2014

Causes and Consequences of Genetic Robustness and Fragility in HIV-1 Proteins

Suzannah J. Rihn

CAUSES AND CONSEQUENCES OF GENETIC ROBUSTNESS AND FRAGILITY IN HIV-1 PROTEINS

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Suzannah J. Rihn

June 2014

CAUSES AND CONSEQUENCES OF GENETIC ROBUSTNESS AND FRAGILITY IN HIV-1 PROTEINS

Suzannah J. Rihn, Ph.D.

The Rockefeller University 2014

Genetic robustness describes a capacity to maintain function in the face of mutation. RNA viruses, like human immunodeficiency virus-1 (HIV-1), experience extremely high mutations rates *in vivo*, which contribute to the evolvability that permits HIV-1 to evade immune defenses. However, the compact nature of the HIV-1 viral genome, in which small or overlapping proteins are often multifunctional and essential, often necessitates a strong pressure to conserve functional roles, which is at opposition with immunological pressures to diversify. Under such conditions, and in consideration of the relatively rapid replication cycle of HIV-1, natural selection for robustness would be predicted to be beneficial. At the same time, robustness can come at the cost of fitness, and it is unclear whether robustness or fitness would constitute the dominant selective force in the natural setting.

Within a given virus, genetic robustness is expected to vary amongst proteins, in accordance with specific function. Therefore, to more accurately examine genetic robustness within HIV-1, we selected one protein in which the competing needs to both conserve function yet diversify sequence is particularly acute. HIV-1 capsid (CA) is under strong pressure to preserve roles in viral

assembly, maturation, uncoating and nuclear import, but is also under intense immunological pressure to diversify, particularly from adaptive immune responses. To evaluate the genetic robustness of HIV-1 CA, we generated a large library of random single amino acid substitution CA mutants. Surprisingly, after measuring the replicative fitness of the CA library mutants, we observed HIV-1 CA to be the most genetically fragile protein analyzed with such an approach, with 70% of random CA mutations resulting in nonviable, replication-defective viruses (<2% of WT fitness). While CA is involved in several steps of HIV-1 replication, analyses of conditionally (temperature-sensitive) and constitutively nonviable mutants indicated that the biological basis for the genetic fragility, or lack of robustness, was principally the need to organize accurate and efficient assembly of mature viral particles.

Examination of the CA library mutations in naturally occurring HIV-1 subtype B populations indicated that all mutations occurring at a frequency >3% *in vivo* had fitness levels that were >40% of WT *in vitro*. Moreover, protective CTL epitopes were observed to preferentially occur in regions of HIV-1 CA that were even more genetically fragile than HIV-1 CA as a whole. Overall, these results suggest the extreme genetic fragility of CA may explain why immune responses to Gag, as opposed to other viral proteins, correlate with better prognosis in HIV-1 infection.

To evaluate whether the fragility we observed in CA was unique to CA or a general property of HIV-1 proteins, and to more firmly establish causes and

consequences of genetic robustness or fragility, we evaluated the robustness of another HIV-1 protein, integrase (IN). Again, a large library of unbiased single amino acid mutants was created. In contrast to CA, IN appeared comparatively robust, with only 35% of 156 single amino acid mutations resulting in nonviable viruses. However, when these nonviable mutants were mapped onto a model of the HIV-1 intasome, we noted a striking localization of nonviable mutants to certain IN subunit interfaces. To this end, single mutants affecting both particle production and IN expression in virions could be mapped to even more specific regions within the intasome model. Furthermore, after examining the prevalence of the IN library mutations in an *in vivo* cohort of subtype B IN sequences, mapping to the intasome helped explain why some *in vitro* library mutations with high fitness (>40% of WT) never occur *in vivo*, and also revealed that residues that are highly variable *in vivo* are more likely to occur in regions in the intasome with more exposed surface area. Despite the relative robustness of HIV-IN, our analyses with an intasome model demonstrate that localized regions of fragility, namely certain subunit interfaces, exist.

Understanding the genetic robustness of HIV-1 is particularly important given the continued need for novel therapeutic interventions. Not only has the evaluation of HIV-1 CA and IN robustness revealed particularly fragile regions in both proteins, which may prove ideal targets, but comparisons of their genetic robustness is suggestive of the direction future therapeutic research should take.

ACKNOWLEDGEMENTS

I would like to first acknowledge the guidance and mentorship I have received from Dr. Paul Bieniasz. Not only has he been integral to the development of my thesis work, but his flexibility and support throughout my time in his lab has meant a great deal.

Importantly, Dr. Sam Wilson must also be acknowledged for his important contributions to this work. As the creator of the mutagenized capsid library, a significant portion of my work relied on what he had previously generated. However, he has also been an extremely valuable source of advice and encouragement.

Furthermore, I would like to thank my collaborators. Both Dr. Rob Gifford and Dr. Joseph Hughes provided tremendous help with bioinformatics analyses. Dr. Frazer Rixon and Dr. Saskia Bakker also provided important electron microscopic analyses. Additionally, our examinations of the HIV-1 intasome model structure would not have been possible without the structure coordinates provided by Dr. Stephen Hughes and his laboratory. I would also like to thank Dr. Mike Malim for generously providing a valuable reagent, the INT4 antibody.

Finally, I would like to thank the members of my thesis committee, Dr. Charles Rice, Dr. Nina Papavasiliou, and Dr. Rob Gifford for their valued discussions and direction.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

Retroviridae and human immunodeficiency virus-1 (HIV-1)

The impact of retroviruses in the modern era has been tremendously broad and polarized, causing both global pandemic and also contributing revolutionary biotechnology. While not formally named until 1974, the retrovirus family had been characterized and implicated in numerous vertebrate diseases since the 19th century [1,2]. After the early recognition of Bovine leukosis and Jaagsiekte in sheep, key investigations of oncogenic retroviruses in chickens, including Rous sarcoma virus (RSV), contributed to Temin's proviral hypothesis in 1964, which was based on his observations that RSV particles containing RNA make a DNA copy that is integrated into the host chromosome [2,3,4]. The ensuing discovery of reverse transcriptase (RT) in 1970 permitted definition of the two distinguishing characteristics of retroviruses, the reverse transcription of the single stranded RNA (ssRNA) genome to yield double stranded DNA (dsDNA), and the subsequent integration of the dsDNA into host cell chromosomes [5,6].

Concordant with these defining features, classification of retrovirus genera is assigned based on phylogenetic analyses of the highly conserved RT sequences (Figure 1). As Figure 1 indicates, retroviruses are also broadly divided by the organization of their genomes into two categories, simple or complex [7,8,9]. The simple retroviruses, which include gamma-retroviruses, alpha-retroviruses, beta-retroviruses, and epsilon-retroviruses, are comprised,

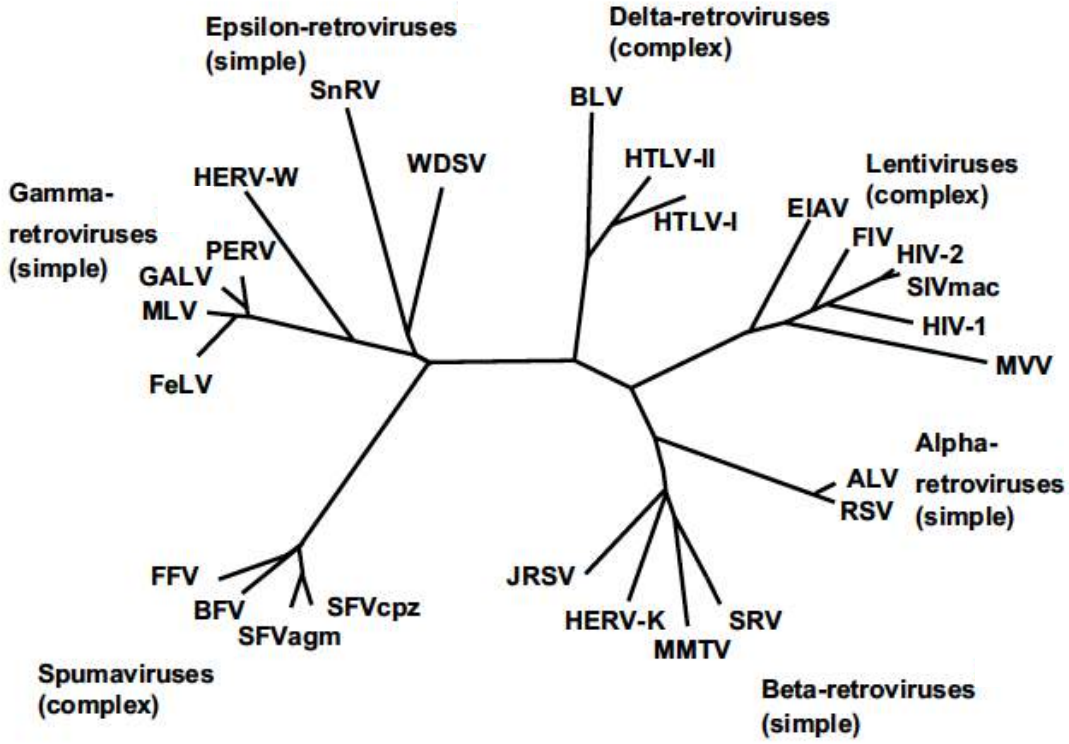


Figure 1. Retrovirus family phylogeny. Classification of retroviruses based on reverse transcriptase sequence. Relevant retrovirus abbreviations discussed in the text. Figure adapted from [2].

primarily, of just three major structural genes. These include: *gag*, the polyprotein which organizes the assembly of viral particles, *pol*, which contains the required viral enzymes for reverse transcription (RT), viral genome integration (integrase), and particle maturation (protease, *pro*, which occurs in a separate open reading frame in beta- and delta-retroviruses), and *env*, which encodes the surface and transmembrane portions of the viral envelope. Complex retroviruses, on the other hand, contain the same three essential structural genes, but typically include additional accessory or regulatory proteins, which are often generated from multiply spliced transcripts.

Despite any differences in genome classification, there are other commonalities amongst retroviruses. All retroviruses are enveloped, nonsegmented, single-stranded, positive sense RNA viruses. Retroviruses are also comparatively small in size, with genomes typically ranging from 7-13 kb, and particles that are generally 80-120 nm in diameter. Furthermore, all retrovirus genomes, aside from containing *gag*, *pol*, and *env*, are flanked on either end by long terminal repeats (LTRs), which contain essential promoter, poly A signal, and TATA components, and through their repeating nature, also help to ensure proper reverse transcription.

Retroviruses also share a common life cycle (Figure 2). To summarize briefly, viral entry (Figure 2A) initiates with the engagement of the retroviral envelope with the specific host cell receptor(s), which triggers membrane fusion at the plasma membrane, or, less frequently, fusion with endosomes.

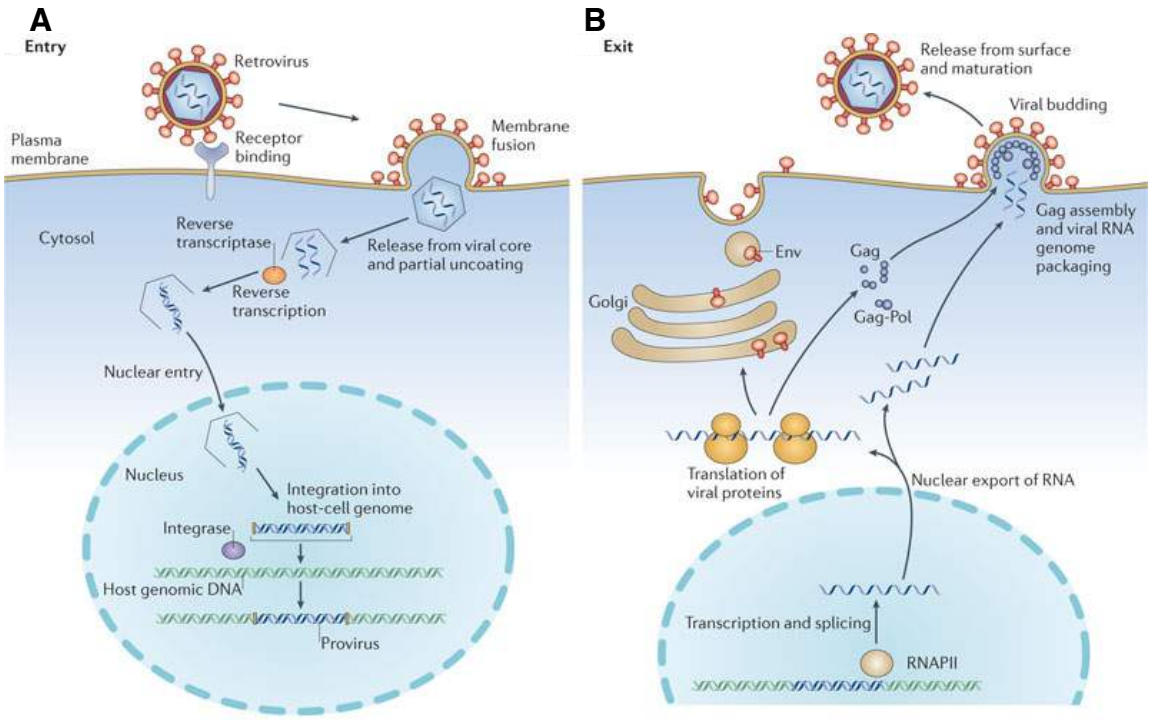


Figure 2. Illustration of the retrovirus lifecycle. (A) Steps in the lifecycle that occur during and after viral entry. These steps include binding of the retrovirus to the host cell receptor, membrane fusion with the plasma membrane (or with endosomes), release of the viral core and subsequent uncoating, reverse transcription, entry through the nucleus, and finally, integration of the viral DNA into the host genomic DNA. (B) Steps in the lifecycle corresponding to assembly and exit of newly made viral particles. These steps include transcription by RNA polymerase II, splicing and export of viral RNA from the nucleus, translation of the viral proteins, RNA packaging and Gag assembly, budding, and release from the cell membrane, and finally, virus maturation. Figure adapted from [10].

Subsequent entry of the viral core into the host cell cytoplasm is followed by the disassembly, or uncoating, of the viral core, after which the reverse transcription complex (RTC) forms, and the ssRNA is reverse transcribed. The newly synthesized dsDNA genome traverses the cytoplasm and forms a preintegration complex (PIC), which permits nuclear entry. Subsequently, the viral dsDNA is integrated into the host genomic DNA to yield the provirus, a process that will later be described in far greater detail. To begin the formation of new retroviral particles (Figure 2B), the promoter in the U3 region of the 5' LTR engages the error-prone RNA polymerase II for transcription. Viral RNA transcripts are then spliced, as necessary, and escorted out of the nucleus, often with the aid of accessory proteins. Some of the exported transcripts are translated as viral proteins or polyproteins that are prepared for Gag assembly, while the remaining full-length RNA transcripts are packaged for inclusion in new particles. As Gag assembly and multimerization completes and viral budding begins, the endosomal sorting complexes required for transport (ESCRT) network of proteins are recruited to aid in the scission and release of the particle from the cell surface. Finally, cleavage by the viral protease ends the life cycle with the formation of a mature retroviral particle. (Basic retroviral characteristics and life cycle reviewed in [9,11,12,13,14]).

Among the retroviruses, one has become the overwhelming focus of intense research. Human immunodeficiency virus-1 (HIV-1) is one of the deadliest human scourges the world has experienced, as infection typically

results in acquired immunodeficiency syndrome (AIDS), which is responsible for the deaths of an estimated 36 million people [15,16]. As of 2012, a further 35.3 million individuals were currently living with HIV, indicating that the clinical impact of the virus remains tremendous, particularly as effective antiretroviral therapies are accessible to just 34% of individuals in the middle- and low-income countries where most infected individuals reside [15], and because vaccine strategies have thus far proved challenging [17,18]. Overall, this is a relatively modern burden, as the first indications of AIDS appeared in 1981 [19,20], and the determination of HIV-1 (often referenced as HTLV-III in the mid 1980s) as the causative agent of AIDS followed in 1983-1984 [21,22,23]. The origins of the AIDS pandemic, caused by the HIV-1 group M lineage, arose from a single cross-species transmission from simian immunodeficiency virus (SIV) in chimpanzees (*Pan troglodytes troglodytes* with SIVcpz) that occurred in southeastern Cameroon around 1910-1930 and spread around Kinshasa [20,24,25,26]. However, several additional HIV-1 transmission events between humans and other primates have occurred, including non-pandemic HIV-1 group N and perhaps O lineage transmissions from SIVcpz, HIV-1 group P lineage transmission from gorillas (SIVgor), and a related cross-species transmission from sooty mangabeys (SIVsmm) is now known to be the source of HIV-2 [20,24,27,28].

Notably, HIV-1, which will form the focus of the research discussed herein, is a lentivirus, along with HIV-2 and primate SIVs (Figure 1). Additional nonprimate lentiviruses are divided into four other groups: ovines-caprines,

bovines, felines, and equines. Lentiviruses derive their name (slow, *lentus*-) from their long incubation periods, and share many additional features, including chronic and persistent viral replication, macrophage tropism, destruction of specific hematologic or immunologic cells, the inclusion of accessory genes in the genome, neurologic manifestations, and periods of latency [9,29,30]. Lentiviruses are also characterized by a common morphology, including a viral core with a unique cone or cylindrical shape [9]. While lentiviruses are not directly oncogenic like many pathogenic retroviruses, the immune disorders they induce can result in oncogenic disorders, such as Kaposi's sarcoma or lymphomas [31,32].

As is true for other primate lentiviruses, the primary host cell receptor for HIV-1 is CD4, and this significantly limits the tropism of the virus, as CD4 expression is generally restricted to macrophages, dendritic cells, and helper T-cells [33,34]. In fact, the slow decline in the population of CD4+ cells over the course of HIV-1 infection contributes to, and defines, the hallmark HIV-1-induced immunodeficiency that is AIDS [35]. In addition to CD4, the HIV-1 surface envelope (gp120) must also bind one of two co-receptors, either chemokine CXCR4 or CCR5 [36,37,38,39]. While most transmitted viruses initially utilize CCR5 (R5-tropic), over time many switch to using CXCR4 (X4-tropic), in a manner linked to disease progression [40].

Not surprisingly, many of the other defining features of HIV-1 involve its unique viral genome (Figure 3), which is just 9-10kb in size. In addition to the

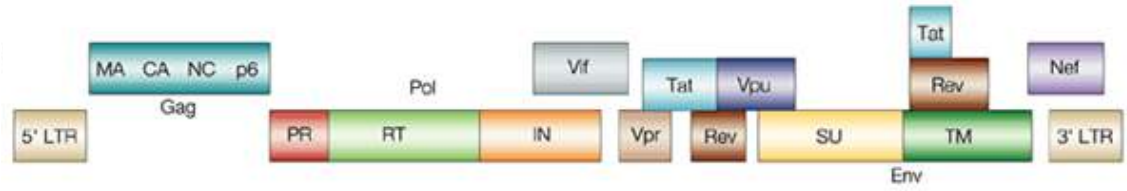


Figure 3. Organization of the HIV-1 viral genome. Illustration of the open reading frames comprising HIV-1. Sizes of proteins are not to scale. Figure adapted from [41].

three structural proteins common to all retroviruses, Gag, Pol, and Env, the HIV-1 genome also includes two regulatory proteins essential for viral replication, Tat and Rev, as well as four accessory proteins, Nef, Vif, Vpu, and Vpr, that can modulate viral replication and affect pathogenesis *in vivo.* As the nature of the work discussed in this thesis includes detailed analyses of the capsid (CA) protein, which is part of Gag, and the integrase protein (IN), which is part of Pol, these polyproteins will be reviewed in far greater detail in separate introductory sections, but basic functions of the other proteins will be summarized. Env is similar to the general retroviral envelope previously described, and consists of surface, trimeric gp120, which elicits considerable virus-neutralizing antibody responses, as well as transmembrane, trimeric gp41 [42,43]. The regulatory protein Tat is primarily involved in nuclear transactivation of the LTR by interactions with TAR sequences, and helps ensure proper activation and elongation of viral genome transcription [44,45]. Both Tat, and the other regulatory protein, Rev, are encoded by multiply spliced mRNAs [16]. Rev is involved in regulating viral mRNA expression; it binds the Rev response element (RRE) located in *env* and impacts nuclear export of either unspliced or singly spliced RNAs [46,47].

One of the common features amongst the HIV-1 accessory proteins is their capacity to act as adapter proteins, as they often bind other host cellular complexes in order to target host proteins (frequently antiviral) for degradation, generally in an ubiquitin-dependent fashion [48]. To this end, Vif, the viral

infectivity factor, partners with cellular proteins (Cul5, elongins B and C, and Rbx1) to promote the degradation of antiviral factor APOBEC3G (as well as APOBEC3C, 3F, and 3H) [49,50,51,52]. Similarly, viral protein U (Vpu) aids virion release through both its interactions with the restriction factor tetherin, in a fashion causing tetherin degradation, and through its contribution to the degradation of CD4 [53,54]. While the Vpr protein does not target a known restriction factor for degradation, it has been speculated that this is likely, and Vpr has very recently been suggested to target the host SLX4 complex to promote arrest of the cell cycle in the G2/M phase as well as escape from innate immune sensing [55]. Vpr may additionally be involved in the activation of transcription, induction of cell death, and enhancement of RT fidelity [48,56]. The final HIV-1 accessory protein, negative factor (Nef), despite its misnomer labeling, is reported to be involved in an increasingly large number of functions that primarily enhance infectivity, alter replication, and delay immune clearance, but it is not essential for viral replication. Nef's wide-ranging roles in HIV-1 are known to impact: signaling transductions pathways, host major histocompatibility complex-1 (MHC-1) and CD4 and other protein expression, antigen presentation, cellular trafficking, and survival of cells [57,58,59].

These descriptions of retroviruses and HIV-1, while basic, provide a context necessary for understanding the subsequently described concepts and experiments.

Genetic robustness and fragility

One of the most remarkable features of HIV-1, and other RNA viruses, is their strikingly high mutation rate. As Figure 4A demonstrates, the mutation rate is higher for RNA viruses than for any other biological entity, with the exception of one documented example of a viroid [60,61]. Furthermore, additional evidence has shown that the mutation rate for ss(+)RNA viruses like HIV-1 is considerably higher than for other RNA viruses, or DNA viruses for that matter (Figure 4B) [62]. Given the neutral theory of molecular evolution, which postulates that the mutation rate is the sole determinant of the rate of evolution [63], as well as the adaptation theory [64,65,66], which also considers mutation rates to be central to evolution rates, it is perhaps not surprising that ss(+)RNA viruses have also been shown to have higher rates of evolution than other viruses (Figure 4C) [62]. In fact, the HIV-1 genome has previously been estimated to evolve approximately one million times faster than mammalian DNA [67,68].

Examination of viral mutation rates and evolvability is of importance in the context of HIV-1 and RNA viruses on several fronts. Mutation rates have already been determined to impact everything from pathogenesis in RNA viruses [69,70], to drug resistance patterns [71,72], antiviral therapy success [73,74], efficacy of vaccines [75,76], and the emergence of new pathogens [77,78]. However, despite the simplicity of the neutral theory of molecular evolution, directly equating mutation and evolution rates is unrealistic given the complex environments in which these rates occur.

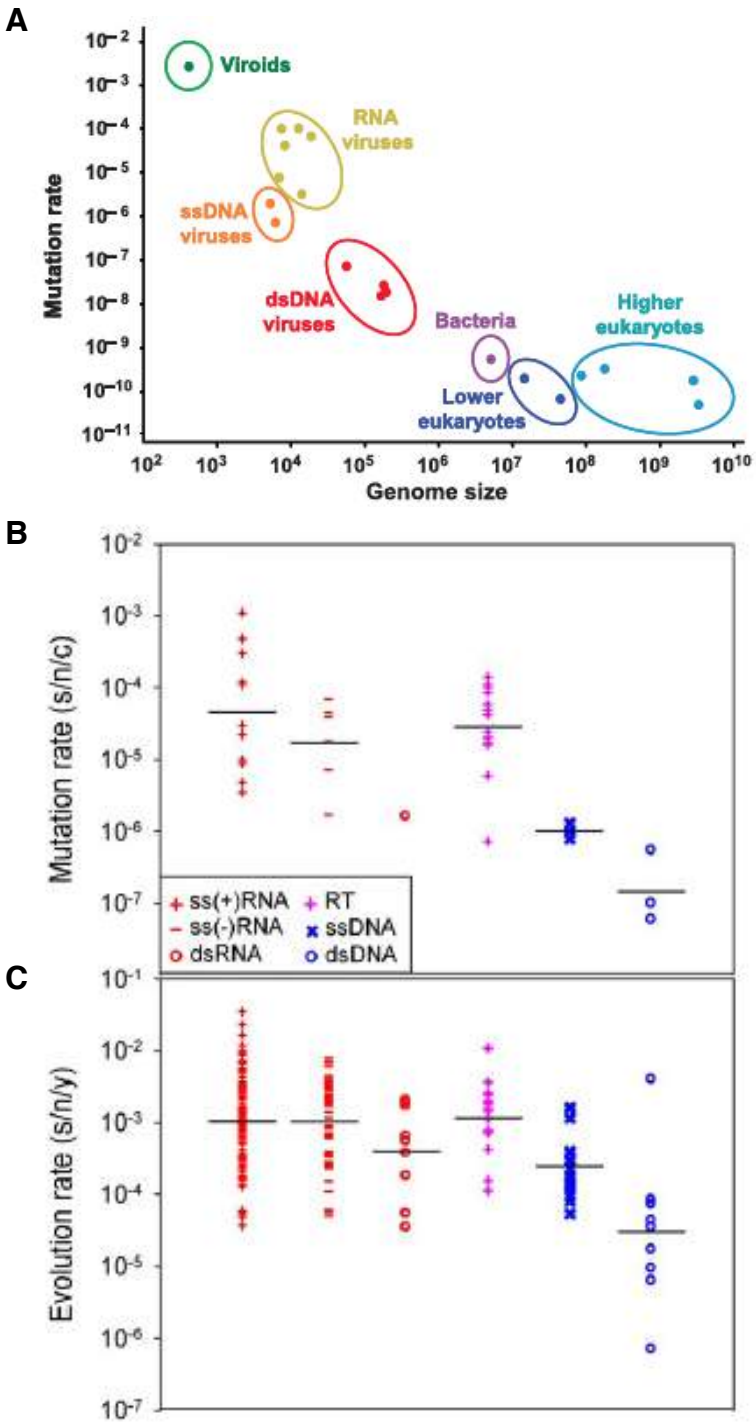


Figure 4. Comparisons of viral mutation and evolution rates. (A) Evaluation of mutation rates, as compared to genome size, for various biological entities. (B, C) Mutation (B) and corresponding evolution (C) rates, as indicated, for the various major virus groups (as defined by the Baltimore classification). Points are indicative of individual viruses. Figure adapted from [61] and [62].

To better understand this environment and the interplay between mutation and evolution, it is helpful to conceive of viral populations as mutant networks (Figure 5A). HIV-1's use of error-prone reverse transcriptase and non-proofreading RNA polymerase II provides characteristic low-fidelity replication, similar to other RNA viruses, and results in a mutation rate for HIV-1 of approximately 1.4-3.4 x 10^{-5} mutations/nucleotide/cycle [79,80]. This, combined with the large population size typical of HIV-1, has been suggested to mean that every single point mutation, and some double mutation combinations, could theoretically be produced during each replication cycle in an infected individual (thus nearly every 1-2 days [81]), leading to an extremely genotypically diverse population over time [82]. However, previous work has already determined that, in addition to high mutation rates, RNA viruses also have comparatively low tolerance to mutation [83]. Therefore, within the framework of a mutant network (Figure 5A), some of the randomly introduced mutations will maintain a fitness level, or replication competency, which is comparable to the parental sequence (illustrated by red circles), while many others will have moderate (green circles) or severe (gray circles) fitness defects. In this way, error-prone replication proceeds as an exploration of available sequence space, and large populations make for rapid, moderately low-risk travel through this space.

Still, not all entities, or RNA viruses for that matter, can tolerate mutation to the same extent. Thus, the term genetic robustness can be employed to describe a biological entity's capacity to maintain function in the face of mutation

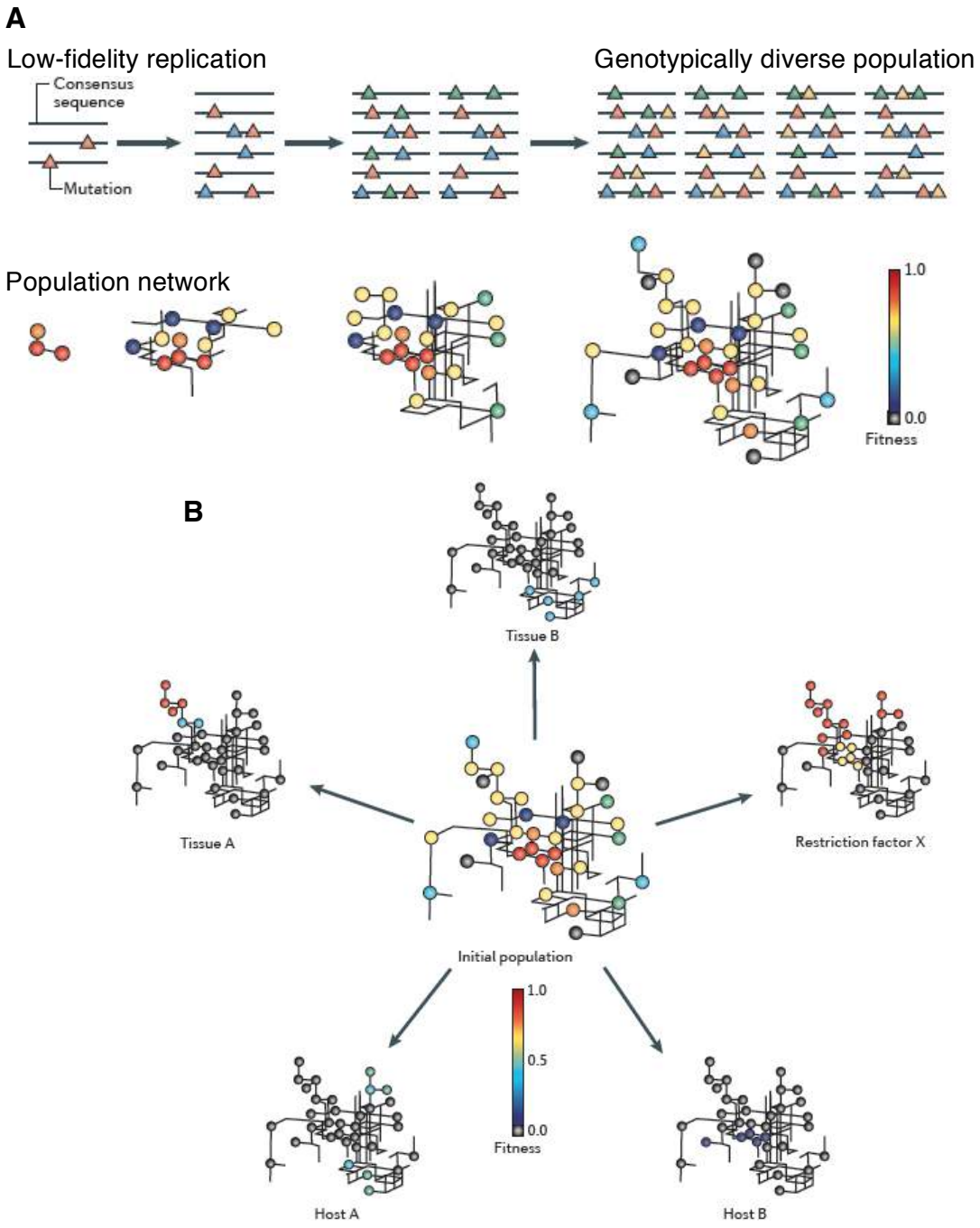


Figure 5. Viral networks and their dynamics in hosts. (A) Schematic demonstrating how mutations (colored triangles) are acquired from the consensus sequence (gray line) due to the characteristic low-fidelity replication in RNA viruses. The population network indicates how RNA virus populations can be illustrated as networks with genetic variants (colored circles) of differing fitness that are linked by mutational pathways. (B) Illustration of how viral networks shown in (A) might behave when under selection in a natural host. The center diagram shows the initial viral population, and variants are indicated by circles (colored by fitness) connected by the network (black lines). Overall, an indication is given of how robustness might be impacted by: differences in host immune pressures (Host A vs. Host B), the presence of a restriction factor, or replication in diverse tissues. Figure adapted from [82].

[84,85]. More robust proteins or organisms can tolerate higher mutation rates, while less robust, or fragile, proteins or organisms are more intolerant of mutation and are likely to lose function or to be driven to extinction by high mutation rates. Determining genetic robustness typically involves quantification of mutational fitness effects for individual mutations, in comparison with wild type fitness. Generally, at least in the case of viruses, this is achieved by the introduction of random point mutations into the genome and the subsequent measurement of the mutation's effect on replicative efficiency [86,87,88]. However, it is also possible to more indirectly measure robustness of viruses by quantifying their sensitivity to nucleoside analogues that artificially increase mutation rates [74,89]. In both cases, (subjectively) large mutational fitness effects are indicative of fragility, while smaller fitness effects suggest robustness. In this way, robustness can be calculated for whole entities, such as entire RNA virus genomes, or for smaller segments such as proteins or domains. However, it is expected that measurement of robustness for large entities like whole genomes may mask the variation in robustness that occurs amongst individual proteins or regions within that genome, in accordance with the specific functions of the individual proteins or domains.

At present, there are no known absolute determinants of an entity's genetic robustness. Instead genetic robustness of an entity like HIV-1, or its proteins, likely depends on a variety of specific characteristics, primarily concerning individual functions and the natural host environment.

Correspondingly, the robustness of either distinct viruses, or proteins or domains within a virus, can vary widely, as the extreme robustness of the HIV-1 LTR, in which data indicated that 100% of point mutants maintained WT-like levels of infectiousness [90], and the relative fragility of HIV-1 protease [91], in which just 46% of mutants maintained viability, illustrate. To this end, it has been previously theorized that proteins that perform multiple or complex functions that are highly dependent on structure, such as enzymes (like protease), should tend to be more genetically fragile and display greater sequence conservation than those that do not [92]. At the same time, population genetics theory predicts that robustness is more likely to be selected for when mutations rates are high, as it should be advantageous [84,92]. Nevertheless, previous data has also shown that robustness can vary even amongst whole virus genomes with high mutation rates, as only 20% of single mutations in bacteriophage ΦX174 were considered lethal [87], while more than twice as many, 41%, of point mutations in tobacco etch potyvirus had the same effect [93].

As Figure 5B demonstrates, the host environment also has a large influence on the available sequence space a population can utilize, and, thereby, its overall robustness. Specific host features including interactions with restriction factors or other cellular factors, tissue tropism, and other immune pressures can impact the fitness of individual mutations, which in turn impact the greater fitness landscape and the further exploration of sequence space [82]. In the case of HIV-1, adaptive immune responses play a particularly significant,

hostile and changing environmental role. These immune responses can take the form of antibodies that target proteins displayed on the surface of virions, or cytotoxic lymphocytes (CTLs) that target epitopes in a number of viral proteins. As a result, some viral proteins, even those that should exhibit genetic fragility as a reflection of their complex or essential nature, are placed under strong evolutionary pressure to diversify in sequence, despite fitness costs, in order to evade adaptive immune responses. Under these conditions, otherwise fragile proteins might evolve higher robustness. Ultimately, the pressures impacting robustness can be narrowed to two broadly competing categories: those inherent pressures imposing conservation of sequence in order to preserve specific and essential functions, and opposing pressures to diversify sequence in order to avoid a hostile host environment. However, the outcomes regarding any potential trade-off between robustness and fitness are not established, so it is unclear whether fitness, or acquisition of robustness, would constitute a dominant selective force on a viral protein in the natural setting. It should be noted, nevertheless, that initial examinations have suggested that, at least under certain circumstances in viral populations, high robustness can be favored at the expense of high fitness, indicating robustness may be the more potent selective force [94,95,96].

Intriguingly, both theories and experiments concerning the evolution of robustness have yielded a common observation: ‘survival of the flattest’ [95,97,98] (Figure 6). In further accordance with Sewall Wright’s 1932 adaptive

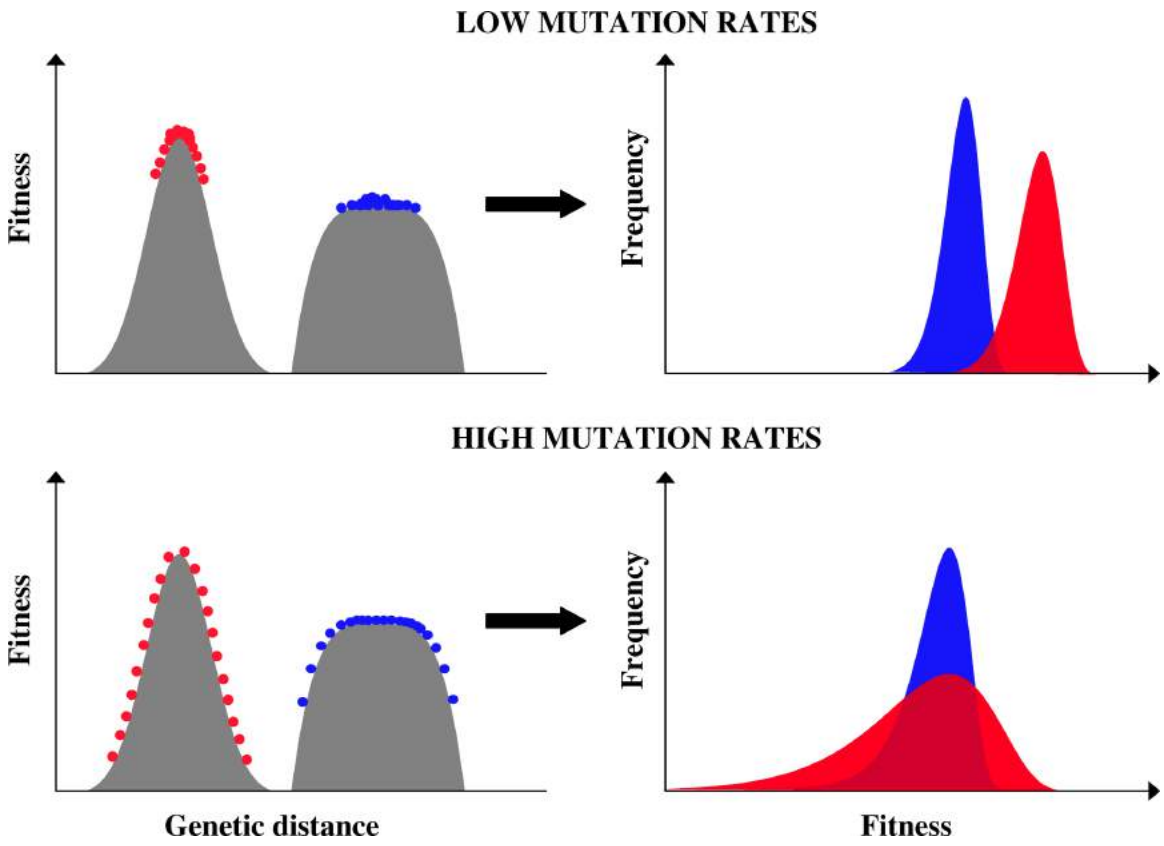


Figure 6. Lethal mutagenesis and the survival of the flattest. Red indicates populations with fast replication, but that are genetically fragile, while the blue shows 'flat' populations that have slower replication, but are more robust. The panels on the left show populations with either high or low mutation rates, and the right panels demonstrate the expected distributions of the two types of populations under the different mutation conditions. When mutation rates are low, the fitter population will always successfully outcompete the 'flatter' population. However, this can be reversed when mutation rates are high, as the tops of the high narrow peak will display the highest fitness, but the other population will have less variance in fitness due to its increased robustness. Figure adapted from [96].

landscape model [99], these results have indicated that at low mutation rates, quickly replicating populations (in red, high fitness populations with little genetic distance) are expected to always outcompete the slower replicating populations (in blue, populations with more genetic distance) that have a low, flat peak and higher robustness [96]. However, this situation can be reversed at high mutation rates (Figure 6, bottom panels), when the flatter population with less variance in fitness may be more adaptive. Consequently, for HIV-1, with its high mutation rate, a flatter, more robust population with greater genetic variance might be expected to be favored over more fit but fragile populations.

At the same time, a number of studies, using either mutagenic nucleoside drugs or variant RNA-dependent RNA polymerases, have shown that the mutation rate in RNA viruses is extremely close to a catastrophic error rate that would lead to extinction [100,101,102], indicating there may be some degree of fragility in these populations. These 'lethal mutagenesis' studies have sought to determine the threshold at which the mutation rate becomes biologically intolerable and viral infectivity is eliminated. From a therapeutic point of view, examination of lethal mutagenesis for HIV-1 offers great potential as an alternative to antiretroviral treatments susceptible to drug resistance [73]. This is especially true given that natural hosts contain their own source of hypermutagenesis in APOBEC3G (normally counteracted by Vif) [103,104,105]. Hence, examination of the genetic robustness or fragility of HIV-1 and its components may contribute to therapeutic vaccine and drug development efforts

not only through the revelation of regions remarkably sensitive to mutation, but also through discovery of how close HIV-1 may already be to its own error catastrophe.

HIV-1 capsid

As previously mentioned, evaluation of the robustness of whole viral genomes can mask significant variability amongst proteins or domains within, in accordance with specific functions or host pressures. For this reason, evaluation of a single essential protein in HIV-1, such as capsid (CA), may provide a more thorough, albeit localized, representation of genetic robustness and its causes and consequences. Capsid is a particularly appropriate protein in which to measure robustness as it experiences competing needs to both preserve function, yet diversify sequence, in an acute manner. As a critically important protein, it is placed under intense selective pressure to maintain its structure and perform several functions. Conversely, as a highly expressed, immunologically visible protein, CA also assumes a competing pressure to diversify its sequence, and it is unclear how these competing pressures might impact the protein's genetic robustness.

