Journal of Antimicrobial Chemotherapy Advance Access published October 19, 2012

J Antimicrob Chemother 2012; **0**: 1–5
doi:10.1093/jac/dks409

Journal of Antimicrobial Chemotherapy

Collinearity of protease mutations in HIV-1 samples with high-level protease inhibitor class resistance

Farbod Babrzadeh[1]†, Vici Varghese[2]*†, Mary Pacold[2], Tommy F. Liu[2], Pål Nyrén[3], Celia Schiffer[4], W. Jeffrey Fessel[5] and Robert W. Shafer[2]

[1]Stanford Genome Technology Center, Stanford University, Stanford, CA, USA; [2]Department of Medicine, Division of Infectious Diseases, Stanford University, Stanford, CA, USA; [3]Department of Biotechnology, AlbaNova University Center, KTH, Stockholm, Sweden; [4]Department of Biochemistry, University of Massachusetts Medical Center, Worcester, MA, USA; [5]Kaiser-Permanente Medical Care Program Northern California, Clinical Trials Unit, San Francisco, CA, USA

*Corresponding author. 300 Pasteur Drive, Grant building Room S-146, Stanford University Medical Center, Stanford, CA, USA-94305. Tel: +1-650-723-8291; Fax: +1-650-725-2395; E-mail: viciv@stanford.edu
†These authors contributed equally to the manuscript.

Received 5 May 2012; returned 5 July 2012; revised 31 July 2012; accepted 12 September 2012

Objectives: To determine whether pan-protease inhibitor (PI)-resistant virus populations are composed predominantly of viruses with resistance to all PIs or of diverse virus populations with resistance to different subsets of PIs.

Methods: We performed deep sequencing of plasma virus samples from nine patients with high-level genotypic and/or phenotypic resistance to all licensed PIs. The nine virus samples had a median of 12 PI resistance mutations by direct PCR Sanger sequencing.

Results: For each of the nine virus samples, deep sequencing showed that each of the individual viruses within a sample contained nearly all of the mutations detected by Sanger sequencing. Indeed, a median of 94.9% of deep sequence reads had each of the PI resistance mutations present as a single chromatographic peak in the Sanger sequence. A median of 5.0% of reads had all but one of the Sanger mutations that were not part of an electrophoretic mixture.

Conclusions: The collinearity of PI resistance mutations in the nine virus samples demonstrated that pan-PI-resistant viruses are able to replicate *in vivo* despite their highly mutated protease enzymes. We hypothesize that the marked collinearity of PI resistance mutations in pan-PI-resistant virus populations results from the unique requirements for multi-PI resistance and the extensive cross-resistance conferred by many of the accessory PI resistance mutations.

Keywords: drug resistance, deep sequencing, minority variants, Sanger sequencing

Introduction

HIV-1 isolates with high-level resistance to most or all protease inhibitors (PIs) are not uncommon in patients who have received multiple PI-containing regimens.[1,2] Considering the decreased fitness associated with PI resistance mutations and their disparate and often antagonistic effects on different PIs,[3–5] we sought to determine whether pan-PI-resistant virus populations comprise viruses with resistance to all PIs or virus subpopulations with different patterns of PI resistance mutations. Because standard genotypic resistance testing using dideoxynucleotide terminator (Sanger) sequencing rarely detects variants present in proportions below 20%–30% of circulating viruses,[6] we performed deep sequencing of pan-PI-resistant virus samples using the 454 Life Sciences/Roche Molecular Systems pyrosequencing technology.

Methods

Patients and samples

Patients included HIV-1-infected individuals in the Kaiser Permanente Medical Care Programme—Northern California undergoing Sanger genotypic resistance testing at Stanford University and phenotypic resistance testing at Monogram Biosciences. Cryopreserved plasma virus samples with HIV-1 RNA levels ≥4.5 log copies/mL and high-level phenotypic or genotypic resistance to all PIs were selected for deep sequencing. The Stanford University Human Subjects Panel approved the study.

