MAJOR ARTICLE

Positive In Vivo Selection of the HIV-1 Envelope Protein gp120 Occurs at Surface-Exposed Regions

Beda Joos,[1] Marek Fischer,[1] Andreas Schweizer,[1] Herbert Kuster,[1] Jürg Böni,[2] Joseph K. Wong,[3] Rainer Weber,[1] Alexandra Trkola,[1] and Huldrych F. Günthard[1]

[1]Department of Medicine, Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zürich, and [2]Swiss National Center for Retroviruses, Zürich, Switzerland; [3]San Francisco Veterans Affairs Medical Center, University of California, San Francisco

The rapid evolution of human immunodeficiency virus (HIV) envelope represents a major challenge to vaccine and drug development, particularly because the underlying mechanisms are not completely understood. To explore whether distinct patterns of positive selection within the envelope glycoprotein (gp) 120 exist and are associated with functionally relevant domains, we conducted a long-term survey of sequence evolution in 20 HIV-1–infected persons who interrupted antiretroviral therapy. In total, 1753 clonal sequences encompassing the C2-V3-C3 region of gp120 were derived. Strikingly, positively selected amino acids mapped almost exclusively ($P = .0003$) to externally accessible residues on the gp120 crystal structure. The current understanding of envelope structure and function associates the main determinants of viral entry and the targets for neutralizing antibodies with these exterior regions of gp120, strongly suggesting that the observed adaptive evolution of these sites occurs in response to respective selective forces.

HIV evolves rapidly in vivo because of its high replication rate of 1×10^8–1×10^{10} per day [1] and because of the relatively poor fidelity of the reverse transcription (RT) process. Compelling evidence exists indicating that intrahost evolution is not random but is driven by continuous selection pressure from immune responses [2]. However, the relative contributions of neutralizing antibodies, helper T cells, and cytotoxic T lymphocytes (CTLs) in controlling viral replication in vivo and inducing selection pressure are incompletely understood.

Among the different HIV genes, *env*, which encodes the viral envelope proteins gp120 and gp41, exhibits the highest evolutionary rate, with increases of ~1% per year in divergence and diversity in the C2-V5 region of gp120 [3]. Positively selected sites were frequently found within C2-V3-C3 in earlier cross-sectional studies [4–6]. The third variable loop (V3) of the HIV-1 envelope protein gp120 contains features essential for HIV binding to its receptors and is a determinant of viral tropism and coreceptor usage (CXCR4 or CCR5) [7, 8]. Type-specific neutralizing responses directed against the V3 loop are commonly found in infected individuals [9, 10] and have been suggested to drive the high evolutionary rate of this variable region [3].

We previously studied viral diversity [11] in patients with chronic HIV-1 infection who underwent controlled treatment interruption after receiving antiretroviral treatment for 16–40 months [12–16]. In this cohort, lower pretreatment diversity was associated with spontaneous control of viremia, reduced viral replication capacity, and higher neutralizing antibody activity against autologous virus but not with the breadth or magnitude of the CTL response [11]. Longitudinal evolutionary analysis suggested that stronger autologous neutralizing antibody responses contributed

Received 13 November 2006; accepted 18 February 2007; electronically published 5 June 2007.

Potential conflicts of interest: none reported.

Presented in part: 15th International HIV Drug Resistance Workshop, Sitges, Spain, 13–17 June 2006 (abstract 67).

Financial support: Swiss National Science Foundation (grants 3345–65168 and 3100AO-103748/1 to H.F.G. and A.T. and grant PP00B-102647 to A.T.); EMDO Foundation (grant to H.F.G.); Kanton of Zürich (research grant to the study); US National Institutes of Health (grant R01 NS51132 to J.K.W.).

Reprints or correspondence: Dr. Beda Joos, University Hospital Zürich, Dept. of Medicine, Div. of Infectious Diseases and Hospital Epidemiology, Rämistrasse 100, CH-8091 Zürich, Switzerland (beda.joos@usz.ch); or Dr. Huldrych Günthard, University Hospital Zürich, Dept. of Medicine, Div. of Infectious Diseases and Hospital Epidemiology, Rämistrasse 100, CH-8091 Zürich, Switzerland (huldrych.guenthard@usz.ch).

The Journal of Infectious Diseases 2007;196:313–20

0022-1899/2007/19602-0020$15.00

DOI: 10.1086/518935

Table 1. Patient characteristics.

