厚生労働科学研究費補助金（政策創薬総合研究事業）
分担研究報告書

ランダムアプローチによるHIV-1 Nef蛋白を標的とした新規抗HIV-1薬のスクリーニング

分担研究者　岡田誠治　熊本大学エイズ学研究センター予防開発分野
研究協力者　鈴　伸也　熊本大学エイズ学研究センター予防開発分野

研究要旨　HIV-1 Nef蛋白を標的とした新たな抗HIV-1薬を開発するために、Nef活性化誘導型細胞株（TF-1-fms－Nef）を用いたバイオアッセイによるスクリーニング系を用いて低分子物質ライブラリーのスクリーニングを行った。約２万種のライブラリーのうち、1.6万種のスクリーニングを終了し、３種類の候補物質を得ることができた。現在候補物質の作用機序について検討中である。

A. 研究目的

HIV-1の感染者は年々その数が増加し、感染者は全世界で４千万人を超え、近年インドや中国では著しくHIV-1感染者が増加している。本邦においてもHIV-1感染者は１万人を突破し、その数は徐々に増えている。有効な治療法がなかったエイズも、有効な薬剤の開発によりその生命予後は著しく改善しいる。しかし、現在の治療法ではHIV-1の完全な排除は事実上不可能であり、薬剤の長期投与による副作用や薬剤耐性など多くの問題が生じている。現在の治療薬は、HIV-1の生活環を阻害する薬剤であり、作用機序の異なる新たな薬剤の開発が望まれている。

HIV-1は、いくつかのアクセサリー蛋白を有しており、これらのアクセサリー蛋白が、エイズの病態に深く関わっていることが知られている。特に、Nef蛋白は、HIV-1の複製能を高めるなどエイズの発症を助長する重要な病原因子である。そのため、Nef蛋白を分子標的とした化合物は、新たな作用機序を持つ抗HIV-1薬として潜伏感染や薬剤耐性ウィルスへの効果が期待される。本研究では、Nef蛋白を標的とした新規薬剤開発を目指して、HIV-1の重要な標的細胞であるマクロファージ系細胞におけるNef蛋白の影響を指標に、低分子物質ライブラリーの大規模スクリーニングを行った。

B. 研究方法

1) Nef蛋白を標的とした抗HIV-1薬のランダムライブラリー大規模スクリーニング

Nef活性化誘導型細胞株（TF-1-fms-Nef）がNefの活性化により増殖抑制をきたすことを指標に、新たなバイオアッセイ系を樹立した。本系では、添加薬剤によりNefの活性化が阻害されると、TF-1-fms-Nefの増殖抑制が解除されることを指標にしたバイオアッセイである。TF-1-fms-Nefを96穴プレートに撒き、合成エストロジェンであるタモキシフェンと低分子物質を添加し、MTT法により細胞増殖を計測し、タモキシフェン単時投与と比べて細胞増殖が回復している低分子化合物を陽性として２次スクリーニングと機能検査をおこなった。

2) Nef蛋白, Hck, M-CSF受容体蛋白の結合阻害を指標とした二次スクリーニング系の樹立

Nef蛋白は、そのProline-rich領域と非受容体型チロシンキナーゼHckのSH3領域を介して結合する事が既に明らかになっている。そこで、GST蛋白質と融合させたNef蛋白を用いたPull-down法による二次スクリーニング法を樹立した。一方、Nef蛋白とチロシンキナーゼの結合阻害以外の機序によるNefの機能阻害薬を、Nef蛋白の機能であるCD4やMHC class Iの発現低下を阻害することを指標にフローサイトメトリーを用いて行う。

（倫理面への配慮）

本年度は、細胞株を用いたバイオアッセイと蛋白解析のみであり、倫理面に配慮が必要な研究は行っていない。

C. 研究結果

1)Nef蛋白を標的とした抗HIV-1薬のランダムライブラリー大規模スクリーニング

Enamine低分子化合物ライブラリーのスクリーニングを行った。２万種類のうち1.6万種類のスクリーニングを終了し、数種類の陽性化合物を得た。これらの化合物を濃度を変えて再検査し、有為に活性を示すものとして、これまでに３種類の化合物を同定している。

2) Nef蛋白, Hck, M-CSF受容体蛋白の結合阻害を指標とした二次スクリーニング系の樹立

GST蛋白質と融合したNef蛋白を作成し、Pull-down法によりHckとの結合阻害を蛋白レベルで解析する系を樹立した。本システムを用いて既知のSH3領域結合阻害剤が、Nef蛋白とHckの結合を阻害することを確認した。今後、本システムを用いて上記一次スクリーニングにより同定された化合物の二次スクリーニングを行う予定である。

D. 考察

HIV-1 Nef蛋白を標的とした新たな抗HIV-1の開発を目的に、バイオアッセイにより低分子物質ライブラリーの大規模スクリーニングを行った。本スクリーニング系は、Nef活性化誘導型細胞株（TF-1-fms-Nef）が、Nefの活性化により細胞増殖が抑制され、Nefの機能阻害剤が存在すると細胞増殖抑制が解除されるというバイオアッセイである。既知のSH3阻害剤を用いると容量依存性に細胞増殖抑制が

解除されたことから、本スクリーニングシステムは充分に機能すると評価した。低分子物質ライブラリーのスクリーニングから、現在までに３種類の候補物質が同定され、すべての一次スクリーニング終了後、二次スクリーニングを行う予定である。

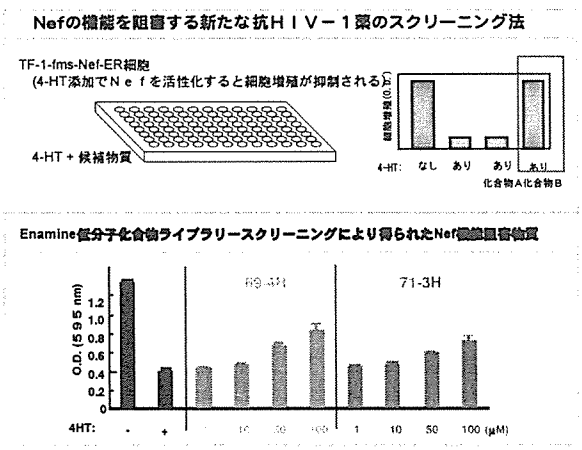
また、Nef 蛋白と Hck の結合阻害を指標として、Pull-down 法による二次スクリーニング系を樹立した。SH3 阻害薬を加えることにより Nef と Hck の結合が阻害されることから、この系が機能することを確認した。今後、一次スクリーニングで同定された低分子物質の解析を行っていく予定である。

一方、Nef 蛋白の機能には、CD4 や MHC class I の発現低下、HIV-1 の複製増加などチロシンキナーゼとの結合を介さない機能もある。今後フローサイトメトリーを用いて、これらの指標を元に新たなスクリーニングを行うことを考えている。

HIV-1 Nef 蛋白を標的とした薬剤は、これまでに開発された薬剤とまったく作用機序が異なる抗 HIV-1 薬として、既存の薬剤との併用療法で効果が期待される。また、その作用機序から潜伏感染や薬剤耐性ウィルスに対する効果も期待できることから、その開発が急がれる。

E. 結論

Nef 蛋白を標的とした抗 HIV-1 薬開発のために、低分子物質ライブラリーの大規模スクリーニングを行い、３種類の候補物質を得ることができた。今後スクリーニングの継続と候補物質の作用機序の解明を行っていく予定である。



F. 研究発表

1. 論文発表

1) Suzu S, Hiyoshi M, Yoshidomi Y, Harada H, Takeya M, Kimura F, Motoyoshi K, and Okada S; M-CSF-mediated macrophage differentiation but not proliferation is correlated with increased and prolonged ERK activation. ***J Cell Physiol*** *in press*

2. 学会発表

1) Seiji Okada, Hideki Harada, Masateru Hiyoshi, Shinya Suzu. Selective inhibition of receptor pathways for macrophage-specific cytokines by HIV-1 Nef protein. XVI International AIDS Conference. (Toronto, Canada) Aug.13-18, 2006.
2) Shinya Suzu, Masateru Hiyoshi, Hideki Harada, Seiji Okada. HIV-1 Nef inhibits M-CSF signals by down-regulating its receptor Fms through Hck. 48th Annual Meeting of American Society of Hematology. (Orlando, USA) Dec. 9-12, 2006.
3) Masateru Hiyoshi, Shinya Suzu, Yuka Yoshidomi, Hideki Harada, Seiji Okada. HIV-1 Nef inhibits macrophage colony-stimulating factor signals by down-regulating its receptor through Src family kinases. 14th Conference on Retroviruses and Opportunistic Infections. (Los Angels, USA), 26-28 Feb. 2007.
4) 鈴　伸也、日吉真照、吉富友香、原田英樹、岡田誠治. HIV-1 Nef による M-CSF/M-CSF レセプター経路の選択的阻害とその分子機構. 第 68 回日本血液学会総会、2006 年 10 月 6-8 日、福岡
5) 日吉真照、鈴　伸也、吉富友香、原田英樹、岡田誠治. Nef は Hck を介して M-CSF 受容体

の細胞内における成熟過程を阻害する. 第 54 回日本ウィルス学会学術集会、2006 年 11 月 19-21 日、名古屋
6) 鈴　伸也、日吉真照、吉富友香、原田英樹、岡田誠治. Nef は M-CSF 受容体の発現低下を誘導する. 第 20 回日本エイズ学会学術集会総会、2006 年 11 月 30 日—12 月 2 日、東京

G. 知的財産権の出願・登録状況

1. 特許取得：なし
2. 実用新案登録：なし
3. その他：なし

Ⅲ. 研究成果の刊行に関する一覧表

平成１８年度 研究成果の刊行に関する一覧表

発表論文リスト（2006-2007 年度）

主任研究者：武部　豊
分担研究者：西郷　薫・駒野　淳・高橋　信弘・山越　智・村上　努・
森川　裕子・増田　道行・岡田　誠治
協力研究員：内藤　雄樹・程　久美子

論文発表

1. Naito, Y., Ui-Tei, K., Nishikawa, T., Takebe, Y., Saigo, K. (2006). siVirus: web-based antiviral siRNA design software for highly divergent viral sequences. Nucl. Acid Res. (Web Server issue) W448-W450.

2. Takebe, Y. and Telesnitsky, A. (2006). Evidence for the acquisition of multidrug resistance by an HIV clinical isolate via human gene transduction. Virology 351: 1-6.

3. Tee, K. K., Li, X.-J., Nohtomi, K., Ng, K. P., Kamarulzaman, A., and Takebe, Y. (2006). Identification of a novel circulating recombinant form (CRF33_01B) disseminating widely among various risk populations in Kuala Lumpur, Malaysia. J. AIDS 43(5): 523-9.

