Diversification of Subtype E Human Immunodeficiency Virus Type 1 env in Heterosexual Seroconverters from Northern Thailand

Zhe Wang, Cynthia M. Lyles, Chris Beyrer, David D. Celentano, David Vlahov, Chawalit Natpratan, Richard Markham, Chirasak Khamboonruang, Kenrad Nelson, and Xiao-Fang Yu

The C2–V3 region of the human immunodeficiency virus (HIV)-1 env was determined from 15 northern Thailand seroconverters between 1993 and 1995. Similar sequences were also determined from 18 seroconverting injection drug users in Baltimore. All seroconverters from northern Thailand were infected with subtype E HIV-1 on the basis of env sequences. Intersubject viral DNA distances increased from 2.3% in asymptomatic HIV-1–infected subjects characterized between 1990 and 1992 to 7.8% in these more recent seroconverters from Thailand. On the other hand, sequences from 18 seroconverters from Baltimore had a mean intersubject distance of 13.2%. The genetic diversity within HIV-1 subtype E in seroconverters in Thailand has increased significantly but is still less than that observed in HIV-1 from seroconverters in the United States, where the epidemic of HIV-1 infection is more mature. These results suggest that continued monitoring of the molecular epidemiology of HIV-1 infection in Thailand will be important for HIV vaccine development and evaluation.

Materials and Methods

Study subjects. Between 1993 and 1995, blood samples were collected from 15 recent HIV-1 seroconverters in northern Thailand as part of the HIVNET study. The median length of time in the HIV-1 seroconversion window was 4 months. Blood samples were collected within 3 months after the first seropositive visit from 9 subjects and within a year after first seropositive visit from the remaining 6 subjects. Among these seroconverters, 10 were female commercial sex workers, 1 was a male military conscript (E11429), 1 was a male sexually transmitted disease patient (B04034), and the other 3 were women attending HIV anonymous testing and counseling sites who had a known or suspected HIV-infected male partner. By self-report, all patients had no risk factors for HIV-1 infection other than heterosexual exposure. Samples were also collected from 18 HIV-1 seroconverters from the Balti-
more ALIVE (AIDS link to intravenous experiences) cohort [11] to study env sequence diversity.

Polymerase chain reaction amplification, DNA sequencing, and phylogenetic analysis of the C2–V3 region of HIV-1 env. Venous blood samples from Thai seroconverters were drawn and heparinized and transported to the United States within 48–72 h of collection. Peripheral blood mononuclear cells were separated from the blood samples for polymerase chain reaction and DNA sequencing as previously described [9]. DNA sequencing included 4–6 clones per sample from the C2–V3 region of gp120. Consistent with previous published observations, intrasubject env sequences from our seroconverters are remarkably homogeneous (data not shown). Intrasubject consensus sequences were generated and aligned using ClustalW [9] and edited by hand to shift gaps to preserve the reading frame. Sites containing gaps were then removed from the alignment as is customary, reducing the length of the aligned sequences from 296 to 280 sites. Phylogenetic trees were generated as previously described [9]. Intersubject distances were calculated by use of the DNADIST program and expressed as the average genetic distance for all pairwise combinations of nucleotide sequences from subjects within each of the following 3 groups: 18 previously reported asymptomatic persons collected between 1990 and 1992 in Thailand [1, 2]; 15 seroconverters in this study; and sequences from 18 seroconverters from a cohort of IDUs in Baltimore.

Statistical methods for nucleotide distance analysis. Each subject’s derived consensus sequence of C2–V3 nucleotide was used for the analysis. Descriptive statistics were calculated to describe the intersubject distances between pairs of viral sequences within each group. The frequency distributions for the intersubject distances are graphically represented for each group in figure 1. A robust statistical method, the generalized estimating equation (GEE), was used to test the equivalence of the mean intersubject distance between 2 groups while controlling for the within-subject correlations. In applying this method, we assumed the clusters were defined by the first subject in the subject pair used to quantify the nucleotide distance. The mean intersubject distances were compared and tested between the recent HIV-1 seroconverters in northern Thailand and each of the other 2 groups.

