



# Within-host co-evolution of Gag P453L and protease D30N/N88D demonstrates virological advantage in a highly protease inhibitor-exposed HIV-1 case

Junko Shibata<sup>a,b,1</sup>, Wataru Sugiura<sup>b,c,d,\*</sup>, Hirotaka Ode<sup>e</sup>, Yasumasa Iwatani<sup>b,c,d</sup>, Hironori Sato<sup>e</sup>, Hsinyi Tsang<sup>b</sup>, Masakazu Matsuda<sup>d,f</sup>, Naoki Hasegawa<sup>a</sup>, Fengrong Ren<sup>a</sup>, Hiroshi Tanaka<sup>a</sup>

<sup>a</sup> School of Biomedical Sciences, Tokyo Medical and Dental University, Tokyo, Japan

<sup>b</sup> Clinical Research Center, National Hospital Organization, Nagoya Medical Center, Nagoya, Japan

<sup>c</sup> Nagoya University Graduate School of Medicine, Nagoya, Japan

<sup>d</sup> AIDS Research Center, National Institute of Infectious Diseases, Tokyo, Japan

<sup>e</sup> Pathogen Genomics Center, National Institute of Infectious Diseases, Tokyo, Japan

<sup>f</sup> Mitsubishi Chemical Medience Corporation, Tokyo, Japan

## ARTICLE INFO

### Article history:

Received 7 September 2010

Received in revised form

28 December 2010

Accepted 11 February 2011

Available online 19 February 2011

### Keywords:

HIV

Protease

Gag

Drug resistance

Co-evolution

## ABSTRACT

To better understand the mechanism of HIV group-specific antigen (Gag) and protease (PR) co-evolution in drug-resistance acquisition, we analyzed a drug-resistance case by both bioinformatics and virological methods. We especially considered the quality of sequence data and analytical accuracy by introducing single-genome sequencing (SGS) and Spidermonkey/Bayesian graphical models (BGM) analysis, respectively. We analyzed 129 HIV-1 Gag–PR linkage sequences obtained from 8 time points, and the resulting sequences were applied to the Spidermonkey co-evolution analysis program, which identified ten mutation pairs as significantly co-evolving. Among these, we focused on associations between Gag-P453L, the P5' position of the p1/p6 cleavage-site mutation, and PR-D30N/N88D nelfinavir-resistant mutations, and attempted to clarify their virological significance *in vitro* by constructing recombinant clones. The results showed that P453L<sup>Gag</sup> has the potential to improve replication capacity and the Gag processing efficiency of viruses with D30N<sup>PR</sup>/N88D<sup>PR</sup> but has little effect on nelfinavir susceptibility. Homology modeling analysis suggested that hydrogen bonds between the 30th PR residue and the R452<sup>Gag</sup> are disturbed by the D30N<sup>PR</sup> mutation, but the impaired interaction is compensated by P453L<sup>Gag</sup> generating new hydrophobic interactions. Furthermore, database analysis indicated that the P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup> association was not specific only to our clinical case, but was common among AIDS patients.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

Major mutations in the human immunodeficiency virus-1 (HIV-1) protease (PR)-coding region selected by protease inhibitors (PIs) are mainly located within the active sites of the PR, and these mutations significantly reduce PR activity and viral replication capacity, i.e., viral fitness, compared to that of wild-type strains (Mahalingam et al., 1999). However, PI-resistant viruses have the potential to undergo further selection and evolution to recover their impaired PR activity and viral fitness by acquiring additional mutations not only in the PR but also in its natural substrate, the Gag protein

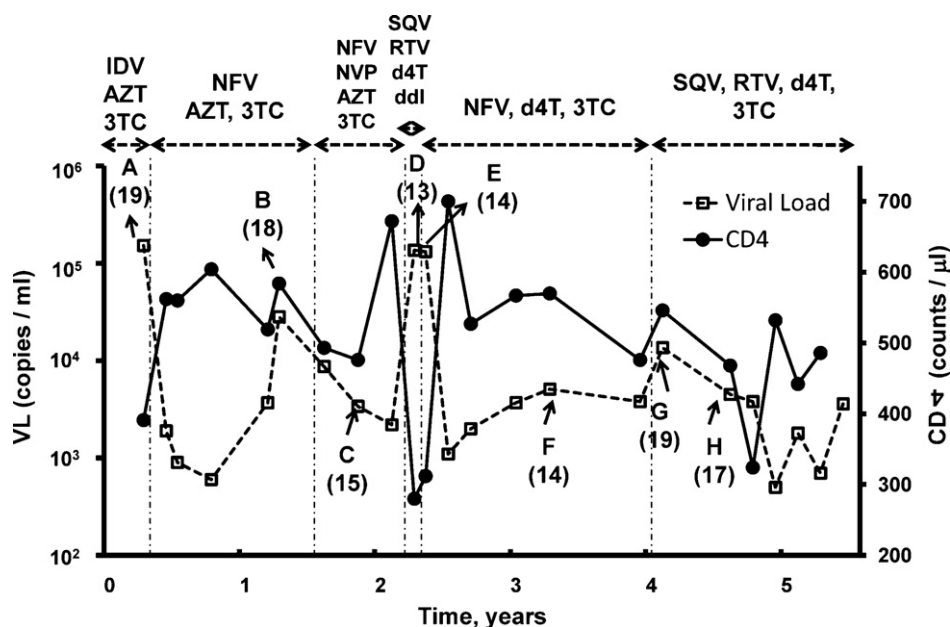
(Nijhuis et al., 1999). In particular, mutations in Gag cleavage sites can improve the replication capacity of PI-resistant viruses. Indeed, tight associations have been demonstrated between PI-resistant mutations and Gag cleavage-site mutations, such as S373Q<sup>Gag</sup> (Malet et al., 2007) and I376V<sup>Gag</sup> (Ho et al., 2008) at p2/NC, A431V<sup>Gag</sup> (Bally et al., 2000; Doyon et al., 1996; Gallego et al., 2003; Koch et al., 2001; Maguire et al., 2002; Malet et al., 2007; Verheyen et al., 2006; Zhang et al., 1997) at NC/p1, L449F<sup>Gag</sup> (Doyon et al., 1996), and P453L<sup>Gag</sup> (Bally et al., 2000; Maguire et al., 2002; Verheyen et al., 2006) at p1/p6. Gag mutations at non-cleavage sites have also been reported to improve the fitness of PI-resistant viruses (Myint et al., 2004). Thus, the selection and evolution of Gag and PR are accepted to significantly interfere with each other. This phenomenon is known as “Gag–PR co-evolution.”

However, previous reports of Gag–PR co-evolution appear to have two technical limitations related to data quality and analytical method. First, the standard population-based genotyping method commonly used for determining HIV-1 variants has limited accuracy for technological reasons. To sequence the viral genome,

\* Corresponding author at: Clinical Research Center, National Hospital Organization, Nagoya Medical Center, 4-1-1, San-no-maru, Naka-ku, Nagoya, Aichi 4600001, Japan. Tel.: +81 52 951 1111; fax: +81 52 963 3970.

E-mail address: [wsugiura@nih.go.jp](mailto:wsugiura@nih.go.jp) (W. Sugiura).

<sup>1</sup> Current address: Inserm U941, IUH, Université Paris Diderot, Hôpital Saint Louis, 75475 Paris Cedex 10, France.



**Fig. 1.** Clinical course of patient-treatment protocols. Open squares and solid circles indicate plasma viral load and CD4+ cell count, respectively. Each treatment regimen is indicated in the upper part of the graph by horizontal dashed lines between arrows. Sample collection points are labelled A to H, and numbers of sequences analyzed at each point are shown in parentheses. AZT, zidovudine; d4T, stavudine; ddI, didanosine; 3TC, lamivudine; NVP, nevirapine; NFV, nelfinavir; RTV, ritonavir; SQV, saquinavir; IDV, indinavir.

viral RNA from patient plasma must be reverse-transcribed to cDNA, and the target regions are then amplified by PCR. Since viral cDNA includes multiple viral populations, using such samples as a template for the PCR step followed by sequencing results in amplification of predominantly the major variants (Gunthard et al., 1998; Hance et al., 2001), which may not represent the diversity of the original population. Furthermore, there is always a risk of artificial recombination of viral cDNAs, which are estimated to occur at rates of 4–70% during the PCR step (Meyerhans et al., 1990) and may disturb the linkage of mutation sites. Thus, this artificial linkage information may significantly affect the results of co-evolution analyses. To circumvent this technical problem, we employed the single-genome sequencing (SGS) technique based on limiting-dilution assays (Palmer et al., 2005). With SGS, artificial recombination can be avoided during PCR because cDNA samples are diluted after the reverse transcriptase reaction to a single cDNA molecule, which is used as the PCR template. Therefore, the viral sequence information obtained by SGS is not only more sensitive for analyzing HIV population diversity and mutation linkages than that obtained by standard genotype analysis, but also more precise for identifying co-evolving sites.