4. Ui-Tei K, Naito Y, Saigo K. Essential notes regarding the design of functional siRNAs for efficient mammalian RNAi. *J. Biomed. Biotechnol.* **2006**, 65052 (2006).

5. Futahashi Y, Komano J, Urano E, Aoki T, Hamatake M, Miyauchi K, Yoshida T, Koyanagi Y, Matsuda Z, Yamamoto N.(2006). Separate elements are required for ligand-dependent and –independent internalization of metastatic potentiator CXCR4. Cancer Science. (In press)

6. Shimizu S, Urano E, Futahashi Y, Miyauchi K, Isogai M, Matsuda Z, Nohtomi K, Onogi T, Takebe Y, Yamamoto N, Komano J. (2006). Inhibiting lentiviral replication by HEXIM1, a cellular inhibitor of cdk9/cyclinT complex. AIDS. (In press)

7. Miyauchi K, Komano J, Myint L, Futahashi Y, Urano E, Matsuda Z, Chiba T, Miura H, Sugiura W, Yamamoto N. (2006). Rapid propagation of low-fitness drug resistant mutants of human immunodeficiency virus type 1 by a streptococcal metabolite sparsomycin. Antivir Chem Chemother. 17(4): 167-174.

8. Miyauchi K, Curran R, Matthews E, Komano J, Hoshino T, Engelman DM, Matsuda Z. (2006). Mutations of conserved glycine residues within the membrane-spanning domain of human immunodeficiency virus type 1 gp41 can inhibit membrane fusion and incorporation of Env onto virions. Jpn J Infect Dis. Apr;59(2):77-84.

9. Sekiguchi, T., Hayano, T., Yanagida, M., Takahashi, N., Nishimoto, T. (2006) NOP132 is required for proper nucleolus localization of DEAD-box RNA helicase DDX47. Nucleic Acids Res. 34(16): 4593-4608.

10. Murakami Y, Yamagoe S, Noguchi K, Takebe Y, Takahashi N, Uehara Y, Fukazawa H. Ets-1-dependent expression of vascular endothelial growth factor receptors is activated by latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus through interaction with Daxx. J Biol Chem. 281: 28113-28121, 2006.

11. Murakami T, and Yamamoto N. AIDS: How Do We Overcome This Social or Biodisaster (Review)? The Journal of Disaster Research, In press, 2007

12. Sakuragi S, Sakuragi J.-I, Morikawa Y, & Shioda T.
Development of a rapid and convenient method for the quantitation of HIV-1 budding. Microbes Infect. 8: 1875-1881 (2006)

13. 森川　裕子. 酵母を用いた動物ウイルスの研究. ウイルス 56: 9-16 (2006)

14. Suzu S, Hiyoshi M, Yoshidomi Y, Harada H, Takeya M, Kimura F, Motoyoshi K, and Okada S; M-CSF-mediated macrophage differentiation but not proliferation is correlated with increased and prolonged ERK activation. J Cell Physiol in press

Ⅳ. 研究成果の刊行物・別刷

W448–W450 *Nucleic Acids Research, 2006, Vol. 34, Web Server issue*
doi:10.1093/nar/gkl214

siVirus: web-based antiviral siRNA design software for highly divergent viral sequences

Yuki Naito[1,*], Kumiko Ui-Tei[1,2], Toru Nishikawa[3], Yutaka Takebe[4] and Kaoru Saigo[1]

[1]Department of Biophysics and Biochemistry, Graduate School of Science, [2]Undergraduate Program for Bioinformatics and Systems Biology, School of Science and, [3]Department of Computer Science, Graduate School of Information Science and Technology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan and [4]Laboratory of Molecular Virology and Epidemiology, AIDS Research Center, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan

Received February 14, 2006; Revised and Accepted March 24, 2006

ABSTRACT

siVirus (http://siVirus.RNAi.jp/) is a web-based online software system that provides efficient short interfering RNA (siRNA) design for antiviral RNA interference (RNAi). siVirus searches for functional, off-target minimized siRNAs targeting highly conserved regions of divergent viral sequences. These siRNAs are expected to resist viral mutational escape, since their highly conserved targets likely contain structurally/functionally constrained elements. siVirus will be a useful tool for designing optimal siRNAs targeting highly divergent pathogens, including human immunodeficiency virus (HIV), hepatitis C virus (HCV), influenza virus and SARS coronavirus, all of which pose enormous threats to global human health.

INTRODUCTION

RNA interference (RNAi) is now widely used to knockdown gene expression in a sequence-specific manner, making it a powerful tool not only for studying gene function, but also for therapeutic purposes, including antiviral treatments (1–4). Currently, the replication of a wide range of viruses can be inhibited successfully using RNAi, with both short interfering RNAs (siRNAs) and siRNA expression vectors (5).

In mammalian RNAi, the efficacy of each siRNA varies widely depending on its sequence; only a limited fraction of randomly designed siRNAs is highly effective. Many experiments have been conducted to clarify possible sequence requirements of functional siRNAs. Of these, our work incorporates guidelines from three major studies (6–8) of selecting functional siRNAs. However, designing functional siRNAs that target viral sequences is problematic because of their extraordinarily high genetic diversity. For example, about 500 entries of near full-length sequences of HIV-1 group M, which is largely responsible for global pandemic, are stored in the sequence databases, but it proved impossible to select a common 21mer from among all of them. Moreover, RNAi-resistant viral mutants achieved through point mutation or deletion emerge rapidly when targeting viruses in cell culture. These problems suggest a strong need to select highly conserved target sites for designing antiviral siRNAs. Furthermore, the off-target silencing effects of siRNA are also a serious problem that could affect host gene expression (9). Off-target silencing effects arise when an siRNA has sequence similarities with unrelated genes. In antiviral RNAi, it is desirable to minimize off-target effects against human genes.

Consequently, only a limited fraction of 21mers is suitable for use as antiviral siRNAs. In this study, we developed a novel web-based online software system, siVirus, which provides functional, off-target minimized siRNAs targeting highly conserved regions of divergent viral sequences.

METHODS

Selection of highly conserved siRNA target sites

Highly conserved siRNA sequences are selected based on their *degree of conservation*, defined as the proportion of viral sequences that are targeted by the corresponding siRNA, with complete matches (i.e. 21/21 matches). All possible siRNA candidates targeting every other position of user-selected viral sequences are generated and their degrees of conservation are computed. Users can arbitrarily specify a set of viral sequences for the computation; e.g. sequences can be selected from a specific geographic region(s) or a specific genotype(s) to design the best siRNAs tailored to specific user needs. siVirus also accepts user's own sequences in a multi-FASTA format and shows whether each siRNA can target the posted sequences.

*To whom correspondence should be addressed. Tel: +81 3 5841 4404; Fax: +81 3 5841 4400; Email: y-naito@RNAi.jp

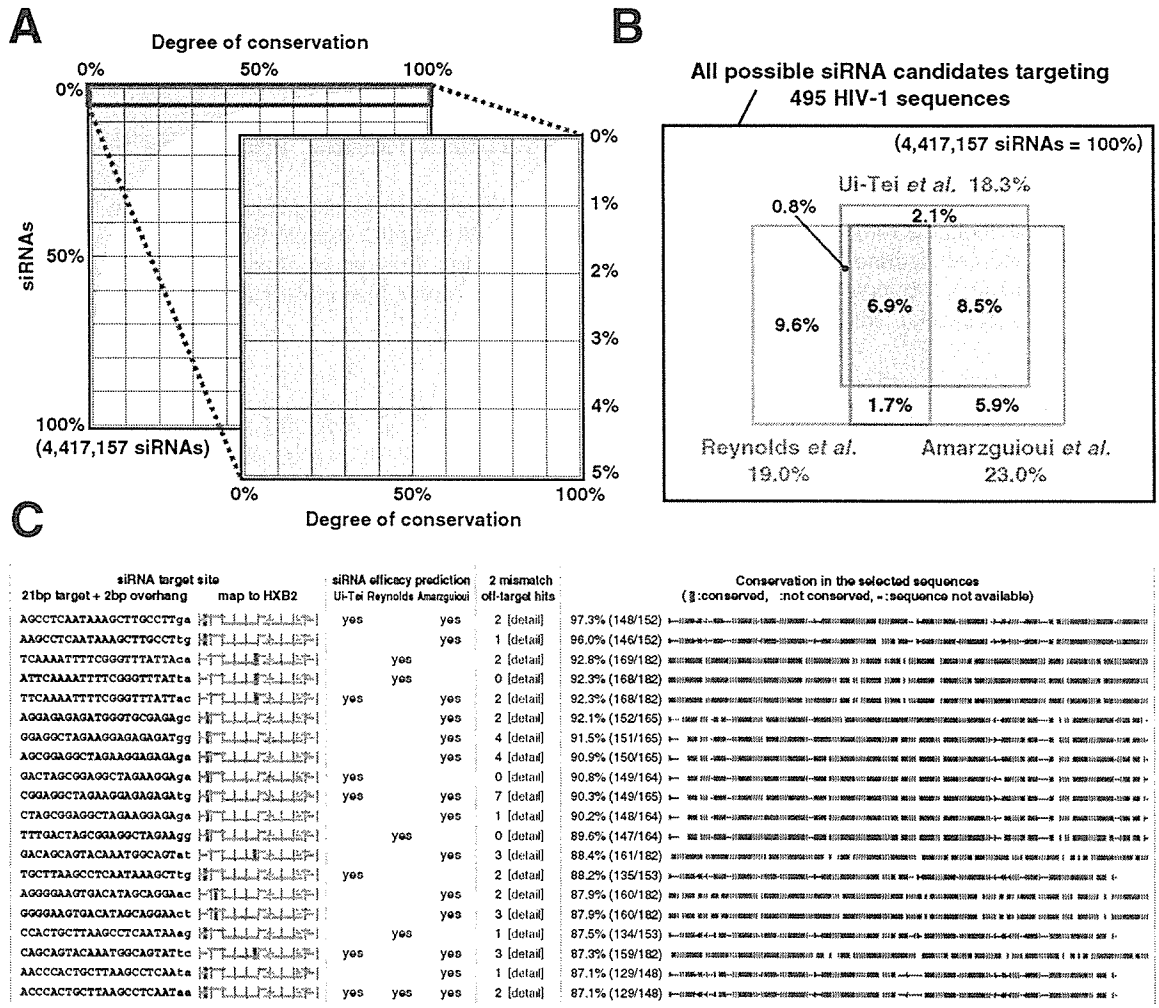


Figure 1. (**A**) The degree of conservation is calculated for all possible siRNA candidates (total 4 417 157) targeting every other position of 495 HIV-1 sequences. (**B**) The efficacy predictions of these 4 417 157 siRNA candidates based on three different guidelines: Ui-Tei *et al.* (6), Reynolds *et al.* (7) and Amarzguioui *et al.* (8). (**C**) Typical output of siVirus for designing anti-HIV siRNAs. Sequence information, efficacy predictions, off-target search results and the degrees of conservation are shown.

siRNA efficacy prediction

In mammalian RNAi, the efficacy of each siRNA varies markedly depending on its sequence; hence, several groups have reported guidelines for selecting functional siRNAs. siVirus incorporates the guidelines of Ui-Tei *et al.* (6), Reynolds *et al.* (7) and Amarzguioui *et al.* (8) and shows whether each siRNA satisfies these guidelines.

