

DOCTORAL THESIS 2013

Evolution of the Human Immunodeficiency Virus type I
Protease and Integrase:
Effects on Viral Replication Capacity and Robustness.

Elena Capel Malo

Evolution of the human immunodeficiency virus type 1 integrase: Effects on viral replication capacity

Relationship between HIV-1 integrase sequence conservation and viral RC

To investigate whether HIV-1 integrase sequence diversification affected virus RC, 94 integrases were randomly selected from the set of 139 HIV-1 infected patients plasma samples, described in the previous study that were naïve for protease and integrase inhibitors. Forty-seven integrases were chosen from the 1993–1994 samples (mean viral load \pm SE = 128255 \pm 30759 copies/ml, mean CD4+ cell count \pm SE = 302 \pm 29 cells/ μ l), and the other 47 were from the 2006–2007 samples (mean viral load \pm SE = 313223 \pm 98542 copies/ml, mean CD4+ cell count \pm SE = 358 \pm 29 cells/ μ l). Bulk patient plasma-derived PCR integrase amplicons were recombined with an integrase-deleted pNL4-3 infectious clone. The integrase region of the chimeric viruses was then sequenced from the resultant chimeric viral stocks and compared to the original source plasma (Fig 35).

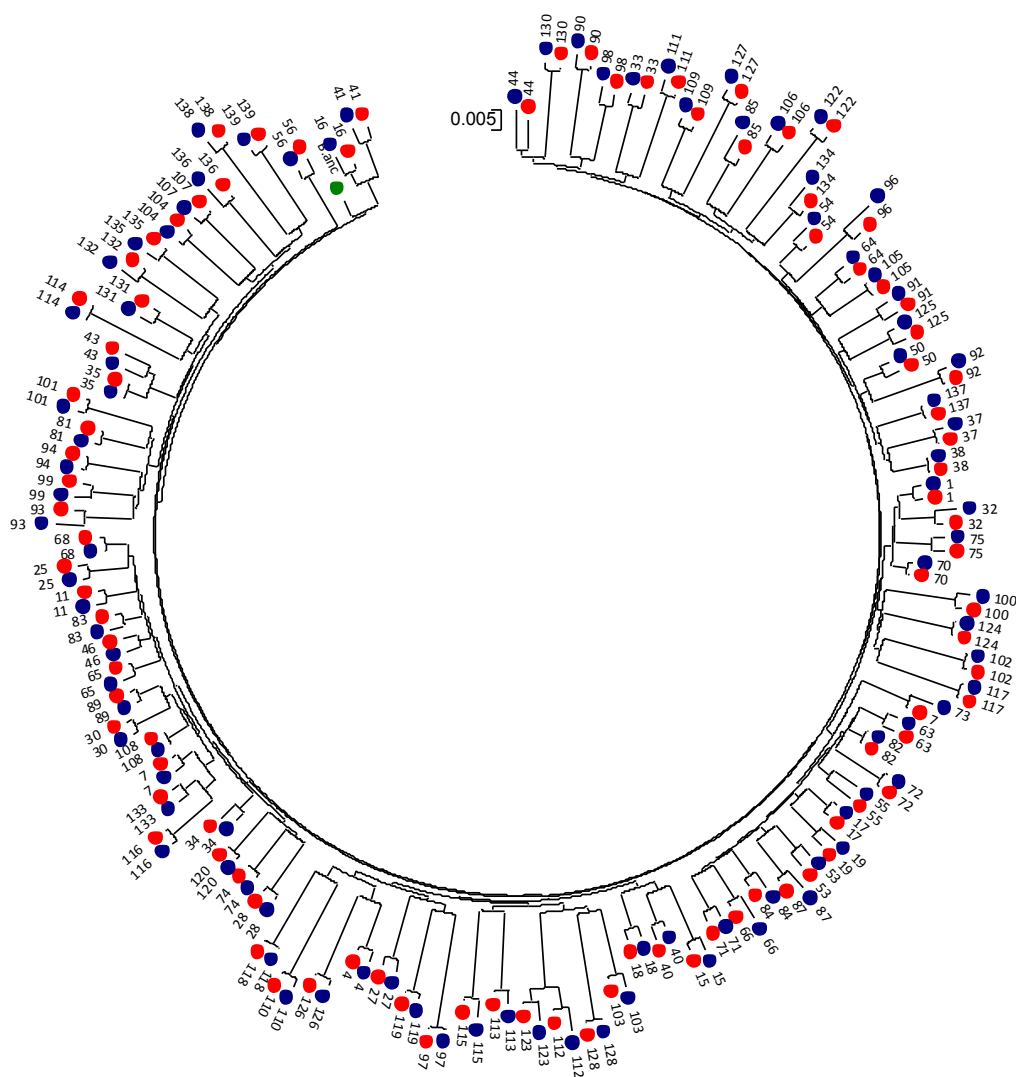


Figure 35: Validation of obtained chimeric pNL4-3 sequences by phylogenetic analysis. The tree was drawn by the neighbour-joining method using the integrase nucleotide sequences. Red taxa symbols indicate sequences from chimeric viral stocks. Blue taxa symbols indicate the original plasma viral sequence. The ancestral B sequence is shown in green. Note that all viral stocks and plasma viral sequences from the same subjects shared the same branches.

A mean \pm SD of $0.17\% \pm 0.14$ and $0.39\% \pm 0.38$ differences at the nucleotide and amino acid level, respectively, was observed between plasma HIV RNA and chimeric virus. These data indicated that although there may have been some selection of chimeric viruses during recombination and/or culture, integrase chimeric viruses represented the dominant form of plasma viruses and selection during culture was unlikely to affect the overall analysis. To quantitatively compare the ex vivo viral RC of the resulting 94 chimeric viruses, CEM-GFP cells were infected and viral growth was measured by quantifying the GFP production on days 3 to 7 after infection (Fig 36, A). Viral RC was calculated as the slope of the natural log of percent GFP-expressing cells between days 3 and 7 after infection (Fig 36, B). The viral RC of the bulk (quasispecies-containing) recombinant viruses was assayed in three independent experiments, and the mean viral RC values were calculated. Chimeric viruses derived from naïve samples isolated in 2006–2007 showed a significantly lower ex vivo viral RC compared to the viral RC of viruses from naïve samples isolated in 1993–1994 (Fig 36, B). The mean \pm SE RC of the early (1993–1994) viruses was 0.97 ± 0.02 , and that of the late (2006–2007) viruses was 0.90 ± 0.03 ($p = 0.0286$).

