SURVEILLANCE OF SUBTYPE AND GENETIC VARIATION OF THE CIRCULATING STRAINS OF HIV-1 IN THAILAND

Piyanot Wirachsilp¹, Wannee Kantakamalakul¹, Suporn Foongladda¹, Thippawan Chuenchitra², Surapol Kohriangudom³, Niracha Athipanyasilp¹, Sombat Tanprasertsuk³, Malee Gasitrong⁴ and Ruengpung Sutthent¹

¹Department of Microbiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok; ²Armed Forces Research Institute of Medical Sciences, Bangkok; ³Bureau of AIDS, TB, and STIs, Ministry of Public Health, Nonthaburi; ⁴Supprasittiprasong Hospital, Ubon Ratchathani, Thailand

Abstract. Two HIV-1 strains, CRF01_AE and subtype B', were reported in Thailand during the early years of the epidemic. Recently, an intersubtype recombination of HIV-1 strain was found in Thailand. Eight-hundred and twenty-eight samples collected during years 1995-2004 from high-risk groups in Bangkok, northern, northeastern, and southern region of Thailand were studied. HIV-1 env nucleotide sequences were used for phylogenetic analysis of the circulating HIV-1 strain. By single HIV-1 region (env) genotyping, CRF01_AE was found in 97.3% and HIV-1 subtype B was found in 2.7%. A predominance of CRF01_AE was found in all geographic regions. Parallel analysis of the HIV-1 gag and env genes demonstrated that 2.1% and 4.0% of recombinant HIV-1 strains were found using p17 and p24 region sequences, respectively. The recombinant gag gene was also found in one southern isolate. Phylogenetic analysis of HIV-1 isolated from 20 provinces in 2002 suggested the northern and northeastern isolates were more related than the southern isolates which had the lowest genetic diversity of 0.13. The GPGQ V3 loop tip was also present in isolates from all regions. The molecular epidemiological data from this study may be useful for surveillance design as well as targeting prevention efforts. It also provides information regarding new antigenic regions of circulating strains responsible for the HIV-1 epidemic in Thailand.

INTRODUCTION

The diversity of human immunodeficiency virus type 1 (HIV-1) is expanding globally; primarily as a result of the low fidelity of the viral reverse-transcriptase (RT) and cellular RNA polymerase. These polymerases are responsible for high rates of substitution, insertion, and deletions during transcription events in the virus replication cycle (Mansky, 1998). There are 3 groups of HIV-1, groups M, N and O, of which, the main group (group M) is responsible for the pandemic. As of 2005, there have been 9 recognized subtypes (A-D, F-H, J-K), 2 sub-subtypes of HIV-1 (A1-A2, F1-F2), with 34 circulating recombinant forms (CRF) reported in the HIV Sequence Database (http://www.hiv-web.lanl.gov). While subtype B predominates in North America and Europe, HIV-1 CRF01_AE appears to be the cause of high proportion of infections in Thailand and Southeast Asia and seems to be largely confined to this area of the world.

The first case of AIDS in Thailand was reported in 1984. The patient was a Thai male who developed full-blown AIDS outside the country and returned home. More AIDS cases were reported during 1984-1988 but all of
them were men who had sex with men (MSM) and had foreign partners (Phanuphak et al., 1985; Wangoongsarb et al., 1985; Limsuwan et al., 1986; Traisupa et al., 1987). In early 1988, HIV sero-surveillance studies in injecting drug users (IDUs) who attended methadone clinics in Bangkok were conducted, and the outbreak of HIV-1 in Thailand was first recognized among IDUs (Smith, 1990). The first HIV-1 molecular epidemiology study conducted in Thailand was reported in 1992 (Ou et al., 1992). Based on the analysis of C2-V3 nucleotide sequences, two HIV-1 genotypes in Thailand were identified. These 2 strains appeared independently in distinct high-risk groups, “genotype A” among heterosexuals and “genotype B” among IDUs; which were later named as subtypes E (and then CRF01_AE) and B, respectively. Further HIV-1 molecular epidemiology studies during year 1993-1994 showed that these 2 strains have been circulating in Thailand in distinct high-risk groups for more than a decade (Ou et al., 1993; Kalish et al., 1994).