Deep sequencing

Extracted plasma viral RNA was reverse transcribed with random primers and Superscript III RT. cDNA titres were estimated by limiting dilution PCR to include >120 amplifiable virus templates in a nested PCR protocol using Expand High Fidelity PLUS DNA polymerase to generate an amplicon encompassing HIV-1 protease. Second-round primers were tailed with a sample-specific barcode. Bidirectional deep sequencing was performed on a GS-FLX sequencer. Each sample pool was loaded in a 70×75 mm PicoTiter plate fitted with an eight-lane gasket.

Flowgram format files were processed to generate paired files containing FASTA sequence reads and Phred-equivalent quality scores. To reduce artefact, trimmed sequence reads with <150 nucleotides, nucleotide quality scores <10 or mean quality scores <25 were excluded.[7] Sequence reads were de-multiplexed using the 5′ primer and barcode, and aligned to the Sanger sequence using MosaikAligner.[8]

Quality control analysis

To distinguish authentic variants from PCR and pyrosequencing artefacts, we estimated the technical error rate by deeply sequencing plasmid HIV-1 pNL43 protease. The plasmid mismatch error rate was 0.20% with a nucleotide position-specific median mismatch error rate of 0.14% (range 0%–1.2% per nucleotide position). Minority variants were defined as mutations not detected by Sanger sequencing occurring in ≥1.0% of sequence reads in a proportion >5.0 times the variant's proportion in the control sequence.

Mutation categories

Minority variant mutations were classified into five categories: (i) PI resistance mutations; (ii) protease variants present in >0.1% of sequences from individuals in the Stanford HIV Drug Resistance Database (HIVDB);[9] (iii) residual wild-type variants defined as consensus B amino acids in samples for which the Sanger sequence contained a mutation; (iv) unusual variants defined as mutations present in <0.01% of non-hypermutated sequences in HIVDB; and (v) silent mutations defined as synonymous nucleotide differences from the Sanger sequence.

Non-polymorphic PI-selected mutations[10] and L10I/V, V11I, L33F and A71V/T were classified as PI resistance mutations. The subset D30N, V32I, M46I/L, I47V/A, G48V/M, I50L/V, I54V/L/M/T/A/S, L76V, V82A/T/F/S/L, I84V, N88D/S and L90M were classified as major mutations.[11]

Sequence heterogeneity

To characterize intra-patient nucleotide and amino acid sequence heterogeneity, we clustered reads with identical or nearly identical sequences defined as sequences differing only in length (provided they encompassed ≥180 nucleotides); a stop codon; a nucleotide deletion; or a mutation present in only one strand in a proportion <0.25%.

Results

Patients and samples

Nine plasma samples met the inclusion criteria (Table 1). The darunavir phenotype was not available for one sample; three samples had intermediate phenotypic tipranavir resistance. The patients had received a median of five PIs for a median of 8.5 years. The median time of uninterrupted therapy prior to sample collection was 12 months (range 5–32 months). The median number of PI resistance mutations per sample was 12 (range 10–14).

We generated a median of 2134 high-quality sequence reads per sample (range 830–923), including a median of 1331 forward reads encompassing positions 10–84 (range 527–933), and 966 reverse reads encompassing positions 30–99 (range 263–1990). No reads encompassed all PI resistance positions (10–90).

PI resistance mutations detected by Sanger sequencing

The second column in Table 2 shows the proportion of deep sequence reads with each PI resistance mutation present in the Sanger sequence: 88 (80%) of the 110 PI resistance mutations occurred in ≥99.0% of sequence reads. Of the remaining 22 PI resistance mutations, the eight underlined mutations were present as electrophoretic mixtures by Sanger sequencing and occurred in a median of 54.5% of reads (range 11.4%–79.1%). The remaining 14 mutations were present in a median of 96.1% of reads (range 79.1%–98.5%).

Table S1 (available as Supplementary data at *JAC* Online) shows that of the PI resistance mutations present in an unmixed form by Sanger sequencing (range 9–12), a median of 94.9% of reads per sample (range 80.6%–97.6%) contained each mutation within its coverage (10–84 for forward reads and 30–90 for reverse reads). A median of 5.0% of reads contained all but one PI resistance mutation (range 2.4%–18.9%) and a median of 0.1% contained all but two mutations (range 0%–1.6%). Figure S1 (available as Supplementary data at *JAC* Online) shows the amino acid alignments of all distinct viral variants in ≥1.0% of forward or reverse deep sequence reads.

