionalized with other moieties for different applications. For example, capsid-binding aptamers could be modified with fluorophores to track the capsid as it moves toward the nucleus in microscopy experiment, as long as aptamer binding does not alter capsid stability. Therefore, the aptamers could be used to track the presence of CA assembly states. Using aptamers to monitor the presence of CA assembly states is complementary to other approaches that monitor the steps from the release of capsid into the cytoplasm to integration of the provirus and subsequent transcription, including the use fluid phase markers to monitor capsid uncoating(11,12), the incorporation of Bgl stem loops into the viral genome to locate HIV-1 transcription sites due to the presence of Bgl-labeled mCherry(12), and the use of fluorescently-labeled host or viral factors to monitor host factor interactions(12,13) and the pre-integration complex(11), respectively. In addition, it is possible that capsid-binding aptamers may be used to monitor the formation of the immature and mature lattices during virus assembly and maturation. Biotinylated aptamers could be used for pull-down experiments, and these experiments could potentially be used to identify host factors that specifically interact with the different CA assembly states. This would be a complementary approach to previous host factor screens that utilized affinity purification methods in combination with mass spectroscopy to identify interacting partners with the Gag polyprotein(14).

Finally, while we have identified aptamers that are lattice-specific or that bind both lattice and hexamer, selections could be done to identify aptamers that are hexamer-

specific or pentamer-specific. Based on the results from the capsid differentiation selection, starting from a random library or a pre-enriched library that had not fully converged may increase the success of these selections. It is likely that aptamers specific for these capsid assembly states will have sequence or structural requirements that are different from the aptamers that have specificity for only the lattice. For example, the lattice and hexamer binding aptamer H7.10.1 is predicted to contain sequence and structural requirements that did not match with any of the other top 1000 seed sequences from the lattice round 15 library or the L17+H3 library nor did it contain any of the recurring motifs found in these libraries identified using MEME Suite(15). These aptamers could be particularly useful for tracking the assembly states of capsid in microscopy experiments.

5.4. Final Conclusions

This thesis highlights the benefits of combining bioinformatics and biochemical methods to study aptamer-protein target interactions. Chapters 2 and 3 describe the identification and characterization of HIV-1 capsid-binding aptamers. Chapter 2 described the initial selection of aptamers to the assembled capsid lattice and characterization of aptamer CA15-2. Chapter 3 described a capsid differentiation selection starting with the lattice round 15 population and subsequent high-throughput sequencing and bioinformatic analyses and the biochemical characterization of the candidate aptamers. Chapter 4 described an aptamer re-selection to identify aptamers that could tolerate different 2′ modifications, and how all 2′-fluoro-pyrimidine-modified RNAs tested strongly inhibited HIV-1 RT and other retroviral RTs. High-throughput sequencing of multiple rounds during an aptamer selection generates a significant amount

of data, and careful analyses of these data can help to identify candidate sequences for further characterization. Based on how an aptamer and its cluster of related sequences enrich or deplete within a selection trajectory, one can make a prediction on an aptamer's binding phenotype. Careful design and execution of the selection, preparation for high-throughput sequencing, and bioinformatic analyses increase the strength of the method. Using this workflow, binding phenotypes of the aptamers from the capsid differentiation selection were predicted with a high success rate. Combining the selection technology with high-throughput sequencing and bioinformatic analyses was also useful for identifying HIV-1 RT aptamers that could tolerate the incorporation of the different 2′-modified pyrimidines. Future aptamer selection studies can benefit from the insights gained from these studies to increase the likelihood of identifying aptamers with desired binding phenotypes.