Patient[a]	Observation period, years	Time without ART, years	Pretreatment diversity[b]	Divergence over time[b]	Adaptation rate[c]
105	7.6	4.2	0.68	1.44	0.100
106	7.8	4.1	0.71	1.42	0.117
107	6.5	3.3	2.12	1.67	0.066
109	3.8	0.3	4.02	2.05	0.000
111	3.2	1.1	4.66	1.29	0.240
112	4.1	0.8	1.61	0.13	0.000
113	7.5	3.9	2.27	1.66	0.201
114	6.7	3.2	2.74	2.09	0.246
115	6.0	3.2	1.60	0.97	0.134
117	7.1	3.2	0.88	0.33	0.050
118	7.1	3.6	0.94	1.97	0.163
119	3.3	1.2	1.60	0.81	0.161
120	6.2	3.1	2.97	1.02	0.110
121	3.8	1.4	1.23	0.97	0.233
122	7.4	4.0	3.34	1.58	0.136
123	6.9	3.2	2.87	4.46	0.177
124	4.2	1.9	2.93	2.51	0.188
126	6.9	3.0	3.87	1.80	0.186
127	3.9	1.1	2.95	0.13	0.081
128	6.6	3.5	4.90	2.35	0.096
Median	6.6	3.2	2.51	1.51	0.135

[a] Other characteristics, such as viral load, CD4 cell count, coreceptor usage, and virus subtype, have been reported elsewhere [11, 15]. Identification codes correspond to those given previously.
[b] Genetic distance estimates were obtained by use of the Tamura-Nei 6-parameter model, and analyses were conducted by use of MEGA software (version 3.1) [18].
[c] Total no. of adaptive events per month, as determined by the method of Williamson [19].

to virus control in a subset of these chronically infected patients [17].

Because our previous observations suggested a potential role of the neutralizing antibody response in restricting viremia in vivo and, hence, driving HIV envelope evolution, we conducted in the present study a long-term survey of the sequence evolution of the envelope region (spanning the V3 loop and its flanking domains) in a cohort of 20 patients. We thereby aimed to formally determine whether positive selection occurs, whether hotspots for positive selection can be located, and whether these mutated codon sites fall within structurally relevant domains. We found that positively selected amino acid changes occurred predominantly at exposed locations on the virion surface, conferring high external accessibility. This suggests that these sites evolved in response to selective pressure induced by neutralizing antibodies. The structural model of gp120 converged remarkably well with our evolutionary analysis in HIV-1–infected patients. Moreover, the results confirm the utility of maximum likelihood–based genetic algorithms in accurately detecting and monitoring selection pressure.

PATIENTS, MATERIALS, AND METHODS

Nomenclature of HIV envelope substructures. The HIV-1 envelope glycoprotein consists of the surface moiety gp120, which is bound to the transmembrane protein gp41. The primary structure of gp120 can be subdivided into 5 relatively conserved regions and 5 variable loops (C1-V1-V2-C2-V3-C3-V4-C4-V5-C5, from amino to carboxy terminus).

Subject selection and study design. In total, 29 patients were enrolled at the University Hospital Zürich into the Swiss-Spanish Intermittent Treatment Trial [12–14]. In the present analysis, we included those who remained without antiretroviral therapy after completion of the trial provided that polymerase chain reaction (PCR) amplification was successful. Detailed patient characteristics [11, 15] (see also table 1) and results of the clinical trial [12] have been reported elsewhere. Written informed consent was obtained from all patients, in accordance with the guidelines of the Ethics Committee of the University Hospital Zürich.

RNA extraction, RT-PCR, cloning, and sequencing. RNA was extracted from 0.8–0.9 mL of plasma; subsequently, for

each time point, 16 representative clonal sequences spanning the C2-V3-C3 region of HIV-1 *env* were derived as described elsewhere [11]. Overall, 109 longitudinal samples from 20 patients were analyzed, resulting in a total of 1753 clonal sequences (31–190/patient) obtained from before antiretroviral therapy as well as during and after treatment interruptions. The average misincorporation rate was 0.18%, as tested by amplification of HIV-1 strain YU2 (GenBank accession number M93258) [11]. The reproducibility of sequencing was shown by duplicate analysis of 32 clones (error rate, 0/26,000 bp). The accuracy of the method was tested by analyzing 2 patient samples (1 of high diversity and 1 of low diversity) in parallel by limiting-dilution cloning, resulting in almost identical quasispecies composition.

Phylogenetic analyses. Sequences were edited and aligned by use of Lasergene software (version 5.08; DNASTAR). Alignments were manually corrected to adjust sequence gaps with the reading frame. Phylogenetic analyses were conducted by use of MEGA software (version 3.1) [18]. Genetic distances were estimated by use of the Tamura-Nei 6-parameter model. A neighbor-joining phylogenetic tree was constructed by use of MEGA using the sequences of HIV strains HXB2 (GenBank accession number K03455) and JR-FL (GenBank accession number U63632) as references and by bootstrapping (500 replications). Screening for recombination was performed on the data from each patient individually by use of single-breakpoint analysis and the Bayesian information criterion of Schwarz (HyPhy software package; version 0.99 [20]). No evidence of recombination was found in any of the data sets.

Positive selection detection. Average numbers of adaptive changes over the entire gene fragment were determined according to the method of Williamson [19]. Site-by-site identification of positively selected positions was assessed by use of single-likelihood ancestor counting, fixed-effects likelihood, and random-effects likelihood methods, executed in the HyPhy software package by use of the algorithms proposed by Kosakovsky Pond and Frost [21, 22]. Briefly, the optimal time-reversible substitution model was determined for each patient after removal of duplicate sequences; then, depending on the size of the data set, each or a subset of the maximum likelihood–based analyses were performed on the datamonkey platform.

