Viral Evolution and Escape during Acute HIV-1 Infection

Christian L. Boutwell, 1 Morgane M. Rolland, 2 Joshua T. Herbeck, 2 James I. Mullins, 2 and Todd M. Allen 1

1 Ragon Institute of MGH, MIT and Harvard, Boston, Massachusetts; 2 Departments of Microbiology and Medicine, University of Washington School of Medicine, Seattle

The extensive genetic diversity of human immunodeficiency virus type 1 (HIV-1) presents a significant barrier to the development of an effective and durable HIV vaccine. This variability not only makes it difficult to identify the targets against which immune responses should be directed, but it also confers on the virus the capacity for rapid escape from effective immune responses. Here, we describe recent investigations of the genetic diversity of HIV-1 at transmission and of the evolution of the virus as it adapts to the host immune environment during the acute phase of HIV-1 infection. These studies increase our understanding of the virology of the earliest stages of HIV-1 infection and provide critical insights into the mechanisms underlying viral replication and immune control of diverse HIV-1 strains. Such knowledge will inform the design of smarter, more effective vaccines capable of inducing immune control of HIV-1.

The acute phase of human immunodeficiency virus type 1 (HIV-1) infection encompasses the initial interaction between virus and host and is of critical interest, because the outcome strongly influences the pathogenesis and clinical course of the infection. Acute HIV-1 infection resulting from sexual transmission begins with the breaching of the mucosal barrier by infectious virions and the establishment of a localized focus of infection in tissue-associated immune cells in the mucosa. As the infection spreads to the regional lymph nodes and is disseminated to other lymphocyte-rich compartments throughout the body, particularly the spleen and the gut-associated lymphoid tissue, the virus encounters high densities of primary target CD4+ T cells, causing a massive burst of viral replication. An ensuing peak of viremia is closely followed by a decrease in viral load, in association with the concomitant induction of HIV-1–specific CD8+ T cells and depletion of susceptible CD4+ T cells. The viral load ultimately levels to a steady-state “set-point” viremia that then remains broadly stable throughout much of the chronic phase of infection.

The immune control of HIV-1 is determined by the intrinsic qualities of both the immune response and the HIV-1 quasispecies. The effectiveness of early CD8+ T cell responses in controlling HIV-1 replication is a critical determinant of the set-point viral load [1], and their importance is further supported by the association between particular HLA class I alleles and an improved course of HIV-1 infection [2]. The exact mechanisms by which some individuals achieve control of viral replication remain unclear, yet their elucidation would provide valuable insight into the immune environment that should be induced by an HIV-1 vaccine. Recent investigations of the virology of acute HIV-1 infection have started to yield critical insights into (1) the virus lineages that establish infection, (2) the earliest viral adaptive changes that enable escape from host immune responses, and (3) the impact of immune escape mutations in HIV-1 on viral replication and pathogenesis.
Here, we summarize some of these recent advances and discuss the application of these findings to the development of more effective HIV-1 vaccine approaches.

CHARACTERIZATION OF TRANSMITTED OR FOUNDER VIRUSES

The extensive genetic diversity of HIV-1 poses a significant challenge to the development of broadly effective prophylactic interventions, because the majority of circulating strains are likely to differ significantly from the antigenic insert in the prophylactic agent. As such, the identification of unique viral genetic characteristics associated with the transmission of HIV-1 would provide important information about the mechanisms of transmission or establishment of infection and may narrow the choices among potential HIV-1 targets. Although a small number of studies have successfully identified viral characteristics associated with successful transmission (eg, shorter V1–V2 length and fewer N-linked glycosylation sites in Env protein of non-B subtypes [3, 4]), the practical difficulties associated with obtaining samples from the earliest days of HIV-1 infection have hampered their characterization. Recent efforts have focused on sampling virus populations during acute infection, especially before seroconversion, to infer the genetic sequence of the founder virus using a combination of sequence analysis, evolutionary modeling, and phylogenetic analyses [5–10]. These studies have demonstrated that ∼75% of sexual transmissions of HIV-1 are attributable to infection with a single viral genetic lineage, whether for HIV-1B transmission in men who have sex with men [7, 10] or for HIV-1A and HIV-1C transmission in heterosexual couples [5, 6, 8]. In the cases in which multiple founder variants were deduced, there were usually 2 and typically ≤5 founder lineages, and such cases were often associated with factors that can disrupt the mucosal barrier, for example, sexually transmitted disease–associated genital inflammation [6]. More recently, a study by Bar et al illustrated that injection drug use was associated with the transmission of a high frequency of multiple sequence variants, suggesting that bypassing the normal mucosal barrier may facilitate transmission of multiple variants [11].