Second, many analytical methods have been developed for detecting co-evolving sites in molecular sequences. However, early methods did not accommodate the evolutionary history among sequences, and risked generating false-positive predictions (Altschuh et al., 1987; Gutell et al., 1992; Martin et al., 2005; Neher, 1994; Pollock and Taylor, 1997; Tillier and Lui, 2003). In recent years, the accuracy of estimating co-evolving sites has greatly improved due to several computational algorithms that incorporate the phylogenetic relationships among molecular data (Bhattacharya et al., 2007; Dutheil et al., 2005; Poon et al., 2008; Tuff and Darlu, 2000; Wollenberg and Atchley, 2000; Yeang and Haussler, 2007). In particular, Spidermonkey analysis (Poon et al., 2007a,b, 2008) not only considered the phylogenetic relationships but also implemented two empirical HIV-1 subtype B amino acid-substitution models for describing between- and within-host HIV evolution (Nickle et al., 2007).

In this study, we clarified the impact of Gag and PR co-evolution in the acquisition of drug resistance by inferring co-evolving sites

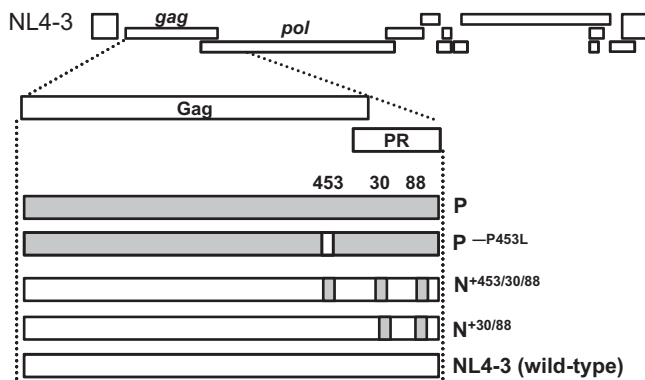
between Gag and PR in a dataset of HIV-1 Gag–PR linkage sequences from a single patient who had undergone highly active antiretroviral therapy (HAART) for a long period. We used the SGS method for sequencing viral samples to avoid artificial recombination, and applied Spidermonkey analysis to our data by using its option for the within-host HIV substitution model. Virological significance of the estimated co-evolving sites was analyzed *in vitro* by constructing recombinant viruses on a pNL4-3 backbone. The replication kinetics, susceptibility to anti-HIV drugs, and the Gag processing efficiency were evaluated for each clone. We also conducted homology modeling analysis to examine Gag–PR interactions and database analysis to confirm the universality of the co-evolving sites.

## 2. Materials and methods

### 2.1. Sample collection and gag-PR-coding region sequencing by single-genome sequencing

From all HIV/AIDS patients monitored from April 1998 to August 2002 at the National Institute of Infectious Diseases in Japan, we selected a virological failure case with a history of multiple antiretroviral treatments. Of all cases, this one had been followed for the longest period and had enough detailed clinical information to perform our analysis. For the selected case, we collected plasma samples and clinical information, such as treatment regimens, changes in viral load, and CD4 counts (Fig. 1).

The single-genome sequencing (SGS) method was used as described (Palmer et al., 2005). Briefly, HIV-1 RNA was extracted from plasma samples (containing a minimum of 1000 copies of HIV-1 RNA) by guanidinium isothiocyanate treatment. cDNA was synthesized using Superscript II RT kit (Invitrogen, Carlsbad, CA) and a random hexamer. cDNA was serially diluted and amplified by PCR and nested PCR using Platinum Taq DNA Polymerase High Fidelity (Invitrogen). The endpoint of reverse-transcribed cDNA was determined as a single clone by the Poisson distribution, with cDNA dilutions yielding PCR products in 3 out of 10 reactions. The following primer sets were used in the first and nested PCR amplifications: first PCR-WGPF1 (5'-CTCTCTCGACGAGGACTCG-



**Fig. 2.** Construction of recombinant viruses. Four different types of recombinant viruses were constructed according to their mutation patterns: (1) P, NL4-3 backbone with patient-derived Gag and protease (PR); (2) P-453L, NL4-3 backbone with patient-derived Gag and PR but P453L<sup>Gag</sup> was converted to wild-type; (3) N<sup>+453/30/88</sup>, NL4-3 backbone including the substitutions P453L<sup>Gag</sup>, D30N<sup>PR</sup>, and N88D<sup>PR</sup>; (4) N<sup>+30/88</sup>, NL4-3 backbone including the substitutions D30N<sup>PR</sup> and N88D<sup>PR</sup>.

3') and 3500- (5'-CTATTAAGTATTTGATGGGTCATAA-3'), nested PCR-WGPF2 (5'-TTGCTGAAGCGCGCACGGCAAGA-3') and 3410- (5'-CAGTTAGTGGTATTACTTCTGTTAGTGCTT-3'). These primer sets amplified a 2.7 kbps fragment containing *gag*, the PR-coding region of *pol*, and the first 900 nucleotides of the reverse transcriptase-coding region of *pol*. Positive PCR products were determined by agarose gel electrophoresis and sequenced using ABI Prism BigDye Terminator version 3.1 dideoxycycle sequencing (Applied Biosystems, Carlsbad, CA). Sequences, including the entire *gag* gene (1500 bp) and PR-coding region (297 bp), were aligned on the HXB2 reference using Sequence Navigator software (Applied Biosystems). Sequences containing mixtures at any position were excluded from analysis. PI-resistant mutations and other PR mutations in our data set were determined using the Calibrated Population Resistance tool Version 4.3 beta (<http://cpr.stanford.edu/cpr/>).

## 2.2. Detecting co-evolution using Spidermonkey analysis

Co-evolving sites between Gag and PR were inferred using Spidermonkey analysis (Poon et al., 2007a,b, 2008) to analyze all sequences (data type = protein), including the entire Gag (500 amino acids) and PR (99 amino acids) sequences. A phylogenetic tree was first reconstructed using the neighbor-joining method (Saitou and Nei, 1987). To infer co-evolving sites, we selected the within-host HIV protein-substitution model (frequencies: model-defined), based on the maximum likelihood method (Nickle et al., 2007). Amino acid sequences at MA/CA, CA/p2, p2/NC, NC/p1, and p1/p6 Gag cleavage-sites (residues P5 to P5') and the complete PR sequence were selected for BGM analysis. We used the default options for calculations and inferred co-evolving sites if the estimated posterior probability for a pair of positions exceeded the cutoff value of 0.5.

## 2.3. Construction of recombinant HIV with a patient-derived gag-PR-coding region

To confirm the contributions of P453L<sup>Gag</sup> in the context of patient-derived gag-PR-coding region, three types of recombinant viruses were constructed (Fig. 2). Recombinant viruses were constructed using the pNL4-3 molecular clone of HIV-1 (GenBank accession No. AF324493). A patient-derived gag-PR-coding region containing P453L<sup>Gag</sup>, D30N<sup>PR</sup>, and N88D<sup>PR</sup> was chosen from sampling point F. The recombinants were (1) P, NL4-3 backbone with patient-derived gag-PR-coding region; (2) P-453L, NL4-3 backbone

with patient-derived gag-PR-coding region but P453L<sup>Gag</sup> was converted to wild-type.

Details of the recombinant virus construction were as follows. The patient-derived 1.9 kbp gag-PR-coding region was subcloned into pCR4-TOPO (Invitrogen) (designated pCR4-TOPO<sup>Patient</sup>). Using pCR4-TOPO<sup>Patient</sup> as a template, P453L<sup>Gag</sup> was substituted with wild-type P453. Each substitution was introduced using the following primer sets: For the Gag-P453 substitution, forward: 5'-TTCAGAACACACAGAGCCATCAGCT-3', reverse: 5'-AGCTGATGGCTCTGGTCTGTTCTGAA-3'. Subsequently, the substituted gag-PR-coding region was amplified using the following primers: WGPF2: 5'-TTGCTGAAGCGCGCACGGCAAGA-3' and DRPRO6: 5'-ACTTTTGGGCCATCCATTCTGGCTTT-3'. The NL4-3-derived RT-coding region was amplified using the following primers: RT-63F: 5'-TAAACAATGGCCATTGACAGAAG-3' and RT-898R: 5'-CTGCTTCTCTGTTAGTGGTACTAC-3'. Primers DRPRO6 and RT-63F are phosphorylated at their 5' ends. Both resulting fragments were ligated and digested with BssHII and SbfI, and cloned back into pNL4-3. All pNL4-3-based recombinant DNAs (3.75 µg) were transfected into 2 × 10<sup>5</sup> HeLa cells using Eugene6 (Roche, Indianapolis, IN), and culture supernatants were harvested at 72 h after transfection, filtered through a 0.45 µm membrane, assayed for reverse transcriptase (RT) activity (Willey et al., 1988), and kept as virus stocks at -80 °C until use. Each virus stock (5 × 10<sup>6</sup> <sup>32</sup>P cpm of RT activity) was used for replication kinetics analyses.