Off-target searches

Off-target searches were performed for each siRNA using siDirect (10,11). siVirus shows the number of off-target hits within two mismatches against the non-redundant database of human transcripts (10).

Database maintenance

Currently, siVirus incorporates viral genome sequences of HIV-1, HCV, influenza A virus and SARS coronavirus. These sequences were downloaded from the Los Alamos HIV Sequence Database (http://hiv-web.lanl.gov/), the Los Alamos HCV Sequence Database (12), the NCBI Influenza Virus Sequence Database (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html), and NCBI GenBank (13), respectively. siVirus will be updated continuously as these databases are revised. We also plan to incorporate other viruses if sufficient numbers of their sequences are available.

RESULTS AND DISCUSSION

To design anti-HIV siRNA, we analyzed the 495 near full-length HIV-1 sequences listed in Supplementary Table 1. A total of 4 417 157 possible siRNA candidates (i.e. substrings of length 21) targeting every other position of the HIV-1 sequences were produced from the 495 viral sequences. The analysis of these siRNA candidates revealed that highly

conserved siRNAs constituted only 0.3% of the possible siRNAs if >90% conservation is expected (Figure 1A). The fraction is still as small as 0.8% even if the threshold of the conservation is relaxed to 80%. On the other hand, siRNAs predicted to be functional by one or more guidelines (6–8) constituted 35.5% of the 4 417 157 siRNAs (Figure 1B). Taken together, siRNAs that are >80% conserved, and satisfy at least one guideline constitute only 0.2% of the siRNAs. In this condition, 20–30 siRNAs can be designed for each full-length sequence of HIV-1. These indicate that most of the randomly designed siRNAs are not suited for targeting HIV-1 efficiently.

Figure 1C shows typical output from siVirus for designing anti-HIV siRNAs. A total of 182 sequences from HIV-1 subtypes B, C and CRF01_AE, which are the most prevalent HIV-1 genotypes circulating in Asia, were selected. The results were sorted by their degree of conservation, and filtered to display siRNAs that satisfy at least one efficacy guideline. The off-target search results against human genes are also shown. It is desirable to select an siRNA that has less off-target hits.

To test the validity of siVirus, 35 siRNAs satisfying the guideline by Ui-Tei *et al.* (6) were designed against the conserved regions of HIV-1 genomes using siVirus and were assayed for inhibition of viral replication. Among them, 31 siRNAs effectively inhibited HIV-1 replication by >80% when each siRNA duplex was transfected at 5 nM (Y. Naito, K. Ui-Tei, K. Saigo and Y. Takebe, unpublished data).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan to K.S., K.U.-T. and Y.T., and by grants from the Ministry of Health, Labour and Welfare of Japan to Y.T. Funding to pay the Open Access publication charges for this article was provided by the Ministry of Education, Culture, Sports, Science and Technology of Japan. Y.N. is a Research Fellow of the Japan Society for the Promotion of Science.

Conflict of interest statement. None declared.

REFERENCES

1. Fire,A., Xu,S., Montgomery,M.K., Kostas,S.A., Driver,S.E. and Mello,C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, **391**, 806–811.
2. Mello,C.C. and Conte,D. Jr (2004) Revealing the world of RNA interference. *Nature*, **431**, 338–342.
3. Hannon,G.J. and Rossi,J.J. (2004) Unlocking the potential of the human genome with RNA interference. *Nature*, **431**, 371–378.
4. Voinnet,O. (2005) Induction and suppression of RNA silencing: insights from viral infections. *Nature Rev. Genet.*, **6**, 206–220.
5. Leonard,J.N. and Schaffer,D.V. (2006) Antiviral RNAi therapy: emerging approaches for hitting a moving target. *Gene Ther.*, **13**, 532–540.
6. Ui-Tei,K., Naito,Y., Takahashi,F., Haraguchi,T., Ohki-Hamazaki,H., Juni,A., Ueda,R. and Saigo,K. (2004) Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res.*, **32**, 936–948.
7. Reynolds,A., Leake,D., Boese,Q., Scaringe,S., Marshall,W.S. and Khvorova,A. (2004) Rational siRNA design for RNA interference. *Nat. Biotechnol.*, **22**, 326–330.
8. Amarzguioui,M. and Prydz,H. (2004) An algorithm for selection of functional siRNA sequences. *Biochem. Biophys. Res. Commun.*, **316**, 1050–1058.
9. Jackson,A.L. and Linsley,P.S. (2004) Noise amidst the silence: off-target effects of siRNAs? *Trends Genet.*, **20**, 521–524.
10. Naito,Y., Yamada,T., Ui-Tei,K., Morishita,S. and Saigo,K. (2004) siDirect: highly effective, target-specific siRNA design software for mammalian RNA interference. *Nucleic Acids Res.*, **32**, W124–W129.
11. Yamada,T. and Morishita,S. (2005) Accelerated off-target search algorithm for siRNA. *Bioinformatics*, **21**, 1316–1324.
12. Kuiken,C., Yusim,K., Boykin,L. and Richardson,R. (2005) The Los Alamos hepatitis C sequence database. *Bioinformatics*, **21**, 379–384.
13. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2006) GenBank. *Nucleic Acids Res.*, **34**, D16–D20.

Available online at www.sciencedirect.com
SCIENCE DIRECT
VIROLOGY
ELSEVIER
Virology 351 (2006) 1–6
www.elsevier.com/locate/yviro

Rapid Communication

Evidence for the acquisition of multi-drug resistance in an HIV-1 clinical isolate via human sequence transduction

Yutaka Takebe [a], Alice Telesnitsky [b,*]

[a] *Laboratory of Molecular Virology and Epidemiology, AIDS Research Center, National Institute of Infectious Diseases, Tokyo 162-8640, Japan*
[b] *Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109-0620, USA*

Received 29 March 2006; returned to author for revision 3 May 2006; accepted 9 May 2006
Available online 14 June 2006

Abstract

Insertions in HIV-1 reverse transcriptase's fingers subdomain can enhance chain terminator excision and confer resistance to multiple nucleoside analogs. Inserts that resemble flanking sequences likely arise by local sequence duplication. However, a remarkable variety of non-repeat fingers insertions have been observed. Here, molecular epidemiology, sequence analyses and mechanistic modeling were employed to show that one Japanese isolate's RT fingers insert likely resulted from non-homologous recombination between virus and host sequences and the transductive copying of 37 nucleotides from human chromosome 17. These findings provide evidence that human sequence transduction can, at least rarely, contribute to genetic and phenotypic variation in pandemic HIV.

Keywords: Genetic recombination; Retroviruses; Reverse transcriptase; Drug resistance

Introduction

Much of HIV-1's genetic variation arises by stepwise – albeit at times hypermutation-accelerated – accumulation of point mutations (Harris et al., 2003; Leitner and Albert, 1999). Genetic recombination also contributes to HIV genetic diversity and occurs about 10-fold more frequently than base substitution (An and Telesnitsky, 2002; Jetzt et al., 2000). Because fewer molecular events are required to introduce panels of mutations by recombination than by serial mutation, clustered genome changes are generally believed to reflect recombination or related rearrangements (Malim and Emerman, 2001; Wain-Hobson et al., 2003). Retroviral recombination results from template switching during reverse transcription and generally occurs in regions of high sequence similarity between the two intact genomes each retrovirus co-packages (An and Telesnitsky, 2002). Non-homologous recombination, often guided by microhomology between donor and acceptor templates, can also occur (Hajjar and Linial, 1993; Zhang and Temin, 1993). Deletions and insertions can arise via non-homologous recombination between discontinuous portions of the viral genome (Parthasarathi et al., 1995; Temin, 1993). Duplications result if template switch occurs from one RNA position to a locus further downstream on the co-packaged RNA, while deletions arise when reverse transcriptase "jumping" bypasses sequences and terminates upstream of the point of template departure (Parthasarathi et al., 1995). Insertion-in-a-deletion or insertion-in-a-duplication mutations can result from a series of non-homologous crossovers (Lobato et al., 2002; Parthasarathi et al., 1995; Pathak and Temin, 1990). Either virus or host sequences can template insertions, as postulated by models for oncogene transduction (Muriaux and Rein, 2003). Although whole gene transduction is rare, incorporation of short host segments into defective viral genomes is observed fairly frequently (Dunn et al., 1992; Fang and Pincus, 1995; Hajjar and Linial, 1993; Mikkelsen and Pedersen, 2000; Pulsinelli and Temin, 1991; Sun et al., 2001). Thus, experimentally, the use of host sequences to bridge non-homologous recombination junctions and the insertion of host segments at strong stop or non-homologous crossover sites is a well-established

* Corresponding author. Fax: +1 734 764 3562.
E-mail addresses: takebe@nih.go.jp (Y. Takebe), ateles@umich.edu (A. Telesnitsky).

0042-6822/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2006.05.009

phenomenon. Possible contributions of host sequences to HIV clinical isolates has been less clear: both because pre- and post-mutation sequence data are seldom available and because inserts maintained in replication-competent virus are generally too short to implicate a specific template (Winters and Merigan, 2005).

Results

In this report, we examined an unusually lengthy insertion mutation in an HIV-1 isolate from a Japanese child, with the goal of elucidating its likely origins (Sato et al., 2001). Epidemiologic investigation provided sequence information for the viral populations observed in the patient, referred to as NH3, during the 6 years prior to and at the onset of clinical drug resistance. These included the multi-drug-resistant isolate itself (designated 99JP-NH3-II) and several co-circulating insertion region variants (Fig. 1A). Additional sequence information was available for isolates from NH3's father, NH1, who had been infected in Thailand, and from his mother, NH2, who had contracted HIV-1 from NH1 and subsequently infected NH3 through maternal transmission. This clinical history confirmed that the insertion arose during highly active antiretroviral treatment in NH3 (Sato et al., 2001). A comparison of virus with and without the insert demonstrated that the 11-amino-acid insertion in the reverse transcriptase of this CRF001_AE circulating recombinant form variant contributed to its high-level resistance to multiple nucleoside analogs.

