Molecular Mechanisms of HIV-1 Prophylaxis Failure Revealed by Single Genome Sequencing

Hui Li1, Lily Blair2, Yalu Chen3, Gerald Learn1, Katja Pfafferott4, Mina John4, Tanmoy Bhattacharya2, Beatrice H Hahn1, Simon Mallal4, George M Shaw1, Katharine J Bar1,#

1Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, U.S.A
2Los Alamos National Laboratory, Los Alamos, New Mexico, U.S.A
3University of Alabama at Birmingham, Birmingham, Alabama, U.S.A
4Institute for Immunology and Infectious Diseases, Murdoch University, Perth, Australia

#Corresponding author: Katharine J. Bar, 502D Johnson Pavilion, 3610 Hamilton Walk, Philadelphia, PA 19104. Tel: 215.573.8497. Fax: 215.573.8976. bark@upenn.edu
ABSTRACT

Trials of HIV-1 pre- and post-exposure prophylaxis show promise. Here, we describe a novel strategy for deciphering mechanisms of prophylaxis failure that could improve therapeutic outcomes. A healthcare worker began antiretroviral prophylaxis immediately after a high-risk needlestick injury but nonetheless became viremic 11 weeks later. Single genome sequencing of plasma viral RNA identified 15 drug-sensitive transmitted/founder HIV-1 genomes responsible for productive infection. Sequences emanating from these genomes exhibited extremely low diversity, suggesting virus sequestration as opposed to low-level replication as the cause of breakthrough infection. Identification of transmitted/founder viruses allows for genome-wide assessment of molecular mechanisms of prophylaxis failure.
Introduction

Prevention of HIV-1 infection is one of the most important and challenging goals of global health. Current efforts are broadly focused, and promising interventions under development include pre- and post-exposure prophylaxis (PrEP; PEP) using antiretroviral drugs, topical microbicides, and vaccines. Recent clinical trials have demonstrated efficacies of 31% to 67% for these prophylactic strategies in different risk groups [1]. There are no prospective randomized trials of PEP, but an earlier retrospective case-control study of occupational exposures suggested an 81% reduction in HIV-1 acquisition [2]. Despite these encouraging findings, protection in all studies was incomplete. Methodologically, it has been challenging to pinpoint specific mechanisms of prophylaxis failure. Nonetheless, elucidation of mechanisms of prophylaxis failure could aid the development of more effective regimens. Here, we describe a novel application of the single genome sequencing (SGS) strategy to identify T/F viruses responsible for prophylaxis failure and implicate a novel mechanism of PEP treatment failure.

Case Report

A 53-year old Australian female healthcare worker (HCW) sustained a needle-stick injury with a hollow-bore 21-gauge needle after drawing blood from an advanced stage HIV-infected patient. She reported the needle pierced the pulp of her gloved index finger and that she inadvertently struck the syringe plunger such that a small amount of blood was likely injected. The HCW had no other risk factors for HIV acquisition. The source patient had been HIV-infected for more than ten years, with prior treatment experience to three classes of antiretrovirals. At the time of the exposure, he was non-adherent to his prescribed regimen of zidovudine, lamivudine and nevirapine. He had a HIV viral load exceeding 100,000 copies per milliliter, a CD4+ T cell count of 279 cells per milliliter and a recent HIV genotype demonstrating a K103N mutation. The HCW initiated PEP with zidovudine, lamivudine and indinavir within two hours of the exposure and
completed four weeks of therapy (days 1-32) that included a four-day interruption between days 22-25. At a routine visit five weeks after the needlestick exposure and several days after completing PEP treatment, the HCW was asymptomatic and had a negative third-generation HIV antibody ELISA test. Eleven weeks after the exposure (six weeks after PEP discontinuation), the HCW presented with new onset fever, arthralgias and exudative pharyngitis. Laboratory examinations revealed an HIV-1 plasma viral load of 2,200,000 copies per milliliter, a positive HIV-1 ELISA and an indeterminate HIV-1 Western blot (faint <1+ gp160 band only). She was diagnosed with acute HIV-1 infection, Fiebig stage IV [3].