## 2.4. Construction of pNL43 with P453L<sup>Gag</sup>, D30N<sup>PR</sup>, and N88D<sup>PR</sup> by site-directed mutagenesis

In addition to constructing patient-derived gag-PR-coding region recombinant viruses, we evaluated the effect of P453L<sup>Gag</sup>, D30N<sup>PR</sup>, and N88D<sup>PR</sup> interference by constructing three types of pNL4-3-based recombinant viruses (Fig. 2). These were (1) N<sup>+453/30/88</sup>, NL4-3 including P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup>; and (2) N<sup>+30/88</sup>, NL4-3 including D30N<sup>PR</sup>/N88D<sup>PR</sup>. These recombinant viruses were constructed as follows. pCR4-TOPO (Invitrogen) including the Apal-SbfI fragment of pNL4-3, which contained the complete p1/p6 of gag and PR-coding region, was constructed as a template for further mutagenesis. This construct was designated as pCR4-TOPO/NL<sup>Apal-SbfI</sup>. Using pCR4-TOPO/NL<sup>Apal-SbfI</sup> as a template, we introduced three mutations (P453L<sup>Gag</sup>, D30N<sup>PR</sup>, and N88D<sup>PR</sup>) by site-directed mutagenesis. The following primer sets were used to introduce each mutation: For P453L<sup>Gag</sup>, forward: 5'-TTCAGAGCAGACTAGAGCCAACAGCC-3' and reverse: 5'-GGCTGTGGCTCTAGTCTGCTCTGAA-3'. For D30N<sup>PR</sup>, forward: 5'-CAGGAGCAGATAATACAGTATTAGAAG-3' and reverse: 5'-CTTCTAATACTGTATTATCTGCTCTG-3'. For N88D<sup>PR</sup>, forward: 5'-CATAATTGGAAGAGATCTGTGACTC-3' and reverse: 5'-GAGTCAACAGATCTCTTCCAATTATG-3'. After each mutagenesis reaction, the entire sequences were verified and cloned back into pNL4-3 using Apal and SbfI restriction enzymes. All pNL4-3-based recombinant DNAs (3.75 µg) were transfected into 2 × 10<sup>5</sup> HeLa cells using Eugene6 (Roche), and culture supernatants were harvested at 72 h after transfection, filtered through a 0.45 µm membrane, assayed for RT activity (Willey et al., 1988), and kept as virus stocks at -80 °C until use. Each virus stock (4 × 10<sup>5</sup> <sup>32</sup>P cpm of RT activity) was used for replication kinetics analysis.

## 2.5. Replication kinetics of recombinant viruses

Replication kinetics of recombinant viruses was evaluated as described (Matsuoka-Aizawa et al., 2003). Briefly, 5 × 10<sup>4</sup> MT-2 cells were infected with each virus stock in the absence or presence of 0.1 µM nelfinavir at 37 °C for 16 h. Cells were then washed once and resuspended in 0.5 ml culture medium with the same concen-



tration of nelfinavir, and cultures were maintained for 12–22 days, changing half of the medium every 2 or 3 days. The titer of each virus was evaluated by both RT activity and p24 amount; as both measures demonstrated good correlation, p24 amount was used to adjust the virus inoculum. Culture supernatants were collected, and residual supernatants were kept at  $-80^{\circ}\text{C}$  until use. Replication kinetics was independently analyzed two times.

## 2.6. Evaluation of nelfinavir susceptibilities of recombinant viruses

Recombinant viruses were evaluated for nelfinavir susceptibility using an in-house drug susceptibility assay with the MaRBLE cell line (MaRBLE assay) as described (Chiba-Mizutani et al., 2007). Briefly,  $1 \times 10^5$  MaRBLE cells were infected with 100 CCID<sub>50</sub> of each recombinant virus, and virus replication was monitored in serial dilutions of nelfinavir from  $1.28 \times 10^{-13}$  to  $1.0 \times 10^{-6}$  M for 7 days in triplicate. At day 7, cells were harvested and lysed in luciferase assay reagent. Firefly (FF) and renilla luciferase (RF) activities produced by the cells were quantified using a Dual-luciferase Reporter Assay System (Promega, Madison, WI). The relative virus replication rate (% replication) at each drug concentration was calculated by the following formula: % replication = (observed FF luciferase activity with drug – background [mock] FF luciferase activity) / (observed FF luciferase activity without drug – background [mock] FF luciferase activity)  $\times 100$ . IC<sub>50</sub> values were calculated with 95% confidence intervals using GraphPad Prism software and nonlinear regression analysis fitted with a sigmoidal dose–response curve with variable slope.

## 2.7. Analysis of Gag processing patterns in recombinant viruses

The Gag processing patterns of recombinant viruses were analyzed as described (Sugiura et al., 2002) with minor modifications. In brief, NL4-3-based recombinant DNAs (18  $\mu\text{g}$ ) was transfected into  $5 \times 10^6$  HeLa cells using Fugene6 (Roche). The culture supernatants were harvested at 48 h after transfection with a culture medium change at 12 h post transfection to the absence or presence of 0.1  $\mu\text{M}$  nelfinavir. For pelleting virus through a sucrose cushion, 25 ml of cell culture medium was layered onto 10 ml of 20% sucrose (wt/vol, in PBS) before centrifugation at 30,000 rpm in a swing rotor for 1.5 h. The medium and cushion were discarded, and the virus pellet was dissolved in 200  $\mu\text{l}$  Laemmli sample buffer (Bio-Rad, Hercules, CA). Viral supernatants were normalized using PETRO-TEK HIV-1 p24 Antigen ELISA (ZeptoMetrix Corporation, Buffalo, NY) and subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE). Proteins in the gels were passively transferred to PVDF membranes (Bio-Rad). The membranes were incubated with a chicken anti-p6 polyclonal antibody (Sigma–Aldrich Corporation, St. Louis, MO) and a murine anti-p24 monoclonal antibody (ZeptoMetrix Corporation) for 2 h, followed by incubation with secondary antibody, a chicken HRP-conjugated, IgG antibody (Bethyl Laboratories Inc., Montgomery, TX), a mouse HRP-conjugated, IgG antibody (Thermo Fisher Scientific, Yokohama, JP), respectively. Finally, proteins were visualized using SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific).

## 2.8. Molecular modeling of Gag p1/p6 peptide–PR complexes

We constructed three-dimensional models of PR in complex with peptide representing Gag-p1/p6 substrate by a homology modeling method (Baker and Sali, 2001; Marti-Renom et al., 2000; Shirakawa et al., 2008) using Molecular Operating Environment (MOE) ver. 2008.10 (<http://www.chemcomp.com/>, Chemical Computing Group Inc., Montreal, Quebec, Canada). For the modeling

template, we used an X-ray crystal structure of inactive D25N PR in complex with the p1/p6 substrate at 2 Å resolution (PDB code: 1KJF) (Prabu-Jeyabalan et al., 2002), as it had the highest similarity (94.2% identity) to the HIV-1 NL4-3 strain among HIV-1 protease-p1/p6 peptide structures in the protein data bank, even though this is an inactive model. Furthermore, the active-site D25N<sup>PR</sup> mutation has been reported to hardly influence the structure of the protease in complex with ligands (Sayer et al., 2008). In the models, nine-amino-acid-length peptides corresponding to the Gag p1/p6 of the NL4-3, N<sup>30/88</sup>, and N<sup>453/30/88</sup> strains were bound to the catalytic sites of PRs of the same strains. We considered effects of a water molecule (HOH11) that mediates important hydrogen bonds between the Gag p1/p6 peptide and the PR. AMBER ff99 force field (Wang et al., 2000) and generalized Born/volume integral (GB/VI) implicit solvent model (Labute, 2008) were applied for intra- and inter-molecular energy calculations.

## 2.9. Database analysis

To confirm the universality of P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup>, we obtained 3249 sequences of HIV-1 subtype B gag-PR-coding region (positions 2146–2516) from the Los Alamos National Laboratory HIV sequence database (<http://www.hiv.lanl.gov/>). For the dataset, we applied Fisher's exact test and investigated associations between the D30N/N88D mutations in PR and the P453L mutation in Gag and between the 30/88 mutations in PR and the P453L mutation in Gag, respectively. Fisher's exact test was implemented using the R 2.6.1 statistical package.

## 3. Results

### 3.1. The majority of patient-derived Gag cleavage-site mutations are in p2/NC and p1/p6

Twenty-three plasma samples were serially collected from a patient who had received HAART for 52 months. The patient's clinical history, changes in viral load, and CD4 counts are depicted in Fig. 1. At 8 sampling points (A–H), 129 gag-PR-coding region sequences were obtained, with p2/NC and p1/p6 Gag cleavage-site mutations observed at each point in more than 60% of clones (Table 1). PI-resistant mutations and other PR mutations are also summarized in Table 1. None of the clones had CA/p2 cleavage-site mutations, and only a few clones had MA/CA or NC/p1 cleavage-site mutations. The Y132F<sup>Gag</sup> mutation in MA/CA was found in 1.6% of the clones, and E428D<sup>Gag</sup> and R429K<sup>Gag</sup> within NC/p1 had prevalences of 1.6% and 5.4%, respectively. On the other hand, cleavage-site mutations were frequently observed within p2/NC and p1/p6. In the p2/NC site, we observed 11 mutations: S373Q<sup>Gag</sup> (83.7%), S373P<sup>Gag</sup> (15.5%), A374G<sup>Gag</sup> (79.8%), A374V<sup>Gag</sup> (16.3%), A374S<sup>Gag</sup> (2.3%), A374R<sup>Gag</sup> (0.8%), T375N<sup>Gag</sup> (100%), I376V<sup>Gag</sup> (3.9%), M378L<sup>Gag</sup> (0.8%), G381S<sup>Gag</sup> (0.8%), and N382Y<sup>Gag</sup> (0.8%). In the p1/p6 site, we observed three mutations: L449F<sup>Gag</sup> (0.8%), S451N<sup>Gag</sup> (100%), and P453L<sup>Gag</sup> (65.1%).