Among the unique features of HIV-1 CA are its multifunctional nature and its presence in different structural orientations—it exists as both a domain within the Gag precursor polyprotein during particle assembly, and as an autonomous protein in mature virions (Figure 7) [11,106,107]. As part of Gag, CA is sandwiched between matrix (MA) at the 5' end, and short spacer peptide 1 (SP1)

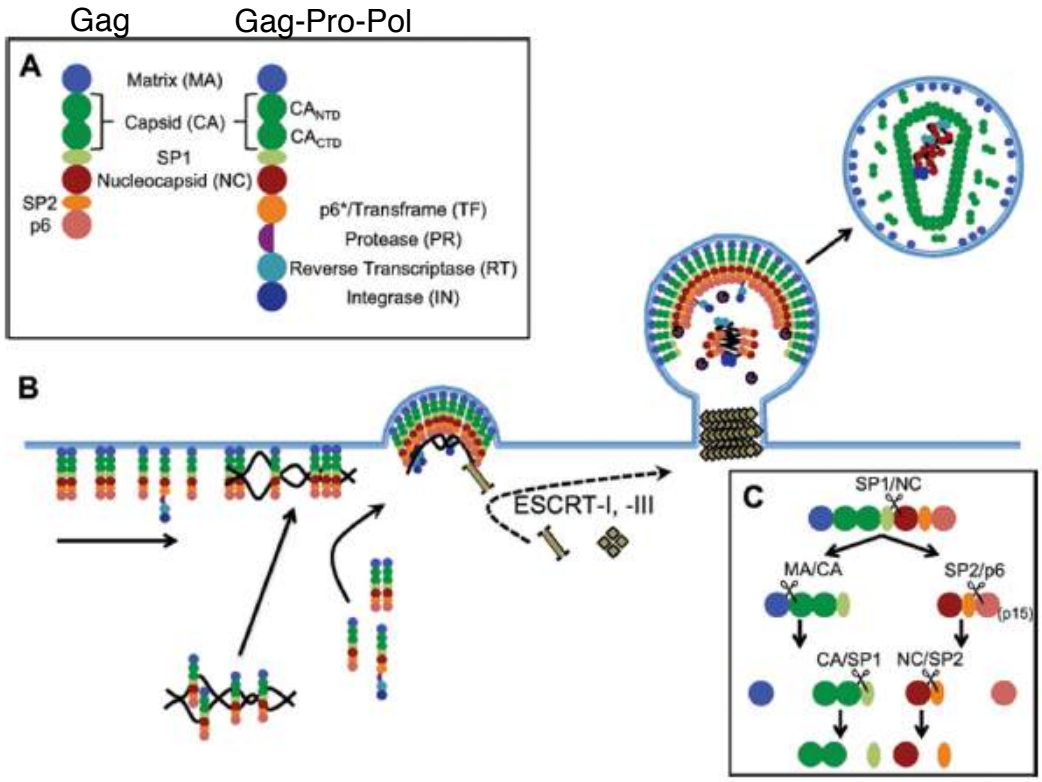


Figure 7. Diagram of HIV-1 assembly. (A) Legend indicating the identity and corresponding color of the respective proteins comprising Gag or Gag-Pro-Pol. Capsid is shown in the darker green color. (B) Outline of the steps that occur during assembly at the plasma membrane and in subsequent viral budding and maturation. (C) Sequential illustration of the HIV-1 Gag polyprotein's proteolytic cleavage. Image adapted slightly from [108].

at the 3' end (Figure 7A). Gag also includes nucleocapsid (NC), spacer peptide 2 (SP2), and p6, which follow SP1 sequentially, whilst Gag-Pro-Pol additionally contains the viral enzymes (protease, reverse transcriptase, and integrase) at the 3' end of Gag (Figure 7A) [108]. In the initial stages of virion morphogenesis, Gag molecules assemble, largely through the targeting of MA, at the plasma membrane, along with additional Gag-Pro-Pol molecules (Figure 7B), to drive the formation of roughly spherical virions, which contain radially arrayed Gag molecules that bud through the plasma membrane [109,110]. During and after subsequent viral budding, protease is activated and catalyzes cleavage of Gag in five positions (Figure 7C), causing extensive morphological transformations and the transition from an immature to mature viral particle [14]. As a result, the liberated CA protein forms a distinctive conical capsid that encapsulates the nucleocapsid-genomic RNA complex [111,112].

Like all retroviral capsids, the 231 amino acid (aa) HIV-1 CA molecule is composed of two domains: the N-terminal domain (NTD), which contains the first 146 aa, and the C-terminal domain (CTD), which contains the remaining 85 aa. The structure of the NTD comprises an N-terminal β hairpin followed by 7 α helices, while the CTD consists of 4 α helices and a final C-terminal unstructured region of 11 residues (Figure 8B, 8C and 8D in white) [113,114,115]. In between the NTD and CTD lies a small interdomain linker region (residues 146-150). Other noteworthy regions in CA include the major homology region (MHR, residues 153-172) of the CTD, which is a highly conserved 20 amino acid region

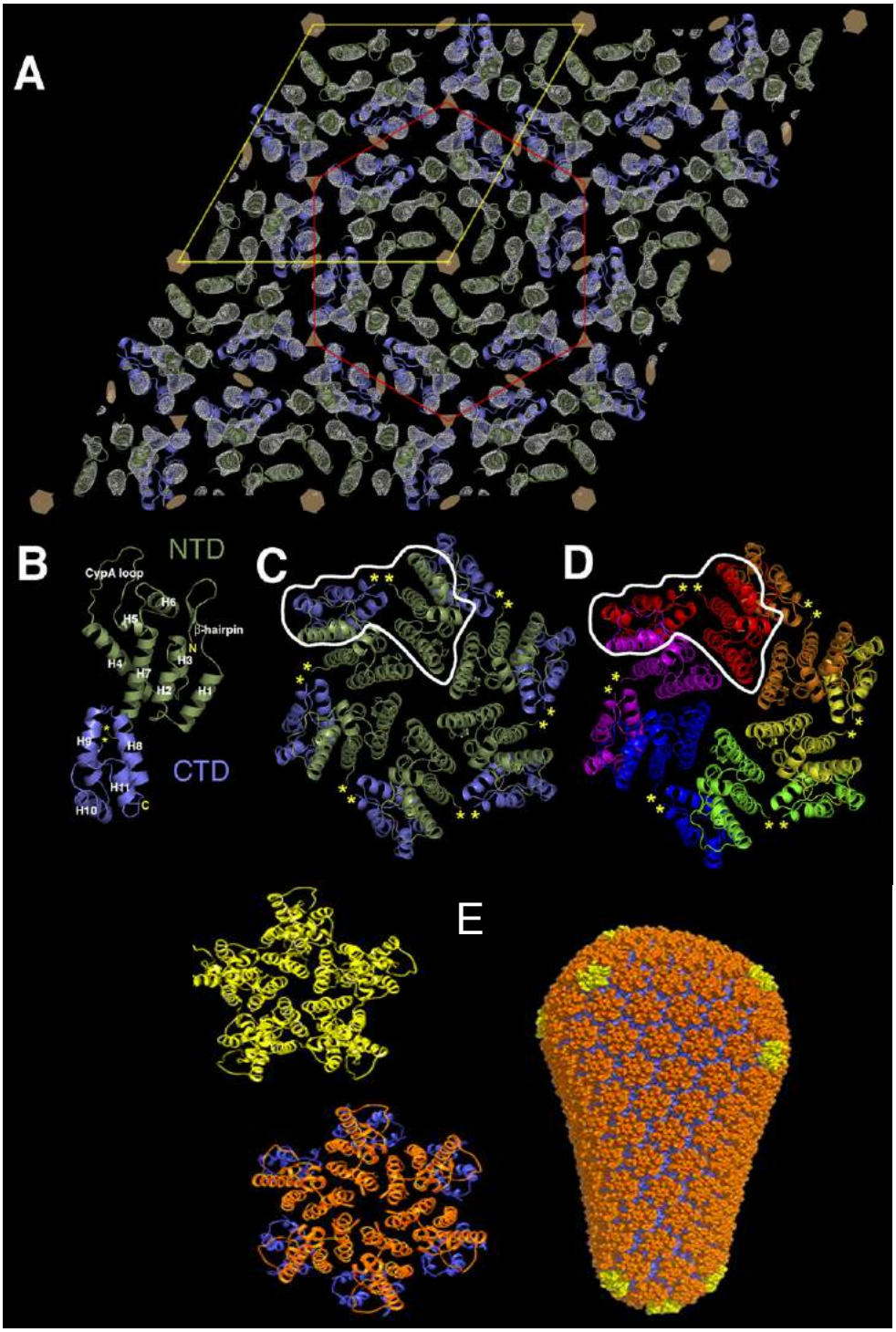


Figure 8. HIV-1 CA structures. (A) A slab view illustrating an experimental map of CA hexamer formation, with the CTD in blue, the NTD in green, and a red line outlining the hexamer. (B) A pseudoatomic model of the full-length CA monomer. Individual helices and loops are labeled accordingly. (C, D) External views of the CA hexamer, in which a single monomer is outlined in white. (E) Stereoview of the complete CA fullerene cone, comprised of approximately 1,056 subunits, and including 12 pentameric declinations. The pentamer structures are shown in yellow, while the hexamer NTD is in orange and the hexamer CTD is in blue. Figure constructed with images from [116] and [117].

found in all retroviruses [118] and the likely binding site of host protein ABCE1 [119], and a loop (NTD residues 85-93) which binds the cellular protein cyclophilin A (hereafter referred to as the cyclophilin binding loop) [120].

As will later become evident, the structure of CA is particularly relevant to its robustness, and thus its components warrant further description. The mature HIV-1 capsid includes ~1056 CA monomers arranged in a hexameric lattice (Figure 8A, and 8E in orange/blue) interspersed with 12 pentameric declinations (Figure 8E, in yellow) which allow the viral capsid to form a fullerene cone (Figure 8E) [111,117]. Within the hexameric lattice (Figure 8A, C, E), the NTD creates the hexamer ring shape (Figure 8A, C in green, 8E in orange) while the CTD forms the dimeric linkers that link neighboring hexamers (Figure 8A, C, E in blue) [116,121]. Correspondingly, there are three primary protein-protein interactions in the mature capsid lattice: NTD-NTD, CTD-CTD, and NTD-CTD. The NTD-NTD hexamerization interface is mediated by a six-fold symmetric interface involving helices 1, 2, and 3 (to form an 18-helix bundle) [116,122]. The homodimeric CTD-CTD interface links the hexameric building blocks and occurs through a parallel association at helix 9 (around residue 185), although other CTD regions likely contribute to the stability of this arrangement [115,123]. Finally, the hexamer is reinforced through stabilizing intermolecular NTD-CTD interactions, primarily involving the extended side chains from CTD helix 8 which maintain interfaces with NTD helices 3-4 (and the intervening loop) [123,124,125]. Protein interactions in the relatively rare pentameric declinations (Figure 8E, in yellow)

are similar to those found in hexamers, with only subtle rearrangements in the subunit interactions [117]. Disruptions to any of the interaction interfaces are likely to disrupt proper capsid assembly, causing irregularly shaped capsids and less- or noninfectious particles [123]. Additionally, the dramatic differences between the conformations of immature and mature capsid suggest the presence of additional protein-protein interfaces and interactions not found in the mature capsid, and if these were disturbed, would also be likely to negatively impact assembly [110].

In addition to contributing a major structural role, CA determines several other biological properties of HIV-1. For instance, HIV-1 capsid has been demonstrated to be the key determinant that permits HIV-1 to infect non-dividing cells [126,127]. Other work has identified capsid as having a number of genetic or physical interactions with various host proteins, including karyopherin β transportin-3 (TNPO3) [128], nucleoporin 153 (NUP 153) [129,130], nucleoporin 358 (NUP358)/RanBP2 [131], and cleavage and polyadenylation factor 6 (CPSF6) [132]. Moreover, the interactions between HIV-1 capsid and host protein cyclophilin A are also known to influence nuclear import and subsequent integration site selectivity, as well as replication efficiency, in a cell type dependent manner [131,133,134,135].

Capsid is also a major target of intrinsic, innate, and adaptive immune defenses. Intrinsically speaking, capsid has recently been shown to be a determinant of HIV-1's sensitivity to the antiviral factor Myxovirus resistance 2

(Mx2) [136], an interferon-induced, dynamin-like GTP-ase that can inhibit early steps of HIV-1 infection *in vitro*. Reports of CA residues affecting Mx2 restriction of HIV-1 have thus far been limited to CA residues within the cyclophilin binding loop (and one other residue, N57, previously reported to affect HIV-1 infection of non-dividing cells [127]) [136,137,138], however the likelihood that other residues within CA determine sensitivity to Mx2 will be further explored in the results section. CA is also targeted by antiretroviral factor TRIM5α, and is thereby a major determinant of its restriction capacities [139,140]. Alternatively, as a target of the innate immune defenses, CA is believed to be detected by undefined sensors in dendritic cells [141].

HIV-1 CA is also under pressure from adaptive immune forces [142,143,144,145]. In particular, host human leukocyte antigen (HLA)-associated CD8+ cytotoxic T lymphocyte (CTL) responses to HIV-1 infection are significant determinants of viral control [146,147], and there is an association between Gag- and capsid-specific CD8+ T-cell responses and the *in vivo* viral burden [148]. Moreover, the emergence of viral 'escape' mutations in CA has been shown to result in higher viremia in infected patients [149,150]. However, CA escape mutations often incur significant fitness costs, which may be subsequently compensated for by secondary mutations in CA, or result in reversion when immune pressure is lifted [144,150,151].

A large number of earlier studies have mutated HIV-1 capsid with the aim of elucidating the importance of particular domains, regions, and residues in CA functions. Such targeted mutagenesis studies have largely relied upon insertion

[152], deletion [153], or alanine or proline scanning [154,155]. While certainly informative of the impact a single residue mutation could have on infectivity, core morphology, or the like, these studies are not necessarily predictive of the phenotypes that might result from naturally-occurring mutations that arise through errors during reverse transcription or from RNA polymerase II. In our CA mutagenesis approach, we will attempt to utilize more unbiased means of inducing mutation, on a far larger scale, and in so doing, measure the genetic robustness of CA and determine it's biological causes and implications. These are, of course, deeply rooted in the many functions and interactions of HIV-1 capsid.

HIV-1 integrase

To establish whether any trends observed from evaluation of the genetic robustness of HIV-1 capsid might be unique to CA, or a larger reflection of the robustness of the entirety of HIV-1, it seemed prudent to evaluate the genetic robustness of another important HIV-1 protein. HIV-1 integrase (IN), like HIV-1 CA, is a highly conserved, essential, multifunctional protein, and is the enzyme responsible for the integration of the viral genome into the host chromosome. Within the framework of the viral life cycle, IN functions just after reverse transcription, and is responsible for two necessary catalytic activities, 3' processing and DNA strand transfer. It has previously been hypothesized that the requirements needed to fulfill enzymatic catalytic activities like these, which involve adopting and changing tertiary structures in subtle, flexible ways, may

mean enzymes are particularly sensitive to mutation [92]. The relatively high degree of conservation in integrase, especially within the protein's catalytic core domain [156], also suggests IN may be not be tolerant of mutation.

While the crystal structure for the functional HIV-1 IN unit, the intasome, has not yet been determined, significant insight exists regarding the domain structures and functions in the 32 kDa, 288 amino acid HIV-1 IN protein (Figure 9A). At the most N-terminal end, the roughly 49 amino acid N-terminal domain (NTD) adopts a helix-turn-helix fold, contains a highly conserved HH/CC zinc binding domain (at residues 12, 16, 40, 43, Figure 9A), and contributes to stabilization, multimerization, and DNA binding [157]. The adjacent catalytic core domain (CCD), residues 50-212 (also sometimes reported as 59-202), is responsible for both 3' processing and strand transfer reactions, and also contributes to multimerization and both viral DNA (vDNA) and host target DNA binding [158,159]. This domain adopts a RNase H fold, and contains a D, D(35)-E motif that binds divalent metal ions, largely Mg^{2+} [159,160]. At the C-terminal end, the C-terminal domain (CTD), which includes residues 213-288 (or 223-270), adopts a beta barrel resembling an Src homology 3 (SH3)-type fold, is significantly involved in binding to host cell target DNA, and like the other domains, also contributes to multimerization [159,161,162]. Ultimately, the three domains in each monomer come together to form a tetramer comprised of a dimer of dimers, which, when bound to the vDNA, forms the intasome [163]

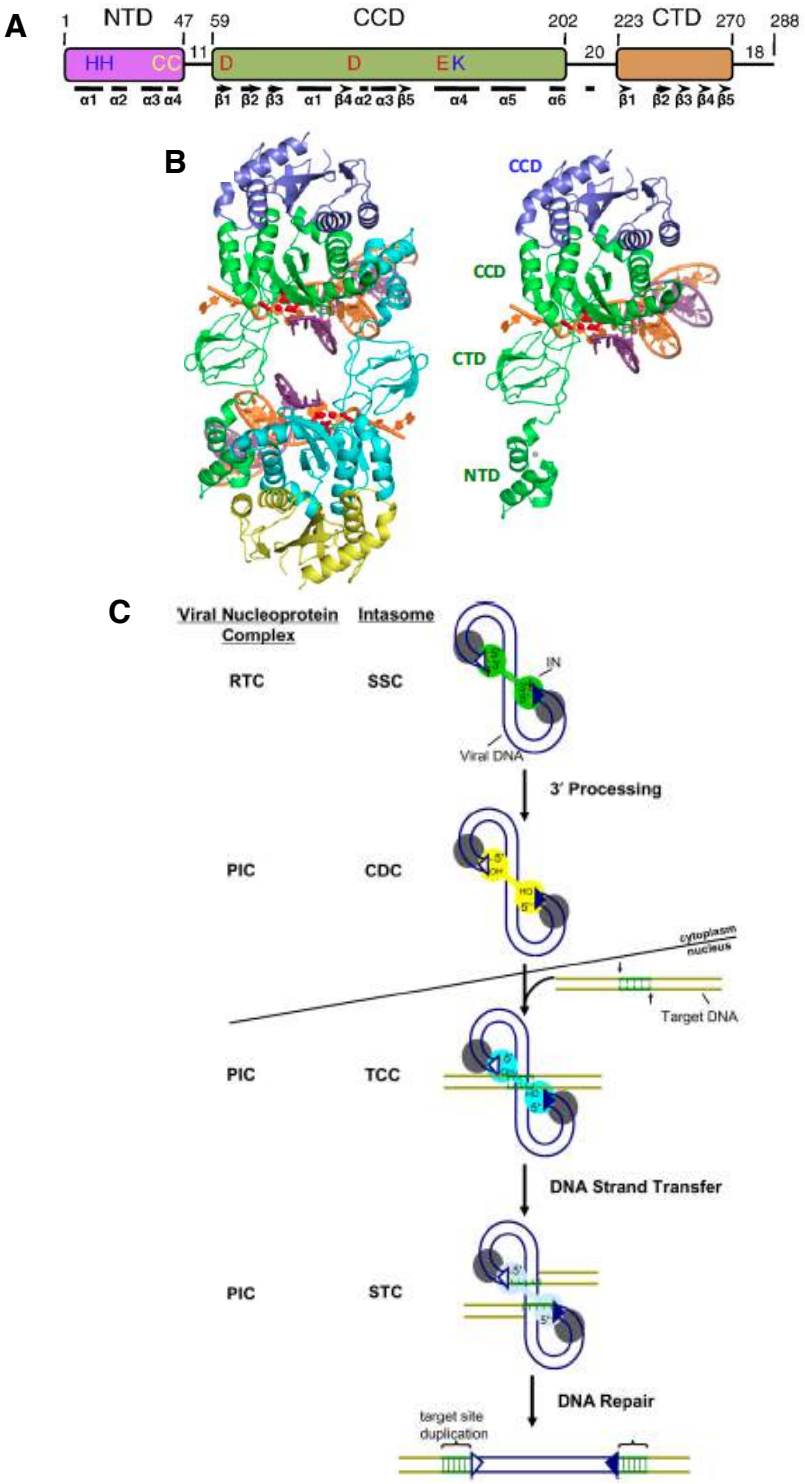


Figure 9. Overview of HIV-1 IN, the intasome, and integration. (A) Organization of the HIV-1 IN protein, indicating the approximate individual domains, their sizes, and the notable elements within. (B) First published model of the HIV-1 intasome, with the image on the left showing the full IN tetramer. The right image indicates the IN dimer and the domains within. (C) Schematic of the retroviral integration processes, indicating the retroviral nucleoprotein complexes formed during the IN-mediated DNA cutting and joining of the DNA. Figure constructed with images from [164], [159], [165].

(Figure 9B). The publication of the crystal structure of the prototype foamy virus (PFV) intasome has provided further insight into the possible arrangement of HIV-1 IN's three domains within the tetramer, and has suggested how the domains might contribute to vDNA and host DNA binding [166,167]. Furthermore, the PFV crystal structure, and the knowledge it has conveyed, has led to new, improved HIV-1 intasome models [164,168] (Figure 9B).

The catalytic activities of integrase begin within the confines of the reverse transcriptase complex (RTC) in the host cell cytoplasm (Figure 9C). After reverse transcriptase copies the single-stranded viral RNA genome into a linear double-stranded DNA molecule that contains a copy of the LTR at each end, multimeric (likely tetrameric) integrase binds the newly synthesized LTR ends to form the retroviral intasome complex known as the stable synaptic complex (SSC) [165,169] (Figure 9C). During the subsequent 3' processing reaction, IN processes each LTR end at a location adjacent to an invariant CA dinucleotide, so as to yield CA_{OH} 3'-hydroxyl groups, which will later serve as the nucleophile in the second reaction [170,171]. This 3' processing reaction results in an intasome configuration known as the cleaved donor complex (CDC). At the same time, the viral replication intermediate now has the capacity to initiate DNA strand transfer activity, and this marks a transition in the viral nucleoprotein complex nomenclature from the RTC to the pre-integration complex (PIC). Upon nuclear entry and the location of a suitable target site on the host cell chromatin, tetrameric IN uses the exposed 3' oxygen atoms to cut phosphodiester bonds on

opposing strands of the target DNA (tDNA), in a staggered fashion, and it also simultaneously joins the viral DNA ends to the tDNA's 5' phosphates [172,173,174] (Figure 9C). The initiation of the DNA strand transfer reaction, in which the CDC engages the tDNA, leads to the formation of the target capture complex (TCC), and the subsequent integration of the viral DNA 3' ends results in the formation of the strand transfer complex (STC). Finally, host cell enzymes repair the unjoined viral DNA 5' ends, to yield the fully integrated provirus.

Determination of these enzymatic functions of IN has often involved mutagenesis of IN. Scanning and site-directed mutagenesis work has been critical to demonstrating the structural and functional contributions of both whole subunits and individual residues within IN [175,176,177,178,179,180,181]. Furthermore, deletion mutants of various subunits, or parts thereof, have also proved individual domain contributions to both HIV-1 IN catalytic activities, and initially indicated the likelihood that IN forms a functional multimer [182,183,184]. Overall, however, mutations in IN have typically been divided into two classes [185]. Class I mutants are those that directly affect either of the enzymatic activities of IN (which in turn, affect infectivity), while class II mutants reduce infectivity at a stage other than integration. In some cases, class II mutants have also been shown to yield virions with abnormal cores [186]. Despite the large quantity of IN mutagenesis studies probing specific function, no study has randomly or extensively mutagenized the entirety of HIV-1 integrase yet, nor has any study sought to evaluate its genetic robustness.

Mutagenesis of IN has also helped demonstrate interactions between IN and its cellular co-factors. Like CA, IN is known to interact with a number of host proteins, some of which are required for efficient levels of infectivity. Perhaps the most important of these co-factors is the essential lens epithelium-derived growth factor (LEDGF/p75), which helps direct integration into chromatin sites within genes, or regions with active transcription units [187,188]. Certain mutations within IN, particularly in the C-terminal region of the catalytic core domain, have the ability to disrupt the interaction between IN and LEDGF and correspondingly impact infectivity [189]. HIV-1 integrase is also believed to interact with IN interactor 1(INI1)/hSNF5 [190], as well as the cellular co-factor transportin SR-2 (TRN-SR2, or TNPO3), a karyopherin which helps mediate nuclear import [191], amongst other less confirmed interactions [192,193].

Despite its interactions with the cellular environment, and unlike HIV-1 CA, HIV-1 IN does not appear to be the target of intense, or, rather, successful immune responses. At present, there are no known intrinsic or innate immune defenses that target IN. Furthermore, while certain human leukocyte antigen (HLA)-associated CD8+ cytotoxic T lymphocytes (CTL) epitopes to CA can be protective, and correlated with reduced viral loads, replication capacity, or non- or delayed progression to AIDS [194,195], the same does not appear to be true for integrase [148]. Integrase can be a target of CTL responses, including some associated with protective HLA class I alleles [196,197,198], and individuals with reduced viral loads can very occasionally show T-cell responses against

integrase [199]. Nevertheless, more conclusive evidence suggests that there is no substantial immune-driven attenuation of integrase during HIV-1 infection [200].

As anti-retroviral treatments for HIV-1 have progressed, integrase has become an increasingly important target of HIV-1 therapeutic treatment. In fact, a multitude of integrase inhibitors targeting the individual integration steps have been developed, and these can be divided into six general classes: DNA-binding inhibitors, 3' processing inhibitors, nuclear translocation/import inhibitors, strand transfer inhibitors, gap repair inhibitors, and the allosteric inhibitors [156,201]. At present, the integrase strand transfer inhibitors (INSTIs), raltegravir, elvitegravir, and dolutegravir, have been the most clinically successful class of inhibitors [165,202]. However, the susceptibility of these drugs to resistance mutations has meant that the search for new integrase inhibitors continues [202]. Work has increasingly focused on the allosteric integrase inhibitors (ALLINIs), which may work by competing with LEDGF/p75 for binding to integrase, thereby disrupting integrase assembly with viral DNA and inhibiting enzyme function [203]. Furthermore, recent work has shown that administration of a long-acting integrase inhibitor can protect macaques from infection by a simian/human immunodeficiency virus [204]. Evaluating the genetic robustness of HIV-1 IN therefore seems a particularly worthy task, as it presents the possibility of uncovering regions of vulnerability or fragility that could prove valuable to novel inhibitors or other therapeutic treatments. Moreover, comparison of IN

robustness with that of CA should reveal more about how the competing pressures to maintain sequence for function preservation, and also to diversify sequence to avoid immune pressure, might vie in different circumstances.

II. MATERIALS AND METHODS

Plasmids

Throughout the described work, production of infectious HIV-1 was achieved using derivatives of a proviral plasmid previously used in our laboratory, pNHG (accession: JQ585717) [205]. Containing full-length HIV-1, with the exception of Nef, which is replaced with EGFP, pNHG expresses NL4.3 GagPol and HXB2 envelope (the remaining portions being previously derived from pNL/HXB and R7/3/GFP [206,207]). All of the work done to construct the CA mutant library and evaluate its genetic robustness was done with a derivative of pNHG, pNHGcapNM (accession: JQ686832), created by Sam Wilson [208]. Almost identical to pNHG, pNHGcapNM differs only in that it contains three additional restriction enzyme sites, *Not*I and *Mlu*I that roughly flank the capsid protein, and a *Spe*I site that lies around the middle of capsid. These sites were created through synonymous substitutions, with the *Not*I site added in matrix approximately 36 nucleotides 5' to capsid, while the *Mlu*I site lies within capsid, approximately 12 nucleotides from its 3' end.

Work done to synthesize and characterize the HIV-1 integrase mutants was done with a similar derivative of pNHGcapNM, pNHGintBS. To create this proviral plasmid the first required step was to eliminate the *Spe*I site in capsid of pNHGcapNM. This was done by amplifying the region between the *Not*I and *Mlu*I sites, using primers with overlapping sequence that eliminated the *Spe*I site in a synonymous fashion (sense primer: 5'- CCA AGG GGA AGT GAC ATA GCA

GGA ACT ACG AGT ACC CTT CAG GAA CAA ATA GG -3' and antisense primer: 5'-CCT ATT TGT TCC TGA AGG GTA CTC GTA GTT CCT GCT ATG TCA CTT CCC CTT GG -3'). Using the overlapping PCR scheme, the two products from a first round of PCRs (the *Not*I-antisense primer product and the sense primer-*Mlu*I product) could be used as templates with external primers incorporating the *Not*I and *Mlu*I sites in another PCR reaction, after which the final PCR product was gel-purified and cloned back into pNHGcapNM. Next, it was necessary to create the new *BstB*I and *Spe*I sites that could flank integrase. Creation of the *BstB*I site was done using a similar overlapping PCR scheme, with amplification of a region between *Mlu*I and *EcoR*I sites. Again, utilizing synonymous nt substitutions to create the *BstB*I site (sense primer: 5'- GGT TGG TCA GTG CTG GAA TTC GAA AAG TAC TAT TTT TAG ATG G -3' and antisense primer 5'- CCA TCT AAA AAT AGT ACT TTT CGA ATT CCA GCA CTG ACC AAC C -3'), two products from a first round of PCRs (containing a *Mlu*I-antisense primer product and a sense primer-*EcoR*I product) were weaved together with external primers including the *Mlu*I and *EcoR*I sites. Following gel purification of the final PCR product, it was cloned back into the pNHGcapNM lacking *Spe*I, using the *Mlu*I and *EcoR*I sites. Addition of a new *Spe*I site was the final step to completing pNHGintBS. Once more, the overlapping PCR scheme was used, with amplification of the region between *Mlu*I and *EcoR*I sites. Synonymous nucleotide substitutions were used to create the *Spe*I site (sense primer: 5'- GGA TTA ACA CAT GGA AAA GAC TAG TAA AAC ACC ATA

TGT ATA TTT CAA GGA AAG C -3' and antisense primer: 5'- GCT TTC CTT GAA ATA TAC ATA TGG TGT TTT ACT AGT CTT TTC CAT GTG TTA ATC C -3'), such that products from a first round of PCRs (containing the *Mlu*I-antisense primer product and a sense primer-*EcoR*I product) were used as templates that could be weaved together with external primers including the *Mlu*I and *EcoR*I sites in a second reaction. The final PCR product was then gel-purified and cloned into the pNHGcapNM containing the *BstB*I site (and the eliminated *Spe*I site in capsid) using the *Mlu*I and *EcoR*I sites. As a result, pNHGintBS was created (from pNHGcapNM), so that a *BstB*I restriction site is at the 5' end 9 nt outside integrase (in reverse transcriptase) and a *Spe*I site is at the 3' end 15 nt outside integrase in Vif.

The final plasmid used in the described work is pNL4.3 Pr- (sometimes referred to just as Pr-), which has also been previously used in our lab [209]. This plasmid is a derivative of pNL4.3 (NIH AIDS Research and Reference Reagent Program, Catalog No. 114), in which protease is inactive, due to the presence of a point mutation in the protease active site [209]. The capsid mutants used in this plasmid were digested from pNHGcapNM and cloned directly into pNL4.3 Pr- using compatible *Not*I and *Mlu*I sites found flanking capsid in each plasmid.

Construction of CA and IN mutant libraries

Construction of the unbiased CA mutant library was completed by Sam Wilson using a low fidelity PCR approach. The Genemorph II random

mutagenesis kit (Agilent) was used with the following oligos: sense – 5' –GTA AGA AAA AGG CAC AGC AAG CGG CCG CTG - 3' and antisense 5' – CTT GGC TC A TTG CTT CAG CCA AAA CGC GTG - 3'. The resulting PCR product was cloned using a TOPO TA Cloning Kit and then plasmid DNA was extracted from approximately ~1x10^4 pooled, insert-positive colonies. Following sequencing of the amplicon from 20 clones, in order to obtain a preliminary estimate of the mutagenesis frequency, the pooled plasmid DNA was digested using *Not*I and *Mlu*I, and the pooled CA-library insert was subcloned into pNHGcapNM (Figure 10A). The *Not*I and *Mlu*I digested proviral plasmid had been prepared in advance and generated >200-fold fewer colonies when ligated without insert than when ligated with insert. Additionally, representative restriction digests indicated that aberrant restriction patterns, which could be caused by recombination, occurred at a frequency of only 1-2% of clones. Finally, proviral plasmid DNA was extracted from individual cultures derived from 1056 colonies and subjected to sequencing and further analysis. For the single mutants listed in Tables 1 and 2, proviral plasmid DNA was freshly isolated, re-sequenced and analyzed by *Not*I and *Mlu*I restriction digest.

Creation of the IN mutant library was completed in a manner similar to that done for the CA mutant library. As before, the Genemorph II random mutagenesis kit (Agilent, Cat. No. 200550) was utilized, this time with the following oligos precisely flanking IN: sense 5' – GGA AAT GAA CAA GTA GAT GGG TTG GTC AGT GCT GGA ATT CGA AAA GTA CTA – 3' and antisense 5' –

GCT TTC CTT GAA ATA TAC ATA TGG TGT TTT ACT AGT CTT TTC CAT GTG TTA – 3', in several separate reactions with varying amounts of template pNHGintBS (400-700 ng), in order to determine the template quantity providing the optimal number of single residue mutants. Both the mutagenized PCR products and the proviral plasmid pNHGintBS were then digested with *BstB*I and *Spe*I, ligated, and cloned (Figure 29A). Cloning in pNHGintBS without insert generated >500-fold fewer colonies than cloning with the mutagenized insert. To ensure the appropriate mutagenesis frequency, preliminary amplicon sequencing was done for approximately 285 mutagenized plasmid clones with varying template quantities (400, 500, 600, or 700 ng). During subsequent production of the library under the conditions yielding the highest number of single nucleotide changes (400 ng template), DNA was extracted from individual cultures derived from ~768 colonies. Amplicon sequencing then revealed which mutant clones contained only a single amino acid mutation, and were suitable for inclusion in the library. Final library proviral plasmid DNA was then freshly isolated for the single IN mutants, and subjected to analysis by restriction digest using *BstB*I and *Spe*I, at which point ~2% of clones were removed for evidence of recombination.

Cell lines and transfection

The 293T adherent human cell line was maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal calf serum (FCS) and gentamicin. Suspension cell lines used in this work were maintained in Roswell Park Memorial Institute medium (RPMI-1640), also supplemented with 10% fetal

calf serum and gentamicin, and these included MT-4 cells and also a variety of other MT-4 lines expressing species variants of the Mx2 protein (either human (Hs) macaque (Mac), African Green Monkey (AGM), ovine, or canine). Those MT-4s with variants of Mx2 were made by Sam Wilson and Idoia Busnadiego, and were generated by transductions with modified LKO-derived lentiviral vectors followed by selection with 2 μg/ml puromycin (Sigma). Expression of Mx2 variants in these cell lines was induced by overnight treatment with 125 ng/ml doxycycline hyclate (Sigma), prior to viral challenge.

In transfection experiments, 293T cells were plated at 2.5 x 10^4 cells/well in 96 well plates or 1.7 x 10^5 cells/well in 24 well plates. To measure the effects of either CA or IN mutations, transfections were done the following day using 4 μg polyethylenimine (PEI, Polysciences) per μg of DNA, and either 100 ng (96 well plates) or 500 ng (24 well plates) of the WT (NHGcapNM or NHGintBS) or NHGcapNM or NHGintBS mutant plasmids, or 500 ng of the pNL4.3 Pr- mutants (all in 24 well plate experiments).

For the temperature-sensitive experiments, plates for all transfection experiments were placed at 35, 37, or 39.5°C, as indicated, immediately upon addition of the transfection mixture.

Primary cells

Peripheral blood mononuclear cells (PBMCs), CD4+ T cells, and macrophages were isolated from buffy coats (from anonymous healthy donors, purchased from the New York Blood Center) using a Ficoll gradient (with

centrifugation at 1500 rpm, for 30 minutes, without brake). Bulk PBMCs were isolated, and washed in RPMI by centrifugation three times. Primary CD4+ T cells were extracted directly from buffy coats using negative selection via the RosetteSep Human CD4+ T-cell enrichment cocktail (Stemcell Technologies, #15062), also followed by centrifugation using a Ficoll gradient, as well as 2 RPMI washes. Macrophages were grown from a subset of the isolated PBMCs by allowing the PBMCs to adhere to plastic plates and treating them with 100 ng/ml of granulocyte/macrophage-colony stimulating factor (GM-CSF) for 96 hours prior to infection. All primary cells were maintained in RPMI supplemented with 10% FCS, L-glutamine, and penicillin/streptomycin. Activation of PBMCs and CD4+ T cells was achieved by the addition of phytohemagglutinin (PHA) at 5 µg/ml for 72 hours prior to infection, with addition of 25 U/ml of interleukin-2 (IL-2) at the time of infection.

Viral replication and infectivity assays

Relatively similar protocols were used to evaluate genetic robustness in CA and IN through replication and infectivity assays. Therefore, unless specifically ascribed to one protein, the referenced assays were used on both IN and CA mutants.

In order to generate viral particles, 293T cells were transfected with the indicated proviral plasmids and given fresh medium 16 hours later. At ~ 48 hours post-transfection, cell supernatants were collected and filtered (0.22 µm). For the single cycle assay, infectivity was measured using MT-4 cells that were seeded

in 96 well plates at 4 x 10^4 cells per well and inoculated with a volume of filtered supernatant that was equivalent to an MOI of ~1 for the WT viral clone. At 16 hours post-infection dextran sulfate was added at 100 μM in order to limit replication to a single cycle, and cells were fixed in 4% paraformaldehyde (PFA) 48 hours after infection. Additionally, for temperature-sensitive mutants, the same single cycle infectivity assay was used, but viral production (and sometimes infection) occurred at 35, 37, or 39.5°C, as indicated in figure legends.

For spreading replication assays, MT-4 cells seeded in 96 well plates at 2 x 10^4 cells per well were inoculated with a much smaller volume of supernatant, equivalent to an MOI of ~0.01 for the WT viral clone, and fixed in 4% PFA at around 80 hours post-infection. FACS analysis for all infectivity and replication assays was carried out using a Guava EasyCyte instrument, and the percent of infected cells was determined by the number that were GFP+.

Some of the temperature-sensitive CA mutations (used in Figures 16, 17, and 18) that were originally detected in viability screens (as part of analysis of Figure 11) were double or triple mutants, and are therefore not listed amongst CA mutants evaluated for fitness in Tables 1 and 2. To isolate single residue mutations responsible for the temperature-sensitive phenotype, CA mutants encoding each of the following individual substitutions were generated from double or triple mutants: S4C, L189M, I91F, Q6P, A78V, I15M, A78V, I15M, A78T, S16T, T48A, S33C, A92V, I91N, D163E, I91V, I124T, M96T, R100S,

Q112L, R132G, R167Q, M214L, AND K227I. All of these CA mutants, along with other CA mutants identified in temperature-sensitive and viability screens, were evaluated for the temperature-sensitive phenotype, and those 11 mutants with the greatest effect were included in Figures 16-18.

For the CA mutant screen examining Mx2 sensitivity, purified plasmid DNA was used from most CA mutants considered viable (>2% WT fitness) in the single cycle infectivity assay, along with NHGcapNM. These mutants included: I2L, N5D, I6T, M10I, M10L, Q13H, A14S, I15V, N21S, S33C, Q50H, N57S, M68L, L69I, E75D, V83M, H87Q, H87R, P90T, I91T, I91V, Q95L, M96I, M96T, E98D, R100S, N121I, R132G, S146C, T148I, S149C, S149G, I150V, L151I, I153T, Y164F, R167Q, A177S, M185I, E187V, L190M, T200S, A204G, G208E, A209V, L211I and T216A, and were used to transfect 293T cells in 24 well plates as previously described. Supernatants were harvested 48 hours later, filtered, and titrated onto doxycycline treated MT-4 cells encoding HsMx2, MacMx2, AGMMx2, OvMx2 or CanMx2, in addition to MT-4 HsMx2 cells that had not been doxycycline treated (as control). At 16 hours post-infection, infected cultures were treated with dextran sulfate, so as to limit infection to a single cycle. Finally, at 48 hours post-infection, all cells were fixed and subjected to flow cytometry, such that infectious titers could be calculated by counting GFP+ cells.

Infections of primary cells necessitated different conditions. For the single cycle infections of CD4+ T cells, 1 x 10^6 cells were inoculated using virus generated from 293T cells, as described above, for the 11 rarely occurring ‘fit’ CA

mutants described in Figure 22C, the WT virus, and 4 randomly selected frequently occurring 'fit' CA mutants from Figure 22B (H87Q, I91V, E98D, T200S) at an MOI of ~5, and were fixed in 4% PFA 36 hours later and subjected to FACS analysis. For infection of macrophages, vesicular stomatitis virus (VSV)-pseudotyped virus was generated in 293T cells for the same 15 mutants plus WT virus, under transfection conditions in 24 well plates the same as previously described, but with the addition of 50 ng of VSV-G. Macrophages were infected at an MOI of ~4 and fixed for FACS analysis 72 hours later. For spreading infection assays in PBMC, 1 x 10^6 cells were infected at an MOI of 0.1, and cells were fixed at the time points shown in Figure 23C and then analyzed by flow cytometry. The MOIs used for primary cells were reference values derived from titrations on MT-4 cells.

Western blotting

Both virions, which were initially pelleted through 20% sucrose by centrifugation (for 2 hours at 14000 rpm at 4°C), and cell lysates were resuspended in SDS sample buffer and separated by electrophoresis on NuPage 4-12% Bis-Tris Gels (Novex). Proteins were then blotted onto nitrocellulose membranes. For the CA mutant experiments, blots were probed with either a primary anti-HIV p24 capsid antibody (183-H12-5C) or a primary anti-HIV MA p17 antibody (VU47 Rabbit anti-p17 [210]) and then probed with goat anti-mouse/anti-rabbit IRDye® 800CW secondary antibody (LI-COR Biosciences). For the IN mutant characterizations, blots were probed with either primary anti-

HIV p24 capsid antibody (183-H12-5C), primary anti-HIV INT-4 antibody ([211], gift of Malim lab), or primary anti-clathrin heavy chain antibody (BD Biosciences, #610499), whilst secondary probing was done using a goat anti-mouse antibody (Thermo Scientific). A LI-COR Odyssey scanner was used to detect and quantify fluorescent signals for all Western blots. A minimum of 3 separate Western blots was produced for each temperature-sensitive mutant, including those in the protease negative background, and representative blots are shown.

Structure analysis

Structural analysis of the HIV-1 capsid hexamer was done using MacPyMOL, with RCSB Protein Data Bank (PDB) reference 3GV2. Examinations using the HIV-1 intasome model were done with structure coordinates provided by the authors of [168], also using MacPyMOL.

Analyses of solvent accessible surface area values for both the CA hexamer and the IN intasome model were done using the UCSF Chimera modeling system.

Thin-section microscopy

In order to prepare samples for thin-section electron microscopy, 293T cells were seeded in 6-well plates at 0.8 x 10^6 cells per well, in duplicate. Transfections were done the following day with the addition of polyethylenimine to 2 µg of either the WT or mutant NHGcapNM plasmids (S33C, T48A, M96T, Q112L, R132G, L189M, G60W, K25I, L52F or N195S) plus 0.5 µg of modified human tetherin previously used in our laboratory (delGI, T45I that is resistant to

antagonism by HIV-1 Vpu [212]) to enhance visualization of virions at the plasma membrane. Immediately after addition of the transfection mixture, plates were kept at either 37 or 39.5°C, as indicated. After 48 hours, the supernatant was removed and the cells were fixed in a solution of 2% PFA, 2% glutaraldehyde. One set of cells was then analyzed by FACS to compare the transfection efficiencies (and these ranged from 23% for WT to 32% for L189M). The second set of cells were fixed with 2.5% glutaraldehyde and 1% osmium tetroxide, and stained with 2% aqueous uranyl acetate. The fixed and stained cells were then harvested into PBS and pelleted through 1% SeaPlaque agarose (Flowgen) at 45°C. The agar was then set at 4°C and the cell pellets were cut into ~2 mm cubes, which were dehydrated through a graded alcohol series and infiltrated with TAAB 812 embedding resin. Following polymerization, thin sections (120 nm) were cut and examined in a JEOL 1200 EX II electron microscope. For each sample, numbers of virus particles associated with 150 randomly selected cells were counted. Thin-section EM analyses were done with the help of Dr. Frazer Rixon.

Cryo-electron microscopy

In order to isolate virions for cryo-electron microscopy, 10 cm plates of approximately 4 x 10^6 293T cells were transfected with 7 μg of the indicated plasmids (NHGcapNM, R18G, K30N, G60W or M215V) using polyethylenimine. Approximately 48 hours later, supernatant was collected and filtered, and virions were pelleted by centrifugation at 14000 rpm through 20% sucrose. Pelleted

virions from either WT or the CA mutants were then fixed in 10 µl of 2% PFA, 2.5% glutaraldehyde solution.

Subsequently, fixed aliquots of 3 µl of each sample were loaded onto freshly glow-discharged c-flat holey carbon grids (CF-22-4C, Protochips, Inc.) held at 4°C and 100% humidity in a Vitrobot vitrification robot (FEI). The grids were blotted for 4 seconds prior to being frozen by plunging into a bath of liquid nitrogen-cooled liquid ethane. Vitrified specimens were then imaged at a low temperature in a JEOL 2200 FS cryo-microscope equipped with Gatan 626 cryo-stages. Finally, low dose (10 e/Å2), energy-filtered images (slit width, 20 eV) were recorded on a Gatan ultrascan 16-megapixel charge-coupled-device camera at a magnification of 50,000x. Cryo-EM analyses were done with the help of Dr. Saskia Bakker.