Since 1995, HIV-1 CRF01_AE, which has been mainly found in sexually transmitted cases has caused an increasing proportion of new infections among IDUs in Thailand, which accounts for at least 80% of new infections in Bangkok and more than 90% in northern Thailand. Studies of HIV-1 subtypes among seroconverted individuals in Thailand from 1995 to 2000 reported that HIV-1 CRF01_AE accounted for more than 80% of cases in both IDU and sexual transmission groups (Subbarao et al., 1998). The long stable molecular picture for HIV-1 in Thailand appears to be entering a new phase of complexity when CRF15_01B was reported to be the first new circulating recombinant HIV-1 in Thailand in 2001 (Viputtijul et al., 2002). In 2003, a new recombinant virus between subtype C and CRF01_AE was reported in an HIV-1 infected Thai (Ramos et al., 2003; Tovanabutra et al., 2003).

As of December 2006, a total of 307,114 HIV infected cases have been reported in Thailand by the Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health, Thailand (MOPH). During the past 10 years, heterosexual transmission has been the major route for developing of HIV-1 (83.9%) followed by IDU (4.4%). During the same period, the northern region of Thailand had the highest number of newly infected cases, followed by the central and northeastern regions.

The genetic variability of HIV-1 makes it possible to study the HIV epidemic within populations. Because HIV-1 strains evolve rapidly, only viruses that are epidemiologically related, such as founder viruses, will share sequence similarity that can be highlighted by analysis, such as phylogenetic analysis and signature sequence analysis. Such analyses have been used in various studies tracing global epidemiologic patterns (Nerurkar et al., 1996; Kuiken et al., 2000; Takebe et al., 2003; Sepherd et al., 2005; Tee et al., 2005). In this study, subtype determination and phylogenetic analysis were performed to study the molecular epidemiology of HIV-1 in Thailand.

MATERIALS AND METHODS

Study subjects

Eight-hundred twenty-eight samples were collected from asymptomatic and drug naïve HIV-1 infected individuals in Bangkok (BKK, n=88), northern (NO, n=295), northeastern (NE, n=200), eastern (E, n=2), and southern (SO, n=243) regions of Thailand from 1995 to 2004 through national surveillance programs. Of 828 samples, 786 were collected from the sexual transmission risk group, while 31, 9, and 2 samples were collected from IDU, perinatal risk groups, and blood donors, respectively.

All subjects in this study were asymptomatic with CD4+ count more than 200 cells/mm³ and were antiretroviral drug naïve.
Specimen processing

Unclotted EDTA blood was collected after obtaining written informed consent from the HIV-1 infected individuals through the annual national surveillance program by the Bureau of AIDS, TB, and STIs, MOPH. Five milliliters of blood were collected in an EDTA vacutainer tube. The EDTA blood tube was centrifuged at 1,800 rpm for 5 minutes to separate blood cells from plasma. Plasma was removed and stored at -70°C for HIV-1 viral RNA extraction.

Amplification of HIV-1 gene regions

HIV-1 viral RNA extracted using QIAamp spin columns (QIAGEN GmbH, Hilden, Germany) was performed according to manufacturer’s instruction.

The C2-V4 region of the env gene, and the p17 and p24 regions of the gag gene were amplified by RT-PCR using a One step Superscript RT-PCR with platinum Taq DNA polymerase (Invitrogen, CA, USA) and nested-PCR. The primers for HIV-1 gene amplification are shown in Table 1. The RT-PCR reaction mixture consisted of 25 µl of 2X reaction mix, 1 µl each of 10 pmol/µl outer primer, 2 µl of superscript RT-PCR with platinum Taq DNA polymerase, 10 µl of HIV-1 viral RNA and 11 µl of deionized water. The reaction mixtures were transferred to a 2400 DNA Thermal cycler (Applied Biosystem, CA, USA). The cycles consisted of a reverse transcription step at 55°C for 30 minutes then inactivated RT enzyme at 94°C for 15 minutes and 35 cycles of 94°C denaturation step for 1 minute, a 55°C annealing step for 1 minute, and a 68°C extension step for 3 minutes and the final extension at 68°C for 5 minutes.