### 3.2. Ten Gag–protease co-evolving sites are inferred by Spidermonkey analysis

Among the 129 gag-PR-coding region sequences analyzed by Spidermonkey analysis (Poon et al., 2007a,b, 2008), ten co-evolving sites were inferred. These sites were identified using the default cutoff posterior probability (pp) value of 0.5. The ten co-evolving pairs identified with pp > 0.5 are shown in Table 2. Four pairs, R429K<sup>Gag</sup>/M36V<sup>PR</sup> (pp = 0.87), S373Q<sup>Gag</sup>/T12A<sup>PR</sup> (pp = 0.83), P453L<sup>Gag</sup>/D30N<sup>PR</sup> (pp = 0.63), and P453L<sup>Gag</sup>/N88D<sup>PR</sup> (pp = 0.61), represented Gag/PR inter-molecular co-evolution;

**Table 1**

Gag cleavage-site mutations and PR mutations from a HAART-treated case.

Sampling point	Gag cleavage-site mutation (%) <sup>a</sup>		PR mutation (%) <sup>a</sup>	
	p2/NC	p1/p6	PI-resistant mutations	Other PR mutations
A to H (n = 129)	T375N	S451N	–	I62V, L63P, A71T, I93L
A (n = 19)	S373Q(100), A374G (89),	–	–	M36I(95), I72V(11), V77I(100)
B (n = 18)	S373Q(100), A374G (94)	P453L(100)	D30N(100), M46I(6), N88D(100)	E35D(94), M36I(100), V77I(100),
C (n = 15)	S373Q(100), A374G(100)	P453L(100)	D30N(100), N88D(100)	L10F(80), E35D(100), M36I(100), K45R(27),
				I72T(7), V77I (100),
D (n = 13)	S373P (85), A374V (85)	P453L (8)	D30N (8), N88D (8)	L10F(8), V11I (8), T12A(8), K20R(8), E35D(8),
				M36I(100), H69Y(8), V77I(8)
E (n = 14)	S373P (64), A374V(71)	–	–	T12A(7), M36I(100), K55N (7), V77I(7)
F (n = 14)	S373Q(100), A374G(100)	P453L(100)	D30N(100), N88D(100)	L10F(100), I13V(79), E34G(7), E35D(100), M36I
				(100), N37T (86), K45R (14), Q58E (86), I72T
				(7), V77I (100)
G (n = 19)	S373Q(100), A374G(100)	P453L(100)	D30N(100), N88D(100)	L10F(100), I13V (68), E35D(100), M36I(68),
				M36V(32), N37T(100), Q58E(100), V77I(100)
H (n = 17)	S373Q(100), A374G(100)	P453L(100)	D30N(100), I54V(76), N88D(100),	L10F(100), I13V (94), K20R (76), E35D (100),
			L90M (47)	M36I (94), M36V (6), N37T (100), Q58E (100),
				I72T (53), V77I (100), G78R (6)

Gag mutations refer to HXB2 and PI-resistant mutations and other PR mutations were determined using the Calibrated Population Resistance tool Version 4.3 beta.

<sup>a</sup> Numbers in parentheses are the percentages of mutations at each sampling point.

one pair, S373Q<sup>Gag</sup>/A374G<sup>Gag</sup>, represented Gag/Gag intra-molecular co-evolution; and the other five pairs, N37T<sup>PR</sup>/Q58E<sup>PR</sup> (pp = 0.94), E35D<sup>PR</sup>/M46I<sup>PR</sup> (pp = 0.89), K20R<sup>PR</sup>/I54V<sup>PR</sup> (pp = 0.88), V11I<sup>PR</sup>/K20R<sup>PR</sup> (pp = 0.86), and D30N<sup>PR</sup>/N88D<sup>PR</sup> (pp = 0.62), represented PR/PR intra-molecular co-evolution.

We focused on three pairs, P453L<sup>Gag</sup>/N88D<sup>PR</sup>, P453L<sup>Gag</sup>/D30N<sup>PR</sup>, and D30N<sup>PR</sup>/N88D<sup>PR</sup>, because D30N<sup>PR</sup> and N88D<sup>PR</sup> are well-known major and minor nelfinavir-resistant mutations, respectively (Johnson et al., 2008), and P453L<sup>Gag</sup> is the P5' position of the p1/p6 cleavage-site mutation. Although D30N<sup>PR</sup> has been associated with N88D<sup>PR</sup> (Rhee et al., 2007; Wu et al., 2003), the interactions among P453L<sup>Gag</sup>, D30N<sup>PR</sup>, and N88D<sup>PR</sup> have not been investigated. Since P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup> was frequently observed in the presence of nelfinavir (Fig. 1 and Table 1), we conducted *in vitro* experiments to confirm whether the co-existence of P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup> has a virological advantage in the presence of nelfinavir.

### 3.3. P453L<sup>Gag</sup> improves the replication capacity of viruses with D30N<sup>PR</sup>/N88D<sup>PR</sup> in both patient- and NL4-3-derived genetic backgrounds

To evaluate the virological impact of P453L<sup>Gag</sup> in the patient-derived genetic background, we constructed two types of patient-derived gag-PR-coding region viruses, P and P<sup>-P453L</sup> (Fig. 2). These two recombinant viruses and the wild-type virus (NL4-3) were cultured independently in the absence or presence of nelfinavir, and their replication kinetics was monitored by measuring RT activity in culture supernatants. Assays for replication kinetics were independently performed twice, confirming identical orders of replication kinetics.

**Table 2**

Positions of coevolving pairs inferred by Spidermonkey analysis.

Position	Position	Expected posterior probability	Total number of sequences
429 <sup>Gag</sup>	36 <sup>PR</sup>	0.866	7
373 <sup>Gag</sup>	12 <sup>PR</sup>	0.825	2
453 <sup>Gag</sup>	30 <sup>PR</sup>	0.632	84
453 <sup>Gag</sup>	88 <sup>PR</sup>	0.607	84
373 <sup>Gag</sup>	374 <sup>Gag</sup>	0.895	22
37 <sup>PR</sup>	58 <sup>PR</sup>	0.938	48
35 <sup>PR</sup>	46 <sup>PR</sup>	0.887	1
20 <sup>PR</sup>	54 <sup>PR</sup>	0.877	13
11 <sup>PR</sup>	20 <sup>PR</sup>	0.858	1
30 <sup>PR</sup>	88 <sup>PR</sup>	0.618	84

Significant coevolving sites (pp value &gt; 0.5) are shown.

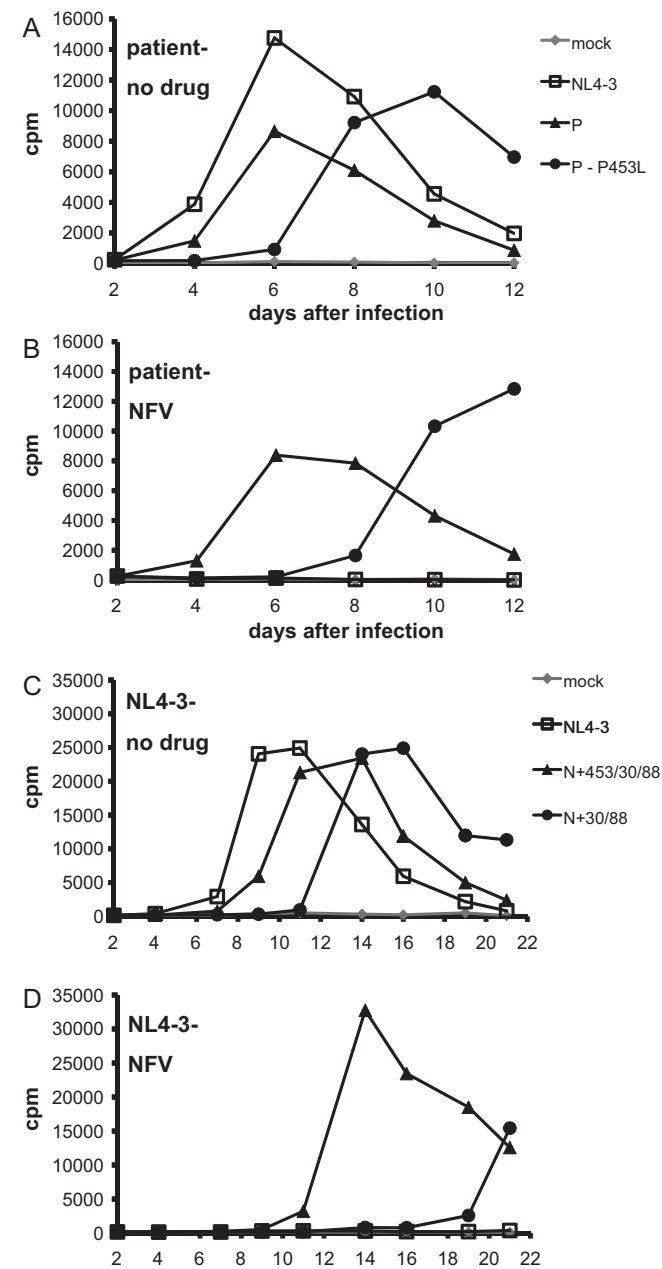
In the absence of nelfinavir (Fig. 3A), the RT activities of the NL4-3 and P viruses peaked at 6 days after infection, and the RT activity of NL4-3 was higher than that of the P virus, whereas viral replication was delayed in the P<sup>-P453L</sup> virus (i.e., P virus without P453L<sup>Gag</sup>), and its RT activity peaked at 10 days after infection. The order of replication kinetics in the absence of nelfinavir was wild-type (NL4-3) > P > P<sup>-P453L</sup>. On the other hand, in the presence of nelfinavir (Fig. 3B), replication of the wild-type virus was completely suppressed, and replication of P virus was the most active, demonstrating peak RT activity at 6 days after infection. The P<sup>-P453L</sup> virus showed lower replication capacity than the P virus. Thus, the order of replication kinetics in the presence of nelfinavir was P > P<sup>-P453L</sup> > wild-type (NL4-3).

