Identifying vulnerable sites of the HIV-1 capsid protein

Siriphan Manocheewa

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

James I. Mullins, Chair

John E. Mittler

Roger E. Bumgarner

Program Authorized to Offer Degree:

Microbiology

University of Washington

Abstract

Identifying vulnerable sites of the HIV-1 capsid protein

Siriphan Manocheewa

Chair of the Supervisory Committee:

Professor James Mullins

Microbiology

One very challenging aspect of developing HIV vaccines and therapies is to overcome the high evolutionary rate and consequent sequence diversity of the virus. HIV is notorious for rapidly acquiring drug resistant and immune escape mutations, which allow the virus to survive and persist against antiviral drug suppression and host immune responses. I hypothesized that functionally and structurally conserved elements of the viral proteome with little-to-no tolerance to mutations would be good candidates for targets of vaccine-induced responses and antiretroviral drugs. To identify such regions in the HIV-1 capsid protein, I analyzed data from 5000 HIV capsid sequences from Genbank and the Los Alamos HIV sequence database to estimate sequence conservation and mutation frequencies. The structural location for each amino acid residue was determined based on the high-resolution X-ray crystal structure of the hexameric form of the capsid protein (the major morphological subunit of the mature HIV-1 capsid). *In vitro* pairwise growth competition assays were then carried out to determine the relative fitness cost of the most frequently observed mutations at capsid

hexamerization interface sites and non-interface sites. Only a weak relationship between sequence conservation and the relative fitness was detected. On the other hand, the most frequently observed mutations at interface sites had larger fitness costs than the mutations at the non-interface sites, suggesting that interface sites are could be suitable targets for HIV vaccines and therapies. In addition to protein interface sites, I used homology protein modeling and two protein stability prediction methods to investigate links between changes in protein stability and the impact of mutations in the capsid protein on viral replication. I found that mutations predicted to induce large alterations in the dimerization of or the structural stability of the CA hexamer were far less likely to be found in the HIV sequence database than those not predicted to alter stability. Destabilizing mutations were also associated with deleterious phenotypes. Compared to mutation frequency, predicted protein stability was a better classifier of deleterious and non-deleterious mutations. However, a newly derived simple composite score, which takes into account both mutation frequency and the proteins stability score, performed better than both protein staibilty and mutation frequency alone. These results suggest that both sequence conservation and *in silico* structural stability should be used to identify potentially inactivating mutations. Utilizing optimized *in vitro* pairwise growth competition assays and *in silico* mutation and protein stability predictions, along with sequence conservation, I identified potentially important sites in the HIV-1 capsid protein that warrant further investigation as candidates for drug and CTL vaccine targets.

DEDICATION

My beloved father, mother, sisters and brother

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to the following people whose supports and guidance enable me to complete my Ph.D. study successfully.

My parents have always encouraged and supported me to pursue my dream. They never set the boundary on what I can achieve. I am very fortunate to have a loving and supportive sisters and brother. Their unconditional love and unwavering supports have allowed me to overcome challenges of living and studying abroad.

I am deeply indebted to Dr. James I Mullins, my Ph.D. mentor and chair of my thesis supervisory committee, for his tremendous help, insightful advice, encouragement and care throughout my graduate study. I would not be able to complete my dissertation without his guidance.

I would like to extend my sincerest thanks and appreciation to my previous and current thesis supervisory committee: Dr. John Mittler, Dr. Ram Samudrala, Dr. Roger Bumgarner, Dr. Shiu-Lok Hu and Dr. Jaisri Lingappa, for their continuous supports, useful advice and discussion.

I would like to express my appreciation to all members of Mullins lab, past and present, for their constant supports. In particular, I thank Morgane Rolland and Robert Smith who have taught me many invaluable lessons in conducting good scientific research. I thank Victor Swain, Mary Campbell, Wenjie Deng, Hong Zhao, Kim Wong and Lennie Chen for their kindness and willingness to spend their times training, discussing and helping me with any problems that I encountered.

Last but not least, I would like to thank all my friends who made my life in the US full of wonderful and memorable experiences.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

Acquired immunodeficiency syndrome (AIDS), a disease caused by infection with Human immunodeficiency virus (HIV), is a major global health burden affecting more than 35 million people. At present, there is no cure treatment, other than expensive bone marrow transplants with CCR5Δ32 donor cells, and a vaccine remains elusive. Rapid viral evolution and high sequence diversity are major obstacles for the development of effective vaccines and therapies, as the virus can readily acquire mutations to escapes these pressures. One strategy to overcome this problem is to focus drugs and vaccines on conserved elements of the viral proteome; i.e., on relatively immutable regions hypothesized to be important for viral replication. However, identifying conserved regions based on sequence conservation is challenging as HIV sequence diversity continues to increase and the sequence conservation of specific residues have been shown to change over the course of HIV epidemic.

Protein evolution is constrained by the need to maintain structural forms that mediate function. Using information obtained from high-resolution three-dimensional structures of the HIV-1 capsid protein, we identified regions in which mutations are likely to alter function or destabilize the protein and capsid structure. The capsid protein is a target for host cellular immune responses that have been associated with viral control in both human and non-human primates. It is also an emerging target for novel HIV drugs.

1.1 HIV

HIV is a member of the Lentivirus genus of the family Retroviridae, characterized by a long incubation period for the development of disease. There are two major types of HIV: HIV-1 and HIV-2. HIV-1 is classified into four groups: M, N, O and P. HIV-2 consists

of eight groups, A-H. HIV-1 group M is the most common type of HIV and is responsible for more than 90% of HIV infections worldwide. Group N, O and P viruses are extremely rare and mostly confined to Cameroon, Gabon and neighboring countries. HIV-2 prevalence rates are much lower than that of HIV-1, and HIV-2 is largely restricted to West Africa. HIV-1 group M is further classified into nine subtypes: A-D, F-H, J and K, and more than 40 circulating recombinant forms (CRFs). A CRF is generated when multiple subtypes infected the same host and viral genomes recombine, with the recombinant then spreading to others [1]. Unless specified otherwise, the rest of this dissertation is related to HIV-1 group M only.

HIV-1 biology is a very active area of ongoing research. The following subsections provide a brief overview of our current understandings of the virus, including virion structure, genome organization, viral protein functions and the replication process. A separate subsection is dedicated to the discussion of the viral capsid protein structure and functions in detail, as it is the primary focus of this thesis.

1.1.1 HIV-1 virion structure

HIV-1 is a spherical enveloped retrovirus composed of two copies of the positive single-stranded RNA genome. The genomic RNA (gRNA) is tightly bound with viral nucleocapsid (NC) proteins. The genome along with other host and viral proteins and enzymes are enclosed in a protein shell, or capsid, made up of about 2,000 copies of viral capsid (CA) protein. The capsid is surrounded by the matrix proteins (MA) and, then, the viral envelope, which is a phospholipid bilayer derived from the host cell membrane when the virus buds from the cell. The viral envelope contains various host cell surface proteins as well as multiple copies of a viral glycoprotein complex that protrudes through the surface of viral particles. The protein complex consists of a cap made of three

glycoprotein 120 (GP120) molecules and a stem made of three glycoprotein 41 (GP41) molecules that anchor the protein complex into the viral envelope (Figure 1.1).

1.1.2 HIV-1 genome

HIV-1 genome is about 9.7 kilobases (Kb) and comprises nine genes: *gag, pol, env, tat, rev, vif, nef, vpr and vpu*, along with two untranslated regulatory regions (UTR), at the 5' and 3' end. The viral genome contains multiple overlapping open reading frames and several splice donor and acceptor sites. These features allow the generation of the multiple transcripts needed to obtain nine different gene products [2] (Figure 1.2). The HIV-1 genome exists in two forms: single-stranded (SS) RNA in the virion, and double-stranded (DS) "proviral" DNA integrated into the host cell genome. The HIV-1 RNA genome contains several regulatory regions including the trans-activating response (TAR) element, a polyadenylation signal (Poly A), a genome packaging signal or psi (Ψ) element (PE), a primer binding site (PBS), a dimer initiation site (DIS), a rev-response element (RRE), and a slippery site (SLIP) or ribosomal frame-shift signal (Figure 1.2). Compared to the RNA genome, the proviral DNA contains an extra copy of non-coding U5 and U3 element, generated during reverse transcription, in the 5' and 3' end of the genome, respectively. The region encompassing U3, R and U5 is called long terminal repeat (LTR). There are two copies of LTR, one at the 5' and the other at the 3' ends of the proviral genome, and hence, are called the 5' LTR and 3' LTR, respectively (Figure 1.2).

The TAR makes up the first 45 or so nucleotides of the viral transcript. It forms a stable stem-loop structure and acts as a binding site for the Tat protein as well as host cell factors required for efficient transcription elongation [3, 4]. The PBS is an 18-nucleotide sequence, located right after the U5 region, that serves as the binding site of cellular tRNA(3Lys), which the virus uses as the primer for reverse transcribing the SS gRNA to

DS DNA [5]. The PE is a set of four stem-loop structures located upstream of the Gag coding region. The PE is recognized by the NC protein and required for the packaging of viral genomic RNA (gRNA) into nascent viral particles [6, 7]. The first stem loop of the PE also contains the DIS element, which promotes the dimerization of SS gRNA [8]. RNA dimerization and genome packaging are tightly linked processes [9], and [10] have proposed that genome packaging is regulated by dimeric RNA recognition.

The SLIP sequence, UUUUUUA, followed by a stem-loop structure located at the 3' end of the *gag* mRNA occasionally induces a -1 translational frameshift that allows the synthesis of the Gag-Pol polyprotein instead of only the Gag polyprotein. Frameshifting is estimated to occur at the frequency of 5%-10% in HIV-1 (Figure 1.3) [11, 12]. The RRE is a 234 nucleotide, highly structured RNA element located within the Env coding region. It contains seven high-affinity binding sites for the Rev protein and only exists in unspliced or singly spliced transcripts. The binding of multiple Rev molecules is required for nuclear exports of unpliced or partially spliced RNA and, hence, the translation of these transcripts [13-15].

1.1.3 HIV-1 genes and proteins

Three genes – *gag, pol, and env* - encode enzymes and structural proteins needed to make a new viral particle. The *gag* gene is transcribed and translated into a polyprotein Gag, which plays a major role in the assembly of a new viral particle. Upon budding from the host cell, the Gag polyprotein is cleaved by the viral protease into four proteins: matrix (MA), capsid (CA), nucleocapsid (NC) and p6, and two spacer peptides - SP1 and SP2 [16]. The MA protein is located at the amino-terminal domain (NTD) of the Gag polyprotein. Its primary functions are targeting the Gag proteins to the plasma membrane, the virion assembly site, and incorporating the envelope glycoproteins into nascent virions [17, 18]. In the mature virion, after protease cleavage, the MA protein

remains associated with the viral envelope. Following the MA protein is the CA protein, which forms the conical mature capsid enclosing viral gRNA and various other proteins. The structure and functions of the CA protein are described later in this chapter, as it is the primary focus of this dissertation.

Following the CA protein and preceding the NC protein is a stretch of 14 amino acids (AA) called the SP1 region. SP1, along with the carboxy-terminal domain (CTD) of the CA protein, serve as the major region of Gag-Gag interaction in viral immature particle assembly [19, 20]. The presence and sequence of the SP1 peptide are crucial to the ability of Gag to assemble into normal viral particles [21-23]. The NC is a nucleic acid (NA) chaperone. It contains two '$CX_2CX_4HX_4C$' zinc finger (ZF) motifs that recognize the PE, or genome-packaging signal, and are indispensable for the selective packaging of two copies of full-length viral gRNA [24]. The NC protein also has a non-specific NA binding activity. It promotes RNA dimerization, condensation of RNA genome in the mature particle, annealing of the appropriate cellular tRNA primer to the PBS, and 5' to 3' DNA strand transfer efficiency during reverse transcription (step 3 in Figure 1.4) [25-28].

Following the NC protein is the SP2 peptide and, lastly, the p6 protein. The functional role of the SP2 region is not well established. A recent study suggested that while the cleavage of the SP2 peptide from the NC and the p6 protein is essential to the proper viral capsid maturation, the SP2 itself might be dispensable for structural maturation and viral infectivity [29]. The p6 protein contains two functional motifs, P_7TAP and $LY_{36}PX_2S_{40}L$, designated as Late (L) domain-1 and L domain-2, respectively. Both motifs serve as docking sites for cellular factors needed for viral particle budding [30-32]. L domain-2 is also the binding site of the Vpr protein and is required for the incorporation of Vpr into nascent virions [33, 34].

The *pol* gene encodes the viral enzymes protease (PR), reverse transcriptase (RT), and integrase (IN). The Pol precursor protein is translated as part of a Gag-Pol polyprotein, generated when -1 translational frameshifting occurs at the 3' end of the Gag coding region (Figure 1.3), and is later processed by the PR to yield individual enzymes. PR cleavage of Gag and Gap-Pol polyproteins is critical to the viral maturation process [35]. Proviral DNA lacking functional PR produces immature, noninfectious viral particles [36]. The RT possesses two enzymatic activities: (1) a DNA polymerase that can use either RNA or DNA as a template and (2) a ribonuclease (RNase) H that cleaves RNA when it is a part of an RNA/DNA duplex. This enables conversion of the SS viral RNA genome into the DS DNA that can be integrated into the host chromosome by the IN. IN catalyzes two reactions: (1) 3'-processing, the cleavage of two nucleotides from the 3' ends of viral DNA, and (2) strand transfer, or the insertion and covalent ligation of the 3'-processed DNA into the host chromosomal DNA [37].

The *env* gene codes for a polyprotein (Env or gp160) that is later processed by a host cell protease, Furin, into a non-covalent complex of the surface (SU) glycoprotein GP120 and the transmembrane (TM) glycoprotein GP41 [38]. GP120-GP41 proteins are associated as a heterotrimeric spike on the host cell surface and viral envelope. GP120 contains the binding sites for the CD4 receptor and the chemokine co-receptors [39, 40], which the virus uses to gain entry into the host cell. GP41 is necessary for mediating fusion between viral and host cell membranes [41-43].

The regulatory genes, *tat* and *rev*, encode for proteins that control HIV-1 gene expression. The first protein, Tat (Trans-Activator of Transcription), binds to the TAR, a RNA stem-loop structure, formed by the 5' end of nascent HIV-1 transcripts. The binding of Tat recruits and activates a host protein called positive transcription elongation complex (P-TEFb). After a transcription initiation, host RNA polymerase II (Pol II)

becomes trapped in HIV promoter proximal region. P-TEFb is a cyclin dependent kinase that phosphorylates Pol II and causes transition into productive elongation leading to production of full length viral mRNA [44]. The second protein, Rev (Regulator of Expression of Virion proteins), acts as a nuclear shuttling protein to facilitate the export of unspliced or singly spliced viral transcripts, which contain introns, from the nucleus to the cytoplasm. In the absence of Rev, the host RNA splicing machinery removes all introns from viral mRNAs. Only Tat, Rev and Nef proteins can be produced from completely spliced RNAs. The unspliced and singly spliced transcripts contain multiple RREs. Rev contains a leucine-rich nuclear export signal (NES) near its carboxy-terminal end. The binding of multiple Rev proteins to RRE-containing transcripts induces nuclear export of the Rev-RNA complex. The exported unspliced or singly-spliced transcripts are then translated into the rest of viral proteins; Gag, Pol, Env, Vif, Vpr and Vpu [45].

The regulatory genes, *vif, vpr, vpu,* and *nef,* encode proteins that are essential to the virus' ability to infect, replicate and cause disease. Vif (Viral infectivity factor) counteracts the antiviral activity of a host restriction factor called APOBEC3G (A3G). In the absence of Vif, A3G is packaged into HIV-1 virions. During reverse transcription, A3G, a cytidine deaminase, induces guanosine-to-adenosine (G->A) hypermutation in nascent viral cDNA, resulting in a debilitating loss of genetic integrity. Vif binds and targets A3G to ubiquitination and protein degradation pathways and, hence, prevents the inclusion of A3G into new viral particles [46]. Vpu (Viral Protein U) is a membrane-associated protein that has been associated with two primary functions. First, Vpu interacts with newly synthesized CD4, the primary cell surface receptor for HIV, in the endoplasmic reticulum (ER) and targets CD4 to the ubiquitination and protein degradation pathway [47]. The second function of Vpu is to enhance virion release from the infected cell by antagonizing a host restriction factor called Tetherin. Tetherin is a

transmembrane protein that traps the newly assembled HIV-1 particles at the cell surface and, hence, prevents virion budding and release. While the presence of Vpu has been clearly shown to interfere with Tetherin functions, the exact molecular mechanism of the antagonism remains unclear [46].

Vpr (Viral protein R) has several proposed functions, including nuclear targeting of pre-integration complexes, inducing cell cycle arrest and apoptosis, modulation of reverse transcription accuracy, and transcriptional control of cellular genes. It directly interacts with the p6 region of the Gag polyprotein and is incorporated into virions [48]. Nef (Negative Factor) was shown to be dispensable for HIV replication *in vitro*, but essential for viral spread and disease progression *in vivo* [49, 50]. Nef's primary function is the downregulation of CD4 and major histocompatibility complex class I (MHC-I) expression on the host cell surface. Nef interacts in clathrin-dependent trafficking pathways to direct MHC-1 and CD4 receptor molecules from the cell surface to endolysosomal pathway where the proteins are degraded at an accelerated rate [51, 52].

1.1.4 HIV-1 replication cycle

HIV-1 gains entry into the host cell via interactions between GP120 and CD4 molecules on the surface of the host cell [53-55]. GP120 first binds with the CD4 molecule. This binding triggers a structural change in the GP120-GP41 complex that enables GP120 to also interact with the coreceptors, CCR5 and CXCR4. This two-step attachment allows GP41 proteins to penetrate the cell membrane and induce viral-cell membrane fusion leading to the release of the viral capsid into cell cytoplasm [56-58] After entry, a series of events occurs that leads to the host chromosomal integration of the viral genomic DNA. The timing and molecular mechanisms of these events are not yet completely understood. The current model posits that the capsid starts to disassociate shortly post-entry. The viral reverse transcriptase then initiates the synthesis of DS DNA.

The nascent full-length HIV-1 DNA, along with viral integrase enzyme, is then imported into the cell nucleus through nuclear pore complexes [59]. The capsid disassembly process, or uncoating, is likely to be a multi-step process and tightly coupled with the reverse transcription and trafficking of the viral DNA-protein complexes, or pre-integration complexes (PICs), to the nucleus [60-64] The essential components and steps of reverse transcription are well-established (Figure 1.4) [65]. However, the factors and mechanisms involved in coordinating productive synthesis and successful delivery of the viral DNA into the nucleus remain unclear.

After nuclear entry, viral DNA is inserted into the host genome through the action of the viral Integrase. First, two nucleotides are cleaved from each 3' end of the viral linear DS DNA generating a reactive intermediate containing 3' hydroxyl group. These 3' ends attack a pair of phosphodiester bonds on opposite strand of chromosomal DNA. The 3' end of viral DNA and the 5' end of host DNA are then covalently joined. The attacks usually occur about 4-6 bps apart creating unpaired regions of DNA, or DNA gaps, that are repaired by cellular enzymes. The polymerase enzyme fills in the gaps. Nuclease enzymes remove dinucleotide overhangs at 5' ends of the viral DNA. Lastly, the DNA ligase enzyme joins 5' ends of viral DNA to the host DNA (Figure 1.5). Post integration, HIV-1 may enter a latent stage of infection, wherein the viral genes are not expressed at detectable levels. At present, our understanding of the molecular mechanisms and factors underlying HIV-1 latency remains incomplete. Nevertheless, latently infected cells are relatively rare *in vivo* [66]. In productive infection, HIV-1 proviral DNA is transcribed and spliced into small intron-free mRNAs, encoding the Tat, Rev and Nef proteins. These small mRNAs are exported to the cytoplasm and translated into proteins, which are then transported back into the nucleus. Inside the nucleus, the Tat protein functions as a transcriptional enhancer and significantly increases the rate of

viral RNA synthesis. The Rev proteins binds with unspliced and singly spliced viral mRNAs, enabling these intron-containing transcripts to be transported out of the cell nucleus into the cytoplasm. Unspliced mRNA encodes structural protein and enzyme precursor, Gag and Gag-Pol. It also functions as the viral gRNA. Env glycoproteins and the other regulatory proteins - Vif, Vpr and Vpu - are synthesized from the same singly spliced transcript [45].

Following protein synthesis, the viral structural proteins – Gag, Gag-Pol and Env – are trafficked to cell plasma membrane (PM), albeit by different mechanism [67], where assembly of new virions takes place. Gag plays a primary role in particle formation with different structural domains interacting with viral gRNA, viral proteins and host cell factors. Gag/Gag-Pol polyproteins interact with each other, mostly via CA-SP1 region, to form a spherical particle [20, 68]. A fatty acid group, myristic acid, is added post-translationally to the NTD of Gag, or the MA, and this myristoylated domain facilitates association of Gag/Gag-Pol polyproteins with phospholipid PI(4,5)P2 molecules located within cytoplasmic leaflet of the PM [69, 70]. The MA also interacts with the TM domain of Env and promotes Env incorporation into the virion [71, 72]. The NC domain contains zinc finger motifs that recognize the genome packaging signal (PE), located within the 5' UTR end of the viral genome, and is responsible for recruiting the viral gRNA dimer to the forming particle [24]. Efficient HIV-1 genome packaging requires RNA dimerization [10]. Study of murine leukemia virus suggests that NC binding sites are usually sequestered and only become exposed upon RNA dimerization [73]. In addition to the RNA genome, cellular tRNA(3Lys) and viral proteins – Pol, Vif, Vpr, Nef, p7 – are packaged, as well as several other host small RNAs and proteins [74-76]. However, it is still unclear whether these host RNA and proteins are specifically recruited and have functional roles in the viral life cycle.

The last steps of the HIV-1 replication cycle are the budding and maturation of new viral particles. HIV-1 hijacks host endosomal sorting complexes required for transport (ESCRT pathway) to terminate Gag polymerization and catalyze membrane fission reactions that release viral particles from the host cell membrane [30, 77]. Upon budding and release, viral PR autocatalytically cleaves Gag and Gag-Pol polyproteins, yielding the fully processed MA, CA, NC, p6, PR, RT, and IN proteins [78-80]. After cleavages, the MA remains anchored to the viral envelope, while CA proteins interact with each other forming a conical capsid. This proteolytic cleavage is essential for viral infectivity. Inhibition of this process, either by deleterious mutations in PR or at the cleavage site, or use of a PR inhibitor, results in an immature, noninfectious virions that retain unprocessed Gag and Gap-Pol polyproteins, and maintenance of a spherical shape [17, 35, 81-83].

1.2 HIV-1 capsid and capsid protein (CA)

Viral capsid is a protein shell that encloses the viral genetic material and proteins necessary for the next round of viral infection and replication. HIV-1 capsid exists in two stages: an immature and a mature particle. The initial, or immature, capsid formed inside a host cell has a spherical shape and is made up of the entire Gag polyprotein. The capsid assembly process is a highly orchestrated event involving Gag proteins, viral gRNA, lipid membrane, and host cell factors [84]. Based on cryo-electron microscopy and tomography structures of immature HIV-1 and Mason-Pfizer monkey viruses, CA provides key protein-protein interactions that stabilize the immature particle [85, 86].

The capsid changes into a conical shape upon protease cleavage of Gag, which occurs simultaneously with or shortly after HIV-1 buds from host cells. The mature capsid consists only of CA [84]. Capsid maturation is essential to HIV-1 infectivity [17, 35, 81-83],

as rearrangement of the capsid structure is required for proper disassembly of capsid in a newly infected cell. Upon entering the new host cell, the capsid begins to disintegrate, following by the initiation of reverse transcription process and the trafficking of PICs to the cell nucleus. Once thought of as a spontaneous process, recent studies have shown that the timing and the rate of HIV-1 capsid disassembly are crucial to the success of downstream processes and tightly regulated [64]. The host factors cyclophilin A (CypA) and the PDZ domain-containing protein 8 (PDZD8) have been shown to help stabilize the capsid and prevent premature uncoating. Knocking down PDZ8 accelerates the disassembly of HIV-1 capsids in infected cells and results in decreased reverse transcription [87, 88]. In contrast, the nonhuman primate restriction factors rhesus macaque (rh) TRIM5α and TRIMCyp, prevent infection of HIV-1 and other retroviruses by binding to incoming viral capsids, accelerating capsid disassociation and impeding reverse transcription [89-93]. Besides the interactions with host cell factors, mature capsid stability also influences reverse transcription and viral infectivity. Several CA mutations that alter *in vitro* capsid stability reduce the formation of stable reverse transcripts and abolish infectivity [63, 94, 95].

The HIV-1 CA consists of 231 amino acids. Despite the high rate of HIV-1 evolution, the CA protein remains highly conserved, with two-thirds of the protein having consensus amino acid frequencies of 0.9 or greater [96]. HIV-1 CA plays important structural and functional roles both in the early and the late steps of viral replication. Small interfering RNA (siRNA) screening and knockdown studies have identified several host cell factors that interact with HIV-1 CA and affect not only the completion of capsid disassembly but also reverse transcription, nuclear import and integration of viral DNA [62, 97-101]. In the late stage of viral replication, CA-CA interactions as part of the Gag polyprotein are also important for the assembly of immature capsid [20, 68, 85]. Recently

obtained high-resolution three-dimensional structures of mature HIV-1 capsid assembly units have provided insight into residues involved in protein-protein interactions that stabilize the mature capsid [102]. However, less is known about the assembly of the immature capsid, as its high-resolution structure has not been elucidated. Hence, only structural details of the mature capsid assembly unit is described in this thesis.

1.2.1 Mature capsid structure

The mature HIV-1 capsid is a fullerene cone, a variably curved hexagonal assembly consisting of approximately 250 hexamers and 12 pentamers of CA proteins [103, 104]. Each CA protein consists of two independently folded domains – the amino-terminal domain (NTD) and the carboxyl-terminal domain (CTD). These two domains are joined by a short unfolded region commonly referred to as flexible linker [105]. The three-dimensional structures of the NTD and the CTD of the HIV-1 CA have been solved, providing atomic-level resolution detail of each structural domain [105-107]. The NTD is composed of a β-hairpin, seven α-helices (helix 1 through helix 7) and an extended loop region between the fourth and fifth helix. The CTD contains a short 3_{10} helix, a loop and four α-helices (helix 8 to helix 11) (Figure 1.6A). The 3_{10} helix has 3 residues (10 atoms) per turn while the α-helices have 3.6 residues (13 atoms) per turn.

A high-resolution crystal structure of the CA hexamer [102] has shown that there are two main interfaces between each CA unit within the hexamer – the NTD-NTD and the NTD-CTD interfaces. As shown in Figure 1.6B, the NTD-NTD interface is between residues on helix 2 of the first CA and helix 1 and 3 of the neighboring CA. Both hydrophobic and hydrophilic contacts are observed in this interface. The NTD-CTD interfaces, between helices 4 and 7 and helices 8 and 11, respectively, also consists of both hydrophobic and hydrophilic contacts [102]. In addition to the NTD-NTD and NTD-

CTD interfaces within hexamer,s CA proteins of neighboring hexamers also interact via the CTD-CTD interface.

Worthylake et al. [108] and Rose et al. [109] have shown that the CA-CTD forms soluble protein dimers. Several high-resolution structures of CTD dimers have been solved [105, 106, 108, 110, 111]. From these structures, four different configurations of CTD dimers were observed. Three of the four contained pairwise parallel packing of helix 9 from each monomer (Figure 1.6C) with different helix 9 crossing angles [112]. The other distinct configuration, the so-called domain-swapped dimer, was observed in a mutant CA-CTD that contains a single amino acid deletion, Δ177 [111]. In this domain-swapped form, the 3_{10}-helix and helix 8 is exchanged between two monomers (see Figure 1.6D).

Recently, an electron cryocrystallography map of mature HIV-1 capsid-like particles was solved. Interestingly, none of the currently available atomic structures of the CTD dimer can be perfectly fitted onto this map. Although one of the parallel packing form dimers fit quite well, the best fit was achieved by fitting CTD monomers into the map individually. This observation led to the suggestion that small conformational changes might occur during mature capsid assembly [112]. Hence, the exact configuration of CTD-CTD interface in HIV-1 mature capsid remains to be established.

Its structural and functional significance, as well as its high sequence conservation, makes the viral capsid and CA attractive targets for HIV-1 drug and vaccine. There are currently 26 FDA-approved drugs for the treatment of HIV-1, which can be classified into six distinct classes based on their target sites [113]. None of the current drugs target HIV-1 capsid or CA. While these drugs are generally effective, several factors including poor adherence, toxicity associated with long-term treatment, and drug resistance can limit their efficacy. A new class of HIV-1 inhibitor is desirable, as it will expand potential treatment options. Several studies have identified potential drug

candidates that efficiently inhibited HIV-1 replication by interfering with CA-CA interactions during one or more of these processes: virus assembly, maturation or disassembly [114].

Cytotoxic T-lymphocyte (CTL) mediated immune responses against peptides derived from HIV-1 CA have been associated with control of viral replication in infected individuals [115, 116]. In particular, two highly conserved CTL epitopes derived from CA, TW10 and KK10, were shown to induce strong immune responses and reversion of escape mutations could be readily detected during the acute-phase of infection [117, 118]. In both cases, viruses containing the escape mutation were shown to have lower replication fitness when compared to the otherwise autologous strain [119-121].

Identification of additional critical residues in HIV-1 CA and their fitness impacts may facilitate the development of new antiretroviral therapies and vaccine immunogens. While details from protein structures provide clues to potentially crucial residues and protein-protein interactions, the fitness costs of mutations at a given site are difficult to decipher from the structure alone. An experimental test is required to obtain a reliable estimate. As a part of this thesis, amino acid sequences of HIV-1 CA derived from clinical isolates were gathered and applied to determine mutation frequencies at each site. The fitness effects of the most frequently observed mutations at CA-CA interacting sites were then studied *in vitro* and analyzed based on the high-resolution structure of the mature capsid assembly unit.

1.3 Viral replication fitness assays

Viral replication fitness is defined as the capacity of a virus to produce infectious progeny in a given environment [122]. By definition, viral variants with high levels of fitness have a selective advantage over less-fit variants. In HIV-1 infection, viral replication fitness is postulated to influence disease progression and clinical outcomes [123-126].

Studies have found that HIV-1 controllers, individuals who maintain low viral load and do not progress to AIDS despite not receiving treatments, are infected by unfit viruses [127-130]. In addition, HIV-1 replication fitness of drug resistant mutations positively correlated with the likelihood of the mutations emerging during treatment [131-136]. Viral fitness is also associated with detection of immune escape variants emerging during early and chronic phases of infection [137-139].

Various *in vitro* and *ex vivo* assays have been developed to study fitness of HIV-1 variants. The nature of these assays differ in many aspects, including, but not limited to, the number of viral replication cycles studied, the biological markers used to measure and compare viral growth and the type of host cells used in cell culture [140, 141]. Among the various different types of assays, viral growth competitions or dual infection assays, in which two viral strains are competed directly in the same cell culture over multiple rounds of viral replication, provide the most sensitive measurement of fitness difference between two viral strains [122, 123, 142].

Several experimental protocols have been proposed for conducting pairwise growth competition fitness assays [143-146]. These protocols employ different experimental parameters such as host cell type, multiplicity of infection (MOI), sampling frequency, duration of cell culture, and time points used to calculate viral replication fitness. The effects of these parameters on assay outcome have not been studied until recently [147, 148]. Human T lymphocyte cell lines are often employed in competition assays due to their consistency compared with the primary cells. However, there are concerns that viral replication fitness determined using cell lines might be different from that in the primary cells. A comparative study of more than twenty HIV-1 mutant viruses detected no significant differences between the replication fitness determined using a T lymphocyte cell line and peripheral blood mononuclear cells (PBMC) [147]. Viral input ratio and sampling frequency also did not show significant effects on the fitness outcome. Viral

input MOIs, on the other hand, was found to impact fitness measurements. Higher variability between experimental replicates was observed at lower input MOI, i.e., 0.0001, compared with higher MOIs, i.e., 0.001 and 0.005. Nevertheless, the use of MOIs higher than 0.005 was not recommended as it increases the risk of recombination between two competing viral strains. Higher MOIs also shorten the exponential growth phase and limited the number of sampling times available [148].

In addition to experimental conditions, readout methods used to compare viral growth and relative fitness may affect end results. Viral CA production, RT activity and the number of infected cells are often used to measure viral growth [141]. However, these methods are not applicable to growth competition assays, as different viral strains cannot easily be distinguished. On the other hand, sequence analysis of proviral DNA or viral RNA [119, 149] and heteroduplex tracking assays (HTAs) [123] can be used to determine the relative amounts of the competing strains but not viral growth. Allele-specific real-time polymerase chain reaction (RT-PCR) [150] allows the relative amounts of viral strains as well as viral growth rates to be quantified simultaneously. One disadvantage of this method is that the genetic variation surrounding primer and probe target sequences can influence the binding efficiency and, hence, the performance of this method. Thus, the assay needs to be optimized and validated for each codon or mutation of interest.

For this thesis, I delineated an optimized protocol for pairwise growth competition fitness assay. The experimental parameters as well as the readout methods were optimized and chosen based on results from previous studies [147, 148, 151] and studies conducted within the scope of the thesis [152]. The results obtained using two different readout methods, RT-qPCR [151] and relative peak-heights on sequence chromatograms [147, 148, 153, 154], were compared and reported.

1.4 Structural-based computational approaches for predicting effect of mutations in HIV-1 CA on viral replication

Functionally and/or structurally important HIV-1 CA residues with low tolerance for mutations are great potential targets for HIV-1 vaccines and therapies. Effects of amino acid changes on protein function and organism phenotype, i.e., replication fitness, are usually evaluated using experimental techniques. However, it is not resource efficient to assess the impact of all possible amino acid substitutions experimentally. A high-throughput approach that can facilitate identification of these crucial residues is desirable. For this purpose, several computational methods have been developed to forecast effects of mutations on protein function, structural stability and phenotype [155-158]. However, these computational tools were mostly developed using information from human proteins and associated phenotypes or clinical outcome data. It is not known whether these methods perform as well in predicting impact of mutations on HIV-1 protein function and viral replication until recently.

Williams et al. [159] and Humphris-Narayanan et al. [160] applied protein structural modeling and scoring functions to predict tolerated mutations, which appeared in more than 1% of the sequences studied, in HIV-1 RT, PR and MA with varying degrees of success. Several factors were shown to impact the prediction accuracy, including the origin of the template structure, experimentally determined versus computational models, the number of template models, as well as the prediction algorithms. In addition to tolerated mutations, the impact of a mutation on RT activity and PR function has been associated with the predicted structural interaction between the protein and its substrates [160]. Existing high-resolution structures of the CA monomer and multimers enable the use of structural-based computational approaches to study the impact of mutations within the core of the protein and at the CA-CA interface on CA hexamer

stability. Optimal capsid stability has also been shown to be important for successful viral replication [63]. Mutations that alter the individual CA or the CA hexamer are likely to have a significant negative impact on overall capsid stability. A study of 29 single mutations in HIV-1 Gag showed a correlation between predicted changes in protein stability and viral fitness [161]. A more comprehensive study would yield further insight into the feasibility of this approach.

Computational methods for predicting protein stability can be classified into two categories: physics-based and knowledge-based. Physics-based methods use an energy function, which uses basic physical estimates of molecular interactions such as long-range electrostatic and van der Waals forces, to estimate the thermodynamic stability of a given protein structure. Knowledge-based methods are based on statistical analyses of known protein sequences, structures, and experimentally derived stability data [162]. Some methods combine both approaches together. For example, CC/PBSA and FOLDEF are energy functions that were optimized to reproduce experimentally derived protein free energies and binding affinities [155, 163]. Several kinds of experimental data have been used to optimize energy functions including native protein sequence, native protein structure, and experimentally determined protein stabilities [155, 162, 164-166]. The change in stability following a single amino acid change is calculated from the difference stability between the wild type and mutant proteins. In this thesis, the feasibility of applying protein stability predictions and *in silico* modeling of single amino acid mutations in HIV-1 CA to predict the impact of mutations on viral replication is evaluated. This approach has potential to be used as a screening step for selecting mutations for further study.