For Nested PCR, 5 µl of first round RT-PCR product was mixed into the reaction tube which consisted of 5 µl of 10X PCR buffer, 3 µl of 25 mM MgCl₂, 6 µl of 10 mM dNTP (Promega, WI, USA), 3 µl each of 10 pmol/µl inner primers, and 2.5 U of Taq DNA polymerase (Promega, WI, USA). The amplification cycles were carried out at 94°C for 5 minutes followed by 35 cycles of a 94°C denaturation step for 1 minute, 55°C annealing step for 1 minute, and 72°C extension step for 3 minutes and final extension at 72°C for 7 minutes.

HIV-1 nucleotide sequencing and analysis

A 530-base pair, 500-base pair, and 800-base pair RT-PCR products of C2-V4 of env gene, p17 and p24 of gag gene, respectively, were purified using the QIAquick Gel Extraction Kit (QIAGEN GmbH, Germany). DNA cycle sequencing was done bidirectionally using BigDye Terminator Ready Reaction Mix (Applied Biosystem, CA, USA) with nested primers for each region of the HIV-1 gene. The reaction mixture contained the template DNA of approximately 100-250 ng, 4-8 pmol/µl of sequencing primers and 4 µl of terminator dye (containing deoxynucleotides, dye labeled dideoxynucleotides, a polymerase enzyme and buffer system). When the reaction was terminated, the product was labeled with the specific dye corresponding to that specific nucleotide. The excess dye-labeled dideoxynucleotide was eliminated by precipitating with 95% ethanol in a 1.5 microcentrifuge tube. After the fluorescent labeled ssDNA was completely dried, the sequencing was carried out by an ABI 310 automatic sequencer.

A chromatogram of HIV-1 nucleotide sequences was performed using Chromas version 1.45. Lasergene Software version 5.0 (DNASTAR, WI, USA) was used to edit and translate the HIV-1 nucleotide sequences. The edited nucleotide sequences were aligned using ClustalX (Thompson et al, 1997). Phylogenetic analysis, statistical robustness of the neighbor-joining tree with F84 evolutionary model and bootstrap test of 500 replicates, were performed with PHYLIP 3.63 software package for Microsoft Windows (distributed by the author, Department of Genome Sci-
ences, University of Washington, Seattle) to determine the subtype of HIV-1 (Felsenstein, 1993). Known subtype HIV-1 reference strains were included in the analysis. The phylogenetic tree was constructed by TREEVIEW, version 1.6.6 (distributed by Roderic DM, Page 2001 in: http://taxonomy.zoology.gla.ac.uk/rod/rod/html).

Bootscanning analysis contained in SimPlot version 3.2 (Lole et al, 1999) was used to identify and to determine the significance of recombination breakpoints. A neighbor-joining algorithm with 500 replicates, using sliding windows of 200 nucleotides with an overlap of 50 nucleotides, was applied to the analysis. The background sequences that were used included CRF01_AE (CM240), subtype B (MN), and subtype C (96BW05), all of which are reference strains available at the HIV Sequence Database website (http://hiv-web.lanl.gov). The breakpoint was determined by a significant (p<0.001) difference in the ratio of informative sites for each HIV-1 subtype on either side of the breakpoint, as assessed by the Maximum Likelihood Chi-squared (Martin et al, 2005).

**RESULTS**

**HIV-1 subtype determination using C2-V4 region of env gene**

Among 828 HIV-1 nucleotide sequences, 806 (97.3%) sequences were HIV-1 CRF01_AE and 22 sequences (2.7%) were HIV-1 subtype B.

The analysis of HIV-1 circulated in each geographic region in Thailand demonstrated that CRF01_AE was predominant in all regions. HIV-1 CRF01_AE was found at 95.5, 99.3, 98.0, 95.1, and 100% in Bangkok, northern, northeastern, southern, and eastern regions, respectively. HIV-1 subtype B was found at 4.5%, 0.7%, 2.0%, and 4.9% in Bangkok, northern, northeastern, and southern regions, respectively. There was no HIV-1 subtype B found in the eastern region for this study.