Analysis of natural CA and IN variants

A cohort of 1,000 HIV-1 subtype B CA sequence isolates was obtained from the Los Alamos HIV sequence database (www.hiv.lanl.gov). All sequences were sampled from distinct infections between 1980 and 2009. In order to minimize risks of sampling biases, multiple sequences from known transmission clusters were excluded. Sequences with frameshift mutations or stop codons that were likely to represent nonfunctional viruses or poor quality sequencing were also excluded. Sequences were aligned using MUSCLE [213], and PERL scripts were used to survey the genetic variation in the resulting sequence alignment. An information-theoretic measure of diversity (Shannon's entropy)

[214] was applied to quantify the amount of amino acid variation at each capsid position. Analysis of the natural variation in CA was done with the help of Dr. Rob Gifford.

Similar to the cohort of sequences analyzed for CA, all HIV-1 subtype B IN sequences from 1980 to 2013 were selected from the Los Alamos HIV sequence database. Sequences with frameshifts, stop codons, or nucleotide ambiguities (N or IUPAC code) were excluded from the dataset prior to randomly selecting 1000 sequences from distinct individuals and locations. As before, IN sequences were aligned using MUSCLE, with the use of PERL scripts to determine the number of mutations at each site in the alignment. Shannon's entropy was used to measure the amino acid diversity at each position within integrase. The frequency of each amino acid was determined for each variable site. Analysis of HIV-1 IN variants was done with the help of Dr. Joseph Hughes.

The hydrophobicity and molecular weight changes were also calculated for all CA and IN library single mutants.

III. RESULTS

Chapter 1. HIV-1 CA and creation of a randomly mutagenized CA library

1.1 Introduction to surveying the robustness of CA through random mutagenesis

HIV-1 CA is under acute competing pressures to both conserve sequence to preserve function, and to diversify sequence to evade immune responses. To evaluate how such pressures ultimately impact genetic robustness or any potential trade-offs with fitness in an environment with a high mutation rate, we utilized an unbiased mutagenesis method to generate a large library of randomly mutagenized CA sequences that could simulate the natural process of random mutation that occurs during HIV-1 replication. Analyses of the CA mutant library uncovered a remarkable degree of genetic fragility in HIV-1 CA, and it is this, along with its basis, *in vivo* correlates, and functional consequences, that will be subsequently analyzed.

1.2 Creation of the CA randomly mutagenized library

Construction of the library of random HIV-1 CA mutants was completed by a former postdoc in the Bieniasz lab, Sam Wilson, but was subsequently passed on for the work described within this thesis. However, as the nature of its construction is highly relevant to the subsequent experiments and analyses herein, the generation will be briefly described.

To begin construction, a low fidelity PCR was approach was used, as depicted in Figure 10A. In this method, sequences encoding capsid were

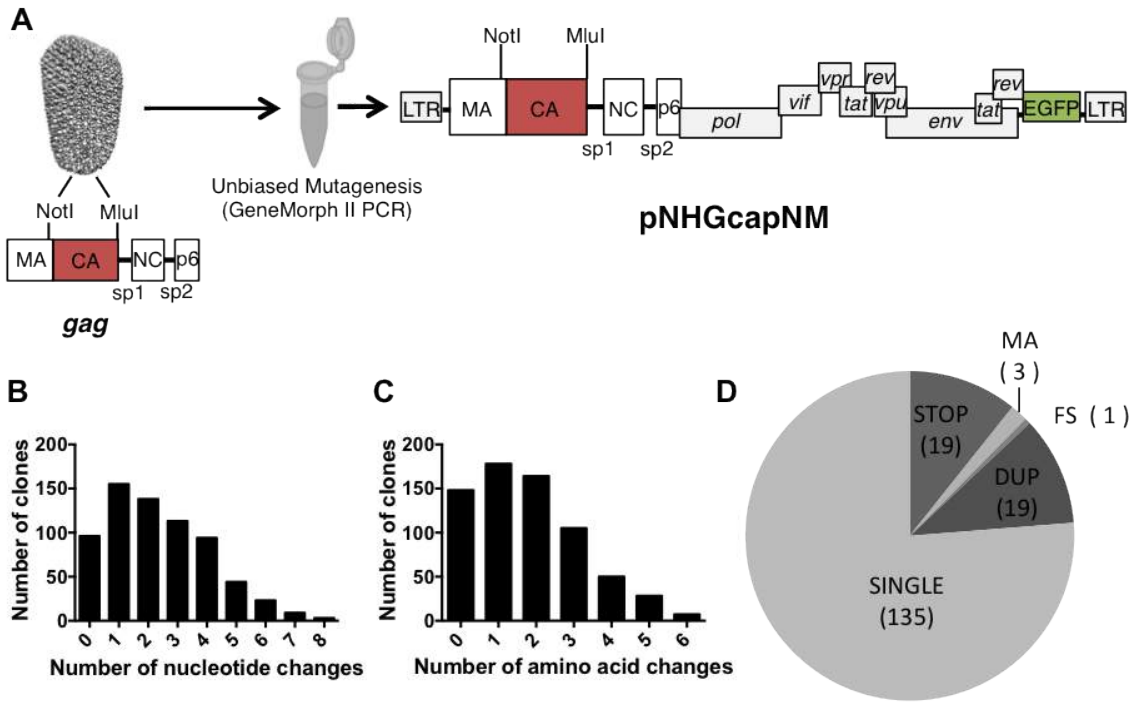


Figure 10. Generation and composition of the CA mutant library. (A) Schematic illustrating the replication competent proviral plasmid pNHGcapNM used for this study. The 717 bp amplicon in Gag flanked by *Not*I and *Mlu*I restriction sites includes nearly all of CA, and a few amino acids in MA. This amplicon was subjected to Genemorph II mutagenesis and subsequently subcloned into pNHGcapNM. The conical capsid image was taken from PDB ID: 1VUT.
(B) Distribution of the number of nucleotide changes found in each library clone (including sequences that occurred more than once). (C) Distribution of the number of amino acid changes in each library clone (including sequences that occurred more than once, mutations in MA, nonsense mutations, and frameshifts). (D) Pie chart depicting all resulting single amino acid changes. SINGLE clones were those library mutants included in further experiments. The remaining clones were discarded for the following reasons: STOP (mutation resulted in a stop codon), MA (mutation occurred in the small fragment of MA included in the amplicon), FS (frameshift), DUP (duplicate mutation of a clone already in the library). Creation of the library was completed by Sam Wilson.

amplified using Genemorph II, an unbiased error-prone PCR system, and cloned using a TOPO TA cloning kit to yield a library with an estimated complexity of 1 x 10^4 clones. Subsequently, plasmid DNA was extracted from the pooled library and digested, so that the mutagenized CA sequences were then inserted into the replication competent proviral clone, pNHGcapNM, using *Not*I and *Mlu*I restriction sites. As noted previously, pNHGcapNM (accession: JQ686832) encodes EGFP in place of Nef, and will hereafter be referred to as the WT or parental virus for all capsid mutagenesis work.

1.3 Composition of the CA library

Approximately 1000 individual colonies were then picked and proviral plasmid DNA was extracted from individual mini-cultures. Each mutagenized CA amplicon was then sequenced for every clone, and after the removal of clones with failed sequencing reactions or whose chromatograms indicated the presence of more than one template, 680 arrayed library proviral plasmids remained. Evaluations of the library composition then revealed the distribution of all nucleotide changes (Figure 10B), missense mutations (Figure 10C), and nonsense or other mutations (Figure 10D). During the initial creation of the library, conditions in pilot experiments were optimized to reduce the frequency of both non-mutated and heavily mutated sequences in the library. As a result, the most common mutation in the library was a single nucleotide or amino acid substitution (Figure 10B, C, D).

Chapter 2. Evaluations of CA library mutant fitness

2.1 Fitness of all library CA mutants

To determine the genetic robustness of CA, we examined the ability of the 680 CA library mutants to replicate. Replicative fitness was analyzed using a spreading replication assay in which MT-4 cells were challenged with a small volume of virus-containing supernatant harvested from WT or CA mutant transfected 293T cells (equivalent to an MOI of ~0.01 for the WT virus). After approximately 80 hours, and multiple rounds of the virus replication cycle, the experiment was concluded once the WT virus had infected more than 50% of cells. Each CA mutant's replicative fitness was then expressed as the number of cells that were infected (EGFP-positive), as a percentage of the number of cells infected by the WT virus (Figure 11A, B). These viability screens were conducted at two temperatures, 35°C (Figure 11A) and 39.5°C (Figure 11B) in order to permit isolation of temperature-sensitive CA mutants that could facilitate further characterization of the replication defects associated with CA mutations (see below).

Not surprisingly, nearly all of the viral clones with WT CA sequence, or containing silent mutations, replicated similarly to that of the WT pNHGcapNM construct. This suggested that there was a low incidence with which defective viruses were generated by cloning artifacts or silent mutations (Figure 11A, B). More importantly, these analyses of CA library replicative fitness demonstrated capsid to be extremely fragile, or intolerant to mutation. If the cut-off for being

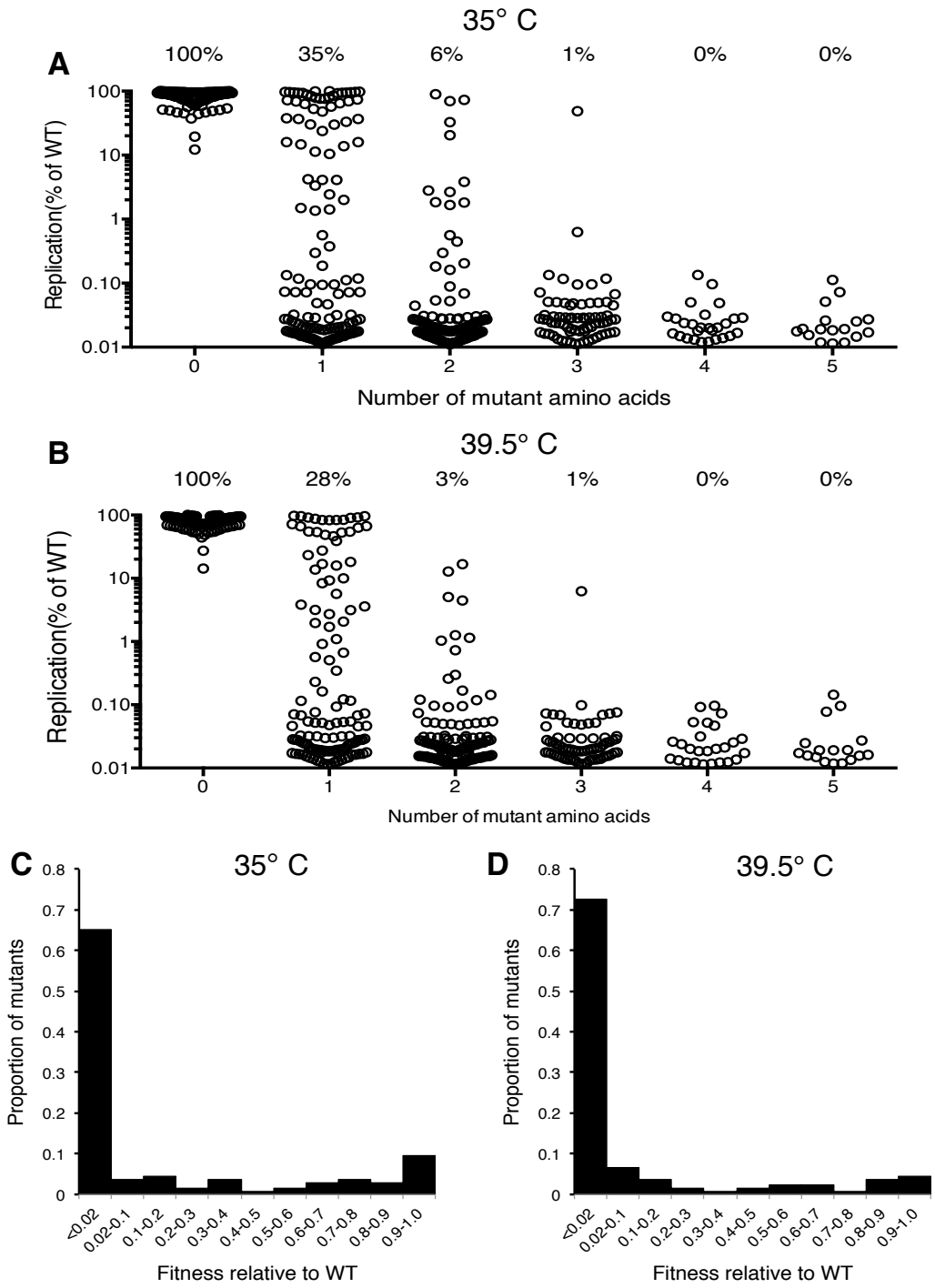


Figure 11. Fitness effects of all CA library mutants. (A) Viability of all CA mutant clones at 35°C (excluding nonsense, frameshift, duplicate and MA mutations), indicating how the number of amino acid substitutions in CA affects fitness. Replication shown is the fraction of infected MT-4 cells (GFP+), as a percentage of the fraction infected by WT virus after a spreading replication assay. (B) Viability of all CA mutant clones at 39.5°C, as in (A). (C) Distribution of mutational fitness effects (DMFE) for all single amino acid CA substitution mutants at 35°C. (D) DMFE for all of the single amino acid substitution mutants at 39.5°C. Data shown in (A) and (B) was measured by Sam Wilson.

viable is arbitrarily set at 2% of WT virus replicative fitness, only 35% of the 135 clones with single amino acid substitutions were viable at 35°C, and even fewer, 28%, were viable at 39.5°C. Given that the requirements for replication in an infected individual may be more stringent than in highly permissive MT-4 cells, these figures may even overestimate the proportion of mutants that are viable *in vivo*. To this end, 2% of WT virus fitness is a rather generous cut-off to earn the designation of 'viable'. Increases in the number of amino acid substitutions yielded even further reductions in viability; just 6% of the 125 clones with two random amino acid substitutions were viable at 35°C, and this figure was only 3% at 39.5°C. Accordingly, just 1 of 73 clones with 3 random amino acid substitutions was viable, and none of the clones with either four or five random substitutions could replicate at all.

It should be noted that the work involved in screening the entire CA library for viability at 35°C and 39.5°C was done by Sam Wilson. However, his involvement in the work ended at this point, and all further experiments described herein were done independently, with the exception of some bioinformatics and electron microscopic analyses requiring specialist assistance.

2.2 Fitness of single residue CA mutants

To better characterize the contributions of discrete residues to the genetic fragility of HIV-1 CA, we next focused our attention on the subset of mutants with unique single amino acid substitutions. These single mutants covered 102 (44%) of the 231 residues in CA, and there were 33 CA residues for which there was

more than one unique mutation. To evaluate robustness in more natural conditions, two additional infectivity assays were performed at 37°C, utilizing only the single residue mutants. In the first MT-4 spreading replication assay, just 40 mutant viruses (30%) exhibited at least 2% of parental virus fitness. These viable mutants, and their corresponding fitness values can be found in Table 1, while the longer list of the 95 mutants (70%) with less than 2% of WT fitness in the spreading infectivity assay can be found in Table 2. In the second assay, single cycle infectivity was measured by transfecting 293T cells with WT or CA mutant proviral plasmids and measuring infectivity using a larger dose of virus (equivalent to an MOI ~ 1 for the WT clone), and treating the MT-4 cells with dextran sulfate 16 h post-infection, in order to limit replication to a single cycle. Fitness in the replicative (spreading) versus the single cycle infectivity assay was generally similar (Table 1), although most mutant viruses had slightly higher fitness in the single cycle assay than in the spreading assay. This phenomenon may be a product of the effect of transfection in the single cycle assay, in which overexpression of viral proteins could mask any defects, as opposed to the spreading infection assay in which lower gene expression levels may better approximate those found in the natural environment. Alternatively, the multiple rounds of replication permitted in the spreading assay may enhance the effect of any defects in the single cycle assay.

Interestingly, the majority of the viable CA mutants still displayed attenuated infectivity, as a remarkable 89% of single amino acid substitutions

Table 1. Fitness measurements for viable HIV-1 CA mutants

Viable HIV CA mutant	Location in CA	Single cycle infectivity (% of WT)[a]	Spreading infectivity (% of WT)[b]
I2L	β-strand	90	65.6
N5D	β-strand	20	5.2
I6T	β-strand	52	56.9
M10I	β-strand	19	2.9
M10L	β-strand	73.7	27.0
M10V	β-strand	10.6	16.2
Q13H	β-strand	27.4	5.8
A14S	Loop	77.1	80.7
A14T	Loop	97.6	86.6
I15V	Loop	32.3	13.6
N21S	Helix 1	55.9	46.6
S33C	Loop	40.2	47.8
Q50H	Helix 3	19.6	3.1
E75D	Helix 4	40.8	22.1
V83M	Helix 4	75.3	64.8
H87Q	Cyclophilin b.l.*	120.3	92.0
H87R	Cyclophilin b.l.	27.5	2.5
I91T	Cyclophilin b.l.	85.2	76.9
I91V	Cyclophilin b.l.	100.0	82.8
M96I	Loop	120.3	92.0
E98D	Loop	101.0	63.4
R100S	Loop	53.9	42.5
S146C	Loop	43.7	34.0
T148I	Loop	65.8	37.3
S149C	Loop	64.8	73.7
S149G	Loop	70.9	30.3
I150V	Loop	18.6	3.1
I153T	Loop	21.2	4.7
Y164F	Helix 8	64.4	30.6
R167Q	Helix 8	11.1	5.9
A177S	Loop	99.4	37.3
M185I	Helix 9	24.2	2.5
E187V	Helix 9	59.7	47.4
L190M	Loop	38.0	32.6
T200S	Helix 10	104.2	90.4
A204G	Helix 10	77.4	89.0
G208E	Loop	15.1	3.8
A209T	Loop	108.3	94.3
A209V	Loop	72.4	84.3
T216A	Helix 11	72.4	37.5

[a] Infectivity assay in which MT-4 cells were inoculated with supernatant from 293T cells (equivalent to MOI = 1 for WT virus) that had been transfected with single residue CA mutant proviral plasmids. Dextran sulfate was added 16h after infection to limit replication to a single cycle. Values are reported as the percentage of the number of infected cells (GFP+) obtained with the WT virus.

[b] Fitness measurement in which MT-4 cells were inoculated with supernatant from 293T cells (equivalent to MOI = 0.01 for WT virus) that had been transfected with single residue CA mutant proviral plasmids. Multiple cycles of replication were allowed over a 72-80 hr time period. Values are reported as the percentage of the number of infected cells (GFP+) obtained with the WT virus.

* Abbreviation for cyclophilin binding loop.

Mutants were considered viable or infectious if they maintained 1/50 (or 2%) of WT infectivity in the spreading infectivity assay.

Table 2. Nonviable HIV-1 CA mutants

Nonviable CA mutant	Location in CA	Nonviable CA mutant	Location in CA
I2T	β-strand	E113V	Helix 6
H12R	β-strand	I115K	Helix 6
Q13R	β-strand	I115L	Helix 6
S16P	Loop	I115T	Helix 6
R18G	Helix 1	M118L	Helix 6
N21Y	Helix 1	M118V	Helix 6
K25I	Helix 1	H120P	Loop
V26E	Helix 1	N121I	Loop
K30E	Loop	I124N	Loop
K30N	Loop	V126A	Helix 7
F32L	Loop	I129M	Helix 7
F32S	Loop	I129T	Helix 7
S33I	Loop	Y130C	Helix 7
P34S	Helix 2	Y130H	Helix 7
V36A	Helix 2	R132G	Helix 7
M39T	Helix 2	G137E	Helix 7
M39V	Helix 2	L138I	Helix 7
F40L	Helix 2	M144V	Helix 7
F40Y	Helix 2	L151I	Loop
L43I	Helix 2	L151Q	Loop
G46R	Loop	I153M	Loop
P49L	Helix 3	E159V	Loop
L52F	Helix 3	F161C	Helix 8
N57S	Helix 3	D163G	Helix 8
V59A	Loop	V165A	Helix 8
V59M	Loop	F168S	Helix 8
G60W	Loop	R173I	Helix 8
Q63R	Helix 4	A174G	Loop
A65V	Helix 4	K182Q	Helix 9
M66I	Helix 4	W184L	Helix 9
M66V	Helix 4	W184R	Helix 9
M68L	Helix 4	T186A	Helix 9
M68T	Helix 4	L190S	Loop
L69I	Helix 4	A194V	Loop
I73N	Helix 4	N195S	Loop
A77D	Helix 4	I201N	Helix 10
W80R	Helix 4	L211I	Helix 11
P90T	Cyclophilin b.l.	E212D	Helix 11
Q95L	Loop	M214L	Helix 11
Q95R	Loop	M215V	Helix 11
M96T	Loop	A217V	Helix 11
P99A	Loop	Q219P	Loop
P99L	Loop	G220V	Loop
R100W	Loop	V221A	Loop
S102R	Helix 5	K227I	Loop
A105G	Helix 5	K227N	Loop
T110I	Loop	K227R	Loop
L111F	Helix 6		

Mutants were considered nonviable if they exhibited <2% of WT fitness in the spreading fitness assay described in Table 1.

caused a greater than 2-fold reduction in infectious virion yield. This observation, in conjunction with the lack of mutants with replicative fitness greater than WT, indicates that it may be difficult to improve upon the fitness of the parental CA, which may already be near a fitness peak. Another way of analyzing the overall robustness of the single residue mutants is to examine the distributions of mutational fitness effects (DMFE). Previous work analyzing panels of RNA virus single amino acid mutants have presented DMFE values that were bimodal in nature, with substitutions typically resulting in either minor reductions in fitness or lethality [86,87,93,94,95,96]. While this phenomenon was also true to some extent for HIV-1 CA (Figure 11C at 35°C, Figure 11D at 39.5°C, Figure 12 at 37°C), the proportion of nonviable CA mutants was significantly larger than has previously been reported [86,87,93,94,95,96], so that the proportion of viable CA mutants was also significantly smaller than earlier reports, and the occurrence of mutants with only minor fitness defects was also greatly reduced in the case of HIV-1 CA (Figure 11C, D, Figure 12).

Assessments of Tables 1 and 2 reveal that different mutations at the same residue in CA, such as I2, can sometimes result in highly variable fitness effects, suggesting that the type of residue change is not entirely insignificant. However, analyses of the mutant residues' changes in hydrophobicity or molecular weight for the entirety of the single residue mutant CA library did not reveal any general associations with replicative fitness (Figure 13A, B). While some polar residues appeared to be more sensitive to change, the lack of any overall or consistent

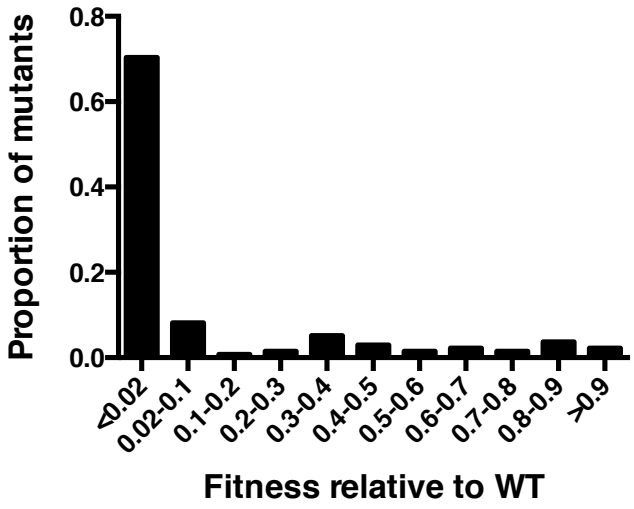


Figure 12. Fitness effects of single residue mutations in the CA mutant library. Distribution of mutational fitness effects (DMFE) for all single amino acid CA substitution mutants, using fitness values obtained from the spreading infectivity assay at 37°C, as described in Table 1.

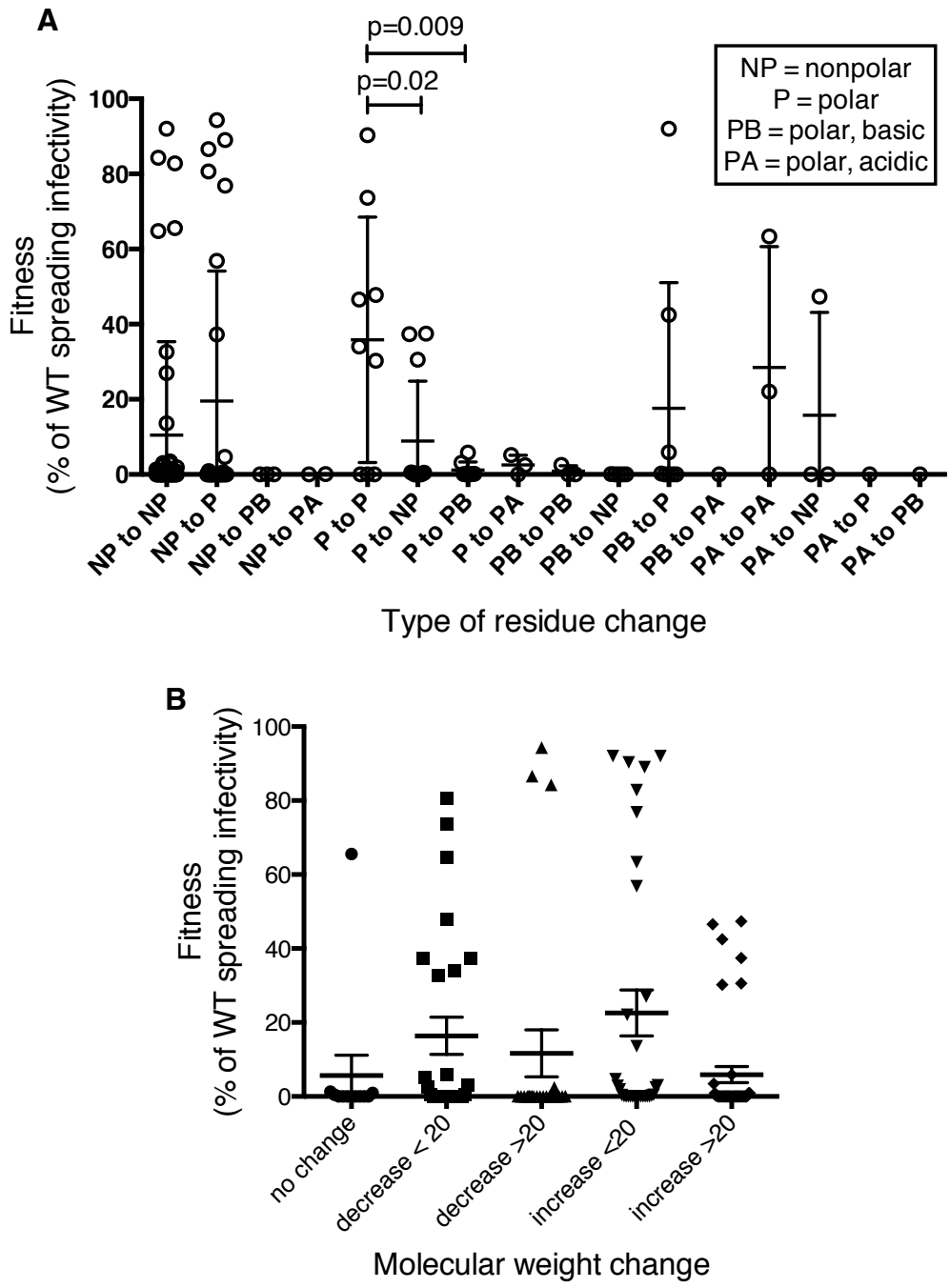


Figure 13. Changes in hydrophobicity or molecular weight are not primary determinants of fitness. (A) Evaluations of the effect of the type of residue change for all library mutations on fitness (in a spreading replication assay). Letters indicate changes from the initial type of residue to the mutated residue type, as indicated in the legend. Analyses of disparities between conservative changes (i.e., polar (P) to polar (P)) and nonconservative changes (i.e., polar to nonpolar (NP)) were performed for all scenarios, and all significant differences are indicated. (B) Evaluations of the effect of the size of mutant residues' molecular weight change on fitness (in a spreading replication assay). Analyses between residue substitutions in which there was no change in molecular weight and those with either increases or decreases in molecular weight indicated no significant differences (significance determined using unpaired, parametric t-tests).

patterns related to hydrophobicity changes suggests that types of residues changes are not significant determinants of CA fitness. Moreover, inspection of Tables 1 and 2 reveals that the position within CA was a far greater determinant of a mutation's fitness effect. Of the 231 residues in CA, 56.3% lie within CA's 11 helices, 5.6% are in the N-terminal β-strand, 3.9% are in the cyclophilin A binding loop, and the remaining 34.2% are in interhelical loops. Together, these figures provide a context in which it is possible to summarize mutant fitness by region, in order to indicate areas of particular genetic fragility in CA (Table 3, Figure 14). Notably, there were distinct regions of relative robustness within CA, as 70-100% of mutants retained some viability within the N-terminal β-strand, the cyclophilin binding loop, and the interdomain linker region. In contrast, just 16% of mutants in helices maintained viability, and an intermediate 39% of mutants retained viability within the interhelical loops. More strikingly, all 25 mutations in helices 2, 5, 6, and 7 resulted in nonviability, suggesting these N-terminal helices are particularly fragile. When these results are graphically represented such that the fitness of each mutant is plotted against its position in the linear CA sequence (Figure 14), it is even more apparent that regions of mutational fragility are non-uniformly distributed in CA, with mutations in NTD helices the most likely to result in inactivation (Figure 14).

2.3 Locations of library mutations on the CA hexamer structure

When CA mutant fitness is displayed within the context of a capsid hexamer structure (PDB: 3GV2, Figure 15A, B), amino acids corresponding to

Table 3. Summary of CA mutant viability by CA region

Region	Number of viable mutants	Number of nonviable mutants	% viable mutants
β-strand	7	3	70
Helix 1	1	4	20
Helix 2	0	7	0
Helix 3	1	3	25
Helix 4	2	10	17
Cyclophilin b.l.	4	1	80
Helix 5	0	2	0
Helix 6	0	7	0
Helix 7	0	9	0
Helix 8	2	5	29
Helix 9	2	4	33
Helix 10	2	1	67
Helix 11	1	5	17
MHR	3	6	33
Interdomain linker	5	0	100
Helices (all)	11	57	16
Loops (all)	22	35	39
NTD	27	68	28
CTD	13	27	33

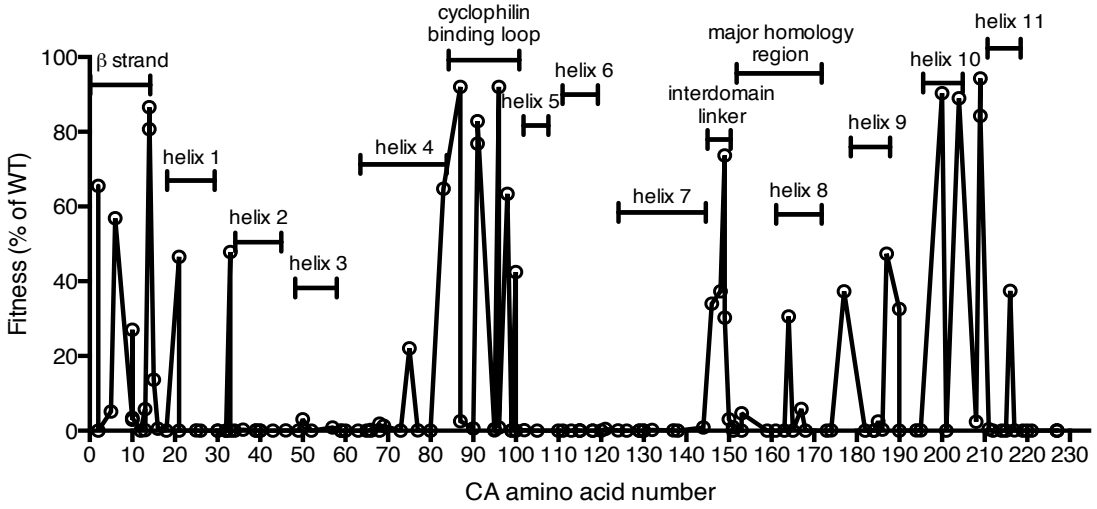


Figure 14. Linear distribution of fitness effects of single residue CA mutations. Plot of fitness values for CA mutants in which the location of the mutation is arranged on the X-axis from left (N-terminal residue) to right (C-terminal residue) with their corresponding fitness values (as the % of WT in the spreading replication assay, as described in Table 1) plotted on the Y-axis.

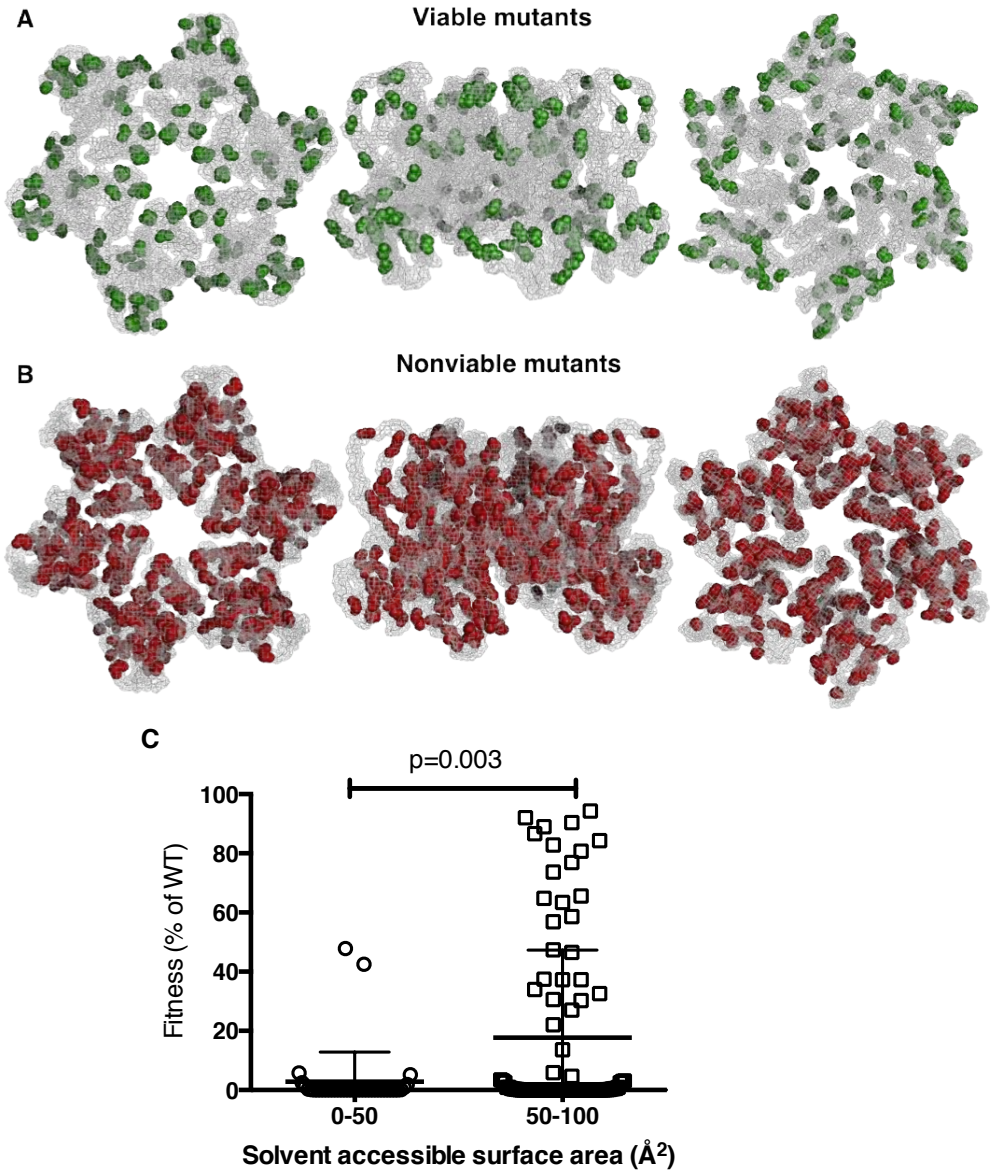


Figure 15. Mapping of CA fitness to the CA hexamer structure. (A, B) Location of single amino acid mutant residues on an X-ray crystal structure of the CA hexamer (PDB: 3GV2). Images on the left show the hexamer as viewed from the exterior of the intact capsid, while center images show the hexamer as viewed from within the plane of the capsid lattice, and rightmost images show the hexamer as viewed from the interior of the intact conical capsid. Green residues, shown in (A), are those that had greater than 2% of WT fitness in a spreading replication assay when mutated (viable mutants). Red residues, shown in (B), are those that had less than 2% of WT fitness in a spreading replication assay when mutated (nonviable mutants). For residues with more than one mutation, the most fragile result was plotted. (C) Comparison of fitness and residue surface exposure, as determined by the solvent accessible surface area (calculated using UCSF Chimera with PDB 3GV2). Mutated residues were grouped as having either a solvent accessible surface area $<50Å^2$ or a solvent accessible surface area $>50Å^2$. Fitness is plotted in accordance with the grouped residue. Significance was determined using unpaired, parametric t-tests.

viable mutants (in green, Figure 15A) appear more likely to occur at surface exposed residues. Conversely, the mutations shown in the interior regions of the CA structure almost always resulted in nonviable mutants (displayed in red, Figure 15B). This observation is particularly evident when the CA hexamer is viewed from a point of view from the interior of the assembled capsid (Figure 15B, rightmost images), and further emphasizes the impression of an inner CA 'core' structure, composed of helices that are particularly intolerant to mutation.

Such findings are further confirmed quantitatively by analysis of the solvent accessible surface area values for individual mutated residues (Figure 15C). Specifically, the mutated residues with solvent accessible surface area values less than 50 $Å^2$ (a suitable cutoff for surface exposure based on previous mutagenesis studies [215]) had significantly larger fitness defects ($p=0.003$). These more buried mutant viruses (with solvent accessible surface areas less than 50 $Å^2$) displayed a mean fitness value of only 2.8% of WT (Figure 15C). However, viruses harboring mutations in more surface exposed residue locations (with solvent accessible surface areas greater than 50 $Å^2$) displayed a much larger mean fitness value, at 18% of WT (Figure 15C).

Chapter 3. Biological basis for the genetic fragility of HIV-1 CA

3.1 Analysis of conditionally nonviable (temperature-sensitive) CA mutants

As has been already mentioned, CA performs numerous critical functions during HIV-1 replication, including its roles mediating the assembly of immature

virions (in the context of the Gag precursor) and the formation of a mature conical capsid. In addition, the optimal stability of capsid is thought to be important to the uncoating step of the HIV-1 cycle in newly infected cells, and CA is also critically involved in the import of the viral genome into an infected cell's nucleus. As it was unclear which, if any, of these functions was responsible for the extreme genetic fragility of CA, a number of assays were performed to elucidate the nature of replication defects in both constitutively and conditionally nonviable CA mutants.

During analysis of the initial viability screens performed on the entire CA library (Figure 11 A, B), we observed that a small subset of CA mutants were viable at a permissive temperature of 35°C, but had severely reduced fitness at a nonpermissive temperature of 39.5°C, whereas most CA mutants were approximately equally fit or unfit at both temperatures. This subset of temperature-sensitive (ts) conditionally nonviable mutants included both single and double residue mutants, and so CA mutants encoding only single amino acid substitutions in isolation were generated. In total, eleven single amino acid CA mutants with large ts replication phenotypes were selected for examination.

Since the initial viability screen in which we identified the presence of ts mutants had measured replicative fitness over multiple rounds of replication, it did not determine which specific step in the viral life cycle was impaired in the constitutively or conditionally nonviable mutants. Thus, it seemed prudent to determine the single cycle infectivity of the ts CA mutants, utilizing virions

generated at 35°C, 37°C, or 39.5°C in 293T cells, which were then used to infect MT-4 cells at 35°C, 37°C, or 39.5°C (Figure 16A, B). As Figure 16A demonstrates, the parental control, pNHGcapNM, has a consistent viral titer regardless of the temperature of production or infection. It is also apparent that most of the ts CA mutants had similar, or modestly reduced, infectivity as compared to the parental virus when virion production and infection were done at the permissive temperature (35°C) (Figure 16A). Strikingly however, when restrictive temperatures were applied during virion production, single cycle replication of all the ts mutants was inhibited at least 50 to 1000-fold (Figure 16A, B). Conversely, when restrictive temperatures were applied only during infection, and not production, the single cycle replication of all of the ts mutants was completely unaffected. These results demonstrate that the defects associated with conditionally nonviable ts CA mutants occurred exclusively and irreversibly, during, or immediately after, particle production, and not during earlier steps of the viral life cycle (such as uncoating or nuclear import).

Subsequent examination of Gag expression and processing by the ts CA mutants and WT virus showed similar levels of cell-associated Gag precursor Pr55 and capsid p24 were present regardless of the temperature of production (Figure 17A). Alternatively, analysis of extracellular virion-associated p24 showed 57% to 88% reductions in p24 levels at the restrictive temperature for the ts CA mutants only (Figure 17A). Since it was possible that the ts CA mutants could have affected either CA stability or recognition by the p24 antibody,

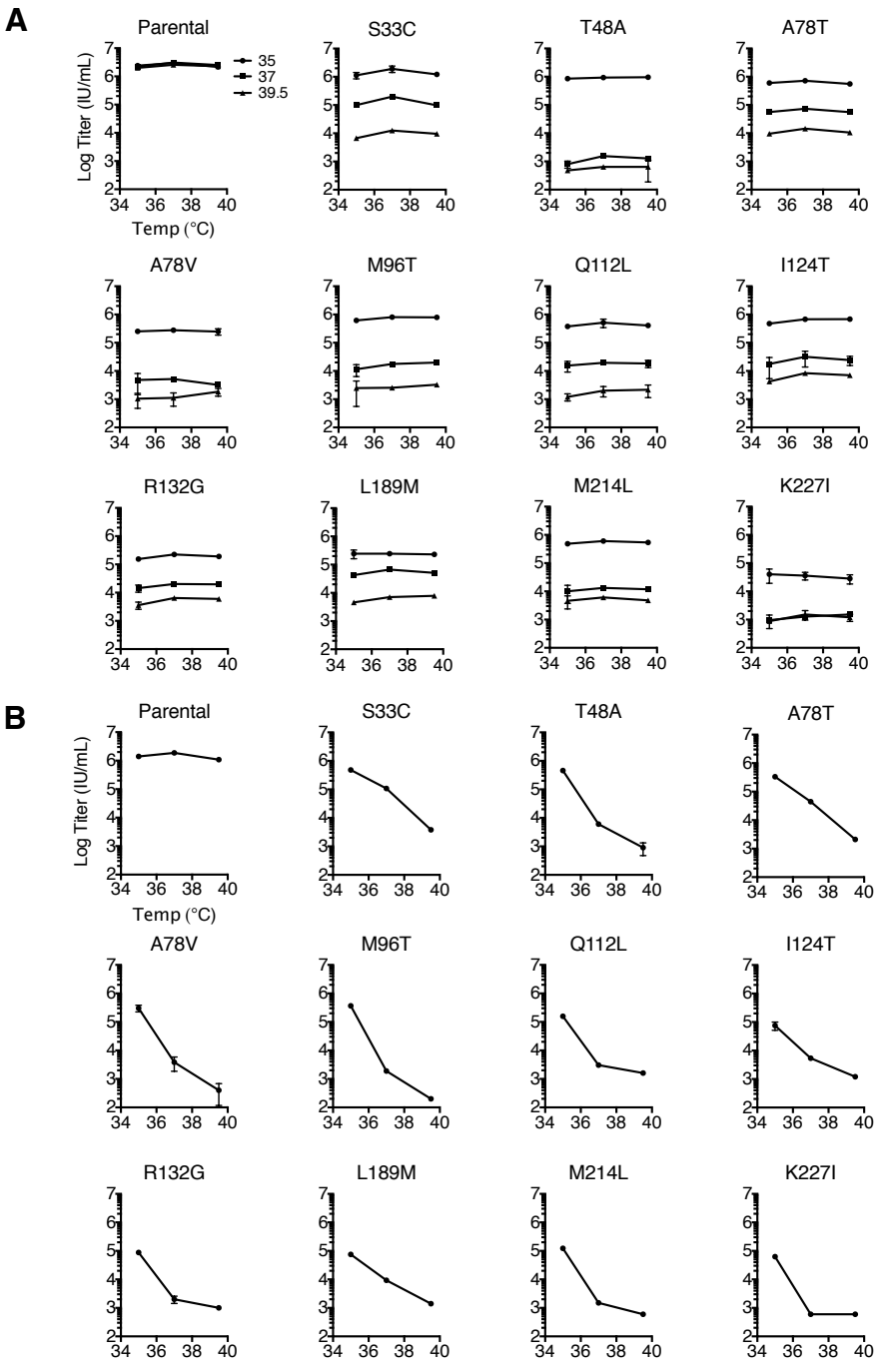


Figure 16. Conditionally viable (ts) CA mutants display defects only when assembled at a nonpermissive temperature. (A) Measurements of infectious virion titers from 293T cells transfected with either the parental or ts single residue mutant proviral plasmids, and placed at either 35°C (filled circles), 37°C (filled squares), or 39.5°C (filled triangles) for production of virions, as the distinct plot lines indicate. Titers of viruses from resulting supernatants from each mutant were measured in MT-4 cells at 35, 37, or 39.5°C, as displayed on the x-axis. Dextran sulfate was added 16h post-infection to restrict replication to a single cycle. (B) As in (A), but here the temperature of virus production is indicated on the x-axis and inoculations were done in MT-4 cells at 37°C only.