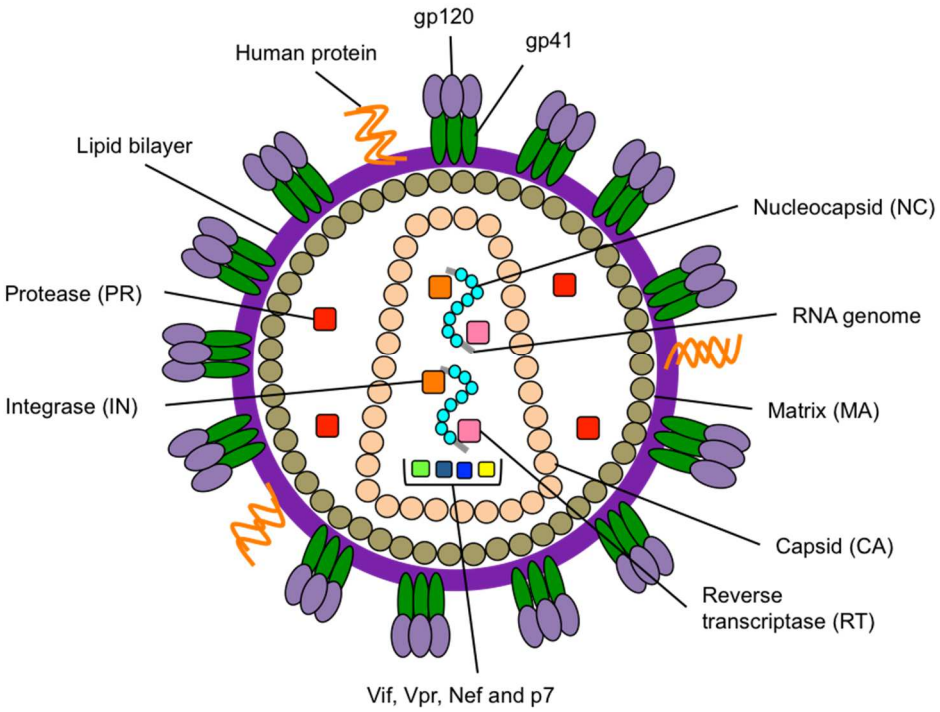

Figure 1.1. Diagram of HIV-1 virion

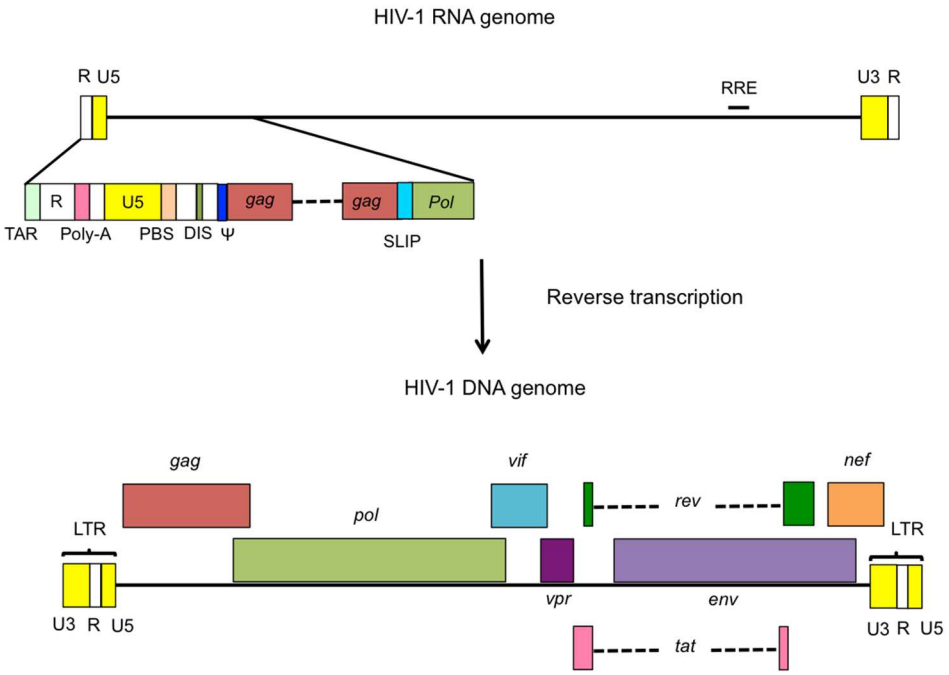

Figure 1.2. Diagram of HIV-1 RNA and proviral DNA genome

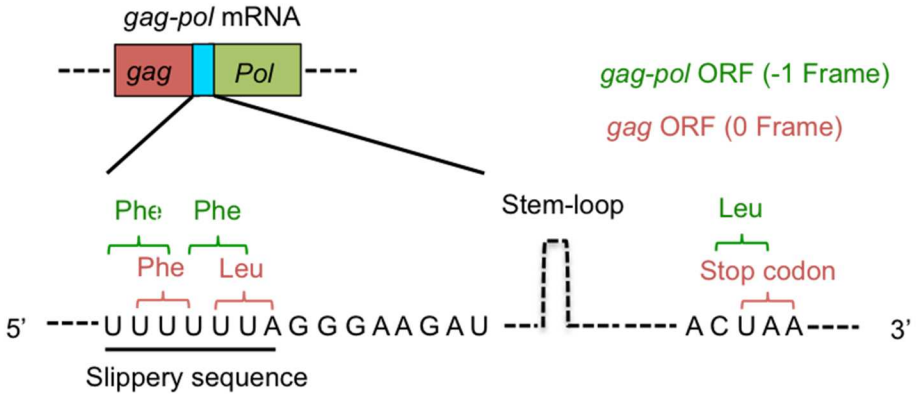

Figure 1.3. HIV-1 translational frameshift signal

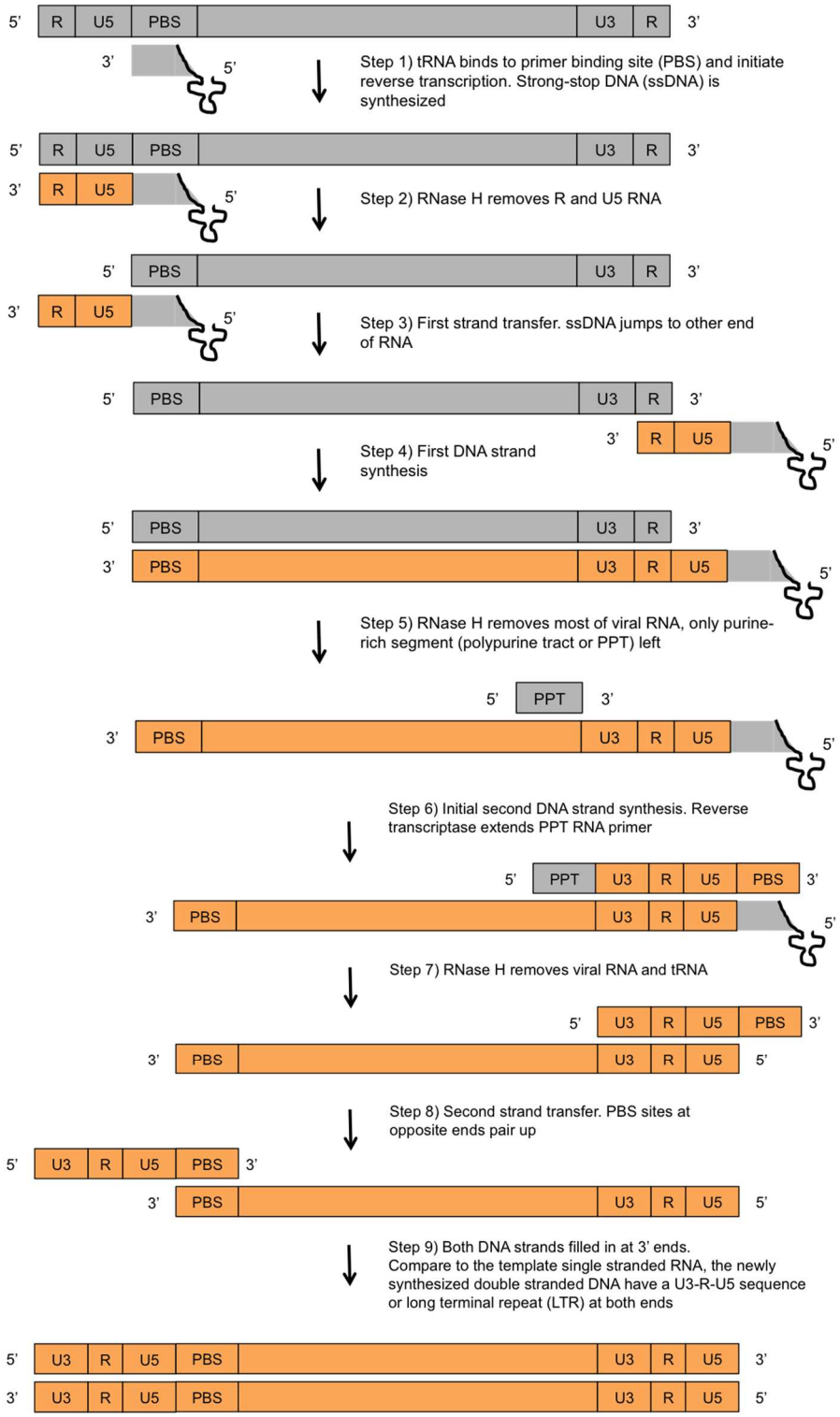

Figure 1.4. HIV-1 reverse transcription process

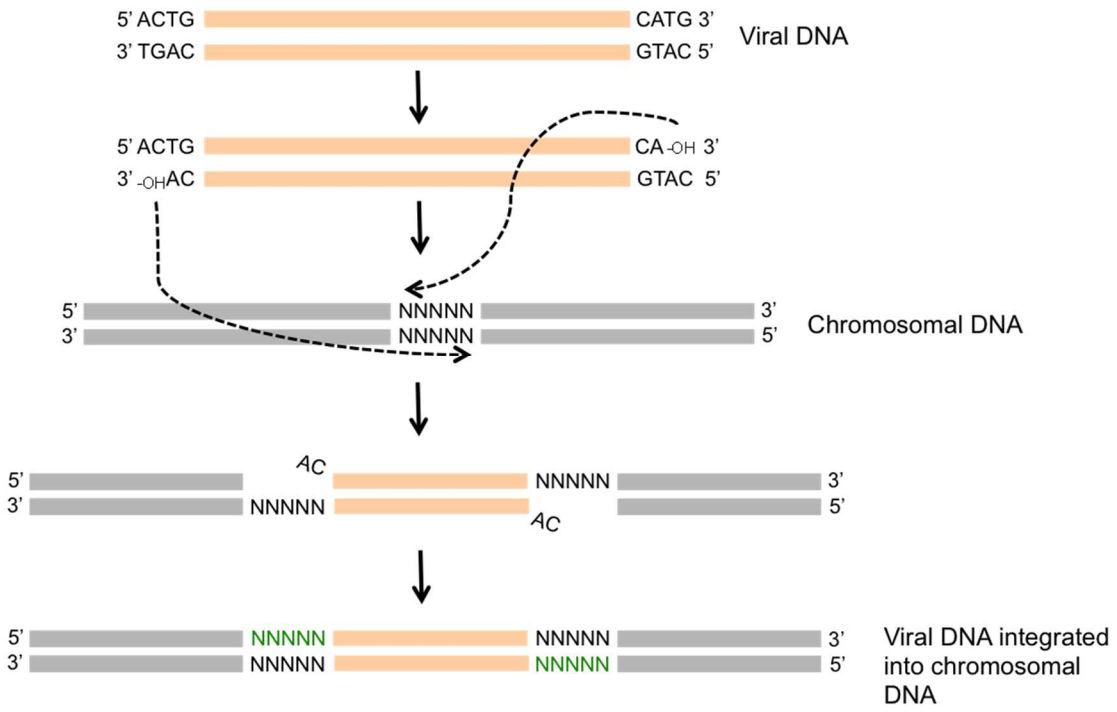

Figure 1.5 HIV-1 DNA integration into host chromosomal DNA

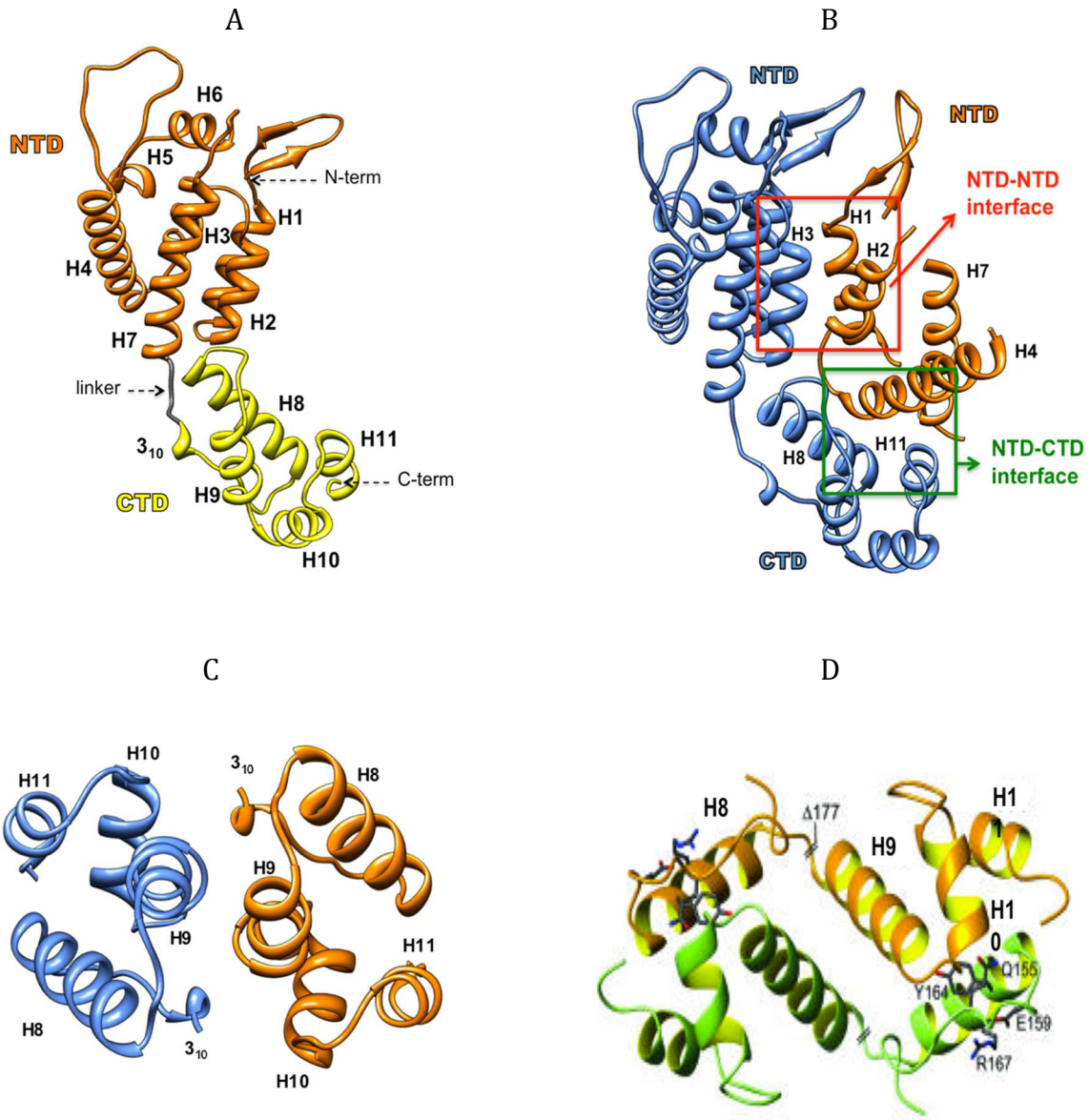

Figure 1.6. A) HIV-1 CA protein. B) NTD-NTD and NTD-CTD interface between two CA proteins C) CTD dimer in parallel packing form D) CTD dimer in domain swapped form (adapted from [111]).

CHAPTER 2

Pairwise Growth Competition Assay for Determining the Replication Fitness of Human Immunodeficiency Viruses

2.1 Introduction

Viral replication fitness is defined as the capacity of a virus to produce infectious progeny in a given environment [122] and is an important contributing factor in determining the prevalence of a virus variant at the population level over time [167]. As experimental *in vivo* fitness studies are not feasible with pathogenic human viruses, such as HIV-1, various *in vitro* and ex *vivo* replication fitness assays have been developed to study the effects on fitness arising from drug resistance and immune escape mutations, epistasis and the evolution of viral populations [143-146]. Among different fitness assays, growth competition assays are recognized to yield more sensitive and valid measures of fitness differences, as two or more viral variants compete for the same cell population under precisely the same environmental conditions, as occurs *in vivo* [122, 123, 142]. Before starting growth competition experiments, several variables need to be determined, including the use of different multiplicities of infection (MOI), viral input ratio, and timing of sampling for analysis. We have studied the effects of these parameters on viral growth kinetics and on the outcome of competition experiments, and have identified key factors necessary for robust measurements of HIV-1 fitness in cell culture [148].

In addition to assay variables, there are a variety of methods for quantitating viral variants in growth competition experiments. Bulk [149, 168] or clonal sequencing [153, 154] has been used to determine the ratio of the competing viruses based on the nucleotide frequencies at the site(s) of interest. Relative fitness is derived from changes in this ratio

over time. This method is convenient as DNA sequencing services are widely available. The parallel allele-specific sequencing (PASS) method enables sequencing at multiple sites and the detection of recombinants [169], but it also requires specifically developed reagents and detection systems. Other methods use a small reporter gene [170] or synonymous mutations [137, 144, 171, 172] as tags to distinguish the competing viruses by sequencing, heteroduplex tracking assays (HTA) [123, 137, 144, 173] or reverse transcription-quantitative real-time polymerase chain reaction (RT-qPCR) [151, 170-172], all of which can be made applicable to study competing strains regardless of sequence similarity. An additional step is required to introduce tags into viral genomes and the RT-qPCR assay also requires specific reagents and instrumentation. We have found that bulk Sanger sequencing yields comparable results [148].

Following growth competition, viral replication fitness is presented as relative fitness, or a fitness ratio between two viral variants. The relative fitness of a virus can be defined as the final proportion of a viral variant normalized by its initial proportion in the inoculum or as the net growth rate difference between the two competing viruses. We found that the latter method, using longitudinal data points only within the exponential growth phase, produced the most robust results [148, 151].

In vitro fitness assays are used primarily to study biological clones [123, 142, 145] and infectious molecular clones of HIV-1. The latter, being amenable to genetic manipulation, are often employed to study the effect on fitness from particular mutations or specific sequences of interest [128, 143, 144, 146, 174]. The following protocols describe a workflow from the point of constructing new full-length infectious HIV-1 molecular clones using HIV-1 vectors containing a sequence tag, introducing mutations of interest, making viral stocks and establishing viral growth kinetics, to performing the growth competition assay and calculating relative fitness (Figure 2.1).

Using our optimized procedures, we created three recombinant HIV-1 mutants and determined their replication fitness. The recombinant molecular clone was first constructed by replacing the HIV-1 *gag-p24* gene region of pNL4-3, a plasmid containing a full length infectious genome of HIV-1 lab strain NL4-3, with a synthetic COTB (Center-Of-Tree, subtype B) *gag-p24* sequence [147] to create the prototype strain. Single amino changes (T186M, T242N, and I256V) were then introduced to create three mutant clones. Each mutant was competed against the prototype virus to observe the fitness impact of each mutation in the given genetic background. The three mutants demonstrated varying levels of replication fitnesses, ranging from slightly to significantly lower than the prototype virus. The T242N mutation was previously reported to have a moderate fitness cost [119, 175, 176], similar to the results shown in this study. The fitness cost of the other two mutations had not been reported previously.

2.2 Protocol

The protocol, as described below, does not include any patient identifiable information and is thus not considered Human Subjects Research by the University of Washington Institutional Review Board or Human Subjects Division.

1) Construction of chimeric HIV-1 NL4-3 molecular clones

1.1) Amplify insert DNA fragment

1.1.1) Design chimeric primers. The 5' halves of both forward and reverse primers contain an HIV-1 vector sequence, at which the fragment will be inserted. In the overlap extension protocol to be employed, the 3' half of the primers must contain the end of the insert sequence (Figure 2.2). Make sure that the chimeric primer sequence retains the original open reading frames.

1.1.1.1) Use primers at least 20 bases in length, with a melting temperature greater than or equal to 60 °C, ~50% GC content, and a low tendency to form primer dimers, heterodimers and/or hairpin structures. These properties can be assessed using the OligoAnalyzer web tool (https://www.idtdna.com/analyzer/Applications/OligoAnalyzer/)

1.1.2) Use PCR [177] and the chimeric primers to amplify insert DNA (Figure 2.2). For each PCR reaction, use 1X high fidelity polymerase buffer, 0.2 mM dNTPs, 1 U high fidelity DNA polymerase, 0.5 µM of forward chimeric primer, 0.5 µM of reverse chimeric primer, and 1 pg - 10 ng of DNA sample carrying insert region. Add dH_2O to a final volume of 50 µl.

1.1.3) Set thermal cycling steps as follows: Perform an initial DNA denaturation step at 98 °C for 10 sec. Amplify with 30 cycles of DNA denaturation at 98 °C for 10 sec and DNA annealing at 3 °C above the lowest melting temperature of the two primers for 20 sec. Perform a final extension at 72 °C for 10 min. Store PCR products at 4 °C.

1.1.4) Take 5 µl of the PCR products from the previous step and run agarose gel electrophoresis [178].

1.1.4.1) Use a 0.7% agarose gel, 1X TAE buffer (40 mM Tris-acetate, 1 mM EDTA), 0.5 µg/ml ethidium bromide (EtBr) and a 1 kb ladder as the DNA size marker. Set power source voltage to 5 V/cm distance between electrodes. Stop the electrophoresis when the loading dye migrates through about 2/3 of the gel length. Visualize the gel using a gel documentation system [178].

Note: EtBr is a suspected carcinogen and must be properly disposed of, per institution regulations. Gloves should always be worn when handling gels containing EtBr. Change

to new gloves after finish handling EtBr containing material and before handling other materials or equipment to prevent cross contamination.

1.1.5) If only one DNA band with size corresponding to the desired PCR product is detected, purify the rest of the PCR product using a commercial kit such as QIAquick PCR Purification Kit according to the manufacturer's protocols.

1.1.5.1) If other, non-specific bands are also present, use the rest of the PCR product to run preparative gel electrophoresis. Use the same parameters and conditions specified in step 1.1.4. Ensure that the gel well is large enough to load ~45 µl of PCR products. Cut out the band of interest and extract the DNA from the gel using QIAquick gel extraction kit according to the manufacturer's protocols.

1.2) Introduce the insert fragment into full-length infectious HIV-1 subtype B vector (pNL4-3)

1.2.1) Use purified PCR products from step 1.1.5 as PCR primers. Use pNL4-3VifA [151] as template DNA in one PCR reaction and use pNL4-3VifB [151] as template in the other reaction (Figure 2.2). For each PCR reaction, use 1X high fidelity polymerase buffer, 0.2 mM dNTPs, 2 U high fidelity DNA polymerase, 500 ng of primer DNA and 50 ng of template DNA in a final volume of 50 µl. Set the thermal cycling parameters to: 98 °C for 30 sec, 35 cycles at 98 °C for 10 sec, 48 °C for 1 min and 72 °C for 10 min, followed by 72 °C for 10 min.

1.2.2) Add 10 U of the restriction endonuclease *Dpn*I to 50 µl of the PCR reaction and incubate at 37 °C for 1 hr to digest the template DNA. Ensure that the plasmid DNA is isolated from a methylation competent bacterial strain, e.g., TOP10 chemically competent *Escherichia coli*.

1.2.3) Use *Dpn*I digested product to transform competent bacterial cells. Use heat-shock transformation with TOP10 chemically competent *E. coli*, according to the manufacturer's protocol. To select for bacterial cells containing the recombinant plasmid, use Luria Broth (LB) culture plates containing 100 mg/L carbenicillin.

1.2.3.1) Pick ~10 well-separated colonies and grow each separately in 3 ml LB liquid medium containing 100 mg/L carbenicillin and incubate at 30 °C in a shaker overnight.

1.2.3.2) Use QIAprep Spin Miniprep kit to isolate plasmid DNA from the bacterial liquid culture, according to the manufacturer's protocol.

1.2.4) Use double restriction digestion [179] to determine whether the plasmid DNA contains the proper insert. Ensure that one of the restriction sites exists only within the insert region and the other restriction site exists only once in the HIV-1 vector, outside of the insert region.

1.2.4.1) Digest at least 300 ng of plasmid DNA in a 10 µl final reaction volume. Select restriction buffers, incubation temperature and incubation time according to the manufacturer's protocol of the selected restriction enzymes. Take 9 µl of the digested DNA and run gel electrophoresis as described in step 1.1.4. Select recombinant plasmids that have DNA bands of the predicted sizes.

1.2.5) Confirm sequence integrity of the recombinant plasmids by Sanger sequencing, as rare, unwanted mutation(s) can be introduced during PCR reactions.

1.2.5.1) Sequence both strands of the plasmid DNA. Follow instructions in step 1.1.1.1 to design sequencing primers. In addition, ensure that the forward and reverse

sequencing primers anneal at least 50 bp upstream and downstream of the insert region in the recombinant plasmid, respectively.

1.2.5.2) Submit plasmid DNA and sequencing primers to a commercial DNA sequencing service provider. Prepare DNA sample and primers as specified by the service provider.

1.2.6) Make an endotoxin-free stock of the mutated plasmid DNA using an Endotoxin free plasmid DNA kit according to manufacturer's protocol. Prepare at least 1 µg of endotoxin-free plasmid DNA for transfection in the following step.

1.3) Introduce small-scale mutations via site-directed mutagenesis

1.3.1) Design mutagenic primers with overlapping forward and reverse primers containing the desired mutation(s). Position the base(s) to be substituted, inserted, or deleted in the middle of the primers, flanked by 10-15 homologous bases. Follow instructions in step 1.1.1.1.

1.3.2) Use PCR to synthesize mutant plasmids. For each PCR reaction, use 1X high fidelity polymerase buffer, 10 mM dNTPs, 2 U of high fidelity DNA polymerase, 0.5 µM of forward and reverse mutagenic primers and 50 ng of chimeric pNL4-3VifB, from step 1.2.6, in a final volume of 50 µl. Set the thermal cycling parameters to: 98 °C for 30 sec, 25 cycles at 98 °C for 10 sec, 48 °C for 1 min and 72 °C for 10 min, followed by 72 °C for 10 min.

1.3.3) Repeat step 1.2.2 to 1.2.6.

2) Generation of viral stock using transfection

2.1) Calculate the amount of viral stock desired and plasmid DNA required. With a viral titer of 10^4 IU/ml or higher, 1.8 ml of viral stock is sufficient for two sets of growth

competition assays, including monoinfections, each done in triplicate. For a transfection done in a 6-well plate, about 1.8 ml supernatant is harvested per well. One µg of plasmid DNA is needed for each transfection done in a 6-well plate.

2.2) For each well of a 6-well plate, prepare 100 µl of transfection mixture, e.g., consisting of 1 µl X-tremeGENE 9 transfection reagent (or comparable product), 1 µg of plasmid DNA and serum-free DMEM.

2.2.1) Determine the volume of plasmid DNA needed, using 1 µg plasmid DNA per well. Ensure that the final concentration of the plasmid DNA is at least 50 ng/µl.

2.2.2) Determine how much serum-free medium (DMEM) is needed per well using the formula: Total volume of DMEM in µl = 100 µL - DNA volume in µl.

2.3) Add 10^6 HEK 293T-17 (ATCC) cells/well in 2 ml of propagation medium (DMEM + 10% fetal bovine serum (FBS)) into a 6-well plate. Incubate for 1 hr at 37 ^oC in a 5% CO_2 atmosphere. Seed as many wells as needed (determined in step 2.1).

2.4) To prepare the transfection mixture, aliquot the appropriate volume of serum-free DMEM, as calculated above, into a 1.8 ml polypropylene microcentrifuge tube, and then add the transfection reagent.

2.4.1) Pipette reagent directly into the media solution, do not add it to the plastic surface of the microcentrifuge tube. Add plasmid DNA last. Pipet up and down gently to mix the solution. Incubate for 15 min at room temperature (15 ^oC to 25 ^oC) to allow the formation of transfection complexes.

2.4.2) Add the mixture in a drop-wise manner to cells seeded in the 6-well plate. Gently shake or swirl the wells to ensure even distribution of transfection complexes.

2.4.3) Seal plates with plastic wrap.

2.5) Incubate cultures at 37 ^{o}C in a 5% CO_2 atmosphere for 48 hr.

2.6) Use a pipette to carefully collect and transfer supernatant to a 15 ml tube through a 0.22 µm filter top.

2.7) Use pipette to transfer 250 µl or more of the filtered supernatant to 1.8 ml screw cap microfuge tubes with rubber gaskets in the lids.

2.8) Store filtered supernatants at -80 ^{o}C until use.

3) Determine infectious titer of viral stocks on peripheral blood mononuclear cells (PBMCs)

3.1) Stimulate PBMCs with phytohemagglutinin (PHA). Per one viral stock, seed 2 x 10^6 PBMCs in complete Iscove's Modified Dulbecco's Medium (cIMDM; IMDM supplemented with 20 U/ml of human interleukin 2 (hIL-2), 10% fetal bovine serum and 1% penicillin/streptomycin) supplemented with PHA (1.5 µg/ml). Seed PBMCs at 2 x 10^6 cells/ml. Incubate PBMCs at 37 ^{o}C in a 5% CO_2 atmosphere for 72 hr.

3.2) Harvest PHA stimulated PBMCs. Transfer non-adherent PHA-PBMCs to a 50 ml conical tube. Spin tube at 228 x g for 10 min. Carefully remove supernatant without disrupting the cell pellet. Resuspend the cell pellet at a final concentration of 2 x 10^5 PBMCs/ml in cIMDM.

3.3) Seed 2x10^4 PHA stimulated PBMCs/well in 100 µl/well cIMDM in a round bottom 96-well plate.

3.4) Make a 1:10 dilution of the viral stock. Then, from the first diluted stock, make twelve 3-fold serial dilutions in a 96-well master plate. This dilution scheme is recommended for detecting viral titers in the range of 10^4 to 10^6 infectious unit (IU) per ml. For example, add 20 µl of virus stock to 180 µl media in the 1.5 ml tube. Mix the dilution by pipetting carefully. Then transfer 90 µl of the diluted stock into 180 µl media of the first well and mix well by pipetting. Continue dilution series by transferring 90 µl from the current well to 180 µl media in the next well eleven more times. Increase or decrease the initial dilution if titers higher than 10^6 IU/ml or lower than 10^4 IU/ml are expected, respectively.

3.5) Add 40 µl of the serially diluted viral stock from the master dilution plate to the seeded PBMCs plate (from step 3.3) in quadruplicate. Incubate plates at 37 ^oC in a 5% CO_2 atmosphere for 16-24 hr.

3.6) Carefully remove 100 µl supernatant from each well, and replace with 160 µl fresh cIMDM to a total volume of 200 µl/well. Incubate plates at 37 ^oC with 5% CO_2 atmosphere (day 1).

3.7) On days 4, 7, 10 and 13 transfer 100 µl supernatant from each well to 100 µl disruption buffer (2 % TritonX-100 in PBS), and replace with 100 µl fresh cIMDM. Store the supernatants at -20 ^oC.

3.7.1) Keep sampling and adding fresh cIMDM every three days until the titer stabilizes.

3.8) Determine the 50% tissue culture infectious dose ($TCID_{50}$) of the viral stock by p24 enzyme-linked immunosorbent assay (ELISA) using the day 7 and 13 samples as described in step 4.

3.8.1) If the $TCID_{50}$ obtained from day 13 is clearly higher than the titer from day 7, the virus stock may need a longer time to expand. Repeat the p24 ELISA using later samples until the infectious titers from two sampling time points become stable (or decrease). Stocks should be selected from samples with the highest titers.

4) **ELISA detection of HIV-1 p24 for determining viral infectious titer**

Note: The following protocol was developed using p24 antigen capture plates prepared in our laboratory [180]. Commercial HIV-1 p24 ELISA plate/kits can also be used, following the manufacturer's protocol.

4.1) Prior to working with samples, prepare working stocks of primary antibody (rabbit anti-HIV-1 SF2 p24 antiserum).

4.1.1) Thaw p24 antiserum at room temperature (RT).

4.1.2) Mix 2.5 ml glycerol with 2 ml 10% FBS in phosphate buffer saline (PBS).

4.1.3) Add 0.5 ml antiserum and mix.

4.1.4) Store 1 ml aliquots at -20 ^{o}C.

4.2) Thaw samples from step 3.7 in 37 ^{o}C incubator.

4.3) Wash the p24 capture plate 5 times with wash buffer (1x PBS with 0.05% Tween-20).

4.4) Add 50 µl/well of sample diluent (1% bovine serum albumin (BSA), 0.2% Tween-20 in RPMI-1640), then add 50 µl of sample to appropriate wells. Include at least three wells with sample diluent only as mock / negative controls. Incubate for 2 hr at 37 ^{o}C or overnight at 4 ^{o}C.

4.5) Prepare the primary antibody solution fresh before use. Make a 1:2,000 fold dilution of the primary antibody working stocks using the primary antibody diluent (12% FBS in RPMI-1640). Ensure to prepare enough for the use of 100 µl solution per each sample/control well in a 96-well plate.

4.5.1) For example, to make enough solution for one 96-well plate, add 5 µl of the primary antibody working stocks to the primary antibody diluent for a final volume of 10 ml.

4.6) Wash capture plate 5 times with wash buffer.

4.7) Add 100 µl of the primary antibody solution to each well. Incubate for 1 h at 37 ^{o}C in a 5% CO_2 atmosphere.

4.8) Prepare the secondary antibody solution fresh before use. Make a 1:14,400 fold dilution of the secondary antibody (1 mg/ml Goat anti-rabbit HRP) using the secondary antibody diluent (7 % FBS, 0.01 % Tween-20 in RPMI-1640). To reduce pipetting errors, perform a two-step serial dilution. Ensure to prepare enough for the use of 100 µl solution per each sample/control well in a 96-well plate.

4.8.1) For example, to make enough solution for one 96-well plate, first add 1 µl of the secondary antibody to 99 µl of the secondary antibody diluent. Then add 70 µl of first dilution to the secondary antibody diluent for a final volume of 10 ml.

4.9) Wash capture plate 5 times with wash buffer.

4.10) Add 100 µl of the secondary antibody solution to each well. Incubate for 1 hr at 37 ^{o}C in a 5% CO_2 atmosphere.

4.11) Wash plate 5 times with wash buffer.

4.12) Add 100 µl of room temperature TMB substrate. Incubate 30 min at room temperature in a closed container to protect from light.

4.13) Add 100 µl of room temperature stop solution (1 N H_2SO_4).

4.14) Read absorbance at 450-650 nm in each well using a microplate reader. Use the absorbance value to score each well as infected or uninfected. A well is considered to contain infectious virus if the absorbance value is at least three times higher than the value read from mock / negative control wells. Calculate $TCID_{50}$ of the viral stock using the Reed-Meunch method [181].