To assess whether the complex P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup> conferred an advantage not only in the patient-derived genetic background but also in the HIV-1 molecular clone (NL4-3)-derived genetic background, we constructed two types of NL4-3-based gag-PR-coding region recombinant viruses, N<sup>+453/30/88</sup> and N<sup>+30/88</sup> (Fig. 2). The results of independent culture studies are shown in Fig. 3C and D. In the absence of nelfinavir, the RT activities of NL4-3, NL4-3 with P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup> (N<sup>+453/30/88</sup>) and NL4-3 with D30N<sup>PR</sup>/N88D<sup>PR</sup> (N<sup>+30/88</sup>) viruses peaked at 11 days, 14 days, and 16 days after infection, respectively, and the order of replication kinetics was NL4-3 > N<sup>+453/30/88</sup> > N<sup>+30/88</sup>. On the other hand, in the presence of nelfinavir, the N<sup>+453/30/88</sup> virus grew the most actively with its peak RT activity at 14 days after infection. The N<sup>+30/88</sup> virus was the second most actively replicating, and the wild-type did not replicate. Thus, the order of replication kinetics in the presence of nelfinavir was N<sup>+453/30/88</sup> > N<sup>+30/88</sup> > NL4-3.

In both studies, not only in the patient-derived genetic background but also in the NL4-3-derived genetic background, the virus with D30N<sup>PR</sup>/N88D<sup>PR</sup> showed lower replication capacity than the virus with P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup>, suggesting that P453L<sup>Gag</sup> significantly contributes to the fitness recovery of virus with D30N<sup>PR</sup>/N88D<sup>PR</sup>.

### 3.4. P453L<sup>Gag</sup> does not influence susceptibility to nelfinavir

To clarify whether P453L<sup>Gag</sup> affects nelfinavir susceptibility, the IC<sub>50</sub>s of nelfinavir for NL4-3-based gag-PR-coding region recombinants were determined using MaRBLE cells. Both N<sup>+453/30/88</sup> and N<sup>+30/88</sup> recombinants showed 14.4 and 15.8-fold greater resistance to nelfinavir than wild-type virus, respectively (Table 3). However, the difference in IC<sub>50</sub> between the two recombinants was



**Fig. 3.** Replication kinetics of recombinants. MT-2 cells were infected with patient-derived gag-PR-coding region, (A) in the absence of, and (B) in the presence of 0.1  $\mu$ M nelfinavir. Open squares, solid triangles, and solid circles indicate wild-type NL4-3, NL4-3 with patient gag-PR-coding region insert, and NL4-3 with patient insert without P453L substitution, respectively. MT-2 cells were also infected with NL4-3-based recombinant, (C) in the absence of, and (D) in the presence of 0.1  $\mu$ M nelfinavir. Open squares, solid triangles, and solid circles indicate wild type NL4-3; NL4-3 with P453L, D30N, and N88D; and NL4-3 with D30N and N88D, respectively. Diamonds indicate mock-infected controls. Assays were independently performed twice, and one representative set of results is shown.

not significant, suggesting that P453L<sup>Gag</sup> had no effect on nelfinavir susceptibility.

### 3.5. P453L<sup>Gag</sup> improves Gag p1/p6 processing in virus with D30N<sup>PR</sup>/N88D<sup>PR</sup>

To gain further insights into the virological effects of P453L<sup>Gag</sup>, we examined Gag processing patterns in the absence and presence of 0.1  $\mu$ M nelfinavir by Western blot analysis with an anti-p6 polyclonal antibody and an anti-p24 monoclonal antibody. The amount

**Table 3**  
Nelfinavir susceptibilities of recombinant viruses.

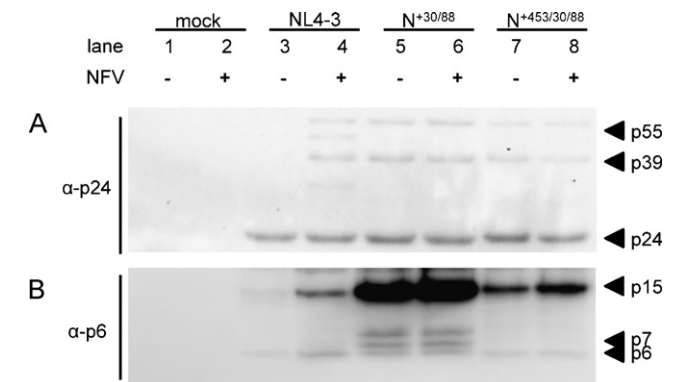
HIV-1	IC <sub>50</sub> (nM)	95% confidence interval	Fold-resistance
NL4-3	1.3	0.8–2.4	1.0
NL <sup>+453/30/88</sup>	18.7	9.2–37.8	14.4
N <sup>+30/88</sup>	20.6	12.6–33.6	15.8

of sample loaded in each lane was normalized by p24 antigen content (600 ng for each lane as determined by ELISA) (Fig. 4A). In the absence of nelfinavir, the partially cleaved Gag intermediate p15 (including NC, p1, and p6) of NL4-3 was efficiently cleaved to the p6 peptide (Fig. 4B, lane 3). In contrast, the processing of NL4-3 was less efficient in the presence of nelfinavir, as indicated by the accumulation of p15 (Fig. 4B, lane 4). N<sup>+30/88</sup> showed defects in cleavage at the p1/p6 site, as demonstrated by the accumulation of p15 and p7 (p1/p6) (Fig. 4B, lanes 5 and 6). On the other hand, lesser p15 and p7 accumulated in N<sup>+453/30/88</sup> than in N<sup>+30/88</sup> (Fig. 4, lanes 7 and 8). Interestingly, a 8–9 kDa band, which is neither p6 nor p7, was observed in N<sup>+30/88</sup> (Fig. 4B, lanes 5 and 6).

### 3.6. Impaired Gag–PR affinity in the N<sup>+30/88</sup> strain is recovered by new L453<sup>Gag</sup> interactions with M46<sup>PR</sup> and F53<sup>PR</sup>

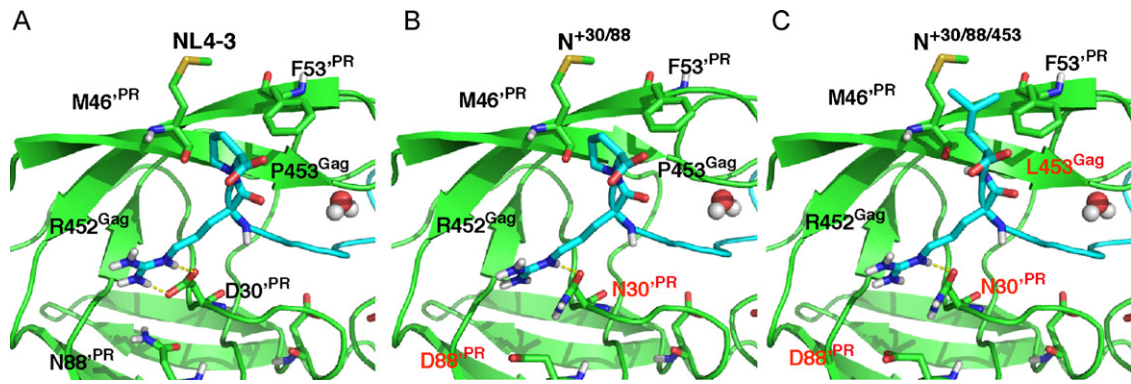
To elucidate the structural impact of the mutations described above on interactions between Gag-p1/p6 substrate and PR, we generated three-dimensional models of the PR in complex with peptide representing Gag-p1/p6 substrate by homology modeling (Baker and Sali, 2001; Marti-Renom et al., 2000; Shirakawa et al., 2008) using Gag and PR sequences of the NL4-3, N<sup>+30/88</sup>, and N<sup>+453/30/88</sup> strains. Comparison of the thermodynamically optimized models showed obvious differences in interactions between side chains of PR and p1/p6 (Fig. 5). First, D30N<sup>PR</sup> mutation resulted in fewer hydrophilic interactions between side chains of the 30th PR and 452nd p1/p6 residues; NL4-3 had two hydrogen bonds between the side chains of D30<sup>PR</sup> and R452<sup>Gag</sup> (Fig. 5A), while N<sup>+30/88</sup> and N<sup>+453/30/88</sup> had only a single hydrogen bond between the side chains of N30<sup>PR</sup> and R452<sup>Gag</sup> (Fig. 5B and C). Second, the P453L<sup>Gag</sup> mutation in the p1/p6 substrate of the N<sup>+453/30/88</sup> strain led to new hydrophobic interactions between side chains of the 46th PR and 453rd p1/p6 residues, as well as of the PR 53F and p1/p6 P453L<sup>Gag</sup> residues (Fig. 5C).