Recently, this approach has been extended to known transmission pairs, which allows for the comparison of the deduced founder virus in the recipient to the entire pool of potential founder lineages that were present in the donor quasispecies. In their study of 10 HIV-1A and HIV-1C heterosexual transmission pairs, Haaland et al [6] analyzed ∼250 amino acids from the V1 loop to the V4 loop of env and identified 2 pairs in which the founder virus was an exact genetic match of a virus present in the donor, 6 pairs in which the founder virus differed from the most closely related virus in the donor by ≤5 amino acids (∼98% identity), and 2 pairs in which the founder virus and the most closely related donor virus differed by 10–11 amino acids (∼96% identity). Interestingly, in 9 of the 10 transmission pairs, the donor variant most closely related (>98% identity) to the founder virus in the recipient comprised only a small fraction of the total donor virus population (<5%). A similar pattern was observed when whole virus genomes were analyzed in a recent study of 3 transmission pairs in which the transmitter had chronic infection and thus had a highly genetically diverse viral population [12].

In addition to these studies of HIV-1 transmission, recent investigations of low-dose, intrarectal infection of rhesus macaques with SIVmac251, a nonclonal SIV inoculum, demonstrated that the majority of infections were founded by a single genetic lineage. Moreover, the genetic relatedness between the inferred founder virus and the most closely related variant in the inoculum was similar to that described above for HIV-1 transmission [13]. In addition to supporting the HIV-1 results, these data suggest the potential value of mucosal challenge approaches for the accurate modeling of HIV-1 transmission for vaccine testing.

The observation that HIV-1 infection is typically established by a single genetic lineage during sexual transmission offers hope that HIV-1 vaccines may have to block a limited number of viral strains in most cases. Unfortunately, to date no unique characteristics of transmitted variants have been identified that would permit more specific targeting.

NEXT-GENERATION SEQUENCING OF ACUTE OR EARLY VIRUSES

The population of genetically diverse viruses that develops in each infected individual confers on HIV-1 the capacity to adapt rapidly and relentlessly to host immune responses. Such immune-adaptation contributes to ineffective immune control of viral replication [14, 15], and experiments in the SIV-macaque model confirm that cytotoxic T lymphocyte (CTL) escape will pose a significant threat to the durability and efficacy of vaccine-induced CD8+ T cell responses [16]. To properly understand immune adaptation of HIV-1, it is critical to characterize the dynamics of viral evolution during acute infection, when numerous adaptive changes take place. Recent studies examining near-full-length genomes [12, 17, 18] showed that at peak viremia the virus population remains virtually homogeneous, with significant evolution occurring by week 6 of infection, including a dramatic increase in CTL immune-adaptive mutations, particularly in highly variable proteins, such as Env and Nef [12, 17]. Interestingly, multiple mutations in targeted CTL epitopes were observed in both studies. Herbeck et al [12] noted that such mutations were often mutually exclusive, which suggests that the virus population can explore multiple adaptive pathways but that there are limits to the plasticity of individual viruses, even in highly variable proteins.

Although such intensive longitudinal analyses of viral evo-
lution during acute or early infection have been limited to a small number of individuals, next-generation sequencing technologies promise to scale up these efforts and to increase our understanding of early HIV-1 adaptations to host selection pressures. Chip-based sequencing technologies, such as pyrosequencing (454 Life Sciences) [19] and 4-color cyclic reversible termination sequencing (Illumina) [20], offer rapid and cost-effective production of up to 10,000-fold sequencing read coverage of the HIV-1 genome. This represents a substantial improvement over the costly and time-consuming amplification and sequencing of individual amplicons for the analysis of viral genetic variation. Indeed, the depth of coverage provided by these approaches allows the quantification of genetic variants at frequencies that are undetectable by standard sequencing approaches, and the increased sensitivity provides an unprecedented view of the earliest events in the adaptation of HIV-1 to frontline immune responses mounted during acute infection. The relatively short read lengths may, however, limit the utility of next-generation sequencing technologies for establishing linkage of mutations except over short regions of the genomes.