5) Establish viral growth kinetics

5.1) Monoinfection

5.1.1) Seed $3x10^5$ PHA-stimulated PBMC/well in 48-well plates in a total volume of 500 µl/well. Keep the culture plates at 37 ^oC in a 5% CO_2 atmosphere until inoculation.

5.1.2) For each virus, prepare an inoculum containing 6,000 IU in 2 ml of cIMDM.

5.1.3) Inoculate wells in triplicate by adding 500 µl of the inoculum (1,500 IU) to the seeded cells. The final volume of the infected cell culture is 1 ml/well and the MOI is 0.005.

5.1.4) Aliquot 200 µl of the remaining inoculum to each of two 96-well plates for RNA isolation, one of which is saved as a backup.

5.1.5) Incubate cultures at 37 ^oC in a 5% CO_2 atmosphere for 16-24 h.

5.1.6) Wash cultures 16-24 h after inoculation.

5.1.6.1) Remove and discard 750 µl of culture supernatant.

5.1.6.2) Add 750 µl of fresh cIMDM. Wrap plates in plastic wrap and spin for 10 minutes at 300 x g. Remove and discard 750 µl supernatant.

5.1.6.3) Add 750 µl of fresh cIMDM. Incubate at 37 ^{o}C with a 5% CO_2 atmosphere (day 1).

5.1.7) Sample cultures daily from day 2 to day 7.

5.1.7.1) Transfer 500 µl of culture supernatant to a 1.8 ml centrifuge tube. Spin for 1 min at 3000 x g.

5.1.7.2) Transfer 200 µl of the cell-free supernatant to two 96-well sample plates for RNA isolation, again saving one plate as backup. Store supernatants at -80 ^{o}C until RNA isolation.

5.1.7.3) Add 500 µl fresh cIMDM to each culture. Incubate at 37 ^{o}C in a 5% CO_2 atmosphere.

5.1.7.4) Isolate RNA from 200 µl of supernatant (use commercial kits such as QIAamp Viral RNA Mini Kit) following the manufacturer's standard protocol. For a large number of samples, use the Qiagen QIAxtractor.

5.1.7.5) Discard cultures into Wescodyne at the end of the experiment.

5.1.7.6) Store RNA samples at -80 ^{o}C until cDNA synthesis.

5.2) cDNA synthesis (Reverse transcription)

5.2.1) For each RNA sample, add 1.2 nmol of dNTP and 1.2 pmol of cDNA synthesis

primer, (5'-GTTGATCCTTTAGGTATCTTTCCACAGC-3', HXB2 nucleotide 7968 to 7995) to 10 µl of viral RNA. Add water to a final volume of 14 µl. Flick the tube to mix and spin briefly to collect liquid at the bottom of the tube.

5.2.2) Incubate mixture for 5 min at 65 ^{o}C, then hold at 4 ^{o}C until the master mix is prepared.

5.2.3) Prepare master mix using 5x first-strand buffer (250 mM Tris-HCl, pH 8.3, 375 mM KCl, 15 mM $MgCl_2$), 5 mM DTT, 120 U of SuperScriptIII and 240 U of Rnase inhibitor. Add water to final volume of 10 µl.

5.2.4) Add 10 µl of master mix to RNA mixture, flick to mix and spin to collect.

5.2.5) Incubate mixture for 90 min at 50 ^{o}C to allow synthesis of cDNA. Incubate for 15 minutes at 70 ^{o}C to inactivate reverse transcriptase. Hold at 4 ^{o}C as needed.

5.2.6) Add 2 U of Rnase H, flick to mix, and then spin to collect.

5.2.7) Incubate 20 min at 37 ^{o}C. Store cDNA at -20 ^{o}C.

5.3) cDNA quantitation using qPCR system

5.3.1) Prepare a DNA standard dilution series. Do 10-fold serial dilution, from 3 x 10^6 copies/µl down to 30 copies/µl, of pNL4-3VifA. Use distilled water for dilutions. Prepare the standard dilution series fresh before use or prepare small batches and keep at -20 ^{o}C. Do not freeze-thaw the standards more than three times.

5.3.2) Set up a 96-well qPCR reaction plate. Ensure that each plate contains at least one well of negative controls, the standard dilution series in triplicate, and each cDNA sample in at least a duplicate .

5.3.3) For each qPCR reaction, use 12.5 µl of qPCR master mix, 0.2 µM probe, 0.8 µM each of the forward and reverse primers and 1 µl of cDNA or the standard dilution series or water/buffer (for the negative control well). Add water to a final volume of 25 µl. The qPCR probe is light sensitive. Keep it in closed container.

5.3.3.1) To detect cDNA derived from a pNL4-3VifA based molecular clone, use the VifA primer-probe: VifAB forward primer (GGTCTGCATACAGGAGAAAGAGACT), VifA reverse primer (5'-AGGGTCTACTTGTGTGCTATATCTCTTTT-3') and VifAB probe (6FAM-5'-CTCCATTCTATGGAGACTC-3'-MGBNFQ). For cDNA derived from a pNL4-3VifB based clone, use the VifB primer-probe: VifAB forward primer, VifAB probe and VifB reverse primer (5'-CACCTGCGTGCTATACCTTTTCT-3').

5.3.4) Set PCR cycling parameters to 50 ^{o}C for 2 min, 95 ^{o}C for 10 min, and 40 cycles of 95 ^{o}C for 15 sec and 60 ^{o}C for 1 min. Consult the manufacturer's support document for operation of the qPCR machine.

5.3.5) Calculate a standard curve using data from the triplicate standard dilution series. Compare amplification data of the cDNA sample to the standard curve to determine the copy number. Consult the manufacturer's support document for data processing.

5.4) Determine viral exponential growth phase

5.4.1) Plot viral growth kinetics with the sampling day along the X-axis and the cDNA copy number along the Y-axis and identify the viral exponential growth phase, i.e., when viral cDNA copy numbers increase in an exponential progression.

5.4.2) Use the GRC web tool (http://indra.mullins.microbiol.washington.edu/grc/) to calculate viral growth rate (*g*). The GRC tool accepts cDNA copy numbers from at least two time points as input, and outputs the viral growth rate (*g*). While the tool can

calculate the net growth rate from two time point data, it is strongly recommended to input data from three or more time points. Use only data obtained from time points within the exponential growth phase (see step 5.4.1) to obtain accurate growth rates. For a detailed description of the mathematical model used in the GRC web tool, see [151].

6) Growth Competition Assay

6.1) Seed 3 x 10^5 PHA-stimulated PBMCs (or 1 x 10^5 CEMx174 cells*)* in 500 µl total volume per well in a 48 well flat-bottomed plate.

6.2) Keep plate in 37 ^oC in a 5% CO_2 atmosphere until inoculation.

6.3) For each virus, prepare 3 ml of inoculum containing 6,000 infectious units (IU).

6.4) Transfer 1.5 ml of each viral inoculum to a sterile tube to create the dual infection inoculum. Add 500 µl of the dual inoculum (1,500 IU) to 3 x 10^5 cells in a 48-well plate. The final culture volume is 1 ml/well. Aliquot 200 µl of the inoculum to two 96-well plates for RNA isolation; save one plate as a back up.

6.5) Incubate inoculated cells at 37 ^oC with a 5% CO_2 atmosphere for 16-24 hr.

6.6) Wash cultures 16-24 h after inoculation.

6.6.1) Remove and discard 750 µl of culture supernatant.

6.6.2) Add 750 µl of fresh cIMDM. Wrap plates in plastic wrap and spin for 10 minutes at 300 x g. Remove and discard 750 µl supernatant.

6.6.3) Add 750 µl of fresh cIMDM. Incubate at 37 ^oC with a 5% CO_2 atmosphere (day 1).

6.7) Select sampling times to include at least 3 time points within the exponential growth phase observed in step 5.4.1.

6.7.1) For each sampling, follow step 5.1.7.

6.8) Perform cDNA synthesis as described in section 5.2.

6.9) Determine the viral variant ratio using qPCR.

6.9.1) Prepare a standard serial dilution series in triplicate, diluting 10-fold in each step from 3 x 10^6 copies/µl to 30 copies/µl of pNL4-3VifA.

6.9.2) Set up a 96-well qPCR reaction plate. Ensure that each plate contains at least one negative control, the standard dilution series in triplicate, and duplicates of each cDNA sample.

6.9.3) For each qPCR reaction, use 12.5 µl of qPCR Master Mix, 0.2 µM probe, 0.8 µM each of the forward and reverse primers and 1 µl of cDNA or the standard dilution series or water/buffer (for the negative control wells). Add water to a final volume of 25 µl. The qPCR probe is light sensitive, keep it in closed container.

6.9.3.1) Use the VifA primer-probe to detect signals in the negative controls and the standard dilution series and with one duplicate of the cDNA sample. Use the VifB primer-probe with the other duplicate of the cDNA sample.

6.9.4) Set PCR cycling parameters to 50 ^oC for 2 min, then 95 ^oC for 10 min, and then 40 cycles of 95 ^oC for 15 sec and 60 ^oC for 1 min. Consult the manufacturer's support document for operation of the qPCR machine.

6.9.5) Calculate a standard curve using amplification data from the standard dilution series. Compare amplification data of the cDNA samples to the standard curve to determine copy number. Consult the manufacturer's support document for data processing.

6.9.6) Use the GRC web tool (http://indra.mullins.microbiol.washington.edu/grc/) to calculate relative viral fitness (*d*). The GRC tool accepts cDNA copy numbers or chromatogram peak heights from at least two time points as the input, and outputs the net growth rate difference (*d*) between the two viruses. While the tool can calculate the net growth rate from two time point data, it is strongly recommended to input data from three or more time points. Use only data obtained from time points within the exponential growth phase (see step 5.4.1) to obtain accurate growth rates. For a detailed description of the mathematical model used in the GRC web tool, see [151].

6.10) Determine viral ratios using chromatogram peak-heights

6.10.1) PCR amplify HIV-1 *vif* fragments containing the VifAB sequence tag using VifFwd (5′ -GAAAGAGACTGGCATTTGGGTCAGGG-3′ ; HXB2 positions 5266–5291) and VifRev primers (5′ -GTCTTCTGGGGCTTGTTCCATCTGTCC-3′ ; HXB2 positions 5579–5553).

6.10.1.1) For each PCR reaction, use 1 µl of cDNA, 1X NH_4 buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTP, 2.5 U of Taq Polymerase, and 0.45 µM of each primer. Add water to a final volume of 50 µl.

6.10.1.2) Set PCR cycling parameters to 3 cycles at 94 °C for 1 min, 55 °C for 1 min, and 70 °C for 1 min, followed by 34 cycles at 94 °C for 15 sec, 58 °C for 30 sec, and 70 °C for 1 min, and then hold at 4 °C.

6.10.2) Purify PCR products using the QIAquick PCR purification kit according to the manufacturer's protocol.

6.10.3) Submit purified PCR products to a DNA sequencing service provider for Sanger sequencing.

6.10.4) Check the average read quality score, which should be provided by the sequencing service. If the average base call accuracy is less than 85%, redo step 6.10.1 to 6.10.3.

6.10.5) Use the ChromatQuan web tool (http://indra.mullins.microbiol.washington.edu/cgi-bin/chromatquant.cgi) to calculate viral ratio at each time point. The tool requires the sequence trace file (*.ab1) and the sequence 5' to the nucleotide site of interest. The tool measures the peak intensity at the specified site. The ratio of the peak intensity corresponds to the ratio of the two viruses present (Figure 2.4).

6.10.6) Use the GRC web tool (http://indra.mullins.microbiol.washington.edu/grc/) and the recorded peak intensities as the input to calculate viral relative fitness (*d*). See step 6.9.6. [151]

2.3 Results

To study the fitness impact of single amino acid changes in HIV-1 Gag-p24, we used oligonucleotide directed mutagenesis to introduce mutations into a pNL4-3 plasmid containing the HIV-1 COTB-p24 gene [147, 182, 183]. Viral stocks were generated by transfection of 293T cells and harvested after 48 hours. We estimated the 50% tissue culture infectious dose ($TCID_{50}$) of each viral stock by the Reed-Muench method [181]. The $TCID_{50}$ of the prototype and mutant viruses ranged from 10^4 to 10^5 IU/ml (Figure 2.3).

The growth kinetics of the recombinant viruses was established in PBMCs from a single donor at an MOI of 0.005. RT-qPCR was used to measure viral cDNA copy number daily for six days. All mutant and prototype viruses grew exponentially between day 2 and day 4, following which viral growth slowed, as indicated by a decreased slope of the viral RNA copy number increase (Figure 2.3). The decrease in cDNA copy number between day 0 (corresponding to the inoculum) and day 2 is due to the absorption of virus to cells and the removal of unbound virions by the day 1 wash (step 5.1.6). In the exponential growth phase, all three mutants had a slower growth rate (*g*) than the prototype virus (Figure 2.3).

All three mutants were competed against the prototype virus in growth competition assays at a total MOI of 0.005. Viral growth kinetics in dually infected cultures were similar to that of monoinfection; the viral exponential growth phase was between days 2 and 4, and viral growth reached a plateau around day 5 (Figure 2.4).

Viral growth rate differences were derived from the change in the viral ratio over time. The viral ratio was calculated based on cDNA copy number of the reference and the mutant viruses using RT-qPCR, and by comparing peak heights in sequence chromatograms at nucleotide sites distinguishing the two viruses (Figure 2.4**)**. The growth rate differences determined using the peak-height and viral RNA copy number methods yielded similar results (Figure 2.4). All three mutants had lower replication fitness than the prototype viruses with the mutant I256V having the lowest fitness (Figure 2.4).

2.4 Discussion

The protocols presented consisted of two main parts: construction of recombinant HIV-1 molecular clones and growth competition assays. In order to distinguish two viruses in a dually infected cell culture, it is important that the competing

molecular clones contain sequence tags, which can be detected by a RT-qPCR primer-probe assay or by Sanger sequencing. This protocol makes use of the VifA and VifB tags, which occupy the same region of HIV-1 NL4-3 *vif* and encode the same amino acid sequence but differ by six synonymous mutations. These mutations were shown not to affect viral replication fitness [151]. The PCR-based cloning method used in this protocol provides more flexibility in selecting cloning sites compared to restriction site based cloning. However, the efficiency of PCR cloning decreases as the insert size increases. The current limit of the insert size is ~5 kb [184]. For PCR-based cloning and site directed mutagenesis, the use of a high-fidelity DNA polymerase is crucial to reduce the likelihood of extra mutations. Use of the minimal number of PCR cycles needed to yield adequate amounts of products is also recommended. In our experience, we did not detect any extra mutations after the PCR-based cloning and site-directed mutagenesis steps described here. Nevertheless, the PCR products should be sequenced to check for any undesired mutations. Ideally, the entire HIV coding region should be resequenced.

At least three sampling time points should be examined within the exponential growth phase [148]. The viral growth kinetics must first be established using daily sampling to determine the appropriate culture period and sampling time points for the growth competition. One factor that affects viral growth kinetics is the multiplicity of infection (MOI). This protocol uses a total MOI of 0.005 for both monoinfection and growth competition, as it was shown to yield more robust results than lower MOIs [148]. Nonetheless, lower MOIs can be used to obtain a longer exponential growth phase if necessary, but at the expense of result consistency. This protocol suggests an initial infection ratio of 50:50, assuming that the fitness differences of the viruses are generally unknown beforehand. However, the use of unequal input ratios is appropriate when

there is preliminary data suggesting significant differences in viral replication kinetics. In these cases, an infection ratio of 70:30 is recommended, where the less fit virus is placed in excess, to allow for the detection of a large fitness difference [148].

The protocol for determining the $TCID_{50}$, viral growth kinetics, and for performing growth competition assays were optimized using HIV-1 subtype B, NL4-3 COTB-p24, and PBMCs from a single donor whose PBMC have demonstrated consistent susceptibility to HIV-1 infection *in vitro*. The culture period and sampling time points presented in this protocol are likely to be suitable for studying HIV-1 group M viruses in human PBMCs. Using PBMCs from a single source is highly recommended to obtain consistent results, as the viral replication can vary in different donor cells [185]. Although less desirable, PBMCs pooled from multiple donors can be used, provided that the same pool is used across all experiments. Another alternative is the use of cell lines. The protocol presented here was used successfully with the T-cell line CEMx174 [147, 152]. However, it is important that the numbers of cells seeded are re-optimized to achieve consistent cell growth. Viral growth kinetics must also be re-established, as it is likely to vary in different cell lines, to determine the appropriate sampling time points in the growth competition steps.

Two different methods to determine the viral ratio for calculating fitness are included in the protocol. The first uses RT-qPCR to measure viral cDNA copy number at each sampling time point. Viral replication fitness was then calculated from the viral ratio, based on cDNA copy number. Alternatively, the viral ratio can be determined based on the ratio of chromatogram peak heights at the VifAB tag sites. The two methods yielded comparable results (Figure 2.4**).** The chromatogram peak height method can be applied to other HIV-1 strains without an engineered sequence tag. For RT-qPCR, the use of primers or probes to distinguish viral variants must first be carefully evaluated (see [151]).

Nevertheless, RT-qPCR provides better sensitivity for samples with a small amount of viral RNA, such as those from the first time point within the exponential growth phase. Direct measurement of viral cDNA also allows detection of technical problems that may arise from RNA extraction and cDNA synthesis. Using molecular clones with sequence tags provides a cost effective solution to the RT-qPCR methods, as only two pairs of primers and probes are needed to study multiple viruses. This strategy also avoids the problem of peak-height variation in sequence chromatograms due to neighboring bases [186], as sequencing is done at the same site across all viruses in the study. The PCR products produced for RT-qPCR and Sanger sequencing can also be use with other methods to determine viral ratios, such as bulk sequencing of individual clones [153, 154], HTA [123, 137, 144, 173], or oligonucleotide ligation assay (OLA) [187].

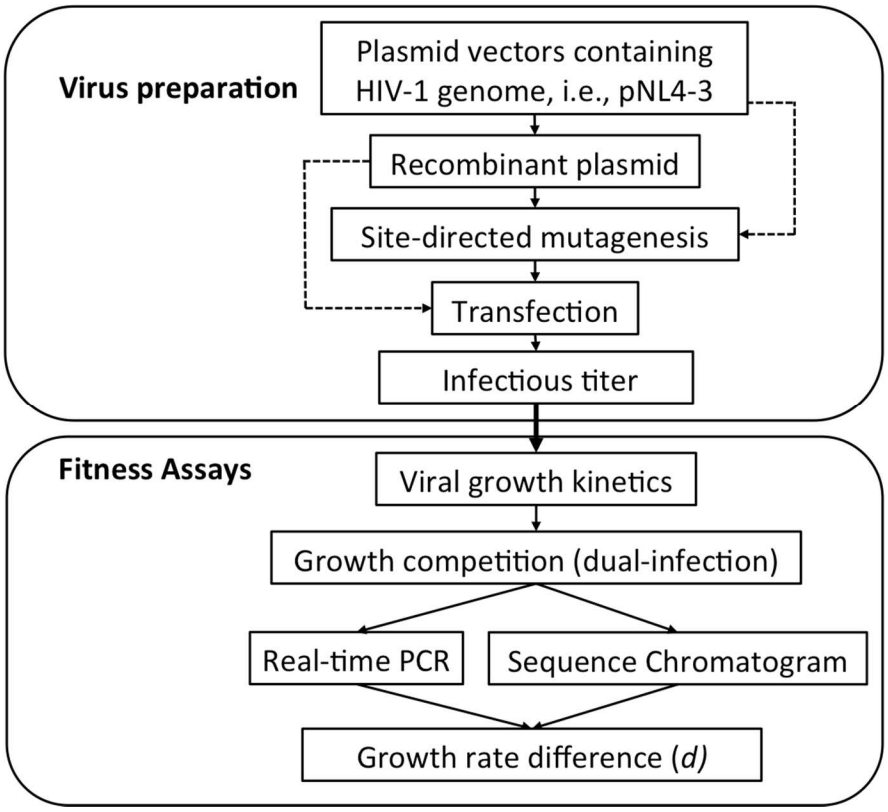

Figure 2.1 Flow diagram of the viral fitness protocols. The Virus preparation protocols concern construction of HIV-1 recombinant clones and generation of viral stocks. The Fitness Assays protocols are for establishing viral growth kinetics and determining relative viral fitness. Dashed lines represent alternative flows for the protocols. For example, a mutation can be introduced directly into the HIV-1 NL4-3 molecular clone without generating a recombinant clone.

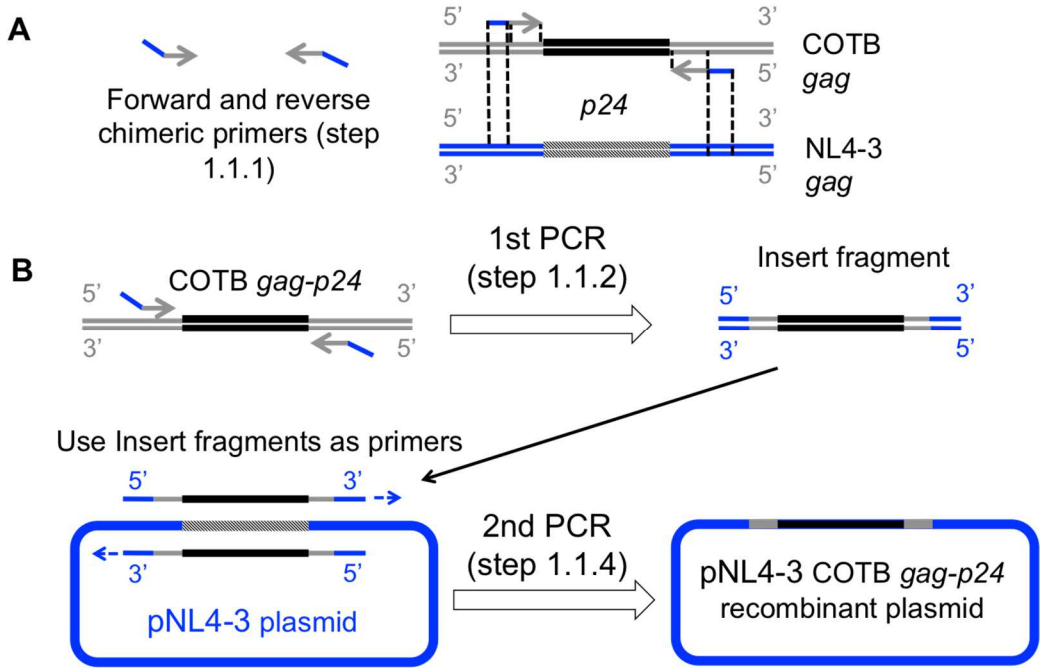

Figure 2.2. Construction of HIV-1 NL4-3 COTB Gag-p24 recombinant molecular clones using overlap extension PCR. A) Design the chimeric primers. The 5' halves of the primers contain the NL4-3 vector sequence and the 3' halves contain the ends of the insert sequence. B) Chimeric primers are used to amplify the insert fragment, COTB Gag-p24. C) The insert fragments are used as primers for a second PCR to generate a new recombinant plasmid.

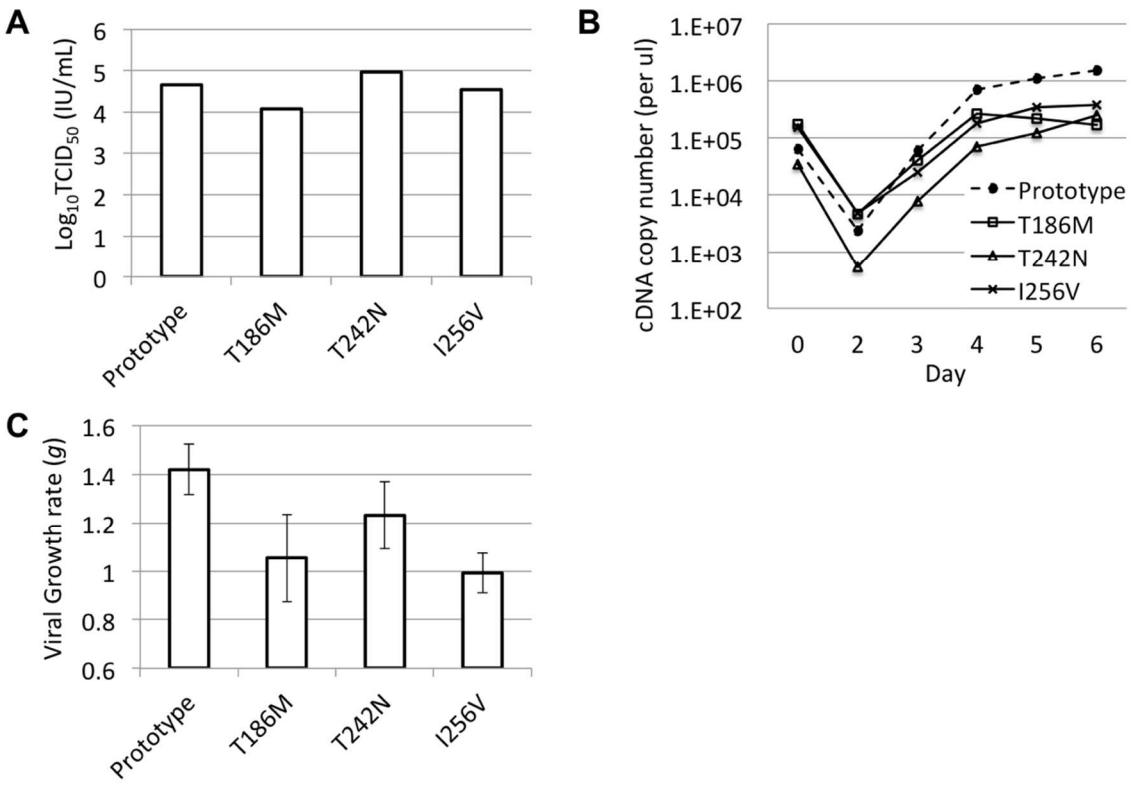

Figure 2.3. Viral growth characteristics in PBMCs. A) Log_{10} $TCID_{50}$ of viral stocks. B) Viral growth kinetics in monoinfections begun at an MOI = 0.005. C) Viral growth rates over six days, including the exponential growth phase (days 2-4) derived from panel B. The values shown represent the average of three replicates from one experiment. The error bars represent 95% confidence intervals.

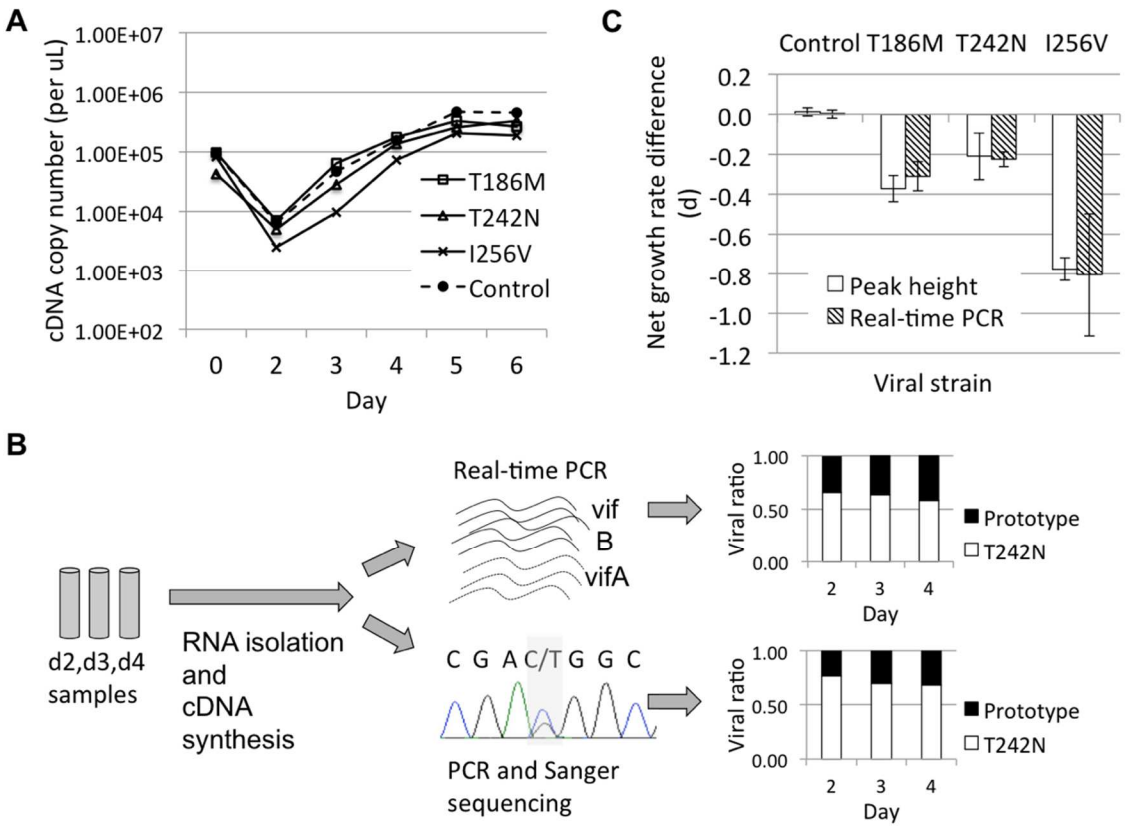

Figure 2.4. Viral growth and relative fitness determination. A) Viral growth in dually infected PBMC cultures. B) Ratios of T242N and prototype viruses determined using RT-qPCR or the sequencing peak-height method. C) Net viral growth rate differences *d* between the mutant and the prototype viruses in dually infected cultures, as shown in A. d values were calculated from the viral ratio data shown in B. The values shown represent the average of three replicates from one experiment. The error bars represent 95% confidence intervals.

CHAPTER 3

Fitness costs of mutations at the HIV-1 capsid hexamerization interface

3.1 Introduction

Human immunodeficiency virus type 1 (HIV-1) is an enveloped retrovirus with a protein shell, or capsid, enclosing the viral genome and its associated enzymes. The immature capsid is formed inside a host cell as a spherical shaped particle made up of Gag polyproteins. Upon budding and release from the host cell, the Gag polyproteins are cleaved by the viral protease, resulting in separation of CA from other functional domains of Gag. The CA units then rearrange to form a more condensed, cone-shaped mature capsid. This process is called maturation [84] and is necessary for successful production of infectious virions. The mature HIV-1 capsid is organized like a fullerene cone, and is made up of approximately 250 hexamer and 12 pentamer units of CA [103, 104]. Each CA protein consists of two independently folded domains – the amino-terminal domain (NTD) and the carboxyl-terminal domain (CTD). The two domains are joined by a short unfolded region called the flexible linker [105]. A recently solved high-resolution crystal structure of the capsid hexamer revealed two main interfaces between CA subunits within the hexamer – the NTD-NTD and NTD-CTD interfaces [102].

Amino acid changes at these interface sites have the potential to interrupt CA-CA interactions and hence destabilize the hexamer structure. Such mutations would also be expected to have a negative impact on viral replication, as previous studies showed that mutations that affected capsid morphology and stability also abolished viral infectivity [63, 94, 188, 189]. Indeed, some mutations at several interface sites were shown to result in less-,

or in some cases, non-infectious virus, while others did not have the same negative effects, suggesting that side-chain interactions had different effects on viral infectivity [94].

Identification of critical residues or side-chain interactions may facilitate the development of novel anti-HIV therapies targeting the capsid [190], including immunogens for cytotoxic T lymphocyte (CTL) based HIV vaccines. CTL responses are crucial for control of viral replication during infection [115, 191]. However, while CTL responses are extensive, their effectiveness is mitigated by immune escape mutations, which allow the virus to evade immune responses and continue to replicate [192, 193]. Focusing immune responses on the functionally constrained regions of the viral proteome can potentially alleviate this problem, as immune escape mutations may be expected to impair viral replication and reduce viral spread [96, 194]. This information would also be useful for development of novel anti-HIV drugs targeting the capsid [190].

A structural characterization of the CA-CA interface within capsid hexameric units was presented in 2009 [102], but the effects of mutations at many of these interface sites have not been investigated. Importantly, most of the previous work utilized alanine-scanning mutagenesis or other non-conservative amino acid changes to study alterations in CA production, capsid formation and viral infectivity. While such approaches are suitable for studying a functional or structural role of specific residues within a protein or protein complex, they did not take into account the likelihood of the observed mutations at a given residue during HIV evolution. Such information is useful when considering potential sites for drug target or vaccine immunogens and plausible drug/immune escape mutations. To gain such information, we introduced the most frequently observed mutation in natural infections at 12 interface sites and 18 non-interface sites and studied their impact on viral fitness. Our results showed that commonly observed mutations at interface sites were more likely to have deleterious

effects on viral replication than mutations at non-interface sites. These results support the potential of CA-CA interfaces as drug targets and for inclusion of these sites in vaccine immunogens.

3.2 Material and Methods

3.2.1 Mutation sites. The capsid hexamerization interface site is defined as any amino acid residue that forms at least one non-bonded contact with a residue from another CA subunit within the same capsid hexamer. The number and type of non-bonded contacts were obtained from PDBsum [195] using the x-ray crystal structure of the capsid hexameric unit (PDB code 3H4E). Twelve sites were selected for mutagenesis: six were located in the NTD-NTD interface, six were in the NTD-CTD interface. Eighteen non-interface sites were selected for mutagenesis to include sites that were located throughout CA, at which no specific functional roles have been reported and which were differentially conserved.

3.2.2 Amino acid substitution. At each selected interface and non-interface site we introduced a mutation in the COTM prototype (see below): usually the most frequently found amino acid was changed to the second most frequently observed one based on a dataset of circulating sequences. The amino acid frequency at each residue was determined as follows. First, all HIV-1 group M *gag* sequences were downloaded from the February 2010 HIV sequence database at the Los Alamos National Laboratory (HIVDB), without regard for duration of infection. The sequences were then screened to carefully exclude all but one sequence per subject or linked pairs of subjects. In addition, only full-length *gag* sequences with a proper start codon, without an early stop codon or frame-shift mutation, were used for the alignment. The multiple sequence alignment was prepared using the MUSCLE program and then manually edited using Mesquite [196]. The database frequency of each amino acid at each site was then calculated using a perl

script (http://indra.mullins.microbiol.washington.edu/perlscript/docs/CountAAFreq.html). In addition, alanine mutations were introduced at two interface sites, T186 and F301.

3.2.3 COTM CA sequence. The impacts of mutations on viral fitness were estimated in the context of the computationally derived Center-Of-Tree group M (COTM) CA sequence. The COTM sequence corresponds to the minimum evolutionary distance to all circulating group M viral sequences while still residing on an evolutionary path. It retains co-varying residues but was not biased toward outlier sequences [96, 182]. The phylogenetic tree and the COTM CA sequence were determined, as part of the COTM Gag sequence, based on the selected group M viral sequences obtained from the HIVDB using DIVEIN [197] (http://indra.mullins.microbiol.washington.edu/DIVEIN/diver.html).

3.2.4 Molecular clone construction. Two restriction sites, *Sfi*I and *Bst*EII, flanking the 5' and 3' ends of the *gag-ca* region, were introduced into pNL4-3 molecular clones using synonymous site changes, with or without a vif tag – a set of six additional synonymous mutations located in HIV-1 *vif* gene (kindly provided by Dr. Eric J. Arts, Case Western Reserve University) – using synthetic oligonucleotide primers with the QuikChange XL II Site-directed mutagenesis kit (Agilent). These nucleotide differences were used to differentiate the prototype and mutant strains in competitive fitness assays. The same restriction sites were also engineered using synonymous site changes into the 222H plasmid, containing COTM *gag* (kindly provided by Dr. Barbara Felber, NCI). Primer sequences are listed in Table 3.1.