HIV-1 nucleotide sequence accession numbers

HIV-1 nucleotide sequences in this study have been submitted to GenBank, the US National Institutes of Health (NIH) genetic sequence database, under accession number AY005164, 66, AF373038-40, AY005172, AY005168, AF373041, AY005174,76, AY00 5170, AF373042-3, AF449824-50, AF449894-908, AF449883, AF449936-63, AF449864-79, AF449884-7, AF449880, AF449888-9, AF44 9912-9,23,27-34, AF449993, AF449805-6, AF449922,24-6, AF449854-8, AF449799-800, AF449808-20, AY243225-36, AY243241-2, AY243180-94, AY243195-216, AY243237-40, AY248123-33, AY248134-8, AY621260, AY621237-54, AY621264-70, EF063390, AY248113-22, AY248107-12, AY248139-55, AY243217-24, AY243135-62, AY243164-8, AY243311, AY243171, AY242288, AY248288, AY248040-4, 51-7, 88-92, AY243173, AY24 8058-60, AY248066-77, AY248093-106, AY24 3174-9, AY248045-50, AY248061-5, AY248 078-87, AY621174-83, AY621198, AY621286- 92, AY243308-10, AY243170, AY248156-69, AY248213-9, AY243283-92, AY248199-208, AY248233-7, AY248252-60, AY248275-87, AY248039, AY248181-8, AY248209-12, AY24 8238-40, AY248261-5, AY243244-8, AY243 266-82, AY243293-301, AY243249-65, AY248 241-6, AY243302-7, AY248189-98, AY248 221-32, AY248248-51, AY248266-74, AY621 261-3, EF063393-438, EF063391-2.

**HIV-1 subtype determination using C2-V4 region of env gene**

Among 828 HIV-1 nucleotide sequences, 806 (97.3%) sequences were HIV-1 CRF01_AE and 22 sequences (2.7%) were HIV-1 subtype B.

The analysis of HIV-1 circulated in each geographic region in Thailand demonstrated that CRF01_AE was predominant in all regions. HIV-1 CRF01_AE was found at 95.5, 99.3, 98.0, 95.1, and 100% in Bangkok, northern, northeastern, southern, and eastern regions, respectively. HIV-1 subtype B was found at 4.5%, 0.7%, 2.0%, and 4.9% in Bangkok, northern, northeastern, and southern regions, respectively. There was no HIV-1 subtype B found in the eastern region for this study.

HIV-1 CRF01_AE was found at 95.9% among males and 96.6% among females in this study; HIV-1 subtype B was found at 4.1% and 3.4% among males and females, respectively (Table 2).

**Parallel phylogenetic analysis of HIV-1 gag and env gene**

HIV-1 gag gene (p17 and p24 regions) of the corresponding samples were used in parallel phylogenetic analysis to identify discordant subtypes. The phylogenetic analysis model for gag gene analysis was the same as
those used for the env gene.

Parallel phylogenetic analysis of 193 nucleotide sequences of p17 region which corresponded to C2-V4 region sequences reveals that 189 samples (97.9%) had the same subtype in both regions, while the remaining 4 (2.1%) had a discordant subtype. Among concordant subtype samples, 10 samples (5.3%) were subtype B and 179 samples (94.7%) were CRF01_AE.

Seventy-six nucleotide sequences of the C2-V4 region, which corresponded to the p24 region.

Table 1

<table>
<thead>
<tr>
<th>Name</th>
<th>Primer sequence (5’-3’)</th>
<th>Location in HXB2</th>
</tr>
</thead>
<tbody>
<tr>
<td>JH35</td>
<td>CACTTCTCCAATTGTCCITCA</td>
<td>6371-6392</td>
</tr>
<tr>
<td>JH44</td>
<td>ACAGTRCARTGYACACATGG</td>
<td>7891-7871</td>
</tr>
<tr>
<td>JH33</td>
<td>CTGTTIAATGGCAGICTAGC</td>
<td>7002-7021</td>
</tr>
<tr>
<td>JH48</td>
<td>RATGGGAAGGRGYATACAT</td>
<td>7540-7524</td>
</tr>
<tr>
<td>Whole gagO1</td>
<td>GCAGAGCTAGAAAGGAGAGA</td>
<td>769-788</td>
</tr>
<tr>
<td>Whole gagO2</td>
<td>GCTCTCCTCGATTCTTACTG</td>
<td>5177-5197</td>
</tr>
<tr>
<td>P17I1</td>
<td>GAGATGGGGTGCGAGAGCG</td>
<td>785-804</td>
</tr>
<tr>
<td>P17I2</td>
<td>GTACCATTGGCCTTTGTCG</td>
<td>1219-1201</td>
</tr>
<tr>
<td>P24I1</td>
<td>GTTCAGCCAAAATTACCCCT</td>
<td>1171-1188</td>
</tr>
<tr>
<td>P24I2</td>
<td>GTTTGTTGGTTGCATCCCACTC</td>
<td>1904-1182</td>
</tr>
</tbody>
</table>