To date, next-generation sequencing of HIV-1 has predominantly been used to identify and quantify antiretroviral drug resistance mutations at frequencies below the limit of detection of commercial drug resistance genotyping assays (to a level of \~1\%), to understand the effect of minor variant resistance mutations on treatment outcomes [21–24]. Recent efforts have been applied to HIV-1 evolution during the acute or early phase of infection. For example, by comparing chip-based pyrosequencing to conventional sequencing for the analysis of the evolution of specific CTL epitopes during acute or early infection in longitudinal samples from 3 patients, Fischer et al [25] showed that next-generation sequencing captured a greater degree of genetic diversity and detected CTL escape variants earlier than conventional approaches. The detection by “deep” sequencing of viral CTL escape variants in early infection than previously appreciated (eg, as early as 17 days after infection), was also recently reported in the SIV-macaque system [26]. Next-generation sequencing also allows the characterization of CD8+ T cell epitopes that “shatter” during viral escape—that is, exhibit multiple, highly variable, low-frequency escape mutations—before coalescing on a single escape pathway, versus epitopes that are limited to escape at a single residue [27]. As such, deep sequencing approaches are now capable of distinguishing between the early pathways of viral escape within different CTL epitopes, revealing critical information about the ease with which HIV-1 can evade these responses.

The recent development of novel assembly, alignment, and analysis tools has enabled high-throughput, next-generation sequencing of near full-length viral genomes (M. R. Henn and T.M.A., unpublished data; W. C. Deng and J.I.M., unpublished data). This approach revealed that the earliest escaping epitopes were in more variable proteins of the virus (Vif, Env, and Nef) [27] and corresponded to highly immunodominant responses (M. R. Henn and T.M.A., unpublished data). These data suggest that viral escape may be tied more closely to the kinetics and specificity of the immune response than to the functionality of the response. These advances will allow future studies to extend beyond analyses of the evolution of specific epitopes in a small number of samples to full-scale, unbiased screening of HIV-1 evolution in large cross-sectional and longitudinal cohorts. Although individual-based studies are critical for providing fine details of the interactions between specific viruses and specific host immune responses, population-based studies are necessary to identify common patterns of HIV-1 escape from the earliest immune responses and, more specifically, to define attributes associated with effective control of HIV-1.

**IMMUNE ESCAPE-ASSOCIATED VIRAL FITNESS COSTS**

The dynamics of immune escape are predominantly governed by the impact of the escape mutation on relative viral fitness, a composite parameter representing the net contribution of all mutation-associated benefits and costs. Although the evasion of immune responses conferred by escape mutations represents a clear fitness benefit to the virus, the HIV-1 proteome is not endlessly plastic, and the same mutations can carry associated fitness costs. The existence of such costs is evidenced by the reversion of transmitted escape mutations during acute and early HIV-1 infection [12, 15, 17, 18, 28], and a number of CTL escape mutations have been identified that disrupt normal virus protein structure and/or function [29–32]. The majority of deleterious escape mutations have been identified in the relatively conserved Gag protein, whereas Troyer et al [33] recently showed that CTL escape mutations in Env protein did not typically carry an associated fitness cost and in several cases enhanced competitive viral fitness. This result is consistent with the lack of a fitness cost associated with virus escape from neutralizing antibodies [34] and suggests that differential functional constraints play a role in determining escape mutation-associated fitness costs.

In addition, a recent analysis of the impact on viral replication capacity of 20 CTL escape mutations in Gag epitopes identified only 3 escape mutations that caused substantial reductions in viral replication capacity, suggesting that high-cost escape mutations are relatively rare [35]. Interestingly, the 3 high-cost CTL escape mutations occurred in epitopes dominantly targeted by protective HLA class I alleles during acute infection [1]. This suggests that the protection afforded by particular HLA class I alleles may result from either maintenance of a dominant and effective CD8+ T cell response because the barrier to viral escape in the targeted epitope is high, and/
or selection of high-cost escape mutations that attenuate virus replication.

Additional evidence of a role for CTL escape–associated fitness costs in control of HIV-1 replication comes from the examination of viruses derived from HIV-1 controllers (individuals who maintain long-term control of HIV-1 viremia), in which rare or novel CTL escape mutations are commonly observed and which exhibit generally reduced replication capacities [36–39]. However, the source of the CTL escape variants in these patients remains unclear; they could result either from early immune selection in the contemporary host or from infection with a founder virus that already contains the escape mutations. Indeed, 2 recent reports describe improved early clinical correlates associated with the transmission of viruses expressing multiple CTL escape mutations known to impair in vitro replication capacity [40, 41]. Importantly, the improved early outcomes associated with such transmissions have been observed to fade as the transmitted escape mutations revert to wild type [32], and the long-term clinical outcome of these transient effects remains uncertain.