The COTM-CA coding fragment was isolated from the modified 222H plasmid by *Sfi*I and *Bst*EII double digestion, gel separation and band extraction. The pNL4-3 vectors were prepared similarly. The COTM-CA coding fragment was inserted into the pNL4-3 vectors using T4 ligase (Invitrogen). We designated the chimeric plasmid with a vif tag

'COTM-CA-vifB' and the original chimeric plasmid 'CotM-CA-vifA'. All amino acid altering mutations were then introduced into the chimeric CotM-CA-vifA plasmid.

The engineered plasmids were transformed and propagated in electrocompetent DH10B *E. coli* cells using the ElectroMax kit (Invitrogen) following the manufacturer's protocol. Plasmid DNA was prepared using the HiSpeed Plasmid Midi Kit (Qiagen). The entire plasmid was sequenced to confirm the presence of the newly engineered restriction sites and the absence of extraneous mutations that could have arisen during the mutagenesis process.

Individual mutations were introduced into the chimeric pNL4-3-COTM-CA-vifA plasmid using the QuikChange II Kit and plasmid DNA was prepared using the HiSpeed Plasmid Midi Kit with an additional endotoxin removal step. The DNA sequence of the entire HIV-1 genome in mutated plasmids was determined to confirm the presence of the desired mutations and the absence of additional mutations.

3.2.5 Generation of HIV-1 Chimeric Viruses. Viral stocks were generated by transfecting HEK 293T-17 cells with 1ug of the chimeric plasmid using the FuGENE6 DNA transfection reagent (Roche). Cell-free supernatants were harvested 48 hours post-transfection, filtered through a 0.22 um filter, and stored in 250ul aliquots at -80C until use. Viral CA production in transfection supernatants was determined using an in-house double-antibody sandwich ELISA specific to the HIV-1 CA antigen [180]. The viral titer for each stock was calculated from quadruplicate measures as the 50% tissue culture infectious dose ($TCID_{50}$) using CEMx174 cells and following the Reed and Muench method [181]. Positive wells were scored based on the presence of cytopathogenic effects (CPE). The titer was considered undetectable for mutants showing no CPE. For confirmation, transfection and $TCID_{50}$ determinations were repeated for mutants showing no CPE.

3.2.6 Relative Fitness Assays. The COTM-CA with the *vifB* tag was considered the prototype virus. The replication fitness of all other viruses was estimated in relation to this prototype.

3.2.6.1 Viral Growth Competition (dual-infection) Assay. All competition experiments were done in the CEMx174 cell line cultured in RPMI-1640 medium (Sigma) supplemented with 10% fetal bovine serum (FBS) (Sigma) and 1% glucose. Cultures were incubated at 37^{o}C with a 5% CO_2 atmosphere. Experiments were done in triplicate at 10^5 cells/well in a 48 well plate using a viral multiplicity of infection (MOI) of 0.005. Viral inocula were prepared with the mutant and reference viruses at a ratio of 1:1 based on $TCID_{50}$. The inoculum volume was 0.5ml and the total volume of viral-cell cultures were 1 ml per well. After 16 hours, 0.75ml of the culture supernatant was removed and replaced with warm complete RPMI-1640. Cultures were spun for 5' at 228xg, and then 0.75ml of the culture supernatant was again replaced. On days 0, 3, 5, 7 and 9, 0.5ml of culture supernatant was replaced and the removed supernatant transferred to a sterile 1.8ml microcentrifuge tube. Supernatants were centrifuged for 5' at 228xg and cell free fluids were stored at -80^{o}C in 200ul aliquots.

3.2.6.2 Viral Growth (mono-infection) Assay. Viral-cell cultures were set up as described above using an MOI of 0.005. For repeated analysis of mutants with undetectable infectious titer, 0.5ml of viral stock was used as inoculum.

3.2.6.3 Nucleic Acid Preparation and Amplification. Viral RNA was extracted from 200ul cell-free supernatants by an ion exchange membrane technology on a QIAxtractor robot following the manufacturer's protocol (Qiagen). RNA was stored at -80^{o}C prior to cDNA synthesis. One 25ul cDNA synthesis reaction consisted of master mix 1 (final concentration: 0.5mM dNTPs, 0.5uM RT-2 primer), master mix 2 (final concentration: 1X first strand buffer, 0.005 M DTT, 10U/ul Superscript III reverse

transcriptase (Invitrogen), 2U/ul RNase Inhibitor (Invitrogen), and 10ul RNA. Four ul of master mix 1 was transferred to each reaction well in a 96 well PCR plate. Ten ul of RNA was added to each reaction well and incubated at 60^{o}C for 5 min. The reaction was then held at 4^{o}C until master mix 2 was prepared. Ten ul of master mix 2 was added to each reaction well and the mixture incubated for 90 minutes at 50^{o}C. Reverse transcriptase was then inactivated by incubating for 15 minutes at 70^{o}C and the reaction was then held at 4^{o}C until real-time PCR quantitation.

3.2.6.4 Real-time quantitative PCR (qPCR). Viral RNA was quantified to determine the growth kinetics of each virus. All mutant viruses were generated using pNL4-3 with a vif tag (vifB) and were competed against the vifA reference virus. Viral RNA from the two viral strains was distinguished based on the presence or the absence of the vif tag mutations [144]. Real-time qPCR was done using an ABI 7300 Real-Time PCR System (Life Technologies). One 25ul PCR reaction contained 12.5ul TaqMan® Gene Expression Master Mix (Life Technologies), 5uM probe, 20uM of the forward and reverse primers [151], and 1ul of cDNA. PCR cycling parameters were 50^{0}C for 2 min, 95^{0}C for 10 min, followed by 40 cycles at 95^{0}C for 15 sec and 60^{0}C for 1 min. A standard curve was generated using pNL4-3 diluted serially from 3x10^{7} to 300 copies/ul run in triplicate wells. The copy number of cDNA and corresponding Ct values was used to extrapolate cDNA copy from each sample. Each sample was run in duplicate wells.

3.2.6.5 Estimating Relative Fitness. All relative fitness estimations were done using the vFitness web tool [198]. The death rate of prototype and mutant viruses were set at 0.5. The least squares method was used to estimate fitness parameters from multiple data points. Data from the time points within the exponential growth phase were used for the relative fitness (1+s) calculation [199]. Viral cDNA copy number measurements from day 3 and 5 samples were used when the COTM virus was competed against the NL4-3,

while the copy number from day 3, 5 and 7 were used for competitions between COTM vifA and the COTM vifB mutants.

3.3 Results

3.3.1 Chimeric HIV-1 COTM-CA viruses. The COTM-*gag* sequence was computationally derived to represent the central point of group M viral sequence diversity [182, 183]. 1,019 full-length *gag* sequences, representing all group-M HIV-1 subtypes and circulating recombinant forms (CRFs), were obtained from the HIVDB: ~40% of the sequences were subtype B, and ~40% were subtype C, all other subtypes together composed ~10%, and the remaining 10% corresponded to circulating recombinant forms (CRFs) (Table 3.2). As the COTM sequence differed from known natural isolates or lab strains, we first confirmed that both COTM-CA-vifA and COTM-CA-vifB viruses were infectious in the CEMx174 T-cell line. We then competed the COTM-CA-vifA against the COTM-CA-vifB virus and showed that the six synonymous mutations in *vif* did not have an impact on viral replication fitness (Figure 3.1). The COTM-CA-vifB virus was also competed against the NL4-3 virus, and was found to have a slower growth rate and hence lower relative fitness than the highly culture adapted NL4-3 (Figure 3.1).

3.3.2 Mutations and CA sequence conservation. The amino acid frequency at each site within CA was determined from the 1,019 codon-aligned sequences. CA sequences were relatively conserved across group M, with an average database frequency of the prototype amino acid of 0.94. Overall, capsid hexamerization interface sites were as conserved as other sites (Wilcoxon rank sum test $p = 0.4317$). However, when restricted to the sites included in this study, the interface sites tended to be slightly more conserved than non-interface sites but the difference did not reach statistical significance (Wilcoxon rank sum test $p = 0.072$) (Table 3.3).

Base on database frequency, the 30 amino acid sites included in this study can be categorized into three groups: 1) 18 sites whose mutation pattern was conserved across group M subtypes, 2) 8 sites whose subtype B mutation pattern is the opposite of subtype C, and 3) 4 sites whose mutation pattern in subtype B and subtype C are similar but different from that of other group M subtypes (Table 3.4). At all 30 sites, the most frequently found amino acid was changed to the second most frequently observed one. In addition, alanine was substituted into two interface sites, T186 and F301.

3.3.3 Impact of single amino acid changes on COTM-CA viral replication. Thirty-two COTM-CA-vifA mutants were generated. CA production was detected in all cultures transfected with mutant plasmids. Transfection with four mutant plasmids (T58I, D166G, F161S, and T200S) yielded 20-fold or lower levels of CA production when compared to the prototype plasmid (Figure 3.2). From each viral stock, viral RNA was extracted, DNase treated and sequenced to confirm that no spurious mutations arose during the transfection process. The infectious titer of each viral stock was then determined (Figure 3.3). Eleven of these mutants did not show detectable cytopathogenic effects (CPE) or CA production in $TCID_{50}$ assays, in contrast to the prototypes and the other twenty-one mutants, suggesting that these mutants could not replicate at detectable levels in CEMx174. We also infected 10^5 CEMx174 cells with 0.5 ml of mutant viral stocks from the apparently defective mutants, along with 500 infectious units (IU) of COTM-CA-vifB and measured the amount of viral RNA from the culture supernatant on day 0, 3, 5, 7 and 9. For all eleven mutants, no growth was detected, although the mutant viral RNA remained above detectable levels (Figure 3.3). Although we cannot completely rule out the possibility of the mutants replicating at very low levels, the viral RNA detected was likely to be residual from the viral RNA in the inoculum as these levels steadily declined throughout the culture period (Figure 3.3). Four of the

eleven non-infectious mutants also showed the lowest CA production in the 293T-transfected culture (Figure 3.2). These mutations may interfere with the CA expression and/or the release of viral particle from host cells.

Next, we assessed the relative fitness of each of the other twenty-one mutants by competing them against the prototype COTM-CA-vifB virus in dual growth competition assays. At each time point, the mutant and prototype viral RNA were quantitated using specific primer-probe sets for vifA and vifB sequences, respectively. Viral replication fitness was calculated based on the change in prototype and mutant viral ratio over time (Figure 3.4). The mean relative fitness (1+s) of these mutants was spread over a wide range, from 0.21 to 1.31 (Table 3.5). The majority of the mutations resulted in decreased replication fitness, with 5: A64G (mean 1+s=0.21), V11I (0.24) D71E (0.25), E128D (0.30), and I124V (0.37), having the largest fitness costs (Figure 3.5). These five mutants exhibited a longer lag phase of viral production in competition experiments and slower growth kinetics (Figure 3.5). Eight mutations: K203R, T216S, F169Y, V230I, S44A, V148T, M68I and I27V, resulted in varying but less severe fitness costs with mean 1+s ranging from 0.60 to 0.87. Three mutants: R154K (mean 1+s±95% CI=0.82±0.24), T110N (0.89±0.14, and L6I (0.92±0.12), also showed a trend of reduced fitness, although not significantly different from the prototype ($p > 0.05$). Mutation I15L resulted in a trend toward higher replication fitness (95% CI $p > 0.05$) while four mutants: T48A (mean 1+s=1.12), G225S (1.21), E98D (1.28) and E45D (1.31), had a small but significantly faster growth rate than the prototype virus in all six replicates from two experiments ($p < 0.05$) (Figure 3.5).

3.3.4 No relationship between sequence conservation and fitness effects. To examine the relationship between amino acid conservation and the mutational impact on viral fitness, the sites chosen for mutagenesis were differentially conserved, from a

database frequency of 0.54 to 1.00 (Table 3.5). We found little direct relationship between the viral fitness and the database frequency of the prototype residue, with a non-significant Spearman's correlation coefficient $\rho = -0.27$ ($p = 0.136$) and $r^2 = 0.05$ ($p = 0.788$) (Figure 3.6). This observation also held true without the inclusion of non-viable mutants (data not shown).

3.3.5 Mutations at interface sites were more likely to be deleterious and had higher fitness costs than mutations at non-interface sites. Amongst the set of mutations we studied, mutations at interface sites were significantly more likely to be lethal, with 8/14 mutants resulting in non-viable viruses compared to 3/18 non-interface mutants (Fisher's exact test $p = 0.027$) (Figure 3.7 and Table 3.5). In addition, mutations at interface sites had significantly greater negative impact on viral fitness among the viable mutant viruses studied (Wilcoxon rank sum test $p = 0.040$) (Figure 3.7 and Table 3.5). Among the fourteen interface mutations, seven were in NTD-NTD and NTD-CTD interfaces (Table 3.5). Mutations at either interface were not significantly different in the generation of non-viable (Fisher's exact test $p = 0.30$) or less fit mutants (Student's t-test $p = 0.32$). For the NTD-NTD interface, five mutations (L20I, A42D, T54M, T54A and T58I) were located in a small patch of hydrophobic contacts between helix 1, 2 and 3, and two mutations (L6I and V11I) were situated in the beta strand at the beginning of the NTD. All five mutations in helix bundles resulted in non-viable viruses. On the other hand, both mutations in the beta strand resulted in viable viruses, with L6I having a small (mean $1+s = 0.85$) and V11I having a substantial ($1+s = 0.24$) negative impact on viral fitness (Figure 3.8). For the NTD-CTD interface, four mutations (A64G, M68I, F169Y and F169A) were situated in a cluster of hydrophobic contacts between helix 4, 8 and 11 and had various impacts on viral replication, from minor decreases to lethality (Figure 3.8). The other three mutations were at sites forming inter-domain helix-capping motifs for

helix 3, 4 and 8. Two of these, R173K and D166G, resulted in a loss of side-chain atoms required for hydrogen bonding (Figure 3.8), and yielded non-viable mutants (Figure 3.3), while the other, D71E, also had a substantial negative impact on viral fitness (Figure 3.5).

3.4 Discussion

We assessed the replication fitness cost of thirty-two single amino acid substitutions in the HIV-1 CA of the HIV-1 M group Center-of-Tree sequence. The COTM-CA sequence was derived from all group M viruses to represent the central point of viral sequence diversity. As such, the COTM-CA sequence is not identical to any natural isolate of HIV-1 but rather a coalescence of prototype amino acids from all subtypes, especially the two best studied, B and C. In pairwise competition assays, the chimeric NL4-3 COTM-CA virus showed slower growth kinetics than the prototype NL4-3, a highly cell culture adapted subtype B virus. The lower relative fitness of the recombinant NL4-3 COTM-CA virus may be due to less optimal interactions between COTM-CA and other NL4-3 viral proteins. Alternatively, by combining multiclade prototype residues, we might have disrupted subtype specific co-evolving residual pairs, and in doing so negatively impacted the CA structure and/or function and hence viral replication fitness. Arguing against this possibility, however, the COTM-CA sequence retains almost all subtype specific co-evolving residues that have been reported previously [200]. The only exception is at the residue 41 and 120 pairing, in which the COTM-CA sequence contains serine at both positions. The mutation S41T, which recovers the naturally observed co-variation of threonine at 41 and serine at 120 [200] (Table 3.6), did not show a significant impact on the replication fitness of the prototype COTM virus (data not shown). Thus, despite slower growth kinetics than the highly cell-line adapted NL4-3 virus, the COTM-CA virus was infectious and applicable as the prototype virus for studying the effect of mutations within CA on viral fitness.

From thirty-two single amino acid substitutions introduced into the COTM-CA virus, eleven were lethal and seventeen had fitness costs, while four appeared to yield faster replicating viruses. The majority of the lethal mutations were located in the interface between CA subunits within a capsid hexamer. X-ray crystallographic studies identified two interfaces between capsid hexamer subunits: NTD-NTD and NTD-CTD. Both interfaces consist of a small cluster of hydrophobic contacts and a more extensive network of water mediated hydrogen bonds [102]. In this study, the hydrophobic contacts at NTD-NTD were highly sensitive to change and even conservative amino acid substitutions were found to be deleterious. On the other hand, mutations at interface residues at the beginning of the NTD, which are outside of the hydrophobic clusters, showed less impact on viral replication. The same cluster of hydrophobic contacts at NTD-NTD is also observed in the x-ray crystal structure of HIV-1 capsid pentamer [201]. Previous studies showed that alanine mutagenesis and substitution of hydrophilic for hydrophobic residues in this otherwise hydrophobic cluster drastically reduced viral infectivity and, in some cases, altered mature capsid morphology but not particle production [63, 94, 189]. Taken together, this suggests that stabilization of capsid hexameric and pentameric subunits through this NTD-NTD hydrophobic cluster is crucial for the capsid function and, hence, essential for viral replication.

Mutations at hydrophobic residues at the NTD-CTD interface also negatively affected viral replication. However, these effects seemed to be smaller, as only the F169A mutation was lethal. The crucial interactions within the NTD-CTD interface were the inter-domain helix capping hydrogen bonds, which have been speculated to act as pivotal points for allowing one domain to move relative to another [102]. A similar structural motif was observed in the capsid hexamer of Rous sarcoma virus [202]. These pivotal points were suggested to be responsible for generating the continuously curved surfaces

observed across retroviral capsids [102]. Five inter-domain helix-capping interactions were observed in the crystal structure of the HIV-1 capsid hexamer, three of which were consistently observed while the other two were found only in some protomers [102]. Among the three consistently observed interactions, R173 and D166 were suggested to be the most critical amino acids, based on the HIV-1 and RSV capsid structures, respectively [102, 202]. Our results support these observations, as the mutations that disrupted these hydrogen bonds, R173K and D166G, were lethal.

Non-interface mutations included in this study were significantly less likely to be lethal (Fisher's exact test $p = 0.027$) (Figure 3.7) and showed less negative effects on viral replication (Wilcoxon rank sum test $p = 0.040$). Nevertheless, three mutations (H12Y, F161S and T200S) resulted in non-infectious mutants and two others (E128D and I124V) had substantial negative impacts on viral fitness, suggesting that these residues are important for CA functions other than mature capsid assembly. On the other hand, four mutations (T48A, G225S, E98D and E45D) had positive effects on viral fitness. It remains to be established how these mutations affect CA structure and function.

In addition to exploring the association between structural localization and relative fitness, we examined the relationship between relative fitness and sequence conservation and found a non-significant trend between these two properties. Although the number of sites was limited to thirty, which accounts for about one-seventh of CA, our observation should not be particularly affected by sampling. The lack of direct relationship between sequence conservation and the relative fitness was surprising but not entirely unexpected [147]. Sequence conservation is influenced by evolutionary constraints, such as structural and functional requirements, and selective pressures, such as host immune pressure, as well as random genetic drift. The results presented

here reflect the immediate impact of mutations on viral replication in a permissive T cell line, which would reflect CA structural and functional constraints but not host-specific selective pressures. In addition, our study focused on characterizing the effect of single mutations and hence did not account for possible occurrences of compensatory mutations, which could be part of the reason why some of the lethal mutations we observed, T54M, T58I and T200S, were found in noticeable fractions (~10%) of the HIVDB. Interestingly, these mutations were more frequently observed in non-B and non-C subtypes (Table 3.4 and Table 3.5). It is possible that their hypothetical compensatory mutations are missing from our recombinant COTM-CA virus. Besides these three mutations, the other eight lethal mutations were rare and observed in ~1% or fewer group M viruses (Table 3.4 and Table 3.5). To assess the influence of genetic background on the fitness effect, we compared results between this study and a fitness study done with subtype B viruses ([147], including 19 overlapping mutations with 6 at capsid hexamerization interface sites. We found a significant positive relationship between fitness effects in the two studies (r^2=0.544, p=0.0003; Spearman's Rho=0.774, p=9.605E-05) (Figure 3.9). Although our results may be affected by the limited capability of the *in vitro* fitness assay to recapitulate evolutionary forces *in vivo*, it indicates that extrapolating the direct impact of mutations on viral replication capability from sequence conservation alone can be inaccurate and that incorporation of structural and functional information is essential.

Our observation that mutations at interface sites were likely to have a large negative impact on viral replication is in agreement with previous studies [94, 188] and indicate the importance of capsid hexamer stabilization on viral replication. However, certain side-chain interactions, such as NTD-NTD hydrophobic contacts and NTD-CTD helix capping hydrogen bonds, were found to be more important than others in

maintaining capsid hexamer structures. As such they represent vulnerable targets that can be exploited in drug development and also vaccine designs that seek to target epitopes encompassing conserved elements critical to viral function [96, 194, 203, 204].

Table 3.1 Primers used to create new restriction sites and CA mutations in pNL4-3 plasmid.

CA mutant/restriction sites	**5'-3' (forward) sequence (mutations in bold)**
*sfi*I	CAAGCAGAAG**G**C**C**CAGCAGGC**C**GCGGCCGACA
*bst*EII	CGAGGCGATGAGCCA**G**GTG**A**C**C**AACACGACGATCATG
L6I	CCGATCGTGCAGAAC**A**TCCAGGGACAGATGG
V11I	TCCAGGGACAGATG**A**TCCACCAGGCCATC
H12Y	CAGGGACAGATGGTC**T**A**T**CAGGCCATCTCCCCA
I15L	CACCAGGCC**C**TCTCCCCACGGACGCTTAAC
L20I	CTCCCCACGGACG**A**TTAACGCGTGGGT
I27V	TAACGCGTGGGTCAAAGTA**G**TCGAGGAGAAGGC
A42D	TCCCCATGTTCTCGG**AT**CTTTCCGAGGGAGCC
S44A	CATGTTCTCGGCACTT**G**CCGAGGGAGCCAC
E45D	GTTCTCGGCACTTTCCGA**T**GGAGCCACCC
T48A	CCGAGGGAGCC**G**CCCCGCAGGAC
T54A	CCGCAGGACCTGAAC**G**CGATGTTGAACACCG
T54M	CCGCAGGACCTGAACA**T**GATGTTGAACACCGTC
T58I	CTGAACACGATGTTGAACA**T**CGTCGGCGGG
A64G	CGGGCACCAGG**G**GGCCATGCAGA
M68I	GGCGGCCATGCAGAT**A**CTTAAGGACACCATC
D71E	CAGATGCTTAAGGA**G**ACCATCAACGAGGAG
E98D	GGGCCAGATGAGAGA**T**CCGCGGGG
T110N	CGGGAACCACCAGCA**A**CTTGCAGGAGCAAAT
I124V	TGACTTCGAACCCGCCA**G**TCCCGGTCG
E128D	CAATCCCGGTCGGGGA**T**ATCTACAAGAGATGGA
V148T	GGATGTACAGCCCT**AC**CAGCATCCTGGAC
R154K	CAGCATCCTGGACATC**AA**ACAGGGACCGAAGGAG

F161S	GGGACCGAAGGAGCCG**AG**CAGAGACTACGTAGAC
D166G	GTTCAGAGACTACGTAG**G**CCGGTTCTTCAAGACTC
F169Y	TACGTAGACCGGTTCT**AT**AAGACTCTCCGGGCGG
F169A	CTACGTAGACCGGTTC**GC**CAAGACTCTCCGGGCG
R173K	GTTCTTCAAGACTCTC**AA**GGCGGAGCAGGCGACG
T200S	CCCGGACTGCAAGA**G**CATCCTGAAGGCTC
K203R	GACTGCAAGACCATCCTGA**G**GGCTCTCGGC
T116S	TGGAAGAGATGATG**T**CGGCGTGCCAGGG
G225S	AGGGAGTCGGGGGACCC**A**GCCACAAGGCG
V230I	CACAAGGCGCGG**A**TCTTGGCCGAGG

Table 3.2 Distribution of HIV-1 subtypes in the dataset used to calculate amino acid sequence conservation and derive the COTM sequence.

HIV-1 Subtypes in Group M	**Number of sequences**	**Percentage**
Subtype A	39	3.83
Subtype B	411	40.33
Subtype C	408	40.04
Subtype D	18	1.77
Subtype F	24	2.36
Subtype G	8	0.79
Subtype H	2	0.20
Subtype J	1	0.10
Subtype K	2	0.20
CRFs	106	10.40
Total	**1019**	**100.00**

Table 3.3 Database frequency of the consensus amino acid of group M HIV-1 CA

Sites (# of sites)	**Average database frequency of consensus residue**
All CA sites (231)	0.941
All interface sites (48)	0.936
Interface sites in this study (12)	0.872
Non-interface sites in this study (21)	0.810

Table 3.4 Amino acid database frequency of subtype B, subtype C and other group M sequences. Base on database frequency, the 30 amino acid sites included in this study can be categorized into three groups (identified by superscripts in "Mutation" column). The first group contains 18 sites, whose mutation pattern was conserved among group M subtypes, i.e., the most frequent and the second most frequent amino acid were the same in both Subtypes B and C. The second group consists of 8 sites, in which the most frequent and the second most frequent residue found in subtype B and subtype C were opposite. For example, at site 71, the most frequent and the second most frequent amino acid was glutamic acid (0.93) and aspartic acid (0.05), respectively, in subtype B sequences. But in subtype C sequences, it was aspartic acid (0.98) and then glutamic acid (0.01). The third group contains 4 mutations, whose mutation pattern in subtype B and subtype C are similar to each other but different from that of other group M subtypes.

	Mutation	Most frequent amino acid frequency				Mutant amino acid frequency			
		Group M (n=1019)	Subtype B (n=411)	Subtype C (n=408)	Others (n=200)	Group M (n=1019)	Subtype B (n=411)	Subtype C (n=408)	Others (n=200)
NTD-NTD	L6I[3]	0.687	0.713	0.860	0.280	0.066	0.114	0.039	0.021
	V11I[1]	0.939	0.976	0.973	0.795	0.037	0.015	0.012	0.135
	L20I[1]	1.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000
	A42D[1]	1.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000
	T54M[3]	0.867	0.971	0.953	0.475	0.096	0.002	0.005	0.475
	T54A[3]	0.867	0.971	0.953	0.475	0.005	0.002	0.005	0.010
	T58I[3]	0.876	0.978	0.968	0.480	0.103	0.002	0.010	0.500
NTD-CTD	A64G[1]	0.999	1.000	1.000	0.995	0.001	0.000	0.000	0.005
	M68I[1]	0.981	0.968	0.990	0.990	0.013	0.024	0.002	0.010
	D71E[2]	0.554	0.054	0.983	0.705	0.439	0.934	0.012	0.290
	D166G[1]	0.998	1.000	0.995	1.000	0.002	0.000	0.005	0.000
	F169Y[2]	0.538	0.000	0.995	0.710	0.456	0.993	0.000	0.285
	F169A[2]	0.538	0.000	0.995	0.710	0.000	0.000	0.000	0.000
	R173K[1]	0.999	1.000	0.998	1.000	0.001	0.000	0.002	0.000

Non-interface	H12Y[1]	0.984	1.000	0.993	0.935	0.013	0.000	0.002	0.060
	I15L[3]	0.617	0.720	0.650	0.340	0.320	0.253	0.321	0.455
	I27V[2]	0.575	0.251	0.880	0.620	0.419	0.742	0.113	0.380
	S44A[1]	0.965	0.925	0.990	0.995	0.029	0.068	0.002	0.005
	E45D[1]	0.987	0.990	0.983	0.990	0.008	0.002	0.012	0.010
	T48A[1]	0.984	0.988	0.983	0.980	0.006	0.000	0.007	0.015
	E98D[1]	0.918	0.912	0.914	0.935	0.074	0.080	0.076	0.055
	T110N[1]	0.861	0.876	0.821	0.910	0.118	0.110	0.145	0.080
	I124V[1]	0.739	0.929	0.556	0.720	0.246	0.056	0.422	0.280
	E128D[2]	0.601	0.895	0.302	0.605	0.390	0.090	0.689	0.395
	V148T[2]	0.663	0.238	0.968	0.915	0.274	0.630	0.015	0.070
	R154K[2]	0.555	0.698	0.373	0.635	0.438	0.294	0.618	0.365
	F161S[1]	0.993	0.993	0.993	0.995	0.002	0.000	0.003	0.005
	T200S[1]	0.842	0.961	0.880	0.520	0.091	0.005	0.025	0.405
	K203R[2]	0.588	0.983	0.196	0.575	0.406	0.010	0.799	0.420
	T216S[1]	0.976	0.964	0.985	0.985	0.009	0.017	0.003	0.005
	G225S[2]	0.562	0.679	0.490	0.470	0.416	0.287	0.500	0.510
	V230I[1]	0.925	0.869	0.971	0.950	0.066	0.119	0.020	0.050
Co-evolving residues	S41T[2]	0.474	0.742	0.037	0.815	0.482	0.168	0.949	0.175
	N120S[2]	0.336	0.535	0.201	0.200	0.432	0.248	0.539	0.590
	G116A[2]	0.391	0.681	0.032	0.525	0.443	0.226	0.789	0.180
	D187E[2]	0.394	0.002	0.787	0.395	0.594	0.985	0.203	0.585
	G208A[2]	0.619	0.236	0.865	0.905	0.365	0.752	0.123	0.065

[1] Amino acid site whose frequency patterns are similar in all group M sequences

[2] Amino acid site whose frequency pattern in subtype B is the opposite of subtype C. The pattern in other group M sequences is either similar to subtype B or subtype C.

[3] Amino acid site, whose frequency pattern in subtype B is similar to subtype C, but different from other group M sequences

Table 3.5 Amino acid database frequency, infectious titer and relative fitness of thirty-two COTM-CA mutations.

[a]Values shown are the average from two experiments done in triplicate.

Location	**Mutation**	**Amino acid frequency**		**Log_{10} TCID50**	**Relative Fitness (1+s)[a]**
		Prototype	**Mutated**		
	Prototype			5.24	1.00
NTD-NTD	L6I	0.687	0.066	5.21	0.92
	V11I	0.939	0.037	3.43	0.24
	L20I	1.000	0.000	Undetectable	-
	A42D	1.000	0.000	Undetectable	-
	T54M	0.867	0.096	Undetectable	-
	T54A	0.867	0.005	Undetectable	-
	T58I	0.876	0.103	Undetectable	-
NTD-CTD	A64G	0.999	0.001	3.07	0.21
	M68I	0.981	0.013	3.9	0.85
	D71E	0.554	0.439	3.43	0.25
	D166G	0.998	0.002	Undetectable	-
	F169Y	0.538	0.456	4.89	0.64
	F169A	0.538	0.000	Undetectable	-
	R173K	0.999	0.001	Undetectable	-
Non-interface	H12Y	0.984	0.013	Undetectable	-
	I15L	0.617	0.320	5.77	1.10
	I27V	0.575	0.419	3.91	0.79

	S44A	0.965	0.029	5.1	0.78
	E45D	0.987	0.008	5.1	1.31
	T48A	0.984	0.006	4.18	1.12
	E98D	0.918	0.074	4.98	1.28
	T110N	0.861	0.118	4.62	0.89
	I124V	0.739	0.246	3.67	0.37
	E128D	0.601	0.390	3.91	0.30
	V148T	0.663	0.274	5.24	0.81
	R154K	0.555	0.438	4.02	0.82
	F161S	0.993	0.002	Undetectable	-
	T200S	0.842	0.091	Undetectable	-
	K203R	0.588	0.406	4.02	0.60
	T216S	0.976	0.009	4.98	0.62
	G225S	0.562	0.416	4.14	1.21
	V230I	0.925	0.066	4.38	0.77

Table 3.6 Subtype B, subtype C consensus and COTM-CA amino acid at the co-evolving residual pair. Taken from [200].

Co-evolving residues	Site[a]	B consensus	C consensus	COTM-CA
27-148	27	V	**I**	**I**
	148	T	**V**	**V**
41-120*	41	**S**	T	**S**
	120	N	**S**	**S**
71-169	71	E	**D**	**D**
	169	Y	**F**	**F**
116-124	116	**G**	A	**G**
	124	**I**	V	**I**
148-169	148	T	**V**	**V**
	169	Y	**F**	**F**
187-208	187	E	**D**	**D**
	208	A	**G**	**G**

[a] Site numbering correspond to Gag protein of HIV-1 HXB2 strain

* Indicates subtype specific co-evolving residual pair that is not maintained in our COTM-CA sequence

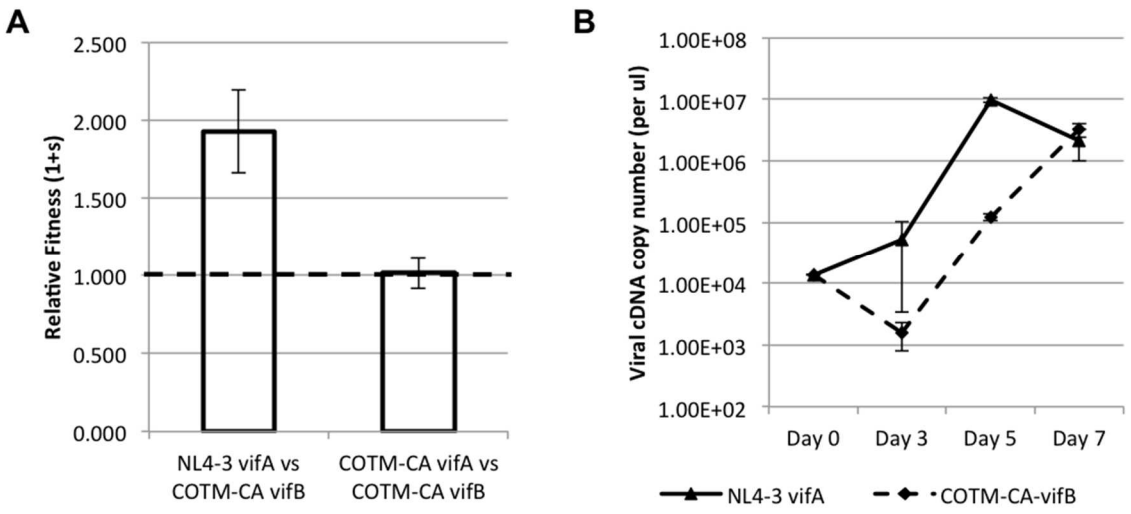

Figure 3.1 Relative fitness and growth kinetics of NL4-3 and COTM-CA viruses. A) COTM-CA vifB, containing six synonymous site changes in vif, relative to COTM-CA vifA, was competed against COTM-CA vifA and the NL4-3 prototype strain. Relative fitness values shown are the average from two experiments done in triplicate. The error bars represent 95% confidence intervals. The dotted line represents neutral fitness. B) Viral growth over seven days in cell culture. Values shown are the average from one experiment done in triplicate. The error bars represent the standard deviation.