5.5. References

1. Gruenke, P.R., Aneja, R., Welbourn, S., Ukah, O.B., Sarafianos, S.G., Burke, D.H. and Lange, M.J. (2022) Selection and identification of an RNA aptamer that specifically binds the HIV-1 capsid lattice and inhibits viral replication. *Nucleic Acids Res.*, **50**, 1701-1717.
2. Alam, K.K., Chang, J.L. and Burke, D.H. (2015) FASTAptamer: a bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Mol. Ther. Nucleic Acids*, **4**, e230.
3. Kramer, S.T., Gruenke, P.R., Alam, K.K., Xu D. and Burke, D.H. (2022) FASTAptameR 2.0: a web tool for combinatorial sequence selections. bioRxiv doi: https://doi.org/10.1101/2022.04.27.489774, 29 April 2022, pre-print: not peer-reviewed.
4. Gruenke, P.R., Alam, K.K., Singh, K. and Burke, D.H. (2020) 2′-fluoro-modified pyrimidines enhance affinity of RNA oligonucleotides to HIV-1 reverse transcriptase. *RNA*, **26**, 1667-1679.
5. McKeague, M., McConnell, E.M., Cruz-Toledo, J., Bernard, E.D., Pach, A., Mastronardi, E., Zhang, X., Beking, M., Francis, T., Giamberardino, A. *et al.* (2015) Analysis of in vitro aptamer selection parameters. *J. Mol. Evol.*, **81**, 150-161.
6. Zhou, W., Huang, P.J., Ding, J. and Liu, J. (2014) Aptamer-based biosensors for biomedical diagnostics. *Analyst*, **139**, 2627-2640.

7. Chandrasekaran, A.R., Anderson, N., Kizer, M., Halvorsen, K. and Wang, X. (2016) Beyond the fold: emerging biological applications of DNA origami. *ChemBioChem*, **17**, 1081-1089.

8. Hong, F., Zhang, F., Liu, Y. and Yan, H. (2017) DNA origami: scaffolds for creating higher order structures. *Chem. Rev.*, **117**, 12584-12640.

9. He, S., Gao, F., Ma, J., Ma, H., Dong, G. and Sheng, C. (2021) Aptamer-PROTAC conjugates (APCs) for tumor-specific targeting in breast cancer. *Angew. Chem. Int. Ed.*, **60**, 23299-23305.

10. Ma, S., Ji, J., Tong, Y., Zhu, Y., Dou, J., Zhang, X., Xu, S., Zhu, T., Xu, X., You, Q. *et al.* (2022) Non-small molecule PROTACs (NSM-PROTACs): protein degradation kaleidoscope. *Acta Pharma. Sin. B* (In Press).

11. Mamede, J.I., Cianci, G.C., Anderson, M.R. and Hope, T.J. (2017) Early cytoplasmic uncoating is associated with infectivity of HIV-1. *Proc. Natl. Acad. Sci. U.S.A.*, **114,** E7169-E7168.

12. Li, C., Burdick, R.C., Nagashima, K., Hu, W.S. and Pathak, V.K. (2021) HIV-1 cores retain their integrity until minutes before uncoating in the nucleus. *Proc. Natl. Acad. Sci. U.S.A.*, **118,** e2019467118.

13. Francis, A.C. and Melikyan, G.B. (2018) Single HIV-1 imaging reveals progression of infection through CA-dependent steps of docking at the nuclear pore, uncoating, and nuclear transport. *Cell Host Microbe*, **23**, 536-548.

14. Engeland, C.E., Brown, N.P., Börner, K., Schümann, M., Krause, E., Kaderali, L., Müller, G.A. and Kräusslich, H.-G. (2014) Proteome analysis of the HIV-1 Gag interactome. *Virology*, **460-461**, 194-206.

15. Bailey, T.L. and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **2**, 28-36.

Vita

Paige Rose Gruenke, the daughter of Glenn and Tamara Gruenke, was born on November 10, 1992. She attended Annunciation BVM Catholic Elementary School and Aurora Central Catholic High School, where she was the valedictorian of her class. She earned her Bachelor of Science in biochemistry from Saint Louis University, St. Louis, MO. She matriculated into the biochemistry doctoral program at the University of Missouri in August 2015. She earned her PhD in May 2022.