As the virology of acute HIV-1 infection is better characterized, it becomes possible to model how differences in founder virus replication and in viral escape from dominant CD8+ T cell responses may contribute to observed variability in the immune control of HIV-1 replication. One hypothesis is that differences in intrinsic founder virus replication, caused by carryover mutations (whether immune escape–derived or otherwise), and differences in the rate of escape from contemporary dominant CD8+ T cell responses, caused by escape-associated fitness costs, combine to influence set-point viral load and the early clinical course of disease (Figure 1). For example, the transmission of a virus attenuated by accumulated carryover mutations to a host that mounts dominant CD8+ T cell responses against epitopes with high escape-associated fitness costs is proposed to contribute to elite control of viral replication. In contrast, transmission of an attenuated virus to a

![Figure 1](image)

**Figure 1.** Differential founder virus replication and immune escape from dominant CD8+ T cell responses combine to influence the human immunodeficiency virus type 1 (HIV-1) set-point viral load and clinical course. The differential replication of wild-type founder virus (solid line) and attenuated founder virus (dashed line) results in different levels of peak viremia and long-term control of HIV-1. The impact of dominant CD8+ T cell responses (red cells) on the viral population is determined by the relative rate of escape in the targeted epitopes (gray triangles), which ranges from immediate escape (top set-point line) because of transmission of escape mutations in the founder virus to no escape (bottom set-point line) because of high escape-associated fitness costs and low viral replication. The differential founder virus replication and dominant CD8+ T cell response effectiveness result in poor control of acute HIV-1 replication and a high set-point viral load (red set-point line), normal control and a normal set point (green lines), and protective control and a low-to-undetectable set point (blue set-point lines). CTL, cytotoxic T lymphocyte.
host with dominant CD8+ T cell responses against epitopes with low, or no, barrier to escape is proposed to result in a more typical set-point viral load and clinical course. Such models may provide a useful framework for considering the relative roles of multiple, highly variable host and virus characteristics.

ESCAPE AND FITNESS-GUIDED VACCINE APPROACHES

The development of an effective vaccine against HIV-1, whether capable of inducing sterilizing or therapeutic immunity, is essential. The consideration of CTL escape in the design of vaccine immunogens will be of particular importance for vaccines that induce effective antiviral immune responses but fail to provide either sterilizing immunity or complete suppression of viral replication. Indeed, there is clear evidence from both human and macaque studies that HIV-1 can escape from narrow vaccine-induced immune responses and that such escape can erode vaccine efficacy [16, 42]. Thus, an increased understanding of barriers to transmission, the nature of transmitted viruses, viral immune escape pathways, and immune escape–associated impairment of replication is required for the development of rational strategies for vaccine design. Variability-inclusive approaches, such as the “mosaic” [43] and the “center-of-tree” approach with high-frequency variants included [44–46], seek to induce immune responses against a broad range of naturally occurring strains and to block common CTL escape states by creating composite immunogens that maximize the 9-mer coverage found in circulating HIV-1 isolates.

Other designs seek to compile the most highly conserved and most immunodominant regions of the HIV-1 proteome into a single antigen to focus immune responses against regions of the virus in which CTL escape carries a severe cost to the virus [47–49]. For example, the immunogen in the “conserved elements” design is generated by an iterative process that includes the most conserved regions of the HIV-1 proteome and known protective CTL epitopes, organizes the elements of the construct to maximize protein expression and epitope processing, and integrates CTL escape–specific modifications on the basis of escape-associated fitness cost data [50]. Finally, other approaches attempt to overcome natural immunodominance hierarchies to focus responses against more conserved subdominant CD8+ T cell epitopes in which there are high escape-associated viral fitness costs rather than against current immunodominant epitopes in which escape occurs with little or no cost [47] (C.L.B. and T.M.A., unpublished data). The common characteristic of all of these approaches is that they seek to leverage our knowledge of the global and intrahost evolution of HIV-1 to overcome the challenges posed by the diversity and adaptive capacity of the virus in order to elicit responses less susceptible to viral escape.

CONCLUSION

In summary, careful characterization of the early events of acute HIV-1 infection has provided key insights into viral-host interactions that influence viral pathogenesis and immune control of viral replication. Continued study of HIV-1 sequence evolution, in particular during acute and early infection and in situations of natural control of HIV-1 replication, is required to refine vaccine design approaches and to identify novel mechanisms of immune control. The recent application of next-generation sequencing approaches to the study of HIV-1 should speed this process substantially.

References