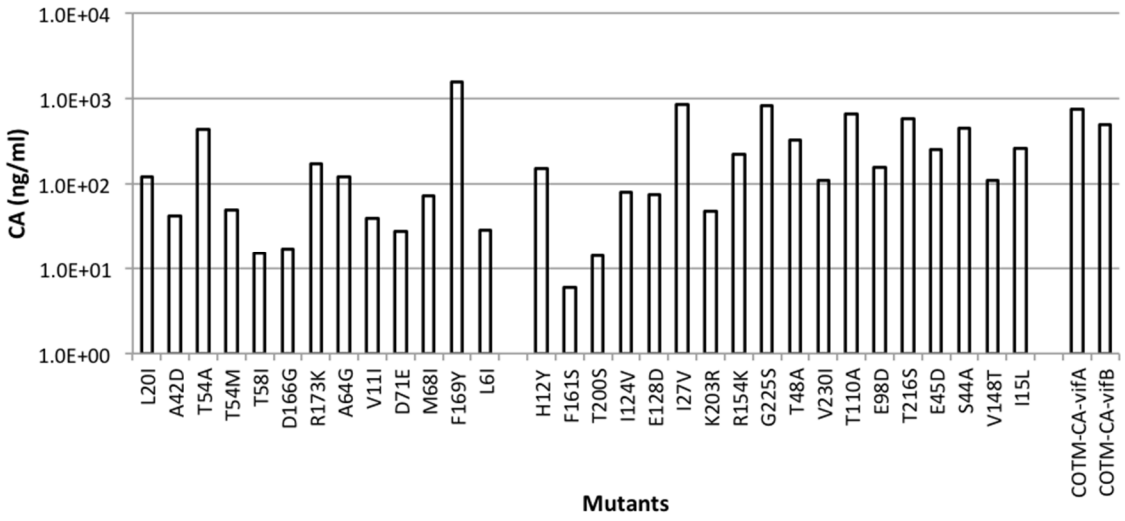

Figure 3.2 CA production in transfected 293T culture supernatants

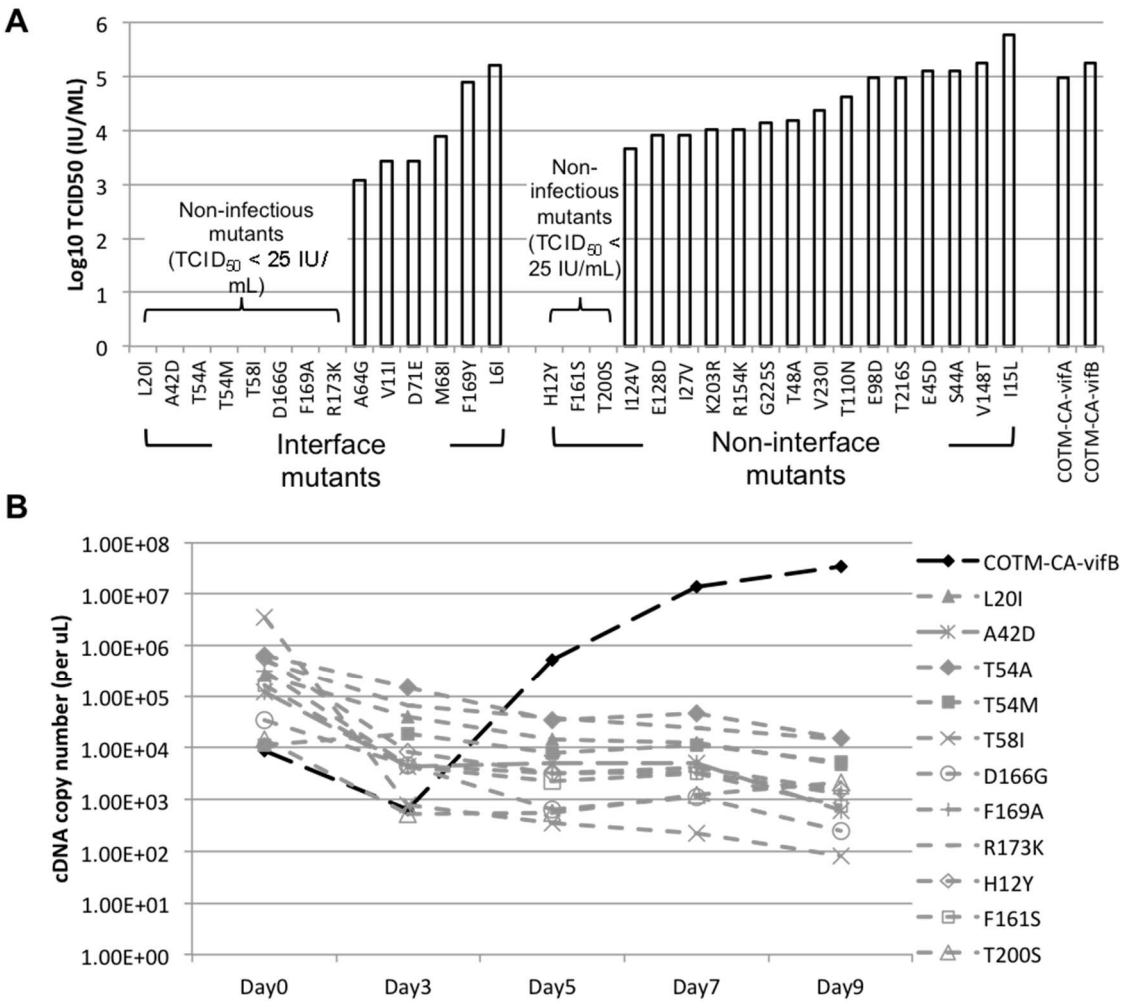

Figure 3.3 Growth kinetics of the prototype COTM-CA virus and its mutants in CEMx174 cells. A) Log_{10} $TCID_{50}$ of viral stocks determined in CEMx174 cells. **B)** Decay of non-infectious mutants in cell culture (gray lines). These mutants showed no cytopathic effects when cultured in CEMx174. Values shown are the average from one experiment done in triplicate.

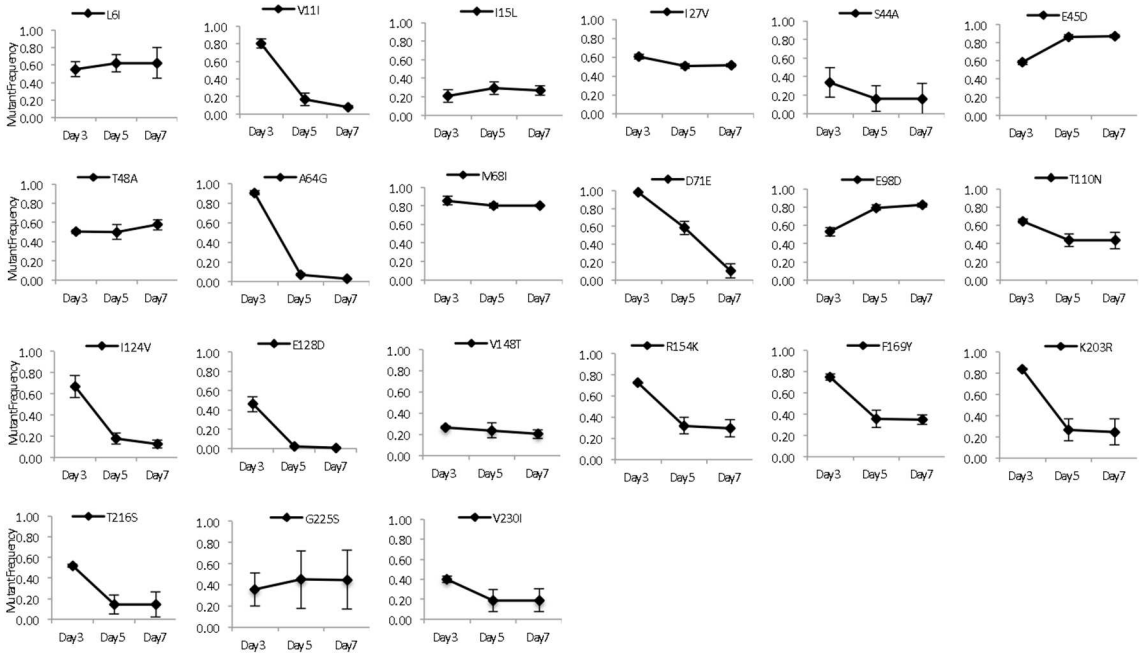

Figure 3.4 Viral ratios in pairwise growth competition assay. Values shown are the average taken from one representative experiment done in triplicate. The error bars give the standard deviation between the triplicates

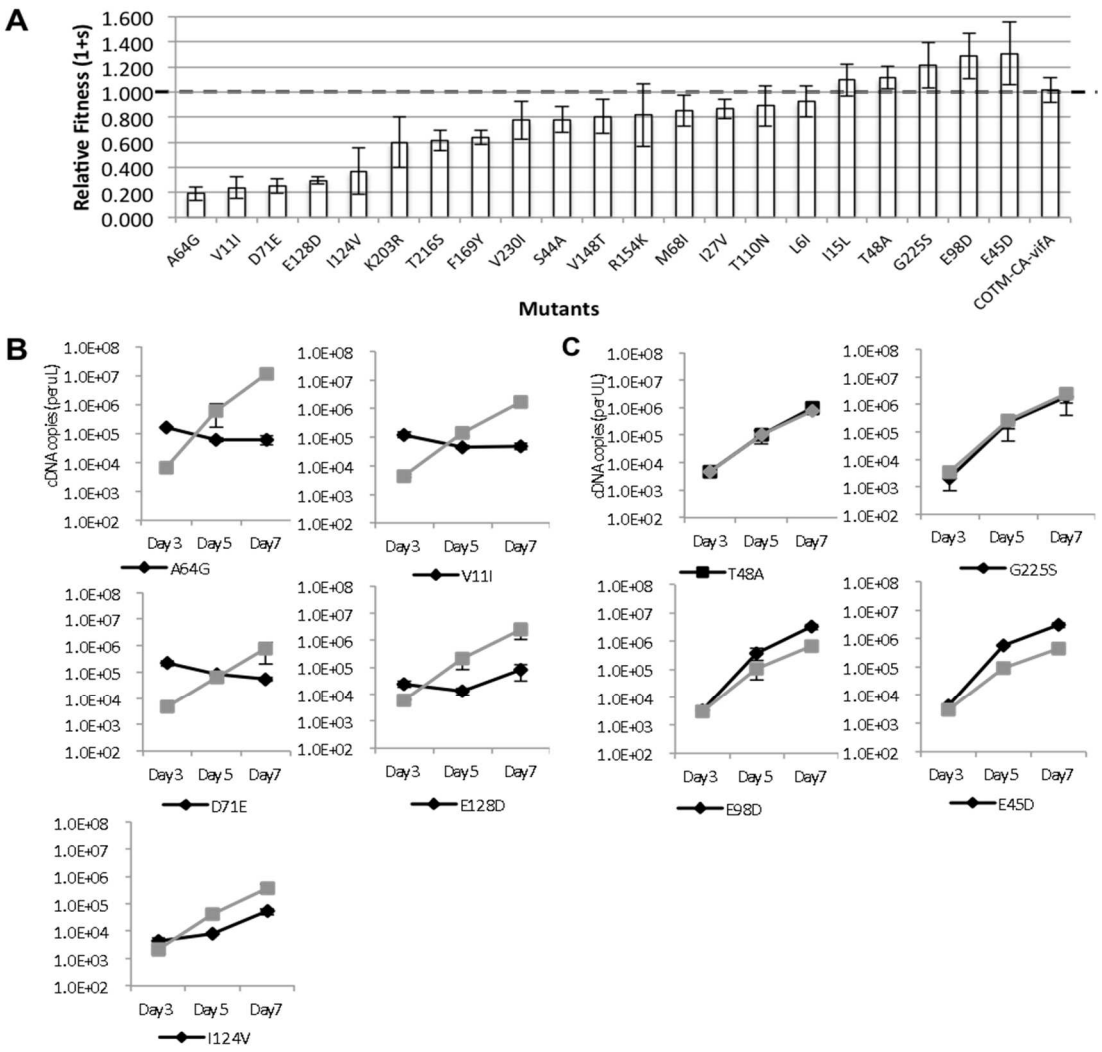

Figure 3.5 Viral replication fitness and growth kinetics of COTM-CA mutants in pairwise competition assays. A) Relative fitness of the 21 viable mutants in CEMx174 cells. Values shown are an average from two experiments, with three replicates each. Error bars represent 95% confidence intervals. The dotted line represents neutral fitness. B) Growth kinetics of the five mutants (black lines) with substantial lower fitness compared to the COTM-CA prototype virus (gray lines). C) Growth kinetics of the four mutants with higher replication fitness than the prototype. Values shown are the average from one selected experiment done in triplicate. The error bars represent standard deviations.

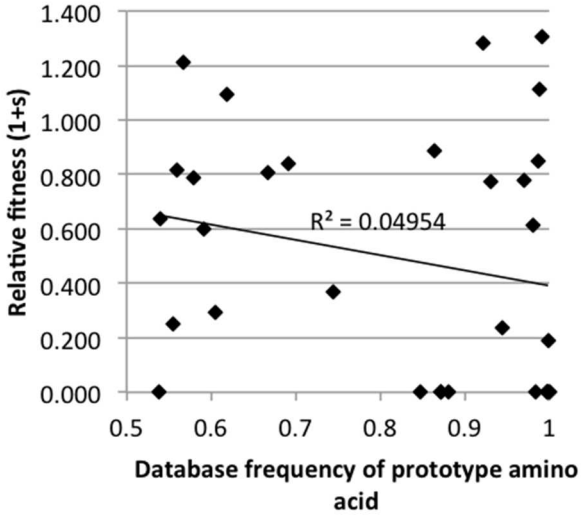

Figure 3.6 Relationship between sequence conservation and replication fitness. Relative fitness of all mutants evaluated as a function of database frequency of the amino acid found in the prototype COTM-CA sequence. Values shown are an average from two experiments, done in triplicate. The replication fitness of non-viable viruses is plotted as zero.

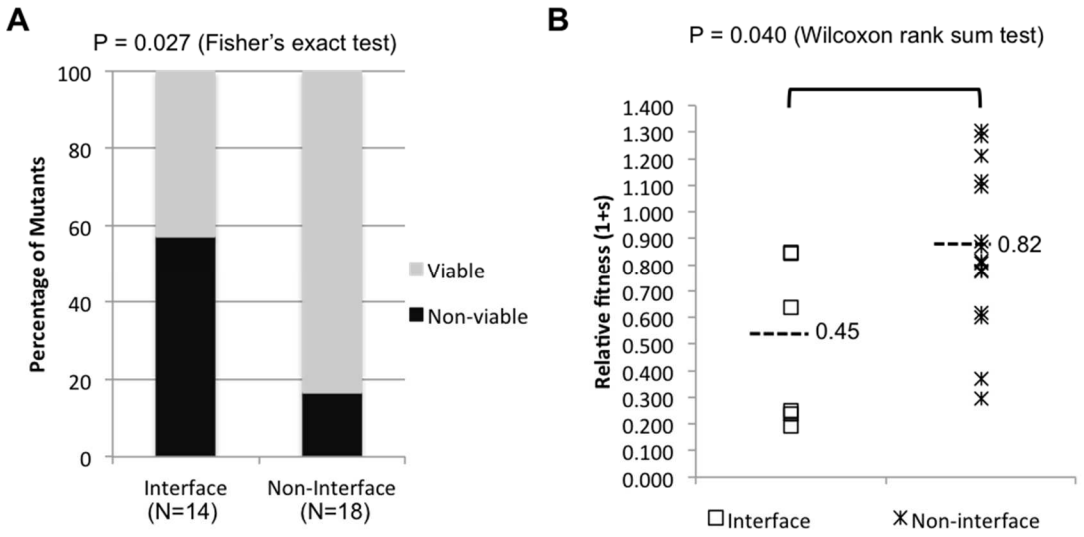

Figure 3.7 Viability and replication fitness of interface and non-interface mutants. A) Fraction of viable and non-viable mutants in each group. B) Relative fitness of viable mutants at interface and non-interface sites. Values shown are an average from two experiments done in triplicate.

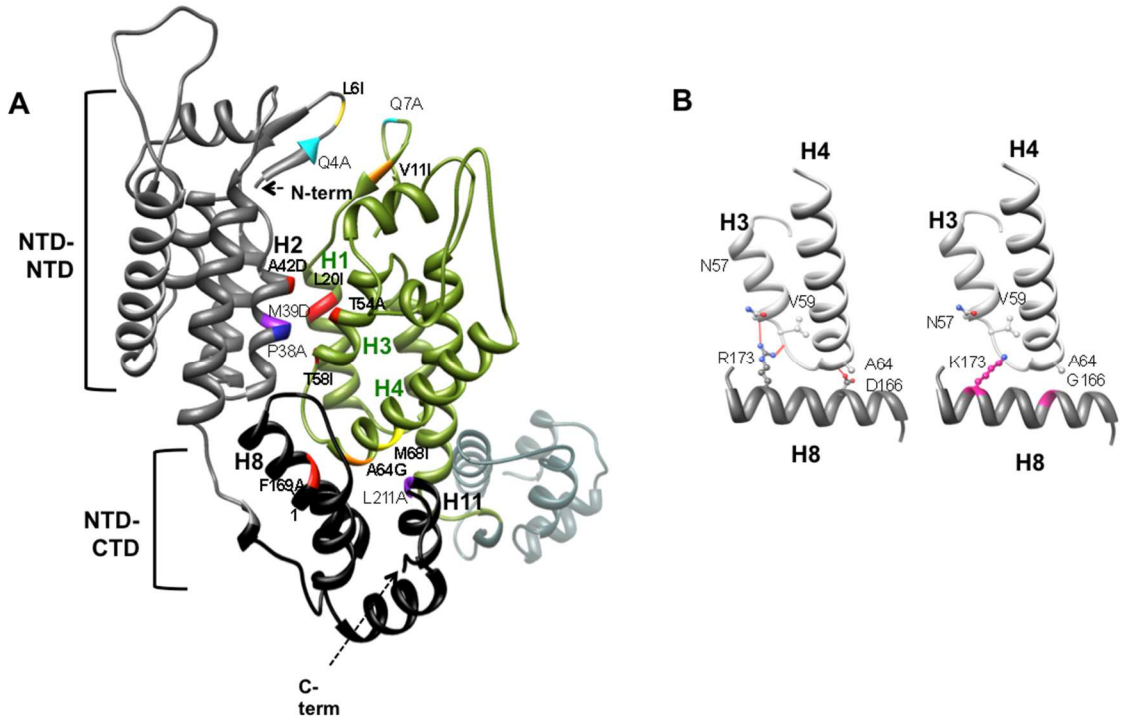

Figure 3.8 Structural localization of interface mutations. A) Two CA chains are shown as gray/black and green ribbons. The interface residues evaluated in this study are highlighted in bold and red-orange-yellow color. Other previously studied residues [94, 188] are highlighted in cyan-blue-purple color. The fitness impact of mutations is represented by the color shade ranging from small (yellow/cyan), moderate (orange/blue) to lethal (red/purple). B) Prototype residues participating in inter-domain helix capping hydrogen bonds, represented by orange lines, are shown on left. The mutations that resulted in loss of hydrogen bonds are modeled and highlighted on the right.

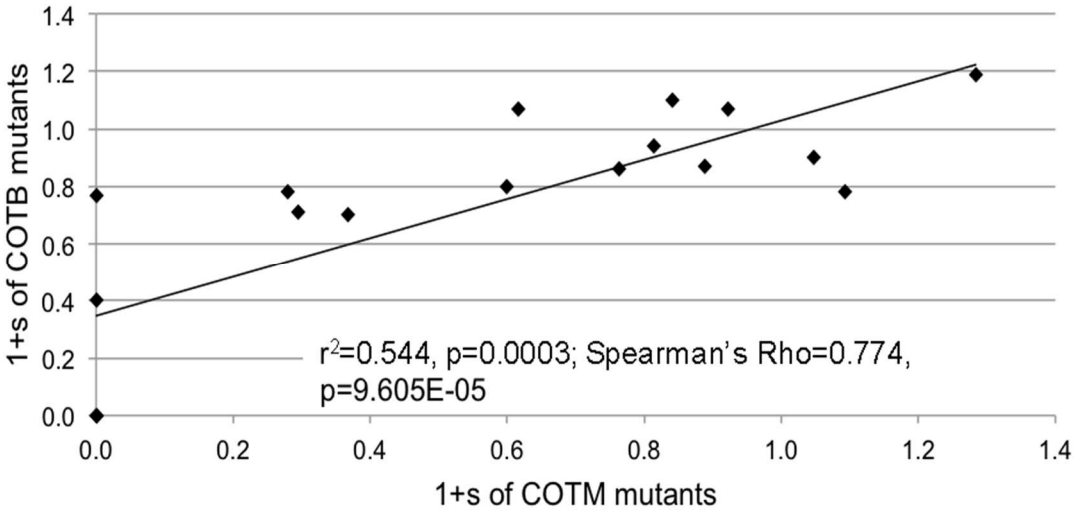

Figure 3.9 Positive relationship between fitness effects in two strains

CHAPTER 4

Composite Sequence-Structure Stability Models as Screening Tools for Identifying Vulnerable Targets for Drug and Vaccine Development

4.1 Introduction

HIV populations are notorious for rapidly acquiring mutations that allow the virus to escape host immune responses and antiretroviral drug pressure. One proposed strategy to overcome this problem is to focus immune responses and therapies on the conserved elements of the viral proteome; i.e., on sites likely to have important structural and/or functional roles [96, 205-208]. This strategy may also increase the likelihood that the suppressive effects induced would work against a broad array of HIV strains. Escape mutations occurring in conserved regions are also thought to be more likely to exact a fitness cost on the virus. Viral replication fitness is defined as the capacity of a virus to produce infectious progeny in a given environment [122] and is an important contributing factor in determining the prevalence of a virus variant at the population level over time [167]. High fitness costs of HIV-1 cytotoxic T lymphocytes (CTL) escape mutations have been associated with improved clinical outcomes [119-121, 209, 210]. However, the success of a conserved elements approach depends on the ability to identify critical features of viral proteins. Previous studies have shown only a weak relationship between fitness cost and sequence conservation at mutated residues. Specifically, mutations at highly conserved sites showed varying degree of fitness costs from deleterious to negligible [95, 147, 152, 211]. Interestingly, mutations at sites that remain conserved through time have been shown to have greater fitness cost than mutations of amino acids that became dominant

(and hence calculated to be conserved) later in the pandemic [147]. Indeed, HIV strains are being imprinted by the human HLA types encountered in different human populations [212, 213]. Thus, sequence conservation may change as the virus continues to evolve and adapt to host immunity, and contemporary sequence conservation may not be sufficient for pinpointing sites with crucial functional or structural roles.

Despite the rapid evolutionary rate of HIV-1, the viral capsid protein (CA) remains relatively conserved, with two-thirds of the protein having consensus amino acid frequencies of 0.9 or greater. The capsid plays several important roles in the viral replication cycle [77, 214] and is an emerging target for novel HIV drugs (Reviewed in [190]). It is also a target for host cellular immune responses that have been associated with viral control in both humans [115] and in non-human primate model systems [215, 216]. Recent studies suggest that specific regions of the HIV-1 Gag polyprotein, which contains the CA, remain highly conserved due to structural constraints, with sites buried in the core of the protein [217] and at CA multimer interfaces [152] being especially important. Another study reported an association between fitness costs of single amino acid changes in HIV-1 Gag and the absolute change in predicted protein stability [161]. These findings suggest that predicted protein stability might be a useful tool for identifying amino acid residues crucial for CA structure and function. However, only 29 of 500 residues of the Gag polyprotein were evaluated [161].

In this study, we utilized computational tools developed for protein structure modeling and prediction to create mutant CA models for all possible single amino acid changes in the protein, one at a time. The high-resolution crystal structure of the CA hexamer [102] and a dimer of the CA carboxy terminal domain (CTD) [108] were used as templates, as the CA hexamer is the major structural unit of mature HIV-1 capsid and the CA-CTD dimerization interface was postulated to be the interaction sites between

CA hexamers [102, 108, 218]. We then investigated whether predicted changes in CA structural stability are associated with the occurrence of mutations, and the impact of these mutations on viral morphology and infectivity. Overall, we found that the majority of all possible mutations were predicted to alter CA stability, either destabilizing or hyper-stabilizing the structure. Most of these mutations were not observed in the HIV sequence database (HIVDB) (http://www.hiv.lanl.gov/content/sequence/HIV/HIVTools.html). Conversely, mutations observed in the database were predicted to have negligible effects on protein stability. Furthermore, mutations that were predicted to change protein stability were associated with an effect on mature capsid structure and viral infectivity.

As different mutations at the same site can have different effects on the CA and viral replication [95, 218], we identified potentially immutable sites based on the proportion of mutations that were predicted to destabilize the protein structure. Approximately one-fifth of CA amino acid residues were predicted to be prone to destabilizing mutations. These vulnerable sites were all highly conserved and located in secondary structural elements throughout the protein, corresponding primarily to small clusters in the core regions and at protein-protein interfaces. Based on these observations, we derived a composite sequence-structure stability score, which could classify deleterious and non-deleterious changes with high accuracy. The computational approach proposed here might therefore be useful for identifying immutable sites in a protein for experimental validation as potential targets for drug and vaccine development.

4.2 Materials and methods

***In silico* mutagenesis**. All 19 possible amino acid changes were introduced *in silico* at each position in the HIV-1 CA protein, one at a time. Reference structures in which the starting amino acid was re-introduced into the structure were generated and used as a control set. The CA hexamer, PDB ID 3H4E [102], and the dimer form of the carboxy-

terminal domain of the CA, PDB ID 1A43 [108], were used as template structures for mutations in the amino-terminal domain (NTD; residue 1 to 147) and the carboxy-terminal domain (CTD; residue 148 to 219, including the linker region), respectively. While the CA hexamer structure also includes CTD, it does not contain the CTD dimerization interface, shown to be crucial for mature capsid structure and function [94, 218, 219]. Hence, the CTD dimer [108] was also used. Both structures were derived from the HIV-1 subtype B NL4-3 strain. The HIV-1 NL4-3 CA is composed of 231 amino acids, however, the last 13 residues were missing from the C-terminus of the template structures and, hence, excluded from the following analyses.

Two *in silico* mutation modeling approaches were applied: Fixed-backbone models explore the best-fit side-chain conformation while the main chain atoms of the mutated residue are kept unchanged from the original position in the template structure. The side-chain atoms of the mutated residue were replaced with those of the new amino acid. The best-fit side chain conformation was selected using the SCWRL program version 4.0 [220]. The model was then run through 200 steps of energy minimization to remove atomic clashes or internal constraints generated by side-chain replacement using the CHARMM force field, as implemented in the program NAMD version 2.8 [221]. Flexible-backbone models allow the main chain atoms to move along with the side chains and the best-fit combination is selected. These were generated in the FOLDX program suite [222] using the BUILDMODEL function with default parameters. This method allows the neighboring side-chains to be moved in order to explore alternative backbone conformations.

Proteins stability scores. Two different types of protein-scoring functions were used to assess the mutant models: The atomic distance-dependent statistical based scoring function referred to as Discrete Optimized Protein Energy (DOPE) [166] and the empirically

based force field energy function called FOLD-X energy function (FOLDEF) [155]. DOPE is part of the protein-modeling package MODELLER, available at http://salilab.org/modeller/ [223]. FOLDEF is part of the FOLDX program suite, available at http://foldx.crg.es/ [222]. Both programs were run using default parameters.

Sequence dataset and amino acid database frequencies. Full-length HIV-1 CA coding sequences were downloaded from the HIVDB (http://www.hiv.lanl.gov/). Any sequences with hypermutations [224], early stop codons, frame-shift mutations or ambiguous amino acids were excluded. This resulted in a dataset of 5811 subtype B sequences. A multiple sequence alignment was prepared using MUSCLE [225] and then manually edited using Mesquite [196]. The database frequency of each amino acid at each site in the final alignment was then calculated using a perl script available at (http://indra.mullins.microbiol.washington.edu/perlscript/docs/CountAAFreq.html).

Composite sequence-structure stability score. We derived a composite score from the mutation database frequency and FOLDEF score of the mutant flexible backbone model. First, we ranked all mutations by database frequency in ascending order. Mutations with the lowest database frequency, i.e. 0%, were given the lowest frequency rank. We then re-ranked all mutations by FOLDEF score in descending order, i.e. mutant models with the highest stability score were given the lowest stability rank. For each mutation, the two ranks are added to get a composite score.

4.3) Results

Initial explorations of different mutation modeling methods and protein scoring functions. Two different approaches are commonly used in modeling mutations in protein structures, fixed- and flexible-backbone modeling. To examine an effect from using different modeling methods on predicting a change in protein structure stability for a given mutation, we generated two sets of mutant protein structures, each with different

methods, and compared the predicted stabilities. For stability prediction, we applied two protein scoring functions, DOPE and FOLDEF, against each model separately. In total then, we obtained two structural models and four predicted stability scores for each mutation. Overall, use of fixed- versus flexible-backbone methods did not have a significant impact on the predicted stability of single point mutation models. We found very good agreement between predicted stabilities generated by DOPE in both NTD and CTD mutants (Spearman's rho = 0.89 and 0.96, respectively; Figure 4.1). Lesser agreement was observed for FOLDEF stabilities (Spearman's rho = 0.75 (NTD) and 0.68 (CTD); Figure 4.1), and lower correlations were observed between stability levels predicted by the two scoring functions on the same models (Figure 4.2).

Statistical- and empirical-based scoring functions showed different patterns of predicted stabilities. For a given set of models and scoring functions, models were separated into bins based on predicted stabilities. Bins were calculated by dividing predicted stability values from the lowest to the highest by number of bins. The number of bins did not affect the overall stability distribution patterns and we chose 20 for further analyses. All reference structures were clustered in the same bin. We considered this bin to represent the typical range of structural stability of the reference protein. Mutant models were considered to be as stable as the reference protein if they were in this bin. As the NTD and CTD models yielded similar distributions, we combined them for the purpose of presentation (separate NTD and CTD results are presented in Figure 4.3 and 4.4).

Using the DOPE scoring function, the predicted stabilities of mutant models had a normal distribution, with the peak being the same bin including the reference protein. About one-fifth of the flexible-backbone models were predicted to be as stable as the reference structures. Roughly equal numbers of the remaining mutants were predicted to

be more or less stable (Figure 4.6A and Figure 4.3). In contrast, using the FOLDEF scoring function, almost half of the mutant models were predicted to be as stable as the reference models. The other half were predicted to have lower stability and only ~2% of the mutants were predicted to be more stable (Figure 4.6B). Similar predicted stability distributions were observed using fixed-backbone modes, notwithstanding a larger variation in FOLDEF stabilities for the reference models (Figure 4.3).

Mutants observed in the HIVDB were predicted to have stabilities close to the reference models by both scoring functions. When considering all mutations, the mutant models had a large range of predicted stabilities (Figures 4.3, 4.4 and 4.6). However, when focusing on mutations that have been observed in the HIV database, the range of predicted stabilities decreased sharply, with the majority of the observed mutants clustered in the same stability score bin as the reference model. This pattern was most pronounced for the mutations observed in at least 1% of all sequences in the database. In contrast, the predicted stabilities of mutations that have not been observed in the database were widely distributed (Figure 4.6C and 4.6D). These predicted stability patterns occurred with all prediction sets, irrespective of the modeling methods, scoring functions or template (Figure 4.4).

Frequency threshold of tolerated mutations. The observation that the majority of observed mutations were not predicted to alter protein stability, in contrast to unobserved mutations, suggests that any mutation found in the database has a high likelihood of being tolerated. To explore this question we examined the frequency of 184 point mutations with known impact on infectivity in subtype B HIV-1 [63, 94, 95, 120, 147, 175, 188, 189, 209]. Ninety-four (48%) of these mutations resulted in non-infectious viruses. Forty of the 94 (42.5%) mutations that destroyed infectivity were not found in the database, while the remainder were found at least once. However, none were found in more than 0.2%

(11 of 5811) of the sequences. While the database frequency of inactivating mutations ranged from 0 to 0.2%, the frequency of infectivity conserving mutations ranged from 0 to 36%. Nineteen out of 90 of the latter mutations (~21%) had not been observed in the HIVDB while thirty-seven (41%) were present in more than 11 sequences (Figure 4.7). Overall, the inactivating mutations appeared at a significantly lower frequency than tolerated mutations (P=3.03E-8; Mann-Whitney U test). Using mutation frequency as a predictor for the impact on mutant infectivity, the threshold of 0.2% yielded the best prediction accuracy (64%) (Table 4.1).

Genetic barrier influences the emergence but not outcomes of amino acid changes. A considerable proportion of amino acid substitutions have not been observed despite having minimal impact on protein stability (Figure 4.4, 4.6C and 4.6D). Fifty-nine percent of all possible amino acid substitutions from the consensus sequence required two or more nucleotide changes. However, this proportion increased to 72% for undetected mutations and decreased to 13% for observed mutations (Figure 4.8A), suggesting that a genetic barrier, or the number of nucleotide changes required, may play a role in emergence of amino acid changes in the CA. Substitutions requiring single nucleotide changes were not further enriched when considering mutations observed in 1% or more of the sequences. Nevertheless, there was an increase in the proportion of transition mutations compared to transversion mutations (Figure 4.5A). Interestingly, we did not detect any association between genetic barrier and effect of mutations on viral infectivity or protein stability (Figure 4.5B and 4.8B-D).

Predicted stabilities in relation to the impact of mutations on mature capsid morphology and viral infectivity. The low percentage of observed mutations with different stabilities from the reference models hints at optimal protein stability being crucial for CA function. Two datasets were used to explore whether predicted stability

was also predictive of capsid structure and virus infectivity. The first consisted of 56 single amino acid mutations with known mature capsid morphology [63, 94, 188, 189]. Twenty-three were reported to result in aberrant capsid shape, while the other 33 had no obvious impact on capsid morphology. Using flexible backbone models, 88% of mutations resulting in a native conical shape were predicted to be as stable as the reference by FOLDEF. With DOPE, a majority of mutations resulting in a conical capsid shape were predicted to be more stable than the reference models, 76% (versus 21% predicted to be as stable) (Figure 4.9A and 4.9B). Similar results were obtained using fixed-backbone models (data not shown).

The second dataset included an additional 128 point mutants with known impact on infectivity in a subtype B virus backbone [95, 120, 147, 175, 209]. The stability distributions of infectious and non-infectious mutations resembled those of conical capsid and aberrant capsid conferring mutations (Figure 4.9). Using FOLDEF, mutations predicted to be as stable as the reference structures were associated with both conical capsid shape (P=5.4E-04; Fisher's exact test) and infectiousness (P=2.76E-10). We also observed a moderate correlation between absolute changes in FOLDEF, but not DOPE, scores and viral infectivity (Figure 4.10). Compared to mutation frequency, protein structural stability, as predicted by FOLDEF, performed better in classifying infectious versus non-infectious mutations (Figure 4.11). Using the stability level of the reference structures as the threshold, any mutation resulting in a less stable structure was predicted to be non-infectious, with a prediction accuracy of 75% (Table 4.1).

One-fifth of HIV-1 CA is prone to destabilizing mutations. As non-infectious mutants were more likely to be associated with lower stability, we speculated that a large proportion of destabilizing mutations at a site suggests low mutational tolerance. We identified 45 residues in the CA that were prone to destabilizing mutations (Table 4.2). At

least three-fourth of all possible mutations at these sites were predicted to lower protein stability by both scoring functions. All forty-five residues were extremely highly conserved, with consensus amino acid frequencies ranging between 0.98 and 0.99. Most of these residues were located in secondary structure elements of the protein with small solvent accessible surface areas - many had side-chains almost completely buried (Table 4.2). Exceptions were P99, located in the loop region between helix 4 and 5, and I150 and I153, which were located in the unstructured linker region between NTD and CTD [102]. However, I150 and I153, as well as L172 and L189, were shown to be involved in HIV-1 CA CTD-CTD dimerization, a structure missing from the CA hexamer structure. Individual mutations at these sites were shown to destabilize CTD-CTD associations [226]. The other residues were either situated in the core of the CA, or at protein-protein interaction interfaces in the CA hexamer and/or CTD dimer (Table 4.1). These regions were shown to be genetically fragile, with low tolerance for non-synonymous mutations [95, 152].

Composite sequence-structure score for predicting the impact of mutations. Low database frequency and the high stability score were individually associated with deleterious mutations. Conceivably, mutations with both undesirable properties would be more likely to be lethal than mutations with only one negative property or none. We therefore derived a composite sequence-structure score based on the ranking of database frequency and structural stability. Mutations were ranked from low to high based on the degree of undesirable properties, i.e., mutations with the lowest database frequency and highest stability score were assigned the lowest rank. This simple composite score predicted which viruses would be infectious or non-infectious more accurately than either mutation frequency or structure stability (Table 4.1 an Figure 4.11).

