

The following donor and target DNA oligonucleotides were designed and used:

Donor-1 (D1): 5'-ACTGCTAGAGATTTTCCA-CACTGACTAAAAG-3'

Donor-2 (D2): Biotin-5'-CTTTTAGTCAGTGTGGA-AAATCTCTAGCA-3'

Target-1 (T1): 5'-CTAGAGATTTTCCCACTGACT-AAAAG-3'-Digoxigenin (DIG),

Target-2 (T2): 5'-CTTTTAGTCAGTGTGAAAA-TCTCTAG-3'-DIG

To form dsDNA, the D1-D2 pair and the T1-T2 pair were mixed in the presence of 0.1 M NaCl and denatured for 10 min at 95°C, followed by an annealing process, gradual cooling down to room temperature. One pmol biotinylated donor dsDNA (D1-D2), 15 pmol IN protein and 5 µl test compounds (100 µM in DMSO) were mixed together in assay buffer (25 mM 3-(N-morpholino)propanesulfonic acid, pH 7.2, 25 mM NaCl, 10 mM MgCl₂, 10 mM DTT, 5% PEG, 10% DMSO), followed by the addition of 0.75 pmol target dsDNA (T1-T2), and adjusted to a final volume of 100 µl and incubated for 1 h at 37°C. After the incubation, the mixture was adjusted to a final volume of 200 µl with ELISA buffer (20 mM Tris [pH 8.0], 0.4 M NaCl, 10 mM EDTA, 0.1 mg/ml sonicated DNA). To harvest the strand-transfer product, the mixture was transferred into a 96-well micro titre plate coated with streptavidin (PIERCE, Rockford, IL, USA), followed by adding an alkaline phosphatase conjugated anti-DIG antibody (Roche Diagnostics, Mannheim, Germany) and a disodium 3-(4-methoxyspiro[1,2-dioxetane-3,2'-(5'-chloro)tricyclo[3.3.1.1^{3,7}]decan}-4-yl) phenyl phosphate (CSPD) substrate (Roche). The lumino-intensity was quantified with a Luminous CT-9000D luminometer (DIA-IATRON, Tokyo, Japan).

In addition to the above two different strand-transfer assays, a strand-transfer assay with radioisotope labelled target DNA and SDS-PAGE was employed in order to visually confirm the strand-transfer inhibition (Craigie *et al.*, 1995). By use of T4 polynucleotide kinase (TAKARA BIO, Osaka, Japan), the 5' end of 20 mer target oligonucleotide-A (5'-TGTGGAAAATCTCTAGCAGT-3') was labelled with [γ -³²P] ATP (370 MBq/µl, Amersham Bioscience, Tokyo, Japan). After the labelling reaction was terminated by adding EDTA, complementary oligonucleotide-B (5'-ACTGCTAGAGATTTTCCACA-3') was added, and dsDNA was formed by heat denaturation and gradual cooling to room temperature. Unincorporated [γ -³²P]ATP was removed by G-25 Column (Amersham Bioscience, Piscataway, NJ). The reaction products were applied to 20% denatured polyacrylamide gel electrophoresis (300V/25A). The result of the electrophoresis was analysed by BAS-2500 (Fuji film, Tokyo, Japan).

Inhibition kinetics of IN

To analyse the strand-transfer inhibition mechanism of the test compounds, whether the action is competitive inhibition or non-competitive inhibition, Michaelis-Menten constant (K_m) and maximum velocity (V_{max}) were evaluated. Strand-transfer inhibition was evaluated on eight different time points (0, 1, 3, 5, 7.5, 10, 15, and 20 min) with four different compound concentrations (0, 1, 5, 10 µM) and target DNA concentrations (0.167, 0.25, 0.5, and 1 pmol). The initial reaction rate constants of IN were determined by linear regression using linear data points of product concentration-time plots. K_m and V_{max} were calculated from the Y-axis intercept in a plot of the slopes of Lineweaver-Burk analysis.

Intercalative activity evaluation

To clarify the possibility of intercalative activity of test compounds, ethidium bromide (EtBr) displacement assay was carried out following the protocol reported previously (Cain *et al.*, 1978). In brief, 1 µM calf thymus DNA (Invitrogen, Carlsbad, CA, USA) was mixed with EtBr (final concentration at 1.26 µM) and reaction buffer (2 mM HEPES, 10 µM EDTA, 9.4 mM NaCl, pH 7.0), and incubated for 10 min at room temperature. After the incubation, test compounds were added into the calf thymus DNA-EtBr mixture at different concentrations (final concentrations of 0.01-1000 µM). Fluorescence intensity of each mixture was determined by Fluoroskan Ascent FL (Helsinki, Finland. Excited at 544 nm, emitted at 590 nm). Actinomycin D (ICN Biomedical, Aurora, OH, USA), which is known as an intercalator, was employed as the positive control of the assay.

Molecular modelling studies

Molecular modelling studies were carried out using SYBYL software Version 6.9.1 (Tripos, St. Louis, MO, USA) running on an SGI Fuel workstation equipped with 600-MHz R14000 processor (SGI, Mountain View, CA, USA).

Evaluation of *in vitro* antiviral activity.

To evaluate HIV-1 replication inhibition by selected test compounds, *in vitro* antiviral assays were performed using a HeLa4.5/nEGFP reporter cell line. The HeLa4.5/nEGFP reporter cell line was established by transfection of CD4 and LTR driven EGFP reporter protein into the HeLa cell line. HeLa4.5/nEGFP reporter cells were maintained with D-MEM (Sigma) containing 5% FCS (Hyclone, Logan, UT, USA), 500 µg/ml G418, 1 µg/ml blasticidin and 2 µg/ml puromycin.

One day before conducting the assay, 1x10⁴ HeLa4.5/nEGFP cells were seeded into clear bottom black 96-well plates (NUNC, Rochester, NY, USA) with

200 μ l/well medium and incubated at 37°C, 5% CO₂. The next day, 1250 TCID₅₀ HXB2 were added in each well, followed by addition of the test compounds in final concentrations of 5, 1, 0.2, 0.04, 0.008, 0.0016, 0.00032, and 0.000064 μ M. Forty-eight hours after infection, the cells were fixed by 3.2% formaldehyde and the nuclei of cells were stained by 10 μ g/ml Hoechst33342 (Molecular Probes, Engene, OR, USA). EGFP positive cell number (EGFP⁺) and Hoechst33342 positive cell number (hoechst33342⁺) were determined by Cellomics Array Scan, HSC Systems (Beckman Coulter, Tokyo, Japan).

Inhibitory activity of each compound was determined by the following formula:

$$\% \text{ inhibition} = 1 - \frac{\{(\text{EGFP}^+ \text{ cell number with drug} / \text{hoechst33342}^+ \text{ cell number with drug}) - (\text{EGFP}^+ \text{ cell number without infection} / \text{hoechst33342}^+ \text{ cell number without infection})\}}{\{(\text{EGFP}^+ \text{ cell number without drug} / \text{hoechst33342}^+ \text{ cell number without drug}) - (\text{EGFP}^+ \text{ cell number without infection} / \text{hoechst33342}^+ \text{ cell number without infection})\}}$$

Results

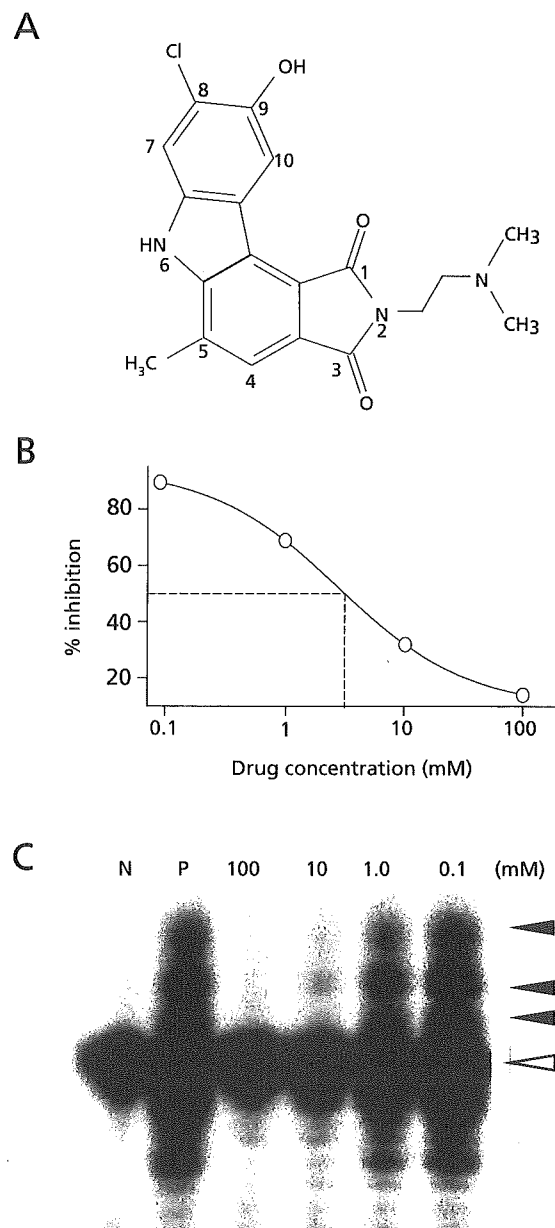
A small molecule bearing a carbazole moiety demonstrated strand-transfer inhibitory activity

A diverse library of 12 000 small-molecule compounds was screened for strand-transfer inhibitory activity at 100 μ M concentration by M8 apparatus. Seventy-two compounds that demonstrated more than 80% strand-transfer-inhibition were selected and applied to the second screening using in-house strand-transfer assay. In the second screening, to confirm dose-dependent inhibition of the test compounds, each compound was tested at four different concentrations. Of the 72 compounds, a compound bearing a carbazole moiety, 8-chloro-2-[2-(dimethylamino)ethyl]-9-hydroxy-5-methylpyrrolo[3,4-c]carbazole-1,3(2H,6H)-dione (coded as **CA-0**), was found to demonstrate potent strand-transfer inhibitory activity (Figure 1A). As shown in Figure 1B, **CA-0** demonstrated clear dose-dependent inhibition of the strand-transfer reaction with an IC₅₀ of 5.00 \pm 3.31 μ M. The dose-dependent inhibition was also confirmed by SDS-PAGE with [γ -³²P] labelled target DNA. As demonstrated in Figure 1C, strand-transferred product bands diminished along with increased concentration of the inhibitor. IC₅₀ value determined from intensities of the bands was 1.24 \pm 0.09 μ M, which was consistent with that evaluated via the plate assay.

Strand-transfer inhibition of 23 carbazole derivatives, and the relationship between their structures and inhibitory activity

To understand the relationship between structure and strand-transfer inhibition activity, we selected 23 carbazole

Figure 1. Structure and strand transfer inhibitory activity of 8-chloro-2-[2-(dimethylamino) ethyl]-9-hydroxy-5-methylpyrrolo[3,4-c]carbazole-1,3(2H,6H)-dione (**CA-0**)