To examine whether these changes in interactions influenced the binding affinity of the p1/p6 substrate to PR, we analyzed



**Fig. 4.** Western blot analysis of Gag processing in the absence or presence of nelfinavir. Western blot analyses in the absence or presence of nelfinavir. HeLa cells were transfected by each recombinant clones and cultured in the absence or presence of NFV (0.1  $\mu$ M). At 48 h post-transfection, virions in culture supernatants were harvested and subjected to Western blot analysis with anti-p24 monoclonal antibody (A) and an anti-p6 polyclonal antibody (B). Each lane was normalized by p24 antigen content (5 ng for each lane as determined by ELISA). Lanes 1 and 2, mock; lanes 3 and 4, NL4-3; lanes 5 and 6, N<sup>+453/30/88</sup>; lanes 7 and 8, N<sup>+30/88</sup>.





**Fig. 5.** Structural models of the Gag–PR complexes. Catalytic sites of the NL4-3 (A), N<sup>+30/88</sup> models (B), and N<sup>+453/30/88</sup> models (C) are highlighted. Green cartoons and sticks represent main and side chains of PR, respectively. Cyan sticks represent 452nd and 453rd residues in Gag corresponding to the p1/p6 region. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

their binding energies using the PR–p1/p6 peptide complex models. The predicted binding energies of the NL4-3, N<sup>+30/88</sup> and N<sup>+453/30/88</sup> models were  $-137.6$  kcal/mol,  $-133.3$  kcal/mol, and  $-137.1$  kcal/mol, respectively, suggesting that p1/p6 substrate has lower affinity with N<sup>+30/88</sup> PR than with NL4-3 and N<sup>+453/30/88</sup> PRs. Taken together, these data suggest that PR mutations in the N<sup>+30/88</sup> strain can reduce Gag–PR affinity primarily via loss of the hydrogen bond between N30<sup>PR</sup> and R452<sup>Gag</sup> and that the Gag p1/p6 mutation in the N<sup>+453/30/88</sup> strain (P453L<sup>Gag</sup>) can recover affinity by generating new hydrophobic interactions of L453<sup>Gag</sup> with M46<sup>PR</sup> and F53<sup>PR</sup>.

### 3.7. P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup> association is commonly observed in a large database

To confirm the prevalence of the P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup> association in other HIV-infected individuals, we investigated 3249 sequences of HIV-1 subtype B gag–PR-coding region from the Los Alamos National Laboratory HIV sequence database (<http://www.hiv.lanl.gov/>). We found that the P453L<sup>Gag</sup> mutation was significantly associated with D30N<sup>PR</sup>/N88D<sup>PR</sup> (Table 4;  $p < 0.001$ , Fisher's exact test). These data support the virological advantage of the P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup> association.

## 4. Discussion

In this study, we analyzed mechanisms of anti-HIV drug-resistant mutation acquisition by investigating crosstalk between Gag and PR mutations. We traced the clinical course and sequence changes in gag and the PR-coding region of a virological-failure case heavily treated with multiple regimens, including different protease inhibitors. To focus on the quality of sequence data and accuracy of analysis, we used SGS and Spidermonkey analysis, respectively.

Among the ten co-evolving Gag–Protease pairs inferred by Spidermonkey analysis, we confirmed a linkage between P453L<sup>Gag</sup> and D30N<sup>PR</sup>/N88D<sup>PR</sup>. D30N<sup>PR</sup> has been reported to associate with N88D<sup>PR</sup> (Rhee et al., 2007; Wu et al., 2003), P453L (Verheyen

et al., 2006), and positively correlate with p1/p6 cleavage-site mutations (Kolli et al., 2006). We were interested in the association between P453L<sup>Gag</sup> and D30N<sup>PR</sup>/N88D<sup>PR</sup> nelfinavir-resistant mutations, as P453L<sup>Gag</sup> is located at the P5' position of the p1/p6 cleavage site, and expected to physically interact with the protease. Thus, we sought to clarify the biological advantage of interference among P453L<sup>Gag</sup>, D30N<sup>PR</sup>, and N88D<sup>PR</sup> in recombinant viruses with patient- and NL4-3-derived genetic backgrounds. Virological advantage was evaluated in three aspects: (1) virological superiority in replication competency, (2) resistance to antiretroviral selective pressure, and (3) Gag processing pattern in virions. Our results indicated that the P453L<sup>Gag</sup> cleavage-site mutation has the potential to improve the replication capacity and Gag processing of viruses with D30N<sup>PR</sup>/N88D<sup>PR</sup>, but has little effect on nelfinavir susceptibility. This latter finding is of interest since Gag cleavage-site mutations have been suggested as a mechanism for protease to develop drug resistance (Dam et al., 2009; Kolli et al., 2009). We also need to consider not only antiretroviral selective pressure but also immune selective pressure. Several of the positions we noted have been described as associated with human leukocyte antigen (HLA) escape mutations. For example, PR codons 12, 35, and 36, and Gag codons 373 and 374 are all potentially HLA-related (Brumme et al., 2009). To confirm the contributions of HLA and immune pressure, further study is required.

Although samples were collected chronologically at multiple times, the order of P453L<sup>Gag</sup>, D30N<sup>PR</sup> and N88D<sup>PR</sup> acquisitions was unclear as all three mutations were detected at the same time. However, a plausible order seems to be the selection of D30N<sup>PR</sup>/N88D<sup>PR</sup> followed by P453L<sup>Gag</sup> acquisition. Nelfinavir appears to select D30N<sup>PR</sup>/N88D<sup>PR</sup> mutations for resistance because these mutations obviously increase drug resistance to nelfinavir (Johnson et al., 2009), whereas the P453L<sup>Gag</sup> mutation without any PR mutations has been reported to have almost no effect on susceptibility to PIs and on viral replication capacity (Maguire et al., 2002). Although D30N<sup>PR</sup> is known as one of the most unstable PI-resistant mutations (Martinez-Picado et al., 1999) and viruses with this mutation have lower PR activity than the wild-type, the impaired replication caused by D30N<sup>PR</sup> has been reported to be compensated by N88D<sup>PR</sup> (Mitsuya et al., 2006; Sugiura et al., 2002). Furthermore, the fitness of virus with D30N<sup>PR</sup>/N88D<sup>PR</sup> was recovered in our study by an additional P453L<sup>Gag</sup> mutation (Fig. 3). However, the P453L<sup>Gag</sup> mutation was not introduced into NL4-3 carrying D30N<sup>PR</sup>/N88D<sup>PR</sup> (N<sup>+30/88</sup>) and patient D30N<sup>PR</sup>/N88D<sup>PR</sup> clones (P<sup>–</sup>P453L) during *in vitro* culture with nelfinavir, suggesting that the P453L<sup>Gag</sup> mutation is a sufficient condition for D30N<sup>PR</sup>/N88D<sup>PR</sup> clones to replicate efficiently.

Virus with D30N<sup>PR</sup>/N88D<sup>PR</sup> was suggested by results of our Western blot analyses to process p1/p6 cleavage inefficiently, as

**Table 4**  
Los Alamos National Laboratory HIV sequence database analysis.

453 <sup>Gag</sup>	30 <sup>PR</sup> /88 <sup>PR</sup>	
	D30/N88	N30N/D88D
P (n = 2801)	n = 2743	n = 8
L (n = 237)	n = 223	n = 8
p-value	<0.001	

demonstrated by the accumulation of p15 and p7 non-cleaved precursors, but addition of P453L<sup>Gag</sup> improved the processing (Fig. 4B). Accumulation of p15 and p7 products has been reported in previous studies using different PR mutant viruses (Doyon et al., 1996; Maguire et al., 2002). In these studies, the additional mutations at NC/p1 or p1/p6 cleavage sites also resulted in efficient processing of these precursors. Interestingly, an aberrant band, which did not match either p6 or p7, was observed in N<sup>30/88</sup> in our study (Fig. 4, lanes 5 and 6). Although further analyses will be required to determine the exact mechanisms, the band suggests inaccurate or alternative recognition of the cleavage site by virus with D30N<sup>PR</sup>/N88D<sup>PR</sup>, and P453L<sup>Gag</sup> may confer an advantage by adjusting the protease to recognize and cleave the right site. As the HIV-1 p6 protein is important for efficient particle budding (von Schwedler et al., 2003), the defect in p1/p6 cleavage may affect viral maturation, which in turn may reduce viral infectivity and replication capacity.