Surface accessibility. Solvent-accessible surface area was quantified by use of the AREAIMOL program (CCP4i, version 5) [23, 24], employing a solvent-sphere radius of 1.4 Å. Relative solvent accessibility was calculated as the accessible area divided by the maximum accessible area of the amino acid in an extended tripeptide conformation (Ala-X-Ala) [25]. Residues with solvent accessibility of 15% or more were considered to be surface exposed.

Glycosylation sites. Potential N-linked glycosylation sites were determined by use of the N-glycosite tool from the Los Alamos HIV sequence database (available at: http://www.hiv.lanl.gov/content/hiv-db/GLYCOSITE/glycosite.html).

Statistical analyses. The number of sequences per patient was variable. Determination of the substitution model, screening for recombination, and detection of positive selection was performed individually for each patient. Spearman's correlation with the Bonferroni adjustment for multiple comparisons was used to test the association between the number of positively selected sites and various patient-specific parameters. Finally, all positively selected sites were pooled for mapping to structural data and solvent-accessible areas. Comparison of the presence or absence of positive selection for both buried ($n = 46$) and surface-exposed ($n = 69$) residues was made by 2-tailed Fisher's exact test. Three positions were excluded (310, 311, and 356), because no surface-accessibility data were available from the respective gp120 structure of JR-FL (Protein Data Bank accession code 2B4C) [26]. The average difference between the nonsynonymous substitution rate and the synonymous substitution rate ($d_N - d_S$) for buried and exposed residues were compared by use of the 2-sided Mann-Whitney *U* test. The χ^2 test and the parametric unpaired *t* test gave comparable results; however, we did not test for normality of distribution. GraphPad Prism (version 4; GraphPad Software) was used for statistical analyses.

Nucleotide sequence accession codes. All reported sequences have been deposited in GenBank and were assigned accession numbers EF424791–EF426096, DQ002058–DQ002345, AY656534–AY656549, AY656440–AY656471, AY656249–AY656312, AY375663–AY375678, AY375616–AY375630, and AY375568–AY375583.

RESULTS

Long-term evolution of HIV envelope C2-V3-C3 in vivo. To study envelope evolution, a total of 1753 clonal sequences of the gp120 C2-V3-C3 region were obtained from the plasma of 20 HIV-1–infected patients over a median of 6.6 (range, 3.2–7.8) years by use of an established RT-PCR and cloning strategy [11]. During the observation period, the patients were enrolled in a clinical trial of treatment interruption after receiving antiretroviral treatment for a median of 2.7 (range, 1.4–3.3) years. The median antiretroviral drug–free interval of the cohort was 3.2 (range, 0.3–4.2) years (table 1). Phylogenetic reconstruction revealed highly specific clustering of individual patient sequences, excluding cross-contamination between subjects and contamination with laboratory strains (figure 1). The inferred net divergence of intraindividual sequences from the first (pretreatment) and last time points sampled averaged 1.5% (range, 0.1%–4.5%). This corresponds to an average increase in viral population divergence of 0.8% per year during the period with-