The region encompassing the 33 nucleotides inserted in NH3's RT gene was evaluated for similarity to experimentally described retroviral recombination products. The insertion did not mirror flanking sequences and thus was not a simple duplication (Parthasarathi et al., 1995). The entire sequence of the 99JP-NH3-II genome was available and none of it resembled the 33-nucleotide insert. Thus, sequential template

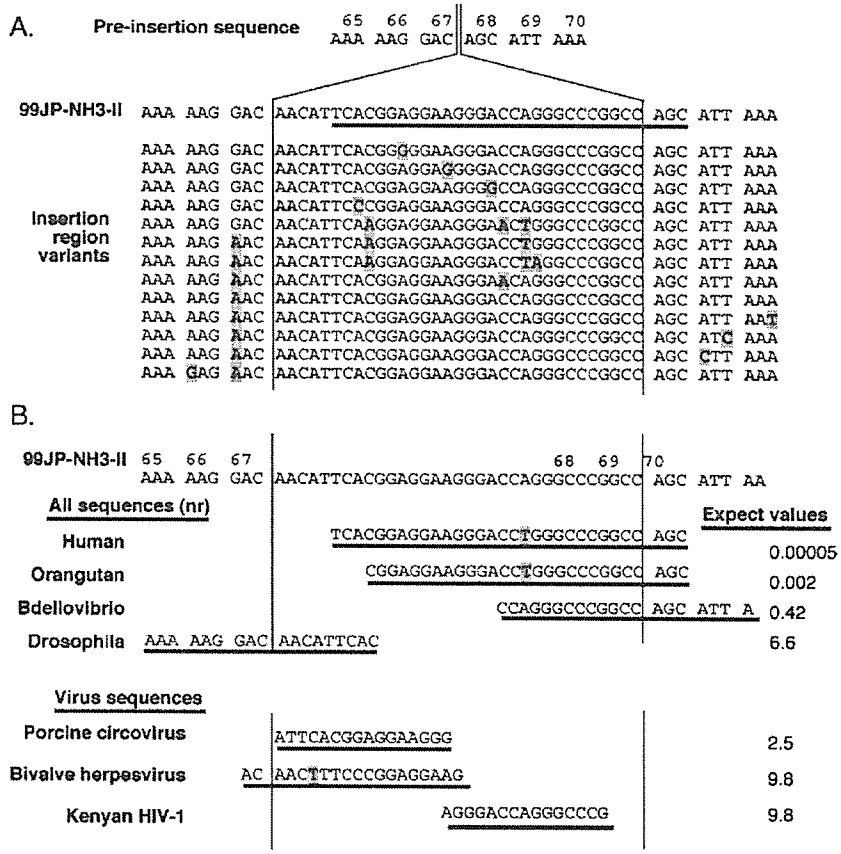


Fig. 1. Sequence analyses of the 99JP-NH3-II insertion region. (A) Deduced preinsertion sequence, sequence insertion in 99JP-NH3-II, and insertion region variants. The NH3 isolates shown are GenBank accession nos. AB053002, AB053074, AB053071, AB053005, AB053038, AB053073, AB053065, AB053035, AB053084, AB053003, AB053077, AB053068, AB053058, and AB053083. Numbers at top represent amino acid residue numbers within reverse transcriptase. Line under 99JP-NH3-II is region of greatest human match. Shaded sequences are positions that differ from 99JP-NH3-II. (B) Comparisons of the 99JP-NH3-II insertion region to other sequences in GenBank. Underlined sequences indicate greatest match to the 51-base insertion region of 99JP-NH3-II. At the top ('All sequences'): blastn results from nr GenBank. The human match encompasses an intron/exon junction in AC135178 (Chromosome 17 open reading frame 68; also found in NC_00017); the orangutan match lies within CR857769; *Bdellovibrio*: BX842654, *Drosophila*: AY183918. Expect values at left are those assigned by blastn, as described in Materials and methods. Note that although human and chimpanzee sequences are highly related in the 3′ regions of human chromosome 17 orf 68, the orf 68 5′ region differs markedly between human and chimpanzee, and thus, no match between the 99JP-NH3-II insert and GenBank chimpanzee sequences is detectable. At the bottom of panel B: blastn results from nr GenBank limited to viruses, after manual removal of NH3 sequences. Accession number for the circovirus: AY321985; for herpesvirus: AY509253; for Kenyan HIV-1: AF457073.

switches among viral segments could not account for NH3's RT insert.

Support for possible non-viral origins of the 33-base insertion came from its striking nucleotide composition. Whereas typical HIV sequences have a <40% G + C content (Berkhout et al., 2002), this insert was 67% G + C. Because abnormal G + C content is a hallmark of horizontal gene transfer (Hacker and Kaper, 2000), we explored possible human origins for the insert. If microhomology guided template switch had contributed to insert generation, viral sequences flanking the insert would be predicted to retain human sequence homology. To accommodate this, a 51-base segment comprised of the 33-base insert plus nine nucleotides from either flank was used in the database searches summarized in Fig. 1B.

When this 51-base segment was used to query unrestricted databases in GenBank, the best match was to 99JP-NH3-II itself, followed closely by the insert variants isolated from NH3 (Fig. 1A). The similarity of these NH3 isolates to one another and the differences between their inserts and all other HIV sequences (see below) makes it reasonable to assume that a single insertion event gave rise to all insert variants in patient NH3: an interpretation also supported by our previous studies on viral molecular evolution in NH3 (Sato et al., 2001).

Excluding NH3's HIV strains, the closest GenBank match to this 51-base sequence was a 30/31-base match to a non-repetitive sequence on human chromosome 17 (Fig. 1B). This was followed by a 27/28 match to the orthologous locus in orangutan, with the remaining top matches including a 20/20 match to the genome of the predatory bacterium *Bdellovibrio*, followed by less extensive matches to sequences from *Drosophila*, mouse, dog, and rice, but notably not from any non-NH3 HIV-1 isolates.

To assess similarity of the 99JP-NH3-II insert to sequences in other HIV isolates, the 51-base segment was subsequently BLASTed against all virus entries in GenBank ('Virus sequences' in Fig. 1B). Besides 99JP-NH3-II and the co-circulating isolates from NH3, no viral sequences in GenBank yielded expect values of <1. Because expect values are measures of probability roughly equivalent to *P* values, the absence of <1 matches in any isolate of any type of virus suggests that the insert in 99JP-NH3-II is even less similar to any known virus sequence than would be predicted by random chance. Whereas BLAST had assigned the 30/31 human genome match an expect score of 0.00005, the only HIV match assigned an expect score <10 was a 15/15-base match (expect score 6.6) to a portion of *env* in a clinical isolate from Kenya (Fig. 1B).

That the virus–human match included a few bases downstream of the insertion was consistent with the possibility that microhomology guided recombination between HIV-1 and the BLAST-identified human sequence generated this insertion. These observations and the mutation's structure suggested that the NH3 insertion was generated via the splinted non-homologous recombination model shown in Fig. 2. Briefly, an HIV-1 provirus was established on chromosome 17 just upstream of the putative insert-encoding sequences (Fig. 2A). Viral polyadenylation signal read-through generated a chimeric HIV-human RNA that became encapsidated. During subsequent reverse transcription, microhomology-guided template switching between portions of the RT gene and human sequences on the read-through RNA generated the observed insertion-in-a-duplication structure (Fig. 2B).

This model is based on experimental outcomes of non-homologous recombination, models for retroviral transduction, and properties of the putative human bridging template (An and Telesnitsky, 2002; Mikkelsen and Pedersen, 2000; Muriaux and Rein, 2003). Specifically, the putative human template straddles an intron/exon junction that lies in the antisense orientation of a mapped and verified mRNA encoding a putative protein of unknown function (Gao et al., 2005; Ota et al., 2004). Thus, although recombination between viral and unlinked host sequences has been reported previously (Sun et al., 2001), for the unspliced antisense sequences in this case, host sequences were more likely to have become encapsidated on a read-through transcript than as a free RNA (Muriaux and Rein, 2003).

Because it differs at three nucleotide positions from the founder strain predicted by our model (Fig. 2C), the postulated recombination events alone cannot explain the JP-NH3-II RT insertion mutation. However, the spectra of insert variants isolated from NH3 differed from one another at up to 4 positions within the examined sequence interval (Fig. 1A). Thus, these variants' sequence heterogeneity demonstrates that the extent of viral diversification required to generate JP-NH3-II's insert from the putative founder strain indisputably did occur within patient NH3 after initial insert acquisition. This supports the notion that JP-NH3-II arose via the mechanism outlined in Figs. 2A and B, followed by the introduction of point mutations at the positions boxed in Fig. 2C.

Discussion

The findings here analyzed a drug resistance-associated sequence insertion that is not closely related to sequences in any other HIV-1 isolate in GenBank. The closest match to this insert among all sequences in GenBank was to a portion of human chromosome 17. The structure of this mutation resembles insertion-within-a-duplication mutations that are well represented among defective retroviral replication products in the experimental literature. This insert and its flanking sequences are so dissimilar from one another that the alternate possibility for insert generation – local sequence duplication followed by mutation – can in large part be ruled out because the number of rare events required to generate the observed structure would far exceed those required by the postulated splinted recombination mechanism. Because RT is not known to polymerize more than a single nucleotide or two without a template, it is likely that all retroviral insertions longer than a couple of bases are synthesized using some form of template (Pathak and Temin, 1990; Preston and Dougherty, 1996). Thus, there is no precedence for *de novo* generation of a heteropolymeric insert of this length.