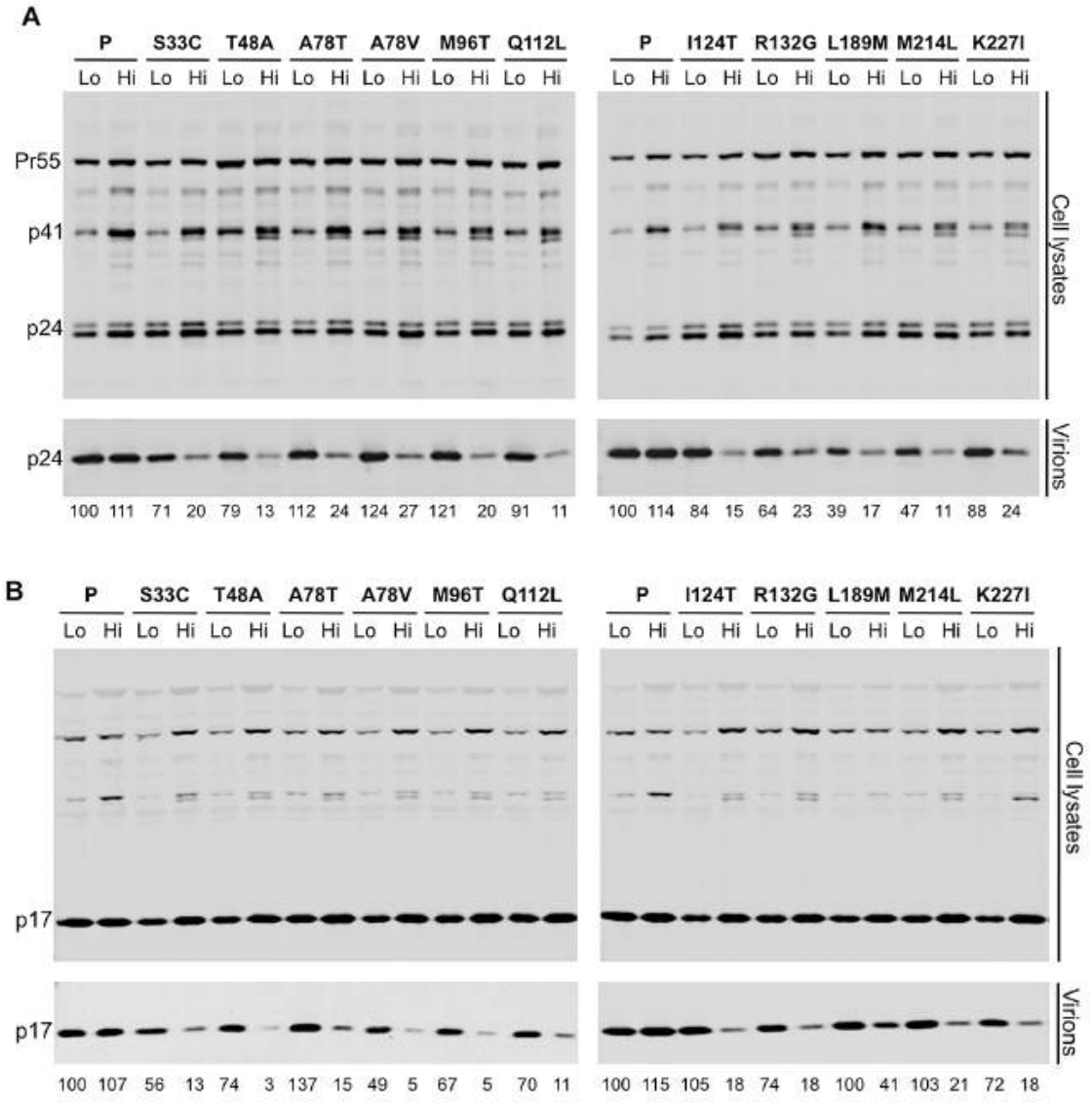


Figure 17. All conditional (ts) CA mutants display reductions in extracellular particle yield at the nonpermissive temperature. (A) Western blot analysis, using an anti-CA antibody, of cell lysates and virions generated in 293T cells transfected with the indicated ts CA mutant or WT proviral plasmid. For each mutant, two lanes are shown: on the left is the sample from cells incubated at 35°C following transfection (Lo, permissive), and on the right is the sample from cells incubated at 39.5°C following transfection (Hi, nonpermissive). The numbers shown below each lane specify the fluorescence intensities (LI-COR) associated with the CA protein pelleted from virion-containing supernatant. (B) Western blot analysis, as in (A), but instead utilizing an anti-HIV-1-MA antibody (p17) to show expression in cell lysates and virions for the same mutant panel. The numbers shown below each lane specify the fluorescence intensities (LI-COR) associated with the MA protein pelleted from virion-containing supernatant.

analysis of Gag expression, processing, and particle production by the ts mutants was also done using an anti-p17 matrix (MA) antibody (Figure 17B). Again we observed uniform levels of p17 in cell lysates regardless of the temperature of production. However, there were large reductions in p17 extracellular particulate expression for the ts CA mutants at the nonpermissive temperature, as the 2.4-25 fold decreases in p17 indicated that particle production was decreased by 59-96%.

While the 2.3 to 25-fold reductions in particle production observed for the ts CA mutants is smaller than the corresponding 50 to 1000-fold reductions in infectious virion yield, it should be noted that the majority (57 to 96%) of the reduction in infectivity must be due to an inability for these mutants to efficiently generate particles. However, these results do also suggest that since the block in the generation of particles was not absolute, there was also some degree of generation of poorly infectious particles, such that attenuation of particle production by ts CA mutants at the nonpermissive temperature varied, rather than being entirely defective. Ultimately, the inability of the ts CA mutants to efficiently generate particles at the restrictive temperature is in alignment with the observation that it is the temperature during particle production, not infection, that determined the phenotype we observed for all of the ts mutants, and that all of these mutants result in significant defects that occur during virion morphogenesis.

Interestingly, inspection of both anti-CA and anti-MA ts mutant blots revealed a specific irregularity in Gag processing. As is typical, the parental virus generated a single ~41 kDa band that was recognized by both anti-CA and anti-MA antibodies at permissive and restrictive temperatures (presumably the p41 MA-CA-p2 processing intermediate, Figure 17A, B). The same was also true for the ts CA mutants at the permissive temperature, however, at the restrictive temperature, one to two additional protein species appeared at a similar, but not identical mobility, to p41. Therefore, for all ts CA mutants, attenuated particle production was accompanied by a disruption in Gag processing, which may be further linked to the overall defects in particle yield and infectiousness. Remarkably, for all eleven ts CA mutants, elimination of the activity of the viral protease also eliminated the temperature-dependent reduction in particle yield (Figure 18). This permits the conclusion that the capsid mutations that instigate temperature-sensitive attenuations of particle production do not do so prior to activation of the viral protease. Combining this observation with the observation that the temperature of inoculation did not impact infectivity, permits the conclusion that all eleven conditionally nonviable CA mutants have defects that are manifested during, and not before or after, virion morphogenesis.

3.2 Analysis of constitutively nonviable CA mutants

Investigation of the eleven conditionally nonviable CA mutants suggested that the fragility of HIV-1 CA arises from requirements imposed during production of particles, and not another portion of the viral life cycle. However, it was

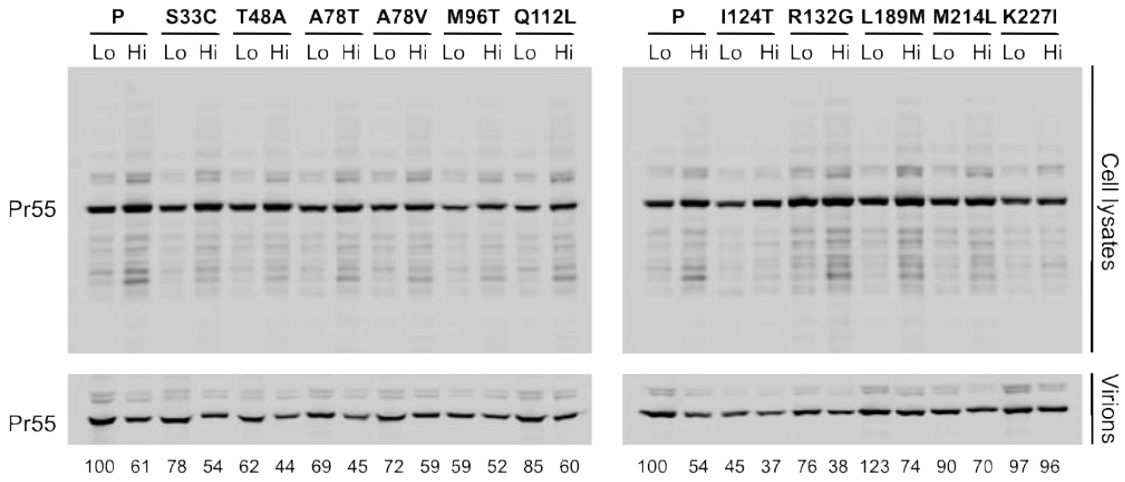


Figure 18. The conditional (ts) CA mutant reduction in particle yield is protease dependent. Western blot analysis of the cell lysates and virions generated by 293T cells transfected with the ts CA mutants (and WT) expressed in a protease-defective proviral plasmid (Pr-). Numbers below each lane indicate the fluorescence intensities (LI-COR) associated with the CA protein pelleted from virion-containing supernatant.

conceivable that the examination of conditionally, rather than constitutively, nonviable mutants could have biased this conclusion. To remedy this, a larger set of 81 constitutively nonviable mutants was examined for their ability to generate extracellular particles by western blot. (Figure 19A).

A striking number (74%) of constitutively nonviable mutants displayed greater than 5-fold reductions in extracellular particle yields. Still, a minority (20%) of the lethal mutations did not significantly impact the magnitude of particle production (less than 2-fold reductions in total particle yield), but instead resulted in the efficient generation of noninfectious particles (Figure 19A). The other 6% of the constitutively nonviable mutants had an intermediate phenotype in which particle reduction was reduced 2 to 5-fold. Occasionally, such as at residues K30, M39, and I115, a change to one amino acid resulted in efficient generation of noninfectious particles, while a change to a different amino acid caused a large reduction in particle yield. However, there again appeared to be no trends in amino acid property that could definitively determine what might cause mutants to efficiently generate noninfectious particles as opposed to causing attenuated particle formation (Figure 13). In alignment with the hexamer mapping distribution of the inactivating CA mutants, constitutively nonviable mutants that yielded WT levels of noninfectious particles also mapped to the inner core of the hexamer (compare Figures 19B and 15B). Furthermore, we also observed that many of the constitutively nonviable CA mutants displayed evidence of disrupted Gag processing, as demonstrated by the presence of additional ~40 kDa Gag

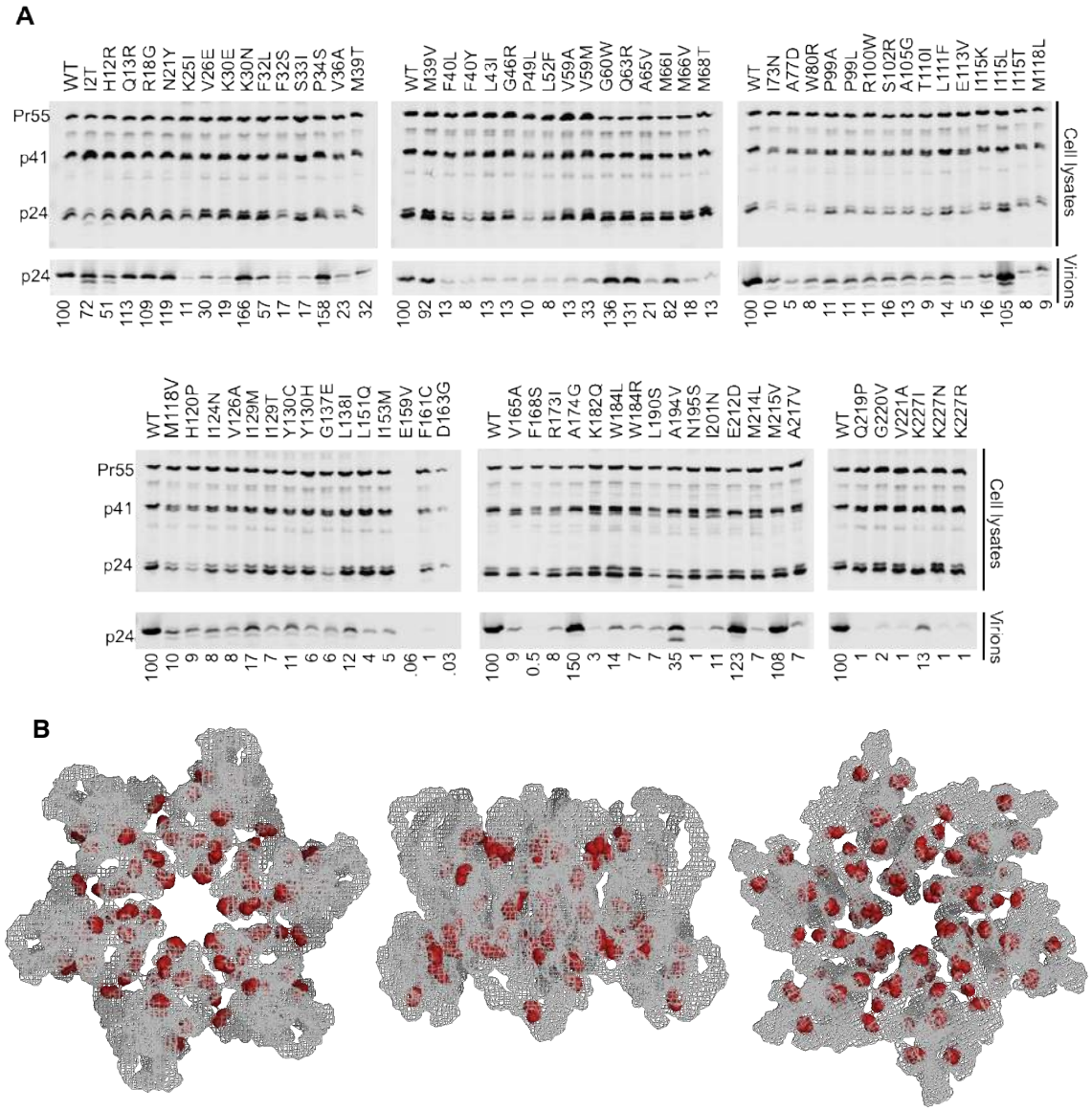


Figure 19. Most constitutively nonviable CA mutants display attenuated particle formation. (A) Western blots showing cell lysates and virions from nonviable mutants (those with <2% of WT fitness in a spreading replication assay) probed with an anti-CA antibody. It should be noted that multiple individual substitutions between residues 159 and 168 yielded diminished signal when the monoclonal anti-CA antibody was used (derived from hybridoma 183-H12-5C). To this end, CA carrying a R167Q mutation is infectious, but is not recognized by this antibody (unpublished observation) suggesting that the epitope of the antibody lies in this region. Therefore, it is likely that the E159V substitution, shown in A, may also disrupt antigen recognition, accounting for the lack of signal in this lane. (B) Locations of mutations in the CA hexamer that are constitutively nonviable, but express near WT levels of particle formation. The image on the left shows the hexamer as viewed from the exterior of the intact conical capsid, the center image shows the hexamer as viewed from within the plane of capsid lattice, and the rightmost image shows the hexamer as viewed from the interior of the intact conical capsid.

protein species, very similar to those present in conditionally nonviable mutants at the restrictive temperature (Figures 19A, 17A, and 17B). Taken collectively, these data reinforce the conclusion that the primary determinant of mutational fragility in HIV-1 CA is the requirement to mediate the efficient and accurate assembly of particles.

3.3 Evaluations of fragility by electron microscopy

To further investigate the nature of the defects in conditionally and constitutively nonviable CA mutants, we utilized electron microscopy (EM) to both quantify and visualize particle morphogenesis. First, to examine the ts CA mutants, 293T cells were transfected at 39.5°C with either the parental proviral plasmid or one of six randomly selected ts CA mutant proviral plasmids, plus a plasmid expressing a modified Vpu-resistant human tetherin (delGI, T45I) [212], in order to maintain virions at the plasma membrane to facilitate visualization. Examination of 150 cells for each ts CA mutant revealed that these mutants generate drastically fewer particles at the restrictive temperature (Figure 20A, shaded bars), with reductions ranging from 4 to 221-fold (75 to 99.5% fewer particles). Visualization of the particles revealed that some of the particles generated by the ts CA mutants seemed morphologically normal, but there also appeared to be a higher fraction of irregular particles and budding structures as compared to the parental virus (Figure 20B). Overall, the smaller number of discernible viral structures produced by the ts CA mutants meant it was difficult to

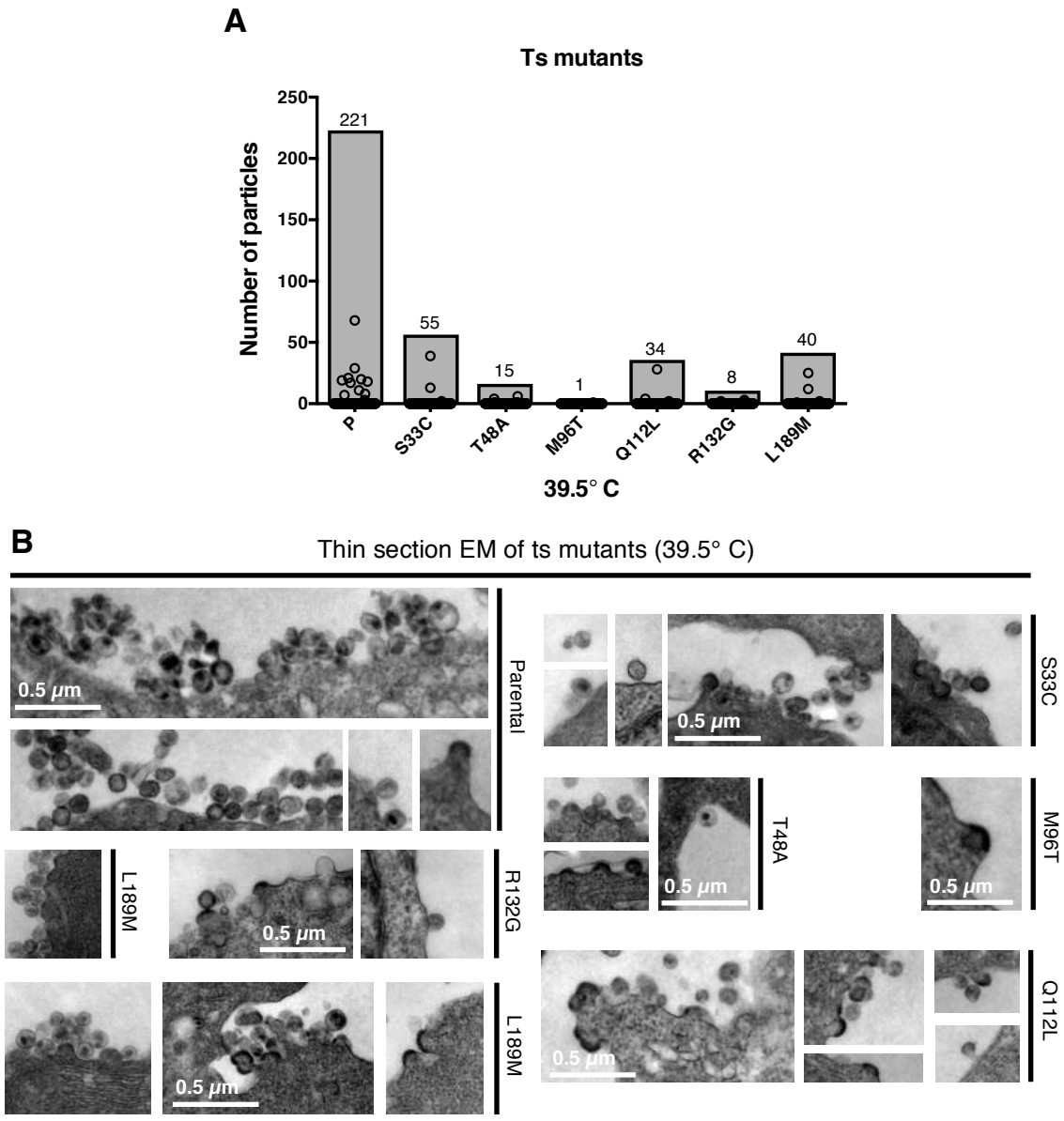


Figure 20. Electron microscopic analysis of particle formation by conditionally nonviable CA mutants. (A) Analysis of thin-section EM images showing the total number of particles produced by either parental virus or ts CA mutants at the restrictive temperature (39.5°C) in 293T cells. Circles represent the number of particles produced by individual cells (n=150 cells for each sample), while the shaded bars indicate the total number of particles counted in the 150 cells. (B) Representative images of the thin-section EM samples analyzed in (A). Scale bars apply to all images.

draw definitive conclusions about the morphological characteristics of the particles that were produced.

Thin-section EM analyses were also conducted on 293T cells transfected with three randomly selected constitutively nonviable mutants (K25I, L52F, and N195S) that had shown significantly reduced extracellular particle yields (Figure 21A, B). As before, this was done in the presence of tetherin (delGI, T45I) to enable inspection and quantification of viral particles. These analyses revealed large reductions in particle yields similar to those that characterized the conditional ts CA mutants. In this case, particle yields were reduced ~6 to 28-fold (82 to 96% fewer particles) as compared to the parental virus. For mutants K25I and L52F, the small number of particles generated appeared to have morphologies similar to the parental virus, but the budding structures produced by mutant N195S were obviously abnormal (Figure 21A, B). Intriguingly, irregular morphologies very similar to those observed for N195S were previously reported for mutations at a nearby CA residue, D197 [155].

So as to provide an additional control for these experiments, we also investigated particle production by EM for a constitutively nonviable mutant (G60W) that displayed high levels of noninfectious particulate p24 in western blot analyses (Figure 21A, B). This mutant actually produced even more numerous particles than the parental virus. Together, the data obtained from the EM analyses of particle formation broadly supported the estimates of particle formation efficiency obtained from western blot assays (Figure 17A, B, and 19A),

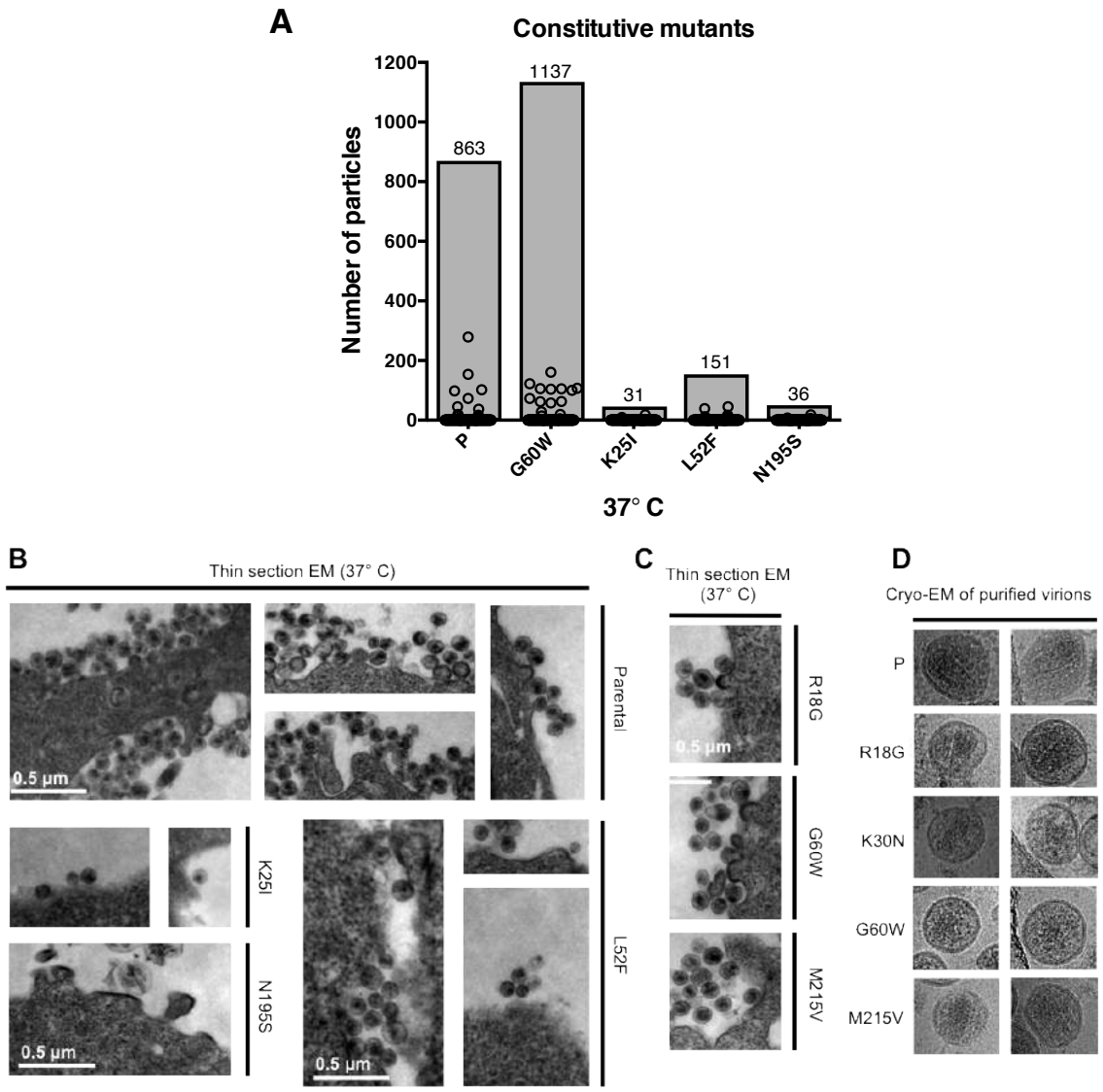


Figure 21. Electron microscopic analysis of particle formation by constitutively nonviable CA mutants. (A) Analysis of thin-section EM images showing the total number of particles produced by either parental or constitutively nonviable CA mutants in 293T cells at 37°C. Circles represent the number of particles produced by individual cells (n=150 cells for each sample), while the shaded bars indicate the total number of particles counted in the 150 cells. (B) Representative images of the thin-section EM samples shown in (A) for WT and samples with reduced particle yields. Scale bars shown apply to all images. (C) Representative thin-section EM images for three constitutively nonviable CA mutants that had high yields of noninfectious particles. (D) Cryo-EM images of purified virions from either parental virus or four constitutively nonviable CA mutants that had high yields of noninfectious particles.

and further emphasizes the notion that most randomly introduced CA mutations cause substantial attenuation, but not complete elimination, of particle formation.

Since a minority (20%) of constitutively nonviable mutants resulted in normal levels of extracellular particles that were noninfectious, we randomly selected an additional three such mutants (R18G, G60W, and M215V) for thin-section EM examination (Figure 21C). None of these mutants displayed particle morphologies that notably differed from particles produced by the parental virus. Moreover, inspection of cell free virions produced by these same mutants, plus one additional nonviable mutant (K30N) was also done by cryo-electron microscopy (Figure 21D). Particle morphologies for these mutants again appeared similar to the WT, perhaps with the exception of G60W, which displayed cores that were more difficult to discern.

Chapter 4. Analyses of CA library mutations in natural HIV-1 subtype B populations

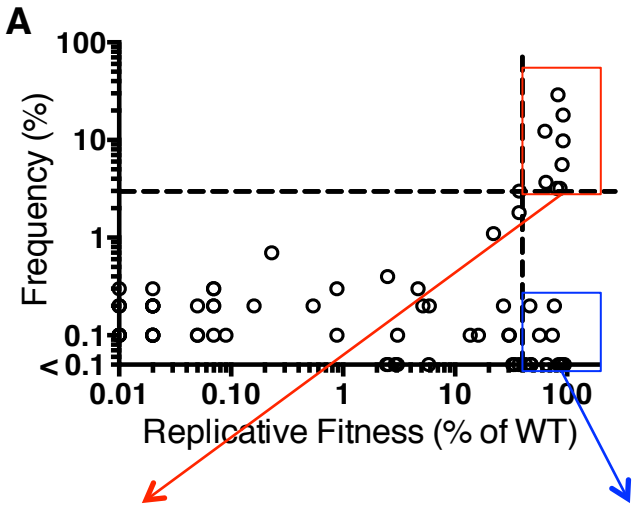
4.1 Occurrence of library CA mutations in an *in vivo* cohort

The *in vitro* measurements of HIV-1 CA mutant fitness suggested that the high degree of fragility could place tight constraints on the sequences of CA that are found in natural populations. Consequently, we obtained a set of 1,000 CA sequences isolated from HIV-1 subtype B infected individuals, so as to compare the fitness of our individual library single amino acid HIV-1 CA mutants with the frequency of their occurrence *in vivo*. From a plot in which *in vitro* replicative

fitness of CA mutations is measured against the frequency with which the mutations occur in natural populations (Figure 22A), it is clear that only mutants that exhibited at least 40% of parental virus fitness in our spreading infectivity assays occurred at a frequency greater than 1% in natural populations. Curiously, however, there was a distinct bimodal distribution regarding the *in vivo* occurrence of the nineteen random mutants that exhibited >40% parental virus fitness. While some of these fit mutants occurred relatively frequently (i.e. in >3% of natural sequences, Figure 22B), others occurred extremely rarely or never (<0.3% of natural sequences, Figure 22C). This indicates that there exists a subset of mutations that incur little or no obvious fitness cost, but are still largely absent from HIV-1 subtype B natural populations. Consistent with their lack of effect on fitness, these seemingly innocuous, but rare, mutant residues occurred preferentially on the outer surfaces of the hexamer, and avoided the inner hexamer core (Figure 22D).

4.2 Examination of 'rare' and 'frequent' mutations in primary cells

The observation that more than half of the random CA mutants that were fit *in vitro* occurred either extremely rarely or never *in vivo* was surprising given the expectation that HIV-1 will take advantage of all available sequence space that does not eliminate fitness. Data from these analyses suggested that some mutations may be selected against *in vivo* in a manner not initially revealed by our fitness assays. Earlier studies have determined that certain CA mutations



B

Mutant	Frequency (%)	Spreading fitness (% of WT)
A14S	3.2	80.69
A14T	3.2	86.62
V83M	3.7	64.78
H87Q	18	92.04
I91V	29	82.84
M96I	9.8	92.04
E98D	12.3	63.37
T200S	5.6	90.35

C

Mutant	Frequency (%)	Spreading fitness (% of WT)
I2L	0	65.67
I6T	0.1	56.88
N21S	0.2	46.55
S33C	0	47.82
I91T	0.2	76.87
R100S	0	42.51
S149C	0.1	73.68
E187V	0	47.39
A204G	0	89.01
A209T	0	94.30
A209V	0	84.29

D

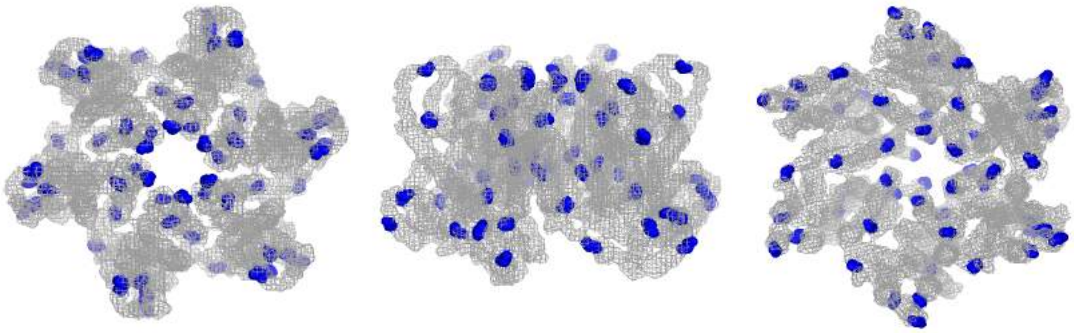


Figure 22. Analysis of CA library mutations in natural HIV-1 subtype B populations. (A) Plot indicating the frequency with which the library mutations occur in 1000 natural HIV-1 subtype B CA sequences (Y-axis) versus the fitness of the same mutations in a spreading replication assay (X-axis). The vertical dashed line shows 40% of WT fitness, below which mutants occurred rarely in nature, while the horizontal dashed line indicates 3% frequency, above which mutants were considered to occur frequently *in vivo*. Mutants with less than 0.01% WT fitness and that were not observed in nature would appear at the origin of the graph, and are not shown. (B) Library mutants that occurred at a frequency greater than 3% in 1000 subtype B sequences and their corresponding fitness values. (C) Fit library mutants that occurred rarely (less than 0.3%) or never in 1000 subtype B isolates, and their corresponding fitness values. (D) Mapping of the residues, shown in (C) on the CA hexamer, where mutations resulted in minimal reductions in fitness (>40% of WT spreading replication), but occurred in less than 0.3% of subtype B isolates. The image on the left shows the hexamer as viewed from the exterior of the intact conical capsid, the center image shows the hexamer from within the plane of the capsid lattice, and the rightmost image shows the hexamer as viewed from within the interior of the intact conical capsid.

can have cell-type dependent effects on HIV-1 infectiousness [133,134,216,217], so it was conceivable that the fit but 'rare' mutations might exhibit fitness defects that are manifested in primary cells, but not in permissive MT-4 cells. To evaluate this possibility, replication assays were performed in multiple primary cell types, using 11 apparently fit mutants viruses that occurred rarely in natural populations (Figure 22C) and 4 randomly selected fit viruses that occurred frequently (Figure 22B).

Our analyses indicated no significant differences between the ability of rarely or frequently occurring fit mutations to infect peripheral blood mononuclear cells (PBMCs), primary CD4+ T cells, or macrophages in short term (semi-single cycle) infection assays (Figure 23A, B, C). Moreover, there was no significant difference in the capacity of the rarely and frequently occurring mutations to replicate in a spreading infection assay in PBMCs (Figure 23C). Ultimately, it appears that the rarity of certain apparently fit mutations in HIV-1 subtype B populations is not related to differences in their capacities to replicate in primary cells. Instead, these findings suggest that there is some unknown selective pressure that permits some intrinsically fit CA variants, but not others, to persist *in vivo.*

4.3 Correlation between fitness and variation in CA

To further examine the relationship between the CA mutations' fitness *in vitro* and their occurrence in natural HIV-1 subtype B populations, analyses were done to compare naturally occurring variability across the CA sequence (as

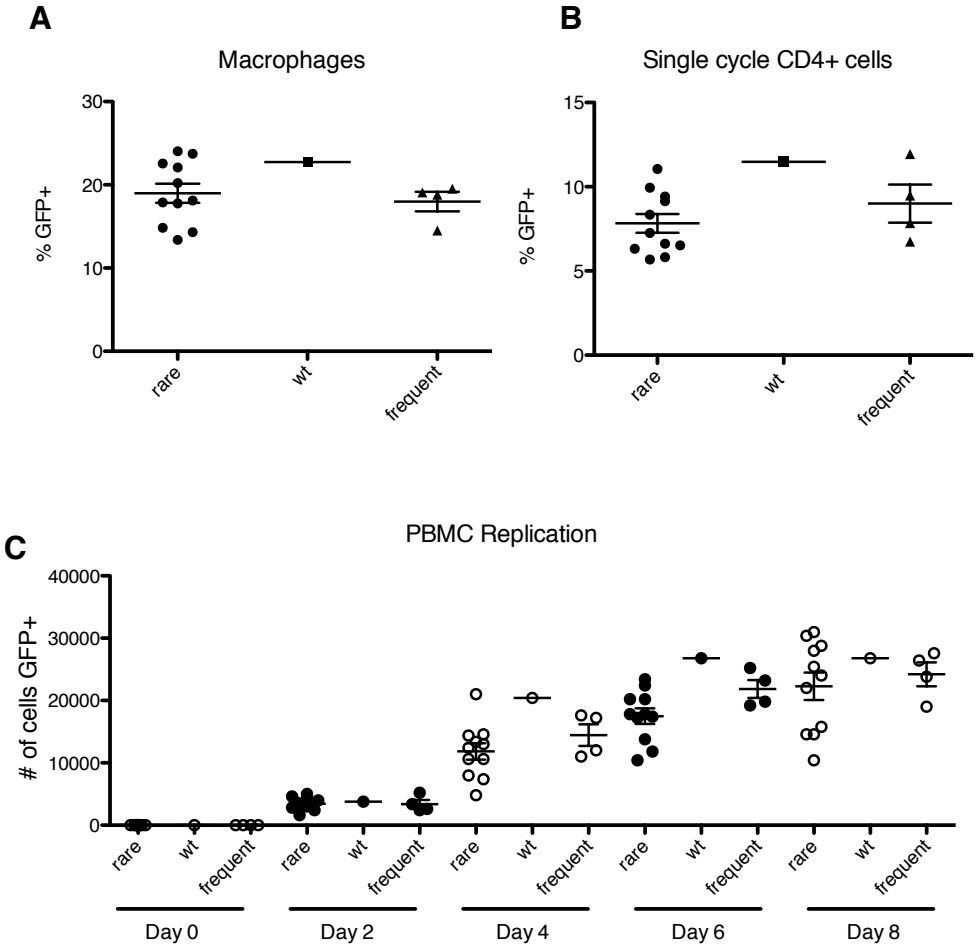


Figure 23. 'Rare but fit' CA mutants are indistinguishable from frequently occurring fit mutants in primary cells. (A) Plot indicating the percentage of infected macrophages after infection with VSV-G pseudotyped WT virus (NHGcapNM, filled squares) or VSV-G pseudotyped fit CA mutations (>40% WT fitness) that occurred either rarely (filled circles) or frequently (filled triangles) in natural sequences. Mutants are those found in Figure 22. (B) Plot indicating the percentage of infected stimulated CD4+ T cells, as indicated in (A) (not pseudotyped). (C) Plot showing spreading replication in PBMCs, as indicated by the number of GFP+ cells, after infection (MOI = 0.1) of stimulated peripheral blood mononuclear cells with virus from either WT or fit CA mutants that occurred rarely or frequently in isolate sequences.

measured by Shannon entropy values) with the propensity of CA domains or regions to tolerate mutation. It seemed a logical expectation that the regions that are more robust (Figure 14) would have higher levels of variability in natural populations. Similar to what was shown in a previous analysis of the Shannon entropy of the N-terminal of subtype C HIV-1 CA [142], our analyses revealed regions of high entropy in CA (>0.4), and other regions that were more conserved (Figure 24A). Clearly, there is also a degree of accord between the capacity of CA domains to tolerate mutation and the Shannon entropy values. This is particularly evident in regions of comparatively high robustness and high entropy, such as the cyclophilin binding loop, residues ~5 to ~35, and residues ~175 to ~210. There were also regions in which especially low robustness corresponded with low variability *in vivo*, such as helices 2 and 3, and also in the major homology region (Figure 24A). Nevertheless, there were also some discrepancies between the Shannon entropy and robustness profiles throughout CA. For example, fitness measurements revealed helices 5, 6, and 7 to be highly genetically fragile, but the Shannon entropy values in this region were often quite high. Ultimately this finding suggests that it is likely that selective pressures in addition to fitness act on the HIV-1 CA sequence *in vivo.*

One such selective pressure likely to impose pressure on CA is that of the adaptive immune system, and in particular, cytotoxic T-lymphocytes (CTLs), which are able to recognize peptides derived from HIV-1 CA. Several previous studies have already shown that the selective pressures inflicted by CTLs drive

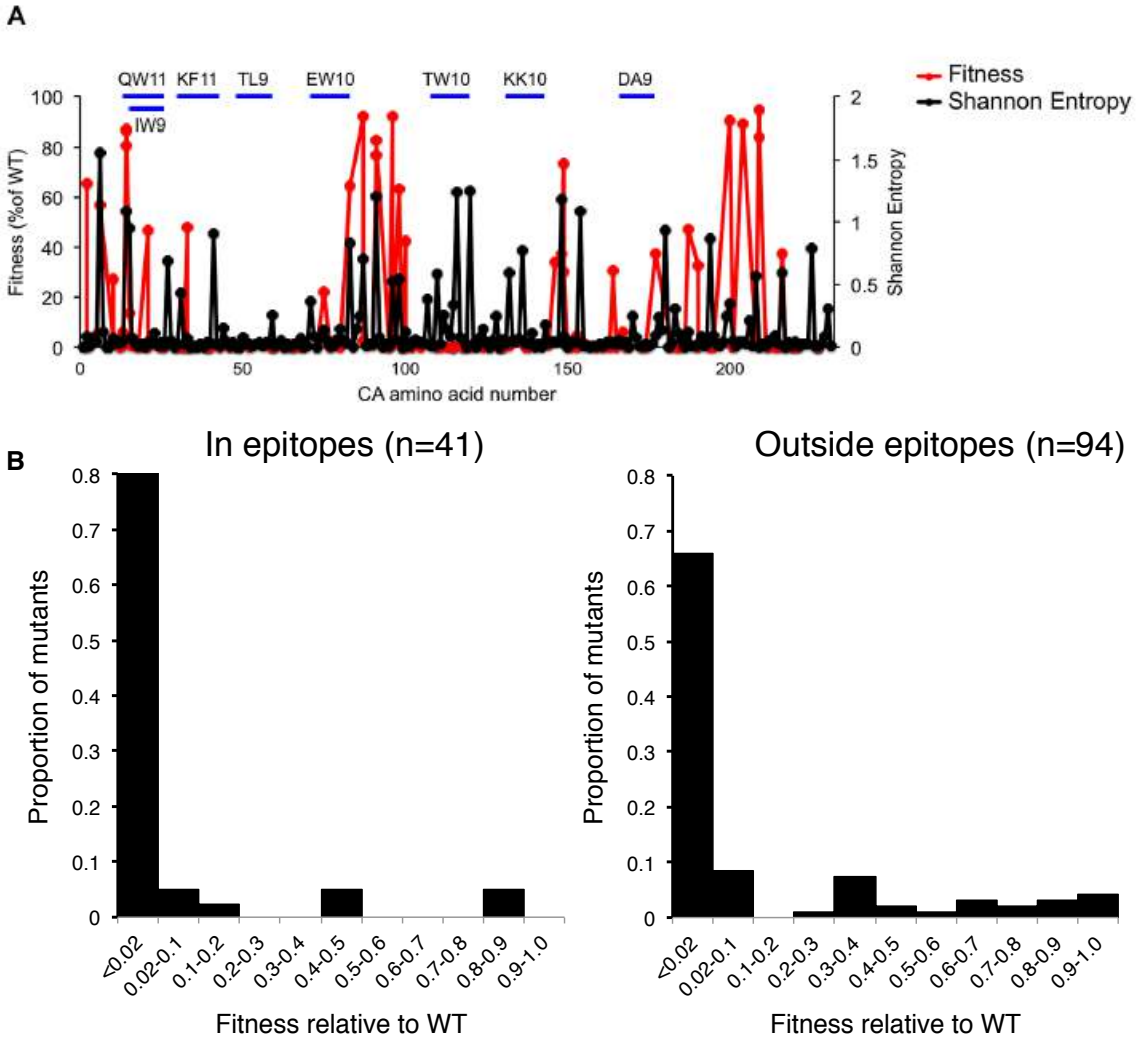


Figure 24. Correlation between CA fitness and natural variation. (A) Plot in which the fitness of the library of CA mutants is overlaid with the Shannon entropy values, derived from 1000 subtype B CA isolate sequences, for each residue in capsid. Locations of CA residues are arranged on the X-axis from left (N-terminal residue) to right (C-terminal residue), and their corresponding replicative fitness (as % of WT in a spreading replication assay) and Shannon entropy values are shown on the Y-axes. The positions of 'protective' CTL epitopes are indicated by horizontal blue lines. (B) Distribution of mutational fitness effects (DMFE), at 37°C, for all single amino acid CA substitution mutants that occurred either in or outside the protective epitopes shown in (A).

certain mutations in CA [142,143,144,145]. More importantly, some of these studies have demonstrated that immune responses to certain epitopes in CA can be protective, such that their presence is associated with slower disease progression. Interestingly, some of the most frequently studied, and most protective epitopes appear to occur preferentially in regions of extreme genetic fragility, as can be seen for HLA-B*57-restricted epitopes KF11 (residues 30-40) and TW10 (residues 108-117), and the HLA-B*27-restricted epitope KK10 (residues 131-140) (Figure 24A, Table 2). Furthermore, a comparison of the DMFE values for the 41 library mutations that occurred in so-called protective epitopes (specifically, QW11 (residues 13-23), IW9 (residues 15-23), KF11 (residues 30-40), TL9 (residues 48-56), EW10 (residues 71-80), TW10 (residues 108-117), DA9 (residues 166-174), KK10 (residues 131-140)) with the 94 library mutations that occurred outside these epitopes indicated that the protective epitopes occur in regions of CA that are more genetically fragile than the rest of the CA sequence (Figure 24B). Indeed, whereas the mean fitness value for mutations occurring outside these protective epitopes was 15.1% of WT, the mean fitness value for mutations within the epitopes was just 7.9% of WT.