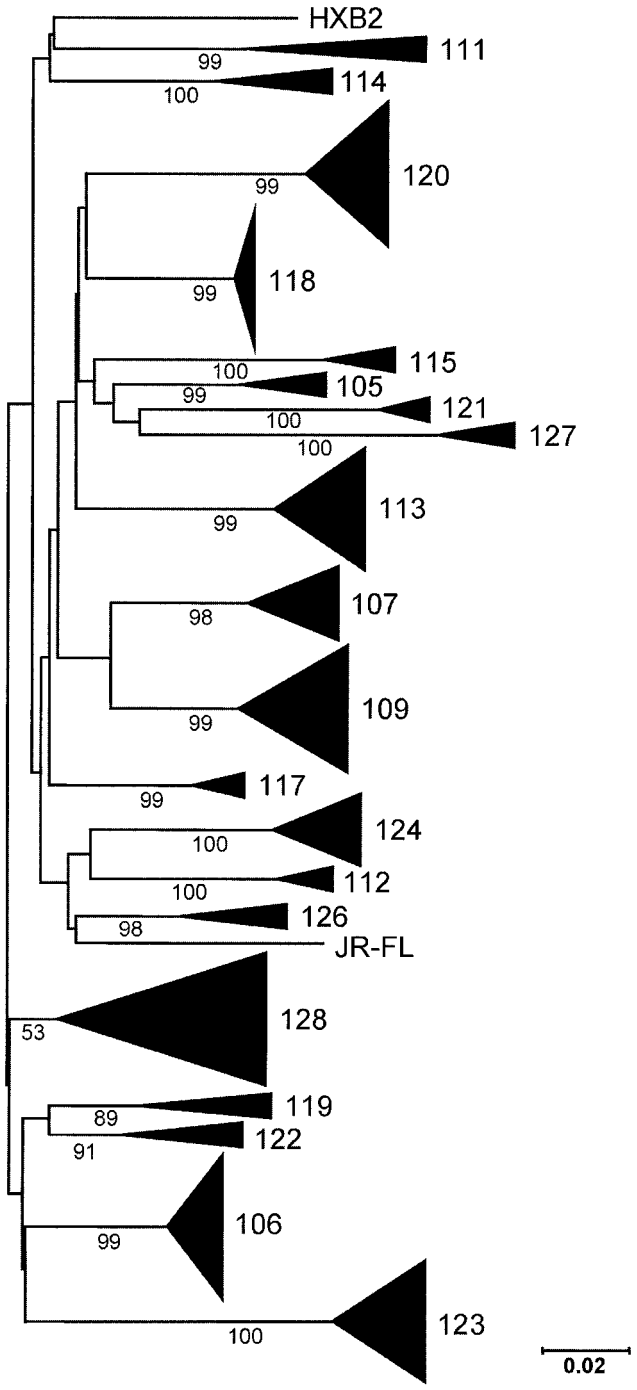

Figure 1. Phylogenetic tree. Shown is an inferred neighbor-joining phylogenetic tree of clonal HIV-1 *env* C2-V3-C3 sequences derived from 20 patients. Elongated triangles represent the compressed subtrees containing 31–190 clones isolated from the plasma of each patient before antiretroviral therapy and after treatment interruption. The length of the triangle corresponds to the respective intrapatient diversity; its thickness is proportional to the no. of taxa. HXB2 and JR-FL were included as external reference strains. The bar denotes 2% nucleotide divergence and diversity (determined by use of MEGA software); bootstrap values are indicated below the branches and correspond to 500 replications.

out antiretroviral therapy and is in agreement with previously reported divergence rates of 1% per year [3].

Detection of positive selection. To quantify the positive selective pressure on the envelope C2-V3-C3 region, we defined patient-specific adaptation rates as the number of adaptive changes introduced per time interval compared with the pretreatment consensus sequence, according to the method of Williamson [19]. This conservative approach is based on counting the total number of nonsynonymous substitutions and adaptive mutations that have reached a high frequency (>50%) and corrects for selectively neutral changes. We found an average of 4.6 (range, 0–10) adaptive events evolving between the first and last time point analyzed. The corresponding rate of 0.135 (range, 0–0.25) adaptive changes per month of untreated infection closely reflects the rate of 0.106 changes per month determined in HIV *env* C2-V3 sequences in adult chronic infection [19]. To identify positions where positive selection had occurred, we next evaluated the in vivo amino acid changes in the C2-V3-C3 region on a site-by-site basis by use of single-likelihood ancestor counting, fixed-effects likelihood, and/or random-effects likelihood methods, executed in the HyPhy software package [21]. On the basis of the algorithms proposed by Kosakovsky Pond and Frost [21, 22], 3 (range, 0–10) amino acid sites per patient were subject to positive selection on average. Across all 20 patients, 35 positively selected positions were detected (figure 2).

In some patients, only the result of the most conservative single-likelihood ancestor counting method could be obtained, because the data set was too large to run the computationally more complex methods. In contrast, for data sets of small or moderate size, the fixed-effects and random-effects likelihood methods appeared to yield more power. However, when only amino acids identified as positively selected by single-likelihood ancestor counting were examined, the conclusions of the statistical analyses remained unchanged (i.e., P values remained $<.05$). This confirms the finding of Kosakovsky Pond and Frost [21], who analyzed simulated data as well as HIV-1 *env* and *pol* sequences and found that the 3 methods arrived at similar conclusions.

No correlation was found between the number of positively selected amino acid sites and plasma viral load observed over time in these 20 patients (pretreatment, rebounds during treatment interruption, postinterruption plateau, and area under the curve). We also evaluated neutralizing antibody titers, CTL response, helper T cell response, gp120 titers, HIV DNA levels in peripheral blood mononuclear cells, and duration of HIV infection; a variety of genetic host polymorphisms in the genes for CCR2, CCR5, CX3CR1, RANTES, macrophage inflammatory protein–1α, stromal-derived factor–1, and interleukin-10; and HLA types obtained in our previous study [11], but none of these parameters was found to be associated with the