Minority variant mutations

Columns 3–7 in Table 2 show the five categories of minority variant mutations. The nine samples contained 54 amino acid and 78 silent minority variants. A median of six amino acid minority variants was detected per sample (range 2–11). Of the 38 minority variant amino acid mutations between codons 30 and 84, all but three were present in both forward and reverse reads.

Eighteen of 54 minority variant amino acid mutations were PI resistance mutations, including two major mutations: N88S (1.7% of 1456 reads) and I50V (20.2% of 38129 reads). Three minority variant major PI resistance mutations were present in 0.5%–1.0% of reads, including V82T (0.8%) and I47A (0.6%) in sample 6585 and V82A (0.8%) in sample 4736 (data not shown).

Twenty-four (43%) minority variants were known protease variants and seven (13%) were residual wild-type amino acids. The six unusual variants comprised 11% of amino acid minority variants and 4.5% of all minority variants.

Inter- and intra-sample sequence diversity

Although the pan-PI-resistant viruses shared many PI resistance mutations, including L10I/V/F, V32I, L33F, M46I, I47V, I54M/L/V, A71V/I, G73S/T/C, I84V, L89V and L90M, their sequences were divergent. The mean nucleotide distance between the nine Sanger sequences was 12.4%±2.1% and between deep sequence reads from different samples was 11.6%±0.6% (Figure S2 available as Supplementary data at *JAC* Online). By contrast, the mean intra-sample diversity was 1.2% (range 0.7%–2.0%).

Table 1. Protease inhibitor (PI) history, protease mutations and PI susceptibilities of nine pan-PI-resistant HIV-1 isolates

Sample	VL	CD4	PIs (years)	PIs (number)	PIs[a]	Mutations detected by direct PCR Sanger sequencing[b]	RC (%)	ATV[c]	FPV	IDV	LPV	NFV	SQV	TPV	DRV
1456	4.5	173	9	6	SQV, NFV, IDV/r, APV/r, TPV/r, LPV/r	10I/V, **32I**, 33F, 35D, 36M/I, 37S, **46I**, **47V**, 53F/L, **54L**, 55R, 63P, 64V, 66V, 71V, 73T, 79P/A, **84V**, **90M**, 93I/L	9	**98**	**120**	**24**	**140**	**65**	**58**	3.3	NA
1459	4.5	255	8	5	IDV, NFV, SQV/r, APV/r, LPV/r	10I, 13V, 19I, **32I**, 33F, **46I**, **47V**, **54M**, 64V, 71V, 73C, 77I, **84V, 90M**, 95L	2	**38**	**15**	**25**	**85**	**50**	**9.4**	6.3	**142**
1556	5.7	12	9	7	IDV, NFV, SQV/r, APV/r, LPV/r, TPV/r, DRV/r	10F, 11I, 12P, 13V, 15V, 19P, 20T, **32I**, 33F, 35G, 36I, **54I/V**, 62V, 63P, 70T, 71I, 73S, 79A, **84V**, 89V, **90M**	52	**172**	**108**	**27**	**62**	**83**	**41**	**9.8**	**134**
4736	4.7	208	12	≥3	IDV, LPV/r, TPV/r	10V, 11I, 13V, 14R, 15V, 20T, **32I**, 33F, 36I, 41K, **46L**, **54L**, 57K, 60E, 63P, 68E, 70T, 71I, 72I/M, 73S, **84V**, 89V, **90M**, 93L	13	**109**	**>200**	**21**	**87**	**94**	**93**	7	**239**
6585	4.7	41	3.5	5	IDV, SQV/r, NFV APV/r, LPV/r	10V, 12V/D, 13V, 15V, 20M, **32I**, 33F, 43T, **46I**, **47V**, **54M**, 60E, 61D, 62V, 63P, 67Y, 69K, 71I, 72L, 73S, 77I, **82A**, 89V**, 90M**	96	**88**	**>200**	**88**	**>200**	**76**	**12**	**12**	**112**
7118	5.1	105	7.5	5	SQV/r, LPV/r, LPV/r +APV, ATV/r, DRV/r	10F, 11V/I, 13V, 16A, 19L/V, 33F, 34Q, 43I, **46L**, 51A, **54M**, 63P, 64M, 71V, 72M, 73A, **84V**, **90M**	NA	**111**	**>200**	**47**	**>200**	**51**	**>200**	**9.5**	**140**
7859	4.5	162	11	5	NFV, IDV, APV, LPV/r, ATV/r	10I, 13V, 15V/I, 19I/L, 20A, **32I**, 33F, 36I, **46I**, **47V**, **54M**, 63P, 71V, 74P, 77I, **84V**, 89V, **90M**, 93L	3	**50**	**>200**	**40**	**>200**	**52**	**37**	**17**	**286**
14311	4.8	136	7	3	APV/r, ATV/r, LPV/r, FPV+LPV/r	10F, 11I, **32I**, 35A, 36I, **46I**, **47V**, **54M,** 57K, 62V, 63P, 64V, 73T, 74A, **84V,** 89V, 90M	4	**120**	**>200**	**48**	**>200**	**50**	**22**	**8**	**245**
38129	4.5	22	≥4	≥3	IDV, ATV/r, DRV/r	10I, 11I, 12K, 13V, 20V, **32I**, 33F/L, 35G, 36I, 37D/N, **46I**, **47V**, **54M**, 57K, 58E, 63P, 64V, 71V/I, 73S, **84V/I**, 89V, **90M**	3	**97**	**>200**	**55**	**>200**	**80**	**26**	**>200**	**>200**