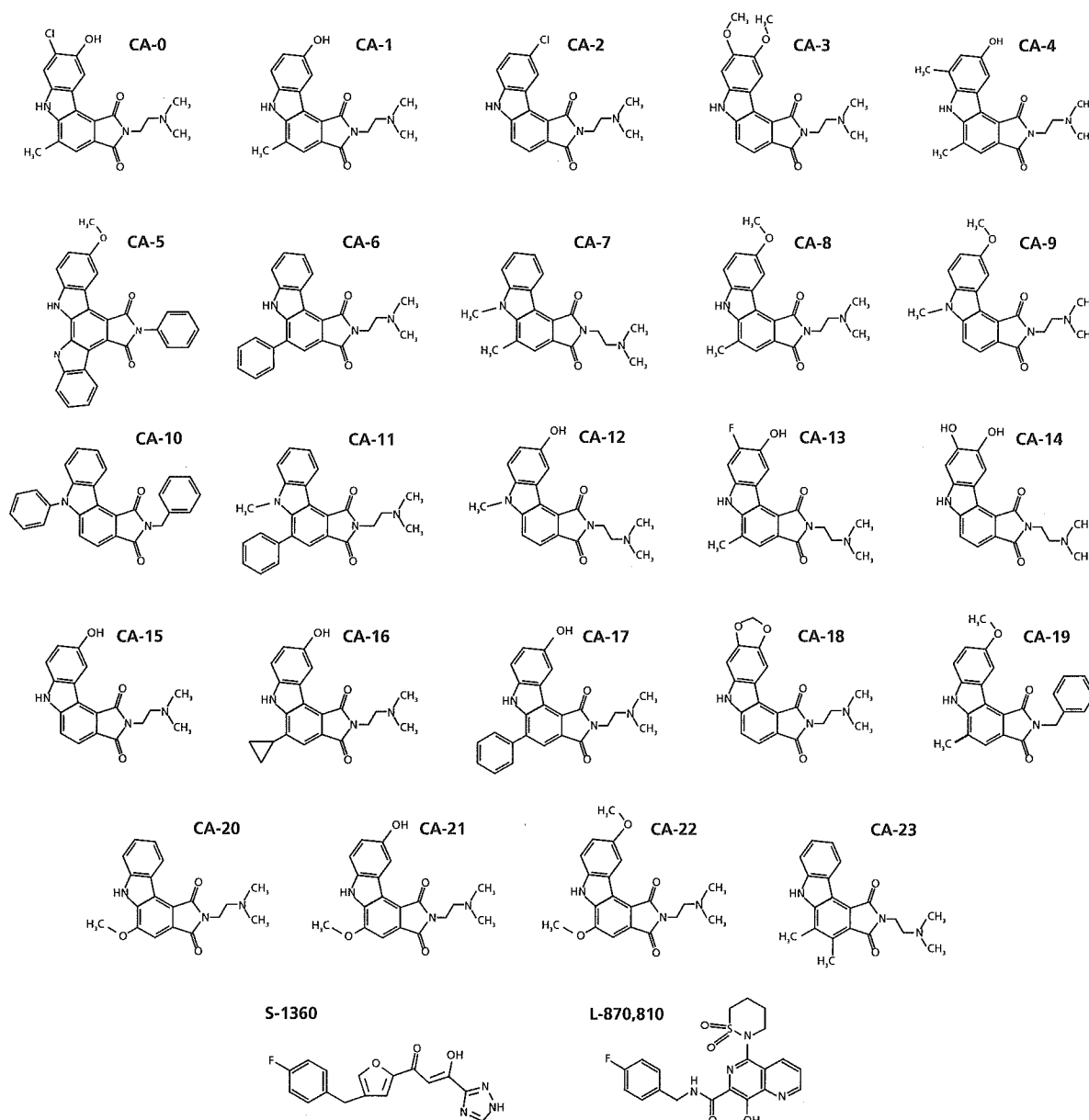


(A) The structure of **CA-0**, a strand transfer inhibitory compound identified from among a library of 12 000 small molecular weight compounds. It has a carbazole structure as a scaffold. The small numbers written beside the structure indicate the residue number of the compound. **(B)** A dose-response curve of **CA-0**. The dotted line indicates the IC₅₀ point of the chemical, which was 5.00 \pm 3.31 μ M. **(C)** A strand transfer assay by radioisotope-labelled oligonucleotide. Lane 1 "N" stands for the negative control, with only a radioisotope-labelled nucleotide. Lane 2 "P" stands for positive control, with radioisotope-labelled nucleotide and recombinant integrase. Lanes 3 to 6 were with inhibitor. The open triangle and solid triangle indicate labelled oligonucleotide and strand transfer products, respectively.

derivatives with different substituents. As demonstrated in Figure 2, all compounds had pyrrolo[3,4-c]carbazole structures as scaffolds, and all except CA-5, CA-10 and CA-19 had 2-dimethylaminoethyl group at position R2. Six of the 23 compounds demonstrated potent strand-transfer inhibition comparable to that of CA-0. These compounds were CA-1, CA-4, CA-8, CA-9, CA-12 and CA-13. IC₅₀

values of these test compounds were similar with positive control S-1360. Moderate inhibitory activities were observed in twelve compounds, CA-2, CA-3, CA-7, CA-11, CA-14, CA-15, CA-16, CA-17, CA-18, CA-21, CA-22 and CA-23. Five compounds, CA-5, CA-6, CA-10, CA-19 and CA-20, did not show significant inhibition, even at the highest concentration tested

Figure 2. Structures of CA-0 and 23 carbazole derivatives evaluated for strand transfer inhibitory activity



CA-0 and 23 related compounds with carbazole scaffold tested for strand-transfer inhibitory activities are depicted. S-1360 and L-870,810, which have previously been reported as potent IN inhibitors, are also shown.

(100 μM). The compounds that demonstrated potent strand-transfer inhibitory activity were also confirmed by gel-based assay, and IC_{50} values determined from the gel-based assay were consistent with the values determined via in-house plate assay (Table 1).

Carbazole derivatives are competitive inhibitors of integrase

To investigate the strand-transfer inhibitory mechanisms and kinetics of the compounds, we determined V_{max} and K_m of the inhibition by Lineweaver–Burke plot analyses. We selected two compounds, **CA-0** and **CA-13**, for the analyses. As summarized in Table 2, larger K_m values (nM)

Table 1. Strand transfer and *in vitro* viral replication inhibitory activities of carbazole derivatives

	IC_{50} in strand transfer assay		Anti-HIV activity
	Plate assay (μM)	Gel assay (μM)	IC_{50} (μM)
<i>(A) High-inhibitory group</i>			
CA-0	5.00 \pm 3.31	1.24 \pm 0.09	0.48 \pm 0.06
CA-13	4.38 \pm 2.78	1.13 \pm 0.21	0.51 \pm 0.12
CA-1	7.94 \pm 4.12	2.97 \pm 0.21	0.92 \pm 0.15
CA-4	8.99 \pm 3.39	6.34 \pm 0.89	1.52 \pm 0.46
CA-8	6.61 \pm 4.17	6.38 \pm 0.32	0.79 \pm 0.07
CA-9	4.42 \pm 1.87	4.10 \pm 0.46	0.80 \pm 0.11
CA-12	5.93 \pm 3.53	3.14 \pm 0.04	0.69 \pm 0.15
<i>(B) Intermediate-inhibitory group</i>			
CA-2	22.50 \pm 2.27	ND	ND
CA-3	72.69 \pm 5.44	ND	ND
CA-7	11.88 \pm 7.66	ND	ND
CA-11	57.00 \pm 3.13	ND	ND
CA-14	17.37 \pm 1.79	ND	ND
CA-15	27.28 \pm 9.10	ND	ND
CA-16	20.51 \pm 15.11	ND	ND
CA-17	50.64 \pm 19.02	ND	ND
CA-18	10.68 \pm 8.88	ND	ND
CA-21	25.01 \pm 10.60	ND	ND
CA-22	16.92 \pm 7.32	ND	ND
CA-23	16.94 \pm 7.82	ND	ND
<i>(C) Intermediate-inhibitory group</i>			
CA-5	>100	ND	ND
CA-6	>100	ND	ND
CA-10	>100	ND	ND
CA-19	>100	ND	ND
CA-20	>100	ND	ND
<i>(D) Previously reported inhibitor</i>			
S-1360	4.67 \pm 1.89	ND	ND

Underline, indicates original compound; IC_{50} , 50% inhibition concentration; ND, not done.

were observed with higher inhibitory concentration, whereas V_{max} values (RU/min) did not change and remained consistent at any inhibitory concentration (Figure 3). As shown in Figure 3A and 3B, data-fitted lines of different time points converged on the Y axis, indicating that **CA-0** and **CA-13** inhibited strand-transfer in a competitive manner.

Carbazole derivatives have not shown intercalative activity

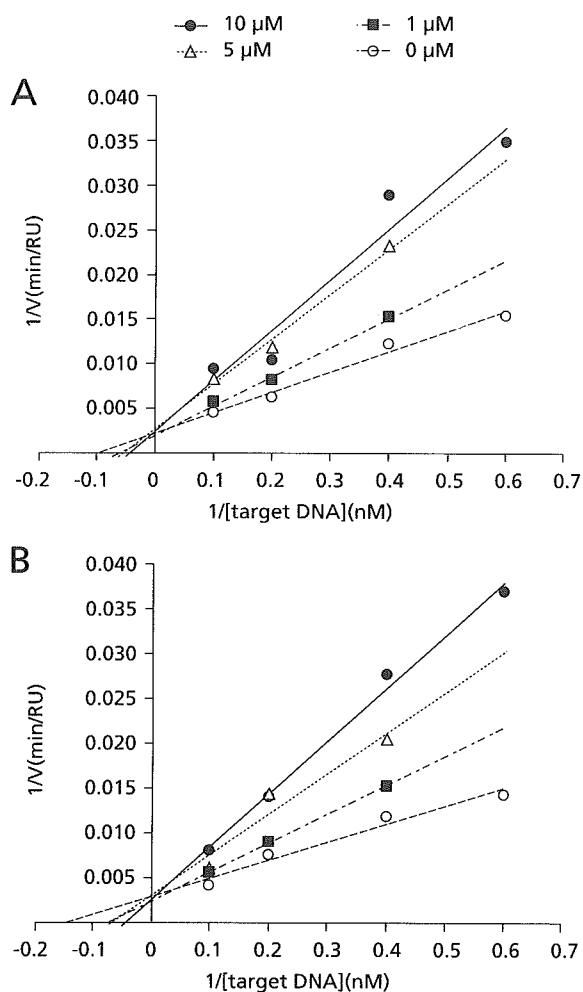
Due to their planar structure and their manner of competitive inhibition, we were concerned that the compounds might have the intercalative activity to destroy substrate dsDNA, rather than binding to the IN to block its enzyme activity. To clear the possibility of the intercalation, EtBr displacement assay was carried out. Since EtBr intercalates into dsDNA and makes visualization possible by growing fluorescence under UV light, intercalative activity of the test compounds can be evaluated by whether the test compounds displace incorporated EtBr out from dsDNA. As shown in Figure 4, fluorescence intensity diminished in a dose-dependent manner by actinomycin D, a compound known as a potent intercalator. In contrast, our two test compounds **CA-0** and **CA-13** did not affect fluorescence intensity, even at the highest concentration of 1 mM, suggesting that **CA-0** and **CA-13** were not intercalators.

Antiviral activity

We employed a single replication infectivity assay using HeLa4.5/EGFP cells to investigate the potency of antiviral activity. IC_{50} values of **CA-0** and the six compounds were 0.48, 0.92, 1.52, 0.79, 0.8, 0.69, 0.51 μM , respectively. The IC_{50} values of all seven compounds were 5.5 to 10.4-fold lower than that of the strand transfer assay (Table 1A). The discrepancy in IC_{50} between the two assays can be explained by stoichiometry of the inhibitor and the target enzyme in the two assays, and the estimated amount of IN in-strand transfer assay was higher than in the

Table 2. Inhibition kinetics of representative carbazole compounds **CA-0** and **CA-13**

Chemical	Concentration	V_{max} (RU/min)	K_m (nM)
CA-0	10 μM	463.16 \pm 63.16	30.40 \pm 7.80
	5 μM	402.58 \pm 32.21	26.21 \pm 7.40
	1 μM	370.14 \pm 84.42	12.71 \pm 2.02
	0 μM	454.55 \pm 0.02	9.18 \pm 1.18
CA-13	10 μM	409.70 \pm 35.47	19.31 \pm 4.68
	5 μM	439.07 \pm 164.74	14.83 \pm 0.24
	1 μM	438.08 \pm 53.85	11.09 \pm 2.42
	0 μM	429.83 \pm 136.46	7.08 \pm 0.64

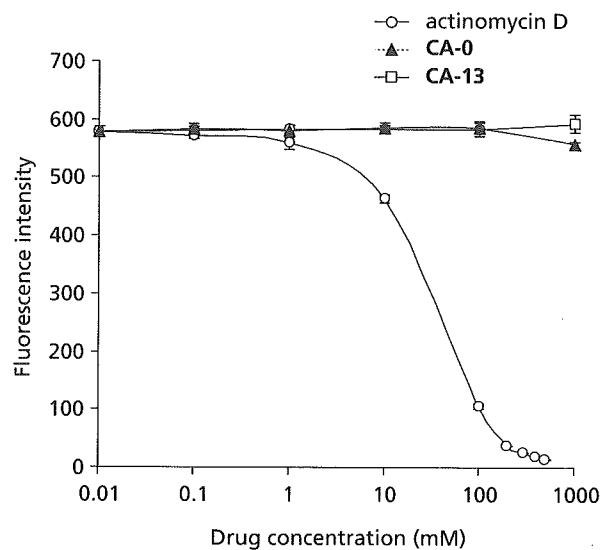
Figure 3. Inhibition kinetics assays of two representative carbazole derivatives, CA-0 and CA-13

Lineweaver-Burke plot analyses of (A) CA-0 and (B) CA-13 are depicted.

HeLa4.5/EGFP assay. Seven compounds exhibited considerable toxicity, suggesting that efforts toward decreasing toxicity are necessary for the further development of carbazole-based inhibitors.

Discussion

Carbazole, a fused phenyl-ring structure with hydrophobicity, has provided an interesting scaffold for the development of novel drugs. Staurosporine, discovered among microbial alkaloids, was the first carbazole derivative reported to demonstrate biological activity (Omura *et al.*, 1977; Furusaki *et al.*, 1978; Furusaki *et al.*, 1982), which was protein kinase C inhibition (Tamaoki *et al.*, 1986).