Results

Comparison of nucleotide sequence diversity among HIV-1–infected persons from Thailand. Phylogenetic trees were constructed on the basis of the consensus nucleotide sequence from the C2–V3 region of gp120 from each of the subjects’ samples and compared with a selection of previously published HIV-1 reference sequences A through H [8]. All these sequences clustered distinctively with E subtype viruses, and the bootstrap value for grouping with known subtype E sequences was 88% (data not shown). The intersubject viral sequence variations were studied by nucleotide distance analysis. The mean nucleotide diversity of pairwise comparisons of intersubject sequences among these 15 seroconverters was 7.8% (range, 3.4%–16.3%). The mean nucleotide diversity of pairwise comparisons of intersubject sequences among previously published

Comparison of nucleotide sequence diversity among seroconverters from northern Thailand and seroconverters from the United States. The relatively low genetic heterogeneity of subtype E HIV-1 sequences in Thailand has been attributed to the relatively short duration of the HIV-1 epidemic there, which is considered a desirable attribute for a vaccine trial. We therefore compared env sequence diversity among seroconverters from northern Thailand with that of seroconverters from the Baltimore ALIVE cohort. C2–V3 sequences were determined by use of uncultured peripheral blood mononuclear cell samples from 18 seroconverters (S1–S18) within 1 month after the first seropositive visit. All subjects were infected with subtype B HIV-1 on the basis of phylogenetic analysis of their env sequences (data not shown).

The mean nucleotide diversity of pairwise comparisons of intersubject sequences among 18 seroconverters from the Baltimore ALIVE cohort was 13.2% (range, 0.4%–25.0%). The difference in mean nucleotide distance between the Baltimore (13.2%) and Thailand (7.8%) HIV-1 seroconverters was statistically significant (P < .001). The comparison of the distributions of intersubject nucleotide distances in the C2–V3 region between these 2 groups also suggests that seroconverters from Baltimore have higher viral genetic diversities than do seroconverters from northern Thailand (figure 1).

Comparison of amino acid sequence diversities in the C2–V3 region of the env gene among HIV-1–infected persons. As shown in figure 2, the derived C2–V3 amino acid sequences from the HIV-1 seroconverters from northern Thailand were aligned with the consensus subtype E HIV-1 sequence used in this analysis. Compared with the previously published subtype E HIV-1 sequences [1, 2] from asymptomatic persons from Thailand (figure 2, top), sequences from some of the seroconverters in our study (figure 2, middle) appeared to have diversified significantly from the consensus sequence. Most of the variations were clustered in the V3 loop and the region immediately after the V3 loop. More heterogeneity in the C2–V3 sequences was detected in seroconverters from Baltimore (figure 2, bottom) compared with seroconverters from northern Thailand (figure 2, middle). The C2 region immediately before the V3 loop is more heterogeneous among seroconverters from Baltimore than is the same region among seroconverters from northern Thailand.

In addition to amino acid diversification in the C2–V3 region, two features regarding the tip of the V3 loop sequence and the pattern of N-linked glycosylation in this region are worth noting. The GPGQ motif at the tip of the V3 loop was predominant in previously published subtype E HIV-1 sequences from 18 asymptomatic persons (100%) [1, 2], 10 heterosexually HIV-1–infected persons from northern Thailand (100%) [12], and 14 IDUs (93%) from Thailand [3]. However, sequences from 4 of 15 recent seroconverters (27%) in our current study had GPGGR, GPGK, or GPGH at the tip of the V3 loop (figure 2, middle). We have previously observed [9] a higher prevalence (59%) of GPGGR and GPGH at the tip of the V3 loop sequences from AIDS patients in northern Thailand.

Seven potential N-linked glycosylation sites (NXT/S) were present in the region that was sequenced (figure 2). With the exception of glycosylation site 6, the other six N-linked glycosylation sites were highly conserved in the previously published subtype E HIV-1 sequences [8] and sequences from the seroconverters in this study (figure 2). Only 1 (5.6%) of 18 previously published subtype E HIV-1 sequences from Thailand had glycosylation site 6 (figure 2, top), although it is present in most other subtypes of HIV-1, including subtype E HIV-1 from Africa [13]. In contrast, sequences from 7 (47%) of 15 recent seroconverters in our study acquired this potential N-linked glycosylation site (figure 2, middle). We also have observed a relatively high proportion (7/22, 32%) of sequences from Thai AIDS patients containing this glycosylation site [9].