VL, plasma HIV-1 RNA level (log copies/mL); CD4, CD4+ lymphocytes (cells/mm^3); RC, replication capacity; ATV, atazanavir; DRV, darunavir; FPV, fosamprenavir; IDV, indinavir; LPV, lopinavir; NA, not available; NFV, nelfinavir; SQV, saquinavir; TPV, tipranavir.

The GenBank accession numbers for the direct PCR sequences are as follows: GQ211137 (1456), AY796708 (1459), pending (1556), GQ212652 (4736), AY797430 (6585), pending (7118), GQ213748 (7859), pending (14311), GQ213273 (38129).

[a]'/r' following the PI abbreviation indicates ritonavir co-administration for pharmacokinetic boosting; '+' indicates the simultaneous use of two PIs.

[b]Mutations in bold are major PI resistance mutations. Underlined mutations are non-polymorphic PI-selected mutations and the minimally polymorphic PI-selected mutations L10IV, V11I, L33F and A71VT.

[c]The last eight columns indicate the fold decrease in susceptibility for each PI as determined by the PhenoSense assay. Those in bold are considered highly resistant according to the PhenoSense clinical cut-off. Each of the nine viruses displayed high-level ritonavir resistance (data not shown).

Table 2. Percentage of deep sequence reads with mutations detected by direct PCR sequencing and with minority variant mutations