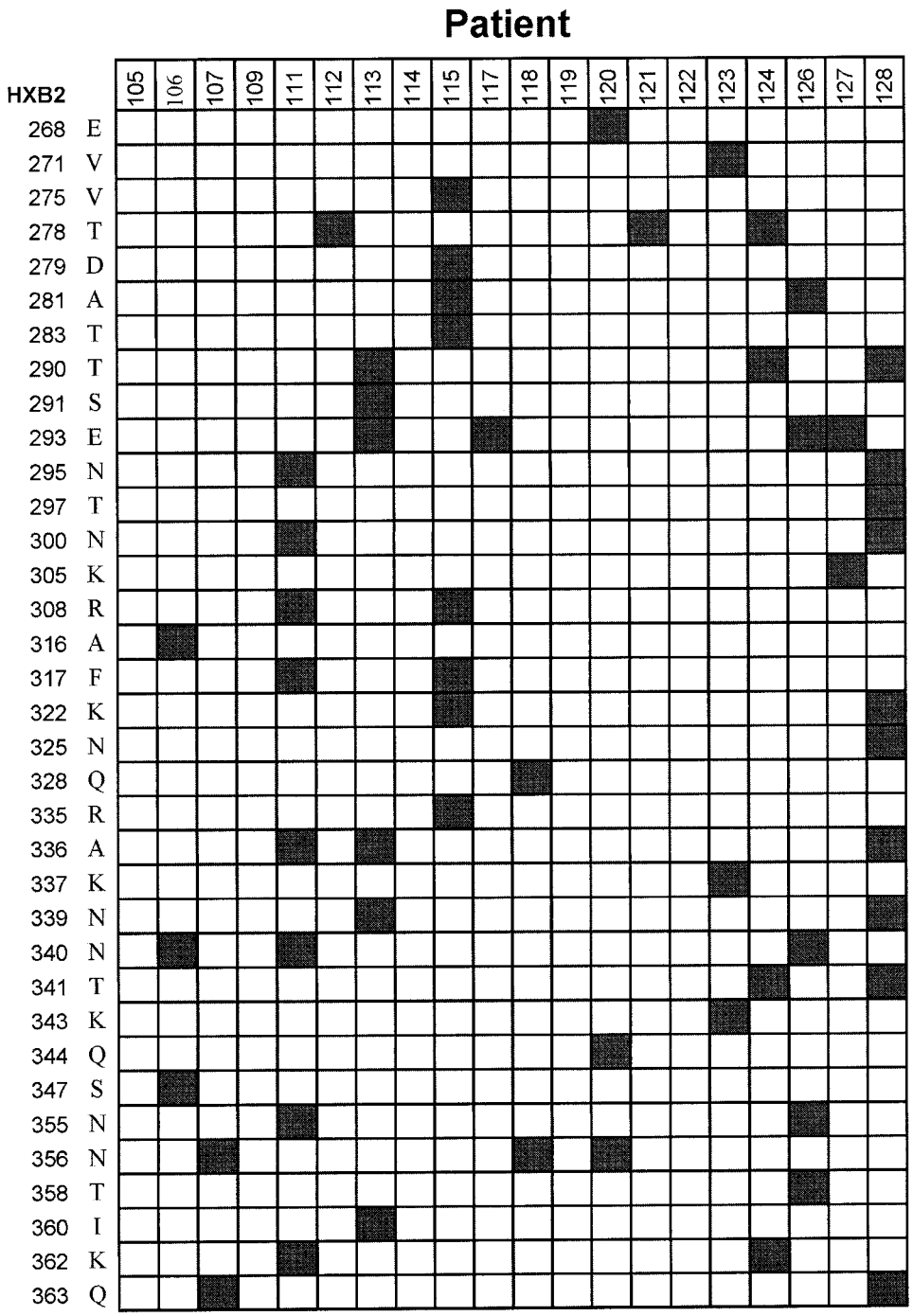

Figure 2. Positive selection. Shown are positively selected sites *(gray)* in the gp120 C2-V3-C3 region, as determined by HyPhy software [20, 21]. The analysis included 1753 clonal sequences obtained from the plasma of 20 HIV-1–infected patients. Position nos. and amino acid residues relative to HXB2 gp120 are given for reference.

degree of positive selection. However, this analysis was based on data limited to a relatively short period of time. Larger longitudinal studies would be required to define such correlations.

Mapping to gp120 structure. Although structural information on gp120 has been limited, the crystal structure of the V3-containing envelope glycoprotein core was recently resolved [26]. Truncated gp120 of the R5-tropic strain JR-FL was crystallized in complex with 2 domains of the CD4 receptor and the Fab fragment of an antibody (X5) to the coreceptor binding site. The resulting structural data allowed us to analyze amino acid residues that had evolved under positive selection in vivo in terms of their spatial orientation. Strikingly, most residues

