Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection

Simon D. W. Frost*,†, Terri Wrin*, Davey M. Smith*, Sergei L. Kosakovsky Pond*, Yang Liu‡, Ellen Paxinos‡, Colombe Chappey‡, Justin Galovich‡, Jeff Beauchaine‡, Christos J. Petropoulos‡, Susan J. Little*, and Douglas D. Richman*‡

*Departments of Pathology and Medicine, University of California at San Diego, La Jolla, CA 92039-0679; †Monogram Biosciences, Inc., 345 Oyster Point Boulevard, South San Francisco, CA 94080; and ‡Veterans Affairs San Diego Healthcare System, San Diego, CA 92161

Edited by Malcolm A. Martin, National Institute of Allergy and Infectious Diseases, Bethesda, MD, and approved October 21, 2005 (received for review June 17, 2005)

HIV type 1 (HIV-1) can rapidly escape from neutralizing antibody responses. The genetic basis of this escape in vivo is poorly understood. We compared the pattern of evolution of the HIV-1 env gene between individuals with recent HIV infection whose virus exhibited either a low or a high rate of escape from neutralizing antibody responses. We demonstrate that the rate of viral escape at a phenotypic level is highly variable among individuals, and is strongly correlated with the rate of amino acid substitutions. We show that dramatic escape from neutralizing antibodies can occur in the relative absence of changes in glycosylation or insertions and deletions (“indels”) in the envelope; conversely, changes in glycosylation and indels occur even in the absence of neutralizing antibody responses. Comparison of our data with the predictions of a mathematical model support a mechanism in which escape from neutralizing antibodies occurs via many amino acid substitutions, with low cross-neutralization between closely related viral strains. Our results suggest that autologous neutralizing antibody responses may play a pivotal role in the diversification of HIV-1 envelope during the early stages of infection.

selection | glycosylation

HIV has the potential to evolve at an extremely high rate due to the combination of a high mutation rate (1), a short generation time (2–5), and a large population of productively infected cells within the host (6). Among genes within the HIV-1 genome, the env gene evolves at a particularly high rate (1–2% per year) at the population (7, 8) and the individual level. After infection, genetic diversity in env begins low (9–11), undergoes a drop (12), and then increases to a peak several years into the infection (13). Genetic divergence from the infecting strain increases during infection, finally reaching a plateau several years into infection. Consequently, env is highly genetically diverse, posing a significant challenge to vaccine development.

Although there is evidence that the rapid evolution of env is caused by diversifying selection (14–23), it remains unclear which mechanism is the driving force of envelope diversification. Selection to use the CXCR4 coreceptor plays a part in generating diversity in HIV-1 env; however, CXCR4 usage evolves later into infection, is not observed in all individuals (13, 24, 25), and involves a small number of residues within env (26–29). Evolution of escape to cellular immune responses may also play a role in driving the rapid divergence of env, although cellular responses may be stronger and/or more common to other genes such as gag, nef, and tat, especially during recent HIV infection (30–32). In contrast, selection by neutralizing antibody responses has been shown to result in rapid, continuous in vivo evolution of viral escape at the phenotypic level (33–36), and therefore may contribute to the rapid evolution of HIV-1 envelope.

Escape from neutralizing antibody responses may occur through a combination of point mutations, changes in glycosylation patterns, and insertions and deletions in the viral envelope. In an early study by Wahlberg et al. (37), no correlation between the accumulation of amino acid mutations in the V3 region of env and the rate of phenotypic escape was detected. However, this does not preclude rapid evolution in regions of envelope other than V3. Specific N- and O-linked glycosylation modifications in the envelope V1 domain of simian immunodeficiency virus variants can alter recognition by neutralizing antibodies (38). Preferential transmission of neutralization sensitive virus, containing fewer N-linked glycosylation sites, has been reported by Derdeyn et al. (39) in a study of subtype C HIV-1, although this process may not hold for transmission of subtype B HIV-1 (40, 41). Wei et al. (35) argued for a mechanism of neutralizing antibody escape during recent infection in which a shifting “glycan shield” protects the virus from neutralization, supported by the demonstration in vitro of escape mutants generated by mutating N-linked glycosylation sites. However, multiple mutations were required to generate a neutralization resistant virus, suggesting that changes at N-linked glycosylation sites may be secondary to selective forces driving single amino acid changes. Insertions and deletions in variable loops within the envelope glycoprotein may also contribute to neutralization escape. Although all three mechanisms may contribute to escape from neutralizing antibodies, the relative contribution of each in vivo is poorly understood.