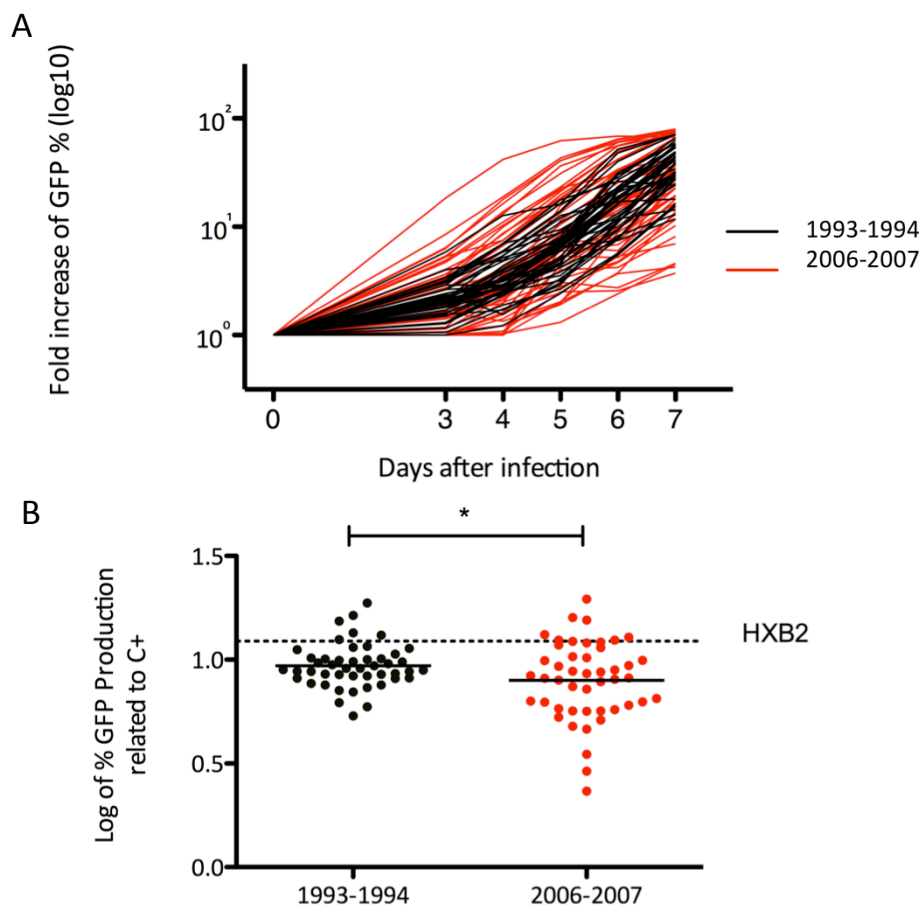


Figure 36: Comparison of viral RC of the 94 chimeric integrase viruses from naïve samples isolated in 1993–1994 and from naïve samples isolated in 2006–2007. (A) The kinetics of viral replication. Black and red lines indicate viruses isolated in 1993–1994 and in 2006–2007, respectively. (B) Comparison of the slopes of the natural logs from days 3 to 7 after infection. The viral RC of each variant was compared with that of the virus with the recombinant integrase from patient 1 (1.0). The values represent the means of at least three independent experiments. P values were determined using the unpaired t test. Dashed line indicates viral RC of wild-type HXB2 recombinant virus.

Viral RC was also analysed by determining how conserved the integrases sequences were relative to an ancestral subtype B sequence. No correlation was found between the starting intersample diversity from early isolates and ex vivo viral RC ($R = 0.0279$, $p\text{-value} = 0.2615$, linear regression) (Fig 37), whereas a low correlation was found between the starting intersample diversity from late isolates and ex vivo viral RC ($R = 0.0873$, $p\text{-value} = 0.0437$, linear regression) (Fig 37). Remarkably, the two recombinant viruses from the late isolates (97 and 132) with the lowest viral RC also had the lowest conservation values at nucleotide and amino acid level.

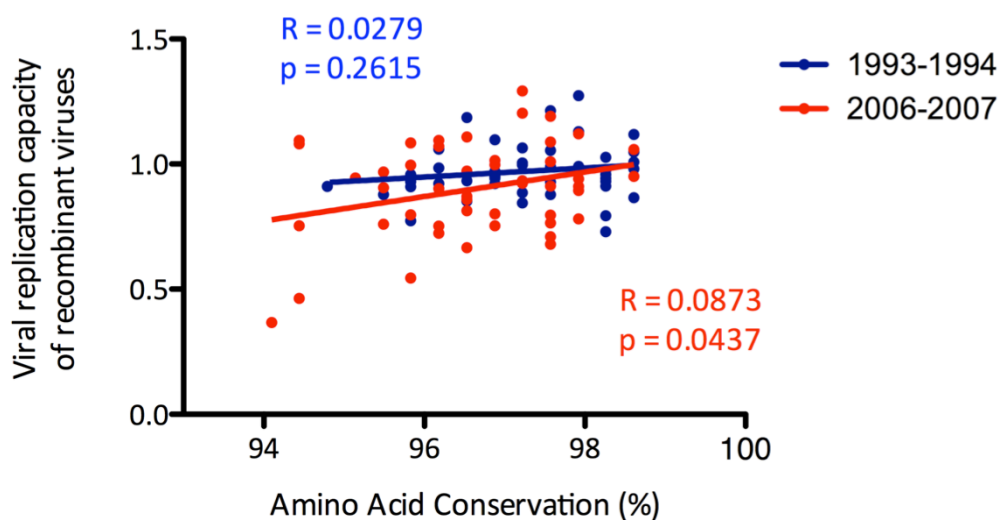


Figure 37: Relationship between VRC and amino acid sequence conservation of the integrase recombinant viruses. ($R = 0.0279$, p -value = 0.2615 for the samples isolated in 1993-1994, and $R = 0.0873$, p -value = 0.0437 for the samples isolated in 2006-2007).

In addition, there was no positive correlation between viral RC and plasma viral load ($R = 0.0077$, $p = 0.3996$, linear regression) (Fig 38) or CD4 counts in all patients ($R = 0.0050$, $p = 0.5020$, linear regression) (Fig 39). Taken together, these results demonstrate that HIV-1 integrase diversification over time has influenced ex vivo virus RC.