Figure 4. Ethidium bromide displacement assays of two representative carbazole derivatives, CA-0 and CA-13

To evaluate intercalative activities of carbazole derivatives, ethidium bromide displacement assays were carried out for two representative compounds, CA-0 and CA-13.

Other carbazole derivatives have demonstrated various other activities, such as topoisomerase inhibition (Marotto *et al.*, 2002; Facompre *et al.*, 2002; Carrasco *et al.*, 2001), hypotensive activity (Furusaki *et al.*, 1982), platelet aggregation inhibition (Oka *et al.*, 1986), and anti-fungal activity (Sunthitikawinsakul *et al.*, 2003). In this report we present another possible activity of carbazole derivatives, that of HIV-1 integrase inhibitor.

As compounds with three or four fused aromatic ring structures have been reported to demonstrate intercalative activity (Fukui & Tanaka, 1996; Dziegielewski *et al.*, 2002), we initially suspected that our carbazole derivatives also have intercalative activities, penetrating and disturbing target dsDNA, resulting in pseudo strand-transfer inhibition. Indeed, several carbazole derivatives have been recognized to demonstrate intercalative activity (Facompre *et al.*, 2002; Long *et al.*, 2002). We confirmed that actinomycin D, which is a well-known intercalator (Ross *et al.*, 1979; Wilson & Jones, 1982), demonstrated strand-transfer inhibition in our assay (data not shown). However, taking into consideration the data that our carbazole derivatives inhibited strand-transfer in a competitive manner, and also that the compounds could not displace EtBr out from dsDNA, we assume that our derivatives bind to part of the IN molecule, to the region responsible for DNA target

binding or to the catalytic site responsible for strand-transfer activity.

To understand in greater detail the substituents responsible for strand-transfer inhibitory activity, we analysed 23 carbazole derivatives, and classified them into three categories according to their levels of inhibition (Table 1). Six compounds were classified as the high-inhibition group, which demonstrated IC_{50} of less than 10 μM , 12 compounds were classified as the intermediate group, which demonstrated IC_{50} of greater than 10 μM and less than 100 μM , and five compounds were classified as the non-inhibition group, in which we did not observe significant inhibition even at the highest concentration tested (100 μM).

Comparing the compounds between and within these three categories, we recognized three factors responsible for strand-transfer inhibition. The first and most important factor is the incidence of a 2-dimethylaminoethyl group at position R2 (Figure 1A).

CA-8, which possesses a 2-dimethylaminoethyl group at position R2, demonstrated high inhibitory activity (IC_{50} : $6.61 \pm 4.17 \mu M$), but CA-19 (IC_{50} : $>100 \mu M$), which possesses a phenyl ring structure at the same R2 position, did not demonstrate inhibitory activity. Thus, it is clear that the incidence of a 2-dimethylaminoethyl group, which has a basic property, is critical for strand-transfer inhibition activity. Indeed, we recognized that all compounds in the "high-inhibitory group" and "intermediate-inhibitory group" had this basic substituent at position R2 (Table 1A, 1B, Figure 2). In contrast, three of five compounds in the "non-inhibitory group" had the phenyl ring structure at R2 position. It is thought that these compounds might bind to the acidic region on the IN molecule and compete with the target dsDNA.

The second factor is the incidence of a methyl (Me) group at position R5, R6 or R7. We recognized that compounds in the high inhibitory group had at least one Me group at the R5, R6 or R7 position (Table 1A, Figure 2). Comparing CA-1 (IC_{50} : $7.94 \pm 4.12 \mu M$), CA-4 (IC_{50} : $8.99 \pm 3.39 \mu M$), and CA-12 (IC_{50} : $5.93 \pm 3.53 \mu M$) with CA-15 (IC_{50} : $27.28 \pm 9.10 \mu M$), it is clear that the incidence of an Me group within the R5 to R7 positions was an important factor for enhanced inhibitory activity. It seems that the position of the substituent may not be critical between R5 and R6, as we did not see significant differences between CA-1 (IC_{50} : $7.94 \pm 4.12 \mu M$) and CA-12 (IC_{50} : $5.93 \pm 3.53 \mu M$), and also between CA-8 (IC_{50} : $6.61 \pm 4.17 \mu M$) and CA-9 (IC_{50} : $4.42 \pm 1.87 \mu M$).

According to the IC_{50} levels of CA-5 ($>100 \mu M$), CA-6 ($>100 \mu M$) and CA-11 ($>100 \mu M$), it appears that bulky substituents at the R5 position have a negative effect on inhibition (Table 1C, Figure 2). Furthermore, the inhibition potential of the three compounds CA-1 (IC_{50} :

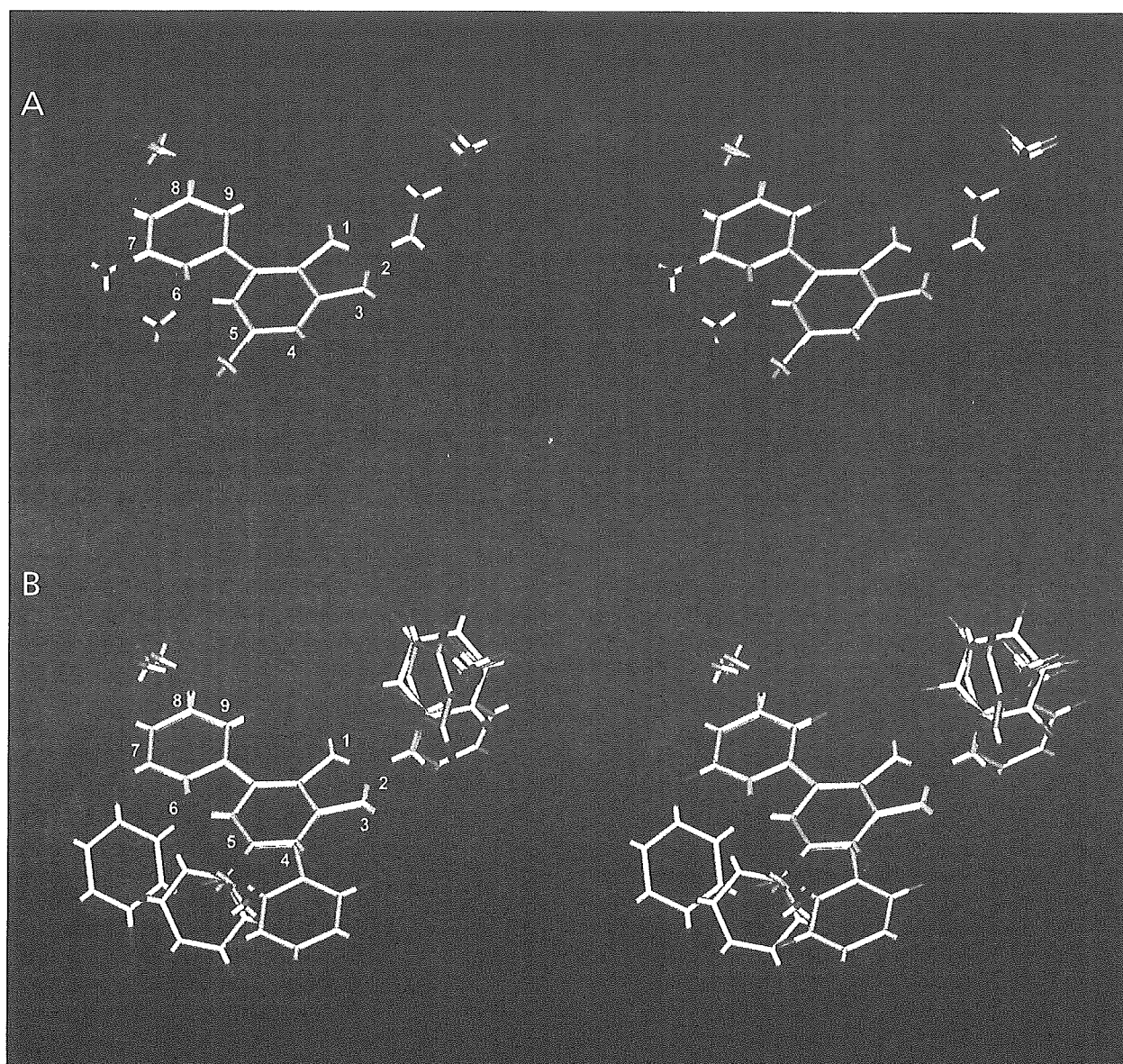
$7.94 \pm 4.12 \mu M$), CA-16 (IC_{50} : $20.51 \pm 15.11 \mu M$) and CA-17 (IC_{50} : $50.64 \pm 19.02 \mu M$) depended on the molecular size of their R5 substituents. It is probable that the R5 substituents of these compounds were too large and that they interfered with surrounding molecules forming the binding site (Table 1A, 1B, Figure 2). These data indicate that the binding site of carbazole might have a space limitation, and thus the size and shape of the molecules may be important factors for inhibitor activity.

The third factor is the substituent at position R9. Comparing CA-20 (IC_{50} : $>100 \mu M$), CA-21 (IC_{50} : $25.01 \pm 10.60 \mu M$) and CA-22 (IC_{50} : $16.92 \pm 7.32 \mu M$), these three compounds were identical, with the exception of the substituent at position R9 (Table 1B, 1C, Figure 2). CA-21 and CA-22 have hydroxyl residue and a methoxy group at position R9, respectively. We noticed a significant difference in inhibitory activity between CA-20 and CA-21, and between CA-20 and CA-22, suggesting the possibility that both the hydroxyl group and the methoxy group at R9 formed hydrogen bonds with the amino acid molecules forming the binding sites, as these two substituents have the potential to be hydrogen bond acceptors. It appears that hydroxyl and methoxy groups have similar effects on strand-transfer inhibitory activities. In addition to the above three factors, we found that molecular interaction between R8 and R9 substituents, and their arrangement, are also important determinants for efficient inhibitory activity. CA-3, with two methoxy groups at R8 and R9, appears to have a bulky arrangement of the two side chains, and demonstrated an IC_{50} of $72.69 \pm 5.44 \mu M$, whereas CA-14 and CA-18, which were expected to have horizontal arrangements, demonstrated lower IC_{50} values of $17.37 \pm 1.79 \mu M$ and $10.68 \pm 8.88 \mu M$, respectively (Table 1B, Figure 2).

To summarize these structural elements, and to understand the common structure of molecules that demonstrated strand-transfer inhibitory activity, we superposed inhibitor structures having significant strand-transfer inhibition (CA-0, CA-1, CA-4, CA-8, CA-9, CA-12 and CA-13) (Figure 5A), and the structures of compounds with no inhibition (CA-5, CA-6, CA-10, CA-19 and CA-20) (Figure 5B). In comparing these two overlapped figures, we found that the compounds with inhibitory activity share a largely identical structure and similar molecular size. In contrast, the non-inhibitory compounds had larger and more uneven-shaped side chains. Overall, the superposed structures indicate that the molecules should be planar and have basic diethylaminoethyl groups to demonstrate strand-transfer inhibitory activity.

In conclusion, we have identified a small molecular weight compound with a carbazole scaffold, which can be the lead compound for developing novel IN inhibitors. Furthermore, analysing the IN inhibitory mechanisms of

Figure 5. A structural comparison between high/intermediate inhibitory compounds and non-inhibitory compounds



Superposed structures of (A) five non-inhibitory compounds, CA-5, 6, 10, 19 and 20, and (B) seven inhibitory compounds, CA-0, 1, 4, 8, 9, 12 and 13, are demonstrated in stereo-view images. In both figures, residue numbers are indicated beside the structures. Red, dark blue and light blue indicate oxygen, nitrogen and hydrogen molecules, respectively. Green indicates chlorine or fluorine molecules. SYBYL software Version 6.9.1 running on an SGI Fuel workstation was used to construct the figures.

carbazole derivatives may yield more detailed information regarding HIV-1 IN structure and function.

Acknowledgements

This study was supported by a grant from the Human Sciences Foundation, the Organization of Pharmaceutical

Safety and Research of Japan and the Ministry of Health, Labor and Welfare of Japanese Government. This study was partly supported by the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NIBIO)

We would like to thank Dr. Haruo Tanaka and Takuro Shiomi, professor and associate professor of Kitazato

Institute, for their valuable advice and comments. We would also like to thank the laboratory members of Toyama Chemical Co. Ltd. for supplying the compounds in the study. Finally, we would like to thank Ms. Mary Phillips and Ms. Yumi Fujiuji for preparing the manuscript.