METHODS

Study subjects: Blood samples from the index HCW and source patient were obtained after receiving informed consent under the Royal Perth Hospital Human Research Ethics Committee. Blood specimens were obtained from the HCW 81 days after exposure and from the source patient 8 days after the exposure.

Viral RNA and genomic DNA extraction, cDNA synthesis, single genome amplification and direct sequencing: vRNA was extracted from plasma, DNA was extracted from PBMCs, cDNA was synthesized from env and pol vRNA, and SGS performed as previously described [4-6]. Sequences from the index subject are identified throughout with the prefix 43715, and from the source subject, the prefix 43690.

Sequence analysis: Sequences were aligned with ClustalW and hand-checked using MacClade 4.08. Phylogenetic trees were generated by the neighbor-joining (NJ) method using ClustalW or maximum likelihood (ML) using PhyML v.3. Sequences were assessed for APOBEC3G/F signatures and recombination (see supplement for details).
RESULTS

Phylogenetic Analyses

A total of 179 full-length pol and gp41 env sequences from the HCW plasma and PBMCs and 61 full-length pol and gp41 env sequences from the source patient plasma were determined. Figure 1A illustrates a NJ phylogeny of 23 gp41 env sequences from the source patient, revealing broad genetic diversity (maximum 4.44%) and quasispecies complexity typical of chronic HIV-1 infection [4, 7]. Figure 1B reveals a strikingly different pattern of env diversity in the recently infected HCW. Here, multiple sets of identical or nearly identical sequences (in blue) are interspersed with sequences of recombinant origin (in orange). This phylogenetic pattern is characteristic of acute infection by multiple genetically diverse viruses [4-6], where each low diversity sequence lineage represents the progeny of a distinct T/F virus and interspersed viruses generally represent discrete recombinants. Occasional T/F viruses represented by a single sequence can also be identified based on their genetic distance from other sequences. The two largest lineages among the gp41 sequences had sufficient numbers of sequences to allow for robust modeling of sequence diversification [4, 8]. These sequences conformed to a model of neutral evolution, including a star-like phylogeny and a Poisson distribution of mutations (Table S1). Sequences from all low-diversity lineages coalesced to unambiguous T/F sequences. Thus, we can conclude that the env sequences from the index subject depicted in Figure 1B represent the progeny of at least 14 discrete T/F viruses.

These 14 T/F genomes were next analyzed together with sequences from the source patient and unlinked reference subjects (Fig 1C). Source and index sequences formed a single monophyletic radiation supported by high bootstrap values (100%), indicating clear epidemiological linkage. Source and index sequences exhibited similar maximum diversities (4.44% and 4.49%, respectively), and were interspersed in the phylogenetic tree, again indicating epidemiological linkage and absence of obvious selection for particular variants in the transmission event.
Figure 2 depicts the NJ phylogeny of source (A) and phylogeny and Highlighter plot of index (B) subjects’ viral pol sequences, which corroborate and extend findings from the env analyses. Again, the pattern of sequence diversity in the chronically-infected source patient was strikingly different from that in the HCW, with sequences from the latter again consisting of multiple sets of identical or nearly identical sequences characteristic of T/F virus lineages. As with env, the pol sequences comprising the most populated lineages conformed to a model of random sequence evolution (Table S1), with each lineage’s sequences coalescing to unambiguous T/F sequences [4]. The enumerated T/F viruses represent minimum estimates, which are substantially affected by sampling depth (see supplement). Our estimates of at least 14-15 T/F viruses based on env and pol analyses are in good agreement.

**Mathematical Modeling of T/F Sequence Evolution**

Sequence lineages with sufficient numbers of sequences for analysis conformed to a model of random HIV-1 evolution [4], and we could thus use well-established parameters for HIV-1 reverse transcriptase (RT) error rate and virus generation time to estimate the time to a most recent common ancestor (MRCA) [8]. Analysis of the most populated T/F lineages in both env and pol revealed far shorter times to an estimated MRCA than the 81 days documented clinically. For example, for gp41 env lineages v1 and v5 (Fig 1B), the estimated times to a MRCA were 37 and 18 days, respectively; for pol lineages v1 and v2 (Fig 2B), MRCA estimates were 14 and 23 days (Table S1). These low MRCA estimates suggest that each sequence lineage evolved from a discrete T/F virus that began to replicate only after PEP therapy was discontinued. This finding implies a substantial period of virus sequestration and evolutionary arrest during PEP therapy.
Drug Resistance Does Not Contribute to Virus Breakthrough