4.4 Discussion

Identifying and targeting critical elements of the HIV proteome remains a challenging task for therapy and vaccine development. Sequence conservation is often considered a proxy for functional or structural importance of a given amino acid site. However, only weak relationships between sequence conservation and fitness cost of mutations in the HIV-1 CA have been reported [95, 147, 152, 161]. In one study, a stronger association was found between the predicted changes in protein stability and the fitness cost of mutation; though, the number of mutation studied was relative small [161]. In this study, we expanded the analysis to a much larger set of HIV-1 CA mutations with two known phenotypes: capsid morphology and infectivity, to more accurately assess the ability of protein structure-stability models to predict the impact of mutations. We also analyzed results using two different protein structure-stability modeling approaches. Our results support the previous observation that *in silico* structural stability can be applied to identify potentially inactivating mutations in HIV-1 CA. More importantly, we found that the composite score derived herein, which takes into account both mutation frequency and the proteins stability score, performed better at classifying deleterious and non-deleterious mutations than either approach considered alone. The combined sequence-structure approach described here has the potential to serve as a target-screening tool for HIV drug and vaccine development.

To assess results from previous sequence conservation studies, the database frequency of 184 HIV-1 CA mutations with reported phenotypes was analyzed. None of the non-infectious mutations appeared in more than 0.2% of the CA sequences in the database, while the database frequency of infectious mutations varied much more widely (Figure 4.7). This observation suggests that database frequency is partially informative of the mutation outcome and that the frequency threshold for tolerated mutations is likely to be lower than 1%, a level presumed in previous studies [159, 160]. By using a mutation

frequency of 0.2% as the cutoff value, infectious and non-infectious mutations could be classified with 64% accuracy, compared to 75% and 80% accuracy achieved by the structural stability and the composite score approach (Table 4.1). The threshold mutation frequency is likely to fluctuate over time and across different viral populations, as the virus continues to evolve to escape host immune pressures and these mutations become fixed in circulating viruses. Antiviral treatment imposed selective pressures are also being introduced as evidenced by the increasing prevalence of drug resistant mutations, which were rarely observed before antiretroviral therapy become widely available [227]. Some of these mutations are being fixed along with compensatory mutations if the immune/drug escape mutations reduce viral fitness [209, 228-230]. As noted previously, amino acids conserved through time is a better indicator of necessary function than mutations that have emerged as consensus more recently [147]. It remains challenging to determine the threshold value most effective for predicting deleterious mutations.

In this study, mutations predicted to induce large alterations in the dimerization of or the structural stability of the CA hexamer were far less likely to be found in the HIV sequence database than those not predicted to alter stability. Among the set of 184 HIV-1 CA variants that had been characterized *in vitro*, all viruses with very high stability scores had aberrant capsids and were non-viable (Figure 4.9). These results support the idea that drugs or vaccines that target regions of the capsid in which putative escape mutants would reduce stability would be particularly effective. The optimal structural stabilities required of HIV-1 CA monomer and multimers to retain proper function are likely to be maintained in the same host cellular environment, regardless of extracellular immune or treatment selective pressures. All sites predicted to be prone to destabilizing mutations were highly conserved (Table 4.2). These sites were located in the core of

individual CA molecules or at protein-protein interfaces found in CA hexamers and CA-CTD dimers. Interface residues, prone to destabilizing mutations, are good candidate targets for HIV drugs, whereas HIV vaccines inducing cytotoxic T lymphocyte responses can target multiple immutable sites, regardless of the structural location of these sites within virions.

The structural stability based approach has disadvantages. It does not directly address stabilizing yet deleterious mutations. About twenty percent of debilitating mutations did not alter CA hexamer or CTD dimer stability. Additionally, about 20% of undetected mutations also had a minimal impact on protein stability. While it is likely that the stability prediction was not entirely accurate, it is also probable that some mutations confer deleterious effect by disrupting other molecular processes, such as CA interactions with other proteins or protein expression and processing [94], and hydrogen bonding [152], without significantly altering protein structural stability. In addition, we hypothesize that some portion of these undetected mutations may not have been observed due to a higher genetic barrier, since amino acid changes requiring a lower number of nucleotide changes were highly associated with the emergence of mutations. The bias toward mutations with low genetic barrier may actually incorporate a bias toward more conservative amino acid changes, which are generally believed to induce smaller changes to protein structure and function. Despite this expectation, we did not find any association between the level of genetic barrier and the deleterious effects of mutations (Figure 4.8B-D).

The performance of the structure-stability approach depends on the accuracy of the protein modeling and stability prediction methods employed. Different scoring functions resulted in greater variation in the predicted stability of single point mutations than different modeling methods. These results support FOLDEF as a more suitable scoring

function than DOPE for predicting the impact of mutations. FOLDEF performed better at classifying deleterious and non-deleterious mutations, and the absolute change in FOLDEF score between the mutant and the reference structure also negatively correlated with change in viral infectivity, which coincided with results from a previous study [161]. However, further studies are necessary to assess the performance consistency of these scoring functions across different proteins. Using FOLDEF, we achieved 75% accuracy in predicting deleterious mutations (Table 4.1). As computational techniques used for protein structure and stability prediction continue to improve, the performance of our approach could likewise be improved. In addition to stability scoring functions, the functional state and origin of the template structure can affect predictions and the interpretation of results. As no high-resolution structure of the HIV-1 immature capsid has yet been elucidated, we only employed the CA hexamer and CTD dimer, which are the functional subunits of the mature capsid, as template structures. Thus, our predictions may not cover all mutations that affect immature capsid formation.

One limitation of this work is that only single amino acid changes were studied, whereas compensatory mutations can arise during viral infection that can restore protein stability and function [231-233]. Also, all sequence, structure and experimental data used in our analyses were obtained using HIV-1 subtype B viruses. The impact of mutations may be different in other different genetic backgrounds. Thus, studies of more complicated mutational patterns on protein stability will provide further insight for identifying desirable targets for HIV vaccines and therapies.

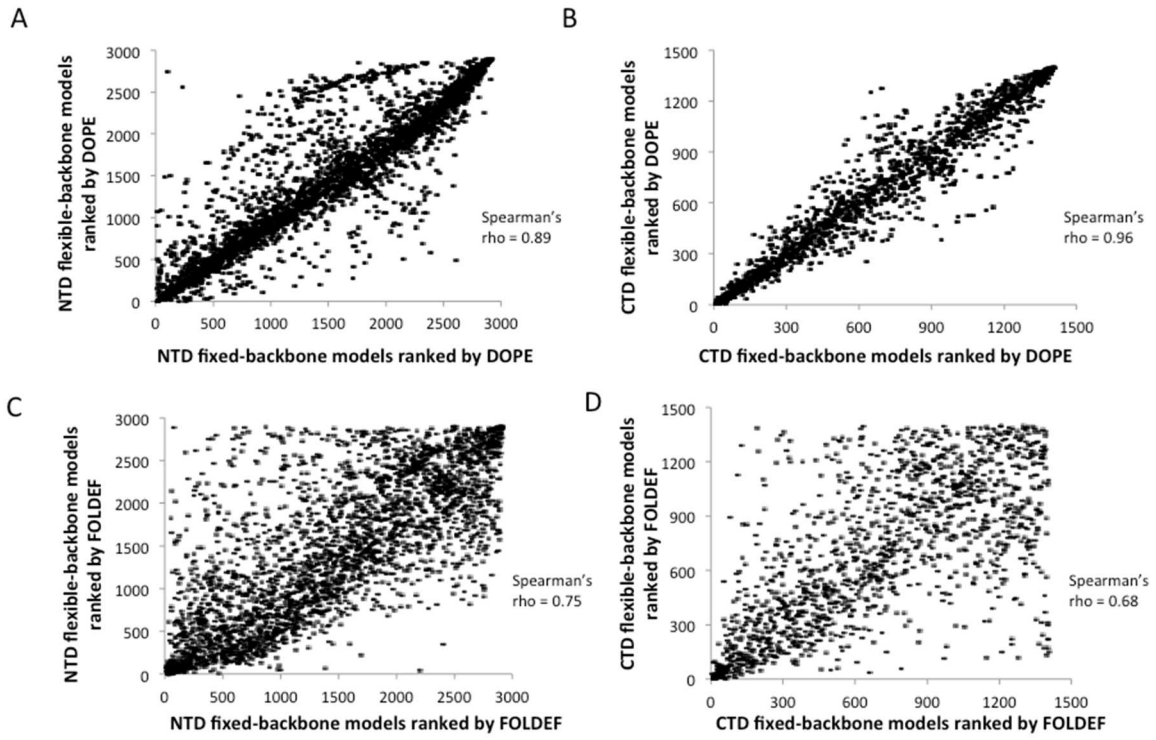

Figure 4.1 Fixed- and flexible-backbone models have highly correlated predicted stabilities. Each dash represents a model. The models were ranked based on stability from the lowest to the highest. Rank correlation of NTD and CTD mutants as predicted by DOPE (A, B), and FOLDEF (C, D).

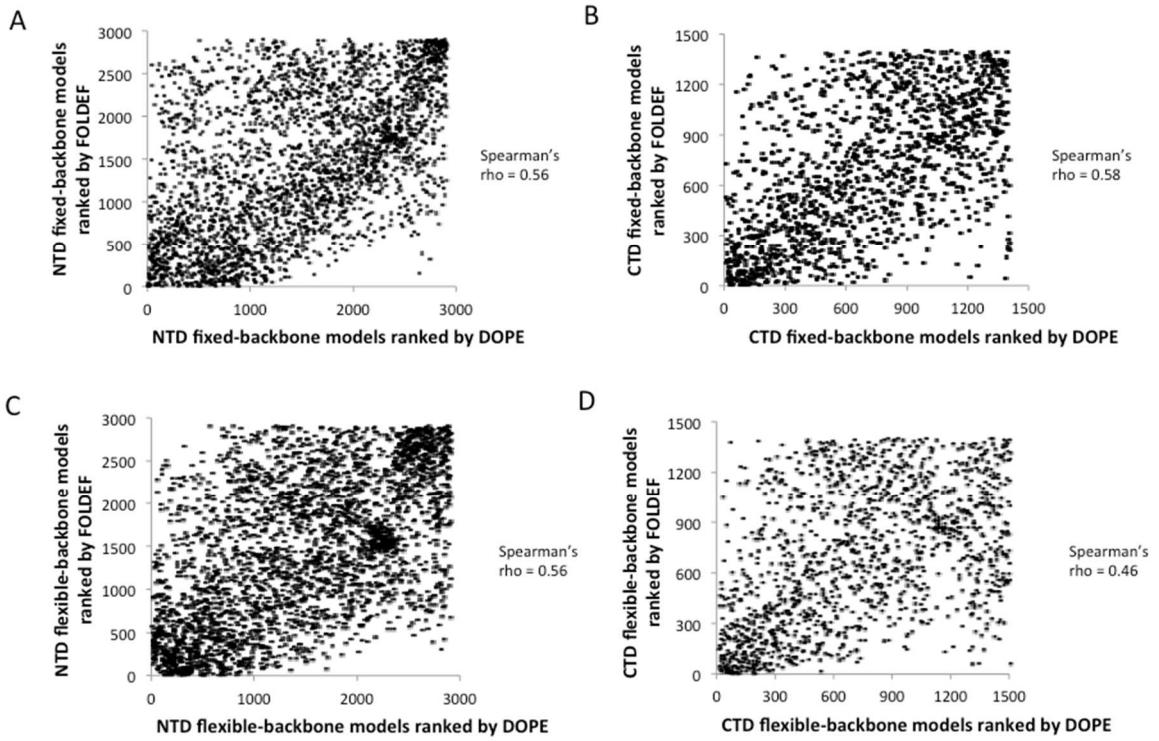

Figure 4.2 Rank correlation between DOPE and FOLDEF stabilities. Each dash represents a model. The models were ranked based on stability from the lowest to the highest. Rank correlation of NTD and CTD using a fixed- (A, B) or flexible-backbone model (C, D).

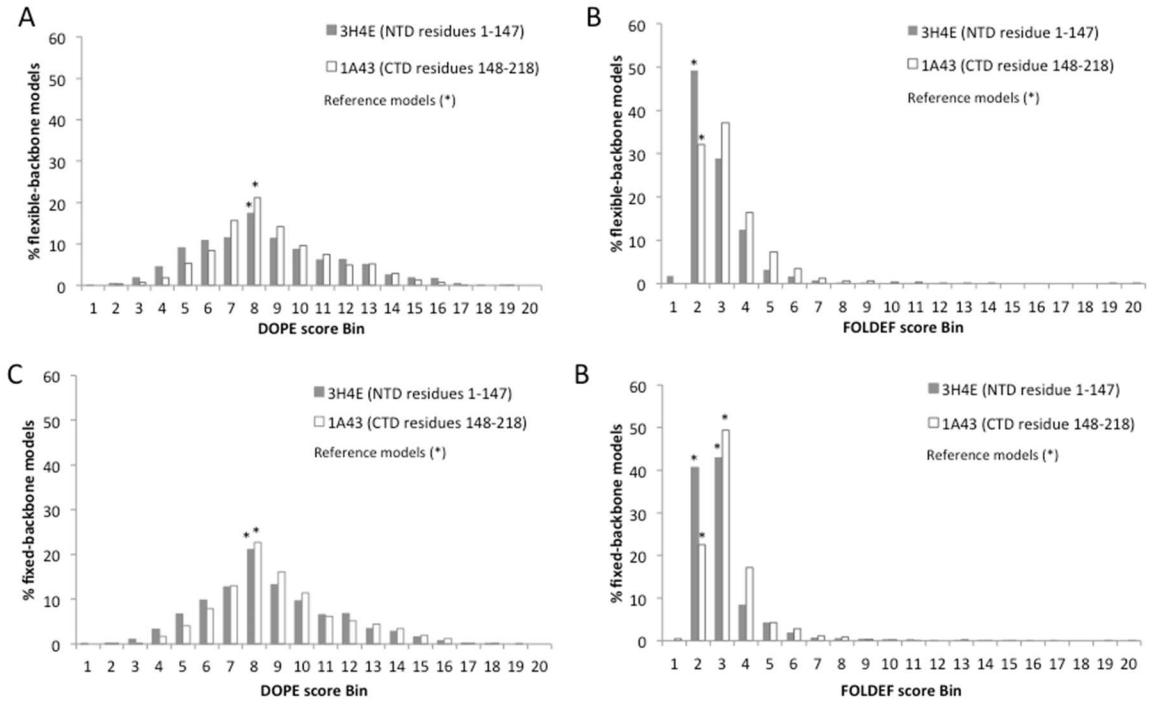

Figure 4.3 Distribution of the NTD and CTD mutant stabilities. Stabilities of flexible-backbone models as predicted by (A) DOPE and (B) FOLDEF. Stabilities of fixed-backbone models as predicted by (C) DOPE and (D) FOLDEF. The score bin reflects the structural stability from higher to lower. '*' Indicates stability of reference structures.

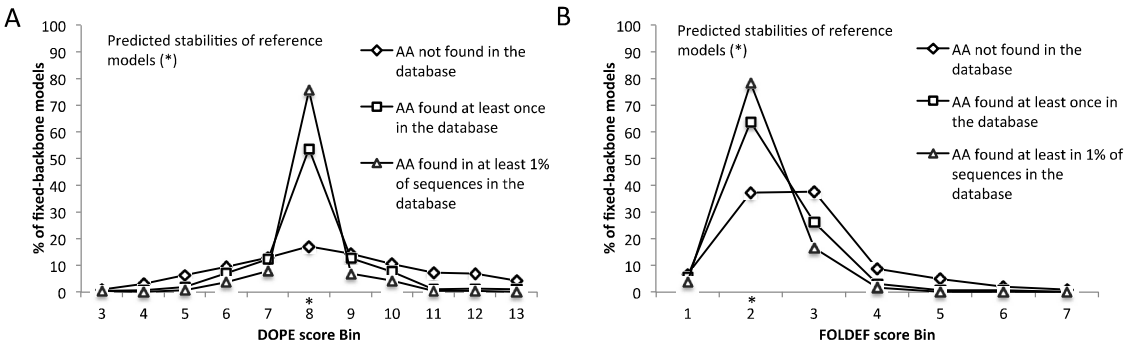

Figure 4.4 Fixed-backbone models of observed mutations were predicted to have stabilities similar to wild type models. Stabilities of the fixed backbone models predicted by (A) DOPE and (B) FOLDEF. The score bin reflects the structural stability from higher to lower. '*' Indicates stability of reference structures. Only results from five higher, five lower and the reference model bins are shown, as they accounted for more than 98% of all models.

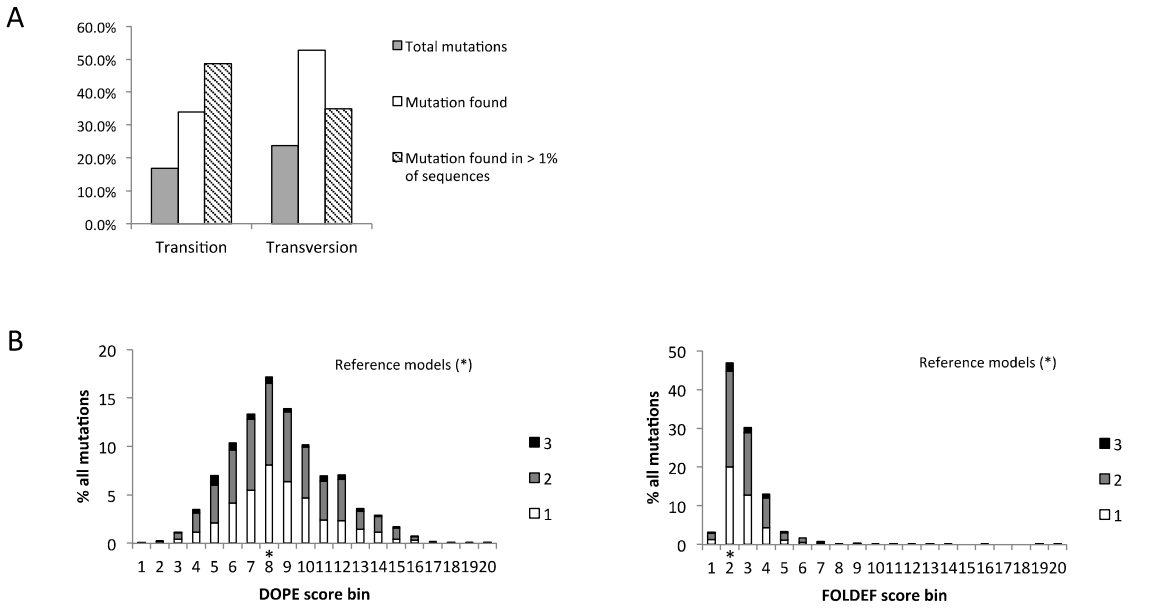

Figure 4.5 Genetic barriers of mutations and their stabilities (A) Division of mutations requiring only single nucleotide change into transitions and transversions. (B) Distribution of genetic barrier across all stability levels. Point mutations in each stability bin were classified based on the least number of nucleotide changes needed. Portion of mutations requiring one nucleotide change is shown in white, two in grey, and three in black, respectively.

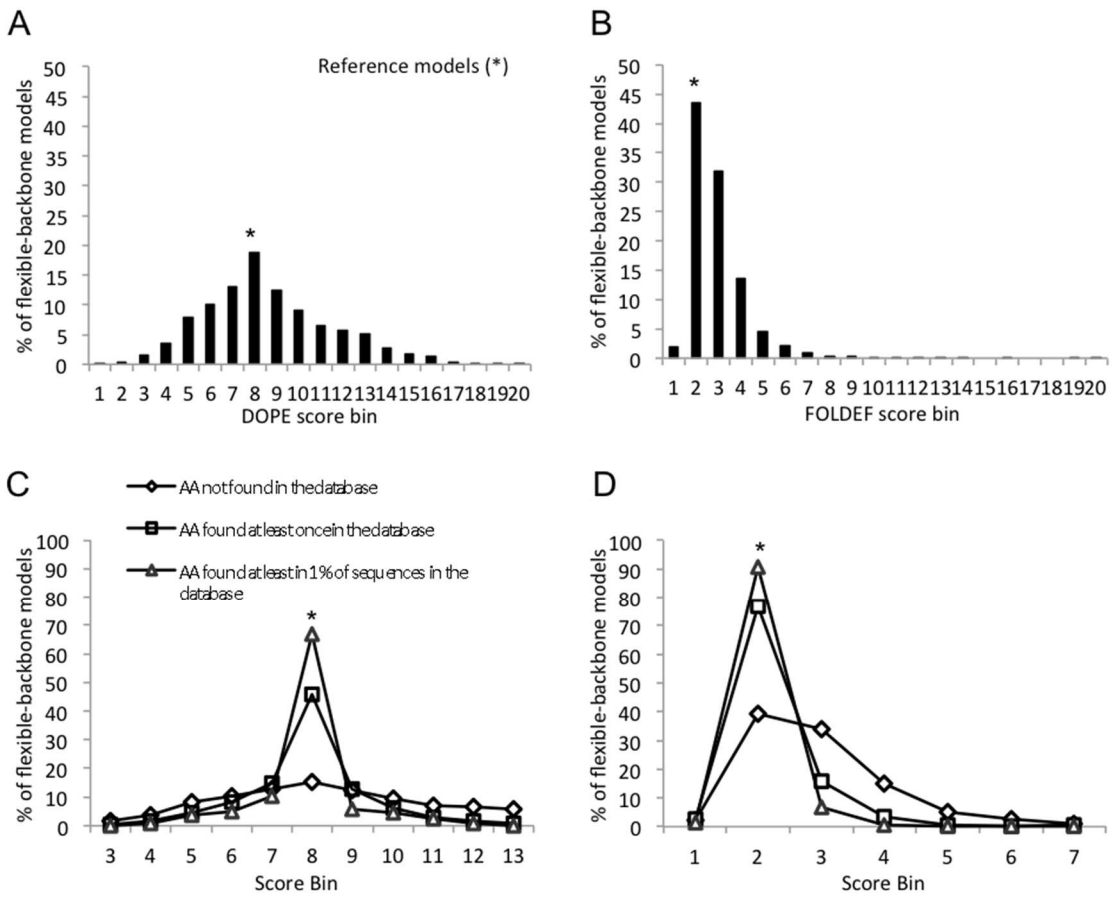

Figure 4.6. Distribution of capsid protein mutant stabilities based on flexible-backbone models. The score bin reflects the structural stability from higher (left) to lower (right) levels. '*' indicates the bin in which reference structures were found. All mutations predicted by DOPE (A,B) and FOLDEF (C,D) were classified into three groups based on their frequency in the HIV sequence database. Only results from five higher, five lower and the reference model bins are shown, as together they accounted for more than 98% of all models.

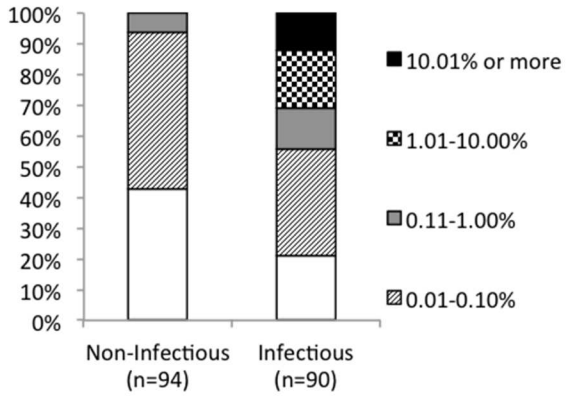

Figure 4.7. Mutations observed in a database of 5811 HIV-1 capsid sequences. Mutations resulting in non-infectious or infectious viruses are shown separately and stratified based on their database frequencies. White boxes represent the percentage of mutations not found in the database.

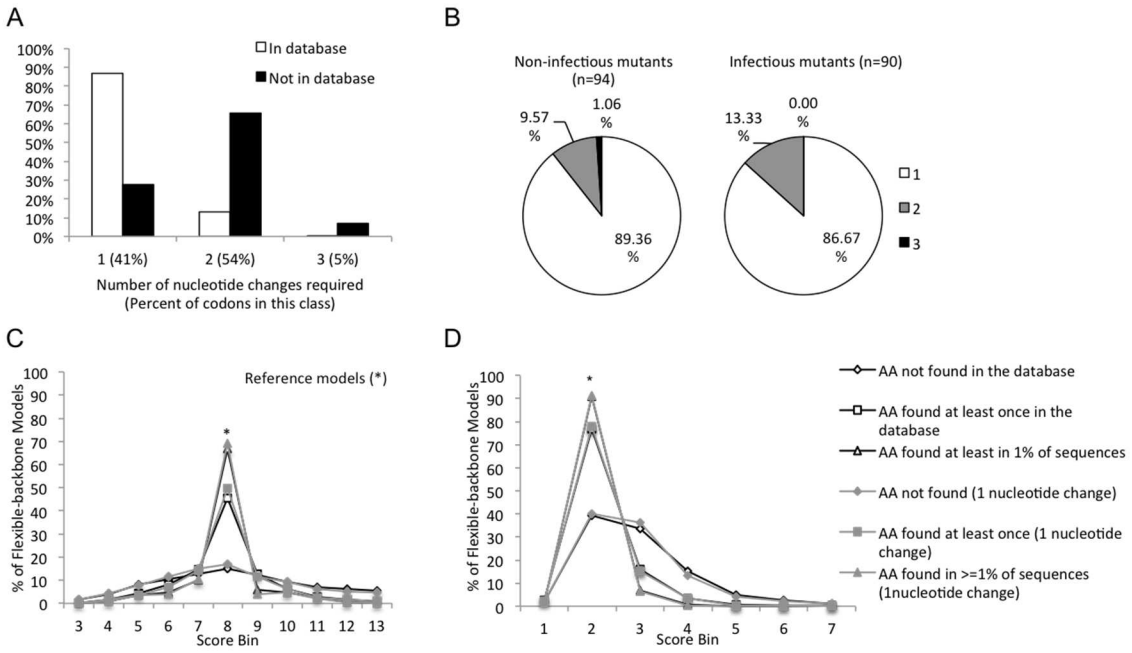

Figure 4.8. Genetic barrier influences the emergence of amino acid mutations but not impact of mutations on viral infectivity. (A) Number of nucleotide changes required for mutations observed and not observed in the HIV database of 5811 sequences. (B) Number of nucleotide changes required for mutations whose impact on infectivity was tested. AA mutations classified into three groups based on their frequency in the HIV database with stability scores predicted by DOPE (C) and FOLDEF (D).

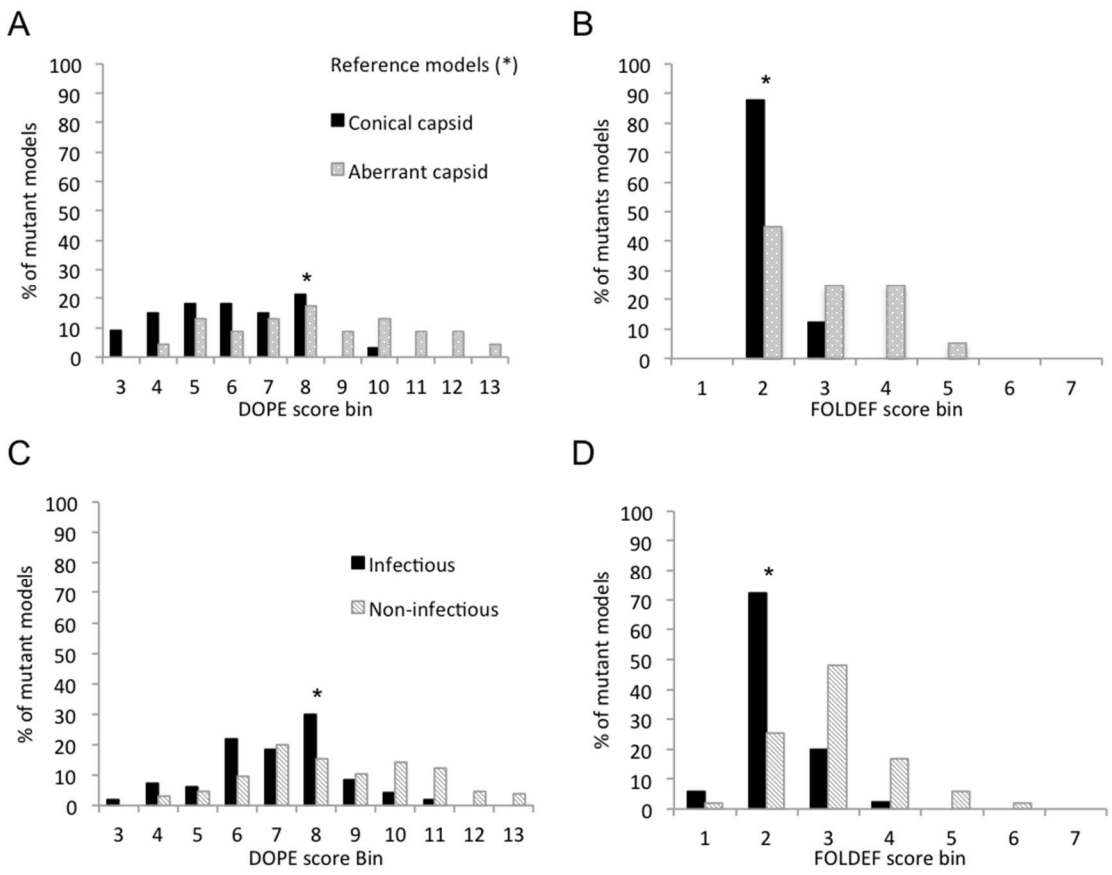

Figure 4.9. Predicted stabilities of mutations with known phenotypes. Flexible-backbone models were predicted by DOPE (A and C) and FOLDEF (B and D) and compared to capsid structure (A and B) and virus infectivity (C and D).

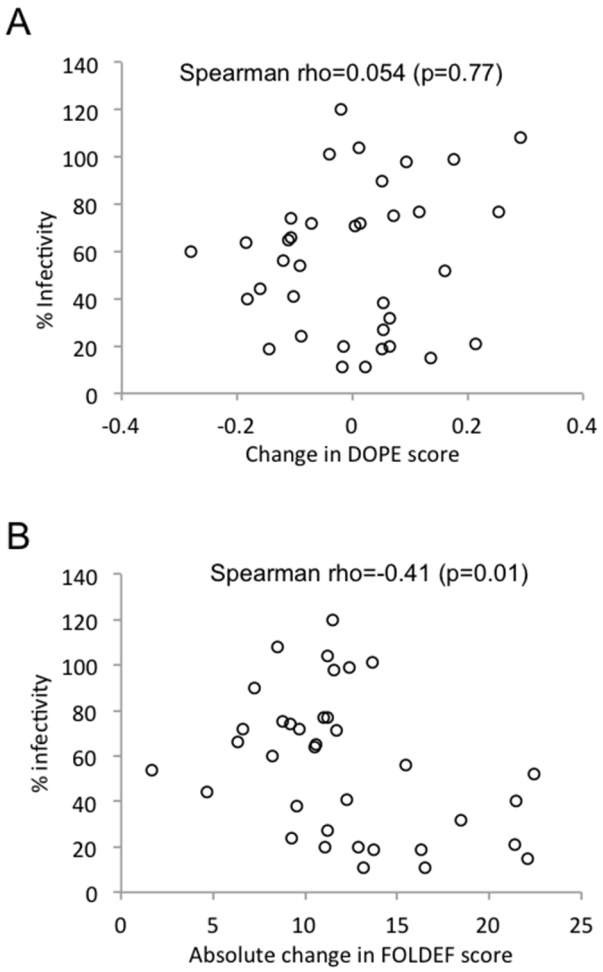

Figure 4.10 Relationship between viral infectivity and change in stability compared to the reference structure. As predicted by (A) DOPE and (B) FOLDEF. Each symbol represents a point mutation (viral infectivity data taken from [95]).

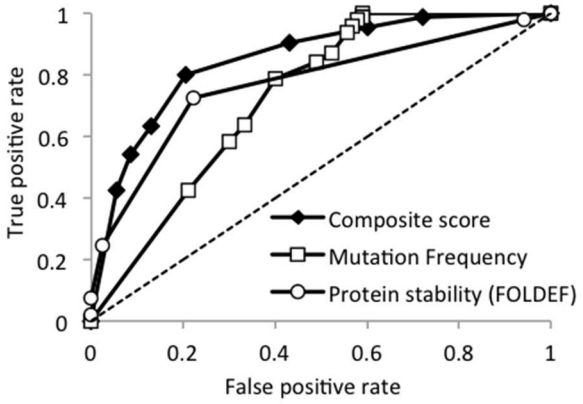

Figure 4.11 Receiver operating characteristic curve of HIV-1 subtype B non-infectious mutations predictions. The composite score is the sum of the FOLDEF-score rank and the mutation-frequency rank.

Table 4.1. Accuracy of using mutation frequency or change in structural stability to predict viral infectivity in binary classification manner

Predictor	Sensitivity[d]	Specificity[e]	Precision[f]	Accuracy[g]
Mutation frequency[a]	59.5%	70.0%	67.10%	64.10%
Stability of reference models[b]	72.34%	77.78%	77.30%	75.00%
Composite Score[c]	80.00%	79.57%	78.16%	79.78%

[a] Mutations with the database frequency of 0.02% or less are predicted to result in non-infectious virus and vice versa

[b] Mutations with structural stability higher than the reference models are predicted to result in non-infectious virus and vice versa

[c] Mutations with the composite score (sum of ranks of frequency and stability scores) higher than 175 are predicted to result in non-infectious virus and vice versa

[d] Sensitivity = (True positive)/(True positive + False negative)

[e] Specificity = (True negative)/(True negative + False positive)

[f] Precision = (True positive)/(True positive + False positive)

[g] Accuracy = (True positive + True negative)/(True positive + True negative + False positive + False negative)

Table 4.2. Amino acid sites prone to destabilizing mutations

Site (HXB2)	Consensus AA	Frequency	Structural element and location in capsid hexamer and/or CTD dimer	Solvent accessibility surface area (A^2)
2	ILE	0.98	β-hairpin	10.22
8	GLY	0.99	β-hairpin	40.98
20	LEU	0.99	Helix 1; capsid hexamerization (NTD-NTD) interface	2.1
23	TRP	0.99	Helix1	2.81
32	PHE	0.99	β turn in loop region between helix 1 and 2	5.97
36	VAL	0.97	Helix 2	0.65
37	ILE	0.97	Helix 2	17.74
40	PHE	0.99	Helix 2	0
43	LEU	0.98	Helix 2; capsid hexamerization (NTD-NTD) interface	13.74
49	PRO	0.99	Helix 3	0
52	LEU	0.99	Helix 3	0.02
55	MET	0.99	Helix 3	0.69
56	LEU	0.99	Helix 3	12.75
65	ALA	0.99	Helix 4	0
66	MET	0.99	Helix 4	13.09
69	LEU	0.99	Helix 4	2.91
73	ILE	0.99	Helix 4	13.8
80	TRP	0.99	Helix 4	40.65
99	PRO	0.99	Loop between helix 4 and 5	8.6
101	GLY	0.99	Helix 5	0.1
104	ILE	0.99	Helix 5	0
106	GLY	0.99	β turn in loop region between helix 5 and 6	18.65
109	SER	0.99	β turn in loop region between helix 5 and 6	13.19
111	LEU	0.99	Helix 6	53.08
117	TRP	0.99	Helix 6	0.24
126	VAL	0.99	Helix 7	0
133	TRP	0.99	Helix 7	10.88
134	ILE	0.99	Helix 7	0
138	LEU	0.98	Helix 7	0.66
141	ILE	0.99	Helix 7	0.22
142	VAL	0.99	Helix 7	0.84
144	MET	0.99	Helix 7	20.6

150	ILE	0.99	Linker region between NTD and CTD; CTD dimerization interface	23.25
153	ILE	0.99	Linker region between NTD and CTD; CTD dimerization interface	11.75
161	PHE	0.99	Helix 8; MHR	6.88
165	VAL	0.99	Helix 8; MHR; capsid hexamerization (NTD-CTD) interface	0
168	PHE	0.99	Helix 8; MHR	0.12
169	TYR	0.99	Helix 8; MHR; capsid hexamerization (NTD-CTD) interface	30.73
172	LEU	0.99	Helix 8; MHR; CTD dimerization interface	5.4
189	LEU	0.99	Helix 9; CTD dimerization interface	24.94
198	CYS	0.99	Helix 10	12.6
202	LEU	0.99	Helix 10	7.67
205	LEU	0.98	Helix 10	77.49
206	GLY	0.99	β turn in loop region between helix 10 and 11	28.12
211	LEU	0.99	Helix 11; capsid hexamerization (NTD-CTD) interface	4.46

CHAPTER 5

Discussion

One very challenging aspect of developing an HIV vaccine or therapy is to overcome the high sequence diversity and evolutionary rate of the virus. HIV is notorious for rapidly acquiring drug resistance and immune escape mutations that allow the virus to survive and persist against the clinical treatment and host immune responses. The use of combination therapy consisting of three or more antiretroviral drugs is advised to prevent the emergence of drug resistance mutations and achieve near total suppression of viral replication and spread. Nevertheless, the cost and short- and long-term side effects of antiretroviral therapy (ART) remain disconcerting issues [234-237]. A therapeutic vaccine that provides long-term control of viral replication is a desirable alternative.