Discussion

The nucleotide sequences in the C2–V3 region of HIV-1 gp120 from a few early seroconverters were substantially different from previously published subtype E HIV-1 sequences from Thailand. Also, the mean intersubject viral sequence distance was significantly higher in these 15 seroconverters (7.8%) obtained from 1993 to 1995 than in earlier Thai sequences from HIV-1–infected, asymptomatic persons (2.3%) obtained from 1990 to 1992 [1, 2]. All pairwise comparisons of intersubject sequences among previously published subtype E HIV-1 sequences from 18 asymptomatic persons [1, 2] were <6% (figure 1, hatched bars). On the other hand, >69% of the pairwise comparisons of intersubject sequences among these 15 seroconverters were >6% (figure 1, open bars).

Samples in this study were collected ~3 years after the earlier studies [1, 2]. These data again suggest that the nucleotide sequences in C2–V3 region of gp120 are more diversified in recent seroconverters from northern Thailand than in those from asymptomatic persons collected in Thailand between 1990 and 1992 [1, 2]. The 5.5% increase in mean intersubject viral sequence differences may not be simply due to the time difference if assuming an evolutionary rate of 0.5%–1.0% per year in HIV-1 env [14]. It is possible that we were sampling only a small region in HIV-1 env that may have a higher rate of evolution than other regions in env. Alternatively, introduction of more divergent viruses from AIDS patients [9] to the general population may have contributed to a rapid increase in sequence diversity among the recently transmitted viruses.
Figure 2. Comparison of C2–V3 amino acid sequences among Thai seroprevalent asymptomatic persons, Thai recent seroconverters, and seroconverters from Baltimore. Top, Thai sequences from 18 HIV-1-seroprevalent subjects published previously [1, 2]. Middle, sequences from 15 recent seroconverters from northern Thailand. Sequences in top and middle panels were aligned with subtype E HIV-1 consensus sequence derived from all 33 Thai subjects. Residues identical to consensus sequence are shown as dashes. Deletions are shown as periods. Potential N-linked glycosylation sites are labeled as ^^^. Bottom, sequences from 18 seroconverters from Baltimore ALIVE cohort aligned with subtype B HIV-1 consensus sequence.
Only the hypervariable V3 loop and flanking regions were monitored in this study. Increased diversity in hypervariable regions such as V3 does not guarantee overall high diversity in gp120. However, previous studies using Thai sequences from asymptomatic HIV-infected and AIDS patients did find a good correlation between an increased diversity in the C2–V3 region of gp120 and an increased diversity in the whole gp120 or gp160 [9, 10, 15]. Therefore, our results suggest that the diversity of subtype E HIV-1 env sequences that are circulating in the infected population in northern Thailand is increasing.

Increased diversity in variable regions of gp120 may affect the immunogenicity of the viral antigens. Subtype E HIV-1 isolates with more diversified C2–V3 sequences (such as CMU02, CMU06, CMU08, and CMU10) are harder to neutralize by pooled human sera against Thai subtype E HIV-1 than are virus isolates with less diversity [16]. Increased diversity within subtype E HIV-1 as the epidemic continues in Thailand should be considered in the framework of vaccine development in this country. The inverse correlation of genetic diversity with potential efficacy of a subtype E vaccine candidate must be recognized as a concern to vaccine developers.

The overall viral genetic diversity of the C2–V3 region of subtype E HIV-1 gp120 in infected persons from northern Thailand (7.8%) is still significantly lower than that of HIV-1 seroconverters from a cohort of IDUs in the United States (13.2%). Differences in the ages of the epidemics, modes of transmission, and subtypes of HIV-1 may have contributed to the differences in viral genetic diversity between these 2 groups. Selection at mucosal surfaces might lead to a more homogeneous viral quasispecies in sexually transmitted viruses. Areas in which the epidemic is relatively new, such as Thailand, still represent a more favorable site to conduct subunit vaccine trials based on the HIV-1 gp120 or gp160 molecule than do sites where the epidemic is older. However, delay in implementing vaccination efforts in Thailand increases the likelihood of vaccine failure due to increasing expansion of viral diversity in the Thai population.

Acknowledgments

We thank Zheng Chen, Homayoon Farzadegan, Beth Masters, and Jacque Astemborski from Johns Hopkins School of Hygiene and Public Health, Baltimore, for technical assistance.

References