To understand the relevance of P453L<sup>Gag</sup> from a structural viewpoint, we used homology modeling with the published X-ray crystal structure of the PR–p1/p6 substrate complex as a template (Fig. 5). Although P453L<sup>Gag</sup> is located at the P5' position and does not directly interfere with the protease active site or subsites, the modeling demonstrated that P453L<sup>Gag</sup> can compensate for the binding affinity of PR and p1/p6. This mechanism is interesting because it suggests that a mutation outside the cleavage site interferes with the PR–Gag interaction. Indeed, Prabu-Jeyabalan et al. (2004) documented that a Gag mutation (A431V<sup>Gag</sup>) compensates for a PR mutation (V82A<sup>PR</sup>), which is not in direct contact with A431V<sup>Gag</sup>. Thus, our data confirm the virological and structural advantages of P453L<sup>Gag</sup> in viruses possessing D30N<sup>PR</sup>/N88D<sup>PR</sup>. Furthermore, this association appeared to be quite common as the frequency of P453L<sup>Gag</sup> is 7.3% in the Los Alamos National Laboratory HIV sequence database. Though nelfinavir is no longer recommended as a first-line antiretroviral in the guidelines of developed countries, many cases previously exposed to nelfinavir have acquired D30N/N88D mutations. Indeed, the prevalence of the D30N mutation in PI-treated persons infected with subtype B viruses ( $n = 7396$ ) and in nelfinavir-treated persons ( $n = 1128$ ) is 7.9% and 28.1%, respectively, in the Stanford HIV drug resistance database (<http://hivdb.stanford.edu/>).

Regarding the bioinformatics analysis strategy, we selected the within-host substitution model in Spidermonkey analysis to infer the co-evolving sites (Nickle et al., 2007) as the data were sequences serially collected over 5 years from a single patient under anti-HIV treatment. One disadvantage of this program is that it does not account for the number of descendant clones. Often, mutation pairs on few viral clones might be determined as co-evolving pairs. In our study, 46<sup>PR</sup> and 35<sup>PR</sup> mutations were determined as a co-evolving pair with high posterior probability, but only one clone with this pair was observed among 129 sequences (Table 2), suggesting this co-mutation pair could not become “fixed” in a viral population. Thus, it is important to confirm the significance of the program output.

## 5. Conclusions

In conclusion, we successfully determined the Gag–protease associated sites P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup> by applying single-genome sequencing, suggesting the usefulness of this method. However, as SGS is a more expensive method than direct sequencing, researchers need to consider which method has the best advantage for their samples. Extrapolating from our data, the relationships between the major mutations found by SGS may not differ significantly from direct sequencing results, but we might have a greater chance of seeing a variety of minority clones with minor mutations. In addition, our observation of higher variation at later

sampling points suggests that cases with longer treatment histories are good sample candidates.

We found that the viruses acquiring P453L<sup>Gag</sup>/D30N<sup>PR</sup>/N88D<sup>PR</sup> distinctly showed biological advantages. From results obtained using both viral experiments and bioinformatics, we speculate that the P453L<sup>Gag</sup> mutation does not necessarily occur in the presence of nelfinavir, but if it occurs with D30N<sup>PR</sup>/N88D<sup>PR</sup> mutations, viral fitness can be improved, which may eventually lead to worse clinical outcomes of anti-HIV therapy. We believe that the findings of this study provide new insight into the mechanism of within-patient HIV-1 co-evolution and into the acquisition of resistance to anti-HIV drugs.

## Acknowledgements

The authors thank Dr. Akira Shirahata and Mr. Yuki Kitamura for their support. We thank patients who contributed to our study. We also thank Ms. Claire Baldwin for her help in preparing the manuscript. This study was supported by a Grant-in-Aid for AIDS research from the Ministry of Health, Labor and Welfare of Japan (H19-AIDS-007), and also by Scientific Research from the Ministry of Education, Culture, Sports, and Technology of Japan (Project number: 19510208).

## References

- Altschuh, D., Lesk, A.M., Bloomer, A.C., Klug, A., 1987. Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus. *J. Mol. Biol.* 193, 693–707.
- Baker, D., Sali, A., 2001. Protein structure prediction and structural genomics. *Science* 294, 93–96.
- Bally, F., Martinez, R., Peters, S., Sudre, P., Telenti, A., 2000. Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: clinical significance and implications for resistance to protease inhibitors. *AIDS Res. Hum. Retroviruses* 16, 1209–1213.
- Bhattacharya, T., Daniels, M., Heckerman, D., Foley, B., Frahm, N., Kadie, C., Carlson, J., Yusim, K., McMahon, B., Gaschen, B., Mallal, S., Mullins, J.L., Nickle, D.C., Herbeck, J., Rousseau, C., Learn, G.H., Miura, T., Brander, C., Walker, B., Korber, B., 2007. Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science* 315, 1583–1586.
- Brumme, Z.L., John, M., Carlson, J.M., Brumme, C.J., Chan, D., Brockman, M.A., Swenson, L.C., Tao, I., Szeto, S., Rosato, P., Sela, J., Kadie, C.M., Frahm, N., Brander, C., Haas, D.W., Riddler, S.A., Haubrich, R., Walker, B.D., Harrigan, P.R., Heckerman, D., Mallal, S., 2009. HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *PLoS One* 4, e6687.
- Chiba-Mizutani, T., Miura, H., Matsuda, M., Matsuda, Z., Yokomaku, Y., Miyauchi, K., Nishizawa, M., Yamamoto, N., Sugiura, W., 2007. Use of new T-cell-based cell lines expressing two luciferase reporters for accurately evaluating susceptibility to anti-human immunodeficiency virus type 1 drugs. *J. Clin. Microbiol.* 45, 477–487.
- Dam, E., Quercia, R., Glass, B., Descamps, D., Launay, O., Duval, X., Krausslich, H.G., Hance, A.J., Clavel, F., 2009. Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. *PLoS Pathog.* 5, e1000345.
- Doyon, L., Croteau, G., Thibeault, D., Poulin, F., Pilote, L., Lamarre, D., 1996. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J. Virol.* 70, 3763–3769.
- Dutheil, J., Pupko, T., Jean-Marie, A., Galtier, N., 2005. A model-based approach for detecting coevolving positions in a molecule. *Mol. Biol. Evol.* 22, 1919–1928.
- Gallego, O., de Mendoza, C., Corral, A., Soriano, V., 2003. Changes in the human immunodeficiency virus p7–p1–p6 gag gene in drug-naïve and pretreated patients. *J. Clin. Microbiol.* 41, 1245–1247.
- Gunthard, H.F., Wong, J.K., Ignacio, C.C., Havlir, D.V., Richman, D.D., 1998. Comparative performance of high-density oligonucleotide sequencing and dideoxynucleotide sequencing of HIV type 1 pol from clinical samples. *AIDS Res. Hum. Retroviruses* 14, 869–876.
- Gutell, R.R., Power, A., Hertz, G.Z., Putz, E.J., Stormo, G.D., 1992. Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Res.* 20, 5785–5795.
- Hance, A.J., Lemiale, V., Izopet, J., Lecossier, D., Joly, V., Massip, P., Mammano, F., Descamps, D., Brun-Vézinet, F., Clavel, F., 2001. Changes in human immunodeficiency virus type 1 populations after treatment interruption in patients failing antiretroviral therapy. *J. Virol.* 75, 6410–6417.
- Ho, S.K., Coman, R.M., Bunger, J.C., Rose, S.L., O'Brien, P., Munoz, I., Dunn, B.M., Sleasman, J.W., Goodenow, M.M., 2008. Drug-associated changes in amino acid residues in Gag p2, p7(NC), and p6(Gag)/p6(Pol) in human immunodeficiency virus type 1 (HIV-1) display a dominant effect on replicative fitness and drug response. *Virology*.