References

- Balzarini J (2004) Current status of the non-nucleoside reverse transcriptase inhibitors of human immunodeficiency virus type 1. *Current Topics in Medicinal Chemistry* **4**:921–944.
- Burke CJ, Sanyal G, Bruner MW, Ryan JA, LaFemina RL, Robbins HL, Zeft AS, Middaugh CR & Cordingley MG (1992) Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. *The Journal of Biological Chemistry* **267**:9639–9644.
- Cain BF, Baguley BC & Denny WA (1978) Potential antitumor agent. 28. deoxyribonucleic acid polyintercalating agents. *Journal of Medicinal Chemistry* **21**:658–668.
- Carrasco C, Vezin H, Wilson WD, Ren J, Chaires JB & Bailly C (2001) DNA binding properties of the indolocarbazole antitumor drug NB-506. *Anticancer Drug Design* **16**:99–107.
- Chun TW, Finzi D, Margolick J, Chadwick K, Schwartz D & Siliciano RF (1995) *In vivo* fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. *Nature Medicine* **1**:1284–1290.
- Craigie R, Hickman AB & Engelman A (1995) Integrase. in *HIV: A Practical Approach – Volume 2: Biochemistry, Molecular Biology, and Drug Discovery*, pp. 53–71. Edited by J Karn. New York: Oxford University Press.
- Dayam R & Neamati N (2003) Small-molecule HIV-1 integrase inhibitors: the 2001–2002 update. *Current Pharmacology Design* **9**:1789–1802.
- De Clercq E (1992) HIV inhibitors targeted at the reverse transcriptase. *AIDS Research and Human Retroviruses* **8**:119–134.
- Dziegielewska J, Slusarski B, Konitz A, Skladanowski A & Konopa J (2002) Intercalation of imidazoacridinones to DNA and its relevance to cytotoxic and antitumor activity. *Biochemical Pharmacology* **63**:1653–1662.
- Engelman A & Craigie R (1992) Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function *in vitro*. *Journal of Virology* **66**:6361–6369.
- Engelman A, Englund G, Orenstein JM, Martin MA & Craigie R (1995) Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *Journal of Virology* **69**:2729–2736.
- Engelman A, Hickman AB & Craigie R (1994) The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *Journal of Virology* **68**:5911–5917.
- Facompre M, Carrasco C, Colson P, Houssier C, Chisholm JD, Van Vranken DL & Bailly C (2002) DNA binding and topoisomerase I poisoning activities of novel disaccharide indolocarbazoles. *Molecular Pharmacology* **62**:1215–1227.
- Fukui K & Tanaka K (1996) The acridine ring selectively intercalated into a DNA helix at various types of abasic sites: double strand formation and photophysical properties. *Nucleic Acids Research* **24**:3962–3967.
- Furusaki A, Hashiba N, Matsumoto T, Hirano A, Iwai Y & Omura S (1978) X-ray crystal structure of staurosporine: a new alkaloid from a Streptomyces strains. *Journal of the Chemical Society. Chemical Communications* 800–801.
- Furusaki A, Hashiba N, Matsumoto T, Hirano A, Iwai Y & Omura S (1982) The crystal and molecular structure of staurosporine, a new alkaloid from a Streptomyces strains. *Bulletin of the Chemical Society of Japan* **55**:3681–3685.
- Goldgur Y, Craigie R, Cohen GH, Fujiwara T, Yoshinaga T, Fujishita T, Sugimoto H, Endo T, Murai H & Davies DR (1999) Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. *Proceedings of the National Academy of Sciences, USA* **96**:13040–13043.
- Grobler JA, Stillmock K, Hu B, Witmer M, Felock P, Espeseth AS, Wolfe A, Egbertson M, Bourgeois M, Melamed J, Wai JS, Young S, Vacca J & Hazuda DJ (2002) Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. *Proceedings of the National Academy of Sciences USA* **99**:6661–6666.
- Hazuda DJ, Anthony NJ, Gomez RP, Jolly SM, Wai JS, Zhuang L, Fisher TE, Embrey M, Guare JP, Jr., Egbertson MS, Vacca JP, Huff JR, Felock PJ, Witmer MV, Stillmock KA, Danovich R, Grobler J, Miller MD, Espeseth AS, Jin L, Chen IW, Lin JH, Kassahun K, Ellis JD, Wong BK, Xu W, Pearson PG, Schleif WA, Cortese R, Emimi E, Summa V, Holloway MK & Young SD (2004) A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proceedings of the National Academy of Sciences USA* **101**:11233–11238.
- Hazuda DJ, Felock P, Witmer M, Wolfe A, Stillmock K, Grobler JA, Espeseth A, Gabryelski L, Schleif W, Blau C & Miller MD (2000) Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **287**:646–650.
- Imamichi T (2004) Action of anti-HIV drugs and resistance: reverse transcriptase inhibitors and protease inhibitors. *Current Pharmaceutical Design* **10**:4039–4053.
- Johnson AA, Marchand C & Pommier Y (2004) HIV-1 integrase inhibitors: a decade of research and two drugs in clinical trial. *Current Topics in Medicinal Chemistry* **4**:1059–1077.
- Khan E, Mack JP, Katz RA, Kulkosky J & Skalka AM (1991) Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Research* **19**:851–860.
- Kohl NE, Emimi EA, Schleif WA, Davis LJ, Heimbach JC, Dixon RA, Scolnick EM & Sigal IS (1988) Active human immunodeficiency virus protease is required for viral infectivity. *Proceedings of National Academy of Sciences USA* **85**:4686–4690.
- LaFemina RL, Schneider CL, Robbins HL, Callahan PL, LeGrow K, Roth E, Schleif WA & Emimi EA (1992) Requirement of active human immunodeficiency virus type 1 integrase enzyme for productive infection of human T-lymphoid cells. *Journal of Virology* **66**:7414–7419.
- Long BH, Rose WC, Vyas DM, Matson JA & Forenza S (2002) Discovery of antitumor indolocarbazoles: rebeccamycin, NSC 655649, and fluoroindolocarbazoles. *Current Medicinal Chemistry. Anti-Cancer Agents* **2**:255–266.
- Marotto A, Kim YS, Schulze E & Pindur U (2002) New indolocarbazoles as antitumor active compounds: evaluation of the target by experimental and theoretical studies. *Pharmazie* **57**:194–197.
- Oka S, Kodama M, Takeda H, Tomizuka N & Suzuki H (1986) Staurosporine, a potent platelet aggregation inhibitor from a Streptomyces species. *Agricultural and Biological Chemistry* **50**:2723–2727.
- Omura S, Iwai Y, Hirano A, Nakagawa A, Awaya J, Tsuchiya H, Takahashi Y & Masuma R (1977) A new alkaloid AM-2282 of Streptomyces origin. Taxonomy, fermentation, isolation and preliminary characterization. *The Journal of Antibiotics (Tokyo)* **30**:275–282.

- Pluymers W, Pais G, Van Maele B, Pannecouque C, Fikkert V, Burke TR, Jr., De Clercq E, Witvrouw M, Neamati N & Debyser Z (2002) Inhibition of human immunodeficiency virus type 1 integration by diketo derivatives. *Antimicrobial Agents and Chemotherapy* **46**:3292–3297.
- Pommier Y, Marchand C & Neamati N (2000) Retroviral integrase inhibitors year 2000: update and perspectives. *Antiviral Research* **47**:139–148.
- Ross WE, Glaubiger D & Kohn KW (1979) Qualitative and quantitative aspects of intercalator-induced DNA strand breaks. *Biochimica et Biophysica Acta* **562**:41–50.
- Ruscetti FW (1985) Immunopathology associated with human lymphotropic retroviruses. *Survey and Synthesis of Pathology Research* **4**:216–226.
- Schauer M & Billich A (1992) The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA-binding. *Biochemical and Biophysical Research Communications* **185**:874–880.
- Sunthitikawinsakul A, Kongkathip N, Kongkathip B, Phonnakhu S, Daly JW, Spande TF, Nimit Y & Rochanaruangrai S (2003) Coumarins and carbazoles from *Clausena excavata* exhibited antimycobacterial and antifungal activities. *Planta Medica* **69**:155–157.
- Tamaoki T, Nomoto H, Takahashi I, Kato Y, Morimoto M & Tomita F (1986) Staurosporine, a potent inhibitor of phospholipid/Ca²⁺ dependent protein kinase. *Biochemical and Biophysical Research Communications* **135**:397–402.
- Tronchet JM & Seman M (2003) Non-nucleoside inhibitors of HIV-1 reverse transcriptase: from the biology of reverse transcription to molecular design. *Current Topics in Medicinal Chemistry* **3**:1496–1511.
- Wilson WD & Jones RL (1982) Interaction of actinomycin D, ethidium, quinacrine, daunorubicin, and tetralysine with DNA: 31P NMR chemical shift and relaxation investigation. *Nucleic Acids Research* **10**:1399–1410.
- Woerner AM & Marcus-Sekura CJ (1993) Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. *Nucleic Acids Research* **21**:3507–3511.

Received 22 August 2005, accepted 27 September 2005

Impact of HIV-1 Subtype and Antiretroviral Therapy on Protease and Reverse Transcriptase Genotype: Results of a Global Collaboration

Rami Kantor^{1*}, David A. Katzenstein¹, Brad Efron², Ana Patricia Carvalho³, Brian Wynhoven⁴, Patricia Cane⁵, John Clarke⁶, Sunee Sirivichayukul⁷, Marcelo A. Soares⁸, Joke Snoeck⁹, Candice Pillay¹⁰, Hagit Rudich¹¹, Rosangela Rodrigues¹², Africa Holguin¹³, Koya Ariyoshi¹⁴, Maria Belen Bouzas¹⁵, Pedro Cahn¹⁵, Wataru Sugiura¹⁴, Vincent Soriano¹³, Luis F. Brigido¹², Zehava Grossman¹¹, Lynn Morris¹⁰, Anne-Mieke Vandamme⁹, Amilcar Tanuri⁸, Praphan Phanuphak⁷, Jonathan N. Weber⁶, Deenan Pillay¹⁶, P. Richard Harrigan⁴, Ricardo Camacho³, Jonathan M. Schapiro¹, Robert W. Shafer¹

1 Division of Infectious Disease and Center for AIDS Research, Stanford University, Stanford, California, United States of America **2** Department of Statistics and Division of Biostatistics, Stanford University, Stanford, California, United States of America **3** Hospital Egas Moniz, Lisbon, Portugal, **4** BC Centre for Excellence in HIV/AIDS, Vancouver, British Columbia, Canada, **5** Health Protection Agency, Porton Down, United Kingdom, **6** Wright Fleming Institute, Imperial College, St. Mary's Hospital, London, United Kingdom **7** Chulalongkorn University, Bangkok, Thailand, **8** Universidade Federal do Rio de Janeiro, Brazil, **9** Rega Institute for Medical Research, Leuven, Belgium, **10** National Institute of Communicable Diseases, Johannesburg, South Africa, **11** Central Virology, Public Health Laboratories, Ministry of Health, Tel Hashomer, Israel, **12** Instituto Adolfo Lutz, Sao Paulo, Brazil, **13** Hospital Carlos III, Madrid, Spain, **14** National Institute of Infectious Diseases, Tokyo, Japan, **15** Fundación Huesped, Buenos Aires, Argentina, **16** University College London and Health Protection Agency, London, United Kingdom

Competing Interests: The authors have declared that no competing interests exist.

Author Contributions:
See Acknowledgments.

Academic Editor: David D. Ho, The Rockefeller University, United States of America

Citation: Kantor R, Katzenstein DA, Efron B, Carvalho AP, Wynhoven B, et al. (2005) Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: Results of a global collaboration. *PLoS Med* 2(4): e112.

Received: October 8, 2004

Accepted: March 7, 2005

Published: April 26, 2005

DOI:

10.1371/journal.pmed.0020112

Copyright: © 2005 Kantor et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abbreviations: NRTI, nucleoside reverse transcriptase inhibitor; NNRTI, non-nucleoside reverse transcriptase inhibitor; PI, protease inhibitor; RT, reverse transcriptase; RTI, reverse transcriptase inhibitor

*To whom correspondence should be addressed. E-mail: rkantor@brown.edu

†Current affiliation: Division of Infectious Diseases, Brown University, Providence, Rhode Island, United States of America

ABSTRACT

Background

The genetic differences among HIV-1 subtypes may be critical to clinical management and drug resistance surveillance as antiretroviral treatment is expanded to regions of the world where diverse non-subtype-B viruses predominate.