4.4 Summary of effects of random mutations in HIV-1 CA.

By utilizing the scope of the mutagenesis in these studies of CA, in which almost half of the amino acids in CA were individually mutated, in conjunction with the analyses of the frequency of these mutations in natural populations, it is possible to make generalized predictions about the fate of randomly introduced

nucleotide substitutions (Figure 25A) or amino acid substitutions (Figure 25B) in CA. For cases of random nucleotide changes (Figure 25A), 22% should result in synonymous changes that, in most cases, will not affect fitness (Figure 11). If fitness was the only negative selection pressure *in vivo*, then 10% of nucleotide substitutions would be expected to result in a nonsynonymous change that is sufficiently replication competent (>40% of WT fitness *in vitro*) to potentially flourish in natural populations. However the results from Figure 22 suggest that only a minority of the mutations that yield variants with >40% of WT fitness *in vitro* are actually capable of thriving *in vivo*, and by this limitation, just 4% of all nucleotide substitutions would be expected to yield nonsynonymous changes that could flourish in a natural environment.

If only nonsynonymous substitutions are considered (Figure 25B), then 87% should result in nonviable virus, with most mutations (59%) causing >5-fold reductions in particle formation, and a smaller portion (21%) resulting in the generation of particles (at at least 20% of the level of WT) that would be noninfectious. Just 13% of nonsynonymous substitutions are predicted to be adequately replication competent (at least 40% of WT replicative fitness) to potentially flourish *in vivo*, based on intrinsic fitness measurements. Again, however, only a fraction of intrinsically fit mutants were shown to persist *in vivo*, such that only 5% of all CA amino acid substitutions are actually expected to thrive in a natural environment.

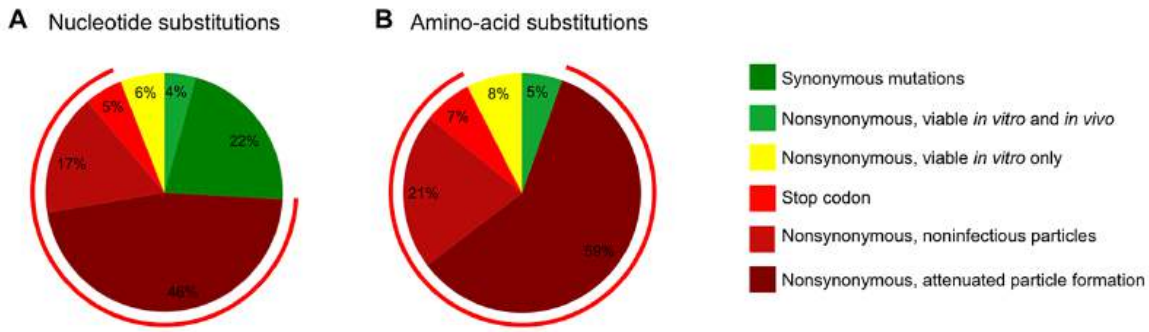


Figure 25. Summary of the effects of random mutations in HIV-1 CA. (A) Assessment, as estimated from the *in vitro* and *in vivo* data shown previously, of the likelihood of various outcomes for random single nucleotide substitutions (A) and random single amino acid changes (B). The charts indicate changes that fall into one of the following categories: synonymous (A only), nonsynonymous, or nonsense (as determined computationally using the pNHGcapNM CA sequence). Since the only mutants frequently observed in natural populations are those with fitness >40% of WT *in vitro*, this threshold was used to consider which mutants could thrive *in vivo* (the proportion of random mutants with fitness less than 40% of WT was derived from data in Tables 1 and 2). The proportion of mutants expected to thrive (nonsynonymous, viable *in vitro* and *in vivo*) and not to thrive (nonsynonymous, viable *in vitro* only) was derived from the data in Figure 22 showing the occurrence of library mutations in natural populations. A red circumferential line indicates the fraction of random mutants expected to be nonviable *in vivo*. Fates of the remaining mutants were derived from the data shown in Figure 19 (nonsynonymous, noninfectious particles or nonsynonymous attenuated (>5-fold deficit) particle formation).

Chapter 5. Additional applications of the CA mutant library

5.1 Screen of CA library for resistance to Mx2 restriction

Aside from demonstrating a rather extreme intolerance to mutation, the randomly mutagenized CA library we have created can also be utilized in other screening approaches to test the relevance of capsid and its residues in numerous other HIV-1 functions. Among the biological properties of capsid are multiple physical interactions with host cell proteins, including karyopherin β transportin-3 (TNPO3), nucleoporin 153 (NUP153), nucleoporin 358 (NUP358)/RanBP2, cleavage and polyadenylation factor 6 (CPSF6), and cyclophilin A [128,129,130,131,132,133,218], and it seems likely further proteins will be added to this list. Work from our laboratory, and others, has recently highlighted how Myxovirus resistance protein 2 (Mx2, sometimes referred to as MxB) can efficiently inhibit early steps of HIV-1 infection *in vitro* [136,137,138]. While the mechanisms by which the interferon-mediated Mx2 suppresses HIV-1 replication have not been clear, these studies have all reported HIV-1 capsid to be a significant determinant of Mx2 sensitivity, as a few capsid amino acid substitutions were able to confer either partial or complete resistance to Mx2 activity. These reported mutations reducing sensitivity to Mx2 were largely localized to the cyclophilin binding loop (at residues 88-90, 92 and 94), but also included one mutation, N57S, which has previously been reported to impact HIV-1 infection of non-dividing cells [127,136,137,138].

To identify other CA residues that could confer resistance to the antiviral activities of Mx2, 46 of the CA library mutants that had greater than 2% of WT infectivity in the single cycle infectivity assay were selected and used to challenge MT-4 cells expressing Mx2 from various species (Figure 26A). The ovine and canine variants of Mx2 displayed almost no antiviral activity against either WT or any of the CA mutant viruses. In contrast, most of the CA mutants were inhibited by primate Mx2 species variants from human (Hs), macaque (Mac), and African Green Monkey (Agm) (Figure 26A). Interestingly, however, seven of the CA mutants conferred significant resistance to at least one of the primate Mx2s (Figure 26A, B).

Like the previous studies of Mx2, several mutations in the cyclophilin binding loop were able to confer resistance to Mx2. These included the P90T substitution, which had not previously been reported, and conferred total resistance (in Hs), as well as substitutions at newly implicated residues, H87R and Q95L, that conferred significant partial resistance, to varying degrees, in the different primates. Identification of these two mutations, together with the previously established Mx2 resistant capsid mutations, suggests that the cyclophilin binding loop is likely an important Mx2 sensitivity determinant.

Perhaps more interesting was the identification of the other CA mutants promoting Mx2 resistance. Two mutants in the N-terminal domain conferred at least partial resistance, M10I and N57S. While the N57S mutant was previously reported, the M10I mutant, located in the N-terminal β-strand, was a far more

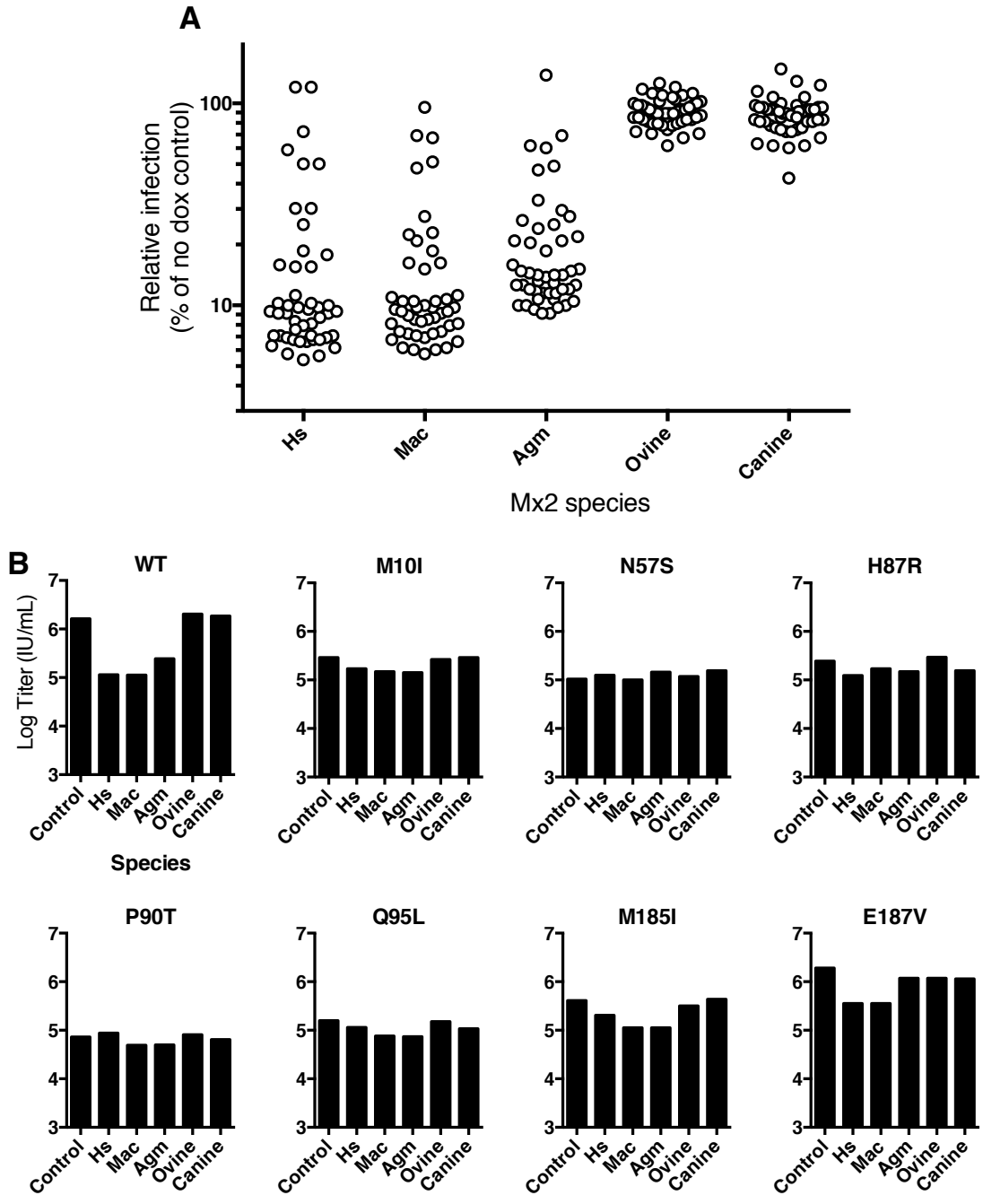


Figure 26. Screening of viable CA library mutants for resistance to Mx2. (A) Screen indicating the level of infection, as compared to non-doxycycline treated controls, for 46 CA mutants (those viable in single cycle infectivity assays) in doxycycline treated MT-4 T-cells expressing doxycycline inducible Mx2. Multiple species variants of Mx2 were tested for resistance, as shown on the X-axis. Mutants are represented by individual circles. (B) The levels of infection for the 7 CA mutants conferring the greatest resistance to Mx2 identified in (A). Viral titers were measured in MT-4 T-cells expressing doxycycline inducible Mx2 species variants with either WT (NHGcapNM) or CA mutants, in the presence of pre-treatment doxycycline, except in the case of the control (HsMx2-MT-4 with no doxycycline treatment).

surprising result. Intriguingly, this residue, like the residues in the cyclophilin binding loop, is a part of the cytosolic surface of mature CA that is known to be available to inhibitory interactions. Furthermore, M10 and cyclophilin binding loop residues have been previously reported to be involved in escape from another antiviral factor, TRIM5α [140,219]. Outside of the N-terminal domain, there were two previously unreported mutants in the C-terminal domain that also conferred partial resistance to Mx2, M185I and E187V. Interestingly, residue M185 is known to be involved in a vital hydrophobic contact at the CA dimerization interface [220]. As for mutant E187V, it appeared to convey only partial resistance to Hs and Mac Mx2, but gave nearly complete resistance to Agm Mx2 (Figure 26B).

When the seven mutant residues conferring resistance to Mx2 in the screen were mapped onto the capsid hexamer, a certain trend was apparent (Figure 27). In general, residues conferring resistance were localized to very exposed surface areas. This is perhaps not entirely surprising given the propensity for residues in the cyclophilin binding loop to be involved in Mx2 escape, but it was intriguing that the trend also extended to the C-terminal exposed residues, and the N-terminal M10 residue. The higher surface exposure of these residues within the mature CA hexamer may mean they are more accessible for direct interactions with host factors like Mx2. As the CA library is not inclusive of mutants for all of the exposed surface residues, it is likely that there are other residues in these locations that could confer resistance to Mx2.

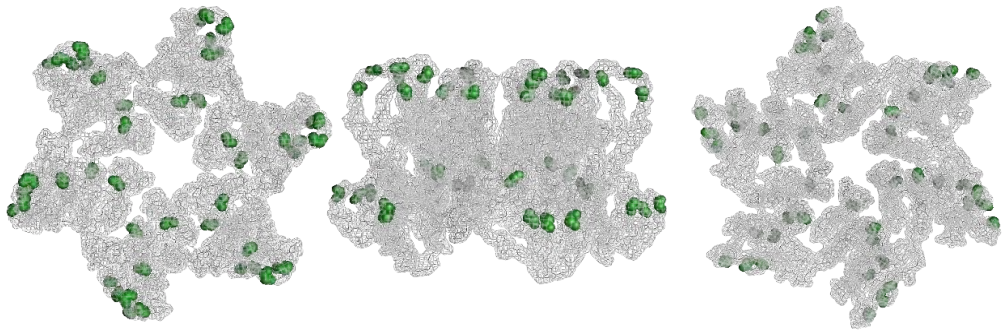


Figure 27. CA mutant residues conferring Mx2 resistance typically occur in surface exposed regions of the CA hexamer. The locations in the CA hexamer (PDB: 3GV2) of the seven residues identified in the CA-Mx2 resistance screen conferring the greatest degree of resistance are shown in green. The image on the left shows the hexamer as viewed from the exterior of the intact conical capsid, the center image shows the hexamer from within the plane of the capsid lattice, and the rightmost image shows the hexamer as viewed from within the interior of the intact conical capsid.

In summary, this screen of CA library mutants for Mx2 resistance highlights how valuable a resource the CA random mutants can be for probing functions and interactions with a variety of proteins and molecules.

Chapter 6. HIV-1 IN and creation of a randomly mutagenized IN library to better understand genetic robustness and fragility

6.1 Introduction to genetic robustness in HIV-1 IN

Having evaluated the robustness of HIV-1 CA, determined the biological basis for its extreme fragility, and investigated correlations between *in vitro* CA mutant fitness and *in vivo* occurrence, we were curious as to how unique CA's fragility may be. Satisfying this interest required evaluating the robustness of another HIV-1 protein in a similar manner. Evaluation of the genetic robustness of another essential HIV-1 protein could also reveal more significant patterns about how competing selection pressures, such as those to both maintain structure and function, and to also diversify sequence to evade immune pressure, are likely to play out in other HIV-1 proteins.

Like HIV-1 CA, HIV-1 integrase (IN) is a highly conserved, essential, and multifunctional protein, and it is also the enzyme responsible for the integration of the viral genome into the host chromosome. It has previously been suggested that enzymes like IN may be particularly sensitive to mutation given their requirement to adopt and change their tertiary structure in very subtle and flexible ways [92]. The relatively highly conserved nature of IN also suggests it could be

vulnerable to substitutions [156]. Finally, the requirement for IN to interact with a number host factors, as is similarly true for CA, may also contribute to genetic fragility.

As the studies of CA robustness demonstrated, a protein's structure can influence its genetic robustness. While the X-ray crystal structure of the HIV-1 intasome, the tetrameric dimer of dimer IN-viral DNA functional structure, has not yet been determined, the publication of the crystal structure of the prototype foamy virus (PFV) intasome [166,167] has provided further insight into the possible arrangement of HIV-1 IN's domains within the tetramer. It has also led to new and improved HIV-1 intasome models [164], one of which will be utilized in the subsequently described robustness studies [168].

Understanding the robustness of HIV-1 IN is particularly valuable given the highly beneficial clinical contributions integrase inhibitors have already made, and the further therapeutic potential IN holds [221]. In this second study of genetic robustness in HIV-1, the generation of an even larger unbiased mutagenesis library will be described, in which the *in vitro* robustness of HIV-1 IN and CA can be compared, and further analyzed in conjunction with the *in vivo* occurrence of IN mutants. To this end, creation of a large, random IN mutant library will also permit the utilization of a recent HIV-1 intasome model in which it will be possible to identify the striking vulnerability of particular IN regions to mutation.

6.2 Comparison of IN and CA variability *in vivo*

In order for comparison of the robustness of HIV-1 IN and CA to be meaningful, it was necessary to establish, as quantifiably as possible, how similarly the two proteins might maintain mutations *in vivo*. As robustness describes a tolerance to mutation, it seemed prudent to compare how frequently variants, introduced by reverse transcription or RNA polymerase II, actually appeared in separate cohorts of 1,000 HIV-1 subtype B sequences isolated from infected individuals for each protein. By measuring the number of sequences that varied from the parental sequence at each residue in both proteins, it seemed possible to establish how similarly each protein might tolerate mutations.

When variation amongst the 1,000 HIV-1 subtype B sequences for CA and IN is mapped in a landscape fashion, CA and IN appear to have relatively similar mutation frequencies *in vivo* (Figure 28A, B). The peaks denoting quite high variability from WT are present at comparable rates—around 9 individual residues in IN and 11 unique residues in CA contain at least 20% variation (such that of the 1,000 sequences for each protein, at least 200 sequences varied from the parental sequence at those distinct residues). IN and CA also share a similar incidence for the residues indicating little or no variance. The similarities between IN and CA variation are also particularly apparent when the frequencies of variants at a residue are arbitrarily binned (Figure 28C, D). The percentage of residues with a given frequency range of variability differs between IN and CA the most for residues with high variability (3.1-30% variants at a residue), as 13% of

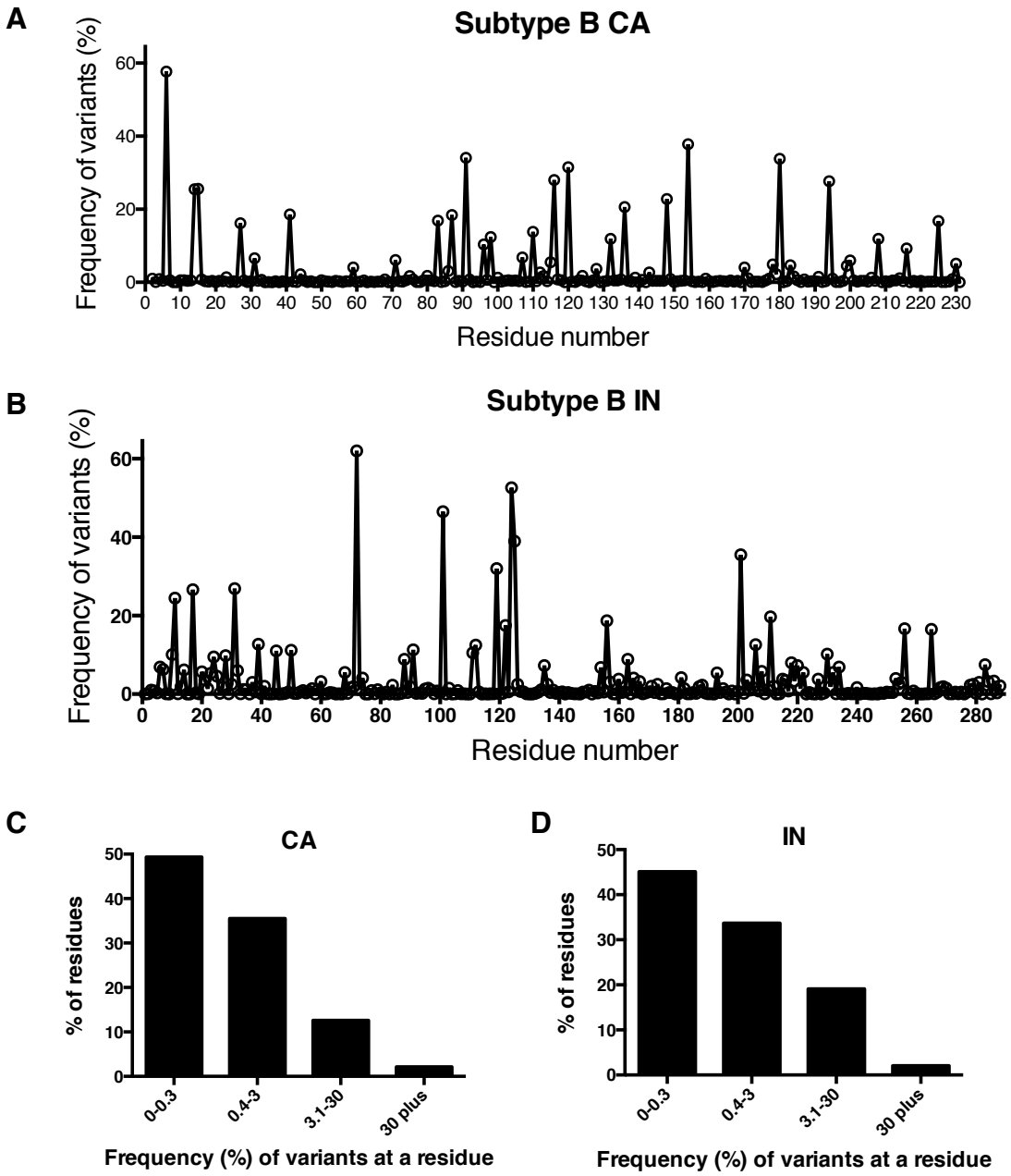


Figure 28. Comparison of naturally occurring variation in HIV-1 IN and CA. (A, B) Survey of the frequency (%) of variants from WT that occurred at individual residues within 1000 HIV-1 subtype B CA (A) or IN (B) isolates. Residue position is shown on the X-axis from left (N-terminal residue) to right (C-terminal residue). (C, D) The variability of IN and CA (as shown in A, B) binned such that the Y-axis indicates the percentage of residues within CA (C) or IN (D) with the indicated frequency of variants at a residue shown on the X-axis.

CA residues and 19% of IN residues have this degree of variability. Still, the percentage of residues with low (0-0.3%), intermediate (0.4-3%), and very high (>30%) variability is strikingly similar. Ultimately, surveying the naturally occurring variation in IN and CA suggests that similar proportions of mutations are tolerated *in vivo*, implying that comparing IN and CA robustness is not only fitting, but an endeavor that may reveal similar degrees of fragility.

6.3 Creation of the randomly mutagenized IN library

For construction of the randomly mutagenized IN library, the procedures were analogous to those used for generation of the CA library (Figure 29A, B). As before, a low fidelity PCR method, Genemorph II, was used to produce error-prone IN encoding sequences. These mutagenized sequences were digested and inserted into the replication competent clone, pNHGintBS, using the unique *BstBI* and *SpeI* restriction sites. The pNHGintBS is nearly identical to the pNHGcapNM (accession: JQ686832) used for the construction of the CA library, and differs only by the creation of the *BstBI* site and reconstruction of the *SpeI* site to flank IN. Like pNHGcapNM, pNHGintBS encodes EGFP in place of Nef, and it will hereafter be referred to as the parental or WT virus for experiments investigating IN, as will pNHGcapNM for experiments involving CA.

6.4 Composition of the IN library, and comparison with the CA library

Subsequently during the IN library construction, around 770 single colonies were isolated and cultured, and proviral plasmid DNA was extracted and sequenced. After removal of clones that had either failed sequencing reactions,

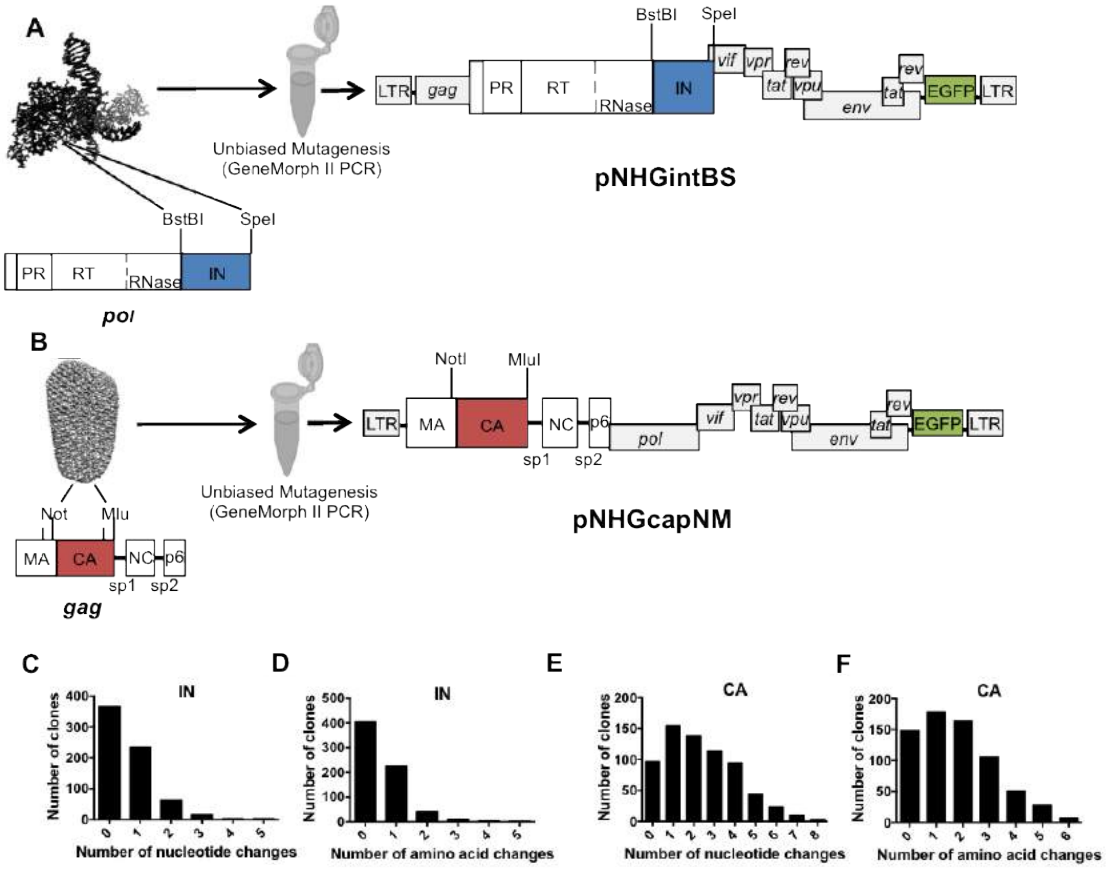


Figure 29. Generation, characterization, and comparison of the IN and CA randomly mutagenized libraries. (A) Schematic that illustrates the PCR-based mutagenesis of the 864 bp amplicon containing HIV-1 IN, which is flanked by *BstB*I and *Spe*I restriction sites. Subcloning into the replication-competent proviral plasmid pNHGintBS followed Genemorph II mutagenesis of the amplicon. (B) Schematic that illustrates the PCR-based mutagenesis of a 717 bp amplicon containing HIV-1 CA (and a few residues of MA), flanked by *Not*I and *Mlu*I restriction sites. Subcloning into the replication competent proviral plasmid pNHGintBS followed Genemorph II mutagenesis of the amplicon. (C, E) Distributions of the number of nucleotide substitutions found in each clone in the IN (C) and CA (E) mutant libraries (including nonsense and duplicate mutations, mutations in MA and frameshifts). (D, F) Distributions of the number of amino acid substitutions found in each clone in the IN (D) and CA (F) mutant libraries (including nonsense and duplicate mutations, mutations in MA, and frameshifts).

chromatograms that indicated the presence of more than one template, or that showed evidence of recombination, 683 clones remained, and the distribution of their nucleotide changes (Figure 29C) and amino acid substitutions (Figure 29D) was determined. Comparing these IN library composition analyses with the distributions of CA library nucleotide changes (Figure 29E) and amino acid substitutions (Figure 29F) exposes the difference in the error or mutagenesis rates used during the creation of the two libraries. Whereas the conditions for the CA library led to a single nucleotide change being the most frequent mutational outcome, the optimized lower mutagenesis rate utilized for the generation of the IN library meant far fewer clones had multiple nucleotide changes. Thus, the IN library yielded more single nucleotide substitution clones from a smaller pool of clones. Accordingly, this trend is reiterated in the distribution of amino acid substitutions (Figure 29D, F), and therefore the IN library included more mutants with single amino acid substitutions than the CA library. It is these mutants that will be employed in the work described hereafter.

From the resulting single amino acid mutants in each library, additional clones were removed due to the presence of nonsense mutations, duplicate mutations of clones already in the library, frameshifts, or mutations outside the gene (only the case for the CA library, in which a very small fraction of matrix was in the mutagenized amplicon). As a result, the final IN library contains 156 single mutants, which cover 118 (41%) of the 288 IN residues, and, as previously

described, the CA library contains 135 single mutants covering 102 (44%) of CA's 231 residues.

Chapter 7. Evaluations of IN library robustness and comparison with CA

7.1 Fitness of single residue IN mutants

To compare the robustness of HIV-1 IN and CA, the fitness of single mutants was analyzed in two assays, as described and completed previously for the CA single amino acid mutants. To briefly reiterate, in the single cycle infectivity assay, 293T cells were transfected with either WT or single residue IN mutant proviral plasmids, and the resulting virus-containing supernatant was added to MT-4 cells (at an MOI of ~1 for the WT virus), which were treated with dextran sulfate 16h post-infection to restrict infection to a single cycle. Alternatively, to measure replicative fitness, virus-containing supernatant harvested from WT or single IN mutant transfected 293T cells was added to MT-4 cells in a much smaller dose (at an MOI of ~0.01 for the WT virus), after which multiple rounds of replication were permitted until the WT virus had infected >50% of cells (approximately 80 hours later) and the experiment concluded. In both assays, fitness was calculated using the number of infected (EGFP-positive) cells, expressed as the percentage of the number of cells infected by the WT virus. So as to distinguish between robust and fragile mutants in these assays, the same arbitrary cut-off for viability used for evaluating the CA library was again employed and generously set at 2% of parental virus fitness, below which it is

unlikely a virus would be able to thrive *in vivo*. Since the replicative fitness assay incorporates a low-dose challenge and multiple replication cycles, this assay is likely a better determinant of how a mutant virus might behave *in vivo*, and for this reason, mutants with less than 2% of WT fitness in this assay were considered nonviable in evaluations of both the CA and IN libraries.

Considering the indispensable nature of HIV-1 IN, and its resemblance in variability to CA, it was surprising to discover that just 35% of single IN amino acid substitutions resulted in nonviable virus by the replicative fitness assay, whereas twice as many, 70%, of single CA amino acid substitutions were nonviable in the same assay. The comprehensive list of the 101 viable IN mutations, with replicative fitness >2% of WT, can be found in Table 4, while the 55 nonviable IN mutants can be found in Table 5. The analogous single cycle and replicative fitness measurements are available for CA in Tables 1 and 2. As was also true for CA, comparison of IN Tables 4 and 5 reveals how different substitutions at the same residue can sometimes result in vastly different fitness outcomes, as is the case for residue C65, indicating that the type of residue change can be important. More often, however, multiple unique aa changes at the same residue resulted in relatively similar fitness. Analysis of the role of an IN mutant's type of residue change, with regard to hydrophobicity or molecular weight, in determining fitness did not reveal consistently significant trends (Figure 30A, C). However, polar basic IN residues were generally more sensitive to

Table 4. Fitness measurements of viable HIV-1 IN mutants

IN mutant	Single cycle infectivity (% of WT)[a]	Spreading replication infectivity (% of WT)[b]	IN mutant	Single cycle infectivity (% of WT)[a]	Spreading replication infectivity (% of WT)[b]
E10D	91.5	107.9	Q148R	33.0	13.8
E13V	78.4	93.9	I151T	32.4	12.0
S17N	91.8	141.2	K156T	36.0	15.7
P30H	9.5	3.9	K159M	44.7	29.5
P30S	11.3	2.3	K160E	26.5	11.8
K42I	71.8	71.2	I162T	6.0	3.1
K42R	77.6	87.7	G163R	117.1	105.1
A49V	77.4	79.9	Q164H	65.1	107.2
M50I	89.3	75.2	D167E	92.6	132.9
M50L	100.2	86.2	E170K	64.9	88.8
M50V	80.6	71.1	L172F	91.0	143.4
Q53L	63.9	39.7	T174I	44.3	57.9
V54A	87.0	89.9	V176I	9.5	6.6
D55E	45.4	30.3	V180I	51.6	63.6
D55N	79.9	65.7	F181Y	68.3	95.1
C56Y	108.6	91.1	H183Q	11.8	9.6
G59E	95.2	65.0	G189E	24.6	8.6
C65S	80.7	60.6	G197W	13.2	8.4
L68S	94.2	103.6	E198Q	89.8	91.5
V72D	35.9	15.6	I200M	85.5	98.5
L74M	73.6	64.5	I203L	86.5	98.0
V75I	99.7	73.4	I203M	101.5	101.5
V75L	90.1	68.8	I204V	82.6	87.0
V77I	82.5	56.8	A205S	84.3	79.0
Y83H	77.9	52.5	T210I	45.3	44.6
P90L	38.2	16.9	K211R	63.0	55.7
P90Q	73.8	66.4	Q214E	79.1	79.1
A91T	73.3	44.6	Q214H	79.0	54.8
A91V	81.6	46.5	Q214L	13.5	5.2
T93I	19.5	15.5	K215R	86.4	86.3
Q95L	42.0	35.3	K215T	54.7	25.9
L102F	18.2	13.3	Q216H	83.6	102.1
T112I	92.9	61.8	N222I	14.8	5.8
T112R	66.6	44.1	N222Y	85.2	93.8
V113L	105.0	86.9	K236I	65.1	23.0
T115I	18.1	7.9	K244N	39.7	21.4
S119N	5.5	2.2	I251V	74.1	92.0
S123N	54.6	40.4	Q252H	77.4	89.6
S123T	13.5	3.3	S255I	87.1	70.1
T124I	74.8	58.1	D256V	98.5	53.3
T124S	110.1	94.2	I257L	77.0	65.3
T125A	119.8	92.0	I257M	66.3	34.4
V126F	86.9	82.9	I257V	79.0	48.7
K127T	58.8	39.1	K258R	88.4	62.6
A129S	36.1	28.6	P261S	102.0	85.8
A129T	43.6	35.0	M275L	65.0	47.7
A129V	20.3	9.6	G277S	67.3	49.8
W131C	24.6	15.2	A282T	99.2	87.2
Q137H	97.9	87.4	D286N	101.1	122.4
E138D	103.7	90.4	D288G	107.8	86.4
P145S	19.6	3.3			

Viable, or infectious, mutants were those with at least 2% of WT infectivity in the spreading replication assay.

[a] Fitness measurement in which MT-4 cells were inoculated, at MOI=1 for the WT virus, with virus-containing supernatant from 293T cells transfected with single residue IN mutant proviral plasmids. At 16h post-infection, dextran sulfate was added to limit replication to a single cycle. Values shown are the percentage of infected cells (GFP+) as compared to WT.

[b] Measurement of infectivity in which MT-4 cells were inoculated, at MOI=0.01 for the WT virus, with virus-containing supernatant from 293T cells transfected with single residue IN mutant proviral plasmids. Multiple replication cycles were allowed over an 80h period. Values shown are the percentage of infected cells (GFP+) as compared to WT.

Table 5. Nonviable HIV-1 IN mutants

IN mutants	
G4E	N144I
N18I	E152D
K34E	N155Y
V37E	V165L
C40F	R166I
G52R	M178V
D64E	A179V
C65F	I182M
C65Y	I182N
A76S	K188E
A76T	K188N
V77F	G192E
H78L	A196G
S81N	G197E
I84V	G197R
A86T	E198K
A86V	D202E
E87G	I204K
L102S	E212D
K103E	L213F
W108R	R228W
P109Q	K244I
P109S	V249E
V113A	I251K
G118D	K258E
W132C	K266N
I135F	R269G
I135T	

Mutants were considered nonviable if they had less than 2% of WT infectivity in the spreading replication assay, as described in Table 4.

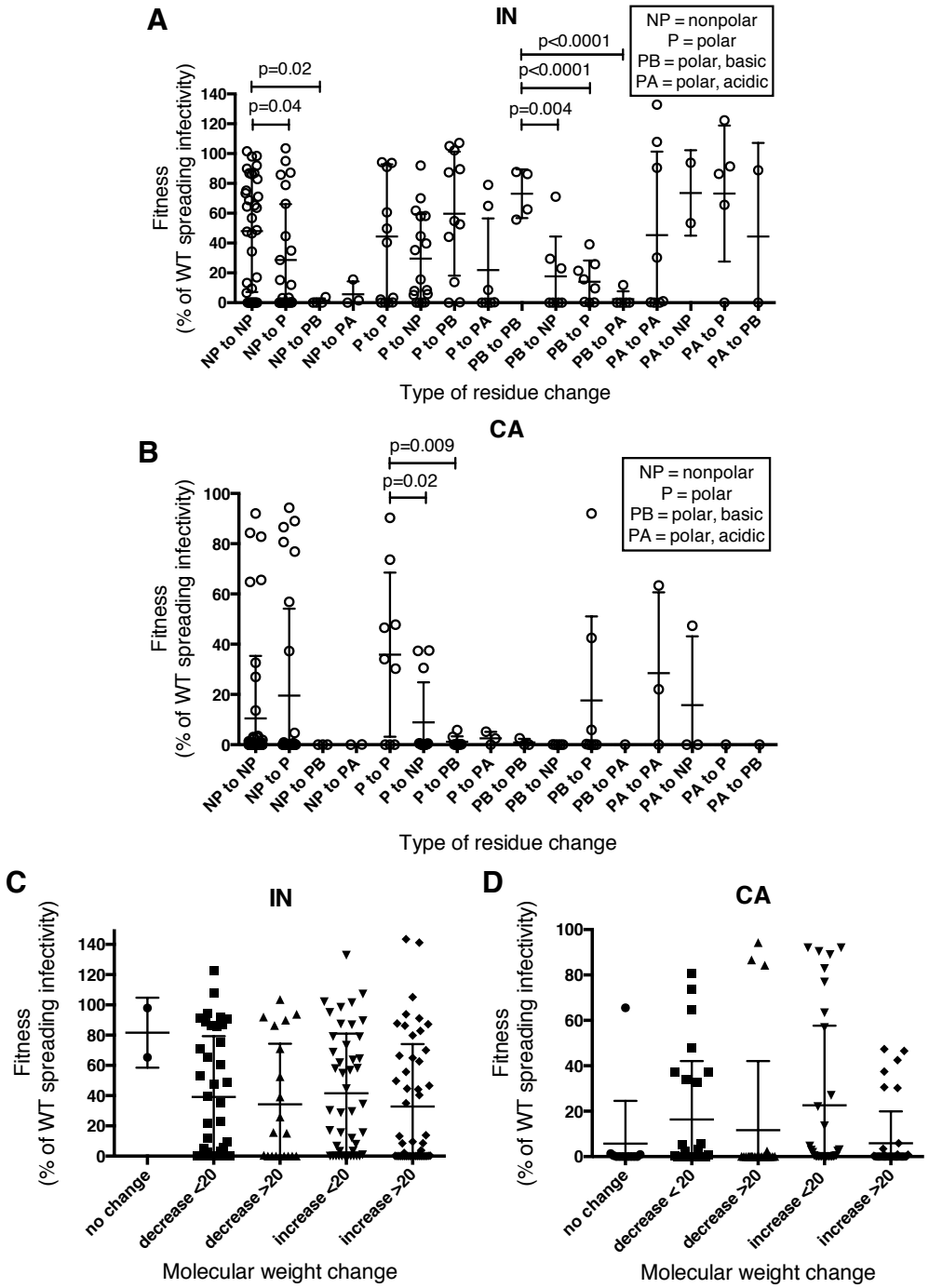


Figure 30. Mutational fitness in IN and CA is not definitively determined by changes in hydrophobicity or molecular weight. (A, B) Evaluation of the effect of mutant residue changes in hydrophobicity in IN (A) and CA (B) on fitness. Figure legends indicate the abbreviations used to designate the initial hydrophobicity of a residue and the resulting change. Analyses of disparities between conservative changes (i.e., polar (P) to polar (P)) and nonconservative changes (i.e., polar to nonpolar (NP)) were performed for all scenarios, and all significant differences are indicated. (C, D) Evaluation of the effect of molecular weight change in IN (C) and CA (D) library mutants on fitness. Analyses between residue substitutions in which there was no change in molecular weight and those with either increases or decreases in molecular weight indicated no significant differences. Significance determined using unpaired, parametric t-tests.

substitution (see residues 34, 42, 78, 103, 166, 188, 228, 244, 258, 266, 269), perhaps because these residues are more likely to be involved in DNA binding [222], but were not entirely sensitive, as not all substitutions at polar basic residues resulted in significant differences in fitness (residues 127, 156, 159, 160, 215, 236). Overall, analyses of the fitness effects of the type of residue change in both the IN and CA libraries (Figure 30A-D) suggest that there are no overarching patterns that can determine how a substitution's change in hydrophobicity or molecular weight will affect its fitness. Together, the values provided in Tables 4 and 5 also provide an opportunity to make more general observations regarding IN domain fragility. As Table 6 indicates, the CTD domain is notably more tolerant of mutation than the CCD or NTD.