To investigate the underlying genetic basis of neutralization escape in vivo, we compared patterns of genetic variation in terms of amino acid changes, insertions and deletions, and N-linked glycosylation sites between groups of recently HIV-infected individuals with different rates of phenotypic escape from neutralizing antibody responses, and interpreted our results by using a simple model of neutralizing antibody responses developed by Haraguchi and Sasaki (42).

Methods

Study Population. We studied 13 subjects with a diagnosis of recent HIV infection, recruited as part of the San Diego Acute Infection and Early Disease Research Program. All subjects were male; 12

This work was presented in part at the 11th Conference on Retroviruses and Opportunistic Infections, San Francisco, CA, February 8–11, 2004.

Conflict of interest statement: D.D.R. is on the Scientific Advisory Board of Monogram Biosciences, Inc., who performed the neutralization assays and sequencing, and S.D.W.F. is a consultant for Monogram Biosciences, Inc.

This paper was submitted directly (Track II) to the PNAS office. Freely available online through the PNAS open access option.

Abbreviations: PANGS, potential N-linked glycosylation sites; nAb, neutralization antibody.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ296677–DQ297052).

*To whom correspondence should be addressed. E-mail: sfrost@ucsd.edu.

© 2005 by The National Academy of Sciences of the USA
identified themselves as white, one as African American; the median age was 44 (range, 29–64); the median baseline CD4+ cell count was 492 (range, 311–1,119); and the median HIV plasma RNA was 175,909 copies/ml (range, 562–1,805,790).

**Specimen Processing.** Blood samples from peripheral venipuncture in ACD tubes were collected and processed with whole blood. Blood plasma was aliquoted and frozen and stored at −80°C until analysis. Viral loads were determined by using the Roche Diagnostics Amplicor Monitor assay (lower limit of detection of 50 copies/ml).

**Measurement of HIV-Specific Neutralizing Antibody Responses.** HIV-specific neutralizing antibody responses were measured by using a recombinant virus assay as described (34), using virus particles containing patient virus envelope proteins plus an HIV genomic vector that contains a firefly luciferase indicator gene. Virus infectivity was determined by measuring the amount of luciferase activity expressed in infected cells compared with an antibody-negative control after a single cycle of replication. Neutralizing antibody titers were calculated as the reciprocal of the plasma dilution conferring 50% inhibition (IC50). In addition to virus isolated from the study subjects, we also measured heterologous neutralization against the HIV-1 strains NL4–3 (GenBank accession no. AY669735) and JR-CSF (GenBank accession no. AY669726) as well as a virus pseudotyped with a non-HIV envelope (amphotropic murine leukemia virus) as a specificity control.

**Sequencing of the HIV-1 env Gene.** HIV genomic RNA was isolated from patient plasma by using oligo(dT) magnetic beads, and first-strand cDNA was synthesized in a standard reverse transcription reaction using oligo(dT) primers. Env DNA (gp160) was PCR amplified by using forward and reverse primers located immediately upstream and downstream of the env initiation and termination codons, respectively. The forward and reverse primers contain unique recognition sites for PinAI and MluI and ligated into the pCXAS expression vector, which uses the cytomegalovirus immediate-early promoter enhancer to drive expression of the env insert in transfected cells. Ligation products were introduced into competent Escherichia coli cells (Invitrogen) by transformation, and DNA was purified from bacterial culture. An aliquot of each transformation was plated onto agar plates, and colony counts were used to estimate the number of env sequences represented in each library (500–5,000 colonies). DNA was diluted, retransformed, and replated to allow easy picking of individual colonies (usually 100–200 per plate). Sequencing analysis was performed by using a thermo-cycling method with fluorescent dye-labeled deoxyxynucleotide chain terminator chemistry (Applied Biosystems). Sequencing reaction products were resolved by using a 96 parallel capillary gel electrophoresis system (Applied Biosystems 37000). Bulk sequences in one patient, 01-0201, exhibited an unusually high rate of ambiguous nucleotides, and were excluded from the analysis. We also screened the sequences for possible G to A hypermutation events (43). Sequence alignment was performed by using a Smith and Waterman (44) algorithm using affine gap costs (see Supporting Text, which is published as supporting information on the PNAS web site).