Methods and Findings

To assess the impact of HIV-1 subtype and antiretroviral treatment on the distribution of mutations in protease and reverse transcriptase, a binomial response model using subtype and treatment as explanatory variables was used to analyze a large compiled dataset of non-subtype-B HIV-1 sequences. Non-subtype-B sequences from 3,686 persons with well characterized antiretroviral treatment histories were analyzed in comparison to subtype B sequences from 4,769 persons. The non-subtype-B sequences included 461 with subtype A, 1,185 with C, 331 with D, 245 with F, 293 with G, 513 with CRF01_AE, and 618 with CRF02_AG. Each of the 55 known subtype B drug-resistance mutations occurred in at least one non-B isolate, and 44 (80%) of these mutations were significantly associated with antiretroviral treatment in at least one non-B subtype. Conversely, of 67 mutations found to be associated with antiretroviral therapy in at least one non-B subtype, 61 were also associated with antiretroviral therapy in subtype B isolates.

Conclusion

Global surveillance and genotypic assessment of drug resistance should focus primarily on the known subtype B drug-resistance mutations.

Introduction

The HIV-1 pandemic resulted from the cross-species transmission of a lentivirus, most likely of chimpanzee origin, that began spreading among humans during the first half of the previous century [1,2,3]. The progeny of this zoonotic infection—designated HIV-1 group M (main) viruses—make up the vast majority of HIV-1 infections. During their spread among humans, group M viruses have developed an extraordinary degree of genetic diversity, and most can be segregated into nine pure subtypes and several commonly circulating recombinant forms [4].

HIV-1 subtype B is the predominant subtype in North America, Western Europe, and Australia. The antiretroviral drugs used to treat HIV were developed using biophysical, biochemical, and in vitro studies of subtype B isolates, and most data on the genetic mechanisms of HIV-1 drug resistance are from subtype B viruses. However, HIV-1 subtype B viruses account for only approximately 12% of the global HIV pandemic [5], and as therapy is introduced into developing countries, the number of persons with non-B viruses initiating therapy will increase dramatically.

HIV-1 subtypes differ from one another by 10%–12% of their nucleotides and 5%–6% of their amino acids in protease and reverse transcriptase (RT) [6]. Intersubtype nucleotide differences influence the spectrum of amino acid substitutions resulting from point mutations, and intersubtype amino acid differences influence the biochemical and biophysical microenvironment within the protease and RT [7,8]. These differences among subtypes therefore could influence the spectrum of mutations that develop during selective drug pressure.

An increasing number of observational studies, in vitro and in vivo, suggest that the currently available protease and RT inhibitors are as active against non-B viruses as they are against subtype B viruses [9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26]. However, fewer data are available on the genetic mechanisms of drug resistance in non-B viruses, and some in vitro and in vivo observations suggest that the various subtypes may respond differently to certain antiretroviral drugs [27,28,29,30,31,32,33,34,35].

Identifying the relevant drug-resistance mutations among non-B subtypes will be important for monitoring the evolution and transmission of drug resistance, for determining initial treatment strategies for persons infected with non-B viruses, and for interpreting genetic resistance among patients who fail antiretroviral therapy.

In this study, we characterize protease and RT mutations in non-B HIV-1 subtypes from persons receiving antiretroviral therapy, and attempt to answer the following two questions. (i) Do the mutations that cause drug resistance in subtype B viruses also develop in non-B viruses exposed to antiretroviral drugs? (ii) Do novel mutations emerge in non-subtype-B viruses during antiretroviral drug failure that are not recognized in subtype B viruses?

Methods

HIV-1 Sequences and Antiretroviral Treatments

Sequences of HIV-1 protease (positions 1–99) and RT (positions 1–240) from persons whose antiretroviral treatment history was known were collected from the published

literature and from 14 laboratories in 12 countries. Persons were considered untreated if they had never been exposed to antiretroviral drugs, and treated if they were receiving RT inhibitors (RTIs) and/or protease inhibitors (PIs) at the time the isolate was obtained. Sequences from treated persons were included for analysis only if they were obtained from persons whose entire treatment histories were known. If multiple isolates from the same person were sequenced, only the latest isolate was included for analysis. Only sequences determined using dideoxyterminator sequencing were included in the analysis. In all, 99% of sequences were determined using direct PCR (population-based sequencing), and 1% of sequences represented the consensus sequence of multiple clones.

Samples obtained from patients were submitted to clinical and research laboratories for resistance testing in the course of evaluation and care of HIV infection. Data analyzed included published and presented data obtained under protocols approved by national and local institutional review boards or ethical review panels in each country. Sequence, demographic, and treatment data, unlinked from all personal identifiers, were analyzed at Stanford University under a protocol approved by the Stanford University Panel on Human Subjects.

Subtype Assignment

Similarity plotting and bootscanning using a window size of 400 nucleotides and a step size of 40 nucleotides were performed using reference sequences for each of the nine pure subtypes (A, B, C, D, F, G, H, J, and K) and two recombinant forms (CRF01_AE and CRF02_AG) [36]. Isolates that contained a combination of more than one subtype were excluded from analysis, except when subtypes A and C were detected in a pattern consistent with CRF02_AG. Because CRF01_AE *pol* sequences do not contain recombinant breakpoints, subtype assignment was based on the fact that *pol* CRF01_AE and pure A sequences are divergent. This approach had an accuracy of 96% when applied to the protease and RT genes of 137 well characterized subtype A, CRF01_AE, and CRF02_AG isolates with known subtypes based on *pol* and *gag* and/or *env*, with most errors resulting from the misclassification of subtype A protease sequences as CRF01_AE (data not shown).

Reference sequences used were U455 (subtype A), CM240 (CRF01_AE), IbNG (CRF02_AG), HXB2 (subtype B), C2220 (subtype C), NDK (subtype D), 93BR020 (subtype F), SE6165 (subtype G), 90CR056 (subtype H), SE9173c (subtype J), 97EQTB11C (subtype K), YBF30 (Group N), and ANT70C (Group O). A total of 223 protease and 307 RT sequences of indeterminate subtype were excluded from the analysis.

Mutation Definitions

Each sequence was translated and compared to the consensus B protease and RT sequences in the Los Alamos HIV Sequence database (<http://hiv-web.lanl.gov>) using the HIVSeq program [37]. Mutations were defined as differences from the wild-type consensus B sequence. Known subtype B drug-resistance mutations were defined as follows: 18 nucleoside RTI (NRTI)-resistance positions at 41, 44, 62, 65, 67, 69, 70, 74, 75, 77, 115, 116, 118, 151, 184, 210, 215, and 219; 15 non-nucleoside RTI (NNRTI)-resistance positions at 98, 100,

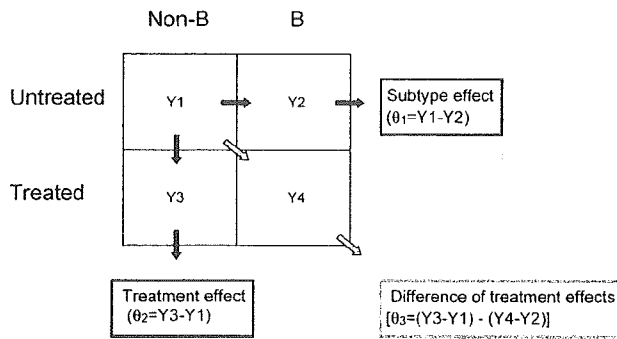


Figure 1. Binomial Response Model Used to Evaluate Subtype and Treatment Effects on Genotypic Evolution for Each Protease and RT Position

A separate model was created for each non-B subtype. The frequencies of mutations at each position in four patient groups (untreated subtype B, treated subtype B, untreated non-B, and treated non-B) were converted to Y scores using a cube root transformation (similar to a logistic transform). Subtype effect was evaluated by calculating θ_1 , the score differences between non-B and B subtypes in untreated persons. The treatment effect was evaluated by calculating θ_2 , the score differences between treated and untreated persons within the same subtype. The subtype-treatment interaction was evaluated by calculating θ_3 , the difference of differences in the 2×2 table, or the difference in treatment effects between non-B and B subtypes.

DOI: 10.1371/journal.pmed.0020112.g001

101, 103, 106, 108, 179, 181, 188, 190, 225, 227, 230, 236, and 238; and 22 protease inhibitor (PI)-resistance positions at 10, 20, 24, 30, 32, 33, 36, 46, 47, 48, 50, 53, 54, 63, 71, 73, 77, 82, 84, 88, 90, and 93 [38,39]. Mutations also included differences from consensus B that were present as part of a nucleotide mixture.

Polymorphisms were defined as mutations that occurred in more than 1% of sequences from untreated persons. Subtype-specific polymorphisms were defined as mutations that were significantly more prevalent in each non-B subtype than in subtype B viruses from untreated persons. Subtype-specific treatment-related mutations were defined as mutations that were significantly more prevalent in HIV-1 isolates from treated persons than in isolates from untreated persons infected with the same subtype.

Quality Control

Phylogenetic analysis to detect duplicate sequences identified 23 pairs of identical sequences and 1,039 pairs of sequences that differed from one another by no more than 1% of their nucleotides. To reduce the likelihood that isolates from different persons with similar mutations resulted from duplicate reporting, laboratory contamination, or HIV-1 transmission, only one sequence from each of these 1,062 sequence pairs was included in the analyses in this study. Although all extant HIV-1 isolates are epidemiologically linked through chains of transmission, protease or RT sequences that diverge by 1% or less appear to be more likely to result from direct transmission than those that diverge by more than 1% [40].

To distinguish mutations developing in multiple individuals from mutations that developed in a smaller number of founder viruses, we reconstructed the ancestral sequences at

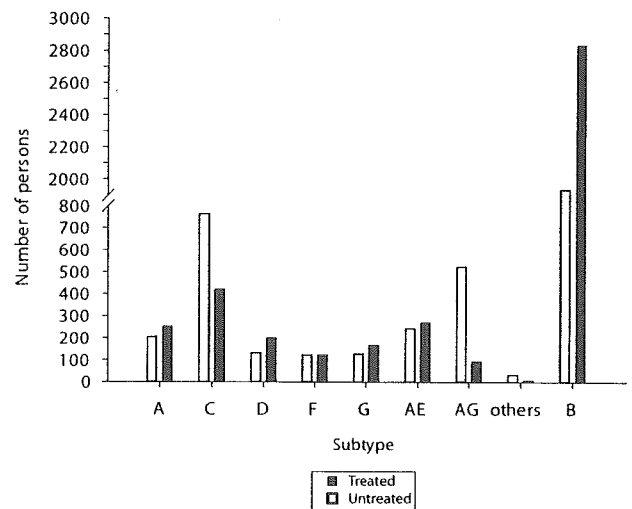


Figure 2. Number of Treated and Untreated Persons Infected with B and Non-B HIV-1 Subtypes from Whom Protease and/or RT Sequences Were Available for Analysis.

DOI: 10.1371/journal.pmed.0020112.g002

each node of the phylogenetic trees for each subtype and counted the number of times each mutation was predicted to have developed within a subtype. Mutations for which founder viruses accounted for 75% or more of occurrences (i.e., the number of branches on which the mutation has developed divided by the number of sequences with the mutation is less than 75%) were not considered treatment-

Table 1. Geographical Origin of Persons Infected with Non-B Subtypes

Subtype	Country
A	Uganda (215), United Kingdom (49), Kenya (41), Rwanda (34), Canada (23), France (21), Belgium (14), Sweden (10), Cameroon (10), Other (43)
C	India (225), South Africa (220), United Kingdom (145), Brazil (106), Canada (93), Zimbabwe (71), Israel (59), Botswana (42), Zambia (29), Sweden (29), Belgium (24), Uganda (19), Denmark (17), Burundi (16), France (14), Other (73)
D	Uganda (189), Cameroon (21), Cuba (21), Kenya (20), Canada (18), United Kingdom (14), Other (46)
F	Brazil (127), Romania (46), Cameroon (21), Cuba (10), Other (40)
G	Portugal (193), Cameroon (20), Spain (14), France (14), Other (53)
CRF01_AE	Thailand (260), Vietnam (98), Japan (35), Cameroon (27), Canada (11), United Kingdom (10), Uganda (10), France (10), Other (52)
CRF02_AG	Cameroon (207), Ivory Coast (128), France (47), Spain (46), Portugal (37), Senegal (24), United Kingdom (18), Italy (12), Belgium (12), Canada (11), Gabon (10), Other (63)
Others ^a	Cameroon (14), Other (26)

Numbers of persons from countries for which ten or more sequences were available are shown in parentheses.

^aSubtypes H₁ and K.