The insert examined in this study was located in RT's β3-β4 hairpin, a region where multiple drug resistance-associated

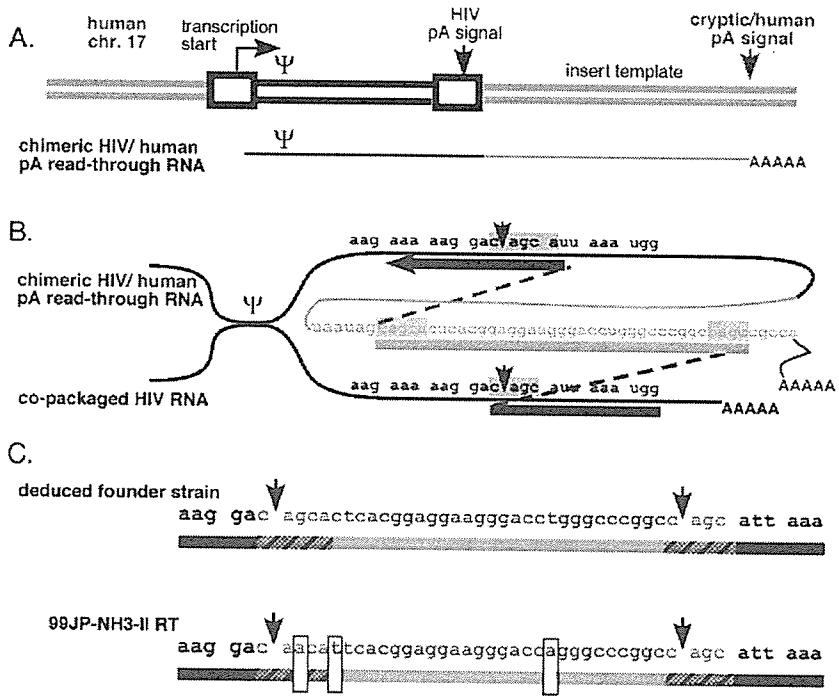


Fig. 2. Splinted recombination model for human sequence insertion. (A) Proviral insertion and read-through RNA. Not to scale. Boxes indicate HIV-1 long terminal repeats; Ψ indicates RNA packaging/dimerization signal. Thick lines are DNA; thin lines are RNA. Black lines are viral sequences; shaded lines are human sequences. Arrows indicate direction of transcription and positions of polyadenylation signals. (B) Template switches postulated to generate 99JP-NH3-II. Shown are two co-packaged RNAs joined at Ψ, and DNA products of template switching (thick and dotted lines: shaded boxes indicate regions of microhomology between viral and human sequences). Note that experimental studies have demonstrated that polyadenylation read-through occurs during the synthesis of about 10% of all HIV RNAs and that even unlinked host sequences can become incorporated in proviruses (Olsen et al., 1990; Sun et al., 2001; An and Telesnitsky, 2004 and references therein). (C) Genetic differences between 99JP-NH3-II and the predicted founder strain. Arrowheads indicate borders of the 33 base insert; hatched sequences were templated by microhomology, boxed residues are sites of mutations evoked to generate 99JP-NH3-II from the deduced founder strain.

insertions are observed in roughly 1% of the viral strains in treated HIV/AIDS patients world-wide (Winters and Merigan, 2005). Although a simple duplication of flanking sequences is sufficient to enhance drug resistance and may be the most frequent cause of β3- β4 insertions, a wide variety of β3- β4 insert sequences – varying in both length and sequence composition – have been observed (Winters and Merigan, 2005).

Identifying likely human origins of the duplication-flanked insert in 99JP-NH3-II was possible because this insert was unusually lengthy. However, alternate means of searching reveal that Genbank contains additional examples of direct repeat-flanked RT insertions, albeit with shorter sequence regions with uncertain origins (Masquelier et al., 2001; van der Hoek et al., 2005) (see Materials and methods). Although these latter inserts are too short for homology searching to implicate specific sequences, it seems reasonable to postulate that their synthesis involved either virus or host bridging templates.

Not surprisingly, when the human genome is used to query all HIV-1 sequences in GenBank, the few isolates that appear genuine and contain long (>50 nt) human inserts are annotated as replication-defective (for example, GenBank accession no. AY561239; see Materials and methods). Among replication-competent HIV-1 isolates, the inserts with strongest virus–human match belong to the class of mutations called AVT codon-rich *env* variable region extensions (Kitrinos et al., 2003). The codon bias differences between these inserts and other HIV-1 sequences, as well as the similarity of these inserts to repetitive human microsatellite sequences, have been described previously (An and Telesnitsky, 2004; Bosch et al., 1994; Kitrinos et al., 2003). However, most HIV AVT-rich extensions identified by BLAST using human query sequences have fewer than 30 contiguous bases of match to the human genome reference sequence. Assessing possible human origins for these AVT-rich inserts is complicated by these inserts' genetic plasticity, the extent of microsatellite variation within the human population, and the relatively high frequency with which AVT-rich extensions are observed: leaving open the possibility that some arose via recombination between viral isolates (Kitrinos et al., 2003; Zhivotovsky et al., 2003). Nonetheless, observations such as RT β3- β4 inserts that lack genetically linked regions of sequence identity (Masquelier et al., 2001) and the *de novo* generation of AVT-rich extensions during virus replication in tissue culture (Kuhmann et al., 2004) suggest that on rare occasions additional instances of short patch human sequence transduction, like that reported here for 99JP-NH3-II, may contribute to HIV-1 genetic variation.

Materials and methods

Sequences and sequence analysis

Analyses shown were performed using default settings in nucleotide-nucleotide BLAST (blastn) via NCBI (http://www.ncbi.nlm.nih.gov). All sequences analyzed in this report were previously deposited in GenBank. Searches described as unrestricted databases in GenBank or "nr" were searches of (all GenBank + RefSeq Nucleotides + EMBL + DDBJ + PDB sequences). Expect values presented in Fig. 1B are those assigned by NCBI blastn using default gap and mismatch penalties.

Additional examples of β3- β4 insert splinted recombinants (GenBank accession nos. AY877315 and AF315271) mentioned in the discussion were identified using blastn by querying Genbank with an artificial sequence comprised of NH3's preinsertion sequence (Fig. 1A) modified to contain a β3- β4 insert of arbitrary length (24 b) while applying reduced mismatch and increased gap penalty values and then by manually screening isolates for strong epidemiologic support and structures predicted by splinted recombination.

Methods and findings from querying all HIV sequences in GenBank with the human genome will be described elsewhere.

Acknowledgments

The authors thank Yuki Naito and James Gergel for their help with the sequence analyses, and David Friedman and Mary Jane Wieland for their useful discussions. This work was supported by NIH no. GM64479 to AT.

References

An, W., Telesnitsky, A., 2002. HIV-1 genetic recombination: experimental approaches and observations. AIDS Rev. 4, 195–212.

An, W., Telesnitsky, A., 2004. Human immunodeficiency virus type 1 transductive recombination can occur frequently and in proportion to polyadenylation signal readthrough. J. Virol. 78, 3419–3428.

Berkhout, B., Grigoriev, A., Bakker, M., Lukashov, V.V., 2002. Codon and amino acid usage in retroviral genomes is consistent with virus-specific nucleotide pressure. AIDS Res. Hum. Retroviruses 18, 133–141.

Bosch, M.L., Andeweg, A.C., Schipper, R., Kenter, M., 1994. Insertion of N-linked glycosylation sites in the variable regions of the human immunodeficiency virus type 1 surface glycoprotein through AAT triplet reiteration. J. Virol. 68, 7566–7569.

Dunn, M.M., Olsen, J.C., Swanstrom, R., 1992. Characterization of unintegrated retroviral DNA with long terminal repeat-associated cell-derived inserts. J. Virol. 66, 5735–5743.

Fang, H., Pincus, S.H., 1995. Unique insertion sequence and pattern of CD4 expression in variants selected with immunotoxins from human immunodeficiency virus type 1-infected T cells. J. Virol. 69, 75–81.

Gao, X.-S., Cheng, J., Zhen, Z., Guo, J., Zhang, L.-Y., Tao, M.-L., 2005. Cloning of hepatitis B virus DNAPTP1 transactivating genes by suppression subtractive hybridization technique. Shijie Huaren Xiaohua Zazhi 13, 2371–2374.

Hacker, J., Kaper, J.B., 2000. Pathogenicity islands and the evolution of microbes. Annu. Rev. Microbiol. 54, 641–679.

Hajjar, A.M., Linial, M.L., 1993. A model system for nonhomologous recombination between retroviral and cellular RNA. J. Virol. 67, 3845–3853.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., Malim, M.H., 2003. DNA deamination mediates Innate immunity to retroviral infection. Cell 113, 803–809.

Jetzt, A.E., Yu, H., Klarmann, G.J., Ron, Y., Preston, B.D., Dougherty, J.P., 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome. J. Virol. 74, 1234–1240.

Kitrinos, K.M., Hoffman, N.G., Nelson, J.A., Swanstrom, R., 2003. Turnover of env variable region 1 and 2 genotypes in subjects with late-stage human immunodeficiency virus type 1 infection. J. Virol. 77, 6811–6822.

Kuhmann, S.E., Pugach, P., Kunstman, K.J., Taylor, J., Stanfield, R.L., Snyder, A., Strizki, J.M., Riley, J., Baroudy, B.M., Wilson, I.A., Korber, B.T., Wolinsky, S.M., Moore, J.P., 2004. Genetic and phenotypic analyses of human immunodeficiency virus type 1 escape from a small-molecule CCR5 inhibitor. J. Virol. 78, 2790–2807.

Leitner, T., Albert, J., 1999. The molecular clock of HIV-1 unveiled through analysis of a known transmission history. Proc. Natl. Acad. Sci. U.S.A. 96, 10752–10757.

Lobato, R.L., Kim, E.Y., Kagan, R.M., Merigan, T.C., 2002. Genotypic and phenotypic analysis of a novel 15-base insertion occurring between codons 69 and 70 of HIV type 1 reverse transcriptase. AIDS Res. Hum. Retroviruses 18, 733–736.

Malim, M.H., Emerman, M., 2001. HIV-1 sequence variation: drift, shift, and attenuation. Cell 104 (4), 469–472.

Masquelier, B., Race, E., Tamalet, C., Descamps, D., Izopet, J., Buffet-Janvresse, C., Ruffault, A., Mohammed, A.S., Cottalorda, J., Schmuck, A., Calvez, V., Dam, E., Fleury, H., Brun-Vezinet, F., ANRS AC11 Resistance study group, French Agence Nationale de Recherches sur le SIDA, 2001. Genotypic and phenotypic resistance patterns of human immunodeficiency virus type 1 variants with insertions or deletions in the reverse transcriptase (RT): multicenter study of patients treated with RT inhibitors. Antimicrob. Agents Chemother. 45, 1836–1842.

Mikkelsen, J.G., Pedersen, F.S., 2000. Genetic reassortment and patch repair by recombination in retroviruses. J. Biomed. Sci. 7, 77–99.

Muriaux, D., Rein, A., 2003. Encapsidation and transduction of cellular genes by retroviruses. Front. Biosci. 8, 135–142.

Olsen, J.C., Bova-Hill, C., Grandgenett, D.P., Quinn, T.P., Manfredi, J.P., Swanstrom, R., 1990. Rearrangements in unintegrated retroviral DNA are complex and are the result of multiple genetic determinants. J. Virol. 64 (11), 5475–5484.

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., 2004. Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat. Genet. 36, 40–45.

Parthasarathi, S., Varela-Echavarria, A., Ron, Y., Preston, B.D., Dougherty, J.P., 1995. Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications. J. Virol. 69, 7991–8000.

Pathak, V.K., Temin, H.M., 1990. Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. Proc. Natl. Acad. Sci. U.S.A. 87, 6024–6028.

Preston, B.D., Dougherty, J.P., 1996. Mechanisms of retroviral mutation. Trends Microbiol. 4 (1), 16–21.

Pulsinelli, G.A., Temin, H.M., 1991. Characterization of large deletions occurring during a single round of retrovirus vector replication: novel deletion mechanism involving errors in strand transfer. J. Virol. 65, 4786–4797.