The JH primer set was used to amplify C2-V4 region. Whole gagO1/O2 and p17I1/I2 were used to amplify the p17 region. Whole gagO1/O2 and p24I1/I2 were used to amplify the p24 region.

Table 2

<table>
<thead>
<tr>
<th></th>
<th>%CRF01_AE (no./total)</th>
<th>%Subtype B (no./total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>By year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995-1996</td>
<td>100 (10/10)</td>
<td>0 (0/10)</td>
</tr>
<tr>
<td>1997-1998</td>
<td>100 (48/48)</td>
<td>0 (0/48)</td>
</tr>
<tr>
<td>1999-2000</td>
<td>97.8 (132/135)</td>
<td>2.2 (3/135)</td>
</tr>
<tr>
<td>2001-2002</td>
<td>96.9 (567/585)</td>
<td>3.1 (18/585)</td>
</tr>
<tr>
<td>2003-2004</td>
<td>98.0 (49/50)</td>
<td>2.0 (1/50)</td>
</tr>
<tr>
<td>By geographic region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bangkok</td>
<td>95.5 (84/88)</td>
<td>4.5 (4/88)</td>
</tr>
<tr>
<td>North</td>
<td>99.3 (293/295)</td>
<td>0.7 (2/295)</td>
</tr>
<tr>
<td>Northeast</td>
<td>98.0 (196/200)</td>
<td>2.0 (4/200)</td>
</tr>
<tr>
<td>South</td>
<td>95.1 (231/243)</td>
<td>4.9 (12/243)</td>
</tr>
<tr>
<td>East</td>
<td>100 (2/2)</td>
<td>0 (0/2)</td>
</tr>
<tr>
<td>By gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>95.9 (210/219)</td>
<td>4.1 (9/219)</td>
</tr>
<tr>
<td>Female</td>
<td>96.6 (375/388)</td>
<td>3.4 (13/388)</td>
</tr>
</tbody>
</table>
Table 3
Phylogenetic classification of HIV-1 isolates with discordant subtype.

<table>
<thead>
<tr>
<th>No.</th>
<th>p17</th>
<th>p24</th>
<th>Pol</th>
<th>C2-V4</th>
<th>Year</th>
<th>Area</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO350</td>
<td>01_AE</td>
<td>01_AE</td>
<td>-</td>
<td>B</td>
<td>2002</td>
<td>Phitsanulok</td>
<td>Female</td>
</tr>
<tr>
<td>SO469</td>
<td>01_AE</td>
<td>01_AE</td>
<td>B</td>
<td>B</td>
<td>2002</td>
<td>Nakhon Si Thammarat</td>
<td>Female</td>
</tr>
<tr>
<td>SO868</td>
<td>B</td>
<td>01_AE</td>
<td>B</td>
<td>B</td>
<td>2002</td>
<td>Nakhon Si Thammarat</td>
<td>Female</td>
</tr>
<tr>
<td>NE1036</td>
<td>01_AE</td>
<td>01_AE</td>
<td>B</td>
<td>B</td>
<td>2002</td>
<td>Nong Khai</td>
<td>Female</td>
</tr>
</tbody>
</table>

Phylogenetic classification of HIV-1 isolates with discordant subtype.

Samples discordant between the gag and env genes were further analyzed by phylogenetic analysis of the partial reverse-transcriptase (RT) region nucleotide sequences of the pol gene. A neighbor-joining tree of the partial RT region nucleotide sequences of 600 base-pairs of 3 subtype discordant samples and HIV-1 reference strain nucleotide sequences demonstrated that all 3 samples contained the HIV-1 subtype B genome in the RT region of the pol gene.