7.2 Distribution of IN fitness and divergence with CA fitness

The fitness measurements of IN differ from those of CA extensively. As the distributions of mutational fitness effects (DMFE) measurements for IN and CA demonstrate, not only are there fewer mutants with severe fitness attenuations in IN, there are also more IN mutants with intermediate, small, or no fitness reductions, and this reinforces the notion that IN is more robust than CA (Figure 31A, B). In many ways, IN mutant fitness distributions are similar to the DMFE measurements reported for other viruses, including RNA viruses, and this further strengthens the idea that capsid is unique in its extreme fragility [86,87,93]. In fact, whereas 89% of random CA single mutants caused >2-fold reductions in fitness, only 62% of IN single mutants had this effect. When the

Table 6. IN mutant phenotypes by region

Region	Number of viable mutants	Number of nonviable mutants	Fragility (%)[a]	Mutants resulting in reduced particle yield (%)[b]	Mutants resulting in reduced IN expression in particles (%)[c]	Mutants affecting clathrin particle incorporation (%)[d]	Nonviable mutants only affected by loss of infectiousness (%)[e]
NTD	8	5	38	0 (0%)	6 (46%)	2 (15%)	0 (0%)
CCD	69	42	38	9 (8%)	25 (23%)	11 (10%)	25 (23%)
CTD	24	8	25	2 (6%)	3 (9%)	2 (6%)	5 (16%)
IN total	101	55	35	11 (7%)	34 (22%)	15 (10%)	30 (19%)

[a] The percentage of mutants considered nonviable (<2% of WT replicative fitness).
[b] Mutants with at least 2-fold reductions in western blot virion IN expression.
[c] Mutants with at least 2-fold reductions in western blot particle production, as measured by levels of p24 in virions.
[d] Mutants with at least 2-fold reductions in western blot clathrin incorporation into virions.
[e] All nonviable mutants with less than 2-fold reductions in both western blot virion IN expression and western blot particle production.

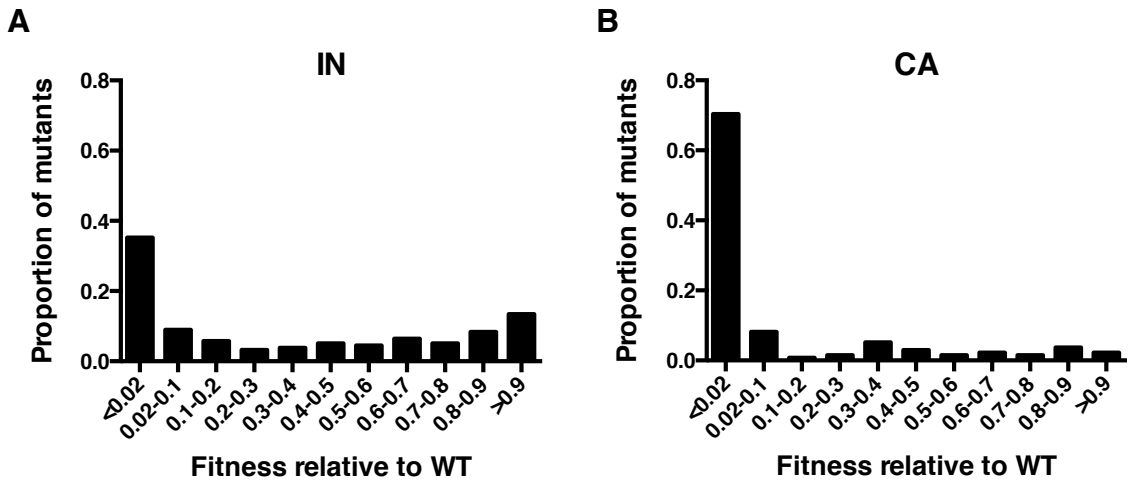


Figure 31. Distributions of mutational fitness effects in IN and CA. (A, B) Distribution of mutational fitness effects (DMFE) for the single residue IN (A) and CA (B) substitution mutants at 37°C.

fitness of each IN and CA mutant is plotted against its linear position in the protein (for residues with multiple substitutions, the smallest fitness value is selected), differences in robustness are further highlighted (Figure 32A, B). These graphical representations expose stark disparities that exist between IN and CA robustness. Integrase obviously includes significantly more high fitness mutants, including some with fitness levels above the parental virus. Perhaps even more intriguingly, the higher levels of IN fitness in the linear landscape map relatively randomly, whereas CA tends to display small, localized regions of more relative robustness (corresponding largely to the β-strand, cyclophilin binding loop, interdomain linker region, and CTD helix 10) dispersed between regions of fragility.

7.3 Structural analyses of IN fitness on an HIV-1 intasome model

As noted previously, mapping of viable and nonviable capsid mutants to the mature capsid hexameric structure revealed remarkable patterns in which viable mutants were preferentially localized to regions with greater solvent accessible surface area. To determine if analogous patterns might exist for integrase mutants, IN library viability results needed to be mapped onto an integrase model structure. Limited by the lack of crystal structure for the HIV-1 intasome, the most recently published HIV-1 intasome model, based largely on the crystal structure of the PFV intasome, was employed to model the IN mutant library fitness results [168]. First, it was necessary to establish the location of the individual IN subunits within the dimer-of-dimer IN tetramer intasome model

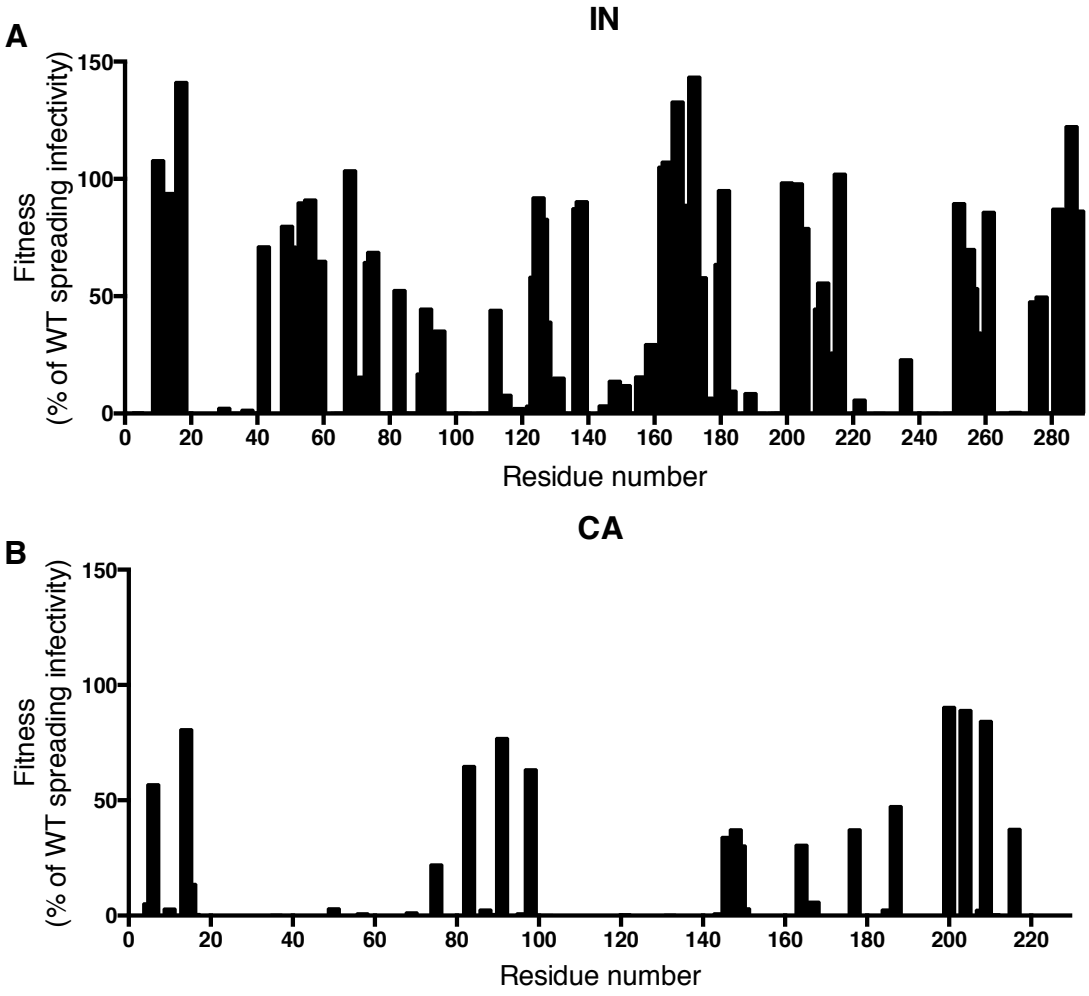


Figure 32. Comparison of linear fitness distributions throughout IN and CA sequences. (A, B) Plots with replicative fitness, as the % of WT in a spreading infectivity assay, on the Y-axes, against the position of mutation on the X-axes, organized from left (N-terminal residue) to right (C-terminal residue). For individual residues in which the IN or CA library contained more than one mutation, the smallest fitness value is plotted, so as to represent any potential fragility that could be observed in nature. For the IN mutants (A), the plot displays results shown in Tables 4 and 5, and for CA mutants (B) the plot displays results shown in Tables 1 and 2.

(Figure 33A). In this structure, unique colors indicate the individual IN subunits, while grey regions designate DNA, and magnesium and zinc molecules are hidden from view. Images, from left to right, model theoretical top, side, and underside profiles. It should be noted that in this model, like the published HIV-1 intasome models before it, the 2 outer subunits (in yellow and green) comprise only the CCD and a tiny portion of the CTD (including residues 50-230), while the inner subunits (red and blue) contain full-length IN [164,168].

Upon mapping the nonviable IN mutants onto the HIV-1 intasome model structure two remarkable trends were apparent—the first was that the nonviable IN mutants appeared to be predominantly localized to the interfaces between IN subunits, in particular the dimerization interface between the inner and outer subunits (Figure 33A, B). The second observation was that the nonviable IN mutants were generally more buried within the intasome structure (Figure 33A, B). Mapping of the viable IN mutants onto the intasome model confirmed both of these observations, as viable mutants seemed to occur predominantly in surface locations, away from subunit interfaces (Figure 33C). If the viable and nonviable IN mutants are mapped together on the same intasome structure, these trends are even more evident (Figure 33D). To quantitatively verify the observation that viable residues appeared in more surface exposed positions, measurements were made for the solvent accessible surface area for each IN residue, in each of the four monomer subunits. As Figure 33E conclusively illustrates, the nonviable IN residues, have, on average, significantly lower solvent accessible surface area

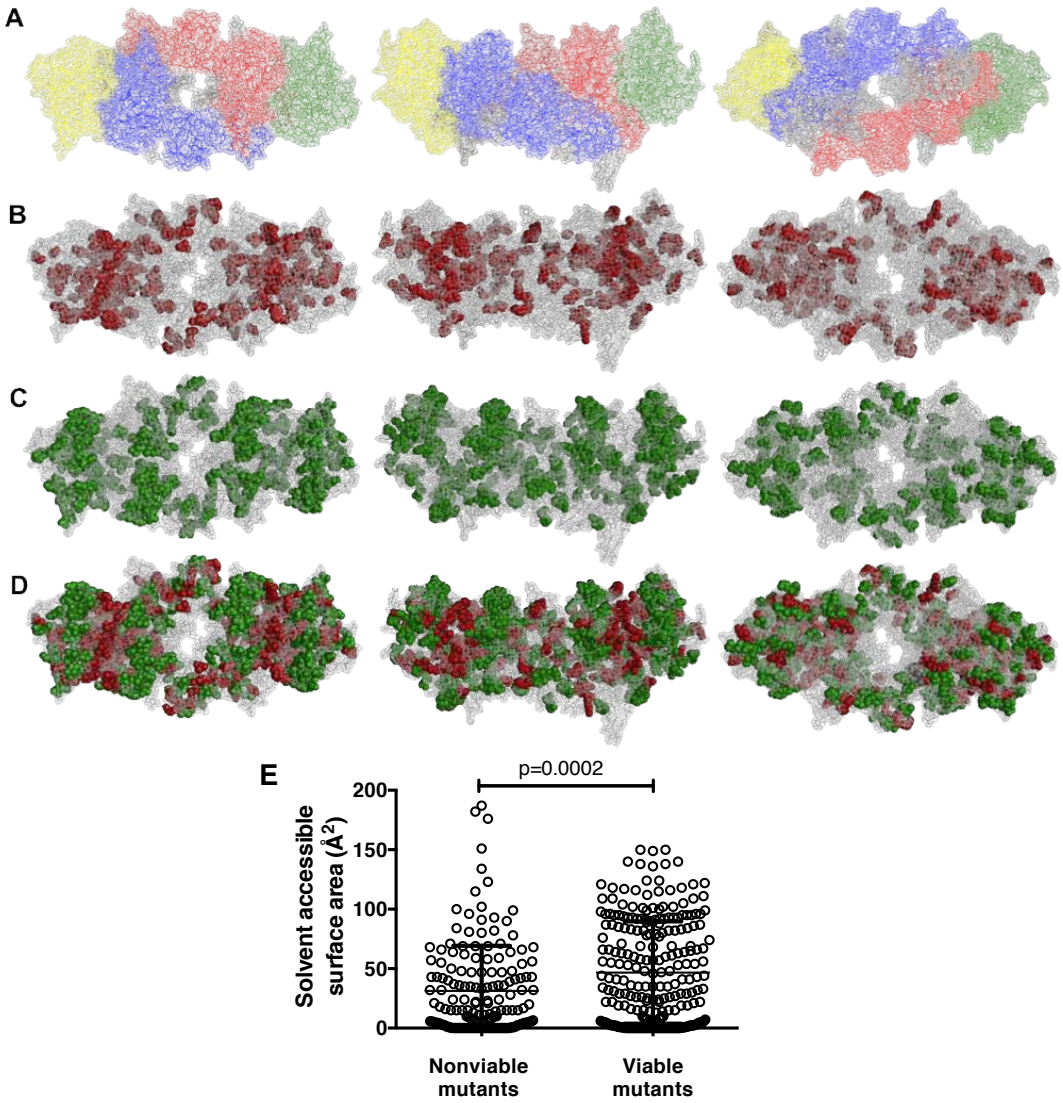


Figure 33. Mapping of IN mutant fitness on an HIV-1 intasome model. (A) Images of an HIV-1 intasome model [168] showing the individual IN subunits comprising the tetramer. Inner subunits are shown in red and blue, while outer subunits are in yellow and green. Outer subunits model only the CCD, while the inner subunits include full-length IN. Magnesium and zinc binding pockets are not visible, but DNA is shown in gray. The image on the left shows the intasome as viewed from the top, the center image displays a side profile, and the rightmost image displays a view from underneath. This intasome model will be utilized for all subsequent intasome images. (B) Locations of the nonviable (<2% of WT replicative fitness) IN mutants mapped on the intasome model. (C) Locations of the viable (>2% of WT replicative fitness) IN mutants mapped on the intasome model. (D) Locations of both viable and nonviable IN mutants mapped on the same intasome model. (E) Analysis of the correlation between residue viability and solvent accessible surface area (in $Å^2$). Mutated residues were either nonviable or viable, and corresponding solvent accessible surface area values for each IN subunit for every mutated residue were plotted, so as to represent the varying degree of surface exposure of individual residues in distinct subunits. Significance was determined using unpaired, parametric t-tests.

values than the viable IN mutants (p=0.0002). If the outer subunits of the model were to include full-length IN, it is likely this observation would only be further enhanced (given that many of the nonviable mutants are in the CCD, and if the outer subunits were to contain full-length IN, these nonviable mutants would likely be more buried by NTD and CTD arrangements, further lowering their solvent accessible surface area values).

Chapter 8. Other functional impacts of IN mutations

8.1 Characterization of IN mutant phenotypes by western blot

Measuring the *in vitro* fitness of the IN mutants was a good way of evaluating IN robustness, but there are other functional measures that could further inform its robustness *in vivo*. Using western blot analyses, it was possible to quantify the effects of all 156 IN library single residue mutants on: particle production, IN expression in particles, and clathrin incorporation into particles (Figure 34). Integrase mutants affecting all of these properties have already been reported, and it seemed prudent to investigate the extent of mutants that might exhibit such effects [186,209,223]. While the reasons for measuring particle production and IN expression in the IN mutants are more obvious, evaluation of clathrin incorporation is interesting as clathrin has previously been shown to help regulate proteolytic processing of HIV-1 virions during particle assembly, in a manner dependent on Pol, such that certain mutations in integrase and reverse transcriptase could abolish clathrin incorporation [209].

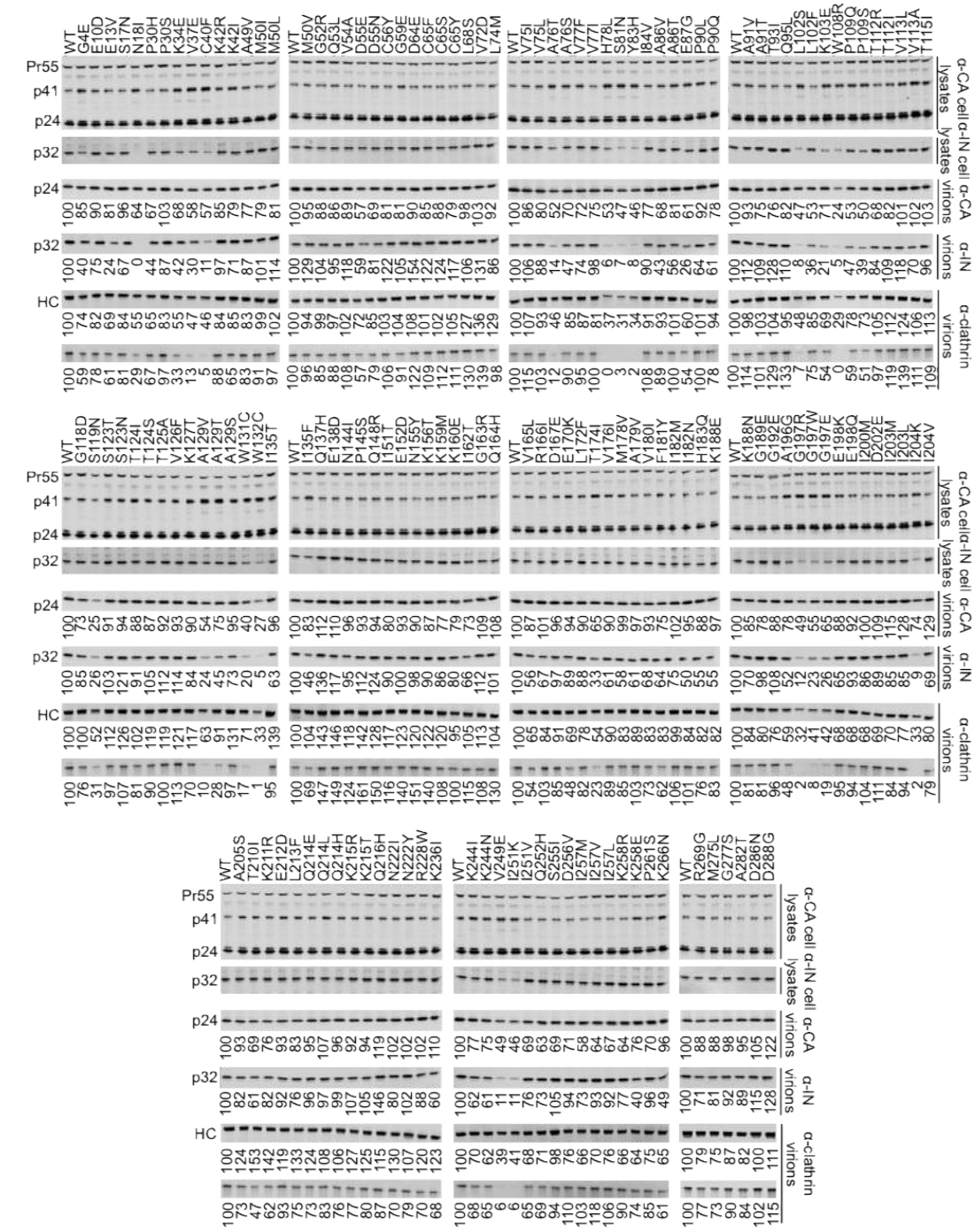


Figure 34. Analyses of the robustness of IN mutants by western blot. Western blots of virions and cell lysates generated in 293T cells transfected with WT and IN mutant plasmids, and probed with anti-CA, anti-IN, and anti-clathrin heavy chain antibodies, as indicated. Numbers shown below each band indicate the fluorescence intensities (LI-COR) associated with the WT or mutant IN protein that was pelleted from virion-containing supernatant. Virion band blots shown below the anti-clathrin heavy chain (HC) are a secondary clathrin band appearing on the same blot, shown for correspondence to virion IN expression. Mutant N18I likely lost the epitope for the anti-IN antibody, so this mutant was not included in Table 6 conclusions. Phenotypes were considered reduced from WT if expression levels dropped by at least 2-fold.

Table 6 shows the summaries of the evaluated phenotypes, as observed by domain.

To begin more in-depth examination of these blots, it should be noted that cell-associated Gag expression in both the WT and all IN mutants revealed consistent levels of both capsid p24 and Gag precursor Pr55, ensuring transfection efficiencies should be equivalent, and giving confidence to the subsequent analyses. More interesting is the observation that, consistent with the *in vitro* infectivity fitness measurements, only a minority of IN mutants appeared to have deficiencies by western blot. In total, 34 (22%) mutants had at least 2-fold reductions in IN expression in virions, and just 13 (8%) had a more stringent 5-fold reduction (mutant N18I was not included in these calculations, as it seemed to have lost the epitope for the IN antibody). Notably, a larger fraction, 46%, of mutants in the NTD appeared to have substantial reductions in IN expression in virions than in the CCD (23%), while a far smaller proportion had IN expression reductions in the CTD (9%) (Table 6). Additionally, mutants with decreases in IN expression often occurred in clusters in the CCD (residues 78-83, 102-109, 129-132), suggesting certain, smaller regions of IN are particularly vulnerable in this regard. Not surprisingly, the mutants with reductions in virion IN expression also had decreases in cell-associated IN expression as well. Similarly, there was a large degree of concurrence between the mutants with reductions in IN expression in virions, and those with diminished clathrin incorporation. While not all mutants with IN expression reductions had clathrin

incorporation reductions in virions, all mutants displaying reductions in clathrin incorporation also had reductions in IN expression in virions. This was true for the traditional heavy chain clathrin band denoted HC (Figure 34), while a smaller clathrin band on the same blot often indicated comparatively lower levels of clathrin expression (Figure 34, below HC). Still, there were no viable IN mutants with at least 2-fold reductions in virion clathrin incorporation, so there was no further investigation of the clathrin incorporation phenotype.

Very few IN library mutants appeared to display reductions in particle production. Only 11 mutants (7%) had at least 2-fold decreases in virion-associated p24, and none had the more stringent 5-fold reduction in p24. Interestingly, none of the mutants in the NTD had considerable reductions in particle production, nor were any affected by loss of infectiousness only, suggesting, not entirely surprisingly given past research, that this region is unlikely to contain many class II IN mutants. Inspection of the mutants with significant (>2-fold) reductions in particle production also revealed that all of these mutants also had reductions in virion IN expression (Figure 34), but not all mutants with reductions in IN expression had reductions in particle production. Also intriguing is the observation that, of the 55 nonviable single amino acid IN mutants, 24 (44%) had no significant effect (< 2-fold) on either IN expression in virions or particle production. Instead, these nonviable mutants impacted infectiousness only.

8.2 Mapping of IN mutations with functional disruptions on an HIV-1 intasome model

Given how informative the mapping of IN mutant viability on the intasome model had proved, it seemed worthwhile to investigate whether more structural trends could be illuminated by mapping the functional IN phenotypes observed by western blot onto the intasome model. Strikingly, mapping of the nonviable mutants with at least 2-fold reductions in particle production exposed a rather narrow and distinct pattern in which most mutants displaying this phenotype congregated at the dimerization interface between the inner and outer subunits (Figure 35A). This localization is particularly apparent when compared to the mapping of the nonviable mutants without IN expression or particle production reductions, the so-called 'infectiousness only' nonviable mutants, which were scattered throughout the intasome in a relatively buried fashion (Figure 35B). Equally remarkable was the mapping of the nonviable IN mutants with at least 2-fold reductions in virion IN expression, in which an even greater number of mutants are localized to the same IN subunit interface region (Figure 36A), which is again especially clear when compared to the mapping of the nonviable 'infectiousness only' mutants (Figure 36B). The degree to which these nonviable mutants that had no reductions in either particle production or IN expression in virions occurred in buried locations in the intasome model was significant when compared to nonviable mutants with reduced IN incorporation ($p=0.04$) (Figure 36C), emphasizing that the 'infectiousness only' mutants are even more buried

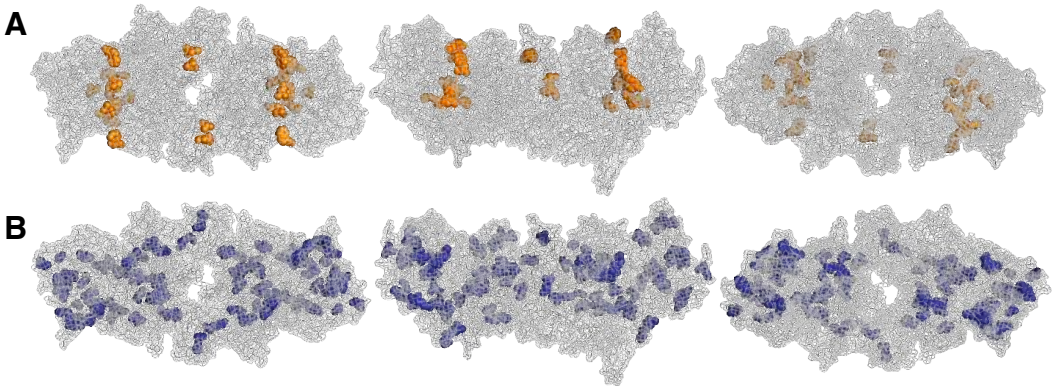


Figure 35. Localization of nonviable IN mutants affecting particle production in the intasome. (A) Intasome model images showing the locations of the nonviable IN mutants shown by western blot to have at least a 2-fold reduction in particle production, as measured by levels of expression of p24 in virions. (B) Intasome model images showing the locations of nonviable IN mutants for which western blot analyses indicated no significant reduction in either particle production (as measured by p24 expression in virions) or IN expression in virions. These nonviable mutants were affected only by a loss of infectiousness.

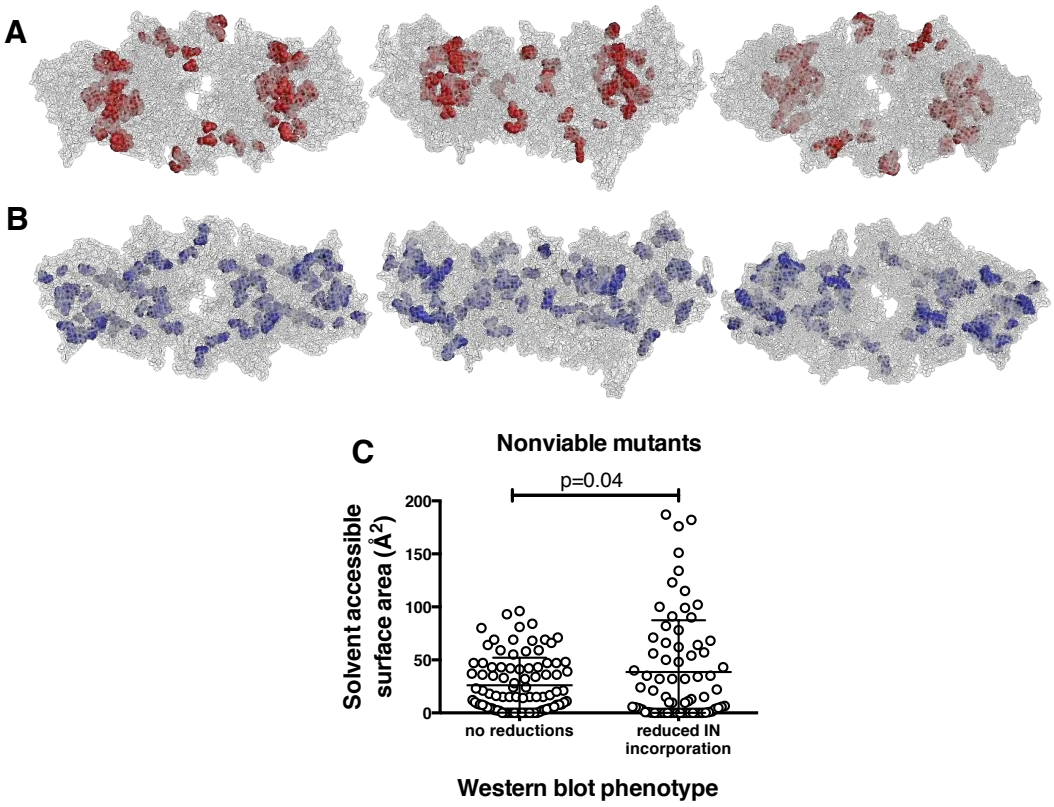


Figure 36. Nonviable IN mutants disrupting IN expression in virions are also highly localized to IN subunit interfaces. (A) Intasome model images showing the locations of the nonviable IN mutants shown by western blot to have at least a 2-fold reduction in IN expression in virions. (B) Intasome model images showing the locations of nonviable mutants for which western blot analyses indicated no significant reduction in either particle production (as measured by p24 expression in virions) or IN expression in virions. These nonviable mutants were affected only by a loss of infectiousness. (C) Comparison of the solvent accessible surface area values (in $Å^2$) for nonviable IN mutants that result in either a reduction in IN expression in virions, or no reductions in either virion IN expression or particle production (those mutant residues mapped in (B)). All available subunit values were plotted for the relevant residues in each group. Significance was determined using unpaired, parametric t-tests.

than other nonviable IN mutants. Together, the mapping on these intasome structures highlights surprising and novel findings indicating that nonviable IN mutations causing reductions in particle production and IN expression in virions are highly localized to certain IN subunit interfaces.

Chapter 9. Analyses of IN library mutations in natural HIV-1 subtype B populations

9.1 Occurrence of library IN mutations in an *in vivo* cohort and comparison with CA

As earlier demonstrated, the fragility of HIV-1 CA places tight constraints on the sequences of CA that naturally occur. However, the relative robustness of HIV-1 IN indicates that IN may have more available sequence space to exploit. By utilizing the same cohorts of 1,000 HIV-1 subtype B IN and CA sequences used to evaluate the variability in Figure 28, the *in vivo* occurrence of the library IN and CA single amino acid mutants could be determined, and contrasted. Plots of the frequency with which the IN and CA library mutations naturally occurred against the replicative fitness of the same mutants, revealed a number of similarities between the behavior of IN and CA mutants *in vivo* (Figure 37A, B). Remarkably, in both IN and CA plots, library mutants required the same amount of replicative fitness, at least 40% of WT, in order to occur at a frequency >1% in natural populations (Figure 37A, B). Also notable was the observation that for both IN and CA mutants with replicative fitness >40% of WT, more mutants never

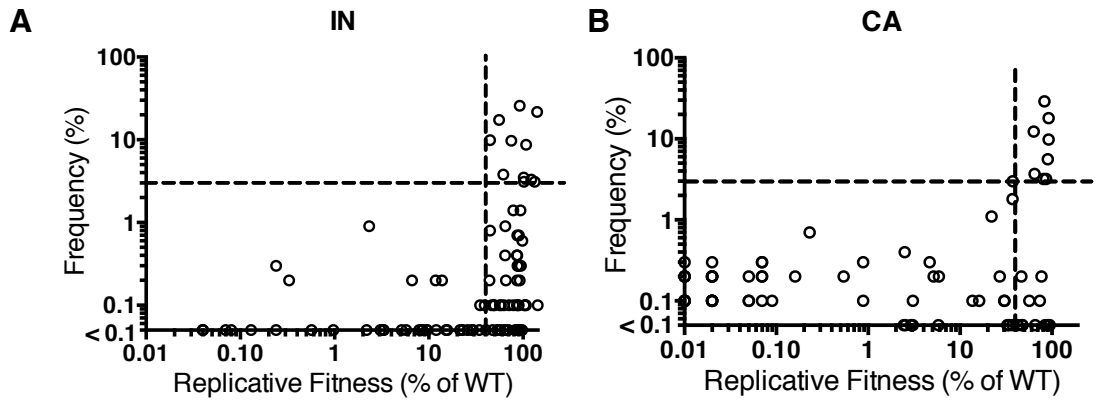


Figure 37. Occurrence of library IN and CA mutations in natural HIV-1 subtype B populations. (A, B) Plots indicating the frequencies with which IN (A) and CA (B) library single residue mutations occur in 1000 HIV-1 subtype B isolate IN and CA sequences (Y-axes) versus the replicative fitness of the same mutations (X-axes). The horizontal dashed lines designate a frequency of 3%, above which mutations were considered to be frequently occurring. Vertical dashed lines indicate 40% of WT replicative fitness, below which mutants typically occurred rarely in isolate sequences. Mutants that had fitness values below 0.01% of WT, and that did not occur in 1000 isolate sequences, are not plotted.

or rarely (frequency <0.3%) appeared *in vivo* than occurred at a high frequency (>3%). Table 7 lists the 11 fit IN mutants that occur frequently, and the longer list of 22 IN mutants that are equally fit, yet never occur in the 1,000 naturally occurring IN isolate sequences, along with their corresponding fitness values. Again, these results are surprising given the generally held notion that HIV-1 will take advantage of all available sequence space that does not incur significant fitness costs, and indicates that additional selective pressures may be at play for both IN and CA.

9.2 Mapping of correlations between fitness, *in vivo* frequency, and variability on an HIV-1 intasome model

To explore whether the location of the fit (>40% WT replicative fitness) IN mutant residues correlated with the frequency with which they occur *in vivo*, intasome mapping was used to compare the locations of the subsets of 'fit and frequent' (>40% WT replicative fitness *in vitro* and >3% frequency *in vivo*) and 'fit and never' (>40% WT replicative fitness *in vitro* and 0% frequency *in vivo*) mutants. Fit and frequently naturally occurring IN mutants appeared to be primarily localized in quite surface exposed locations throughout the intasome (Figure 38A). Alternatively, the fit and never naturally occurring mutants were typically located in more buried positions throughout the intasome (Figure 38B). These observations were verified quantitatively by plotting the solvent accessible surface area values for each IN subunit residue for the fit 'frequent' and 'never' residues, and the results confirm that the fit and never naturally occurring IN

Table 7. Opposing frequencies of fit IN mutants *in vivo*

Fit IN mutants occurring frequently *in vivo* [a]	Spreading replication infectivity (% of WT) [b]	Frequency in subtype B isolates (%) [c]	Fit IN mutants never occurring *in vivo* [d]	Spreading replication infectivity (% of WT) [b]
E10D	107.9	8.7	E13V	93.9
S17N	141.2	21.7	K42I	71.2
M50I	75.2	9.7	A49V	79.9
A91T	44.6	9.9	C65S	60.6
T112I	61.8	3.8	V75I	73.4
T125A	92.0	25.7	V75L	68.8
D167E	132.9	3.1	Y83H	52.5
I203M	101.5	3.5	P90Q	66.4
K211R	55.7	17.4	S123N	40.4
Q216H	102.1	3.1	V126F	82.9
D286N	122.4	3.3	T174I	57.9
			V180I	63.6
			E198Q	91.5
			I203L	98.0
			Q214H	54.8
			N222Y	93.8
			Q252H	89.6
			S2551	70.1
			D256V	53.3
			I257L	65.3
			P261S	85.8
			M275L	47.7
			A282T	87.2

[a] Fit mutants occurring frequently are those single residue mutants with fitness >40% of WT that are found in >3% of 1000 IN subtype B isolate sequences.
[b] Spreading replication infectivity is as described in Table 4.
[c] Frequency in subtype B isolates is as described in Figure 37.
[d] Fit mutants that never occur are mutations that were not found in any of the 1000 subtype B IN sequences.

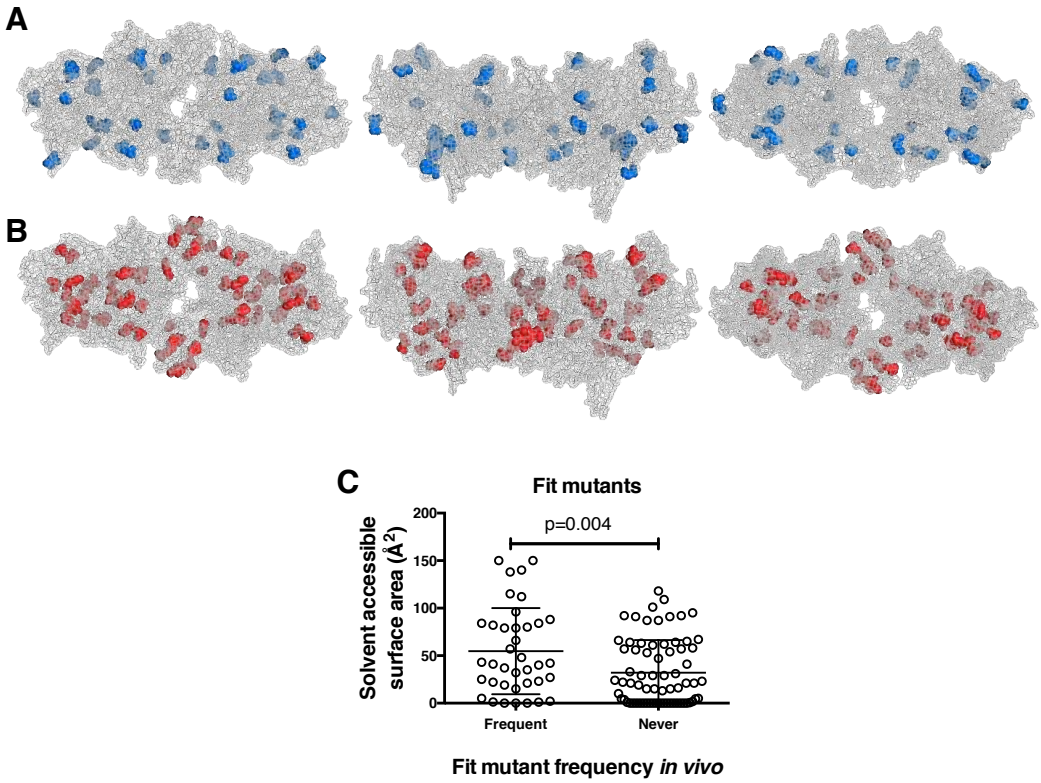


Figure 38. Residue surface exposure is a determinant of fit IN mutant frequency *in vivo*. (A) Intasome model images showing the locations of the fit (>40% of WT fitness) IN library mutations that occurred frequently (>3%) in 1000 HIV-1 subtype B IN isolate sequences. (B) Intasome model images showing the locations of the fit (>40% of WT fitness) IN library mutations that never occurred in 1000 HIV-1 subtype B IN isolate sequences. (C) Comparison of solvent accessible surface area values (in $Å^2$) for the fit IN library mutants occurring either frequently (as shown in A) or never (as shown in B). All available IN subunit solvent accessible surface area values were plotted for the individual 'frequent' or 'never' occurring fit mutants. Significance was determined using unpaired, parametric t-tests.

residues are significantly more buried (p=0.004) (Figure 38C). While the degree of surface exposure may not completely explain why some fit IN mutants seem to not occur *in vivo*, it appears to be a significantly contributing factor. It is not entirely clear why more buried positions would be less tolerant to mutation *in vivo* in the absence of significant fitness effects. It is possible that fit surface residues are more variable because their exposure makes them more vulnerable to unknown pressures, possibly of immune origin. However, it may also be true that relatively small differences in fitness may still impact the frequency of certain mutations *in vivo*, as the average replicative fitness of the fit mutants that occurred frequently *in vivo* was 94.3% of WT, but the average replicative fitness of the fit mutants that never occurred *in vivo* was just 71.7% of WT.

Having established that the degree of surface exposure contributed to the naturally occurring frequency of IN library mutations, there was good reason to investigate whether the degree of surface exposure of IN residues might more generally explain why some residues are more variable than others *in vivo*. Again utilizing the cohort of 1,000 naturally occurring IN sequences, as well as the variability results found in Figure 28, residues for which isolate sequences varied from the parental sequence in more than 10% of sequences (variability >10%) were determined and mapped onto the intasome model (Figure 39A). Whereas these highly variable residues seemed to be preferentially localized to exposed locations, residues for which there was no variability observed appeared more buried (Figure 39A, B). Measurements of solvent accessible surface area

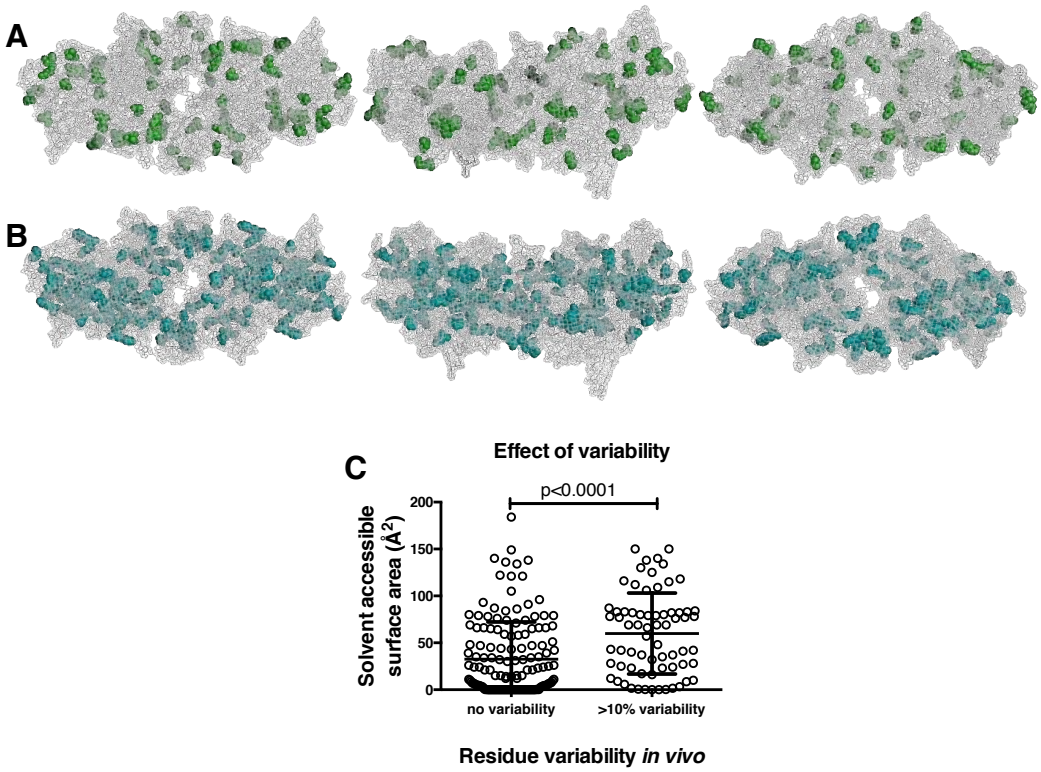


Figure 39. HIV-1 IN natural variability is highly correlated with residue surface exposure. (A) Intasome model images showing the location of the residues in IN with >10% variability (as analyzed in 1000 HIV-1 IN subtype B isolate sequences, and indicated in Figure 28). (B) Intasome model images showing the location of the residues in IN with no variants (as analyzed in 1000 HIV-1 IN subtype B isolate sequences, and indicated in Figure 28). (C) Comparison of solvent accessible surface area values (in $Å^2$) for the IN residues that were either highly variable (shown in A) or not variable at all (shown in B). All available IN subunit solvent accessible surface area values were plotted for the relevant '>10% variability' or 'no variability' residues. Significance was determined using unpaired, parametric t-tests.

again confirmed the observations, as residues with no variability had significantly lower values ($p<0.0001$) (Figure 39C). Ultimately, within this HIV-1 intasome model, the degree of IN residue surface exposure seems to be a remarkably good indicator of both how fit a mutant is, and of how likely it is to occur *in vivo*.