Focusing vaccine-induced immune responses on functionally and structurally conserved elements of the viral proteome has been proposed as a strategy to battle this highly diverse and rapidly evolving virus [96, 205-208]. Gag and Pol polypeptides are relatively highly conserved, compared to the rest of the viral proteome. Rolland et al. and others [96, 194, 203, 204] have hypothesized that mutations in the conserved regions would impose a fitness cost on the virus, impair viral replication and, thus, hamper viral spread in infected hosts. CTL escape mutations and drug resistance mutations shown to have fitness cost were often accompanied or followed by compensatory mutations, which restore the viral fitness [120, 169, 175, 209, 228-230]. Conceivably, CTL responses against multiple target sites, whose escape mutations have high fitness cost, would hinder the occurrence of all necessary compensatory mutations.

To confirm this hypothesis, identification of structurally and/or functionally important amino acid sites with limited tolerance for mutations is essential. Sequence

conservation is considered to be a surrogate for the importance of a particular site and, hence, was deemed to be a predictor of how deleterious mutations are likely to be at different sites. Intriguingly, results from recent *in vitro* and *ex vivo* studies had identified only weak relationships between sequence conservation and the impact of mutations on viral replication fitness. Mutations at highly conserved sites exhibited various degree of fitness loss: from large, as would be expected, to unexpectedly negligible [95, 147, 152, 161]. These observations support the notion that viral replication fitnesses measured using *in vitro* and *ex vivo* viral fitness assay are not entirely reflective of fitness *in vivo*. Currently available assays do not encompass all biological factors and selective pressures presented in host organisms, such as host immune responses. Nonetheless, fitness costs detected in primary cell cultures that represent fundamental defects in viral replication are likely to persist *in vivo* regardless of host cell-specific environments. Previous studies have found associations between the *in vitro* and/or *ex vivo* viral fitness and disease progression, clinical outcomes and emergence of drug resistant mutations [123-126, 131-136]. Therefore, the amino acid sites with low tolerance for mutations or whose mutations have large fitness cost, as identified by *in vitro* and *ex vivo* viral fitness assays, are target sites for vaccine-induced immune responses.

A variety of viral-cell culture assays have been developed to measure levels of viral infectivity and rates of viral replication. The assay outcome for a given HIV strain is typically reported in relation to the same measurement obtained from a wild type or reference strain, such as a percent infectivity or relative fitness. Among these assays, the pairwise growth competition assay provides the most sensitive and reliable measurement of the fitness difference, as two viral strains are competed directly in the same cell culture [140, 141]. Several experimental protocols have been developed for growth competition assays. These protocols utilize different experimental conditions including

but not limited to: viral input MOI, types of host cells, time period of cell culture, sampling time points, methods for measuring viral growth and techniques for determining viral ratios [143-146, 151].

Our previous study [148] showed that the viral input MOI affects the consistency of the assay outcome, the relative fitness value. MOIs lower than 0.001 yield more variable results. However, the higher the MOI the faster the virus exhaust the available target cells and the shorter the period of exponential viral growth. Shorter growth phases limit the number of sampling time points, which could affect the robustness of assay results. In addition to input MOIs, the type of host cell also influences viral growth kinetics. Hence, the input MOI must be chosen carefully based on the viral growth kinetics observed in the same host cells used in the growth competition assay. The type of host cell that should be used in growth competitions is a subject of debate. Primary cell cultures are considered to better resemble host cell environments *in vivo* than cell lines. However, our previous study [147] demonstrated that there was a highly significant positive relationship between the fitness cost of HIV-1 CA mutants determined in PBMCs and the CEMx174 T cell line, suggesting that this cell line can be used as a substitute for PBMC.

Another important component of the growth competition assay is the method used to determine viral growth and virus ratios. The number of infected cells, viral CA production and RT activity are commonly used surrogates for viral growth, but these can only be use to measure combined growth of two viruses in the same cell culture. The growth of one virus cannot be differentiated from the growth of the other. Instead, several techniques have been developed to distinguish the two viruses and determine virus ratios based on the difference in their DNA or RNA sequence [143-146, 151]. Among these techniques, only quantitative real-time PCR (qPCR) offers the ability to measure the growth of both viruses as well as ratio simultaneously. However, the qPCR primer

and probe set must be optimized and validated for each specific DNA region to yield accurate results. It is not a very cost-effective method unless the same set of primers and probes can be applied with multiple different viral strains. It is suitable for studying HIV molecular clones, which are amendable to genetic manipulation. A sequence tag, such as synonymous mutations that have no effect on viral replication, can be engineered into a mutant molecular clone allowing a mutant virus to be distinguished from a reference virus that contains no sequence tag [151, 170-172]. For a study of biologically cloned HIV isolates from infected individuals, where the genetic engineering of a sequence tag into the viral genome is not applicable, direct sequencing methods, such as bulk Sanger sequencing [149, 168], are more cost-effective approaches for measuring viral ratios, though, an additional assay, such as qPCR, is required to monitor viral growth in parallel to the viral ratio.

Taking results from previous studies [147, 148, 151], I developed a guideline for determining viral replication fitness was developed (Chapter 2). This guideline encompasses optimized protocols for every necessary step, starting from constructing sequence tagged HIV-1 molecular clones to carrying out viral growth competition assays and calculating relative fitness. Although the growth competition assay protocol was optimized using molecularly cloned viruses, they are readily applicable to biologically cloned viruses. Two viral ratio detection methods, qPRC and peak-height chromatogram, are included. The results, or the relative fitness, obtained from these two methods were comparable, suggesting that they can be used interchangeably. These protocols were used successfully to investigate the fitness cost of HIV-1 CA mutations in different genetic backgrounds (Chapter 3, [147, 233]).

Previous mutagenesis studies had identified several deleterious mutations in HIV-1 CA protein-protein interface residues [63, 94, 188, 189]. Still, only one mutation was

studied at most sites. Moreover, these mutations were mostly non-conservative amino acid changes that are not known to occur in HIV sequences. It is unclear whether there exist other mutations with smaller or no fitness cost at these sites. It is impractical to study all possible mutations at a particular site *in vitro*. One alternative is to study the impact of the most frequently observed mutations, which likely reflects the minimal cost of amino acid changes at the site. This approach has a few potential pitfalls. First, the identity of the most frequent mutations can be ambiguous and vary from one sequence dataset to the others at highly conserved sites, as there often are a few amino acids appearing at about the same but very low frequencies. Second, the supposedly minimal fitness effect may be exaggerated, as the frequencies of independent mutations and co-evolving mutations are not distinguished and the study of single mutations does not account for the role of the compensatory mutations in restoring viral fitness. Last, but not least, the amino acid frequencies can vary between different HIV-1 subtypes, as shown in Table 3.4 of Chapter 3. For example, I observed a switch between the consensus amino acid and the most frequently observed mutation between subtype B and C sequences at 8 out of 30 studied sites. As HIV sequence diversity continues to increase, the degree of amino acid variation is likely to rise. The subtype of the viral backbone must be kept in consideration when planning and interpreting results from mutagenesis and viral replication fitness studies.

The observed fitness effect of a mutation can differ vastly, from a lethal to a small negative impact, in different genetic backgrounds due to compensatory mutations [233]. Nevertheless, such a disparity was observed in only 1 of 10 HIV-1 CA mutations studied. To further explore the likelihood and the degree of variation in fitness effects, I introduced the same 19 CA mutations, not overlapping with the previous study, into two different HIV-1 strains, COTM-p24 (Chapter 3) and COTB-p24 [147] and the fitness costs

assessed using the same protocols (Chapter 2) in the same T-cell line, CEMx174. Overall, a strong positive correlation of fitness effects was observed (Figure 5.1). However, similar to the previous study [233], 2 of 19 mutations showed very different levels of fitness cost in different HIV-1 strains. These results suggest that the generalization of fitness effects observed in a study should be considered carefully. Amino acid co-variation studies can be used to identify potential compensatory mutations for mutations of interest. The fitness effect of co-varying mutations can be evaluated by additional viral fitness studies [146]. Nonetheless, the fitness effects of the majority of mutations studied were relatively consistent. Therefore, evaluating the fitness effect of the most frequent mutations will provide useful guideline for the purpose of identifying the potential vaccine target sites and plausible drug/immune escape mutations.

In addition to sequence conservation, protein structures provide clues for functionally and structurally important residues. X-ray crystal structures of HIV-1 CA hexamers, which is the major structural unit of the viral capsid, yields insights into the CA-CA interactions within the hexamer. CA-CA interactions found in the dimer of the carboxy terminal domain (CTD) of CA dimer were speculated to resemble the protein-protein interactions between the CA hexamers [102]. These interactions are likely to be crucial for stabilization of the CA hexamer and the overall capsid structure and, therefore, are potential targets for HIV vaccines and therapies. Interface residues found in the CA hexamer and CA-CTD dimer, account for about a quarter of the CA (55 out of 231) and overall were as conserved as the rest of the protein. Although interface residues can be identified from high-resolution crystal structures, the impact of mutations at these residues on viral replication remains unclear.

CA-CA interactions within the CA hexamer consist mainly of hydrophobic clusters and a small number, 3 to 6, of hydrogen bonds [102]. To pinpoint the crucial interactions

with low tolerance for mutations, the fitness costs of the most frequent mutations were determined at twelve interface sites by my optimized pairwise growth competition assay (Chapter 3). Mutations that disrupted hydrogen bonds at the NTD-CTD interface resulted in non-infectious viruses, while mutations of hydrophobic residues showed varying levels of deleterious impacts on viral replication. In comparison, the most frequently observed mutations at the non-interface sites had, on average, less negative effects on viral replication (Figure 3.7**)**. Similar results were observed when analyzing data from another random mutagenesis study **(**Figure 5.2**)** using a different HIV-1 strain [95]. These observations suggest that CA-CA interactions are potential target sites for novel HIV therapies. In particular, the hydrogen bonds between the residual pair N189-R305 and V191-R305, and the hydrophobic contacts involving residue L152 and F169 should be considered. Additional viral fitness studies on other interface sites not included in previous studies [63, 94, 95, 161, 188, 189] may reveal more potential target sites.

In vitro methods can be lengthy and resource intensive. To facilitate selection of sites for further studies, several computational approaches have been developed to predict the impact of mutations on protein function, structural stability and organism phenotype [155-158]. Several tools have been developed to predict important sites based on protein sequence conservation. However, previous studies detected no relationship between replication fitness cost of HIV-1 mutations and sequence conservation of the mutated sites [152, 161]. By contrast, recent studies have shown that structure-based predictions may be useful for identifying tolerated mutations in HIV-1 RT, PR and MA [159, 160]. Furthermore, another study reported an association between fitness costs of single amino acid changes in HIV-1 Gag and the absolute change in predicted protein stability of the CA hexamer [95, 161]. Capsid stability has been shown to be important for HIV replication [63]. Mutations that alter the structural stability of the CA hexamer are also likely

to affect overall capsid formation, structure and stability. These findings suggest that structure-based predictions of protein stability might be useful screening tools for identifying potentially important amino acid sites in the CA prior to *in vitro* or *ex vivo* studies. To expand on the previous study, which evaluated changes at 29 of 500 residues of HIV-1 Gag [95, 161], and evaluate the feasibility of this approach, mutant protein structures for all possible single amino acid changes in the CA were generated *in silico,* except for the last 13 residues which were missing from the currently available HIV-1 CA crystal structures. The stabilities of the mutant proteins were then predicted using two different methods, FOLDEF [155] and DOPE [238]. FOLDEF was also used in the previous study [161].

Overall, the two methods yielded moderately correlated predictions of protein stability (Figure 4.1 and 4.2). Interestingly, the mutations that were observed in more than 1% of the HIV-1 CA sequences were predicted to have similar structural stabilities to the reference CA structures by both methods. On the other hand, other CA mutations exhibited larger stability deviations from the references (Figure 4.6). These results suggest that CA protein stability is a good predictor of tolerated mutations. However, only the FOLDEF stability scores correlated well with viral phenotypes (i.e., capsid morphology and infectivity) of CA mutations. Destabilizing mutations were associated with the aberrant capsid shape and non-infectious virus (Figure 4.9). Based on the FOLDEF score, deleterious versus neutral mutations could be classified with 75% accuracy (Table 4.1). We find that 45 residues, one-fifth of the CA, are prone to destabilizing mutations. These forty-five residues were very highly conserved, with consensus amino acid frequencies ranging between 0.98 and 0.99 in subtype HIV-1 subtype B. Most of these residues were located in secondary structure elements of the protein with small solvent accessible surface areas - many had side-chains that were

almost completely buried (Table 4.2). Such low accessibility may pose a problem for the development of drugs targeting these residues, but not for inducing CTL responses.

In silico methods allow very large numbers of mutations to be investigated efficiently. Structural based *in silico* approaches are more computationally intensive than sequence-based approaches. Nonetheless, the former provided a better predictor of the effect of mutations in CA than mutation database frequency (Table 4.1). Moreover, it is difficult to define the effective sequence conservation threshold for predicting the outcome of mutations, as the virus continues to evolve and adapt to the host cellular environment and immune system. On the other hand, protein structure and function are more conserved than sequence. The effect of mutations on CA structural stability is unlikely to change drastically. More importantly, a simple composite ranking score that considers both sequence conservation and protein structure stability improved the accuracy of predicting deleterious mutations (Table 4.1 an Figure 4.11). These results support the notion that protein stability prediction methods, when combined with sequence frequency data, can be a useful tool for identifying candidate targets for drug and vaccine development.

Still, higher accuracy in predicting deleterious mutations is desirable. As HIV continues to evolve and viral sequence data accumulates, analyses of a larger and more comprehensive sequence sets will shed more light on sites that remain conserved despite host immune pressures. From a structure stability perspective, a mutation may disrupt capsid assembly or immature capsid formation instead of capsid maturation or mature capsid structure. Stability analysis of the immature CA may provide insight into additional important sites with low tolerance for mutations, though such analyses requires a high-resolution structure of HIV-1 immature capsid, which is not yet available. Determining fitness and protein stability effects of multiple mutations may also help

improve predictions, as compensatory mutations can arise during viral infection that can restore protein stability [231, 232]. However, the selection of mutations to be studied in such compensation studies must be done carefully as number of combinations increases exponentially with the number of mutations. One approach is to rely on the co-evolving residues or coupled mutation analyses [146, 200, 239].

In conclusion, using the optimized *in vitro* pairwise growth competition assay, I evaluated the fitness cost of the most frequently found mutations at 30 sites in HIV-1 CA. Overall, mutations at the CA hexamerization interface had significantly higher fitness cost and were more likely to be lethal than mutations at non-interface sites. Similar results were obtained when comparing the effect of random mutations at the CA interface and non-interface sites on viral replication in another HIV-1 strain. Hydrogen bonds at the inter-subunit NTD-CTD interface of the CA hexamer were shown to be particularly important. As *in vitro* laboratory studies are resource and time intensive, computational screening tools for identifying potentially important amino acid sites can be extremely helpful. Incorporating sequence information with *in silico* mutation and protein stability predictions, I identified potentially important sites that warrant further investigation as candidates for drug and CTL vaccine targets.

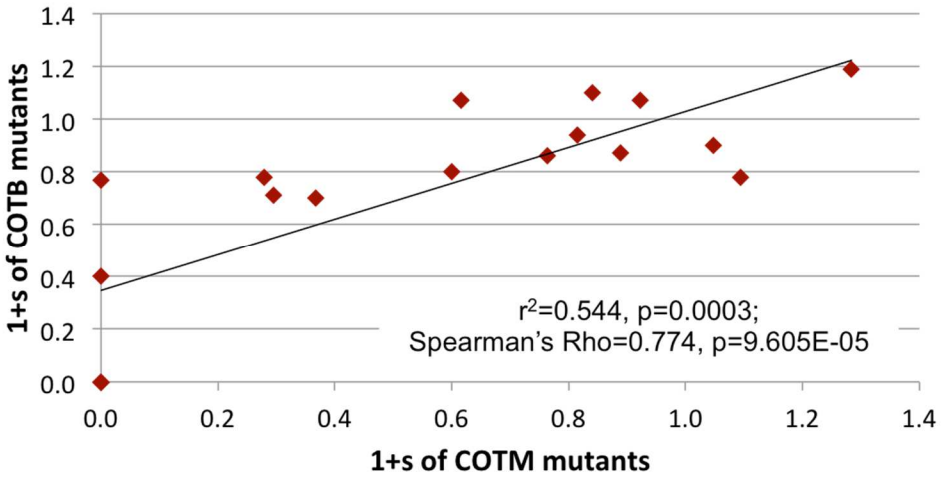

Figure 5.1 Correlation of fitness effects of nineteen single amino acid changes in two HIV-1 CA genetic backgrounds: COTB and COTM.

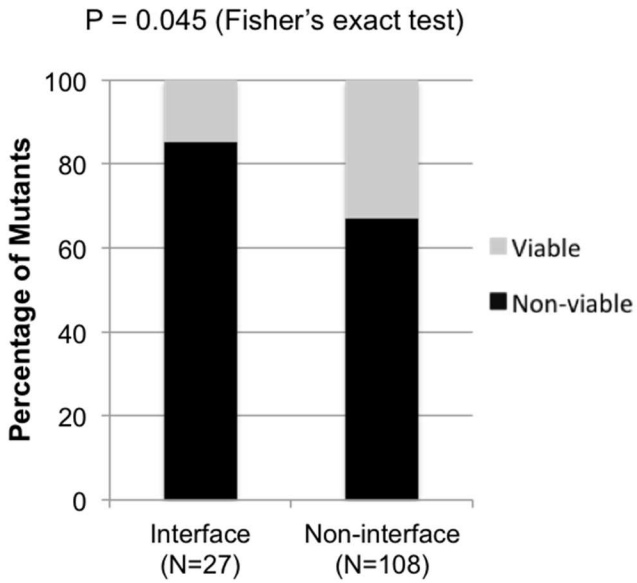

Figure 5.2 Viability of interface and non-interface mutants (data taken from [95])

BIBLIOGRAPHY

1. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med.* Sep 2011;1(1):a006841.

2. Tazi J, Bakkour N, Marchand V, Ayadi L, Aboufirassi A, Branlant C. Alternative splicing: regulation of HIV-1 multiplication as a target for therapeutic action. *FEBS J.* Feb 2010;277(4):867-876.

3. Garcia JA, Harrich D, Soultanakis E, Wu F, Mitsuyasu R, Gaynor RB. Human immunodeficiency virus type 1 LTR TATA and TAR region sequences required for transcriptional regulation. *EMBO J.* 1989;8:765-778.

4. Wei P, Garber ME, Fang SM, Fischer WH, Jones KA. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell.* Feb 20 1998;92(4):451-462.

5. Weiss R, Teich N, Varmus H, Coffin J. *Molecular biology of tumor viruses: RNA tumor viruses*: Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; 1982.

6. McBride MS, Panganiban AT. Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation in vivo. *J. Virol.* 1997;71:2050-2058.

7. McBride MS, Schwartz MD, Panganiban AT. Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis-elements required for encapsidation. *J. Virol.* 1997;71:4544-4554.

8. Skripkin E, Paillart JC, Marquet R, Ehresmann B, Ehresmann C. Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. *Proc Natl Acad Sci U S A.* May 24 1994;91(11):4945-4949.

9. Berkhout B, van Wamel JL. Role of the DIS hairpin in replication of human immunodeficiency virus type 1. *J Virol.* Oct 1996;70(10):6723-6732.

10. Nikolaitchik OA, Dilley KA, Fu W, et al. Dimeric RNA recognition regulates HIV-1 genome packaging. *PLoS Pathog.* Mar 2013;9(3):e1003249.

11. Vaishnav YN, Wong-Staal F. The biochemistry of AIDS. *Annu Rev Biochem.* 1991;60:577-630.

12. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature.* Jan 21 1988;331(6153):280-283.

13. Malim MH, Cullen BR. Rev and the fate of pre-mRNA in the nucleus: implications for the regulation of RNA processing in eukaryotes. *Mol Cell Biol.* Oct 1993;13(10):6180-6189.

14. Fischer U, Pollard VW, Luhrmann R, et al. Rev-mediated nuclear export of RNA is dominant over nuclear retention and is coupled to the Ran-GTPase cycle. *Nucleic Acids Res.* Nov 1 1999;27(21):4128-4134.

15. D'Agostino DM, Felber BK, Harrison JE, Pavlakis GN. The Rev protein of human immunodeficiency virus type 1 promotes polysomal association and translation of gag/pol and vpu/env mRNAs. *Mol Cell Biol.* Mar 1992;12(3):1375-1386.

16. Freed EO. HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology.* Nov 10 1998;251(1):1-15.

17. Gottlinger HG, Sodroski JG, Haseltine WA. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc Natl Acad Sci U S A.* Aug 1989;86(15):5781-5785.

18. Yu X, Yuan X, Matsuda Z, Lee TH, Essex M. The matrix protein of human immunodeficiency virus type 1 is required for incorporation of viral envelope protein into mature virions. *J Virol.* Aug 1992;66(8):4966-4971.

19. Ako-Adjei D, Johnson MC, Vogt VM. The retroviral capsid domain dictates virion size, morphology, and coassembly of gag into virus-like particles. *J Virol.* Nov 2005;79(21):13463-13472.

20. Briggs JA, Riches JD, Glass B, Bartonova V, Zanetti G, Krausslich HG. Structure and assembly of immature HIV. *Proc Natl Acad Sci U S A.* Jul 7 2009;106(27):11090-11095.

21. Krausslich HG, Facke M, Heuser AM, Konvalinka J, Zentgraf H. The spacer peptide between human immunodeficiency virus capsid and nucleocapsid proteins is essential for ordered assembly and viral infectivity. *J Virol.* Jun 1995;69(6):3407-3419.

22. Accola MA, Hoglund S, Gottlinger HG. A putative alpha-helical structure which overlaps the capsid-p2 boundary in the human immunodeficiency virus type 1 Gag precursor is crucial for viral particle assembly. *J Virol.* Mar 1998;72(3):2072-2078.

23. Datta SA, Temeselew LG, Crist RM, et al. On the role of the SP1 domain in HIV-1 particle assembly: a molecular switch? *J Virol.* May 2011;85(9):4111-4121.

24. Berkowitz R, Fisher J, Goff SP. RNA packaging. *Curr. Top. Microbiol. Immunol.* 1996;214:177-218.

25. Feng YX, Copeland TD, Henderson LE, et al. HIV-1 nucleocapsid protein induces "maturation" of dimeric retroviral RNA in vitro. *Proc Natl Acad Sci U S A.* Jul 23 1996;93(15):7577-7581.

26. Tsuchihashi Z, Brown PO. DNA strand exchange and selective DNA annealing promoted by the human immunodeficiency virus type 1 nucleocapsid protein. *J Virol.* Sep 1994;68(9):5863-5870.

27. Darlix JL, Vincent A, Gabus C, de Rocquigny H, Roques B. Trans-activation of the 5' to 3' viral DNA strand transfer by nucleocapsid protein during reverse transcription of HIV1 RNA. *C R Acad Sci III.* Aug 1993;316(8):763-771.

28. Feng YX, Campbell S, Harvin D, Ehresmann B, Ehresmann C, Rein A. The human immunodeficiency virus type 1 Gag polyprotein has nucleic acid chaperone activity: possible role in dimerization of genomic RNA and placement of tRNA on the primer binding site. *J Virol.* May 1999;73(5):4251-4256.

29. de Marco A, Heuser AM, Glass B, Krausslich HG, Muller B, Briggs JA. Role of the SP2 domain and its proteolytic cleavage in HIV-1 structural maturation and infectivity. *J Virol.* Dec 2012;86(24):13708-13716.

30. Garrus JE, von Schwedler UK, Pornillos OW, et al. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell.* 2001;107:55-65.

31. VerPlank L, Bouamr F, LaGrassa TJ, et al. Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). *Proc Natl Acad Sci U S A.* Jul 3 2001;98(14):7724-7729.

32. Strack B, Calistri A, Craig S, Popova E, Gottlinger HG. AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell.* Sep 19 2003;114(6):689-699.

33. Jenkins Y, Pornillos O, Rich RL, Myszka DG, Sundquist WI, Malim MH. Biochemical analyses of the interactions between human immunodeficiency virus type 1 Vpr and p6(Gag). *J Virol.* Nov 2001;75(21):10537-10542.

34. Kondo E, Mammano F, Cohen EA, Gottlinger HG. The p6gag domain of human immunodeficiency virus type 1 is sufficient for the incorporation of Vpr into heterologous viral particles. *J Virol.* May 1995;69(5):2759-2764.

35. Kohl NE, Emini EA, Schleif WA, et al. Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci U S A.* Jul 1988;85(13):4686-4690.

36. Le Grice SF, Mills J, Mous J. Active site mutagenesis of the AIDS virus protease and its alleviation by trans complementation. *EMBO J.* Aug 1988;7(8):2547-2553.

37. Engelman A, Mizuuchi K, Craigie R. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell.* Dec 20 1991;67(6):1211-1221.

38. Hallenberger S, Bosch V, Angliker H, Shaw E, Klenk HD, Garten W. Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature.* Nov 26 1992;360(6402):358-361.

39. Lasky LA, Nakamura G, Smith DH, et al. Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor. *Cell.* Sep 11 1987;50(6):975-985.

40. Kowalski M, Potz J, Basiripour L, et al. Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. *Science.* Sep 11 1987;237(4820):1351-1355.

41. Freed EO, Myers DJ, Risser R. Characterization of the fusion domain of the human immunodeficiency virus type 1 envelope glycoprotein gp41. *Proc Natl Acad Sci U S A.* Jun 1990;87(12):4650-4654.

42. Hwang SS, Boyle TJ, Lyerly HK, Cullen BR. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science.* Jul 5 1991;253(5015):71-74.

43. Freed EO, Delwart EL, Buchschacher GL, Jr., Panganiban AT. A mutation in the human immunodeficiency virus type 1 transmembrane glycoprotein gp41 dominantly interferes with fusion and infectivity. *Proc Natl Acad Sci U S A.* Jan 1 1992;89(1):70-74.

44. Ott M, Geyer M, Zhou Q. The control of HIV transcription: keeping RNA polymerase II on track. *Cell Host Microbe.* Nov 17 2011;10(5):426-435.

45. Karn J, Stoltzfus CM. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med.* Feb 2012;2(2):a006916.

46. Malim MH, Bieniasz PD. HIV Restriction Factors and Mechanisms of Evasion. *Cold Spring Harb Perspect Med.* May 2012;2(5):a006940.

47. Dube M, Bego MG, Paquay C, Cohen EA. Modulation of HIV-1-host interaction: role of the Vpu accessory protein. *Retrovirology.* 2010;7:114.

48. Guenzel CA, Herate C, Benichou S. HIV-1 Vpr-a still "enigmatic multitasker". *Front Microbiol.* 2014;5:127.

49. Kirchhoff F, Greenough TC, Brettler DB, Sullivan JL, Desrosiers RC. Brief report: absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection. *N Engl J Med.* Jan 26 1995;332(4):228-232.

50. Deacon NJ, Tsykin A, Solomon A, et al. Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients. *Science.* Nov 10 1995;270(5238):988-991.

51. Roeth JF, Williams M, Kasper MR, Filzen TM, Collins KL. HIV-1 Nef disrupts MHC-I trafficking by recruiting AP-1 to the MHC-I cytoplasmic tail. *J Cell Biol.* Dec 6 2004;167(5):903-913.

52. Leonard JA, Filzen T, Carter CC, Schaefer M, Collins KL. HIV-1 Nef disrupts intracellular trafficking of major histocompatibility complex class I, CD4, CD8, and CD28 by distinct pathways that share common elements. *J Virol.* Jul 2011;85(14):6867-6881.

53. Sattentau QJ, Weiss RA. The CD4 antigen: physiological ligand and HIV receptor. *Cell.* Mar 11 1988;52(5):631-633.

54. Deng H, Liu R, Ellmeier W, et al. Identification of a major co-receptor for primary isolates of HIV-1. *Nature.* Jun 20 1996;381(6584):661-666.

55. Feng Y, Broder C, Kennedy P, Berger E. HIV-1 entry cofactor: functional cDNA cloning of seven transmembrane, G protein-coupled receptor. *Science.* 1996;272:872-877.

56. Sattentau QJ, Moore JP. Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *J Exp Med.* Aug 1 1991;174(2):407-415.

57. Trkola A, Dragic T, Arthos J, et al. CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature.* Nov 14 1996;384(6605):184-187.

58. Wu L, Gerard NP, Wyatt R, et al. CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature.* Nov 14 1996;384(6605):179-183.

59. Matreyek KA, Engelman A. Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. *Viruses.* Oct 2013;5(10):2483-2511.

60. Roa A, Hayashi F, Yang Y, et al. RING domain mutations uncouple TRIM5alpha restriction of HIV-1 from inhibition of reverse transcription and acceleration of uncoating. *J Virol.* Feb 2012;86(3):1717-1727.

61. Shi J, Zhou J, Shah VB, Aiken C, Whitby K. Small-molecule inhibition of human immunodeficiency virus type 1 infection by virus capsid destabilization. *J Virol.* Jan 2011;85(1):542-549.

62. Schaller T, Ocwieja KE, Rasaiyaah J, et al. HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathog.* Dec 2011;7(12):e1002439.

63. Forshey BM, von Schwedler U, Sundquist WI, Aiken C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J Virol.* Jun 2002;76(11):5667-5677.

64. Ambrose Z, Aiken C. HIV-1 uncoating: connection to nuclear entry and regulation by host proteins. *Virology.* Apr 2014;454-455:371-379.

65. Hu WS, Hughes SH. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med.* Oct 2012;2(10).

66. Siliciano RF, Greene WC. HIV latency. *Cold Spring Harb Perspect Med.* Sep 2011;1(1):a007096.

67. Checkley MA, Luttge BG, Freed EO. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *J Mol Biol.* Jul 22 2011;410(4):582-608.

68. Wright ER, Schooler JB, Ding HJ, et al. Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. *EMBO J.* Apr 18 2007;26(8):2218-2226.

69. Ono A, Ablan SD, Lockett SJ, Nagashima K, Freed EO. Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the plasma membrane. *Proc Natl Acad Sci U S A.* Oct 12 2004;101(41):14889-14894.

70. Chukkapalli V, Hogue IB, Boyko V, Hu WS, Ono A. Interaction between the human immunodeficiency virus type 1 Gag matrix domain and phosphatidylinositol-(4,5)-bisphosphate is essential for efficient gag membrane binding. *J Virol.* Mar 2008;82(5):2405-2417.

71. Yu X, Yuan X, McLane MF, Lee TH, Essex M. Mutations in the cytoplasmic domain of human immunodeficiency virus type 1 transmembrane protein impair the incorporation of Env proteins into mature virions. *J Virol.* Jan 1993;67(1):213-221.

72. Cosson P. Direct interaction between the envelope and matrix proteins of HIV-1. *EMBO J.* Nov 1 1996;15(21):5783-5788.

73. Miyazaki Y, Garcia EL, King SR, et al. An RNA structural switch regulates diploid genome packaging by Moloney murine leukemia virus. *J Mol Biol.* Feb 12 2010;396(1):141-152.

74. Kleiman L, Jones CP, Musier-Forsyth K. Formation of the tRNALys packaging complex in HIV-1. *FEBS Lett.* Jan 21 2010;584(2):359-365.

75. Bachand F, Yao XJ, Hrimech M, Rougeau N, Cohen EA. Incorporation of Vpr into human immunodeficiency virus type 1 requires a direct interaction with the p6 domain of the p55 gag precursor. *J Biol Chem.* Mar 26 1999;274(13):9083-9091.

76. Onafuwa-Nuga AA, Telesnitsky A, King SR. 7SL RNA, but not the 54-kd signal recognition particle protein, is an abundant component of both infectious HIV-1 and minimal virus-like particles. *RNA.* Apr 2006;12(4):542-546.

77. Sundquist WI, Krausslich HG. HIV-1 assembly, budding, and maturation. *Cold Spring Harb Perspect Med.* Jul 2012;2(7):a006924.

78. Erickson-Viitanen S, Manfredi J, Viitanen P, et al. Cleavage of HIV-1 gag polyprotein synthesized in vitro: sequential cleavage by the viral protease. *AIDS Res Hum Retroviruses.* Dec 1989;5(6):577-591.

79. Swanstrom R, Wills JW. Synthesis, Assembly, and Processing of Viral Proteins. In: Coffin JM, Hughes SH, Varmus HE, eds. *Retroviruses.* Cold Spring Harbor (NY); 1997.

80. Pettit SC, Lindquist JN, Kaplan AH, Swanstrom R. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology.* 2005;2:66.

81. Peng C, Ho BK, Chang TW, Chang NT. Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity. *J Virol.* Jun 1989;63(6):2550-2556.

82. Goto T, Ikuta K, Zhang JJ, et al. The budding of defective human immunodeficiency virus type 1 (HIV-1) particles from cell clones persistently infected with HIV-1. *Arch Virol.* 1990;111(1-2):87-101.

83. Ashorn P, McQuade TJ, Thaisrivongs S, Tomasselli AG, Tarpley WG, Moss B. An inhibitor of the protease blocks maturation of human and simian immunodeficiency viruses and spread of infection. *Proc Natl Acad Sci U S A.* Oct 1990;87(19):7472-7476.

84. Ganser-Pornillos BK, Yeager M, Sundquist WI. The structural biology of HIV assembly. *Curr Opin Struct Biol.* Apr 2008;18(2):203-217.

85. Bharat TA, Davey NE, Ulbrich P, et al. Structure of the immature retroviral capsid at 8 A resolution by cryo-electron microscopy. *Nature.* Jul 19 2012;487(7407):385-389.

86. Schur FK, Hagen WJ, Rumlova M, et al. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 A resolution. *Nature.* Jan 22 2015;517(7535):505-508.

87. Guth CA, Sodroski J. Contribution of PDZD8 to stabilization of the human immunodeficiency virus type 1 capsid. *J Virol.* May 2014;88(9):4612-4623.

88. Li Y, Kar AK, Sodroski J. Target cell type-dependent modulation of human immunodeficiency virus type 1 capsid disassembly by cyclophilin A. *J Virol.* Nov 2009;83(21):10951-10962.

89. Sayah DM, Sokolskaja E, Berthoux L, Luban J. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature.* Jul 29 2004;430(6999):569-573.

90. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature.* Feb 26 2004;427(6977):848-853.

91. Stremlau M, Perron M, Lee M, et al. Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. *Proc Natl Acad Sci U S A.* Apr 4 2006;103(14):5514-5519.