DOI: 10.1371/journal.pmed.0020112.t001

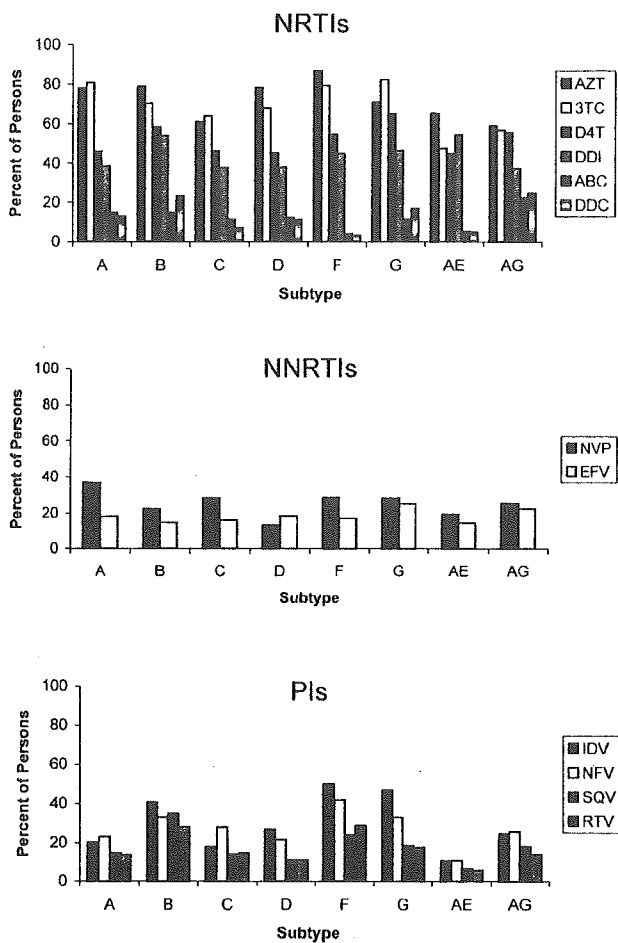


Figure 3. Proportions of Persons Receiving Treatment with Specific NRTIs, NNRTIs, and PIs

The number of persons with non-B virus receiving the NRTI tenofovir (ten), the NNRTI delavirdine (three), and the PIs amprenavir (13) and lopinavir (28) are not shown. 3TC, lamivudine; ABC, abacavir; AZT, zidovudine; D4T, stavudine; DDC, zalcitabine; DDI, didanosine; EFV, efavirenz; IDV, indinavir; NFV, nelfinavir; NVP, nevirapine; RTV, ritonavir; SQV, saquinavir.
DOI: 10.1371/journal.pmed.0020112.g003

related mutations. For this analysis, phylogenetic trees were created using the neighbor-joining method using the HKY85 model with gamma distribution within PAUP* version 4.0b10 for each subtype and each gene. Ancestral sequences were reconstructed using MESQUITE version 1.02 (<http://www.mesquiteproject.org>).

To further reduce the influence of transmitted drug resistance on this analysis and to exclude the possibility that some untreated persons were classified incorrectly, sequences from untreated persons containing two or more non-

polymorphic resistance mutations were excluded from the analysis. This approach was predicated on the strong likelihood that the presence of two mutations at highly conserved sites in untreated persons does not reflect natural variation in protease or RT but rather is most consistent with previous selective drug pressure. Based on this criterion, 23 protease and 25 RT sequences from 47 persons were excluded from the analysis: 22 CRF01_AE sequences, 15 subtype C sequences, six CRF02_AG sequences, five subtype G sequences, four subtype A sequences, three subtype F sequences, and one subtype D sequence. The mutations in the excluded sequences consisted almost entirely of the NRTI-resistance mutations at positions 41, 67, 70, 210, 215, and 219; the NNRTI-resistance mutations at positions 103, 181, and 190; and the PI-resistance mutations at positions 48, 82, and 90. An analysis that included these 56 sequences did not affect any of the significant findings in the study because these mutations were so much more common in treated than in untreated persons in multiple different subtypes (data not shown).

Statistical Analysis

Frequencies of mutations at each RT and protease codon were analyzed by a binomial response model employing a cube root transformation (similar to a logistic transform, with higher accuracy for extreme values) to identify significant differences in polymorphisms and treatment-related mutations between subtypes. Mutation frequencies for treated and untreated persons were compared for each subtype (Figure 1). This analysis defined three parameters: (i) a subtype parameter (θ_1), comparing codons between untreated persons infected with B and each of the non-B subtypes; (ii) a treatment parameter (θ_2), comparing codons between treated and untreated sequences of the same subtype; and (iii) an interaction parameter (θ_3), comparing the effect of treatment between subtype B and each of the non-B subtypes.

To increase the statistical power of our analysis, we made two simplifications. First, we did not distinguish between distinct substitutions at the same position; all differences from consensus B were considered mutations. Second, viruses were categorized only according to the classes of drugs (PI, NRTI, NNRTI) to which they had been exposed.

To correct for multiple comparisons between subtype B and each non-B subtype, significant results included those z values exceeding three in absolute value, according to a 0.05 Benjamini-Hochberg false discovery rate criterion [41]. This method is a sequential Bonferroni-type procedure that is appropriate for situations in which multiple statistically significant associations are expected. The coefficients in the binomial response model were ranked in ascending order, and each hypothesis of rank r was compared with a significance cutoff of 0.05 (false discovery rate) multiplied by r/n , where n was 99 for the protease mutations and 240 for the RT mutations (i.e., number of comparisons).

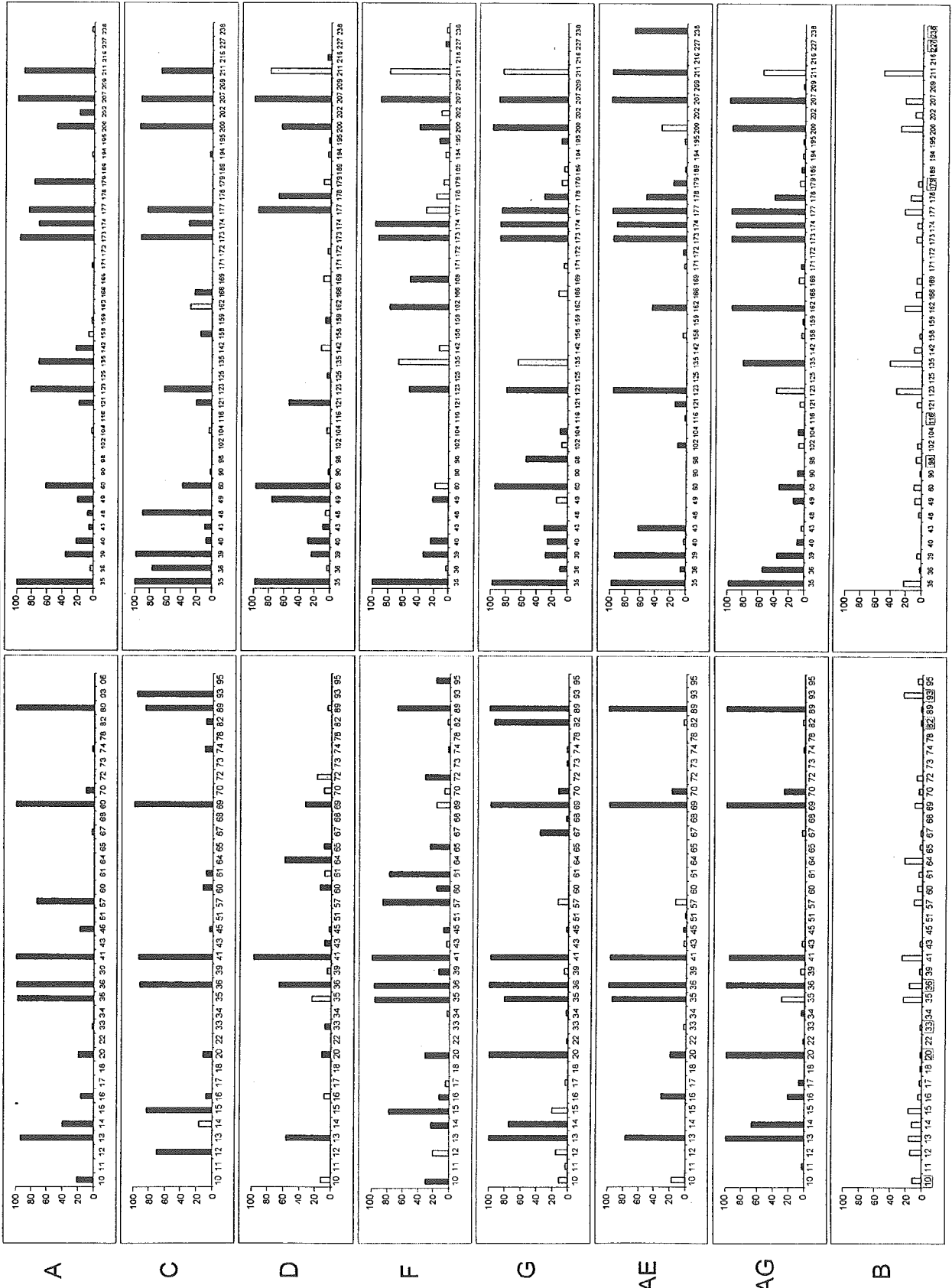
Figure 4. Subtype-Specific Polymorphisms

Positions in protease (left) and RT (right) at which mutation frequency varied significantly between subtype B and at least one non-B subtype in untreated persons. Positions are shown along the x-axes, and the frequency of mutation for each subtype is shown along the y-axes. Positions related to drug resistance in subtype B are boxed. Bar colors denote statistical significance: black is statistically significant ($Z_{01} \geq 3$); gray is borderline significant ($1 \leq Z_{01} < 3$); white is not statistically significant ($Z_{01} < 1$).

DOI: 10.1371/journal.pmed.0020112.g004

Reverse Transcriptase

Polymorphisms



Results

HIV-1 Subtypes

Sequences were obtained from 3,686 persons, in 56 countries, infected with non-B HIV-1 subtypes (Figure 2; Table 1), including 1,997 persons from whom both protease and RT sequences were available, 908 persons from whom only protease sequences were available, and 933 persons from whom only RT sequences were available. Sequences from untreated individuals were isolated between 1983 and 2003. Sequences from treated individuals were isolated between 1993 and 2003. A total of 2,311 (82%) protease and 2,381 (83%) RT sequences were obtained from plasma samples. The remaining sequences were obtained from peripheral blood mononuclear cells.

Antiretroviral Treatments

Of the participants with non-subtype-B viruses, 1,533 (42%) were receiving antiretroviral drugs at the time of sequencing: 1,140 were receiving NRTIs, 527 PIs, and 766 NNRTIs. According to subtype, 89% to 100% had received one or more NRTIs, 22% to 76% had received one or more PIs, and 32% to 55% had received one or more NNRTIs. Among treated persons infected with subtype B viruses, 98% had received NRTIs, 66% had received PIs, and 34% had received NNRTIs (Figure 3).

Mutation Prevalence

Twenty-two (22%) protease and 87 (36%) RT positions were conserved in all subtypes regardless of the presence or absence of therapy. Twenty-four (24%) protease and 38 (16%) RT positions were conserved in untreated persons but were mutant in more than 1% of treated persons. The remaining 53 (53%) protease and 115 (48%) RT positions were polymorphic, or mutant in more than 1% of untreated persons.

To assess the impact of viral subtype and treatment on the distribution of mutations in protease and RT, a binomial response model using subtype and treatment as explanatory variables was used to predict whether a position was wild-type (matching the consensus B site) or mutant. This model identified three types of positions: (i) positions in sequences from untreated people more likely to be mutated in non-B than in B subtypes (subtype-specific polymorphisms); (ii) positions in sequences of the same subtype more likely to be mutated in treated than in untreated people (subtype-specific treatment-related positions); and (iii) positions for which the effect of treatment differed significantly between non-B and B subtypes (subtype-treatment interactions).

Subtype-specific polymorphisms. Figure 4 shows the mutation prevalence according to subtype for 37 protease and 41 RT subtype-specific polymorphisms (significant θ_1 ; see Methods). Twenty-eight of the protease and 26 of the RT subtype-

specific polymorphisms were polymorphic in untreated subtype B viruses, whereas nine of the protease and 15 of the RT were conserved in subtype B. Subtype-specific polymorphisms at conserved positions in untreated subtype B viruses were generally present in a small number of subtypes at low levels (<5%). Notable exceptions included protease positions 45 and 74 and RT positions 40, 43, 104, 195 and 238.

Six subtype-specific polymorphisms in protease (positions 10, 20, 33, 36, 82, and 93) and five in RT (98, 116, 179, 227, and 238) occurred at sites known to be associated with drug resistance in subtype B viruses. These positions (with the exception of positions 116, 227, and 238 in RT) were also polymorphic in subtype B. M184I was present in six monophyletic CRF01_AE isolates from untreated persons [42,43] and was therefore not considered to be a subtype-specific polymorphism. These six sequences also displayed G→A hypermutation [44], possibly explaining the M184I change (ATG→ATA) and further complicating the significance of this finding.