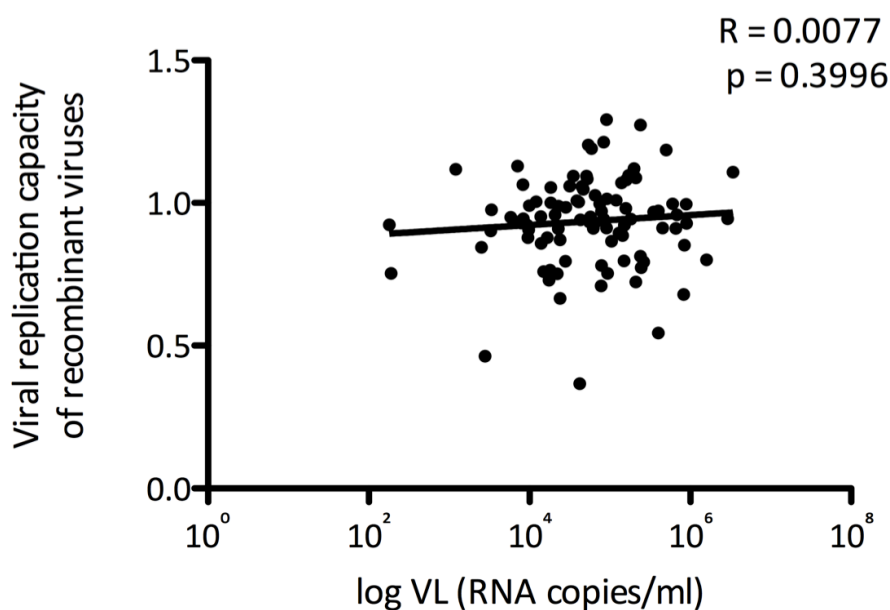


Figure 38: Relationship between VRC and plasma viral load of the integrase recombinant viruses ($R = 0.0077$, p -value = 0.3996).

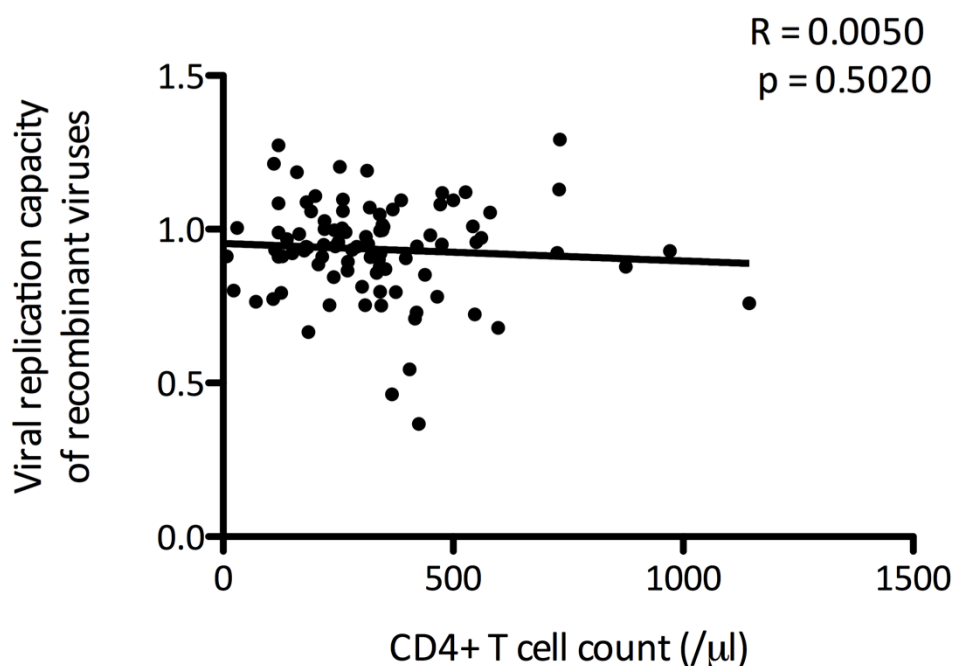


Figure 39: Relationship between VRC and CD4+ T cell count of the integrase recombinant viruses ($R = 0.0050$, p -value = 0.5020).