92. Yap MW, Nisole S, Lynch C, Stoye JP. Trim5alpha protein restricts both HIV-1 and murine leukemia virus. *Proc Natl Acad Sci U S A.* Jul 20 2004;101(29):10786-10791.

93. Ganser-Pornillos BK, Chandrasekaran V, Pornillos O, Sodroski JG, Sundquist WI, Yeager M. Hexagonal assembly of a restricting TRIM5alpha protein. *Proc Natl Acad Sci U S A.* Jan 11 2011;108(2):534-539.

94. von Schwedler UK, Stray KM, Garrus JE, Sundquist WI. Functional surfaces of the human immunodeficiency virus type 1 capsid protein. *J Virol.* May 2003;77(9):5439-5450.

95. Rihn SJ, Wilson SJ, Loman NJ, et al. Extreme Genetic Fragility of the HIV-1 Capsid. *PLoS Pathog.* 2013;9(6):e1003461.

96. Rolland M, Nickle DC, Mullins JI. HIV-1 Group M Conserved Elements Vaccine. *PLoS Pathogens.* 2007;3(11):e157.

97. Brass AL, Dykxhoorn DM, Benita Y, et al. Identification of host proteins required for HIV infection through a functional genomic screen. *Science.* Feb 15 2008;319(5865):921-926.

98. König R, Zhou Y, Elleder D, et al. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell.* Oct 3 2008;135(1):49-60.

99. Krishnan L, Matreyek KA, Oztop I, et al. 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.* Jan 2010;84(1):397-406.

100. Price AJ, Fletcher AJ, Schaller T, et al. CPSF6 defines a conserved capsid interface that modulates HIV-1 replication. *PLoS Pathog.* 2012;8(8):e1002896.

101. Matreyek KA, Yucel SS, Li X, Engelman A. Nucleoporin NUP153 phenylalanine-glycine motifs engage a common binding pocket within the HIV-1 capsid protein to mediate lentiviral infectivity. *PLoS Pathog.* 2013;9(10):e1003693.

102. Pornillos O, Ganser-Pornillos BK, Kelly BN, et al. X-ray structures of the hexameric building block of the HIV capsid. *Cell.* Jun 26 2009;137(7):1282-1292.

103. Ganser BK, Li S, Klishko VY, Finch JT, Sundquist WI. Assembly and analysis of conical models for the HIV-1 core. *Science.* Jan 1 1999;283(5398):80-83.

104. Briggs JA, Wilk T, Welker R, Krausslich HG, Fuller SD. Structural organization of authentic, mature HIV-1 virions and cores. *EMBO J.* Apr 1 2003;22(7):1707-1715.

105. Berthet-Colominas C, Monaco S, Novelli A, Sibai G, Mallet F, Cusack S. Head-to-tail dimers and interdomain flexibility revealed by the crystal structure of HIV-1 capsid

protein (p24) complexed with a monoclonal antibody Fab. *EMBO J.* Mar 1 1999;18(5):1124-1136.

106. Gamble TR, Yoo S, Vajdos FF, et al. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science.* Oct 31 1997;278(5339):849-853.

107. Gitti RK, Lee BM, Walker J, Summers MF, Yoo S, Sundquist WI. Structure of the amino-terminal core domain of the HIV-1 capsid protein. *Science.* Jul 12 1996;273(5272):231-235.

108. Worthylake DK, Wang H, Yoo S, Sundquist WI, Hill CP. Structures of the HIV-1 capsid protein dimerization domain at 2.6 A resolution. *Acta Crystallogr D Biol Crystallogr.* Jan 1999;55(Pt 1):85-92.

109. Rose S, Hensley P, O'Shannessy DJ, Culp J, Debouck C, Chaiken I. Characterization of HIV-1 p24 self-association using analytical affinity chromatography. *Proteins.* Apr 1992;13(2):112-119.

110. Ternois F, Sticht J, Duquerroy S, Krausslich HG, Rey FA. The HIV-1 capsid protein C-terminal domain in complex with a virus assembly inhibitor. *Nat Struct Mol Biol.* Aug 2005;12(8):678-682.

111. Ivanov D, Tsodikov OV, Kasanov J, Ellenberger T, Wagner G, Collins T. Domain-swapped dimerization of the HIV-1 capsid C-terminal domain. *Proc Natl Acad Sci U S A.* Mar 13 2007;104(11):4353-4358.

112. Ganser-Pornillos BK, Cheng A, Yeager M. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell.* Oct 5 2007;131(1):70-79.

113. Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med.* Apr 2012;2(4):a007161.

114. Bocanegra R, Rodriguez-Huete A, Fuertes MA, Del Alamo M, Mateu MG. Molecular recognition in the human immunodeficiency virus capsid and antiviral design. *Virus Res.* Nov 2012;169(2):388-410.

115. Kiepiela P, Ngumbela K, Thobakgale C, et al. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat Med.* Jan 2007;13(1):46-53.

116. Rolland M, Heckerman D, Deng W, et al. Broad and Gag-biased HIV-1 epitope repertoires are associated with lower viral loads. *PLoS One.* 2008;3(1):e1424.

117. Leslie AJ, Pfafferott KJ, Chetty P, et al. HIV evolution: CTL escape mutation and reversion after transmission. *Nat Med.* Mar 2004;10(3):282-289.

118. Li B, Gladden AD, Altfeld M, et al. Rapid reversion of sequence polymorphisms dominates early human immunodeficiency virus type 1 evolution. *J Virol.* Jan 2007;81(1):193-201.

119. Martinez-Picado J, Prado JG, Fry EE, et al. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. *J Virol.* Apr 2006;80(7):3617-3623.

120. Schneidewind A, Brockman MA, Yang R, et al. Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. *J Virol.* Nov 2007;81(22):12382-12393.

121. Boutwell CL, Rowley CF, Essex M. Reduced viral replication capacity of human immunodeficiency virus type 1 subtype C caused by cytotoxic-T-lymphocyte escape mutations in HLA-B57 epitopes of capsid protein. *J Virol.* Mar 2009;83(6):2460-2468.

122. Domingo E, Holland JJ. RNA virus mutations and fitness for survival. *Annu Rev Microbiol.* 1997;51:151-178.

123. Quinones-Mateu ME, Ball SC, Marozsan AJ, et al. A dual infection/competition assay shows a correlation between ex vivo human immunodeficiency virus type 1 fitness and disease progression. *J Virol.* Oct 2000;74(19):9222-9233.

124. Troyer RM, Collins KR, Abraha A, et al. Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression. *J Virol.* Jul 2005;79(14):9006-9018.

125. Quinones-Mateu ME, Arts EJ, Kuiken C, et al. HIV-1 Fitness: Implications for Drug Resistance, Disease Progression, and Global Epidemic Evolution. *HIV Sequence Compendium*. Los Alamos: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory; 2001:In press.

126. Henry KR, Weber J, Quinones-Mateu ME, Arts EJ. The impact of viral and host elements on HIV fitness and disease progression. *Curr HIV/AIDS Rep.* Feb 2007;4(1):36-41.

127. Tang J, Tang S, Lobashevsky E, et al. HLA allele sharing and HIV type 1 viremia in seroconverting Zambians with known transmitting partners. *AIDS Res Hum Retroviruses.* Jan 2004;20(1):19-25.

128. Brumme ZL, Li C, Miura T, et al. Reduced replication capacity of NL4-3 recombinant viruses encoding reverse transcriptase-integrase sequences from HIV-1 elite controllers. *J Acquir Immune Defic Syndr.* Feb 1 2011;56(2):100-108.

129. Miura T, Brumme ZL, Brockman MA, et al. Impaired replication capacity of acute/early viruses in persons who become HIV controllers. *J Virol.* Aug 2010;84(15):7581-7591.

130. Lobritz MA, Lassen KG, Arts EJ. HIV-1 replicative fitness in elite controllers. *Curr Opin HIV AIDS.* May 2011;6(3):214-220.

131. Harrigan PR, Bloor S, Larder BA. Relative replicative fitness of zidovudine-resistant human immunodeficiency virus type 1 isolates in vitro. *J Virol.* May 1998;72(5):3773-3778.

132. Koval CE, Dykes C, Wang J, Demeter LM. Relative replication fitness of efavirenz-resistant mutants of HIV-1: correlation with frequency during clinical therapy and evidence of compensation for the reduced fitness of K103N + L100I by the nucleoside resistance mutation L74V. *Virology.* Sep 15 2006;353(1):184-192.

133. Hu Z, Giguel F, Hatano H, Reid P, Lu J, Kuritzkes DR. Fitness comparison of thymidine analog resistance pathways in human immunodeficiency virus type 1. *J Virol.* Jul 2006;80(14):7020-7027.

134. Goudsmit J, De Ronde A, Ho DD, Perelson AS. Human immunodeficiency virus fitness in vivo: calculations based on a single zidovudine resistance mutation at codon 215 of reverse transcriptase. *J Virol.* Aug 1996;70(8):5662-5664.

135. Gerondelis P, Archer RH, Palaniappan C, et al. The P236L delavirdine-resistant human immunodeficiency virus type 1 mutant is replication defective and demonstrates alterations in both RNA 5'-end- and DNA 3'-end-directed RNase H activities. *J Virol.* Jul 1999;73(7):5803-5813.

136. Sugiura W, Matsuda Z, Yokomaku Y, et al. Interference between D30N and L90M in selection and development of protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob Agents Chemother.* Mar 2002;46(3):708-715.

137. Liu Y, McNevin J, Zhao H, et al. Evolution of human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitopes: fitness-balanced escape. *J Virol.* Nov 2007;81(22):12179-12188.

138. Crawford H, Lumm W, Leslie A, et al. Evolution of HLA-B*5703 HIV-1 escape mutations in HLA-B*5703-positive individuals and their transmission recipients. *J Exp Med.* Apr 13 2009;206(4):909-921.

139. Ganusov VV, Goonetilleke N, Liu MK, et al. Fitness costs and diversity of the cytotoxic T lymphocyte (CTL) response determine the rate of CTL escape during acute and chronic phases of HIV infection. *J Virol.* Oct 2011;85(20):10518-10528.

140. Quinones-Mateu ME, Arts EJ. Virus fitness: concept, quantification, and application to HIV population dynamics. *Curr Top Microbiol Immunol.* 2006;299:83-140.

141. Dykes C, Demeter LM. Clinical significance of human immunodeficiency virus type 1 replication fitness. *Clin Microbiol Rev.* Oct 2007;20(4):550-578.

142. Ball SC, Abraha A, Collins KR, et al. Comparing the ex vivo fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C. *J Virol.* Jan 2003;77(2):1021-1038.

143. Martinez-Picado J, Martinez MA. HIV-1 reverse transcriptase inhibitor resistance mutations and fitness: a view from the clinic and ex vivo. *Virus Res.* Jun 2008;134(1-2):104-123.

144. Troyer RM, McNevin J, Liu Y, et al. Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response. *PLoS Pathog.* Apr 2009;5(4):e1000365.

145. Abraha A, Nankya IL, Gibson R, et al. CCR5- and CXCR4-tropic subtype C human immunodeficiency virus type 1 isolates have a lower level of pathogenic fitness than other dominant group M subtypes: implications for the epidemic. *J Virol.* Jun 2009;83(11):5592-5605.

146. Rolland M, Carlson JM, Manocheewa S, et al. Amino-acid co-variation in HIV-1 Gag subtype C: HLA-mediated selection pressure and compensatory dynamics. *PLoS One.* 2010;5(9):e12463.

147. Rolland M, Manocheewa S, Swain JV, et al. HIV-1 conserved-element vaccines: relationship between sequence conservation and replicative capacity. *J Virol.* May 2013;87(10):5461-5467.

148. Lanxon-Cookson EC, Swain JV, Manocheewa S, et al. Factors affecting relative fitness measurements in pairwise competition assays of human immunodeficiency viruses. *J Virol Methods.* Dec 2013;194(1-2):7-13.

149. Garcia-Lerma JG, MacInnes H, Bennett D, Weinstock H, Heneine W. Transmitted human immunodeficiency virus type 1 carrying the D67N or K219Q/E mutation evolves rapidly to zidovudine resistance in vitro and shows a high replicative fitness in the presence of zidovudine. *J Virol.* Jul 2004;78(14):7545-7552.

150. Rangel HR, Weber J, Chakraborty B, et al. Role of the human immunodeficiency virus type 1 envelope gene in viral fitness. *J Virol.* Aug 2003;77(16):9069-9073.

151. Liu Y, Holte S, Rao U, et al. A sensitive real-time PCR based assay to estimate the impact of amino acid substitutions on the competitive replication fitness of human immunodeficiency virus type 1 in cell culture. *J Virol Methods.* Apr 2013;189(1):157-166.

152. Manocheewa S, Swain JV, Lanxon-Cookson E, Rolland M, Mullins JI. Fitness costs of mutations at the HIV-1 capsid hexamerization interface. *PLoS One.* 2013;8(6):e66065.

153. Martinez-Picado J, Savara AV, Sutton L, D'Aquila RT. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *J Virol.* May 1999;73(5):3744-3752.

154. Wang J, Dykes C, Domaoal RA, Koval CE, Bambara RA, Demeter LM. The HIV-1 reverse transcriptase mutants G190S and G190A, which confer resistance to non-nucleoside reverse transcriptase inhibitors, demonstrate reductions in RNase H activity and DNA synthesis from tRNA(Lys, 3) that correlate with reductions in replication efficiency. *Virology.* May 10 2006;348(2):462-474.

155. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol.* Jul 5 2002;320(2):369-387.

156. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073-1081.

157. Worth CL, Bickerton GR, Schreyer A, et al. A structural bioinformatics approach to the analysis of nonsynonymous single nucleotide polymorphisms (nsSNPs) and their relation to disease. *J Bioinform Comput Biol.* Dec 2007;5(6):1297-1318.

158. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* Apr 2010;7(4):248-249.

159. Williams SG, Madan R, Norris MG, et al. Using knowledge of protein structural constraints to predict the evolution of HIV-1. *J Mol Biol.* Jul 29 2011;410(5):1023-1034.

160. Humphris-Narayanan E, Akiva E, Varela R, S OC, Kortemme T. Prediction of mutational tolerance in HIV-1 protease and reverse transcriptase using flexible backbone protein design. *PLoS Comput Biol.* 2012;8(8):e1002639.

161. Boutwell CL, Carlson JM, Lin TH, et al. Frequent and variable cytotoxic-T-lymphocyte escape-associated fitness costs in the human immunodeficiency virus type 1 subtype B Gag proteins. *J Virol.* Apr 2013;87(7):3952-3965.

162. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol.* Oct 2009;19(5):596-604.

163. Benedix A, Becker CM, de Groot BL, Caflisch A, Bockmann RA. Predicting free energy changes using structural ensembles. *Nat Methods.* Jan 2009;6(1):3-4.

164. Das R, Baker D. Macromolecular modeling with rosetta. *Annu Rev Biochem.* 2008;77:363-382.

165. Samudrala R, Moult J. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. *J Mol Biol.* Feb 6 1998;275(5):895-916.

166. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* Nov 2006;15(11):2507-2524.

167. Domingo E. Mechanisms of viral emergence. *Vet Res.* Nov-Dec 2010;41(6):38.

168. Sharma PL, Crumpacker CS. Attenuated replication of human immunodeficiency virus type 1 with a didanosine-selected reverse transcriptase mutation. *J Virol.* Nov 1997;71(11):8846-8851.

169. Song H, Pavlicek JW, Cai F, et al. Impact of immune escape mutations on HIV-1 fitness in the context of the cognate transmitted/founder genome. *Retrovirology.* 2012;9:89.

170. Ali A, Yang O. A novel small reporter gene and HIV-1 fitness assay. *Journal of Virological Methods.* 2006;133(1):41-47.

171. van Maarseveen NM, Huigen MC, de Jong D, Smits AM, Boucher CA, Nijhuis M. A novel real-time PCR assay to determine relative replication capacity for HIV-1 protease variants and/or reverse transcriptase variants. *J Virol Methods.* May 2006;133(2):185-194.

172. Anastassopoulou CG, Ketas TJ, Klasse PJ, Moore JP. Resistance to CCR5 inhibitors caused by sequence changes in the fusion peptide of HIV-1 gp41. *Proc Natl Acad Sci U S A.* Mar 31 2009;106(13):5318-5323.

173. Abraha A, Troyer RM, Quinones-Mateu ME, Arts EJ. Methods to determine HIV-1 ex vivo fitness. *Methods Mol Biol.* 2005;304:355-368.

174. Nomura S, Hosoya N, Brumme ZL, et al. Significant reductions in Gag-protease-mediated HIV-1 replication capacity during the course of the epidemic in Japan. *J Virol.* Feb 2013;87(3):1465-1476.

175. Brockman MA, Schneidewind A, Lahaie M, et al. Escape and Compensation from Early HLA-B57-Mediated Cytotoxic T-Lymphocyte Pressure on Human Immunodeficiency Virus Type 1 Gag Alter Capsid Interactions with Cyclophilin A. *J Virol.* Nov 2007;81(22):12608-12618.

176. Miura T, Brockman MA, Schneidewind A, et al. HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphocyte [corrected] recognition. *J Virol.* Mar 2009;83(6):2743-2755.

177. JoVE Science Education Database. PCR: The Polymerase Chain Reaction. *Basic Methods in Cellular and Molecular Biology*; 2014.

178. Lee PY, Costumbrado J, Hsu CY, Kim YH. Agarose gel electrophoresis for the separation of DNA fragments. *J Vis Exp.* 2012;62(Apr 20).

179. JoVE Science Education Database. Restriction Enzyme Digests. *Basic Methods in Cellular and Molecular Biology*; 2014.

180. McClure J, van't Wout AB, Tran T, Mittler JE. Granulocyte-monocyte colony-stimulating factor upregulates HIV-1 replication in monocyte-derived macrophages cultured at low density. *J Acquir Immune Defic Syndr.* Mar 1 2007;44(3):254-261.

181. Reed LJ, Muench H. A simple method of estimating fifty percent endpoints. *Am J Hyg.* 1938;27(3):493-497.

182. Nickle DC, Jensen MA, Gottlieb GS, et al. Consensus and ancestral state HIV vaccines. *Science.* Mar 7 2003;299(5612):1515-1518; author reply 1515-1518.

183. Rolland M, Jensen MA, Nickle DC, et al. Reconstruction and function of ancestral center-of-tree human immunodeficiency virus type 1 proteins. *J Virol.* 2007;81(16):8507-8514.

184. Bryksin AV, Matsumura I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques.* Jun 2010;48(6):463-465.

185. Spira AI, Ho DD. Effect of different donor cells on human immunodeficiency virus type 1 replication and selection in vitro. *J. Virol.* 1995;69(1):422-429.

186. Zakeri H, Amparo G, Chen SM, Spurgeon S, Kwok PY. Peak height pattern in dichloro-rhodamine and energy transfer dye terminator sequencing. *Biotechniques.* Sep 1998;25(3):406-410, 412-404.

187. Lalonde MS, Troyer RM, Syed AR, et al. Sensitive oligonucleotide ligation assay for low-level detection of nevirapine resistance mutations in human immunodeficiency virus type 1 quasispecies. *J Clin Microbiol.* Aug 2007;45(8):2604-2615.

188. Bartonova V, Igonet S, Sticht J, et al. Residues in the HIV-1 capsid assembly inhibitor binding site are essential for maintaining the assembly-competent quaternary structure of the capsid protein. *J Biol Chem.* Nov 14 2008;283(46):32024-32033.

189. Tang S, Murakami T, Agresta BE, Campbell S, Freed EO, Levin JG. Human immunodeficiency virus type 1 N-terminal capsid mutants that exhibit aberrant core morphology and are blocked in initiation of reverse transcription in infected cells. *J Virol.* Oct 2001;75(19):9357-9366.

190. Bocanegra R, Rodriguez-Huete A, Fuertes MA, Del Alamo M, Mateu MG. Molecular recognition in the human immunodeficiency virus capsid and antiviral design. *Virus Res.* Jun 21 2012.

191. Koup RA, Safrit JT, Cao Y, et al. Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J Virol.* 1994;68(7):4650-4655.

192. Liu Y, Mullins JI, Mittler JE. Waiting times for the appearance of cytotoxic T-lymphocyte escape mutants in chronic HIV-1 infection. *Virology.* 2006;347:140-146.

193. Allen TM, Yu XG, Kalife ET, et al. De novo generation of escape variant-specific CD8+ T-cell responses following cytotoxic T-lymphocyte escape in chronic human immunodeficiency virus type 1 infection. *J Virol.* Oct 2005;79(20):12952-12960.

194. Mothe B, Llano A, Ibarrondo J, et al. CTL responses of high functional avidity and broad variant cross-reactivity are associated with HIV control. *PLoS ONE.* 2012;7(1):e29717.

195. Laskowski RA. PDBsum new things. *Nucleic Acids Res.* Jan 2009;37(Database issue):D355-359.

196. *Mesquite: a modular system for evolutionary analysis* [computer program]. Version 2.75; 2011.

197. Deng W, Maust BS, Nickle DC, et al. DIVEIN: A Web Server to Analyze Phylogenies, Sequence Divergence, Diversity, and Informative Sites. *Biotechniques.* 2010;48(5):405-408.

198. Ma J, Dykes C, Wu T, Huang Y, Demeter L, Wu H. vFitness: a web-based computing tool for improving estimation of in vitro HIV-1 fitness experiments. *BMC Bioinformatics.* 2010;11:261.

199. Wu H, Huang Y, Dykes C, et al. Modeling and estimation of replication fitness of human immunodeficiency virus type 1 in vitro experiments by using a growth competition assay. *J Virol.* Mar 2006;80(5):2380-2389.

200. Carlson JM, Brumme ZL, Rousseau CM, et al. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. *PLoS Comput Biol.* Nov 2008;4(11):e1000225.

201. Pornillos O, Ganser-Pornillos BK, Yeager M. Atomic-level modelling of the HIV capsid. *Nature.* Jan 20 2011;469(7330):424-427.

202. Bailey GD, Hyun JK, Mitra AK, Kingston RL. A structural model for the generation of continuous curvature on the surface of a retroviral capsid. *J Mol Biol.* Mar 30 2012;417(3):212-223.

203. Mothe B, Llano A, Ibarrondo J, et al. Definition of the viral targets of protective HIV-1-specific T cell responses. *J Transl Med.* 2011;9:208.

204. Niu L, Termini JM, Kanagavelu SK, et al. Preclinical evaluation of HIV-1 therapeutic ex vivo dendritic cell vaccines expressing consensus Gag antigens and conserved Gag epitopes. *Vaccine.* Mar 3 2011;29(11):2110-2119.

205. Altfeld M, Allen TM. Hitting HIV where it hurts: an alternative approach to HIV vaccine design. *Trends Immunol.* Nov 2006;27(11):504-510.

206. Letourneau S, Im EJ, Mashishi T, et al. Design and pre-clinical evaluation of a universal HIV-1 vaccine. *PLoS One.* 2007;2(10):e984.

207. Yang OO. Candidate vaccine sequences to represent intra- and inter-clade HIV-1 variation. *PLoS One.* 2009;4(10):e7388.

208. Kunwar P, Hawkins N, Dinges WL, et al. Superior control of HIV-1 replication by CD8+ T cells targeting conserved epitopes: implications for HIV vaccine design. *PLoS One.* 2013;8(5):e64405.

209. Crawford H, Prado JG, Leslie A, et al. Compensatory Mutation Partially Restores Fitness And Delays Reversion Of Escape Mutation Within The Immunodominant Hla-B*5703-Restricted Gag Epitope In Chronic Hiv-1 Infection. *J Virol.* 2007;81(15):8346-8351.

210. Boutwell CL, Schneidewind A, Brumme Z, et al. CTL escape mutations in Gag epitopes restricted by protective HLA class I alleles cause substantial reductions in viral replication capacity. Paper presented at: AIDS Vaccine, P09-19 LB, 2009; Paris, France.

211. Tieu H-V, Rolland M, Hammer SM, Sobieszczyk ME. Translational Research Insights From Completed HIV Vaccine Efficacy Trials. *J Acquir Immune Defic Syndr.* 2013;63:S150-S154.

212. Kawashima Y, Pfafferott K, Frater J, et al. Adaptation of HIV-1 to human leukocyte antigen class I. *Nature.* Apr 2 2009;458(7238):641-645.

213. Rousseau CM, Lockhart DW, Listgarten J, et al. Rare HLA drive additional HIV evolution compared to more frequent alleles. *AIDS Res Hum Retroviruses.* Mar 2009;25(3):297-303.

214. Fassati A. Multiple roles of the capsid protein in the early steps of HIV-1 infection. *Virus Res.* Dec 2012;170(1-2):15-24.

215. Stephenson KE, Li H, Walker BD, Michael NL, Barouch DH. Gag-specific cellular immunity determines in vitro viral inhibition and in vivo virologic control following simian immunodeficiency virus challenges of vaccinated rhesus monkeys. *J Virol.* Sep 2012;86(18):9583-9589.

216. Iwamoto N, Takahashi N, Seki S, et al. Control of simian immunodeficiency virus replication by vaccine-induced Gag- and Vif-specific CD8+ T cells. *J Virol.* Jan 2014;88(1):425-433.

217. Woo J, Robertson DL, Lovell SC. Constraints on HIV-1 diversity from protein structure. *J Virol.* Dec 2010;84(24):12995-13003.

218. Ganser-Pornillos BK, von Schwedler UK, Stray KM, Aiken C, Sundquist WI. Assembly properties of the human immunodeficiency virus type 1 CA protein. *J Virol.* Mar 2004;78(5):2545-2552.

219. Zhao G, Perilla JR, Yufenyuy EL, et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature.* May 30 2013;497(7451):643-646.

220. Krivov GG, Shapovalov MV, Dunbrack RL, Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins.* Dec 2009;77(4):778-795.

221. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem.* Dec 2005;26(16):1781-1802.

222. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res.* Jul 1 2005;33(Web Server issue):W382-388.

223. Eswar N, Webb B, Marti-Renom MA, et al. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics.* Oct 2006;Chapter 5:Unit 5 6.

224. Rose PP, Korber BT. Detecting hypermutations in viral sequences with an emphasis on G --> A hypermutation. *Bioinformatics.* Apr 2000;16(4):400-401.

225. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* Aug 19 2004;5:113.

226. del Alamo M, Neira JL, Mateu MG. Thermodynamic dissection of a low affinity protein-protein interface involved in human immunodeficiency virus assembly. *J Biol Chem.* Jul 25 2003;278(30):27923-27929.

227. Gupta RK, Jordan MR, Sultan BJ, et al. Global trends in antiretroviral resistance in treatment-naive individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. *Lancet.* Oct 6 2012;380(9849):1250-1258.

228. Schneidewind A, Brockman MA, Sidney J, et al. Structural and functional constraints limit options for cytotoxic T-lymphocyte escape in the immunodominant HLA-B27-restricted epitope in human immunodeficiency virus type 1 capsid. *J Virol.* Jun 2008;82(11):5594-5605.

229. Yeh WW, Cale EM, Jaru-Ampornpan P, Lord CI, Peyerl FW, Letvin NL. Compensatory substitutions restore normal core assembly in simian immunodeficiency virus isolates with Gag epitope cytotoxic T-lymphocyte escape mutations. *J Virol.* Aug 2006;80(16):8168-8177.

230. Nijhuis M, Schuurman R, de Jong D, et al. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS.* Dec 3 1999;13(17):2349-2359.

231. Chang MW, Torbett BE. Accessory mutations maintain stability in drug-resistant HIV-1 protease. *J Mol Biol.* Jul 22 2011;410(4):756-760.

232. Gong LI, Suchard MA, Bloom JD. Stability-mediated epistasis constrains the evolution of an influenza protein. *Elife.* 2013;2:e00631.

233. Liu Y, Rao U, McClure J, et al. Impact of mutations in highly conserved amino acids of the HIV-1 Gag-p24 and Env-gp120 proteins on viral replication in different genetic backgrounds. *PLoS One.* 2014;9(4):e94240.

234. Calza L. Renal toxicity associated with antiretroviral therapy. *HIV Clin Trials.* Jul-Aug 2012;13(4):189-211.

235. Gutierrez F, Masia M. The role of HIV and antiretroviral therapy in bone disease. *AIDS Rev.* Apr-Jun 2011;13(2):109-118.

236. Torres RA, Lewis W. Aging and HIV/AIDS: pathogenetic role of therapeutic side effects. *Lab Invest.* Feb 2014;94(2):120-128.

237. Garg H, Joshi A, Mukherjee D. Cardiovascular complications of HIV infection and treatment. *Cardiovasc Hematol Agents Med Chem.* Mar 2013;11(1):58-66.

238. Bachis A, Aden SA, Nosheny RL, Andrews PM, Mocchetti I. Axonal transport of human immunodeficiency virus type 1 envelope protein glycoprotein 120 is found in association with neuronal apoptosis. *J Neurosci.* 2006;26:6771-6780.

239. Gultas M, Haubrock M, Tuysuz N, Waack S. Coupled mutation finder: a new entropy-based method quantifying phylogenetic noise for the detection of compensatory mutations. *BMC Bioinformatics.* 2012;13:225.

CURRICULUM VITAE

Siriphan Manocheewa

Education

Undergraduate	**Sirindhorn International Institute of Technology**
2001-2004	**Thammasat University, Thailand**
	B.S. Information Technology (First Honor)
	Full Scholarship 2001-04
Graduate	**King Mongkut's University of Technology Thonburi**
2004-2007	**Thailand**
	M.S Bioinformatics
	Full Scholarship 2004-07
2007-2015	**University of Washington**
	Ph.D. Microbiolgoy

Research Experience

Cell culture	Trained in BSL-2/3 laboratory practices for handling HIV-1 and HTLV-1 molecular clones and infected cell lines for HIV replication fitness assays
Molecular Biology	Protein detection / quantification assays: ELISA
	Molecular cloning and site directed mutagenesis
	Nucleic acid extraction / detection / quantification
Bioinformatics	Programming and database development: Perl, Python, C/C++, Java.
	Others: PHP, XML, HTML, SQL.

Protein/DNA sequence analyses: Sanger sequencing, Next generation sequencing, Sequence alignment, Phylogenetic trees, T-cell epitope prediction

Molecular structure visualization and modeling programs and tools: MODELLER, NAMD, Chimera, PyMOL, FOLDX, DOPE and RAPDF

Teaching Experience (Graduate Student Teaching Assistant)

- MICROM 302 General Microbiology Laboratory (Fall 2008)
    - Lab sessions: Give lecture review and instruct laboratory procedures
    - Grade homework and exams
- MICROM 431 Prokaryotic Recombinant DNA Techniques
    - Lab sessions: Give lecture review and instruct laboratory procedures
    - Grade homework and exams

Publications

Manocheewa S, Mittler JE, Samudrala R, and Mullins JI. Composite Sequence-Structure Stability Models as Screening Tools for Identifying Mutational Targets for Drug and Vaccine Development. *In submission*

Manocheewa S, Lanxon-Cookson EC, Liu Y, Swain JV, Ushnal R, Maust B, Wenjie D, Sunshine JE, Kim M, Morgane R and Mullins JI. Pairwise Growth Competition Assay for Determining the Replication Fitness of Human Immunodeficiency Viruses. JoVE (*In press)*

Liu Y, Rao U, McClure J, Konopa P, **Manocheewa S**, Kim M, Chen L, Troyer RM, Tebit DM, Holte S, Arts EJ, Mullins JI. Impact of mutations in highly conserved amino acids of the HIV-1 Gag-p24 and Env-GP120 proteins on viral replication in different genetic backgrounds. PloS one. 2014 April 9(4), e94240

Lanxon-Cookson EC, Swain JV, **Manocheewa S,** Smith RA, Maust B, Westfall, DH, Rolland M, Mullins JI. Factors affecting relative fitness measurements in pairwise competition assays. J Virol Methods. 2013 Dec;194(1-2):7-13

Manocheewa S, Swain JV, Lanxon-Cookson, EC, Rolland M, Mullins JI. Fitness costs of mutations at the HIV-1 capsid hexamerization interface. PLoS One. 2013 June 13:8(6).

Rolland M, **Manocheewa S,** Swain JV, Lanxon-Cookson EC, Kim M, Westfall DH, Larsen BB, Gilbert PB, Mullins JI. HIV-1 conserved element vaccines: relationship between sequence conservation and replicative capacity. J Virol. 2013 May:87(10):5461-7

Rolland M, Carlson JM, **Manocheewa S**, Swain JV, Lanxon-Cookson EC, Deng W, Rousseau CM, Raugi DN, Learn GH, Maust BS, Coovadia H, Ndung'u T, Goulder PJ, Walker BD, Brander C, Heckerman DE, Mullins JI. Amino-acid co-variation in HIV-1 Gag subtype C: HLA-mediated selection pressure and compensatory dynamics. PLoS One. 2010 Sep 1:5(9).

Oral Presentations

March 2006 **Siriphan Manocheewa**, Taweerat Vichitsoonthonkul, Pravech Ajawatanawong, and Supawadee Ingsriswang (2006), "Combination

Approach for Predicting Prokaryotic Sigma Factor Binding Sites", Proceeding of the 10th Annual National Symposium on Computational Science and Engineering (ANSCSE10).

Poster Presentations

March 2013 **Manocheewa S,** Swain JV, Lanxon-Cookson EC, Rolland M, Mullins JI. Influence of HIV-1 genetic background on the fitness cost of single amino acid mutations in Gag-p24. 20th Conference on Retroviruses and Opportunistic Infections. ***Young Investigator Awardee***

March 2012 **Manocheewa S,** Swain JV, Lanxon-Cookson EC, Kim M, Westfall DH, Rolland M, Mullins JI. A single amino acid change at p24-p24 interface within HIV-1 capsid incurs a strong fitness cost and can inhibit viral replication. 19th Conference on Retroviruses and Opportunistic Infections. ***Young Investigator Awardee***

March 2012 Lanxon-Cookson EC, Swain JV, **Manocheewa S**, Kim M, Westfall DH, Rolland M, Maust B, Mullins JI. HIV-1 viral fitness determined using pairwise growth competition assays. 19th Conference on Retroviruses and Opportunistic Infections.

October 2011 GN Pavlakis, M Rolland, BK Felber, V Kulkarni, B Mothe, C Brander, J Termini, GW Stone, S Le Gall, J Yan, DB Weiner, **S Manocheewa**, JV Swain, E Lanxon-Cookson, JI Mullins. Conserved elements (CE) vaccine. AIDS Vaccine 2011.

March 2011 Rolland M, Swain JV, Lanxon-Cookson EC, **Manocheewa S**, Kim M, Westfall DH, Larsen BB, Gilbert PB, Mullins JI. HIV-1 conserved elements vaccine: relationship between sequence conservation and viral fitness. 18th Conference on Retroviruses and Opportunistic Infections.

October 2010 JI Mullins, GN Pavlakis, BK Felber, C Brander, S Le Gall, D Weiner, G Stone, M Rolland, V Kulkarni, B Mothe, J Yan, D Heckerman, B Ganneru, G Zhang, A Singh, M Rosati, P Pankhong, E Lanxon-Cookson, V Swain, S Elliot, **S Manocheewa**, S Zhang, T Gonzalez, J Zamarreno, A Llano, J Ibarrondo, J Herrero, E Palou, B Clotet, S Gel. A global AIDS vaccine based on conserved elements of the viral proteome. AIDS Vaccine 2010

October 2010 V Kulkarni, A Valentin, J Yan, M Rolland, S Le Gall, B Mothe, B Ganneru, V Patel, G Zhang, M Rosati, P Pankhong, S Elliott, A Harris, **S Manocheewa**, NY Sardesai, DB Weiner, C Brander, GN Pavlakis, BK Felber, JI Mullins. Conserved elements vaccine for HIV-1 p24gag is immunogenic in mice and macaques. AIDS Vaccine 2010.