To determine if transmission of drug resistant viruses or selection for drug resistance could have contributed to PEP failure, we analyzed all sequences from the index and source patients for genotypic evidence of antiretroviral resistance. In the RT gene, a K103N mutation was detected in all source and index patient sequences (Fig. S1A). This mutation could be explained by the source patient’s prior therapy but is irrelevant to the PEP regimen administered to the HCW. In protease, the HCW sequences contained a single M46I mutation in only one of 15 T/F lineages (2 of 72 sequences total; see variant 15 in Fig 2). Importantly, M46I confers resistance to protease inhibitors only in combination with other resistance-associated mutations [9]. Thus, drug resistance mutations relevant to the prescribed PEP regimen were neither transmitted nor evolved during or after the administration of PEP.

DISCUSSION

Understanding molecular mechanisms of HIV-1 prophylaxis failure has important implications for public health and for the personal health of individuals exposed to HIV-1. Here, we apply a novel experimental strategy for characterizing and discriminating between potential mechanisms of PEP failure. Despite timely initiation of a potent three-drug PEP regimen, the HCW demonstrated delayed viremia and seroconversion resulting from transmission of at least 15 T/F viruses. This is the first example of occupational HIV-1 exposure and PEP failure where the molecular identities and minimum numbers of actual T/F viral genomes responsible for infection have been unambiguously determined.

The surprising findings of 15 drug sensitive T/F viruses, together with minimal diversification of their progeny and the absence of relevant transmitted or selected drug resistance-associated mutations, suggest virus sequestration and associated evolutionary arrest underlying virus breakthrough in this subject. Several mechanisms could potentially explain the sequestration and evolutionary stasis of transmitted viral
genomes throughout PEP treatment. These can be divided into (i) pre-integration latency, (ii) post-integration latency, (iii) incomplete suppression of virus replication, and (iv) virus trapping. HIV-1 virion-associated RNA can enter cells and be arrested prior to proviral integration (pre-integration latency), however, the lifespan of pre-integration complexes is generally believed to be only a few days, far less than the four-week period of PEP in this subject. Similarly, post-integration latency would require the unlikely occurrence of infection of many thousands, if not millions, of CD4+ T-cells in the first hours after virus exposure in order to generate the 15 or more latently-infected memory CD4+ T-cells required to explain our findings. A third possibility is low-level viral replication via cell-to-cell spread, which has been postulated to be less sensitive to drug suppression than cell-free virus spread [10]. This mechanism also seems unlikely given the evolutionary stasis and absence of accumulated resistance mutations in our 15 T/F virus lineages.

A fourth possibility, which we favor, is virus trapping and sequestration by follicular dendritic cells (FDCs) or other antigen presenting cells. The initial WB immunoreactivity to gp160 (only) is consistent with this hypothesis. FDCs are known to bind and retain infectious HIV-1 in the form of immune complexes for at least 9 months in murine models following footpad inoculation. Such virus can prime B-cells for antibody production and initiate productive virus infection upon co-culture of FDCs and permissive human CD4+ T cells [11].

The literature contains many reports of combination PEP failure in occupational exposures in resource-rich countries. Notably, the majority of these HCWs also experienced significant delays in symptoms of acute retroviral syndrome (mean 46, range 23-79 days) and antibody seroconversion (mean 67, range 23-97 days) [12]. Because clinical symptoms and systemic viremia typically develop within 10 days (95% CI 7-21 days) of virus exposure [3], this protracted time course suggests a scenario similar to our case, where exponential virus replication began after completion of PEP. Similar delays in viremia and antibody seroconversion were observed in macaque studies where tenofovir-based PEP failed to protect against
intravenous inoculation [13]. Thus delayed viremia, clinical symptoms and antibody seroconversion may be common in HIV infection arising after failed PEP.