Sato, H., Tomita, Y., Ebisawa, K., Hachiya, A., Shibamura, K., Shiino, T., Yang, R., Tatsumi, M., Gushi, K., Umeyama, H., Oka, S., Takebe, Y., Nagai, Y., 2001. Augmentation of human immunodeficiency virus type 1 subtype E (CRF01_AE) multiple-drug resistance by insertion of a foreign 11-amino-acid fragment into the reverse transcriptase. J. Virol. 75 (12), 5604–5613.

Sun, G., O'Neil, P.K., Yu, H., Ron, Y., Preston, B.D., Dougherty, J.P., 2001. Transduction of cellular sequence by a human immunodeficiency virus type 1-derived vector. J. Virol. 75, 11902–11906.

Temin, H.M., 1993. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proc. Natl. Acad. Sci. U.S.A. 90 (15), 6900–6903.

van der Hoek, L., Back, N., Jebbink, M.F., de Ronde, A., Bakker, M., Jurriaans, S., Reiss, P., Parkin, N., Berkhout, B., 2005. Increased mulitnucleoside drug resistance and decreased replicative capacity of a human immunodeficiency virus type 1 variant with an 8-amino-acid insert in the reverse transcriptase. J. Virol. 79, 3536–3543.

Wain-Hobson, S., Renoux-Elbe, C., Vartanian, J.P., Meyerhans, A., 2003. Network analysis of human and simian immunodeficiency virus sequence sets reveals massive recombination resulting in shorter pathways. J. Gen. Virol. 84, 885–895.

Winters, M.A., Merigan, T.C., 2005. Insertions in the human immunodeficiency virus type 1 protease and reverse transcriptase genes: Clinical impact and molecular mechanisms. Antimicrob. Agents Chemother. 49, 2575–2582.

Zhang, J., Temin, H.M., 1993. 3′ junctions of oncogene-virus sequences and the mechanisms for formation of highly oncogenic retroviruses. J. Virol. 67 (4), 1747–1751.

Zhivotovsky, L.S., Rosenberg, N.A., Feldman, M.W., 2003. Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. Am. J. Hum. Genet. 72, 1171–1186.

Identification of a Novel Circulating Recombinant Form (CRF33_01B) Disseminating Widely Among Various Risk Populations in Kuala Lumpur, Malaysia

Kok Keng Tee, MMedSc,† Xiao-Jie Li, MD, PhD,* Kyoko Nohtomi, BSc,* Kee Peng Ng, MBBS, PhD,‡ Adeeba Kamarulzaman, MBBS, FRACP,† and Yutaka Takebe, MD, PhD**

Summary: A molecular epidemiological investigation was conducted among various risk populations (n = 184) in Kuala Lumpur, Malaysia, in 2003 to 2005, on the basis of nucleotide sequences of protease and reverse transcriptase regions. In addition to circulating HIV-1 strains, including CRF01_AE (57.1%), subtype B (20.1%), and subtype C (0.5%), we detected a candidate with a new circulating recombinant form (CRF). We determined four near-full-length nucleotide sequences with identical subtype structure from epidemiologically unlinked individuals of different risk and ethnic groups. In this chimera, two short subtype B segments were inserted into the *gag*-RT region in a backbone of CRF01_AE. The recombinant structure was distinct from previously identified CRF15_01B in Thailand. In agreement with the current HIV nomenclature system, this constitutes a novel CRF (CRF33_01B). The overall prevalence of CRF33_01B is 19.0% (35/184). Although the prevalence of CRF33_01B is particularly high among injecting drug users (42.0%, 21/50), it is also detected in a substantial proportion of homo-/bisexual males (18.8%, 3/16) and heterosexuals (9.8%, 9/92). Moreover, unique recombinant forms composed of CRF01_AE and subtype B that have a significant structural relationship with CRF33_01B were detected in 1.6% (3/184) of study subjects, suggesting an ongoing recombination process in Malaysia. This new CRF seems to be bridging viral transmission between different risk populations in this country.

Key Words: HIV-1 circulating recombinant form (CRF), unique recombinant form (URF), CRF01_AE, subtype B, CRF33_01B, Malaysia, Asia

(*J Acquir Immune Defic Syndr* 2006;43:523–529)

Received for publication May 10, 2006; accepted August 7, 2006.

From the *Laboratory of Molecular Virology and Epidemiology, AIDS Research Center, National Institute of Infectious Diseases, Shinjuku-ku, Tokyo, Japan; †Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia; and ‡Department of Medical Microbiology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia.

Funded by grants from the Ministry of Health, Labour and Welfare, the Ministry of Education, Science and Technology, and the Japanese Foundation for AIDS Prevention (JFAP).

Reprints: Yutaka Takebe, MD, PhD, Laboratory of Molecular Virology and Epidemiology, AIDS Research Center, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan (e-mail: takebe@nih.go.jp).

Since the first case of AIDS was reported in Malaysia in 1986,[1] a total of 67,438 HIV infections had been identified nationwide by the Ministry of Health by June 2005. Although the adult HIV prevalence rate remains low (0.5%:0.2%–1.5%), elements exist that could cause the epidemic to erupt suddenly. In 2004 alone, an average of 18 new cases of HIV infections was reported daily. The Joint United Nations Program on HIV/AIDS (UNAIDS) estimated that approximately 69,000 people in Malaysia were living with HIV in 2005.[2] The main route of HIV transmission in Malaysia is through injecting drug use, which accounts for 75% of all reported HIV cases (Fig. 1, inset). The growing proportion of HIV cases attributed to sexual transmission (17% in 2002 compared with 11.5% in 2001 and 7% in 1995; 46% increase between 2001 and 2002), suggesting that HIV is spreading in the general population.[3,4]

HIV-1 exhibits a tremendous genetic diversity that is driven by high rates of mutation (3×30^{-5} sites/genome/replication cycle) and recombination (2–3 crossovers/replication cycle), coupled with high viral turnovers ($>10^9$ per day) and persistent nature of infections (300 replication cycles/year).[5] By these mechanisms, HIV-1 group M, which is largely responsible for the global pandemic, has diversified into 11 subtypes and sub-subtypes (A1, A2, B, C, D, F1, F2, G, H, J, and K) and various types of recombinants (http://hiv-web.lanl.gov/). HIV-1 recombinants with epidemic spread are known as circulating recombinant forms (CRFs). Thirty-two CRFs are currently recognized.[6] Four CRFs have been reported so far in Asia: CRF01_AE and CRF15_01B in Thailand and CRF07_BC and CRF08_BC in China. To define a CRF, at least three epidemiologically unlinked HIV-1 sequences with identical mosaic structures should be characterized, at least two of them in near-full-length genomes (>8 kb).[7] In addition to CRFs, various types of unique recombinant forms (URFs) that were detected in a single individual or a single epidemiologically linked cluster have been identified in the region, where the multiple lineages of HIV-1 strains cocirculate in the same population.

In the early phase of the Thai epidemic, two HIV-1 strains—CRF01_AE and subtype B′ (Thai variant of subtype B)—were circulating relatively independently among different risk populations. CRF01_AE was distributed among persons at risk of sexual exposure, while subtype B′ was distributed mainly among injecting drug users (IDUs).[8,9] However, by 1999 it was reported that CRF01_AE accounted for >50% of new infections among IDUs.[10–12] Cocirculation of CRF01_AE

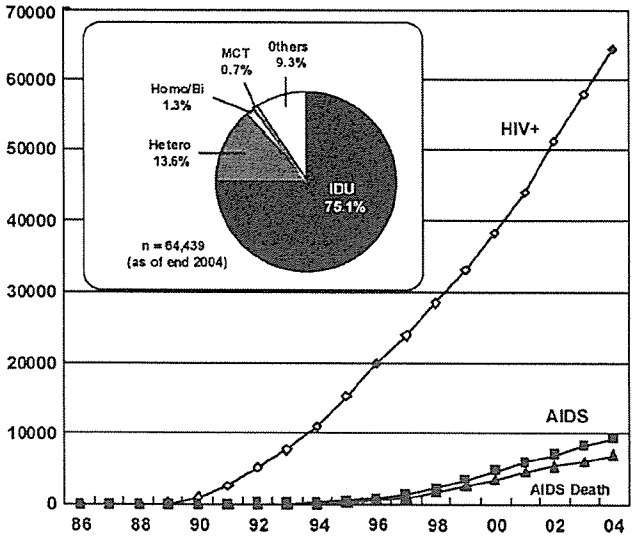


FIGURE 1. Status and distribution of HIV-1 infection in different risk categories in Malaysia. Cumulative numbers of HIV-1 infections, AIDS cases, and AIDS-related deaths reported by the Ministry of Health in Malaysia (1986–December 2004). Inset, Distribution of HIV-1 infections in different risk categories as of the end of 2004 (n = 64,439). Abbreviations: IDU, injecting drug users; Hetero, heterosexuals; Homo/Bi, male homo-/bisexual; MCT, mother-to-child transmission.

and subtype B′ in Thailand led to the generation of various forms of CRF01_AE/B′ recombinants,[13–18] including CRF15_01B.[19] A similar molecular epidemiological trend has been observed in Malaysia. Studies conducted in 1992 to 1997 showed that CRF01_AE and subtype B′ were prevalent among 81% of heterosexuals and 55% to 92% of IDUs, respectively.[20–22] However, more recent studies based on partial protease (Pro) and reverse transcriptase (RT) sequence data suggested the emergence of new forms of CRF01_AE/B recombinants in Malaysia.[23,24]

For the present study, we characterized near-full-length nucleotide sequences of CRF01_AE/B recombinants from eight epidemiologically unlinked individuals to define a novel CRF and other recombinant forms emerging in Malaysia and discuss herein their epidemiological and biological implications.