**Statistical Analysis.** Analysis of the dynamics of neutralizing antibody responses. To estimate the rate at which neutralizing antibodies changed with respect to the time of the virus and plasma sample, we fitted a semiparametric random effects model (45, 46), to the log-transformed neutralization titers. If \( y(t, P, V) \) is the neutralization antibody titer in patient \( i \) using a plasma sample obtained at time \( t_P \) and a virus sample obtained at time \( t_V \), we fitted the model

\[
\log(y(t_P, t_V)) = a_i + b t_P + c t_V + e(t_P - t_V) + e_i(t_V),
\]

where \( a \) is the individual-level intercept, \( b \) is the slope of neutralizing antibody titers with respect to the time of plasma sampling, \( c \) is the slope of neutralizing antibody titers with respect to the time of virus sampling, \( e(t_P - t_V) \) is a smoothing term used to capture nonlinear antibody dynamics, and \( e_i(t_V) \) is the residual error term. A thin plate smoothing spline basis was used for the smoothing term (47). Variation in the parameter values \( a, b, c \) and across individuals was modeled by using a multivariate normal distribution. Within-individual correlations in the residual error term were modeled by using a linear spatial correlation structure. The model was implemented in the statistical programming language R (48) using the NLME (49) and ASSIST (50) libraries.

**Sequence analysis.** We estimated the nonsynonymous to synonymous substitution rate ratio, \( \omega \), using maximum likelihood for different regions of the envelope sequence, based on an MCMC × 100020 model of codon substitution (51). For the clonal sequences obtained from two individuals, we estimated the average nonsynonymous and synonymous divergence between the baseline sample and subsequent viral samples by using the distance-based method of Nei and Gojobori (52), as implemented in the program HYPHY (53). The change in divergence over sampling time was fitted by using a generalized linear model using quasi-likelihood, with a identity link and a \( \mu(1 - \mu) \) variance function (54). Potential N-linked glycosylation sites (PNGS) were identified by using a NXXY motif, where X is any amino acid except proline and Y is either serine or threonine (55–57).

**Mathematical Model of Neutralizing Antibody Responses.** We modeled the evolution of escape to neutralizing antibodies using the model of Haraguchi and Sasaki (42). Mathematical details are given in Supporting Text. The model was simulated by using the program XPP-AUT (58).

**Results.** As previously reported (34), HIV-specific neutralizing antibody responses varied dramatically among study subjects (Fig. 1 a–c). Neutralization antibody (nAb) titers were low for contemporary virus and plasma samples for all patients. Titers generally increased over successive plasma samples for a given virus that had been circulating previously in that patient, and generally decreased over successive virus samples for a given plasma sample. The rate of increase of nAb responses over successive plasma samples to a given autologous virus isolate provides a measure of the speed at which nAb responses are mounted; correspondingly, the rate at which nAb responses decrease against successive autologous virus isolates gives a measure of the rate of viral escape from nAb responses. We fitted a model to the log-transformed neutralization titers, which included the effect of sampling time for virus and plasma and a smoothing term that allowed deviations from exponential growth/decay in neutralization titer. Our model demonstrated a strong, significant correlation between the rate of mounting of nAb responses and the rate of phenotypic escape from nAb (correlation coefficient = −0.87, 95% confidence interval = −0.56 to −0.97) (Fig. 1d), consistent with the low, constant neutralizing titers observed between contemporary virus and plasma samples. There was a significant deviation from exponential changes in neutralization titer (\( P < 0.001 \)), apparently because of the plateauing of neutralization titers in some patients for later plasma samples when tested against early autologous viral isolates.

Using our statistical model, we classified individuals into those with high and those with low rates of viral phenotypic escape from nAb, relative to the mean rate of escape (Fig. 1d). Heterologous neutralization to the HIV-1 strain NL4–3 increased over time (\( P = 0.002 \)). There was no significant difference in heterologous neutralization of the viral strains NL4–3 and JR-CSF between individuals with high or low rates of viral escape (Fig. 5, which is published as supporting information on the PNAS web site). The slope of viral load over time was not significantly different from zero (\( P = 0.95 \)), and although viral loads were 0.52 log_{10} copies higher...
in the individuals with rapid viral escape at baseline, this was not statistically significant ($P = 0.28$).

To elucidate the genetic correlates of escape, we compared patterns of genetic variation between individuals whose virus exhibited either rapid or slow escape from neutralizing antibody responses. For each patient, we obtained a pair of population-based full-length *env* sequences sampled ~1 year apart (median interval, 335 days; range, 182–433 days), corresponding to the first year of infection. Because bulk sequences may not be reliable indicators of genetic variation in areas with large numbers of insertions and deletions ("indels"), we also obtained multiple clonal sequences of the V1-V2 and the V4 region of envelope, both of which exhibit indels when comparing HIV sequences between individuals, at baseline and follow-up (median, nine per patient per time point).

We also obtained full-length clones of virus from patient 01–0127 (72 sequences isolated from nine time points over 740 days of follow-up), whose virus exhibited a high rate of phenotypic escape from nAb responses, and patient 01-0083 (134 sequences isolated from 13 time points over 1,098 days of follow-up), whose virus exhibited a low rate of escape.