Subtype-specific polymorphisms at four protease and three RT drug-resistance positions represented the consensus sequence for at least one non-B subtype: K20I in subtypes G and CRF02_AG, M36I in subtypes A, C, D, F, G, CRF01_AE and CRF02_AG, V82I in subtype G, and I93L in subtype C for protease; and A98S in subtype G, V179I in subtype A, and K238R in CRF01_AE for RT. Each of the non-B subtypes was significantly more polymorphic than subtype B at protease positions 20, 36, and 41 and RT positions 35, 39, and 207.

Subtype-specific treatment-related mutations. Figure 5 shows the mutation prevalence according to subtype for 31 protease and 36 RT subtype-specific treatment-related positions significantly more likely to be mutant in treated than untreated persons in at least one non-B subtype (significant θ_2 ; see Methods). The protease positions included 16 of the 22 known PI resistance positions and 15 additional treatment-related positions. The RT positions included 28 of the known 33 RTI resistance positions and eight additional treatment-related positions. Although each of the known PI- and RTI-resistance positions occurred in at least one non-B subtype, three of the 22 protease positions and five of the 33 RT positions included mutations that occurred too infrequently for a significant association with treatment to be detected in our analysis.

Of the 15 treatment-related protease positions not known to be associated with drug resistance, eight were also significantly associated with treatment in subtype B viruses (positions 13, 23, 43, 45, 62, 66, 74, and 85), and two have been previously reported to be associated with treatment in subtype B viruses (positions 22 and 83) [45]. The remaining five subtype-specific treatment-related protease positions included positions 6, 15, 19, 37 and 64, which—although highly polymorphic in many subtypes—are associated with treatment in subtype C (positions 6 and 64), CRF02_AG

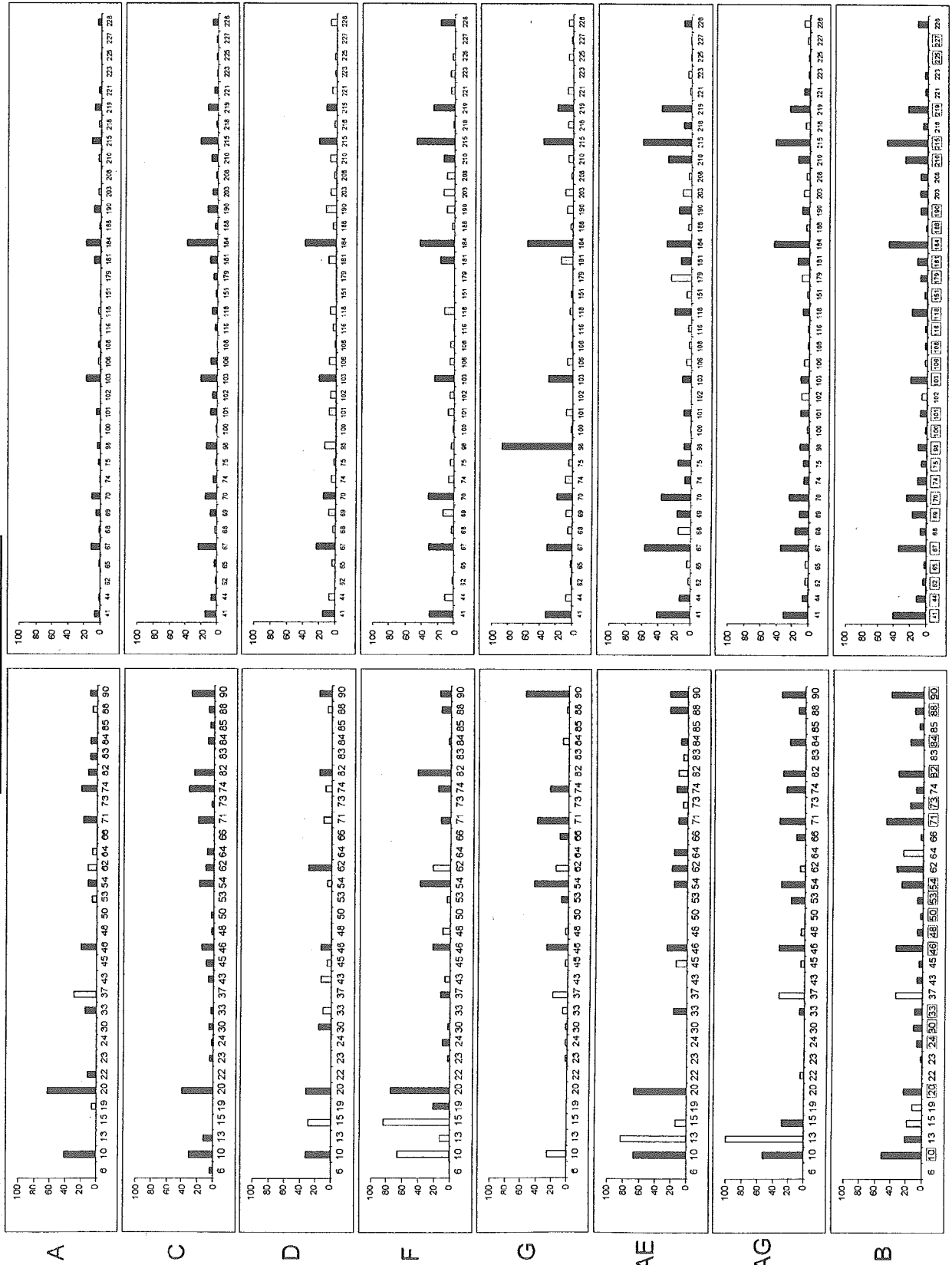
Figure 5. Subtype-Specific Treatment-Related Mutations

Positions in protease (left) and RT (right) at which mutations were significantly more prevalent in HIV-1 isolates from treated than from untreated persons infected with the same subtype. Positions are shown along the x-axes, and the proportion of mutant sequences in treated persons for each subtype is shown along the y-axes. For protease (left), treated persons are those receiving one or more PIs. For RT (right), treated persons are those receiving one or more NRTIs. Positions related to drug resistance in subtype B are boxed. Bar colors denote statistical significance: black is statistically significant ($Z_{02} \geq 3$); gray is borderline significant ($1 \leq Z_{02} < 3$); white is not statistically significant ($Z_{02} < 1$). DOI: 10.1371/journal.pmed.0020112.g005

Reverse Transcriptase

Treatment Related Mutations

Protease



(position 15), subtype F (position 19), subtype A (position 37), and CRF01_AE (position 64).

Of the eight treatment-related RT positions at sites not known to be associated with drug resistance, seven were also significantly associated with treatment in subtype B viruses (positions 68, 203, 208, 218, 221, 223, and 228) and one (position 102) was associated with treatment in subtype C but not B.

Subtype-treatment interactions. The subtype of the sequence significantly influenced the effect of treatment (significant θ_3 ; see Methods) on 20 protease positions (10, 12, 13, 14, 15, 20, 37, 53, 62, 63, 64, 65, 67, 71, 73, 74, 77, 82, 88, and 89) and 11 RT positions (35, 39, 48, 98, 104, 106, 121, 162, 166, 179, and 238). For example, RT position 98 was mutant in 7% of untreated and 16% of treated persons with subtype B viruses (approximately 2-fold difference) and in 1% of untreated and 14% of treated persons with CRF01_AG viruses (14-fold difference). Other positions less likely to be mutated in subtype B than in non-B viruses in response to treatment included protease residues 14 (subtype A); 13 and 64 (subtype C); 37 and 65 (subtype F); 71 (subtype G); 62 and 64 (CRF01_AE); and 15 and 71 (CRF02_AG); and RT residues 35 (subtype A); 98 and 106 (subtype C); 35 and 98 (subtype G); and 98 (CRF02_AG).

At other positions, treatment had a larger effect on subtype B viruses than on one or more non-B subtypes. For example, protease position 20 was mutant in 2% of untreated and 24% of treated persons with subtype B viruses (approximately 12-fold increase with treatment) and in 11% of untreated and 42% of treated persons with subtype C viruses (approximately 4-fold increase with treatment). These positions included protease residues 10, 20, and 63 (subtype A); 20, 53, 63, 74, and 82 (subtype C); 13 and 20 (subtype D); 10, 14, 20, and 77 (subtype F); 20, 67, 73, 82, and 88 (subtype G); 20, 63, 82, and 89 (CRF01_AE); and 20 (CRF02_AG); and RT residues 39 and 179 (subtype A); 35, 48, 121, and 166 (subtype C); 39 (subtype D); 39 (subtype F); 39 and 104 (subtype G); 162 and 238 (CRF01_AE); and 39 (CRF02_AG).

These 31 positions with subtype-treatment interactions included 12 known drug-resistance positions. Of these, seven protease (10, 20, 53, 63, 77, 82, and 88) and two RT (179 and 238) resistance positions were more likely to be mutated in subtype B than in one or more non-B subtypes in response to treatment. One protease position (71) and two RT positions (98 and 106) were more likely to be mutated in one or more non-B subtypes.

Known Drug-Resistance Mutations

Figure 6 shows the amino acid substitutions present at drug-resistance positions in protease and RT sequences from untreated and treated persons infected with B and non-B subtypes. Fourteen of the 22 known PI-resistance positions

occurred in subtype A, 20 in subtype C, 16 in subtype D, 20 in subtype F, 18 in subtype G, 17 in CRF01_AE, and 17 in CRF02_AG. Thirteen of the 18 known NRTI-resistance positions occurred in subtype A, 18 in subtype C, 13 in subtype D, 15 in subtype F, 18 in subtype G, 18 in CRF01_AE, and 16 in CRF02_AG. Ten of the 15 known NNRTI-resistance positions occurred in subtype A, 15 in subtype C, 11 in subtype D, 13 in subtype F, 14 in subtype G, 13 in CRF01_AE, and 12 in CRF02_AG.

In all, 106 of 113 (94%) different amino acid substitutions at 55 known subtype B drug-resistance positions (22 protease and 33 RT) were also present in at least one non-B subtype. In an exploratory analysis, which was not controlled for multiple comparisons, the frequencies of 24 mutations at 14 protease positions and 32 mutations at 19 RT positions differed between subtype B and one or more non-B subtypes.

Discussion

This collaborative analysis was designed to determine whether and to what degree the genetic mechanisms of HIV drug resistance are shared between subtype B and non-B viruses. Mutations responsible for drug resistance in subtype B viruses have been characterized by three types of studies: (i) those that identify mutations selected in viruses of persons receiving antiretroviral therapy, (ii) those that quantify the effect of specific mutations on in vitro drug susceptibility, and (iii) those that examine the effectiveness of treatment regimens in persons with viruses containing known or suspected drug-resistance mutations. This study, which identifies mutations arising in non-B viruses during antiretroviral therapy, is a necessary step for designing laboratory and clinical studies of potential drug-resistance mutations.

Do the known subtype B drug-resistance mutations also occur in non-B subtypes? We found that each of the 55 known subtype B drug-resistance mutations occurred in at least one non-B isolate. Of these, 44 (80%) were significantly associated with drug therapy in non-B isolates. The remaining 11 mutations were uncommon in subtype B and all non-B subtypes, making it difficult to determine whether they were also significantly associated with therapy. Phenotypic susceptibility testing of non-B viruses with treatment-selected mutations is necessary to confirm and quantify the contribution of these mutations to drug resistance in the genetic context in which they arise.