MATERIALS AND METHODS

Study Subjects and Specimens

All subjects (n = 184) were recruited between July 2003 and August 2005 in the HIV clinic of the University Malaya Medical Center (UMMC), Kuala Lumpur. This study was approved by the UMMC Medical Ethics Committee. Information on the patients' clinical and epidemiological backgrounds was collected from the HIV patient-management database. Study subjects included 50 IDUs, 92 heterosexuals, 14 male homosexuals, 2 male bisexuals, 4 persons who were infected from their mothers (mother-to-child transmission, MCT), and 22 cases with unknown risk factors. Subjects consisted of 150 adult males with an age range of 19 to 67 years old (mean: 38.2 ± 8.9 years old) and 31 adult females with an age range of 26 to 53 years old (mean: 36.0 ± 7.9 years old). Three MCT adolescents (2 males, 1 female) ranging in age from 5 to 9 years old (mean: 7.7 ± 2.3 years old) were also included. Specimens from these 184 patients were serologically determined to be HIV-1 positive. No HIV-2 infections were detected. HIV-1 genotypes were screened on the basis of the Pro and RT regions using plasma HIV-1 RNA, as described previously.[23,24]

Viral Isolation, Near-Full-Length DNA Amplification, and Sequencing

Peripheral-blood mononuclear cells (PBMCs) from selected patients were separated on Ficoll-Hypaque density gradient centrifugation (Amersham Biosciences AB, Uppsala, Sweden). For virus isolation, PBMCs from HIV-1–positive individuals were cocultured with phytohemagglutinin-stimulated (1 μg/ml) $CD8^+$ T-cell–depleted PBMCs (Miltenyi Biotec GmbH, Bergisch Gladbach, Germany) from HIV-negative healthy donors in RPMI 1640 containing 10% fetal calf serum and interleukin-2 (20 U/mL). Virus production was detected by a virion-associated RT assay as described previously.[25] HIV-infected PBMCs were harvested, and proviral DNA was isolated with guanidine detergent (Invitrogen, Carlsbad, CA). Near-full-length viral DNA was amplified with primers pbsA (5′-AGT GGC GCC CGA ACA GG-3′; nucleotide positions relative to HXB2: 634–650) and 9KU5B (5′-GGT CTG AGG GAT CTC TAG TTA CCA G-3′; nucleotide positions relative to HXB2: 9666–9690) by the Expand Long Template PCR System (Roche Diagnostic GmbH, Penzberg, Germany), gel-purified, and TA-cloned with the pCR-XL-TOPO vector (Invitrogen).[26,27] Positive clones were selected, and the near-full-length genome of ~9.1 kb was directly sequenced with an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA) using the primer-walking method. For plasma samples, HIV-1 RNA was reverse-transcribed and nested-PCR–amplified to produce 10 overlapping fragments; directly sequenced; and then assembled to generate the near-full-length genomes. Primers were carefully designed to obtain sufficient numbers of overlapping nucleotides to ensure that recombinants were not generated by assembling the sequence fragments derived from different HIV-1 subtypes that had been amplified from an individual (primer sets and PCR parameters are available upon request).

Phylogenetic Analysis

Nucleotide sequences were aligned manually using Se-Al, version 1.0,[28] with HIV-1 reference subtypes and CRFs from the HIV database (http://hiv-web.lanl.gov/). Phylogenetic trees were constructed by the neighbor-joining method[29] based on the Kimura two-parameter model with a transition-to-transversion ratio of 2.0.[30] The reliability of the branching orders was tested by bootstrap analysis of 100 replicates. Bootscanning and informative-site analyses were performed using SimPlot, version 3.5,[31] with a sliding window of 250 nucleotides overlapped by 50 nucleotides to define the recombinant structure. The origin of each segment was analyzed by subregion neighbor-joining tree analysis.

RESULTS

HIV-1 Genotype Distribution in Malaysia

HIV-1 genotypes circulating among various risk populations (n = 184) were determined by phylogenetic and recombination breakpoint analyses based on the nucleotide sequences of the Pro and RT regions. Distribution of HIV-1 genotypes in Malaysia is currently as follows: CRF01_AE (105/184, 57.1%), subtype B′ (37/184, 20.1%), and subtype C (1/184, 0.5%) (Fig. 2). A substantial proportion (41/184, 22.3%) of specimens showed unique subtype structure composed of CRF01_AE and subtype B.[23,24] The majority of them (35/184, 19.0%) harbored the identical subtype structure, and a small proportion of specimens (6/184, 3.3%) displayed different profiles of recombinant structure. We designate these two groups of Malaysian HIV-1 recombinant strains as the major recombinant form (major RF) and the minor RF, respectively. The prevalence of the major RF was highest among IDUs (42.0%, 21/50), but it was also found among other risk populations: male homo-/bisexuals (18.8%, 3/16) and heterosexuals (9.8%, 9/92; Fig. 2).

Near-Full-Length Nucleotide Sequence Analysis Detects New CRF in Malaysia

To characterize the detailed subtype structure of new HIV-1 recombinant strains in Malaysia, we determined near-full-length nucleotide sequences of HIV-1 strains from eight epidemiologically unlinked individuals who are infected with

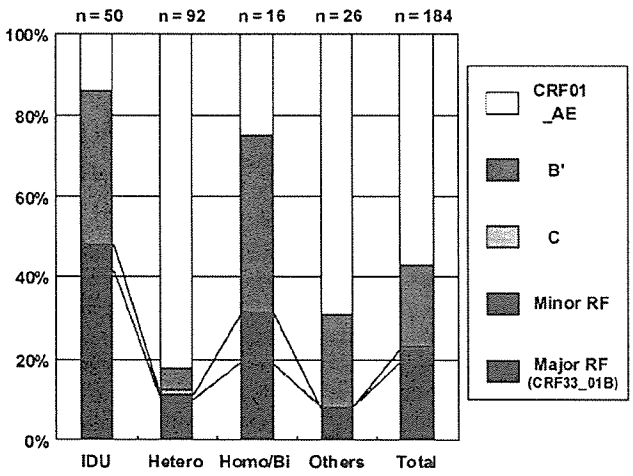


FIGURE 2. Distribution of HIV-1 genotypes in different risk groups in Malaysia. Genotypes were determined by phylogenetic and recombination breakpoint analyses of the nucleotide sequences in protease and reverse transcriptase regions (see text). Bars indicate the relative frequency (%) of HIV-1 genotypes in each risk group shown at the bottom. "n" indicates the number of specimens analyzed in each risk group. "Others" include 22 cases with unknown risk factors and 4 cases of MCT. Abbreviations: B′, HIV-1 subtype B′ (Thai variant of subtype B); C, subtype C; major RF, major recombinant forms (CRF33_01B); minor RF, minor recombinant forms (unique recombinant forms composed of CRF01_AE and subtype B).

either major RFs (n = 4) (05MYKL007.1, 05MYKL015.2, 05MYKL031.1, and 05MYKL045.1) or minor RFs (n = 4) (04MYKL016.1, 03MYKL018.1, 04MYKL019.1, and 05MYKL043.1) (Table 1). The bootscanning plot of near-full-length nucleotide sequences revealed that four major RFs indeed displayed the identical recombinant structure composed of CRF01_AE and subtype B (Fig. 3A). In this chimera, two short subtype B segments are inserted into a backbone of CRF01_AE. The profile of the subtype structure of Malaysian major RFs is distinct from that of CRF15_01B previously reported in Thailand,[19] where most of the *env* region belongs to subtype B in a backbone of CRF01_AE (Figs. 3B, C). In major RFs, a total of four recombination breakpoints (sites I–IV in Fig. 3C) are found to be clustered in a 0.9-kb region in *gag*-RT gene. The subregion tree analysis demonstrated that CRF01_AE segments (regions 1 and 5 in Fig. 3B) in major RFs belonged to the cluster of Thailand CRF01_AE, indicating that these segments indeed originated from CRF01_AE of Thai origin. However, region 3 (121 bp) was too short to discern its origin from either African or Thailand CRF01_AEs. Similarly, we were not able to discern whether subtype B segments (regions 2 and 4) originated from subtype B′ because subtypes B and B′ are phylogenetically indistinguishable in the Pro-RT region (Fig. 3B).

On the other hand, minor RFs (n = 4) showed a different degree of structural relatedness with major RFs. In particular, site II in Pro region is shared between major RFs and three minor RFs (1.6%, 3/184) (04MYKL016.1, 04MYKL019.1, and 05MYKL043.1; Fig. 3C). Similarly, the recombination breakpoints in *gag* p7 (site I) and RT (site IV) regions in major RF are shared with 04MYKL016.1 and 05MYKL043.1, respectively. 03MYKL018.1 is the most distantly related recombinant that shows no obvious structural relatedness with other RFs, while 05MYKL043.1 is most closely related with major RFs, sharing two recombination breakpoints (sites II and IV) but harboring slightly longer subtype B segments in *gag*-Pro and Pro-RT regions than major RFs (Fig. 3C).

Consistent with the results of recombination breakpoint analyses, the phylogenetic tree analysis of near-full-length nucleotide sequences showed that four major RFs (05MYKL007.1, 05MYKL015.2, 05MYKL031.1, and 05MYKL045.1) and the most closely related minor RF (05MYKL043.1) formed a monophyletic cluster with high bootstrap values (93%) (Fig. 3D). In contrast, three other minor RFs (04MYKL016.1, 03MYKL018.1, and 04MYKL019.1) are located outside any of the CRF01_AE-related sequences, including CRF15_01B (Fig. 3D).

Characterization of the Recombination Breakpoints in Malaysian RFs

To search for the possible sequence signatures near the subtype boundaries in Malaysian RFs, we mapped out all four recombination breakpoints (sites I–IV in Fig. 3C) estimated by informative site analysis (Fig. 4). It is noted that the homopolymeric nucleotide tracts (polyadenine ($A_{\geq 4}$), polyguanine ($G_{\geq 3}$), polycytosine ($C_{\geq 4}$), or polythymine ($T_{\geq 4}$)), known to pause in vitro reverse transcription and promote template switch,[32,33] are observed within or adjacent to all four recombination breakpoints. In particular, a characteristic

TABLE 1. Epidemiological and Clinical Background of Study Subjects Harboring Novel CRF01_AE/B Recombinants Emerging in Malaysia

Sample ID	Sex	Age (y)	Ethnicity	Risk Factor*	Date of First Positive HIV-1 Test (mo/yr)	Date of Collection (mo/yr)	Treatment Status†
05MYKL007.1	Male	32	Chinese	Homo	11/04	6/05	ARV-naive
05MYKL015.2	Male	33	Malay	Homo	4/04	8/05	ARV-naive
05MYKL031.1	Male	44	Indian	Hetero	1/05	8/05	3TC d4T NVP
05MYKL045.1	Male	43	Malay	Hetero	7/05	8/05	AZT 3TC EFV
04MYKL016.1	Female	30	Malay	Hetero	9/02	4/04	ARV-naive
03MYKL018.1	Male	32	Chinese	Bisexual	1/97	12/03	ARV-naive
04MYKL019.1	Male	49	Chinese	Hetero + IDU	2/04	6/04	ARV-naive
05MYKL043.1	Male	32	Malay	Hetero + IDU	8/03	8/05	AZT 3TC EFV

Sample ID	CD4 Count (cells/mm^3)	Viral Load (copies/mL)	Specimen Type	HIV-1 Genotype‡	Accession No.
05MYKL007.1	389	14,300	Isolate	CRF33_01B	DQ366659
05MYKL015.2	310	9,900,000	Isolate	CRF33_01B	DQ366660
05MYKL031.1	3	188,000	Isolate	CRF33_01B	DQ366661
05MYKL045.1	121	4300	Isolate	CRF33_01B	DQ366662
04MYKL016.1	401	3290	Plasma	Minor RF	DQ366663
03MYKL018.1	74	328,000	Plasma	Minor RF	DQ366664
04MYKL019.1	78	67,100	Plasma	Minor RF	DQ366665
05MYKL043.1	375	<50	Isolate	Minor RF	DQ366666

*Homo indicates male homosexual; Hetero, heterosexual; IDU, injecting drug user.
†ARV indicates antiretroviral; 3TC, lamivudine; d4T, didanosine; NVP, nevirapine; AZT, zidovudine; EFV, efavirenz.
‡Minor RF refers to the unique recombinant form composed of CRF01_AE and subtype B that showed a different degree of structural relatedness with CRF33_01B (Fig. 3).

homopolymeric nucleotide tract, the A_6GA_6 sequence, was found within site IV (in the RT gene). Similarly, noncontiguous sets of homopolymeric tracts, $A_5/G_5/T_4$ sequence, were observed within site II (in Pro gene). It is also noted that sites I and III are located adjacent to the boundaries between *gag* p7 and p1 and between Pro and RT, respectively (Fig. 4).