One possible genetic mechanism for escape is point mutations of amino acids. Under this hypothesis, a high rate of phenotypic escape is correlated with a high rate of amino acid substitutions in the viral envelope. We estimated the ratio of nonsynonymous to synonymous substitution rates, $\omega$, in individuals with slow or rapid viral escape using maximum likelihood. $\omega$ was significantly higher in individuals with rapid viral escape [8.1 (95% confidence interval, 6.5–10.0) vs. 0.6 (95% confidence interval, 0.4–0.9), $P < 0.00001$] over the whole envelope gene, and over gp120 alone [5.7 (95% confidence interval, 4.4–7.2) vs. 0.67 (95% confidence interval, 0.44–0.98), $P < 0.001$] (Fig. 2a). There were significant differences in $\omega$ between the groups in the constant regions, C1–C5 (5.7 vs. 0.7, $P < 0.001$) and variable regions, V1–V5 (5.3 vs. 0.6, $P < 0.001$), of gp120. Divergence was too low in gp41 to make reliable inferences.

To investigate temporal patterns of evolution in the *env* gene, we obtained clonal sequences from a patient with strong antibody responses (patient 01-0127), and one with weak antibody responses (patient 01-0083). We estimated average pairwise $dN$ and $dS$ between the viruses isolated at baseline and subsequent viral samples. There was a linear correlation between divergence and sampling time (Fig. 3a and b). The rates of nonsynonymous divergence were extremely similar in both patients (0.41 and 0.44% per year in patients 01-0127 and 01-0083, respectively). However, although the rate of nonsynonymous divergence was lower than the synonymous substitution rate in patient 01-0083, with slow viral escape (0.32 vs. 0.44% per year; $P < 0.05$), the rate of nonsynonymous divergence was significantly higher than the synonymous divergence rate in patient 01-0127, with rapid viral escape (1.19 vs. 0.41%; $P < 10^{-8}$). Sites under positive selection were detected in both patients (15 sites in each at the $P = 0.05$ level), and were scattered across the envelope gene (Fig. 3c and d).

A second mechanism for escape from nAb responses involves changes in the glycosylation of the viral envelope (35). However, there was no significant difference in either the total number of PNGS (Fig. 2c; Wilcoxon test, $P = 0.29$) or the number of changes at PNGS between individuals with high and individuals with low rates of escape from nAbs (Fig. 2b). We calculated the number of PNGS for each clone for the two patients for whom we had time series of clonal sequence data. Although there was a long-term change in the number of PNGS in both patients (Fig. 3e and f), patterns of glycosylation were relatively static during the first year of follow up in both patients, despite dramatic phenotype escape from neutralizing antibody responses in one of the patients.

Insertions and deletions ("indels") may also play a role in escape from nAbs, especially in the V1–V2 and V4 regions of envelope. We
generated multiple clonal sequences of these regions, and found no statistically significant associations between the rate of viral escape and the initial length, the final length, or the change in length of the V1–V2 region (Fig. 2c and d) or V4 (results not shown). Because clonal sequence analysis may not detect low-frequency variants, we also used a quantitative PCR technique to estimate the frequency of length variants in the V1–V2 and V4–V5 region (see Supporting Text), which confirmed the lack of association of length variation with rates of neutralization escape (Fig. 6, which is published as supporting information on the PNAS web site). Analysis of patient 01-0127 demonstrated that the pattern of length variation in V1–V2 and V4 remained relatively stable over the first year of infection, despite significant escape from neutralizing antibody responses (Fig. 3g), which was confirmed by using quantitative PCR analysis (Fig. 7, which is published as supporting information on the PNAS web site).

We compared the pattern of evolution of env in vivo with the predictions of a simple mathematical model of viral evolution in response to selection by neutralizing antibodies (42). When only a small number of viral mutations are necessary to escape from the current predominant neutralizing antibody response, the viral load reaches a steady state, despite ongoing viral escape (Fig. 4a). The viral divergence from the infecting strain increases linearly over time, following a brief period of no change (Fig. 4c); a similar observation has been reported in a study of SIV-infected macaques (59). Consistent with our data, the neutralization of contemporary strains in the model remains low, and the neutralization of the infecting viral population increases to a plateau higher than the neutralization of contemporary viruses (Fig. 4e). In contrast, when
a large number of mutations are required for escape, which may be the case for escape by changes in glycosylation (35), the model dynamics of virus and neutralization differ markedly from our data: the viral load undergoes oscillations (Fig. 4b), genotypic divergence from the infecting viral population increases in a step-wise fashion (Fig. 4d), and the neutralization of both the infecting viral strain and the contemporary viral population undergoes fluctuations (Fig. 4f).