- Johnson, V.A., Brun-Vézinet, F., Clotet, B., Gunthard, H.F., Kuritzkes, D.R., Pillay, D., Schapiro, J.M., Richman, D.D., 2008. Update of the drug resistance mutations in HIV-1: Spring 2008. *Top. HIV Med.* 16, 62–68.
- Johnson, V.A., Brun-Vézinet, F., Clotet, B., Gunthard, H.F., Kuritzkes, D.R., Pillay, D., Schapiro, J.M., Richman, D.D., 2009. Update of the drug resistance mutations in HIV-1: December 2009. *Top. HIV Med.* 17, 138–145.
- Koch, N., Yahi, N., Fantini, J., Tamalet, C., 2001. Mutations in HIV-1 gag cleavage sites and their association with protease mutations. *AIDS* 15, 526–528.
- Kolli, M., Lastere, S., Schiffer, C.A., 2006. Co-evolution of nelfinavir-resistant HIV-1 protease and the p1–p6 substrate. *Virology* 347, 405–409.
- Kolli, M., Stawiski, E., Chappey, C., Schiffer, C.A., 2009. Human immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance inhibitor resistance. *J. Virol.* 83, 11027–11042.
- Labute, P., 2008. The generalized Born/volume integral implicit solvent model: estimation of the free energy of hydration using London dispersion instead of atomic surface area. *J. Comput. Chem.* 29, 1693–1698.
- Maguire, M.F., Guinea, R., Griffin, P., Macmanus, S., Elston, R.C., Wolfram, J., Richards, N., Hanlon, M.H., Porter, D.J., Wrin, T., Parkin, N., Tisdale, M., Furfine, E., Petropoulos, C., Snowden, B.W., Kleim, J.P., 2002. Changes in human immunodeficiency virus type 1 Gag at positions 1449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro. *J. Virol.* 76, 7398–7406.
- Mahalingam, B., Louis, J.M., Reed, C.C., Adomat, J.M., Krouse, J., Wang, Y.F., Harrison, R.W., Weber, I.T., 1999. Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. *Eur. J. Biochem.* 263, 238–245.
- Malet, I., Roquebert, B., Dalban, C., Wirten, M., Amellal, B., Agher, R., Simon, A., Katlama, C., Costagliola, D., Calvez, V., Marcelin, A.G., 2007. Association of Gag cleavage sites to protease mutations and to virological response in HIV-1 treated patients. *J. Infect.* 54, 367–374.
- Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., Sali, A., 2000. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291–325.
- Martin, L.C., Gloer, G.B., Dunn, S.D., Wahl, L.M., 2005. Using information theory to search for co-evolving residues in proteins. *Bioinformatics* 21, 4116–4124.
- Martinez-Picado, J., Savara, A.V., Sutton, L., D'Aquila, R.T., 1999. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *J. Virol.* 73, 3744–3752.
- Matsuoka-Aizawa, S., Sato, H., Hachiya, A., Tsuchiya, K., Takebe, Y., Gatanaga, H., Kimura, S., Oka, S., 2003. Isolation and molecular characterization of a nelfinavir (NFV)-resistant human immunodeficiency virus type 1 that exhibits NFV-dependent enhancement of replication. *J. Virol.* 77, 318–327.
- Meyerhans, A., Vartanian, J.P., Wain-Hobson, S., 1990. DNA recombination during PCR. *Nucleic Acids Res.* 18, 1687–1691.
- Mitsuya, Y., Winters, M.A., Fessel, W.J., Rhee, S.Y., Hurley, L., Horberg, M., Schiffer, C.A., Zolopa, A.R., Shafer, R.W., 2006. N88D facilitates the co-occurrence of D30N and L90M and the development of multidrug resistance in HIV type 1 protease following nelfinavir treatment failure. *AIDS Res. Hum. Retroviruses* 22, 1300–1305.
- Myint, L., Matsuda, M., Matsuda, Z., Yokomaku, Y., Chiba, T., Okano, A., Yamada, K., Sugiura, W., 2004. Gag non-cleavage site mutations contribute to full recovery of viral fitness in protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* 48, 444–452.
- Neher, E., 1994. How frequent are correlated changes in families of protein sequences? *Proc. Natl. Acad. Sci. U. S. A.* 91, 98–102.
- Nickle, D.C., Heath, L., Jensen, M.A., Gilbert, P.B., Mullins, J.I., Kosakovsky Pond, S.L., 2007. HIV-specific probabilistic models of protein evolution. *PLoS One* 2, e503.
- Nijhuis, M., Schuurman, R., de Jong, D., Erickson, J., Gustchina, E., Albert, J., Schipper, P., Gulnik, S., Boucher, C.A., 1999. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS* 13, 2349–2359.
- Palmer, S., Kearney, M., Maldarelli, F., Halvas, E.K., Bixby, C.J., Bazmi, H., Rock, D., Falloon, J., Davey Jr., R.T., Dewar, R.L., Metcalf, J.A., Hammer, S., Mellors, J.W., Coffin, J.M., 2005. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J. Clin. Microbiol.* 43, 406–413.
- Pollock, D.D., Taylor, W.R., 1997. Effectiveness of correlation analysis in identifying protein residues undergoing correlated evolution. *Protein Eng.* 10, 647–657.
- Poon, A.F., Kosakovsky Pond, S.L., Richman, D.D., Frost, S.D., 2007a. Mapping protease inhibitor resistance to human immunodeficiency virus type 1 sequence polymorphisms within patients. *J. Virol.* 81, 13598–13607.
- Poon, A.F., Lewis, F.I., Frost, S.D., Pond, S.L., 2008. Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models. *Bioinformatics*.
- Poon, A.F., Lewis, F.I., Pond, S.L., Frost, S.D., 2007b. An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope. *PLoS Comput. Biol.* 3, e231.
- Prabu-Jeyabalan, M., Nalivaika, E., Schiffer, C.A., 2002. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure* 10, 369–381.
- Prabu-Jeyabalan, M., Nalivaika, E.A., King, N.M., Schiffer, C.A., 2004. Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a V82A drug-resistant mutation in viral protease. *J. Virol.* 78, 12446–12454.
- Rhee, S.Y., Liu, T.F., Holmes, S.P., Shafer, R.W., 2007. HIV-1 subtype B protease and reverse transcriptase amino acid covariation. *PLoS Comput. Biol.* 3, e87.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Sayer, J.M., Liu, F., Ishima, R., Weber, I.T., Louis, J.M., 2008. Effect of the active site D25N mutation on the structure, stability, and ligand binding of the mature HIV-1 protease. *J. Biol. Chem.* 283, 13459–13470.
- Shirakawa, K., Takaori-Kondo, A., Yokoyama, M., Izumi, T., Matsui, M., Io, K., Sato, T., Sato, H., Uchiyama, T., 2008. Phosphorylation of APOBEC3G by protein kinase A regulates its interaction with HIV-1 Vif. *Nat. Struct. Mol. Biol.* 15, 1184–1191.
- Sugiura, W., Matsuda, Z., Yokomaku, Y., Hertogs, K., Larder, B., Oishi, T., Okano, A., Shiino, T., Tatsumi, M., Matsuda, M., Abumi, H., Takata, N., Shirahata, S., Yamada, K., Yoshikura, H., Nagai, Y., 2002. Interference between D30N and L90M in selection and development of protease inhibitor-resistant human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* 46, 708–715.
- Tillier, E.R., Lui, T.W., 2003. Using multiple interdependency to separate functional from phylogenetic correlations in protein alignments. *Bioinformatics* 19, 750–755.
- Tuff, P., Darlu, P., 2000. Exploring a phylogenetic approach for the detection of correlated substitutions in proteins. *Mol. Biol. Evol.* 17, 1753–1759.
- Verheyen, J., Litau, E., Sing, T., Daumer, M., Balduin, M., Oette, M., Fatkenheuer, G., Rockstroh, J.K., Schuldenzucker, U., Hoffmann, D., Pfister, H., Kaiser, R., 2006. Compensatory mutations at the HIV cleavage sites p7/p1 and p1/p6-gag in therapy-naïve and therapy-experienced patients. *Antivir. Ther.* 11, 879–887.
- von Schwedler, U.K., Stuchell, M., Muller, B., Ward, D.M., Chung, H.Y., Morita, E., Wang, H.E., Davis, T., He, G.P., Cimbor, D.M., Scott, A., Krausslich, H.G., Kaplan, J., Morham, S.G., Sundquist, W.I., 2003. The protein network of HIV budding. *Cell* 114, 701–713.
- Wang, J., Cieplak, P., Kollman, P.A., 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* 21, 1049–1074.
- Willey, R.L., Smith, D.H., Lasky, L.A., Theodore, T.S., Earl, P.L., Moss, B., Capon, D.J., Martin, M.A., 1988. In vitro mutagenesis identifies a region within the envelope gene of the human immunodeficiency virus that is critical for infectivity. *J. Virol.* 62, 139–147.
- Wollenberg, K.R., Atchley, W.R., 2000. Separation of phylogenetic and functional associations in biological sequences by using the parametric bootstrap. *Proc. Natl. Acad. Sci. U. S. A.* 97, 3288–3291.
- Wu, T.D., Schiffer, C.A., Gonzales, M.J., Taylor, J., Kantor, R., Chou, S., Israelski, D., Zolopa, A.R., Fessel, W.J., Shafer, R.W., 2003. Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. *J. Virol.* 77, 4836–4847.
- Yeang, C.H., Haussler, D., 2007. Detecting coevolution in and among protein domains. *PLoS Comput. Biol.* 3, e211.
- Zhang, Y.M., Imamichi, H., Imamichi, T., Lane, H.C., Falloon, J., Vasudevachari, M.B., Salzman, N.P., 1997. Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J. Virol.* 71, 6662–6670.