9.3 Summary of effects of IN mutations and comparison with effects of CA mutations

Having evaluated the robustness of HIV-1 IN both *in vitro* and in naturally occurring populations, in a fashion very similar to what was done for HIV-1 CA, it was possible to compare generalized predictions of the fate of randomly introduced nucleotide and amino acid substitutions into the two proteins, to better emphasize the stark differences in robustness. For random nucleotide substitutions, an estimated 18% of IN substitutions and 22% of CA substitutions should result in synonymous changes, which should not, in most cases, significantly impact fitness. The difference in the proportion of nucleotide substitutions yielding synonymous changes in IN and CA highlights how the composition of a protein can impact the outcome of any substitutions. In the absence of any negative selection pressure other than fitness, then 43% of IN nucleotide substitutions, and a far smaller 10% of CA nucleotide substitutions, would be expected to result in nonsynonymous changes sufficiently replication competent (>40% of *in vitro* WT fitness) to, at least theoretically, thrive *in vivo*. Yet data from evaluations of the natural occurrence of the fit mutations with >40% of WT fitness (Figure 37) revealed that only a minority of these mutations actually

persist *in vivo* in both data sets, such that 12% of IN nucleotide substitutions, but only 4% of CA nucleotide substitutions would actually be expected to flourish in natural populations (Figure 40).

When only the nonsynonymous nucleotide substitutions are evaluated, then 40% of IN substitutions and 87% of CA substitutions would be expected to be nonviable, with only a few mutations in IN (5%), but most in CA (59%), inducing significant reductions in particle production (Figure 40). This reinforces the idea that the basis for the genetic fragility of CA is the need to efficiently and correctly assemble particles, and that this requirement may be largely unique to CA. While 60% of IN amino acid substitutions are sufficiently replication competent (at least 40% of WT fitness) to potentially thrive *in vivo*, the same is true for just 13% of CA amino acid substitutions, based on *in vitro* fitness measurements. However, these numbers are significantly reduced by the finding that only a fraction of the intrinsically fit mutants actually persist *in vivo* (Figure 37), such that 15% of IN amino acid substitutions and only 5% of CA amino acid substitutions could be expected to thrive in a natural environment (Figure 40).

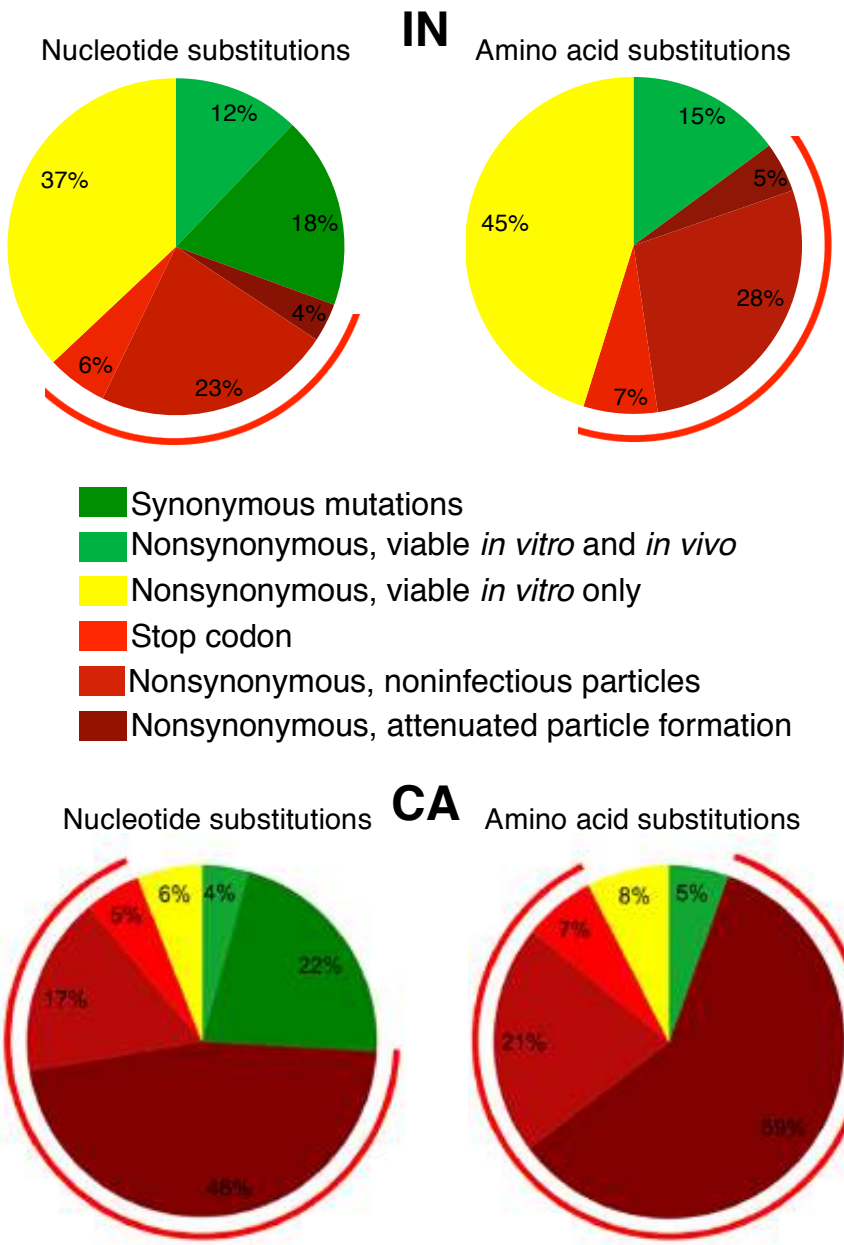


Figure 40. Diverse fates for random mutations in HIV-1 IN and CA. Comparison of various likely outcomes for random nucleotide or amino acid substitutions in IN and CA, as indicated. Charts show changes that are synonymous (nucleotide substitutions only), nonsynonymous, and nonsense (as determined by using the pNHGintBS or pNHGcapNM sequences). Substitutions that resulted in a stop codon were assumed to be nonviable while synonymous changes were assumed to be viable. Given the observation that only mutants with >40% of WT *in vitro* fitness are observed in nature, this threshold was used as a sufficient determinant of mutants that would thrive *in vitro* (proportions of mutants with <40% WT fitness were determined using the viable and nonviable IN and CA mutants in Tables 1, 2, 4, and 5 and are denoted by red circumferential lines). Ratios of fit mutants expected to thrive *in vivo* (nonsynonymous, viable *in vitro* and *in vivo*), or expected not to thrive (nonsynonymous viable *in vitro* only) were derived from data in Figures 22 and 37. Fates of remaining mutants were determined using western blot data from Figures 19 and 34 (nonsynonymous, noninfectious particles or nonsynonymous attenuated particle formation (at least 5-fold deficit for CA and at least 2-fold deficit for IN)).

IV. Discussion

As the circumstances contributing to the fragility of CA and the relative robustness of IN are likely to differ significantly, the situation of CA will first be addressed independently, followed by analyses of the comparative robustness of IN. Thereafter, the implications about the genetic robustness and fragility of HIV-1 proteins will be briefly discussed and given perspective.

The extreme genetic fragility of HIV-1 CA

This study has sought to create a library of random CA mutations representative of those that occur naturally during HIV-1 infection, and to examine their biological effects. In so doing, it has been possible to determine the genetic robustness of HIV-1 CA, and to correlate the effects of amino acid substitutions *in vitro* with their occurrence in natural viral populations. Furthermore, the creation of an extensive CA library constitutes an important resource for investigating various functions and properties of HIV-1 CA, including CA's emerging impact on Mx2 retroviral restriction. Importantly, the results have demonstrated the extreme genetic fragility of the HIV-1 CA protein, in which the majority (70%) of unbiased, single amino acid substitutions result in nonviable viruses (<2% of WT fitness).

In fact, after surveying the literature for all studies that have conducted mutagenesis in such a way that mutational robustness could be determined, HIV-1 CA appears to be the most fragile protein ever described. As Table 8 indicates, capsid is significantly more fragile than other HIV proteins, as in the next most

Table 8. Comparison of the robustness/fragility of HIV sequences*

Protein/Domain	Fragility[#]	Author	Summary/Notes
HIV-1 capsid	**70%**	**Rihn *et al.***	**This study.**
HIV-1 protease flap region (aa 46-56)	61%	Shao *et al*. [224]	80 out of 131 mutants had negative activity, as shown by lack of Western blot p66/p51 bands.
HIV-1 protease	54%	Parera *et al.* [91]	A total 86% of 107 single mutants were deleterious to catalytic efficiency, 54% lethal.
HIV-1 protease	40%	Loeb *et al*. [225]	Out of 330 single missense mutants, around 40% were considered sensitive to mutation.
HIV-1 integrase	35%	**Rihn *et al.***	**This study.**
HIV-1 reverse transcriptase palm subdomain (aa 164-203)	28%	Chao *et al*. [226]	From 154 single mutants, 43 had RT levels that were ≤ 2% of WT.
HIV-2 integrase	13%	van den Ent *et al*. [227]	5 out of 38 mutants display severely attenuated integration.
HIV-1 reverse transcriptase motif B	12%	Smith *et al*. [228]	Using infectivity results, 3 out of 25 mutants had ≤2% of WT levels.
HIV-1 transcriptional promoter (LTR)	0%	van Opijnen *et al*. [90]	None of the 15 single mutants diminished infectivity as much as 2 fold in any cell line.
HIV-1 reverse transcriptase (aa 67-78 of beta3-beta4 loop)	0%	Kim *et al*. [229]	Fragility measured by genetic selection in E. coli, in which no WT aa were essential.

*Table describes single residue mutagenesis studies that were (i) random (i.e. the cohort of mutants was not selected, and did not have amino acid bias) (ii) analyzed proteins or domains greater than 10 aa in length, and (iii) had datasets that included at least 5 mutants. [#]Fragility, when not stated clearly in the published work, was deduced using the % of mutants either indicated as 'lethal' or that had 2% or less fitness or function compared to WT, using the available analyses of fitness or function utilized in the study. In most cases, wording of summary/notes mimics the author's assessment as closely as possible.

fragile protein, HIV-1 protease, only 40-54% of mutations resulted in nonviability [91,225]. HIV-1 CA is also more fragile than the small domains of HIV mutagenized because of their suspected severe fragility, such as the HIV-1 protease flap region or the HIV-1 reverse transcriptase palm domain [224,226]. Likewise, Table 9 demonstrates that HIV-1 CA is also more fragile than proteins or domains in viruses other than HIV, with human papilloma virus 16 protein E1, in which 63% of mutations were lethal, coming the closest to the fragility of capsid (although the study's extremely small mutant sample size may overestimate its fragility) [230]. It is perhaps not surprising then, that CA is also more fragile than the whole viral genomes in which robustness has been evaluated (Table 10), in which 20-41% of mutations were lethal [87,93]. Viruses (particularly ss(+) RNA viruses) not only have very high mutation rates, but also have comparatively low tolerance to mutation [79,83], so it is again perhaps not unexpected that CA is more fragile even outside the viral context, as the most fragile nonviral protein described is the essential E. coli lac repressor, in which 41% of residues were estimated to be intolerant of substitution [231] (Table 11, which is not inclusive of all nonviral random mutagenesis studies, but none were found that were more fragile).

While the biological basis of the extreme mutational fragility of the HIV-1 CA appears to be the need to coordinate accurate and efficient assembly of particles (Figure 16-21), it is not immediately apparent why so many CA mutations result in reduced particle yields. However, it seems likely that CA's

Table 9. Comparison of the robustness/fragility of other viral sequences*

Protein/Domain	Fragility#	Author	Summary/Notes
Human Papilloma Virus 16 E1	63%	Yasugi *et al*. [230]	5/8 single mutants were unable to replicate.
Simian immunodeficiency virus mac239 envelope constant region 4	58%	Morrison *et al*. [232]	Only 11/26 single mutants were able to replicate appreciably.
H3N2 influenza A hemaglutinin	56%	Nakajima *et al*. [233]	44% of 248 single aa changes did not abrogate hemadsorption activity.
H2N2 influenza A neuraminidase	53%	Yano *et al.* [234]	47% of 152 single mutations in NA were tolerated, certain regions were more sensitive.
Bacteriophage λ Cro	53%	Pakula *et al*. [235]	Approximately 37 out of 70 single residue mutants have ≤2% of WT CRM levels.
Mason-Pfizer monkey virus matrix	50%	Rhee *et al*. [236]	4/8 single mutants had significant cell-free infectivity.
Moloney murine leukemia virus capsid central domain	40%	Alin *et al.* [237]	Virus replication was considered blocked based on RT levels for 6 of 15 single mutants.
Bacteriophage f1 gene V protein	35%	Terwilliger *et al.* [238] Masso *et al.* [239]	139 out of 371 single point mutants were inactive (unable to inhibit E. coli growth).
Varicella-Zoster virus thymidine kinase	35%	Suzutani *et al*. [240]	Out of 17 single aa mutants, 6 resulted in a loss of enzyme activity.
Bunyavirus nucleocapsid (N)	4-26%	Eifan *et al*. [241]	5 /115 single mutants had ≤2% WT mini replicon activity, 30/57 tested mutants had titers ≤2% of WT.
Wheat streak mosaic virus HC-Pro	24%	Stenger *et al*. [242]	6 out of 25 nonsynonymous substitutions abolished systemic infectivity.
Bacteriophage T4 lysozyme	16%	Rennell *et al*. [243]	From 2015 single aa substitutions, 328 were deleterious enough to prevent plaque formation.
Mason-Pfizer monkey virus capsid MHR	15%	Strambio-de-Castillia *et al*. [244]	2/13 single mutants have RT levels at ≤2% of WT, all other mutants maintain at least 50%.
Adenovirus VI	9%	Moyer *et al*. [245]	Only 1 out of 11 random single mutants led to a greater than 10 fold reduction in infectivity.
Encephalomyocarditis virus internal ribosome entry site	0%	Van Der Velden *et al.* [246]	Zero out of 13 single random mutants had significantly impaired *in vitro* translation.

*, # Criteria are the same as in Table 8.

Table 10. Comparison of the robustness/fragility of viral genomes*

Virus	Fragility#	Author	Summary/Notes
Tobacco etch potyvirus (whole genome)	41%	Carrasco *et al.* [93]	Out of 66 single mutants, 41% were lethal and an additional 36% were deleterious
Vesicular stomatitis virus (whole genome)	40%	Sanjuán *et al*. [86]	Fitness analysis of 91 single mutants showed a further 29% of mutants were deleterious
Bacteriophage Qβ	29%	Domingo-Calap *et al*. [87]	29% of single mutations in 42 clones of Qβ were considered lethal
Bacteriophage f1 (whole genome)	24%	Peris *et al*. [247]	24% of a total 51 single missense mutants were lethal. Gene lethality varied 7 to 47%
Bacteriophage ΦX174	20%	Domingo-Calap *et al*. [87]	20% of single mutations in 45 clones of ΦX174 were considered lethal

*, # Criteria are the same as in Table 8.

Table 11. Robustness/fragility measurements in nonviral proteins*

ORF/Protein/Region	Fragility#	Author	Summary/Notes
E. coli lac repressor	<41%	Markiewicz *et al*. [231]	From 4000 single mutants, authors conclude 192 of 328 aa should be tolerant to substitution
3-methyladenine DNA glycosylase (AAG)	34%	Guo *et al*. [248]	The calculated probability of a single random amino acid change inactivating AAG is 34%
Campylobacter jejuni oligosaccaryltransferase PglB	33%	Ihssen *et al*. [249]	16 of the 24 single mutants maintain strong PglB expression by ELISA
Human interleukin-3	29%	Olins *et al*. [250]	From analysis of 770 mutants, 71% of residues were considered tolerant, as they had ≥5% of WT bioactivity
Bacillus stearothermophilus α-amylase	20%	Holm *et al*. [215]	12 out of 61 single random mutants had zero enzyme activity
TEM-1 beta-lactamase	16%	Huang *et al*. [251]	Authors conclude that 43 of the 263 residues cannot tolerate substitutions
T15 antibody heavy chain CDR2	7%	Chen *et al*. [252]	Just 1 out of 15 single CDR2 mutants lost binding function
Barnase, a bacterial ribonuclease	5%	Axe *et al*. [253]	5% (of over 600 mutants) were found to render barnase wholly inactive
Human interleukin-1α	2%	Kawashima *et al.* [254]	Only 1 out of 56 single mutants appeared to have significantly impaired activity

*, # Criteria are the same as in Table 8.

fragility is related to the requirements that it maintain a structure that enables several different sets of interactions. Within the mature hexameric CA lattice, each CA monomer must interact, via distinct interfaces, with at least three other CA monomers, and adopt a quite tightly constrained position with respect to its internal structure as well as other CA monomers within a hexamer [123]. At the same time, some CA monomers, that have precisely the same amino acid sequence, must be capable of assuming slightly different structures and positions relative to other CA monomers, in order to form the pentameric declinations that permit closure of the CA lattice [111,117]. Furthermore, another set of CA:CA interactions are likely required during the formation of immature virions [110]. Beyond these constraints, the strength of the various CA:CA interactions in each of the different configurations probably needs to finely tuned to allow proper transitions from immature to mature virions during assembly, and appropriately timed disassembly of the capsid. Indeed, earlier findings have suggested that proper capsid stability is required for HIV-1 infectiousness [255]. Additionally, CA interacts with numerous host proteins, including cyclophilin A [120], ABCE1 [119], and probably with CPSF6 [132] and nuclear pore complex proteins [131]. Together, the number of CA interactions and functions, and the requirement for CA to adopt multiple different structures, and transition between them in an organized manner at multiple, critical stages of replication, likely places tight constraints on sequence and decreases the number of possible mutations that will not result in major fitness costs.

While the lethal CA mutations could, in principle, affect any one of multiple steps in the HIV-1 life cycle, it appeared that most constitutively nonviable mutants exhibited attenuated particle formation. Moreover, all of the conditionally nonviable ts mutants displayed attenuated particle formation at the nonpermissive temperature. Interestingly, these deficits, at least for the conditionally nonviable mutants, could be rescued by inactivation of the viral protease. Ultimately, these results indicate that the critical role of CA in coordinating efficient and accurate particle assembly, particularly during Gag proteolysis, is the principal cause of its genetic fragility.

The processes of particle formation, budding and maturation are not yet fully understood, but over a relatively short period of time (5-6 minutes) [256], thousands of Gag and Gag-Pol molecules coalesce, in a manner requiring cleavage and rearrangement through regimented and ordered processes. Previous work has shown that incorrect, premature activation of the HIV-1 protease can prevent particle formation [257]. Moreover, other perturbations of Gag (such as a partial nucleocapsid deletion) can attenuate particle formation in a way that is rescued by inactivation or inhibition of the viral protease [258,259]. One idea that may explain these findings is that delayed particle formation may result in more extensive, or mistimed, proteolytic cleavage of Gag early in the assembly process, thereby attenuating particle formation [258,259]. Our finding that many of the constitutively nonviable and all of the conditionally nonviable CA mutants showed evidence of disrupted Gag processing, through the presence of

additional ~40 kDa protein species, is in accordance with the idea that CA mutations could similarly delay particle formation, resulting in mistimed Gag proteolysis relative to budding, and, thereby, protease dependent reductions in particle yields. Within this framework, it is easier to understand how even subtle perturbations of Gag structure or multimerization, such as those caused by single amino acid substitutions, could disrupt the delicate process of HIV-1 assembly, and thus how critical a role CA plays in orchestrating virion morphogenesis.

Earlier work has shown that the N-terminal domain of HIV-1 CA is dispensable for particle formation [260], which is intuitively at odds with our finding that many mutations in the N-terminal domain result in diminished particle production. However, it has already been hypothesized that misfolding of the N-terminus (caused by small changes in CA) may be more disruptive to assembly processes than deletion [260]. Even without overt misfolding, inclusion of a CA NTD that does not fit correctly or efficiently into the hexameric lattice is likely to be more disruptive than NTD absence to the other interactions that are essential for particle formation (such as the CTD-CTD interactions).

Previous large-scale mutagenesis studies of retroviral CA proteins have utilized insertional mutagenesis of murine leukemia virus (MLV) CA (using 12 amino acid peptide sequences) [261,262] and alanine substitution of surface exposed residues in HIV-1 CA [155]. The fact that the 12 amino acid insertions into MLV CA were uniformly lethal is consistent with the notion that retroviral CA proteins are generally intolerant of mutation, but insertions are obviously more

likely to disrupt CA structure than single amino acid substitutions. In fact, a study of MLV CA mutants revealed that only 6 of 15 single amino acid substitutions (40%) in the central portion of CA were lethal [237]. Interestingly, this study found that the most frequent phenotype associated with single amino acid substitutions in MLV CA was the generation of noninfectious particles, instead of reduced particle formation, which suggests there may be a difference in the properties of MLV and HIV-1 CA proteins. Prior to our study of HIV-1 CA, the most extensive mutagenesis study of HIV-1 CA [155] described 48 alanine-scanning mutants that targeted surface exposed residues, of which 44% were nonviable. This study also found that single mutations in both the NTD and CTD could cause attenuated or aberrant particle formation, as well as the generation of virions that appear morphologically normal, but were noninfectious [155]. The difference between the fraction of mutations that were lethal in [155] versus our study is in accordance with our finding that viable CA mutants tended to be confined to surface-exposed residues while mutations of the inner core of the CA structure were rarely associated with viability.

Most notably in relationship to the CA structure, we found that sequences encoding CA helices, particularly those in the NTD, were especially genetically fragile. These sequences form the basic structure of the CA hexamer, and in some cases (helices 1, 2, and 3), are involved in the CA:CA interactions within the CA hexamers [123]. Mutations in the CTD helices, and in the interhelical loops (such as the cyclophilin A binding loop) were less likely to cause lethal

defects. This comparative tolerance of mutations in certain portions of CA seems to correlate, at least partially, with conformational variation in crystal structures of CA hexamers [123]. In short, the CA NTD helices appear to be highly structurally constrained and particularly sensitive to mutations that might disrupt the shape or stability of the CA NTD monomer or the CA hexamer.

By correlating our CA library fitness measurements with analyses of the frequency with which the mutations occur in natural populations, we hoped to make observations that neither approach alone could reveal, and also make predictions about the outcomes of hypothetical mutations. For example, all of the mutations that were present in our library at a frequency greater than 3% in natural populations (n=9) were relatively fit *in vitro* (>40% of WT fitness). Conversely, all the mutants with fitness <20% of WT (n=106) occurred at a frequency of <1% in natural populations. Together, these findings indicate that there is indeed some correlation between fitness and occurrence in natural populations, and suggests that below ~40% fitness threshold, mutations are unlikely to confer a selective advantage in a human host. Nevertheless, the correlation between mutant fitness and natural occurrence was not perfect, indicating that there are likely selective pressures other than fitness *in vivo*. Specifically, there was not complete concordance between fitness and Shannon entropy profiles in some portions of the CA sequence, and it was striking that less than half of the mutations that had little or no impact on fitness actually occurred in natural populations ('fit but rare' mutants). These findings could be interpreted

as arising due to either the absence of a selective pressure driving the emergence of 'fit but rare' mutants, or alternatively, because of the presence of an unknown selective pressure against their emergence. Importantly, in our examinations of the fit but rare mutants, we were unable to detect any defect in the ability of these mutants to infect primary cells, including primary T-cells and macrophages. One possible explanation for these findings is that fit but rare mutants are eliminated from natural populations because they confer sensitivity to adaptive or innate immune responses. Taken together, our findings suggest that only ~5% of all possible amino acid substitutions in CA result in viruses that can flourish *in vivo*.

One candidate for selective pressure that induces sequence variation in regions of high genetic fragility, and could possibly suppress 'fit but rare' mutants, is that imposed by cytotoxic T lymphocytes (CTLs). Already a number of examples of CTL escape mutations in protective epitopes have been described, and these have been characterized by their predictability (reoccurrence of the same mutation in distinct infected individuals) [263], their impact on viral fitness (sometimes) [144,264], and by their co-occurrence with additional compensating mutations, suggesting an influence of epistatic interactions that might aid any compromised fitness induced by primary escape mutations [142,143]. These findings are in accordance with the finding that CA is extremely genetically fragile, and suggest that only a small number of possible mutations can permit escape from CTL responses and still maintain fitness. At the same time, it is important to recognize that even though our library of CA mutants should be

representative of random variants, it is still only a sample of the possible mutations that can arise in CA in natural populations. Previous analyses in cell culture indicate that a single cycle of HIV-1 replication (incorporating both transcription and reverse transcription) results in 1.4-3.4 x 10^{-5} mutations per nucleotide [79,80]. Within the 231 aa CA protein, this would yield approximately 0.0097-0.023 errors per replication cycle, which is equivalent to an amino acid substitution once or twice per one hundred replication events. Considering the large population size and high replication rate of HIV-1 *in vivo* [265], this means that every amino acid substitution that can occur by a single nucleotide substitution is generated in nearly every infected individual, many times each day. Therefore selective pressures that are imposed *in vivo* can act on an enormous pool of possible variants. The discrepancies between the fitness and Shannon entropy profiles (Figure 24A) suggest that even in regions of CA in which we found few or no fit mutants amongst 135 library mutants, there are still rare variants that can thrive *in vivo*, even if they may exhibit reduced fitness or require the co-occurrence of compensating mutations.

Previous work has shown that CTL responses to Gag, rather than other HIV-1 proteins, correlate best with reductions in viral loads, and thereby, often disease progression, during HIV-1 infection [148]. Interestingly, we observed that the so-called 'protective' CTL epitopes [143] were more likely to occur in regions of CA that were even more fragile than CA as a whole. As Gag is among the most abundant viral proteins expressed in HIV-1 infected cells, this abundance, and the relative lack of plasticity in CA, likely contribute to the fact

that patients that display strong immune responses to Gag have comparatively favorable clinical outcomes. Still, the capacity of viral populations to explore all available sequence space partly overcomes the genetic fragility of CA, and enables at least some amount of escape from immune responses, even those that target the most vulnerable regions of CA.

Aside from revealing significant insight into the genetic fragility of CA and its biological basis, the CA mutant library is also a useful tool for probing other HIV-1 functions and interactions with cellular factors. Other members of the Bieniasz lab have previously utilized the single amino acid CA mutants to probe for mutants affecting: infection of non-dividing cells, resistance to TRIM5α, and interaction with cyclophilin A. However, it has also been possible to now utilize the CA library to probe the emerging association between CA mutants and Mx2 sensitivity. While previous work had reported CA mutants conferring resistance to Mx2 restriction of HIV-1, these mutants were limited to only a few residues in the cyclophilin binding loop and one other residue (N57, already reported to impact infection of non-dividing cells) [136,137,138]. Surprisingly, we were able to identify CA mutations in multiple new regions of the CA protein that were able to confer resistance to Mx2, some of which had species-dependent effects. Interestingly, comparison of the mutant residues conferring resistance to Mx2 and TRIM5α reveals significant similarities, often involving the same residues or precise mutations (such as M10I and P90T) [140,266]. Considering how many CA mutants have already been shown to impact a number of diverse HIV-1

functions and interactions with host factors, it seems likely that the CA mutant library will continue to prove a useful tool with which HIV-1 can be probed.

Recognizing and understanding the fitness effects of mutations in CA may facilitate drug or vaccine design. For example, a recent study done in the context of a possible vaccination strategy revealed that mutations in the most conserved residues in CA are not necessarily associated with the highest fitness costs [267], as might otherwise be assumed. Thus, an even better grasp of the constraints under which CA evolves may prove beneficial. Furthermore, while CA has not yet been exploited as a target for drugs in the clinic, recent studies have indicated that it is, at least in principle, possible to do so [268,269,270,271,272]. Capsid's genetic fragility should reduce, but not eliminate, the opportunities for drug resistance to occur, and it would thus be ideal if novel therapeutic treatments were to target the CA domains with the greatest fragility. In spite of this, resistance to one recently identified CA-targeted inhibitor has already arisen *in vitro* through mutations in regions of CA that exhibit high fragility. This is a testament to how the sheer number of variants that can be 'tested' or explored in natural viral populations, can, at least sometimes, overcome the lack of robustness in HIV-1 CA. Nevertheless, the high genetic fragility exhibited by CA should make it a target of novel vaccination or therapy strategies.

Uneven mutational robustness of HIV-1 integrase

In our second examination of genetic robustness in the context of HIV-1, we sought to examine the robustness of HIV-1 integrase, in a manner similar

enough to what was done with HIV-1 CA so as to permit comparison, and to also survey how the biological effects of IN mutations could be reflected in an HIV-1 intasome model. Using a large randomly mutagenized library of HIV-1 IN single amino acid mutants, we were able to determine the impact of single residue substitutions *in vitro*, observe their functional impact by western blot, correlate their fitness with their occurrence in natural viral populations, and, perhaps most importantly, meaningfully map all of these phenotypes onto an HIV-1 intasome model. The results reveal a surprising, but significant juxtaposition between the robustness of IN and CA, with just 35% of individual amino acid substitutions in IN resulting in nonviable (<2% of WT fitness) viruses, whereas twice as many mutants in CA (70%) had the same effect. This relative robustness is surprising given the expectation that essential enzymes will be fragile, but is notably similar to results published from a variety of other mutagenesis studies, many of which include other enzymes. For example, comparisons of IN's robustness with other mutagenized HIV-1 proteins (Table 8), reveals how similar the robustness of IN is to protease, in which 40% of single residue mutations were inactivating [225]. Furthermore, other viral proteins have been shown to be equivalently robust to IN (Table 9), such as bacteriophage f1 gene V and Varicella-Zoster virus thymidine kinase, in which 35% of mutations were inactivating [238,239,240]. In fact, the robustness of IN is also similar to several whole viral genomes (Table 10), including Tobacco etch potyvirus, vesicular stomatitis virus, and bacteriophage Qβ, in which 41%, 40%, and 29% of mutations were lethal respectively

[86,87,93]. Finally, the robustness of IN is comparable to nonviral enzymatic proteins (Table 11), such as 3-methyladenine DNA glycosylase (AAG), in which 34% of single amino acid changes were calculated to be inactivating [248], and campylobacter jejuni oligosaccaryltransferase PglB, in which 33% of mutations appeared lethal [249]. While IN is more fragile than other HIV proteins and regions, such as the HIV-1 LTR, in which no mutations (0%) were significantly deleterious [90], or HIV-2 integrase, where just 13% of mutations could be considered lethal [227], HIV-1 IN's relative robustness compared to CA validates the notion that the extreme fragility of CA is exceptional. Nevertheless, the 35% of mutations causing lethal defects in HIV-1 IN are tremendously informative, as mapping to an HIV-1 intasome model indicated their fragility was unevenly distributed, with nonviable mutants, and particularly those affecting particle production or integrase incorporation or expression in virions, highly localized to precise subunit interface regions.

The reasons for the increased robustness of HIV-1 IN, as compared to HIV-1 CA, are not immediately apparent. Both are relatively conserved proteins that are absolutely essential to viral replication, lack redundancy, and appear to have relatively similar rates of variability *in vivo* (Figure 28A-D), all of which could be hypothesized to make them equally vulnerable to mutation. However, there are a number of dissimilarities that could underlie their divergent robustness. As has been mentioned earlier, capsid must maintain a stable structure than can adopt a wide range finely tuned and delicately timed CA:CA interactions that

adjust between immature and mature conformations [110,123,255,256]. Moreover, CA must also form interactions with a wide variety of host proteins, including nuclear pore complex proteins [131], cyclophilin A [120], and CPSF6 [218]. While IN forms critical interactions with LEDGF [189], as well as interactions with TRN-SR2(TNPO3) [273] and 1/hSNF5 [274], and also undergoes multimerization processes [159], there have been no indications that these interactions or structural requirements are as numerous or as stringent as those required for CA. Although any one interaction or structural requirement could theoretically contribute to the observation that significant fitness in CA occurs only within very limited regions, but occurs more frequently and randomly within IN, perhaps it is the larger number of specific and constrained CA interactions (both inter- and intra-protein) and requirements that limits viability to so few regions (Figure 32A, B). Ultimately, it seems likely that the more numerous and restrictive requirements necessary for proper and efficient particle assembly, already established as the basis of HIV-1 CA fragility, result in a unique fragility for HIV-1 CA, and a relative robustness for IN that is more similar to other viral proteins. With more studies of robustness in RNA viruses, more guidelines and critical determinants of robustness will become evident, although these are likely to vary by protein, as we have seen in this work.

Perhaps the most striking finding of the IN mutagenic analyses is the localization of many nonviable IN mutants to specific IN subunit interfaces within the intasome model. Our intention in this work was not to reveal the significance

of single residues within IN, or to investigate IN catalytic activities, but rather to use the comprehensive nature of a randomly mutagenized IN library to derive more extensive understanding of the robustness of IN, and possible regions of vulnerability in the intasome. For this reason, it was gratifying to observe how viable and nonviable mutants can be mapped to distinct regions in the intasome model (Figure 33A-D). It was even more fascinating to observe how precisely many nonviable mutants affecting specific functional phenotypes, namely production of particles and IN expression in virions by western blot, can be mapped to certain subunit interfaces, while nonviable mutants affecting infectiousness only are scattered in buried positions, without any localized pattern in the intasome (Figures 35A, B and 36A, B). Such strong and defined patterns suggest this HIV-1 intasome model is a valid structure for these analyses. The location of most nonviable mutants impacting particle production or IN incorporation in virions lies at the interface between the inner (red and blue regions, Figure 33A) and outer (yellow and green regions, Figure 33A) IN subunits, approximately mutated IN residues 76-87, 102-113, 132-135, and 178-213, and this is largely within the CCD-CCD dimerization interface [159]. Interestingly, the outer subunits of the intasome have previously been reported to lack any clear role in formation of the tetramer, and their function has not been clear [159]. This may explain the considerable viability in these regions, although residues at the edges of the interface with the inner subunit clearly contribute significant fragility (Figure 33C, D). While previous work has suggested that

individual residues at the inner-outer subunit, CCD-CCD interface could be important for multimerization [275,276], none has realized the full extent of the intolerance to mutation in this region, nor has there been intasome mapping that so clearly reveals these specific vulnerabilities. However, this interface is adjacent to, and includes, some critical residues LEDGF uses to bind HIV-1 IN to affect multimerization [277]. This may explain why therapeutic allosteric integrase inhibitors (ALLINIs), which engage with IN at the same LEDGF binding sites (to prevent interaction), are effective at disrupting viral replication [202]. Together, these data suggest that the nonviable IN mutants that affect particle assembly and IN expression in virions occur at localized regions within the CCD-CCD interface on the intasome model, and their fragility may arise because they disturb critical multimerization or binding steps during the assembly of IN. In many ways the suggestion that fragile IN residues are largely those affecting structural dimerization, multimerization, or binding, is analogous to the finding that fragile CA residues are typically those involved in assembly of particles. When combined, the data from both IN and CA suggest that specific structural interactions and requirements during assembly processes are amongst the most significant determinants of robustness and fragility.

While analysis of the location of the fragile IN mutants is informative, so too is the analysis of the robust, viable mutants. An interesting commonality apparent between the IN and CA structural results is the observation that viable single residue mutants were more likely to occur in surface exposed locations

with higher solvent accessible surface areas (Figures 15A, C and 33C, E). Perhaps not intrinsically surprising, as similar data from as far back as the 1960s suggested haemoglobin residues at the surface were less sensitive to replacement than more internal residues [278], it is a trend that may be deserving of more recognition. In the case of HIV-1 IN, these surface residues were also more variable (Figure 39 A, C), an observation that may also be true for other proteins under selective pressure, such as HIV-1 envelope [43,279,280], and which could be used as a predictor of residues more likely to be involved in IN therapeutic resistance. Given the strong association between solvent accessible surface area and viability in both HIV-1 IN and CA, it is likely that the same association exists for other viral proteins, and it is possible that the amount of surface exposure, which could affect a number of structural interactions, is also a determinant of robustness.

Investigations of the relationship between the fitness of IN and CA library mutations and their frequency in naturally occurring HIV-1 subtype B populations revealed a remarkable shared threshold of 40% of WT replicative fitness, below which mutants were unlikely to occur frequently (>3%) *in vivo* (Figure 37A, B). This suggests such a fitness cut-off may exist for other viral proteins and that 40% of WT may be an accurate estimate of *in vivo* fitness restrictions. At the same time, we also noticed that while there were plenty of fit mutants that occurred frequently in natural populations, a large number of fit IN and CA mutants occurred either very rarely or never in our cohorts of 1,000 isolate

sequences (Figure 22C, Table 7). This is intuitively at odds with the expectation that HIV-1 will take advantage of all available sequence space not severely impacting fitness. Interestingly, the differential frequency of fit mutants can be explained for IN, at least in part, by the surface exposure of the fit IN 'frequent' or 'never' mutants (Figure 38A-C). Once again, those IN residues with greater surface exposure and higher solvent accessible surface area values were more likely to occur frequently, whereas more buried residues were more likely to never occur in isolate sequences (Figure 38A-C). However, such an observation did not appear to be true for CA (results not shown), in which no significant differences could be detected between the solvent accessible surface area values for fit 'frequent' and 'rare' residues, although in CA's case, the sample size for analysis was smaller. As previously discussed, the immune pressure created by cytotoxic T lymphocytes (CTLs) could potentially suppress the 'fit but rare' CA mutations. While both IN and CA have been documented to be under selective pressures created by host immune CD8+ CTL responses, there is little to no evidence of CTL-imposed reductions in replication capacity or viral load corresponding to Pol/IN, while multiple CTL responses to Gag/CA have been shown to reduce viral burdens [142,144,148,196,200,264]. The increased robustness of HIV-1 IN should mean it is less vulnerable to immune pressure, as it has more successful 'escape' options, and this may explain why 'fit but rare' mutations in IN are unlikely to be rare because of immune pressure. Ultimately, it

is conceivable that multiple factors, including divergent ones in IN and CA, may contribute to the infrequency of some fit mutants *in vivo*.

The implications of the elevated robustness of HIV-1 IN are not entirely understood. Less vulnerability to immune pressure is one likely outcome, and from a clinical standpoint, lower robustness of CA versus IN should help explain why individuals with immune responses to Gag during HIV-1 infection fare better than those with immune responses to other viral proteins [148]. Increased evolvability is also theorized to be positively correlated with higher robustness [82], as early experiments have already suggested [281].

Summary of HIV-1 IN and CA robustness and associated implications

Through our evaluation of the genetic robustness of two necessary, but functionally divergent HIV-1 proteins, numerous common attributes were revealed. Perhaps most interesting is the observation that in both proteins, fragile residues tended to be those critically involved in the formation of higher order structures—in CA, fragile residues typically impacted particle assembly in a manner likely dependent on CA:CA interactions, and in IN, fragile residues were largely localized to regions associated with multimerization. While perhaps not inherently surprising, this observation may inform of residues or regions in other viral proteins that could be similarly vulnerable to mutation. At the same time, and in accordance with Perutz's 1968 suggestion [278], we also noted that the fragile residues in both IN and CA tended to be those in more buried locations, with lower solvent accessible surface area values. Together, this suggests that,

generally speaking, residues in more buried positions are likely crucial to the structural integrity of the protein, such that disruption by substitution can be a lethal blow to assembling multimers. Finally, we observed a rather surprising shared cutoff for mutant viability *in vivo*: in analyses of both IN and CA libraries, mutants required almost exactly 40% of WT fitness *in vitro* in order to occur frequently in natural populations (>3%). This alludes to the presence of a common threshold at which mutations can confer a selective advantage, and such a limit may also apply to single amino acid substitutions in other HIV-1 open reading frames, or even other viruses. Evaluation of these similarities provides the beginning of a framework in which new structural or functional commonalities amongst viral proteins might be uncovered.

Nevertheless, our genetic robustness studies often revealed significant differences between IN and CA. Most obviously, the two differ greatly in their degree of genetic fragility, with 70% of random single amino acid substitutions in CA resulting in nonviability, while only 35% of single amino acid substitutions in IN had the same effect. The comparative robustness of IN was unexpected given that IN is highly conserved, essential, required to adopt variable conformations, and shows variability similar to that of CA *in vivo*. However, as we have already discussed, the differences in the fragilities of IN and CA likely reflect each protein's individual requirements for inter- and intra- protein interactions. In turn, it seems likely that the disparities between IN and CA's fragilities also underlie the differences in each protein's susceptibility to immune

pressures, including those imposed by CTLs. Ultimately these differences are significant, as they indicate that particularly fragile proteins, like HIV-1 capsid, may be better targets of novel therapies than more robust proteins. Furthermore, the distinctions between IN and CA highlight how further genetic robustness studies of additional proteins in HIV-1 might fill in the significant gaps surrounding our understanding of HIV-1's evolutionary capacity.

Understanding the genetic robustness and fragility of individual proteins in HIV-1, like IN and CA, advances our basic understanding of HIV-1 and reveals how its proteins uniquely cope with pressures induced by high mutation rates and a hostile host environment. Additionally, evaluation of genetic fragility provides insight that could help predict: vulnerable regions within protein structures that could be targeted by novel inhibitors, constraints on sequence space, and resistance to therapeutic treatments. This may prove vital to better estimating the evolutionary path of viruses like HIV-1, and may also aid the development of improved antiviral strategies.

V. REFERENCES

1. Baltimore D (1975) Tumor viruses: 1974. Cold Spring Harbor symposia on quantitative biology 39 Pt 2: 1187-1200.

2. Weiss RA (2006) The discovery of endogenous retroviruses. Retrovirology 3: 67.

3. Crawford LV, Crawford EM (1961) The properties of Rous sarcoma virus purified by density gradient centrifugation. Virology 13: 227-232.

4. Temin (1964) Nature of the provirus of Rous sarcoma. Nat Cancer Inst Monogr: 557-570.

5. Baltimore D (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature 226: 1209-1211.

6. Temin HM, Mizutani S (1970) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature 226: 1211-1213.

7. Cullen BR (1991) Human immunodeficiency virus as a prototypic complex retrovirus. Journal of virology 65: 1053-1056.

8. Coffin JM (1992) Genetic diversity and evolution of retroviruses. Current topics in microbiology and immunology 176: 143-164.