Recombinant HIV-1 gag gene analysis

One sample (SO868), which contained HIV-1 recombinant genome in the gag gene, was further identified for recombination breakpoint by bootscan analysis using SimPlot version 2.5. Nucleotide sequence position 1-380 of HIV-1 gag gene had 90% similarity with HIV-1 subtype B reference strain MN while position 381-1062 had 98% similarity with CRF01_AE reference strain CM240 by SimPlot. The breakpoint where recombination of 2 subtypes occurred was identified at the nucleotide position 380 bp of gag gene with bootstrap support of 47% permuted tree (Fig 1). The phylogenetic tree of the gag gene nucleotide sequence from positions 1-380 and
381-1062 was separately constructed to confirm recombination breakpoint found by bootscan (Fig 2).

Phylogenetic clustering of CRF01_AE circulating in Thailand

HIV-1 CRF01_AE isolates collected from recently seroconverted individuals in northern, northeastern, and southern regions of Thailand were used to analyze the cluster of CRF01_AE strains. These individuals were asymptomatic, contracted HIV heterosexually, and did not receive any anti-retroviral treatments. Five hundred sixty-seven CRF01_AE env (C2-V4) nucleotide sequences isolated in 2002 were used for a cluster study of CRF01_AE strains circulating in Thailand.

All phylogenetic trees of CRF01_AE nucleotide sequences from each geographic region (NO, NE, SO tree) demonstrated the same patterns of phylogenetic clustering (Fig 3). Six phylogenetic clusters were identified among these sequences. These clusters were
Fig 3—Phylogenetic trees of CRF01_AE isolated from northern (3a), northeastern (3b), and southern (3c) Thailand. All CRF01_AE isolates in each region clustered together without geographic distinction. The phylogenetic trees were rooted with SIVcpz. All phylogenetic trees shared the same pattern of clustering. Two clusters, which the circulating strains are clustered with the reference strain, are indicated. The HIV-1 CRF01_AE reference strains, CM244, CM240, and 93TH253, are indicated.
designated as NO1-6, NE1-6, and SO1-6 cluster for clusters from NO, NE, and SO trees, respectively. There was one cluster which CRF01_AE sequences clustered with 2 Thai reference strains, TH293 and CM240, isolated in the years 1993 and 1990, respectively. The other major cluster of these 3 phylogenetic trees was a cluster of CRF01_AE sequences with another Thai reference strain, CM244, isolated in the year 1990. The other 4 clusters consisted of CRF01_AE sequences from provinces in each region clustered together.

The phylogenetic tree for CRF01_AE sequences isolated in the year 2002 was constructed with a Maximum-likelihood method which demonstrated that there was no geographic distinction among these sequences. HIV-1 nucleotide sequences isolated from Bangkok in 2000 were also analyzed with HIV-1 nucleotide sequences from each region. The phylogenetic analysis of HIV-1 Bangkok isolates and HIV-1 NO, NE, and SO isolates demonstrated that there was also no geographic distinction among HIV-1 isolates in Thailand. HIV-1 sequences of Bangkok isolates clustered with HIV-1 sequences isolated from other regions.

Genetic diversity of CRF01_AE circulating in Thailand

The maximum likelihood distance of each group was computed to study the average genetic diversity of NO1-6, NE1-6, and SO1-6 clusters. Mean genetic diversity (±SD) of CRF01_AE circulating in northern, northeastern, and southern Thailand were 0.140±0.36 (0.109-0.201), 0.150±0.22 (0.117-0.194), and 0.126±0.10 (0.117-0.146), respectively (Fig 4). Comparing the genetic diversity of the geographical trees confirmed clustering in each tree that was different in genetic diversity among the 6 clusters in the NO (p<0.0002), NE (p<0.0001), and SO (p<0.0001) trees. CRF01_AE circulating in southern Thailand had the lowest genetic diversity (p<0.05) compared to that circulating in other regions of the country. There was no difference in mean genetic diversity among the CRF01_AE circulating in northern and northeastern Thailand (p<0.05).