Sample	PI resistance mutations detected by direct PCR sequencing[a] (percentage of deep sequencing reads with each mutation)	Mutations not detected by direct PCR sequencing (minority variant mutations)				
		PI resistance mutations[a] (%)	other variants (%)	residual WT[b] (%)	unusual variants[c] (%)	silent variants[d] (no.)
1456	*L10I (52.6)*, *L10V (46.5)*, **V32I (99.4),** L33F (99.7), **M46I (99.9)**, **I47V (99.9)**, *F53L (21.4)*, **I54L (99.0)**, A71V (86.0), G73T (99.9), *P79A (11.6)*, **I84V (99.4)**, **L90M (97.3)**	K43T (15.4), A71I (14.0), **N88S (1.7)**	G16A (7.5), I15V (5.0), Q18E (1.4), K20R (1.3)			9
1459	L10I (100), **V32I (100)**, L33F (99.9), **M46I (99.9)**, **I47V (99.9)**, **I54M (99.2)**, A71V (99.7), G73C (100), **I84V (99.9)**, **L90M (100)**		I13V (12.4), R41K (2.7), R87K (1.6), C67W (1.2), Q18H (1.2)	V77 (1.0)	D29G (2.8)	6
1556	L10F (79.9), V11I (99.7), K20T (99.4), **V32I (99.4)**, L33F (99.8), E35G (98.5) ***I54V (74.0)***, A71I (95.3), G73S (99.6), **I84V (99.8)**, **L89V (99.8)**, **L90M (99.4)**	L10V (18.1), A71V (4.5), T74P (1.8), L10I (1.3)	K70E (7.5), K43R (3.0), R57K (2.2), T12S (1.4), R41K (1.2)	E35 (1.5), L10 (1.1)		12
4736	L10V (99.6), V11I (99.9), K20T (99.0), **V32I (99.9)**, L33F (98.6), **M46L (99.3)**, **I54L (99.3)**, A71I (100), G73S (99.8), **I84V (99.4)**, **L89V (99.1)**, **L90M (99.7)**			L33 (1.3)	G49E (4.1)	3
6585	L10V (100), **V32I (100)**, L33F (94.2), K43T (95.6), **M46I (99.7)**, **I47V (99.4)**, **I54M (99.3)**, A71I (99.0), G73S (100), **V82A (99.1)**, **L89V (100)**, **L90M (100)**	Q58E (3.3), K55R (1.5), V11I (1.1)	Q61N (14.1), L33F_V (5.8), H69N (1.8)	K43 (4.1)		8
7118	L10F (99.5), *V11I (71.2)*, L33F (99.3), **M46L (99.4)**, G51A (100), **I54M (99.2)**, A71V (99.8), G73A (99.7), **I84V (96.5)**, **L90M (97.4)**		L19I (2.5), C67Y (1.1)		L90I (1.6)	8
7859	L10I (99.6), **V32I (99.9)**, L33F (99.9), **M46I (99.7)**, **I47V (99.7)**, **I54M (99.1)**, A71V (100), T74P (99.7), **I84V (99.7)**, L89V (99.7), **L90M (99.6)**	E34G (1.2)	E65K (1.6), C67Y (1.2)		R87S (1.3)	12
14311	L10F (86.8), V11I (99.5), **V32I (99.8)**, **M46I (99.7)**, **I47V (99.9)**, **I54M (99.5)**, G73T (99.9), **I84V (99.6)**, L89V (100), **L90M (99.4)**	L10Y (12.8)	L63T (3.1)			14
38129	L10I (99.7), V11I (84.3), **V32I (99.9)**, *L33F (59.2)*, **M46I (99.9)**, **I47V (99.8)**, **I54M (99.0)**, Q58E (99.6), A71V (59.1), *A71I (34.6)*, G73S (99.8), ***I84V (71.1)***, L89V (79.1), **L90M (99.5)**	**I50V (20.2)**, I66V (17.7), V11L (15.5), E35N (10.7), A71T (6.3), I66F (2.6)	Q92K (3.1), V82I (3.0)	L89 (20.7), R57 (1.6)	G51R (1.7)	6

WT, wild-type.
[a]Mutations in bold are major PI resistance mutations. Underlined italicized mutations were present as part of electrophoretic mixtures.
[b]Variants with the consensus B amino acid in a sample for which the direct PCR sequence had a difference from consensus B.
[c]Variants present in <0.01% of non-hypermutated sequences in the Stanford HIV Drug Resistance Database.
[d]Minority variants that did not result in an amino acid difference from the direct PCR Sanger sequence.

Discussion

Deep sequencing of plasma virus samples from nine patients with high-level phenotypic and/or genotypic resistance to all approved PIs showed that these pan-PI-resistant virus populations were composed of viruses with genotypic resistance to all PIs rather than of subpopulations with resistance to different PIs. Because Sanger sequencing is able to detect variants present in proportions as low as 20%–30%, the absence of electrophoretic mixtures at nearly all PI resistance positions made it likely *a priori* that the proportions of each mutation would be ≥70%–80% of the plasma virus population. However, the finding that >99% of sequence reads contained all or all but one of the median 11 unmixed PI resistance mutations could only be determined by deep sequencing.