DISCUSSIONS

In the present study, we identified a novel CRF (CRF33_01B) composed of CRF01_AE and subtype B in Malaysia. Four sets of near-full-length nucleotide sequences with identical recombinant structure were determined from epidemiologically unlinked individuals with different risk factors and ethnicities in Malaysia (Table 1). This new chimera displays the recombinant structure distinct from any other CRFs reported to date, including CRF15_01B identified in Thailand.[19] The prevalence of CRF33_01B was particularly high among IDUs (42.0%, 21/50), compared with other risk populations (18.8% (3/16) in male homo-/bisexuals; 9.8% (9/92) in heterosexuals; Fig. 2). This may suggest that CRF33_01B first emerged among IDUs and subsequently spread to other risk populations.[23,24]

In addition to this new CRF33_01B, we detected six minor RFs consisting of CRF01_AE and subtype B. Among them, we determined the near-full-length nucleotide sequences of four minor RFs (Table 1). Most of them (except 03MYKL018.1) seem to share the recombination breakpoints with those of CRF33_01B (Fig. 3C). This suggests that they are closely related relatives and are most likely to be secondary recombinants derived from CRF33_01B that were generated by subsequent crossover(s) with either CRF01_AE or subtype B strain circulating in Malaysia.

The sequence characteristics found in or near the recombination breakpoints in CRF33_01B are suggestive of the following traits: (i) The homopolymeric tracts, known to pause in vitro reverse transcription and promote template switch (retroviral recombination),[32,33] are detected within or adjacent to all four recombination breakpoints in CRF33_01B (sites I–IV in Fig. 3C). Particularly, site II recombination breakpoint in Pro gene that contains a characteristic set of noncontiguous homopolymeric tracts, $A_5/G_5/T_4$ sequence (Fig. 4), was shared with three out of four minor RFs (05MYKL043.1, 04MYKL016.1, and 04MYKL019.1; Fig. 3C). Interestingly, this recombination breakpoint (site II) is shared with some of CRF01_AE/B recombinants reported in Thailand (unpublished data). It is thus tempting to speculate that site II may be one of the preferred sites for in vivo recombination. (ii) Some recombination breakpoints are found near the boundaries of functional domains of HIV-1 proteins: sites I and III are located adjacent to the boundaries between *gag* p7 and p1 and between Pro and RT, respectively (Fig. 4). This may reflect the selection pressure to maintain the functional integrity of HIV-1 proteins in recombination processes. Extensive dissemination of this new CRF over CRF01_AE and subtype B particularly among IDUs in Malaysia may suggest the presence of yet undefined selective advantages over parental strains.

Finally, we observed a unique parallelism of the transition in molecular epidemiological features of HIV-1

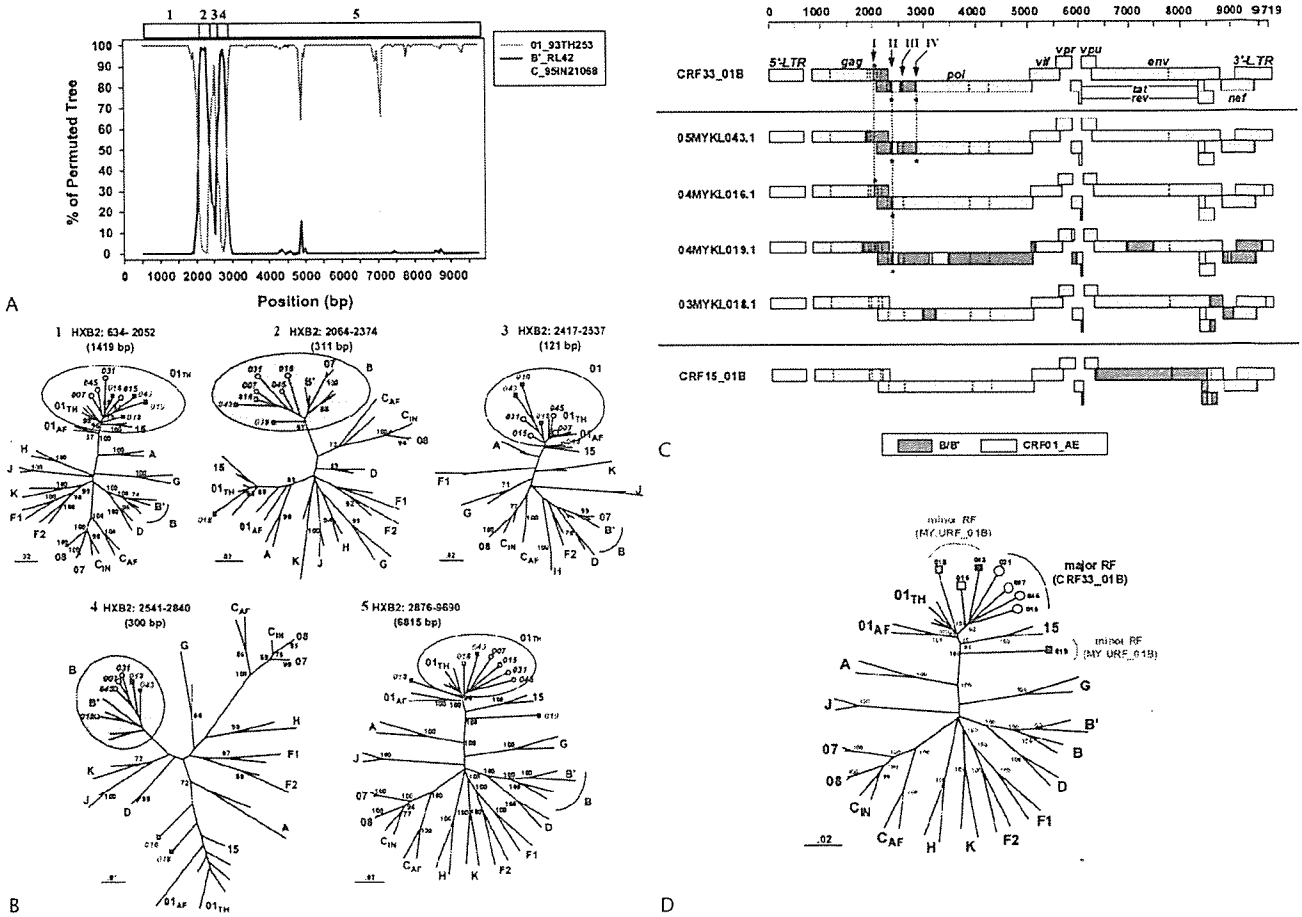


FIGURE 3. Characterization of newly emerging HIV-1 recombinant forms (RFs) in Malaysia. A, Bootscanning analysis of major RF (CRF33_01B). CRF01.93TH253 and B′. RL42 are used as the putative parental subtypes. Bootstrap values of 100 replicates were plotted for a window of 250 bp moving in increments of 50 bp along near-full-length sequences. B, Subregion trees in the segments (1–5) defined by informative site analysis of CRF33_01B are illustrated. Nucleotide positions of each segment relative to HXB2 and the size of segments (in parentheses) are indicated. Bootstrap values (≥70%) are shown. Major RFs (CRF33_01B) (n = 4: 05MYKL007.1, *007*; 05MYKL015.2, *015*; 05MYKL031.1, *031*; 05MYKL045.1, *045*). Minor RFs (n = 4: 03MYKL018.1 *018*; 04MYKL016.1 *016*; 04MYKL019.1 *019*; 05MYKL043.1, *043*). Abbreviations of strain codes for minor RFs are shown in gray. C, Schematic representation of subtype structures of novel HIV-1 recombinants. Recombination breakpoints (sites I–IV) are indicated. Dotted vertical lines indicate the breakpoints (marked with asterisks) shared among Malaysian RFs. D, Neighbor-joining tree of eight near-full-length nucleotide sequences of novel Malaysian recombinants with reference sequences of HIV-1 group M subtypes and sub-subtypes (http://hiv-web.lanl.gov/). Selected HIV-1 genotypes of geographical importance are included: B′, Thai variant of subtype B; C_{IN}, Indian subtype C; C_{AF}, African subtype C; 01_{TH}, Thailand CRF01_AE; 01_{AF}, African CRF01_AE; 07, CRF07_BC; 08, CRF08_BC; 15, CRF15_01B, minor RF, minor recombinant form composed of CRF01_AE and subtype B. Branching orders were assessed by 100 bootstrap analyses, and values ≥70% are shown at the corresponding nodes of the tree. Scale bar represents 2% genetic distance (0.02 substitution/site).

epidemics between Thailand and Malaysia. In the early phase of the Malaysian epidemic, CRF01_AE and subtype B′ were circulating relatively independently among heterosexuals and IDUs, respectively, similar to the early stage of the Thai epidemic.[8,9] However, Tovanabutra and others recently reported that in a high-risk cohort in northern Thailand, CRF01_AE/B recombinants began to be identified in 2001 at 8.3% of incident cases and increased to 57.1% in 2002.[18] Due to the lack of availability of archival HIV specimens in the current study, we were not able to pinpoint the timing of emergence of CRF33_01B. However, the relatively long branch length of the CRF33_01B cluster (interstrain genetic diversity of 6.06 ± 0.50% (n = 4, isolates collected in 2005) compared with the interstrain genetic diversity of 3.10 ± 0.55% for Thailand CRF01_AE cluster (n = 4, isolates collected in 1990–1997) in near-full-length genomes) (Fig. 3D) suggests that the timing of the origin of CRF33_01B could be dated back to mid-1990s (unpublished data).

In summary, we report herein on a novel CRF that is circulating widely among various risk groups in Malaysia.