Discussion

We studied the patterns of genetic variation within full-length sequences of the HIV-1 envelope gene in individuals with recent HIV infection, and correlated patterns of genetic evolution with the rate of escape from neutralizing antibodies at the phenotypic level. Consistent with other studies (60–63), our study population exhibited wide variation in the ability to neutralize early autologous isolates. The rate of evolution in terms of amino acid substitutions, averaged over the whole viral envelope, correlated well with the rate of phenotypic escape, but we failed to demonstrate a correlation between the rate of escape from neutralizing antibodies and the rate of evolution of N-linked glycosylation sites and insertions and deletions. In a detailed longitudinal analysis of viral genetic variation from two individuals, one with rapid and one with slow viral escape from neutralizing antibody responses, we showed that rapid escape from neutralizing antibody responses can occur in the relative absence of changes in length variation or glycosylation patterns. In contrast, the continuous, ongoing accumulation of amino acid mutations spread across the envelope mirror the continuous, ongoing nature of viral escape. Our data are also consistent with observations in SIV-infected macaques (64). A simple model of escape from neutralizing antibody responses (42) further supports a pivotal role for these responses in driving diversification of envelope, despite constant viral loads and low neutralization titers of contemporary strains.

The development of neutralizing antibody responses occurred early, in contrast to other studies (60, 63), perhaps because of the higher sensitivity of the assay used in this study. Because of the similar rates of mounting of neutralizing antibody responses and viral escape from these responses, the titers of neutralizing antibody remain low and constant for matched plasma and virus samples, an example of “Red Queen” evolutionary dynamics (65). Viral loads remain relatively constant throughout the period of escape, consistent with mathematical models (42) and data from HIV-infected children (36) and HIV-infected hu-PBL-SCID mice (66). Hence, neutralizing antibody responses exert “soft” selection pressure (67), by affecting the relative fitness of different strains present in the viral quasispecies. In contrast, cellular immune responses may exert “hard” selection pressure (67), in which the absolute fitness of the population is affected, suggested by the correlation of CTL responses and class I HLA alleles with viral load (68, 69).

We found no correlation between the rates of insertions and deletions and escape from neutralizing antibody responses. Analyses of clonal V1–V2 and V4 sequences demonstrate that patterns of length variation can remain stable even in the presence of rapid escape; conversely, dramatic changes in the pattern of length
variation can occur in individuals who do not exhibit detectable neutralizing antibody responses; this does not imply that such changes have no effect on escape, rather these effects may be secondary to changes at the amino acid level. Similarly, the rate of change by neutralizing antibodies may also occur in individuals who do not exhibit detectable secondary to changes at the amino acid level. Similarly, the rate of such multiple substitutions are unlikely in the absence of selection for individual amino acid changes, evolution at PGNs may be secondary to compare to point mutations, and may give rise to continuous bursts of viral escape, rather than the continuous patterns observed (42).

Identification of individual amino acid changes contributing to escape is complicated by diversity between individuals in terms of both the genetic sequence of the infecting virus and the composition of neutralizing antibodies. We found that differences in selection pressure between individuals with rapid and those with slow viral escape were spread across the whole envelope sequence. Identifying which mutations are contributing to escape, which may act as compensatory mutations, and which

emerge simply because they are linked to advantageous mutations is an important area for further research. Diversifying selection by neutralizing antibodies may also occur in individuals with antibody responses that are limited to the N-terminal domain (~1:45 dilution for a 50% reduction in viral replication). Identification of signatures of neutralization escape at a sequence level may reveal that neutralizing antibodies may have an even more important role in driving HIV-1 envelope diversification than suggested by our data. We thank Drs. Chunlei Ke (St. Jude Medical Center, Fullerton, CA) and Yuedong Wang (University of California, Santa Barbara) for providing an early version of their assist software and invaluable advice and Dr. G. Bard Ermentrout for assistance with XPP-Aut. This work was supported by National Institutes of Health Grants AI27670, AI29164, AI43638, AI47745, AI55276, and AI57167, Adult AIDS Clinical Trials Group Grant AI38858, the Universitywide AIDS Research Developmental Program (IS02-SD-701), University of California at San Diego Center for AIDS Research Developmental Grant AI36214 (to D.W.F.), and the Research Center for AIDS and HIV Infection of the San Diego Veterans Affairs Healthcare System. The development of HIV envelope assay and sequence analysis systems is supported in part by National Institutes of Health Small Business Innovative Research Grants AI48990 and AI57068 (to Monogram Biosciences).