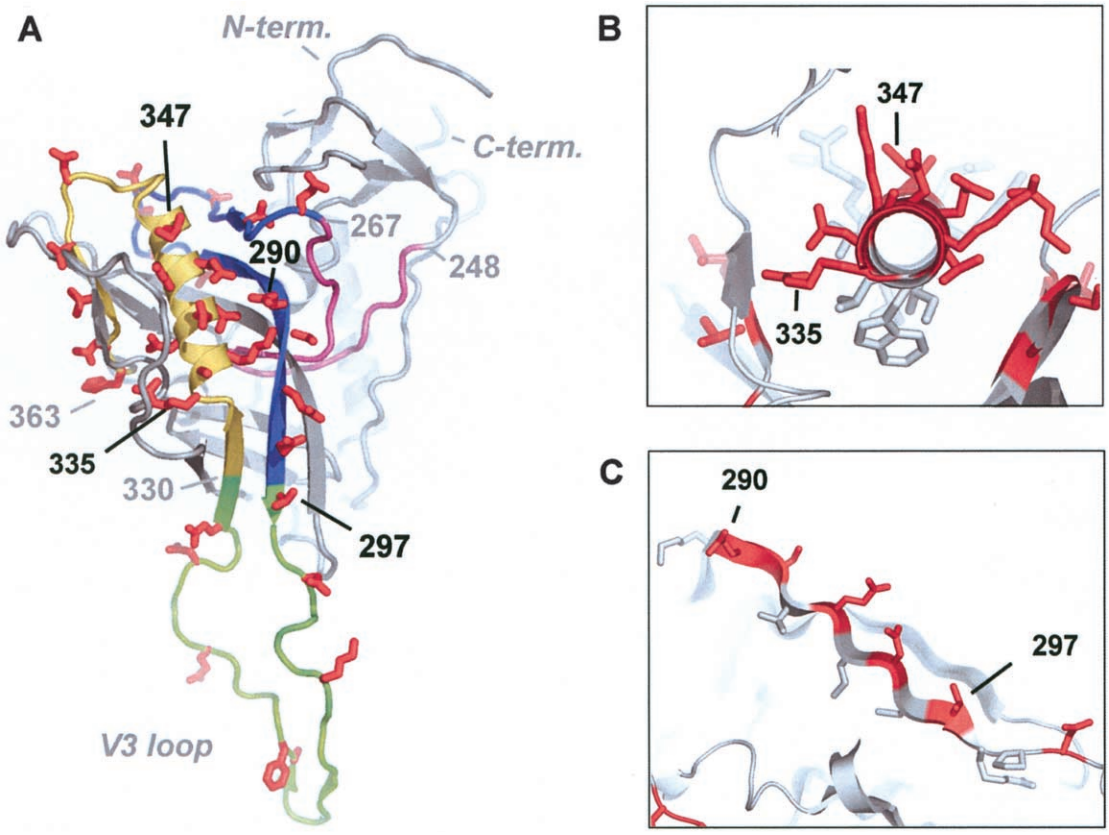

Figure 3. Structure of the HIV envelope protein gp120. Most positively selected amino acids are found at solvent-exposed positions. *A,* Ribbon representation of gp120 JR-FL (Protein Data Bank accession code 2B4C) [26]. Side chains of positively selected amino acids are colored red and are shown as stick models. The stretch of amino acids analyzed in the present study is subdivided into 4 regions, as follows: residues 248–267 *(magenta),* a region that is mostly buried within the glycoprotein; residues 268–296 *(blue),* the N-terminal flank of the V3 loop; residues 297–330 *(green),* the V3 loop; and residues 331–363 *(yellow),* the C-terminal flank adjacent to the V3 loop. *B,* α-Helix, comprising residues 335–355 and viewed along the helical axis. Side chains of amino acids belonging to the helix are shown as stick models, with positively selected residues highlighted in red. *C,* β-Strand, comprising residues 290–297. For clarity, residues 325–347 were removed in this representation. The figure was generated by use of the program PyMOL (available at: http://www.pymol.org/).

at the positively selected sites mapped to solvent-exposed surfaces of the gp120 core (figure 3). Detailed analysis of the spatial positioning of the positively selected amino acids showed that the majority of the changed residues extended into the surrounding space, rendering them available for interactions with the viral receptors or neutralizing antibodies.

To further visualize and quantify the surface accessibility within the C2-V3-C3 of the gp120 core structure model, we assessed the solvent-accessible surface area by use of the method of Lee et al. (AREAIMOL; CCP4i version 5), employing a solvent-sphere radius of 1.4 Å [23, 24]. The accessible area of a given amino acid residue is thereby calculated by a "rolling ball" algorithm, which performs a virtual screen of the 3-dimensional protein structure by use of a probe sphere with the approximate diameter of a water molecule. A residue was considered to be exposed if its relative solvent accessibility (defined as the ratio of the accessible area within the 3-dimensional structure to that in an extended tripeptide conformation [Ala-X-Ala] [25]) was 15% or more. Overall, in our cohort higher rates of nonsynonymous than synonymous substitutions were found to occur at positions coding for the surface-exposed amino acids of the JR-FL gp120 crystal structure model (average of normalized $d_N - d_S$ by single-likelihood ancestor counting; Mann-Whitney $P = .002$). Positively selected changes identified by at least 1 of the 3 likelihood methods were associated almost exclusively (29/34 evaluable sites; Fisher's exact $P = .0003$) with amino acid positions that had high external accessibility (figure 4). Of note, the frequency of positive selection was highest at the locations flanking both ends of the V3 loop (26/62 positions, compared with 9/32 in the V3 loop). The almost exclusive colocalization, as determined by maximum-likelihood algorithms, of positively selected amino acid residues with positions of high surface accessibility in the 3-dimensional structure further validates the use of bioinformatics in sequence analysis and underscores the potential of these methods in detecting biological relevant changes.