9. Coffin JM, Hughes, Stephen H., Varmus, Harold E. (1997) Retroviruses: Cold Spring Harbor Laboratory Press.

10. Stoye JP (2012) Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nature reviews Microbiology 10: 395-406.

11. Bieniasz PD (2009) The cell biology of HIV-1 virion genesis. Cell Host Microbe 5: 550-558.

12. Freed EO, Mouland AJ (2006) The cell biology of HIV-1 and other retroviruses. Retrovirology 3: 77.

13. Goff SP (2007) Host factors exploited by retroviruses. Nature reviews Microbiology 5: 253-263.

14. Sundquist WI, Krausslich HG (2012) HIV-1 Assembly, Budding, and Maturation. Cold Spring Harb Perspect Med 2: a006924.

15. UNAIDS (2013) 2013 Global fact sheet, Report on the Global AIDS Epidemic. UNAIDS.

16. Flint SJ, Enquist, L.W., Racaniello, V.R., Skalka, A.M. (2009) Principles of Virology: American Society for Microbiology.

17. Barouch DH (2008) Challenges in the development of an HIV-1 vaccine. Nature 455: 613-619.

18. Richman DD, Margolis DM, Delaney M, Greene WC, Hazuda D, et al. (2009) The challenge of finding a cure for HIV infection. Science 323: 1304-1307.

19. CDC (1981) Kaposi's sarcoma and Pneumocystis among homosexual men--New York and California. MMWR Morb Mortal Wkly Rep 30: 305-308.

20. Sharp PM, Hahn BH (2011) Origins of HIV and the AIDS pandemic. Cold Spring Harbor perspectives in medicine 1: a006841.

21. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, et al. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220: 868-871.

22. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, et al. (1984) Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science 224: 500-503.

23. Popovic M, Sarngadharan MG, Read E, Gallo RC (1984) Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science 224: 497-500.

24. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, et al. (1999) Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature 397: 436-441.

25. Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, et al. (2006) Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313: 523-526.

26. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, et al. (2008) Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature 455: 661-664.

27. Van Heuverswyn F, Li Y, Neel C, Bailes E, Keele BF, et al. (2006) Human immunodeficiency viruses: SIV infection in wild gorillas. Nature 444: 164.

28. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR (1989) An African primate lentivirus (SIVsm) closely related to HIV-2. Nature 339: 389-392.

29. Haase AT (1986) Pathogenesis of lentivirus infections. Nature 322: 130-136.

30. Fauci AS (1988) The human immunodeficiency virus: infectivity and mechanisms of pathogenesis. Science 239: 617-622.

31. Clumeck N, Sonnet J, Taelman H, Mascart-Lemone F, De Bruyere M, et al. (1984) Acquired immunodeficiency syndrome in African patients. The New England journal of medicine 310: 492-497.

32. Benveniste RE, Arthur LO, Tsai CC, Sowder R, Copeland TD, et al. (1986) Isolation of a lentivirus from a macaque with lymphoma: comparison with HTLV-III/LAV and other lentiviruses. Journal of virology 60: 483-490.

33. Maddon PJ, Dalgleish AG, McDougal JS, Clapham PR, Weiss RA, et al. (1986) The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. Cell 47: 333-348.

34. Clapham PR, McKnight A (2001) HIV-1 receptors and cell tropism. British medical bulletin 58: 43-59.

35. Simon V, Ho DD (2003) HIV-1 dynamics in vivo: implications for therapy. Nature reviews Microbiology 1: 181-190.

36. Fauci AS (1996) Host factors and the pathogenesis of HIV-induced disease. Nature 384: 529-534.

37. Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, et al. (1996) Identification of a major co-receptor for primary isolates of HIV-1. Nature 381: 661-666.

38. Feng Y, Broder CC, Kennedy PE, Berger EA (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science 272: 872-877.

39. Choe H, Farzan M, Sun Y, Sullivan N, Rollins B, et al. (1996) The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. Cell 85: 1135-1148.

40. Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR (1997) Change in coreceptor use correlates with disease progression in HIV-1--infected individuals. The Journal of experimental medicine 185: 621-628.

41. Robinson HL (2002) New hope for an AIDS vaccine. Nature reviews Immunology 2: 239-250.

42. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, et al. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature 393: 648-659.

43. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, et al. (1998) The antigenic structure of the HIV gp120 envelope glycoprotein. Nature 393: 705-711.

44. Cullen BR (1986) Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. Cell 46: 973-982.

45. Gendelman HE, Phelps W, Feigenbaum L, Ostrove JM, Adachi A, et al. (1986) Trans-activation of the human immunodeficiency virus long terminal repeat sequence by DNA viruses. Proceedings of the National Academy of Sciences of the United States of America 83: 9759-9763.

46. Daly TJ, Cook KS, Gray GS, Maione TE, Rusche JR (1989) Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. Nature 342: 816-819.

47. Karn J, Stoltzfus CM (2012) Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harbor perspectives in medicine 2: a006916.

48. Malim MH, Emerman M (2008) HIV-1 accessory proteins--ensuring viral survival in a hostile environment. Cell host & microbe 3: 388-398.

49. Sheehy AM, Gaddis NC, Choi JD, Malim MH (2002) Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 418: 646-650.

50. Mehle A, Goncalves J, Santa-Marta M, McPike M, Gabuzda D (2004) Phosphorylation of a novel SOCS-box regulates assembly of the HIV-1 Vif-Cul5 complex that promotes APOBEC3G degradation. Genes & development 18: 2861-2866.

51. Schrofelbauer B, Yu Q, Landau NR (2004) New insights into the role of Vif in HIV-1 replication. AIDS reviews 6: 34-39.

52. Ooms M, Brayton B, Letko M, Maio SM, Pilcher CD, et al. (2013) HIV-1 Vif adaptation to human APOBEC3H haplotypes. Cell host & microbe 14: 411-421.

53. Neil SJ, Zang T, Bieniasz PD (2008) Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature 451: 425-430.

54. Bour S, Schubert U, Strebel K (1995) The human immunodeficiency virus type 1 Vpu protein specifically binds to the cytoplasmic domain of CD4: implications for the mechanism of degradation. Journal of virology 69: 1510-1520.

55. Laguette N, Bregnard C, Hue P, Basbous J, Yatim A, et al. (2014) Premature activation of the SLX4 complex by Vpr promotes G2/M arrest and escape from innate immune sensing. Cell 156: 134-145.

56. Kirchhoff F (2010) Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. Cell host & microbe 8: 55-67.

57. Arien KK, Verhasselt B (2008) HIV Nef: role in pathogenesis and viral fitness. Current HIV research 6: 200-208.

58. Lenassi M, Cagney G, Liao M, Vaupotic T, Bartholomeeusen K, et al. (2010) HIV Nef is secreted in exosomes and triggers apoptosis in bystander CD4+ T cells. Traffic 11: 110-122.

59. Arhel NJ, Kirchhoff F (2009) Implications of Nef: host cell interactions in viral persistence and progression to AIDS. Current topics in microbiology and immunology 339: 147-175.

60. Duffy S, Shackelton LA, Holmes EC (2008) Rates of evolutionary change in viruses: patterns and determinants. Nature reviews Genetics 9: 267-276.

61. Gago S, Elena SF, Flores R, Sanjuan R (2009) Extremely high mutation rate of a hammerhead viroid. Science 323: 1308.

62. Sanjuan R (2012) From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. PLoS pathogens 8: e1002685.

63. Kimura M (1983) The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

64. Orr HA (2000) Adaptation and the cost of complexity. Evolution; international journal of organic evolution 54: 13-20.

65. Orr HA (2000) The rate of adaptation in asexuals. Genetics 155: 961-968.

66. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, et al. (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303: 327-332.

67. Lemey P, Rambaut A, Pybus OG (2006) HIV evolutionary dynamics within and among hosts. AIDS reviews 8: 125-140.

68. Li WH, Tanimura M, Sharp PM (1988) Rates and dates of divergence between AIDS virus nucleotide sequences. Molecular biology and evolution 5: 313-330.

69. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R (2006) Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature 439: 344-348.

70. Pfeiffer JK, Kirkegaard K (2005) Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. PLoS pathogens 1: e11.

71. Richman DD, Little SJ, Smith DM, Wrin T, Petropoulos C, et al. (2004) HIV evolution and escape. Transactions of the American Clinical and Climatological Association 115: 289-303.

72. Perelson AS (2002) Modelling viral and immune system dynamics. Nature reviews Immunology 2: 28-36.

73. Dapp MJ, Patterson SE, Mansky LM (2013) Back to the future: revisiting HIV-1 lethal mutagenesis. Trends in microbiology 21: 56-62.

74. Anderson JP, Daifuku R, Loeb LA (2004) Viral error catastrophe by mutagenic nucleosides. Annual review of microbiology 58: 183-205.

75. Vignuzzi M, Wendt E, Andino R (2008) Engineering attenuated virus vaccines by controlling replication fidelity. Nature medicine 14: 154-161.

76. Davenport MP, Loh L, Petravic J, Kent SJ (2008) Rates of HIV immune escape and reversion: implications for vaccination. Trends in microbiology 16: 561-566.

77. Pepin KM, Lass S, Pulliam JR, Read AF, Lloyd-Smith JO (2010) Identifying genetic markers of adaptation for surveillance of viral host jumps. Nature reviews Microbiology 8: 802-813.

78. Holmes EC (2009) The evolution and emergence of RNA viruses. Oxford: Oxford University Press.

79. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R (2010) Viral mutation rates. J Virol 84: 9733-9748.

80. Mansky LM, Temin HM (1995) Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. Journal of virology 69: 5087-5094.

81. Murray JM, Kelleher AD, Cooper DA (2011) Timing of the components of the HIV life cycle in productively infected CD4+ T cells in a population of HIV-infected individuals. Journal of virology 85: 10798-10805.

82. Lauring AS, Frydman J, Andino R (2013) The role of mutational robustness in RNA virus evolution. Nat Rev Microbiol 11: 327-336.

83. Sanjuan R (2010) Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. Philos Trans R Soc Lond B Biol Sci 365: 1975-1982.

84. de Visser JA, Hermisson J, Wagner GP, Ancel Meyers L, Bagheri-Chaichian H, et al. (2003) Perspective: Evolution and detection of genetic robustness. Evolution 57: 1959-1972.

85. Wagner A (2005) Robustness, evolvability, and neutrality. FEBS Lett 579: 1772-1778.

86. Sanjuan R, Moya A, Elena SF (2004) The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc Natl Acad Sci U S A 101: 8396-8401.

87. Domingo-Calap P, Cuevas JM, Sanjuan R (2009) The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. PLoS Genet 5: e1000742.

88. Cuevas JM, Moya A, Sanjuan R (2009) A genetic background with low mutational robustness is associated with increased adaptability to a novel host in an RNA virus. J Evol Biol 22: 2041-2048.

89. Graci JD, Gnadig NF, Galarraga JE, Castro C, Vignuzzi M, et al. (2012) Mutational robustness of an RNA virus influences sensitivity to lethal mutagenesis. Journal of virology 86: 2869-2873.

90. van Opijnen T, Boerlijst MC, Berkhout B (2006) Effects of random mutations in the human immunodeficiency virus type 1 transcriptional promoter on viral fitness in different host cell environments. J Virol 80: 6678-6685.

91. Parera M, Fernandez G, Clotet B, Martinez MA (2007) HIV-1 protease catalytic efficiency effects caused by random single amino acid substitutions. Mol Biol Evol 24: 382-387.

92. Wagner A (2005) Robustness and Evolvability in Living Systems: Princeton University Press.

93. Carrasco P, de la Iglesia F, Elena SF (2007) Distribution of fitness and virulence effects caused by single-nucleotide substitutions in Tobacco Etch virus. J Virol 81: 12979-12984.

94. Montville R, Froissart R, Remold SK, Tenaillon O, Turner PE (2005) Evolution of mutational robustness in an RNA virus. PLoS Biol 3: e381.

95. Codoner FM, Daros JA, Sole RV, Elena SF (2006) The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. PLoS Pathog 2: e136.

96. Sanjuan R, Cuevas JM, Furio V, Holmes EC, Moya A (2007) Selection for robustness in mutagenized RNA viruses. PLoS Genet 3: e93.

97. Wilke CO (2001) Selection for fitness versus selection for robustness in RNA secondary structure folding. Evolution; international journal of organic evolution 55: 2412-2420.

98. Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C (2001) Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature 412: 331-333.

99. Svensson EIaC, R. (2012) The Adaptive Landscape in Evolutionary Biology. Oxford: Oxford University Press.

100. Lee CH, Gilbertson DL, Novella IS, Huerta R, Domingo E, et al. (1997) Negative effects of chemical mutagenesis on the adaptive behavior of vesicular stomatitis virus. Journal of virology 71: 3636-3640.

101. Crotty S, Maag D, Arnold JJ, Zhong W, Lau JY, et al. (2000) The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. Nature medicine 6: 1375-1379.

102. Gnadig NF, Beaucourt S, Campagnola G, Borderia AV, Sanz-Ramos M, et al. (2012) Coxsackievirus B3 mutator strains are attenuated in vivo. Proceedings of the National Academy of Sciences of the United States of America 109: E2294-2303.

103. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, et al. (2003) The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature 424: 94-98.

104. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, et al. (2003) Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 424: 99-103.

105. Lecossier D, Bouchonnet F, Clavel F, Hance AJ (2003) Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science 300: 1112.

106. Freed EO (1998) HIV-1 gag proteins: diverse functions in the virus life cycle. Virology 251: 1-15.

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

108. Lee SK, Potempa M, Swanstrom R (2012) The choreography of HIV-1 proteolytic processing and virion assembly. The Journal of biological chemistry 287: 40867-40874.

109. Briggs JA, Krausslich HG (2011) The molecular architecture of HIV. J Mol Biol 410: 491-500.

110. Briggs JA, Riches JD, Glass B, Bartonova V, Zanetti G, et al. (2009) Structure and assembly of immature HIV. Proc Natl Acad Sci U S A 106: 11090-11095.

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

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

113. Gitti RK, Lee BM, Walker J, Summers MF, Yoo S, et al. (1996) Structure of the amino-terminal core domain of the HIV-1 capsid protein. Science 273: 231-235.

114. Gamble TR, Vajdos FF, Yoo S, Worthylake DK, Houseweart M, et al. (1996) Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. Cell 87: 1285-1294.

115. Gamble TR, Yoo S, Vajdos FF, von Schwedler UK, Worthylake DK, et al. (1997) Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. Science 278: 849-853.

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

117. Pornillos O, Ganser-Pornillos BK, Yeager M (2011) Atomic-level modelling of the HIV capsid. Nature 469: 424-427.

118. Chang YF, Wang SM, Huang KJ, Wang CT (2007) Mutations in capsid major homology region affect assembly and membrane affinity of HIV-1 Gag. J Mol Biol 370: 585-597.

119. Zimmerman C, Klein KC, Kiser PK, Singh AR, Firestein BL, et al. (2002) Identification of a host protein essential for assembly of immature HIV-1 capsids. Nature 415: 88-92.

120. Luban J, Bossolt KL, Franke EK, Kalpana GV, Goff SP (1993) Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. Cell 73: 1067-1078.

121. Li S, Hill CP, Sundquist WI, Finch JT (2000) Image reconstructions of helical assemblies of the HIV-1 CA protein. Nature 407: 409-413.

122. Mateu MG (2009) The capsid protein of human immunodeficiency virus: intersubunit interactions during virus assembly. FEBS J 276: 6098-6109.

123. Pornillos O, Ganser-Pornillos BK, Kelly BN, Hua Y, Whitby FG, et al. (2009) X-ray structures of the hexameric building block of the HIV capsid. Cell 137: 1282-1292.

124. Lanman J, Lam TT, Emmett MR, Marshall AG, Sakalian M, et al. (2004) Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange. Nat Struct Mol Biol 11: 676-677.

125. Lanman J, Lam TT, Barnes S, Sakalian M, Emmett MR, et al. (2003) Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry. J Mol Biol 325: 759-772.

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

127. Yamashita M, Perez O, Hope TJ, Emerman M (2007) Evidence for direct involvement of the capsid protein in HIV infection of nondividing cells. PLoS Pathog 3: 1502-1510.

128. Krishnan L, Matreyek KA, Oztop I, Lee K, Tipper CH, et al. (2010) The requirement for cellular transportin 3 (TNPO3 or TRN-SR2) during infection maps to human immunodeficiency virus type 1 capsid and not integrase. J Virol 84: 397-406.

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

130. Lee K, Ambrose Z, Martin TD, Oztop I, Mulky A, et al. (2010) Flexible use of nuclear import pathways by HIV-1. Cell Host Microbe 7: 221-233.

131. Schaller T, Ocwieja KE, Rasaiyaah J, Price AJ, Brady TL, et al. (2011) HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. PLoS Pathog 7: e1002439.

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

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

134. Sokolskaja E, Sayah DM, Luban J (2004) Target cell cyclophilin A modulates human immunodeficiency virus type 1 infectivity. J Virol 78: 12800-12808.

135. Qi M, Yang R, Aiken C (2008) Cyclophilin A-dependent restriction of human immunodeficiency virus type 1 capsid mutants for infection of nondividing cells. J Virol 82: 12001-12008.

136. Kane M, Yadav SS, Bitzegeio J, Kutluay SB, Zang T, et al. (2013) MX2 is an interferon-induced inhibitor of HIV-1 infection. Nature 502: 563-566.

137. Goujon C, Moncorge O, Bauby H, Doyle T, Ward CC, et al. (2013) Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. Nature 502: 559-562.

138. Liu Z, Pan Q, Ding S, Qian J, Xu F, et al. (2013) The interferon-inducible MxB protein inhibits HIV-1 infection. Cell host & microbe 14: 398-410.

139. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, et al. (2004) The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature 427: 848-853.

140. Soll SJ, Wilson SJ, Kutluay SB, Hatziioannou T, Bieniasz PD (2013) Assisted evolution enables HIV-1 to overcome a high TRIM5alpha-imposed genetic barrier to rhesus macaque tropism. PLoS pathogens 9: e1003667.

141. Manel N, Hogstad B, Wang Y, Levy DE, Unutmaz D, et al. (2010) A cryptic sensor for HIV-1 activates antiviral innate immunity in dendritic cells. Nature 467: 214-217.

142. Crawford H, Matthews PC, Schaefer M, Carlson JM, Leslie A, et al. (2011) The hypervariable HIV-1 capsid protein residues comprise HLA-driven CD8+ T-cell escape mutations and covarying HLA-independent polymorphisms. J Virol 85: 1384-1390.

143. Dahirel V, Shekhar K, Pereyra F, Miura T, Artyomov M, et al. (2011) Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. Proc Natl Acad Sci U S A 108: 11530-11535.

144. Brockman MA, Brumme ZL, Brumme CJ, Miura T, Sela J, et al. (2010) Early selection in Gag by protective HLA alleles contributes to reduced HIV-1 replication capacity that may be largely compensated for in chronic infection. Journal of virology 84: 11937-11949.

145. Carlson JM, Brumme CJ, Martin E, Listgarten J, Brockman MA, et al. (2012) Correlates of Protective Cellular Immunity Revealed by Analysis of Population-Level Immune Escape Pathways in HIV-1. J Virol 86: 13202-13216.

146. Borrow P, Lewicki H, Hahn BH, Shaw GM, Oldstone MB (1994) Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. J Virol 68: 6103-6110.

147. Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PI, et al. (2010) The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 330: 1551-1557.

148. Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, et al. (2007) CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. Nat Med 13: 46-53.

149. Kelleher AD, Long C, Holmes EC, Allen RL, Wilson J, et al. (2001) Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. J Exp Med 193: 375-386.

150. Troyer RM, McNevin J, Liu Y, Zhang SC, Krizan RW, et al. (2009) Variable fitness impact of HIV-1 escape mutations to cytotoxic T lymphocyte (CTL) response. PLoS Pathog 5: e1000365.

151. Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, et al. (2004) HIV evolution: CTL escape mutation and reversion after transmission. Nat Med 10: 282-289.

152. Reicin AS, Paik S, Berkowitz RD, Luban J, Lowy I, et al. (1995) Linker insertion mutations in the human immunodeficiency virus type 1 gag gene: effects on virion particle assembly, release, and infectivity. J Virol 69: 642-650.

153. Srinivasakumar N, Hammarskjold ML, Rekosh D (1995) Characterization of deletion mutations in the capsid region of human immunodeficiency virus type 1 that affect particle formation and Gag-Pol precursor incorporation. J Virol 69: 6106-6114.

154. Fitzon T, Leschonsky B, Bieler K, Paulus C, Schroder J, et al. (2000) Proline residues in the HIV-1 NH2-terminal capsid domain: structure determinants for proper core assembly and subsequent steps of early replication. Virology 268: 294-307.

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

156. Ceccherini-Silberstein F, Malet I, D'Arrigo R, Antinori A, Marcelin AG, et al. (2009) Characterization and structural analysis of HIV-1 integrase conservation. AIDS Rev 11: 17-29.

157. Zheng R, Jenkins TM, Craigie R (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. Proc Natl Acad Sci U S A 93: 13659-13664.

158. Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol Cell Biol 12: 2331-2338.

159. Li X, Krishnan L, Cherepanov P, Engelman A (2011) Structural biology of retroviral DNA integration. Virology 411: 194-205.

160. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, et al. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science 266: 1981-1986.

161. Woerner AM, Klutch M, Levin JG, Marcus-Sekura CJ (1992) Localization of DNA binding activity of HIV-1 integrase to the C-terminal half of the protein. AIDS Res Hum Retroviruses 8: 297-304.

162. Eijkelenboom AP, Lutzke RA, Boelens R, Plasterk RH, Kaptein R, et al. (1995) The DNA-binding domain of HIV-1 integrase has an SH3-like fold. Nat Struct Biol 2: 807-810.

163. Guiot E, Carayon K, Delelis O, Simon F, Tauc P, et al. (2006) Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity. J Biol Chem 281: 22707-22719.

164. Krishnan L, Li X, Naraharisetty HL, Hare S, Cherepanov P, et al. (2010) Structure-based modeling of the functional HIV-1 intasome and its inhibition. Proc Natl Acad Sci U S A 107: 15910-15915.

165. Krishnan L, Engelman A (2012) Retroviral integrase proteins and HIV-1 DNA integration. The Journal of biological chemistry 287: 40858-40866.

166. Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P (2010) Retroviral intasome assembly and inhibition of DNA strand transfer. Nature 464: 232-236.

167. Maertens GN, Hare S, Cherepanov P (2010) The mechanism of retroviral integration from X-ray structures of its key intermediates. Nature 468: 326-329.

168. Johnson BC, Metifiot M, Ferris A, Pommier Y, Hughes SH (2013) A homology model of HIV-1 integrase and analysis of mutations designed to test the model. J Mol Biol 425: 2133-2146.

169. Miller MD, Farnet CM, Bushman FD (1997) Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. J Virol 71: 5382-5390.

170. Roth MJ, Schwartzberg PL, Goff SP (1989) Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. Cell 58: 47-54.

171. Fujiwara T, Mizuuchi K (1988) Retroviral DNA integration: structure of an integration intermediate. Cell 54: 497-504.

172. Brown PO, Bowerman B, Varmus HE, Bishop JM (1989) Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. Proceedings of the National Academy of Sciences of the United States of America 86: 2525-2529.

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

174. Bushman FD, Craigie R (1991) Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. Proc Natl Acad Sci U S A 88: 1339-1343.

175. Wiskerchen M, Muesing MA (1995) Identification and characterization of a temperature-sensitive mutant of human immunodeficiency virus type 1 by alanine scanning mutagenesis of the integrase gene. J Virol 69: 597-601.

176. Engelman A, Craigie R (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. J Virol 66: 6361-6369.

177. Engelman A, Liu Y, Chen H, Farzan M, Dyda F (1997) Structure-based mutagenesis of the catalytic domain of human immunodeficiency virus type 1 integrase. J Virol 71: 3507-3514.

178. Jenkins TM, Esposito D, Engelman A, Craigie R (1997) Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. EMBO J 16: 6849-6859.

179. Chen JC, Krucinski J, Miercke LJ, Finer-Moore JS, Tang AH, et al. (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. Proc Natl Acad Sci U S A 97: 8233-8238.

180. Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, et al. (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. Nat Struct Biol 4: 567-577.

181. Johnson AA, Santos W, Pais GC, Marchand C, Amin R, et al. (2006) Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop. J Biol Chem 281: 461-467.

182. Bushman FD, Engelman A, Palmer I, Wingfield P, Craigie R (1993) Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. Proc Natl Acad Sci U S A 90: 3428-3432.

183. Drelich M, Wilhelm R, Mous J (1992) Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein in vitro. Virology 188: 459-468.

184. Vink C, Oude Groeneger AM, Plasterk RH (1993) Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. Nucleic Acids Res 21: 1419-1425.

185. Engelman A (1999) In vivo analysis of retroviral integrase structure and function. Advances in virus research 52: 411-426.

186. Engelman A, Englund G, Orenstein JM, Martin MA, Craigie R (1995) Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. J Virol 69: 2729-2736.

187. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, et al. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. Cell 110: 521-529.

188. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, et al. (2003) HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. The Journal of biological chemistry 278: 372-381.

189. Zheng Y, Ao Z, Jayappa KD, Yao X (2010) Characterization of the HIV-1 integrase chromatin- and LEDGF/p75-binding abilities by mutagenic analysis within the catalytic core domain of integrase. Virol J 7: 68.

190. Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP (1994) Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science 266: 2002-2006.

191. Christ F, Thys W, De Rijck J, Gijsbers R, Albanese A, et al. (2008) Transportin-SR2 imports HIV into the nucleus. Current biology : CB 18: 1192-1202.

192. Taltynov O, Desimmie BA, Demeulemeester J, Christ F, Debyser Z (2012) Cellular cofactors of lentiviral integrase: from target validation to drug discovery. Molecular biology international 2012: 863405.

193. Rain JC, Cribier A, Gerard A, Emiliani S, Benarous R (2009) Yeast two-hybrid detection of integrase-host factor interactions. Methods 47: 291-297.

194. Borghans JA, Molgaard A, de Boer RJ, Kesmir C (2007) HLA alleles associated with slow progression to AIDS truly prefer to present HIV-1 p24. PloS one 2: e920.

195. Tang Y, Huang S, Dunkley-Thompson J, Steel-Duncan JC, Ryland EG, et al. (2010) Correlates of spontaneous viral control among long-term survivors of perinatal HIV-1 infection expressing human leukocyte antigen-B57. AIDS 24: 1425-1435.

196. Rodriguez WR, Addo MM, Rathod A, Fitzpatrick CA, Yu XG, et al. (2004) CD8+ T lymphocyte responses target functionally important regions of Protease and Integrase in HIV-1 infected subjects. J Transl Med 2: 15.

197. Liu Y, McNevin J, Cao J, Zhao H, Genowati I, et al. (2006) Selection on the human immunodeficiency virus type 1 proteome following primary infection. Journal of virology 80: 9519-9529.

198. Payne RP, Kloverpris H, Sacha JB, Brumme Z, Brumme C, et al. (2010) Efficacious early antiviral activity of HIV Gag- and Pol-specific HLA-B 2705-restricted CD8+ T cells. Journal of virology 84: 10543-10557.

199. Mothe B, Llano A, Ibarrondo J, Daniels M, Miranda C, et al. (2011) Definition of the viral targets of protective HIV-1-specific T cell responses. Journal of translational medicine 9: 208.

200. Brockman MA, Chopera DR, Olvera A, Brumme CJ, Sela J, et al. (2012) Uncommon pathways of immune escape attenuate HIV-1 integrase replication capacity. J Virol 86: 6913-6923.

201. Semenova EA, Marchand C, Pommier Y (2008) HIV-1 integrase inhibitors: update and perspectives. Advances in pharmacology 56: 199-228.

202. Engelman A, Kessl JJ, Kvaratskhelia M (2013) Allosteric inhibition of HIV-1 integrase activity. Current opinion in chemical biology 17: 339-345.

203. Kessl JJ, Jena N, Koh Y, Taskent-Sezgin H, Slaughter A, et al. (2012) Multimode, cooperative mechanism of action of allosteric HIV-1 integrase inhibitors. The Journal of biological chemistry 287: 16801-16811.

204. Andrews CD, Spreen WR, Mohri H, Moss L, Ford S, et al. (2014) Long-acting integrase inhibitor protects macaques from intrarectal simian/human immunodeficiency virus. Science 343: 1151-1154.

205. Wilson SJ, Schoggins JW, Zang T, Kutluay SB, Jouvenet N, et al. (2012) Inhibition of HIV-1 particle assembly by 2',3'-cyclic-nucleotide 3'-phosphodiesterase. Cell host & microbe 12: 585-597.

206. Zhang YJ, Hatziioannou T, Zang T, Braaten D, Luban J, et al. (2002) Envelope-dependent, cyclophilin-independent effects of glycosaminoglycans on human immunodeficiency virus type 1 attachment and infection. Journal of virology 76: 6332-6343.

207. Bieniasz PD, Cullen BR (2000) Multiple blocks to human immunodeficiency virus type 1 replication in rodent cells. Journal of virology 74: 9868-9877.

208. Rihn SJ, Wilson SJ, Loman NJ, Alim M, Bakker SE, et al. (2013) Extreme genetic fragility of the HIV-1 capsid. PLoS Pathog 9: e1003461.

209. Zhang F, Zang T, Wilson SJ, Johnson MC, Bieniasz PD (2011) Clathrin facilitates the morphogenesis of retrovirus particles. PLoS Pathog 7: e1002119.

210. Varthakavi V, Browning PJ, Spearman P (1999) Human immunodeficiency virus replication in a primary effusion lymphoma cell line stimulates lytic-phase replication of Kaposi's sarcoma-associated herpesvirus. J Virol 73: 10329-10338.

211. Bouyac-Bertoia M, Dvorin JD, Fouchier RA, Jenkins Y, Meyer BE, et al. (2001) HIV-1 infection requires a functional integrase NLS. Mol Cell 7: 1025-1035.

212. McNatt MW, Zang T, Hatziioannou T, Bartlett M, Fofana IB, et al. (2009) Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants. PLoS Pathog 5: e1000300.

213. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797.

214. Shenkin PS, Erman B, Mastrandrea LD (1991) Information-theoretical entropy as a measure of sequence variability. Proteins 11: 297-313.

215. Holm L, Koivula AK, Lehtovaara PM, Hemminki A, Knowles JK (1990) Random mutagenesis used to probe the structure and function of Bacillus stearothermophilus alpha-amylase. Protein Eng 3: 181-191.

216. Hatziioannou T, Cowan S, Von Schwedler UK, Sundquist WI, Bieniasz PD (2004) Species-specific tropism determinants in the human immunodeficiency virus type 1 capsid. J Virol 78: 6005-6012.

217. Ambrose Z, Lee K, Ndjomou J, Xu H, Oztop I, et al. (2012) Human immunodeficiency virus type 1 capsid mutation N74D alters cyclophilin A dependence and impairs macrophage infection. Journal of virology 86: 4708-4714.

218. Lee K, Mulky A, Yuen W, Martin TD, Meyerson NR, et al. (2012) HIV-1 capsid-targeting domain of cleavage and polyadenylation specificity factor 6. J Virol 86: 3851-3860.

219. McCarthy KR, Schmidt AG, Kirmaier A, Wyand AL, Newman RM, et al. (2013) Gain-of-sensitivity mutations in a Trim5-resistant primary isolate of pathogenic SIV identify two independent conserved determinants of Trim5alpha specificity. PLoS pathogens 9: e1003352.

220. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, et al. (2013) Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature 497: 643-646.

221. Cahn P, Pozniak AL, Mingrone H, Shuldyakov A, Brites C, et al. (2013) Dolutegravir versus raltegravir in antiretroviral-experienced, integrase-inhibitor-naive adults with HIV: week 48 results from the randomised, double-blind, non-inferiority SAILING study. Lancet 382: 700-708.

222. Dolan J, Chen A, Weber IT, Harrison RW, Leis J (2009) Defining the DNA substrate binding sites on HIV-1 integrase. Journal of molecular biology 385: 568-579.

223. Cannon PM, Wilson W, Byles E, Kingsman SM, Kingsman AJ (1994) Human immunodeficiency virus type 1 integrase: effect on viral replication of mutations at highly conserved residues. J Virol 68: 4768-4775.

224. Shao W, Everitt L, Manchester M, Loeb DD, Hutchison CA, 3rd, et al. (1997) Sequence requirements of the HIV-1 protease flap region determined by saturation mutagenesis and kinetic analysis of flap mutants. Proc Natl Acad Sci U S A 94: 2243-2248.

225. Loeb DD, Swanstrom R, Everitt L, Manchester M, Stamper SE, et al. (1989) Complete mutagenesis of the HIV-1 protease. Nature 340: 397-400.

226. Chao SF, Chan VL, Juranka P, Kaplan AH, Swanstrom R, et al. (1995) Mutational sensitivity patterns define critical residues in the palm subdomain of the reverse transcriptase of human immunodeficiency virus type 1. Nucleic Acids Res 23: 803-810.

227. van den Ent FM, Vos A, Plasterk RH (1998) Mutational scan of the human immunodeficiency virus type 2 integrase protein. J Virol 72: 3916-3924.

228. Smith RA, Anderson DJ, Preston BD (2006) Hypersusceptibility to substrate analogs conferred by mutations in human immunodeficiency virus type 1 reverse transcriptase. J Virol 80: 7169-7178.

229. Kim B, Hathaway TR, Loeb LA (1996) Human immunodeficiency virus reverse transcriptase. Functional mutants obtained by random mutagenesis coupled with genetic selection in Escherichia coli. J Biol Chem 271: 4872-4878.

230. Yasugi T, Vidal M, Sakai H, Howley PM, Benson JD (1997) Two classes of human papillomavirus type 16 E1 mutants suggest pleiotropic conformational constraints affecting E1 multimerization, E2 interaction, and interaction with cellular proteins. J Virol 71: 5942-5951.

231. Markiewicz P, Kleina LG, Cruz C, Ehret S, Miller JH (1994) Genetic studies of the lac repressor. XIV. Analysis of 4000 altered Escherichia coli lac repressors reveals essential and non-essential residues, as well as

"spacers" which do not require a specific sequence. J Mol Biol 240: 421-433.

232. Morrison HG, Kirchhoff F, Desrosiers RC (1995) Effects of mutations in constant regions 3 and 4 of envelope of simian immunodeficiency virus. Virology 210: 448-455.

233. Nakajima K, Nobusawa E, Tonegawa K, Nakajima S (2003) Restriction of amino acid change in influenza A virus H3HA: comparison of amino acid changes observed in nature and in vitro. J Virol 77: 10088-10098.

234. Yano T, Nobusawa E, Nagy A, Nakajima S, Nakajima K (2008) Effects of single-point amino acid substitutions on the structure and function neuraminidase proteins in influenza A virus. Microbiol Immunol 52: 216-223.

235. Pakula AA, Young VB, Sauer RT (1986) Bacteriophage lambda cro mutations: effects on activity and intracellular degradation. Proc Natl Acad Sci U S A 83: 8829-8833.

236. Rhee SS, Hunter E (1991) Amino acid substitutions within the matrix protein of type D retroviruses affect assembly, transport and membrane association of a capsid. EMBO J 10: 535-546.

237. Alin K, Goff SP (1996) Amino acid substitutions in the CA protein of Moloney murine leukemia virus that block early events in infection. Virology 222: 339-351.

238. Terwilliger TC, Zabin HB, Horvath MP, Sandberg WS, Schlunk PM (1994) In vivo characterization of mutants of the bacteriophage f1 gene V protein isolated by saturation mutagenesis. J Mol Biol 236: 556-571.

239. Masso M, Mathe E, Parvez N, Hijazi K, Vaisman, II (2009) Modeling the functional consequences of single residue replacements in bacteriophage f1 gene V protein. Protein Eng Des Sel 22: 665-671.

240. Suzutani T, Lacey SF, Powell KL, Purifoy DJ, Honess RW (1992) Random mutagenesis of the thymidine kinase gene of varicella-zoster virus. J Virol 66: 2118-2124.

241. Eifan SA, Elliott RM (2009) Mutational analysis of the Bunyamwera orthobunyavirus nucleocapsid protein gene. J Virol 83: 11307-11317.

242. Stenger DC, Young BA, French R (2006) Random mutagenesis of wheat streak mosaic virus HC-Pro: non-infectious interfering mutations in a gene dispensable for systemic infection of plants. J Gen Virol 87: 2741-2747.

243. Rennell D, Bouvier SE, Hardy LW, Poteete AR (1991) Systematic mutation of bacteriophage T4 lysozyme. J Mol Biol 222: 67-88.

244. Strambio-de-Castillia C, Hunter E (1992) Mutational analysis of the major homology region of Mason-Pfizer monkey virus by use of saturation mutagenesis. J Virol 66: 7021-7032.

245. Moyer CL, Wiethoff CM, Maier O, Smith JG, Nemerow GR (2011) Functional genetic and biophysical analyses of membrane disruption by human adenovirus. J Virol 85: 2631-2641.

246. Van Der Velden A, Kaminski A, Jackson RJ, Belsham GJ (1995) Defective point mutants of the encephalomyocarditis virus internal ribosome entry site can be complemented in trans. Virology 214: 82-90.

247. Peris JB, Davis P, Cuevas JM, Nebot MR, Sanjuan R (2010) Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1. Genetics 185: 603-609.

248. Guo HH, Choe J, Loeb LA (2004) Protein tolerance to random amino acid change. Proc Natl Acad Sci U S A 101: 9205-9210.

249. Ihssen J, Kowarik M, Wiesli L, Reiss R, Wacker M, et al. (2012) Structural insights from random mutagenesis of Campylobacter jejuni oligosaccharyltransferase PglB. BMC Biotechnol 12: 67.

250. Olins PO, Bauer SC, Braford-Goldberg S, Sterbenz K, Polazzi JO, et al. (1995) Saturation mutagenesis of human interleukin-3. J Biol Chem 270: 23754-23760.

251. Huang W, Petrosino J, Hirsch M, Shenkin PS, Palzkill T (1996) Amino acid sequence determinants of beta-lactamase structure and activity. J Mol Biol 258: 688-703.

252. Chen C, Roberts VA, Rittenberg MB (1992) Generation and analysis of random point mutations in an antibody CDR2 sequence: many mutated antibodies lose their ability to bind antigen. J Exp Med 176: 855-866.

253. Axe DD, Foster NW, Fersht AR (1998) A search for single substitutions that eliminate enzymatic function in a bacterial ribonuclease. Biochemistry 37: 7157-7166.

254. Kawashima H, Yamagishi J, Yamayoshi M, Ohue M, Fukui T, et al. (1992) Structure-activity relationships in human interleukin-1 alpha: identification of key residues for expression of biological activities. Protein Eng 5: 171-176.

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

256. Jouvenet N, Bieniasz PD, Simon SM (2008) Imaging the biogenesis of individual HIV-1 virions in live cells. Nature 454: 236-240.

257. Krausslich HG (1991) Human immunodeficiency virus proteinase dimer as component of the viral polyprotein prevents particle assembly and viral infectivity. Proc Natl Acad Sci U S A 88: 3213-3217.

258. Ott DE, Coren LV, Chertova EN, Gagliardi TD, Nagashima K, et al. (2003) Elimination of protease activity restores efficient virion production to a human immunodeficiency virus type 1 nucleocapsid deletion mutant. J Virol 77: 5547-5556.

259. Ott DE, Coren LV, Shatzer T (2009) The nucleocapsid region of human immunodeficiency virus type 1 Gag assists in the coordination of assembly and Gag processing: role for RNA-Gag binding in the early stages of assembly. J Virol 83: 7718-7727.

260. Borsetti A, Ohagen A, Gottlinger HG (1998) The C-terminal half of the human immunodeficiency virus type 1 Gag precursor is sufficient for efficient particle assembly. J Virol 72: 9313-9317.

261. Auerbach MR, Shu C, Kaplan A, Singh IR (2003) Functional characterization of a portion of the Moloney murine leukemia virus gag gene by genetic footprinting. Proc Natl Acad Sci U S A 100: 11678-11683.

262. Auerbach MR, Brown KR, Singh IR (2007) Mutational analysis of the N-terminal domain of Moloney murine leukemia virus capsid protein. J Virol 81: 12337-12347.

263. Allen TM, Altfeld M, Geer SC, Kalife ET, Moore C, et al. (2005) Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. J Virol 79: 13239-13249.

264. Martinez-Picado J, Prado JG, Fry EE, Pfafferott K, Leslie A, et al. (2006) Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. Journal of virology 80: 3617-3623.

265. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD (1996) HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science 271: 1582-1586.

266. Ganser-Pornillos BK, Chandrasekaran V, Pornillos O, Sodroski JG, Sundquist WI, et al. (2011) Hexagonal assembly of a restricting TRIM5alpha protein. Proceedings of the National Academy of Sciences of the United States of America 108: 534-539.

267. Rolland M, Manocheewa S, Swain JV, Lanxon-Cookson EC, Kim M, et al. (2013) HIV-1 Conserved-Element Vaccines: Relationship between Sequence Conservation and Replicative Capacity. J Virol 87: 5461-5467.

268. Tang C, Loeliger E, Kinde I, Kyere S, Mayo K, et al. (2003) Antiviral inhibition of the HIV-1 capsid protein. J Mol Biol 327: 1013-1020.

269. Blair WS, Pickford C, Irving SL, Brown DG, Anderson M, et al. (2010) HIV capsid is a tractable target for small molecule therapeutic intervention. PLoS Pathog 6: e1001220.

270. Sticht J, Humbert M, Findlow S, Bodem J, Muller B, et al. (2005) A peptide inhibitor of HIV-1 assembly in vitro. Nat Struct Mol Biol 12: 671-677.

271. Zhang H, Zhao Q, Bhattacharya S, Waheed AA, Tong X, et al. (2008) A cell-penetrating helical peptide as a potential HIV-1 inhibitor. J Mol Biol 378: 565-580.

272. Lemke CT, Titolo S, von Schwedler U, Goudreau N, Mercier JF, et al. (2012) Distinct effects of two HIV-1 capsid assembly inhibitor families that bind the same site within the N-terminal domain of the viral CA protein. J Virol 86: 6643-6655.

273. De Houwer S, Demeulemeester J, Thys W, Taltynov O, Zmajkovicova K, et al. (2012) Identification of residues in the C-terminal domain of HIV-1 integrase that mediate binding to the transportin-SR2 protein. J Biol Chem 287: 34059-34068.

274. Yung E, Sorin M, Wang EJ, Perumal S, Ott D, et al. (2004) Specificity of interaction of INI1/hSNF5 with retroviral integrases and its functional significance. J Virol 78: 2222-2231.

275. Berthoux L, Sebastian S, Muesing MA, Luban J (2007) The role of lysine 186 in HIV-1 integrase multimerization. Virology 364: 227-236.

276. Al-Mawsawi LQ, Hombrouck A, Dayam R, Debyser Z, Neamati N (2008) Four-tiered pi interaction at the dimeric interface of HIV-1 integrase critical for DNA integration and viral infectivity. Virology 377: 355-363.

277. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A (2005) Structural basis for the recognition between HIV-1 integrase and

transcriptional coactivator p75. Proceedings of the National Academy of Sciences of the United States of America 102: 17308-17313.

278. Perutz MF, Lehmann H (1968) Molecular pathology of human haemoglobin. Nature 219: 902-909.

279. Yamaguchi-Kabata Y, Gojobori T (2000) Reevaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein and prediction of new discontinuous epitopes. Journal of virology 74: 4335-4350.

280. Zolla-Pazner S, Cardozo T (2010) Structure-function relationships of HIV-1 envelope sequence-variable regions refocus vaccine design. Nature reviews Immunology 10: 527-535.

281. McBride RC, Ogbunugafor CB, Turner PE (2008) Robustness promotes evolvability of thermotolerance in an RNA virus. BMC evolutionary biology 8: 231.