Signature sequence pattern

Signature sequence analysis of each cluster in the NO, NE, and SO trees was done using HIV-1 CM244 strain, which is a CRF01_AE reference strain, as a background for the analysis. CRF01_AE circulating in the northeast (NE6 cluster) had the highest number of variable positions (15) compared to CM244, while those circulating in the south (SO4 cluster) had the lowest number of variable positions (4). The variable position identified by signature sequence pattern analysis was consistent with the mean genetic diversity that CRF01_AE circulating in the south had low genetic diversity. Among 6 clusters in the NO tree, NO4 cluster had the highest number of variable amino acid positions. For the southern isolates of CRF01_AE, there were 2 clusters that had the highest number of variable amino acid positions (9), the SO1 and...
Fig 5–Signature sequence patterns of CRF01_AE amino acid sequences of NO1-6, NE1-6, and SO1-6 clusters compared with reference strain CM244. Dots indicate similar amino acid sequences between each cluster and the reference strain. Amino acid position in the alignment was numbered according to the HXB2 numbering engine for the gp160 region of the env gene.
SO5 clusters. These 2 clusters shared the same major population of CRF01_AE isolated from southern coastal provinces in the Gulf of Thailand. GPGQ was the predominant V3 loop tip pattern in all clusters.

N-linked glycosylation prediction

There were 6 acceptor sites for N-linked glycosylation reported for the C2-V4 region sequence of the CM244 reference strain at amino acid positions 275, 289, 295, 301, 333, and 384. N-GLYCOSITE was used to predict N-linked glycosylation sites on amino acid sequences of CRF01_AE in 6 phylogenetic clusters of NO, NE, and SO trees. All 6 N-linked glycosylation sites found in CM244 were also found in 14 clusters of CRF01_AE isolated in 2002. Clusters NO4 and NE6 did not have N-linked glycosylation at position 384 by amino acid substitution from N to T. Two additional N-linked glycosylation sites at amino acid positions 338 and 353/354 were predicted from CRF01_AE isolated from all regions. The additional sites at amino acid position 338 were found in 8 clusters (NO2,4,5; NE1,4; SO1,2,5) by amino acid substitution from A to T at position 340. Another additional site at amino acid position 353/354 was found in 16 clusters. It was caused by amino acid substitution from P to T.

DISCUSSION

In this study, phylogenetic and signature sequence analyses were used to trace the evolution and epidemiology of the circulating HIV-1 strains in Thailand.

The C2-V4 region of HIV-1 env gene nucleotide sequences from infected individuals in major epidemic areas in Thailand were used to study HIV-1 subtypes. The predominant subtype of HIV-1 circulating in all geographic regions of Thailand during the years 1995-2004 was CRF01_AE (97.3%). This finding is consistent with a previous report in 1995 when there were no recombinant strains (Kalish et al., 1994).

Full length analysis of HIV-1 isolate CM237 previously reported to be subtype B by one region genotype, revealed this isolate had a recombinant genome. The discordant subtype between gag and pol genes indicates that HIV isolate CM237 was a CRF01_AE/B recombinant strain (Magiorkinis et al., 2002). Thus, parallel phylogenetic analysis, an analysis of corresponding HIV-1 gag and env genes, was used to screen recombinant genomes of HIV-1 strain in this study. By parallel phylogenetic analysis, 2.1% and 4.0% of HIV-1 strains with discordant subtypes were found by p17 and p24 regions of HIV-1 gag gene analysis, respectively. The percentage of subtype discordance found in this study was lower than that reported during the same period from a cohort study of IDU in Bangkok during 1995-1996 and a prime-boost phase III vaccine trial during 1998-2001 (Ramos et al., 2003; Watanaveeradej et al., 2006). The lower percentage of subtype discordance found in this study may come from analysis of fewer regions of the HIV-1 genome because full length genome analysis was performed in previous studies. The reverse transcriptase region (RT) of the HIV-1 pol gene was used to analyze discordant subtype of isolates that have a recombinant genome. However, no subtype discordance was found between the env and pol genes in this study, which may be due to the lower genetic diversity of the pol gene (Sirivichayakul et al., 2001). The discovery of recombinant strains of HIV-1 in this study by parallel phylogenetic analysis indicates that current subtype screening by simple genotyping may not be sufficiently effective to capture the increasing complexity of recombinant genomes. However, large-scale determination of full-length HIV-1 nucleotide sequences, which is the method used in recombinant genome studies, is not practical for molecular epidemiological investigation. Thus, parallel phylogenetic analysis of HIV-1 env and
gag genes for subtype determination was proposed to be a protocol of choice for an HIV-1 subtype or recombinant strain surveillance program.