Do non-subtype B viruses from persons with virologic failure develop novel mutations? Fifteen protease and eight RT positions not generally considered to be drug-resistance positions were significantly associated with treatment in at least one non-B subtype. However, mutations at 17 of these 23 positions were also associated with treatment in subtype B viruses. Therefore, of the 67 mutations associated with

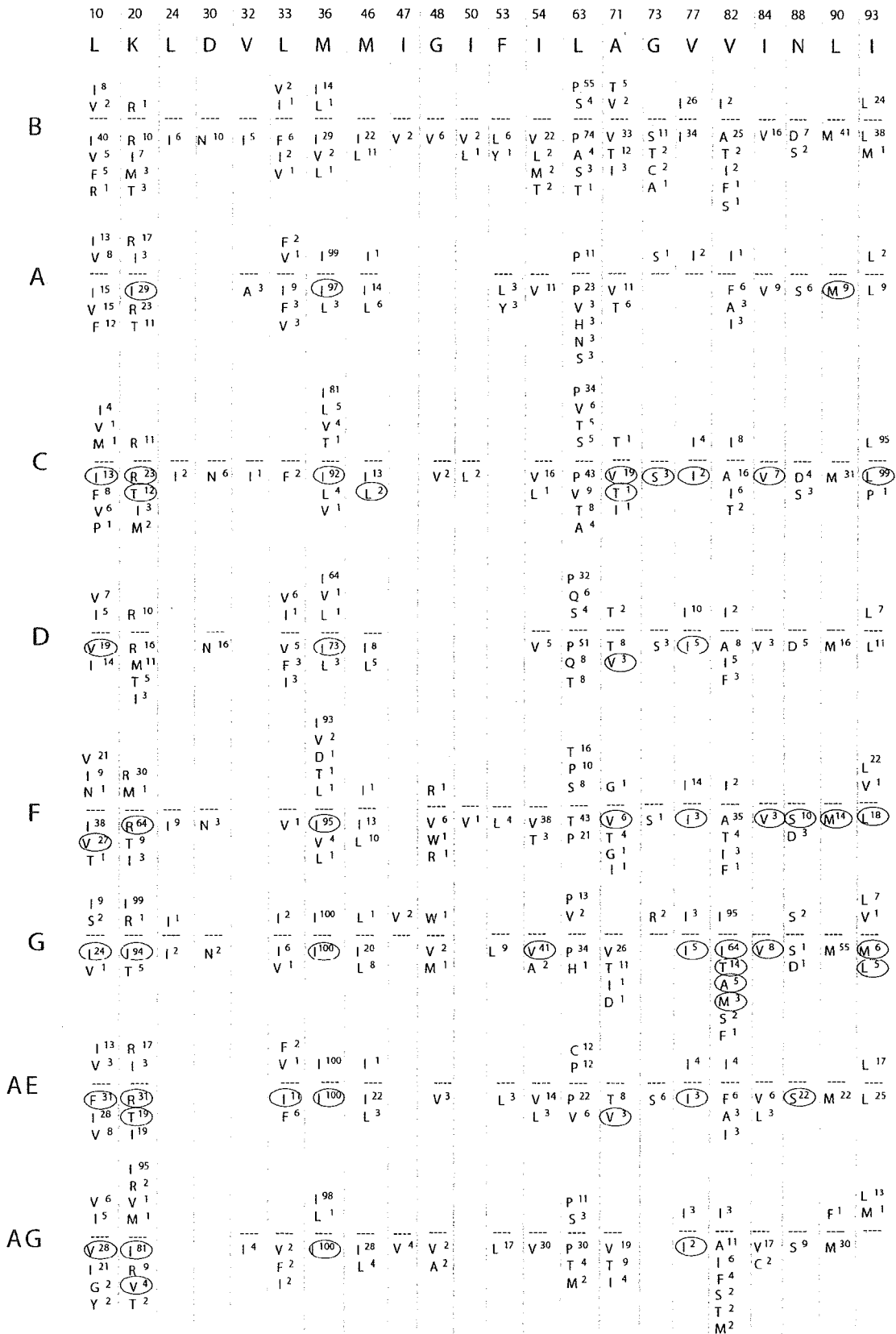
Figure 6. Amino Acid Differences from Consensus B Sequence at Drug-Resistance Positions in Protease and RT according to Subtype

(A) shows data for protease, and (B) shows data for RT. In both, the first line lists the drug-resistance positions. The second line shows single-letter amino acid codes for the consensus B sequence. For each subtype (left column), the percentage of specific mutations in untreated persons is shown above the dashed line, whereas the percentage of specific mutations in treated persons is shown below the dashed line. Positions with significant differences in mutation frequency between B and non-B subtypes ($p < 0.01$, according to χ^2 test with Yate's correction) are circled. A pound sign indicates an insertion.

DOI: 10.1371/journal.pmed.0020112.g006

a

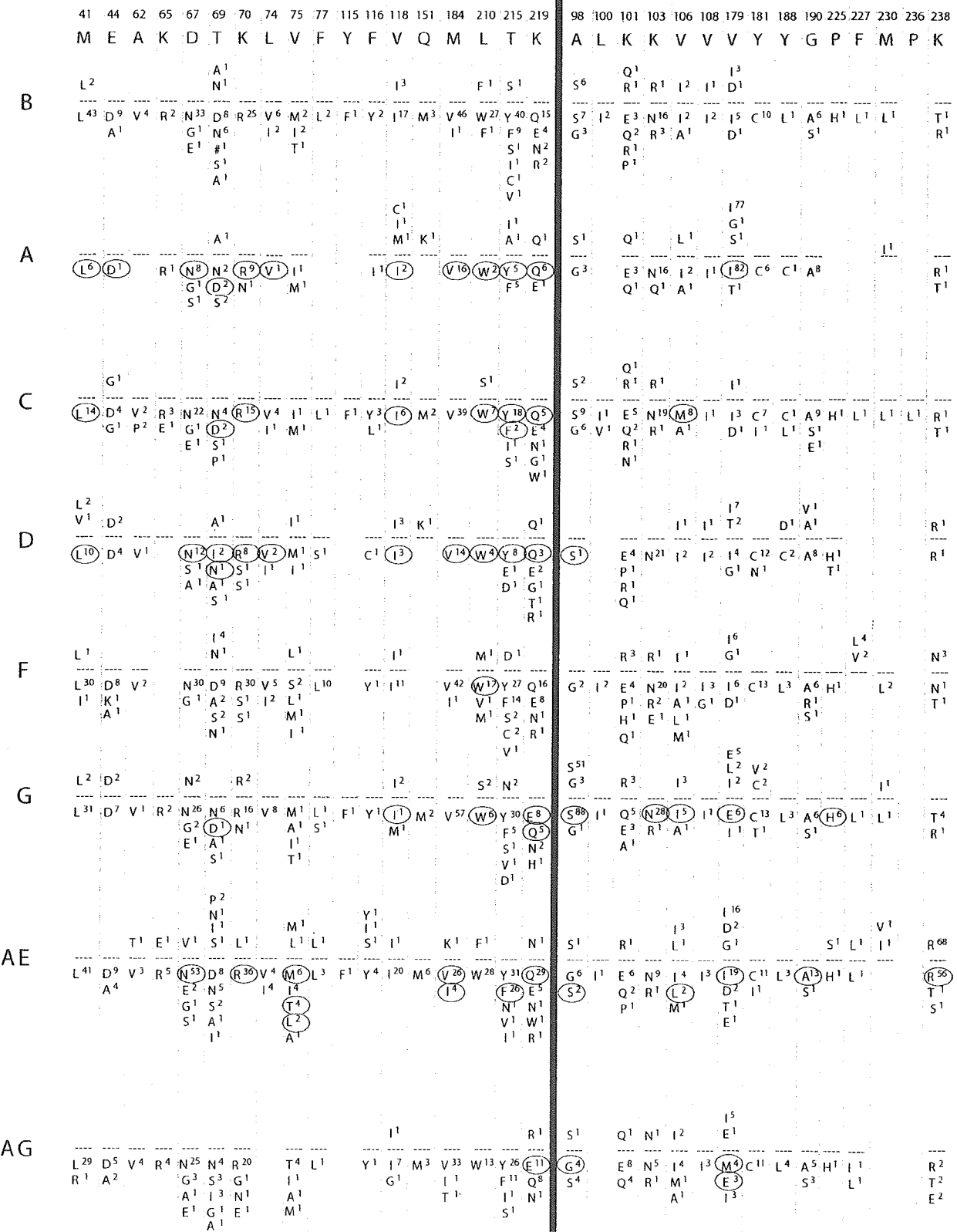
PI-related Mutations



b

NRTI-related Mutations

NNRTI-related Mutations



treatment in at least one non-B subtype, 61 were also associated with treatment in subtype B. For the six mutations associated with therapy in at least one non-B subtype but not in subtype B, the associations were at the borderline of significance and require confirmation.

Among untreated persons, non-B subtype-specific polymorphisms occurred at 37 protease and 41 RT positions. Most of these non-B polymorphic positions are also polymorphic in subtype B viruses, and several act as accessory drug-resistance mutations in subtype B viruses. Phenotypic susceptibility testing of non-B viruses with such polymorphic accessory mutations is needed to confirm that these naturally occurring viruses are fully susceptible to current antiretrovirals—a supposition that appears to be true based on the excellent virologic responses of non-B viruses to antiretroviral treatment in observational studies.

We made two simplifications in this study to increase the statistical power of our analyses. These will become unnecessary in future analyses as sufficient numbers of sequences from persons with well characterized treatment histories become available. First, we did not distinguish between different substitutions at the same position; all differences from consensus B were considered mutations. Second, viruses were classified only by the classes of drugs to which they were exposed rather than by individual drugs or drug regimens. Therefore, our analyses could not detect differences between subtype B and other subtypes that depend on specific mutations or specific drugs. Indeed, two such differences have been reported: (i) V106M is the most common substitution at RT position 106 in subtype C viruses whereas V106A predominates in subtype B viruses [33,35,46], and (ii) although the protease mutations D30N and L90M both develop in non-B viruses during nelfinavir therapy, D30N occurs more commonly in subtype B, whereas L90M occurs more commonly in subtypes C, G, and CRF01_AE [17,34,47,48].

Although the clinical samples in this study were originally obtained for a variety of purposes, including clinical management, the sequences of these samples represent experiments of nature that reveal the mutations associated with continued HIV-1 replication in the presence of selective antiretroviral therapy. The accurate identification of treatment-related mutations in such a cross-sectional analysis is challenging, however, because misclassification can result from the transmission of drug-resistant viruses, differences in specific HIV-1 variants among different human populations (population stratification), and the many statistical comparisons required as a result of HIV-1 sequence variability.

The transmission of drug-resistant HIV-1 viruses weakens cross-sectional analyses because some untreated persons may have been infected with viruses already containing treatment-related mutations. To mitigate this effect, we excluded isolates from untreated persons containing two or more non-polymorphic known drug-resistance mutations, because this pattern is not consistent with natural sequence variation. However, as noted in the Methods, an analysis that included these isolates did not alter any of the significant findings in the study. Conversely, the possibility that resistance mutations transmitted between persons in our dataset inflated the amount of resistance among persons receiving treatment was mitigated by excluding any isolate differing from another isolate at less than 1% of its nucleotides.

HIV-1 evolution is driven by genetic drift, immunologic pressure, and selective drug pressure. Population stratification can be a confounding factor when viral lineages with different founder mutations (resulting from drift or immunologic pressure) are exposed to different degrees of antiretroviral selection pressure. To distinguish mutations developing in multiple individuals as a result of selective drug pressure from mutations originating in a fewer number of founder viruses, we reconstructed the ancestral sequences at each node of a phylogenetic tree for each subtype and counted the number of times each mutation was predicted to have developed within that subtype. Because of the limited ability of phylogenetic methods to estimate accurate trees for large numbers of related sequences (i.e., belonging to the same HIV-1 subtype), only those positions for which the majority of mutations ($\geq 75\%$) appeared to result from new mutations were considered to be selected by antiretroviral therapy.

Because this analysis was, to our knowledge, the first to simultaneously examine all protease and most polymerase-coding RT positions in multiple subtypes, and because multiple associations between mutation and treatment were expected, we used a relatively lenient correction for multiple comparisons in order to minimize the number of missed associations. Nonetheless, of the 67 positive associations detected in this study, 61 were also present in persons with subtype B viruses and have previously been reported [45,49].

In conclusion, most of the protease and RT positions associated with drug resistance in subtype B viruses are selected by antiretroviral therapy in one or more non-B subtypes as well. Conversely, we found no evidence that non-B viruses develop resistance by mutations at positions that are not associated with resistance in subtype B viruses. Based on currently available data, global surveillance efforts and genotypic assessments of drug resistance should focus primarily on the known subtype B drug-resistance mutations.

Supporting Information

Dataset S1. List of GenBank Accession Numbers for Non-Subtype-B Sequences Used in This Study

Found at DOI: 10.1371/journal.pmed.0020112.sd001 (59 KB PDF).

Accession Numbers

The GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>) isolates and accession numbers for the reference subtype specimens discussed in this paper are U455/subtype A (M62320), HXB2/subtype B (K03455), C2220/subtype C (U46016), NDK/subtype D (M27323), 93BR020/subtype F (AF005494), SE6165/subtype G (AF061642), 90CR056/subtype H (AF005496), SE9173c/subtype J (AF082394), 97EQ1B11C/subtype K (AJ249235), CM240/CRF01_AE (U54771), IbNG/CRF02_AG (L39106), YBF30/Group N (AJ006022), and AN170C/Group O (L20587). The accession numbers for the non-subtype-B sequences used in this study are listed in Dataset S1.

Acknowledgments

D. Katzenstein is the recipient of a Doris Duke Distinguished Clinical Scientist award, and the support of the Doris Duke Charitable Foundation for R. Kantor and the project is gratefully acknowledged. R. Shafer was supported by grants from the National Institutes of Allergy and Infectious Diseases (A146148-01) and National Institute of General Medical Sciences (5P01CM066524-02).