Because the GS-FLX reagents used yielded read lengths of 200–250, the forward primer reliably encompassed positions 10–84 and the reverse primer reliably encompassed positions 30–99, but neither encompassed all PI resistance positions. However, considering the overlap between positions 30 and 84, the concordance between forward and reverse reads and the near-complete absence of minority variants at unmixed PI resistance positions, it is overwhelmingly likely that our findings would not have differed had positions 10–90 been covered by both sets of primers.

One limitation of deep sequencing is that PCR-induced template-product bias complicates the quantification of sequenced minority variants. Indeed, a novel method, which attaches a barcode to each cDNA molecule prior to PCR, overcomes this limitation and also helps to distinguish true minority variants from technical artefacts.[12] Although we did not use this method, our limiting dilution analysis showed that >120 cDNA templates were present in each sample prior to PCR. In addition, our *post hoc* analyses showed that >95% of minority variants occurring in ≥1.0% of reads were silent mutations or previously reported variants. This finding suggests that ∼100 genomes per sample were likely to have been successfully amplified and sequenced.

Although at least eight of the nine viruses had PI resistance mutations at positions 10, 32, 33, 46, 54, 71, 73, 84 and 90, the pan-PI-resistant viruses were genetically distinct, indicating that the emergence of this pattern of protease mutations is an extraordinary example of convergent virus evolution. The collinearity of PI resistance mutations in these samples, moreover, indicates that pan-PI-resistant viruses are able to replicate *in vivo* despite their highly mutated protease enzymes.

Funding

The study was funded in part by NIAID AI068581 and California HIV Research Program ID08-ST-031

Transparency declaration

None to declare.

Supplementary data

Table S1 and Figures S1 and S2 are available as Supplementary data at *JAC* Online (http://jac.oxfordjournals.org/).

References

1 Talbot A, Grant P, Taylor J *et al*. Predicting tipranavir and darunavir resistance using genotypic, phenotypic, and virtual phenotypic resistance patterns: an independent cohort analysis of clinical isolates highly resistant to all other protease inhibitors. *Antimicrob Agents Chemother* 2010; **54**: 2473–9.

2 Doherty KM, Nakka P, King BM *et al*. A multifaceted analysis of HIV-1 protease multidrug resistance phenotypes. *BMC Bioinformatics* 2011; **12**: 477.

3 Weinheimer S, Discotto L, Friborg J *et al*. Atazanavir signature I50L resistance substitution accounts for unique phenotype of increased susceptibility to other protease inhibitors in a variety of human immunodeficiency virus type 1 genetic backbones. *Antimicrob Agents Chemother* 2005; **49**: 3816–24.

4 Wiesmann F, Vachta J, Ehret R *et al*. The L76V mutation in HIV-1 protease is potentially associated with hypersusceptibility to protease inhibitors atazanavir and saquinavir: is there a clinical advantage? *AIDS Res Ther* 2011; **8**: 7.

5 Ziermann R, Limoli K, Das K *et al*. A mutation in human immunodeficiency virus type 1 protease, N88S, that causes in vitro hypersensitivity to amprenavir. *J Virol* 2000; **74**: 4414–9.

6 Gianella S, Richman DD. Minority variants of drug-resistant HIV. *J Infect Dis* 2010; **202**: 657–66.

7 Huse SM, Huber JA, Morrison HG *et al*. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* 2007; **8**: R143.

8 Li H, Homer N. A survey of sequence alignment algorithms for next-generation sequencing. *Brief Bioinform* 2010; **11**: 473–83.

9 Rhee SY, Gonzales MJ, Kantor R *et al*. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res* 2003; **31**: 298–303.

10 Shahriar R, Rhee SY, Liu TF *et al*. Nonpolymorphic human immunodeficiency virus type 1 protease and reverse transcriptase treatment-selected mutations. *Antimicrob Agents Chemother* 2009; **53**: 4869–78.

11 Rhee SY, Taylor J, Fessel WJ *et al*. HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrob Agents Chemother* 2010; **54**: 4253–61.

12 Jabara CB, Jones CD, Roach J *et al*. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc Natl Acad Sci USA* 2011; **108**: 20166–71.