Four recombinant HIV-1 isolates were identified in this study from three regions of Thailand, northern, northeastern, and southern regions. Interestingly, all recombinant HIV-1 isolates were from provinces that share their border with neighboring countries. Moreover, the recombination breakpoint could be identified in the gag gene in one recombinant isolate from Nakhon Si Thammarat Province in the south. The recombinant breakpoint of the isolate with the recombinant gag gene was identified at nucleotide sequence position 380 bp which was in the p24 region and the breakpoint was supported by the phylogenetic tree for the corresponding position.

During the years 1995-2000, genetic analysis of HIV-1 CRF01_AE in Thailand was mainly done on strains isolated from Bangkok and Chiang Mai in the north while the epidemiological patterns in the other epidemic areas, including border areas, have not been reported. The HIV-1 CRF01_AE isolated from samples collected in 2002 from 20 provinces in 3 major epidemics including the border areas of Thailand-Loa PDR, Thailand-Vietnam, Thailand-Myanmar, Thailand-Cambodia, and coastal provinces were studied. Phylogenetic trees of CRF01_AE isolates demonstrated that there was no geographic distinction among HIV-1 CRF01_AE isolates in Thailand. The clustering pattern found in the phylogenetic tree for each region were supported by a signature sequence pattern and N-linked glycosylation. The less genetic diversity in the southern CRF01_AE isolates than the northern and northeastern isolates was confirmed by the number of variable positions. Signature sequence pattern analysis demonstrated the GPGQ V3 loop tip was a predominant pattern in HIV-1 isolated from all geographic regions. The differences in both genetic diversity and genetic patterns of CRF01_AE circulated in the south indicates the possibility the HIV-1 epidemic in the south had its own pattern of geographic introduction and distribution.

It was reported that some of HIV-1 CRF01_AE isolates in China spreaded from Thailand to Cambodia to Vietnam and then to China through sexual contact. They were thought to be connected to sexual networks and the HIV-1 epidemic in these locations (Entz et al, 2000). These reports may explain the findings in this study that northern and northeastern isolates form one distinct HIV-1 cluster, which indicates the distribution of HIV-1 CRF01_AE among foreign workers, internal migrants and commercial sex workers along Thailand-Myanmar-Vietnam borders, while southern isolates may be from HIV-1 distribution among fishermen and the population in coastal provinces. The migration of workers or travellers and the link to commercial sex worker contacts may play an important role in the HIV-1 epidemic in Thailand and neighboring countries. Therefore, a better understanding of the social, behavioral and demographic factors associated with HIV-1 transmission among internal migrating workers, which could not be done in this study, will provide insight into the HIV-1 epidemic in the rural area.

Molecular epidemiological data from this study may be used to design surveillance methods and target prevention efforts. The increasing number of recombinant strains of HIV-1 in border and coastal areas suggests that mobile population, such as migrants or foreign workers, and fishermen, play a special role in the HIV-1 epidemic. Moreover, the HIV-1 epidemic in Southeast Asia was reported to be centered in Thailand and becoming more widespread, as well as more complex in terms of the genetic diversity of the circulating strains (Takebe et al, 2003). The molecular epidemiology of circulating strains...
of HIV-1 in Thailand from this study provide information on new antigenic circulating strains, which may complicate the development of an effective vaccine. This study gives insight into the epidemiology of virus strains responsible for the HIV-1 epidemic in Thailand.

ACKNOWLEDGEMENTS

This work was supported by the Bureau of AIDS, TB, and STIs, Ministry of Public Health, Thailand and The Royal Golden Jubilee PhD (RGJ-PhD) Program, The Thailand Research Fund.

REFERENCES


Felsenstein J. PHYLIP. Phylogenetic Inference Package. In. 3.5c ed. Seattle: Department of Genetics, University of Washington, 1993.


Magiorkinis G, Paraskevis D, Magiorkinis E, Vandamme AM, Hatzakis A. Reanalysis of the HIV-1 circulating recombinant form A/E (CRF01_AE): evidence of A/E/G recombina-