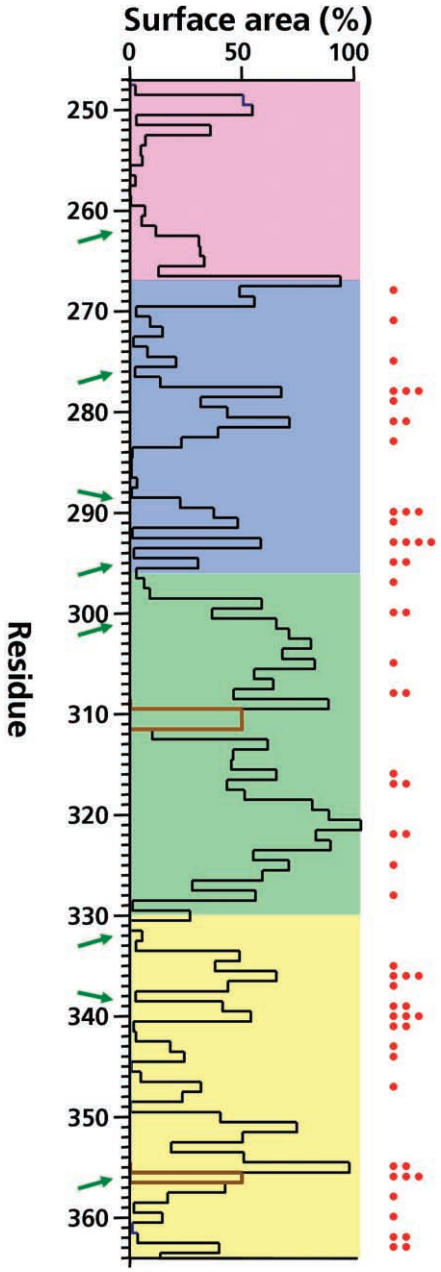

Figure 4. Surface accessibility. Shown is an overlay of relative solvent accessibility calculated for each amino acid of the 3-dimensional JR-FL envelope structure *(black)* and specific homologous sites within the C2-V3-C3 region, in which significant positive selection has been detected in at least 1 patient by codon-based likelihood models *(red dots)*. Amino acid numbering corresponds to gp160 of the reference strain HXB2 (GenBank accession no. K03455). Shaded areas represent the position of the buried region (*magenta;* residues 248–267) and the V3 loop (*green;* 297–330) as well as its N-terminal (*blue;* 268–296) and C-terminal (*yellow;* 331–363) flanking regions. Green arrows indicate potential N-linked glycosylation sites. Three residues missing in the JR-FL structure *(brown)* were excluded from all analyses.

Given that the extent of glycosylation in gp120 is highly variable and that both increases and decreases in glycosylation have been suggested to confer resistance to neutralization [10, 27], we specifically tracked changes in potential N-linked glycosylation sites (figure 4). Although longitudinal variations in multiple glycosylation sites within the C2-V3-C3 region were frequently observed (4 sites per patient affected over time; range, 1–8 sites), positive selection at potential glycosylation sites was found in only half of the patients (total of 13 sites in 10 patients).

DISCUSSION

Our analysis of molecular evolution within the region encompassing the envelope C2-V3-C3 domain has revealed that amino acids at exposed locations on the virion surface with high external accessibility are predominantly positively selected. The present data support the concept that evolution of these sites occurs in response to selective pressure, which can be envisioned to be induced by neutralizing antibodies or adaptations in the receptor interaction necessitated by the expansion of cellular tropism (see Pantophlet and Burton [28] for a comprehensive review of the biological aspects and structural features of the HIV envelope protein gp120). Notably, the structure of gp120 described by Huang et al. [26] and our evolutionary analysis of in vivo–derived sequences converge remarkably well. This underscores that, even though this crystal structure is based on a partially truncated and monomeric form of gp120, it appears to closely represent the structure of the biological relevant oligomeric gp120 on the virus. This is best exemplified by the consistency with which residues that are positively selected in vivo correspond to externally accessible amino acids in the model and highlights the potential these structural models have. Furthermore, the similar mutation patterns observed among widely divergent virus strains (JR-FL and patient specimens) also suggest that the 3-dimensional structure of gp120 is quite conserved. Looking at the different regions of gp120 instead of individual residues, it is not unexpected that the domains flanking the V3 loop undergo a high level of selective pressure, relative to the portions that are buried within the glycoprotein. However, analysis of a larger segment of gp120 would be required for a balanced overview. It is also important to note that the structure was derived from gp120 crystallized in the CD4-bound conformation. Because different residues are exposed in the unbound state or on functional envelope protein oligomers, the observed mutations in the V3 loop region over time must not necessarily imply escape from V3 loop–specific binding antibodies. Additionally, the occurrence of compensatory mutations due to changes elsewhere in the envelope cannot be excluded, and it is also not clear whether the preferential mutability of surface residues is related to a high autologous response in the selected cohort of patients under-

going treatment interruption. Furthermore, host immune-selection pressure could also be exerted by envelope-specific T cell responses, although, on the basis of our data, we could not find a correlation. CTLs have been shown to effectively attenuate viral replication [29]. Consequently, CTL escape mutants may arise early during HIV-1 infection. In a longitudinal study of HIV *env* evolution by Ross and Rodrigo [30], the sites under positive selection appeared to be associated with helper T cell epitopes in some patients.