The remarkably high numbers of T/F genomes seen in this case have not been observed in sexual transmission of HIV-1 [4], nor in the majority of transmissions in injection drug users (IDUs) [6]. The high numbers of T/F viruses in the index case, however, are indistinguishable from what has been observed in a minority of acutely-infected IDUs and in rhesus macaques inoculated intravenously with simian immunodeficiency virus [6, 14]. Thus, while surveys of occupational HIV-1 exposures estimate a low risk of virus acquisition following percutaneous exposure (0.3%) [15], with currently available PEP regimens further diminishing this risk [2], there may be atypical exposures like the described HCW where the risk of virus acquisition is particularly high. In such cases, more prolonged therapy with potent combination antiretroviral therapy may be warranted.

In summary, we have demonstrated the utility of SGS and T/F virus analyses to distinguish potential molecular mechanisms of PEP failure. Future clinical trials of PEP and PrEP may benefit from a SGS approach to the elucidation of mechanisms of virus breakthrough and treatment failure.

Funding: This work was supported by the National Institutes of Health Center for HIV/AIDS Vaccine Immunology (grant number U19AI067854) and the Center for HIV/AIDS Vaccine Immunology and Immunogen Development (grant number UM1AI100645), and the Bill and Melinda Gates Foundation (grant number 37874).

Acknowledgements: We thank Patricia Crystal for manuscript preparation.
Footnote page

1. None of the authors of the manuscript have any commercial or other association that would pose a conflict of interest with this work.

2. This work was supported by the National Institutes of Health Center for HIV/AIDS Vaccine Immunology (U19AI067854), the Center for HIV/AIDS Vaccine Immunology and Immunogen Development (UM1AI100645) and the Bill and Melinda Gates Foundation (grant 37874).

3. This work was presented in part at the Conference on Retroviruses and Opportunistic Infections, March 2011, Boston, MA, abstract #492.

4. Correspondence and requests for reprints should be addressed to Katharine J. Bar, 502D Johnson Pavilion, 3610 Hamilton Walk, Philadelphia, PA 19104, Tel: 215.573.8497. Fax: 215.573.8976, bark@upenn.edu.

5. No changes in author affiliations.
Figure Legends

Figure 1. NJ and ML phylogenies of HIV-1 gp41 env sequences.  A. NJ phylogeny of SGS-derived gp41 env sequences from the chronically-infected source patient’s plasma vRNA reveals unstructured wide genetic diversity typical of chronic HIV-1 infection.  B. NJ phylogeny of SGS-derived sequences from the acutely-infected index subject’s plasma vRNA and PBMC DNA reveals multiple low diversity lineages typical of acute HIV-1 infection shown in blue (variants 1-14) as well as interspersed interlineage recombinant sequences shown in orange. Sequences derived from PBMCs are indicated with asterisks.  C. ML phylogeny of sequences from the source patient (in black) and the 14 T/F consensus genomes from the index subject (in blue) intersperse within a monophyletic lineage separate from all reference sequences (in gray), including Subtype B sequences from Australia.  Scale bars indicate nucleotide distance.

Figure 2. NJ phylogenetic tree and Highlighter plot analysis of pol sequences.  A. NJ phylogeny of SGS-derived pol sequences from the chronically-infected source patient’s plasma vRNA displaying long branch-lengths, with the exception of the single cluster of closely related sequences. Such sequence clusters of near-identity have been reported as minor components of chronically-infected patients’ plasma virus quasispecies and reflect recent clonal expansion of a virus within a diverse chronic virus swarm [7].  B. NJ phylogeny of SGS-derived pol sequences from the acutely-infected index subject’s plasma vRNA reveals multiple low diversity lineages shown in blue (variants 1-15) as well as interspersed interlineage recombinant sequences shown in orange.  C. A Highlighter plot of index case pol sequences represents as colored tics the nucleotide polymorphisms between the top sequence (43715D4) and the remaining pol sequences aligned below. The plot illustrates the homology of each of the T/F lineages (in blue), the heterogeneity between T/F lineages, and the chimeric structure of the recombinant sequences (in orange).
References:


Figure 2