In summary, our analysis of adaptive amino acid changes in the region of gp120 encompassing the V3 loop has brought to light substantial evidence that viral evolution in this domain is shaped by an ongoing competition between host and pathogen. Studies based on even larger longitudinal sequence data sets should be considered to define whether inter- or intrasubtype relevant hotspots of adaptive mutation exists, because these could be substantially important components of future vaccine and drug design.

Acknowledgments

We thank our patients for their commitment; Christine Schneider, Christina Grube, and Roland Hafner for excellent patient care; Esther Beerli, Friederike Burgener, Christine Leemann, Tuyet Trinh Lu, Barbara Niederöst, and Bärbel Sauer for laboratory support; and Ingrid Nievergelt and Christine Vögtli for administrative assistance.

References

1. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature **1995**; 373:123–6.
2. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. Nat Rev Genet **2004**; 5:52–61.
3. Shankarappa R, Margolick JB, Gange SJ, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol **1999**; 73:10489–502.
4. Yamaguchi-Kabata Y, Gojobori T. Reevaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein and prediction of new discontinuous epitopes. J Virol **2000**; 74:4335–50.
5. Yang W, Bielawski JP, Yang Z. Widespread adaptive evolution in the human immunodeficiency virus type 1 genome. J Mol Evol **2003**; 57: 212–21.
6. Choisy M, Woelk CH, Guegan JF, Robertson DL. Comparative study of adaptive molecular evolution in different human immunodeficiency virus groups and subtypes. J Virol **2004**; 78:1962–70.
7. Cocchi F, DeVico AL, Garzino-Demo A, Cara A, Gallo RC, Lusso P. The V3 domain of the HIV-1 gp120 envelope glycoprotein is critical for chemokine-mediated blockade of infection. Nat Med **1996**; 2:1244–7.
8. Edinger AL, Amedee A, Miller K, et al. Differential utilization of CCR5 by macrophage and T cell tropic simian immunodeficiency virus strains. Proc Natl Acad Sci USA **1997**; 94:4005–10.
9. Richman DD, Wrin T, Little SJ, Petropoulos CJ. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc Natl Acad Sci USA **2003**; 100:4144–9.
10. Wei X, Decker JM, Wang S, et al. Antibody neutralization and escape by HIV-1. Nature **2003**; 422:307–12.
11. Joos B, Trkola A, Fischer M, et al. Low human immunodeficiency virus envelope diversity correlates with low in vitro replication capacity and predicts spontaneous control of plasma viremia after treatment interruptions. J Virol **2005**; 79:9026–37.
12. Fagard C, Oxenius A, Gunthard H, et al. A prospective trial of structured treatment interruptions in human immunodeficiency virus infection. Arch Intern Med **2003**; 163:1220–6.
13. Fischer M, Hafner R, Schneider C, et al. HIV RNA in plasma rebounds within days during structured treatment interruptions. AIDS **2003**; 17: 195–9.
14. Fischer M, Joos B, Hirschel B, Bleiber G, Weber R, Gunthard HF. Cellular viral rebound after cessation of potent antiretroviral therapy predicted by levels of multiply spliced HIV-1 RNA encoding *nef*. J Infect Dis **2004**; 190:1979–88.
15. Trkola A, Kuster H, Leemann C, et al. Humoral immunity to HIV-1: kinetics of antibody responses in chronic infection reflects capacity of immune system to improve viral set point. Blood **2004**; 104:1784–92.
16. Oxenius A, McLean AR, Fischer M, et al. Human immunodeficiency virus-specific $CD8^+$ T-cell responses do not predict viral growth and clearance rates during structured intermittent antiretroviral therapy. J Virol **2002**; 76:10169–76.
17. Joos B, Trkola A, Kuster H, et al. Selection pressure by neutralizing antibodies results in higher adaptive mutation rates in gp120. Antiviral Therapy **2004**; 9:S75.
18. Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform **2004**; 5:150–63.
19. Williamson S. Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. Mol Biol Evol **2003**; 20:1318–25.
20. Kosakovsky Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics **2005**; 21:676–9.
21. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol **2005**; 22:1208–22.
22. Kosakovsky Pond SL, Frost SD. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics **2005**; 21:2531–3.
23. Collaborative computational project, number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr **1994**; 50:760–3.
24. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol **1971**; 55:379–400.
25. Ahmad S, Gromiha MM, Sarai A. Real value prediction of solvent accessibility from amino acid sequence. Proteins **2003**; 50:629–35.
26. Huang CC, Tang M, Zhang MY, et al. Structure of a V3-containing HIV-1 gp120 core. Science **2005**; 310:1025–8.
27. Derdeyn CA, Decker JM, Bibollet-Ruche F, et al. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. Science **2004**; 303:2019–22.
28. Pantophlet R, Burton DR. gp120: target for neutralizing HIV-1 antibodies. Annu Rev Immunol **2006**; 24:739–69.
29. Musey L, Hughes J, Schacker T, Shea T, Corey L, McElrath MJ. Cytotoxic-T-cell responses, viral load, and disease progression in early human immunodeficiency virus type 1 infection. N Engl J Med **1997**; 337:1267–74.
30. Ross HA, Rodrigo AG. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. J Virol **2002**; 76:11715–20.