So far, more than 60 integrase mutations have been associated, to some extent, with resistance to all different INIs tested in in vitro and/or in vivo studies (Baldanti *et al.*, 2010; Bedard, 2007; Canducci *et al.*, 2009; 2010; Ceccherini-Silberstein *et al.*, 2010; Cooper *et al.*, 2008; da Silva *et al.*, 2010; Fikkert *et al.*, 2004; Fransen *et al.*, 2008; Goethals *et al.*, 2008; Hazuda *et al.*, 2004a; Johnson *et al.*, 2009; Lataillade *et al.*, 2007; Malet *et al.*, 2008; McColl *et al.*, 2007; Miller *et al.*, 2008; Seki & Kobayashi, 2010; Shafer & Schapiro, 2008; Shimura *et al.*, 2008; Van Baelen *et al.*, 2009; Wai *et al.*, 2007), though many of them can be simply polymorphisms that are co-selected with true resistance integrase mutations. There is no phenotypic evidence that all these changes contribute to resistance (Low *et al.*, 2009; Van Baelen *et al.*, 2008; 2009). We found in our samples 30 amino acid substitutions associated with the former resistance mutations to INSTIs (S17N/T, L68V, I72V, K111T, S119P/R/G, T124N, T125A/N/V, V151I, M154I/L, K156N/H, E157Q, K160G/R, I161V, G163Q/E, H171Y, V201I, I203M, T206S, S230N, D232E, D256E)(Annex IV). These mutations appeared alone or combined with other substitutions. There was no evidence that these substitutions may be influencing the viral RC. For instance, recombinant viruses from patients carrying the accessory resistant mutation K156N (Table 10) (Low *et al.*, 2009), displayed different viral RC. In contrast, no primary mutations associated with resistance to raltegravir (N155H, Q148H/R/K and G140S/A) or

Results

elvitegravir (T66I, E92Q, Q146P and S147G) that affect the HHCC zinc-binding domain or the DDE motif in the enzyme catalytic core were detected. These results agree with recent published data (Ceccherini-Silberstein et al., 2010).

Table 10. Comparison of viral RC from recombinant virus carrying the K156N minor resistance mutation.

Patient	VRC	INSTIs RAM
130	0,797	K156N; S230N; E157Q; V201I
19	0,852	K156N; V201I
71	0,950	K156N; T125V
100	1,203	K156N; V201I; T206S

VRC: Viral replication capacity; RAM: Resistance associated mutation

Identification of integrase amino acids associated with viral replication capacity

We next performed an exploratory codon-by-codon analysis of integrase sequences in the early and late datasets in order to identify polymorphisms that were significantly associated with viral RC. Briefly, after constructing a maximum likelihood tree, ML values were computed (See Materials and Methods, page 57). When the number of nonsynonymous substitutions per site (dN) was higher than 3, we compared the viral RC of the recombinant viruses carrying the amino acid substitution to the viral RC of the recombinant viruses without this amino acid substitution (Table 11). We found that integrase polymorphisms S17N, I72V, S119P, and D256E were each linked with lower RC (all 4 $p < 0.05$; Table 11).

Table 11: Exploratory analysis of integrase polymorphisms and viral RC^a ($P < 0.05$)

Patients	Integrase codon	aa	Number of	Number of	Mean RC \pm SE	Mean RC \pm SE	P value
			patients with aa	patients without aa	with aa	without aa	
1993-1994	S17	N	7	40	0.89 \pm 0.03	0.98 \pm 0.02	0.0317
1993-1994	I72	V	24	23	0.92 \pm 0.02	1.02 \pm 0.02	0.0015
2006-2007	S119	P	10	37	0.79 \pm 0.08	0.93 \pm 0.03	0.0398
2006-2007	D256	E	17	30	0.82 \pm 0.05	0.95 \pm 0.03	0.0246
ALL	D256	E	25	69	0.86 \pm 0.04	0.96 \pm 0.02	0.0033

^a The analysis is restricted to polymorphisms with a frequency greater than or equal to 5.

Abbreviations: aa, amino acid; RC, replication capacity; SE, Standard Error.

To assess the potential additive effects of these 4 polymorphisms on integrase function, we stratified viral RC according to the number of mutations encoded in each patient sequence (0 to 3) and observed a statistically significant inverse correlation ($R^2 = 0.9406$, $p = 0.0301$, linear regression; Spearman $R = -1.00$, $p =$

0.0833; Fig 40), indicating that these four polymorphisms may contribute additively to integrase function.

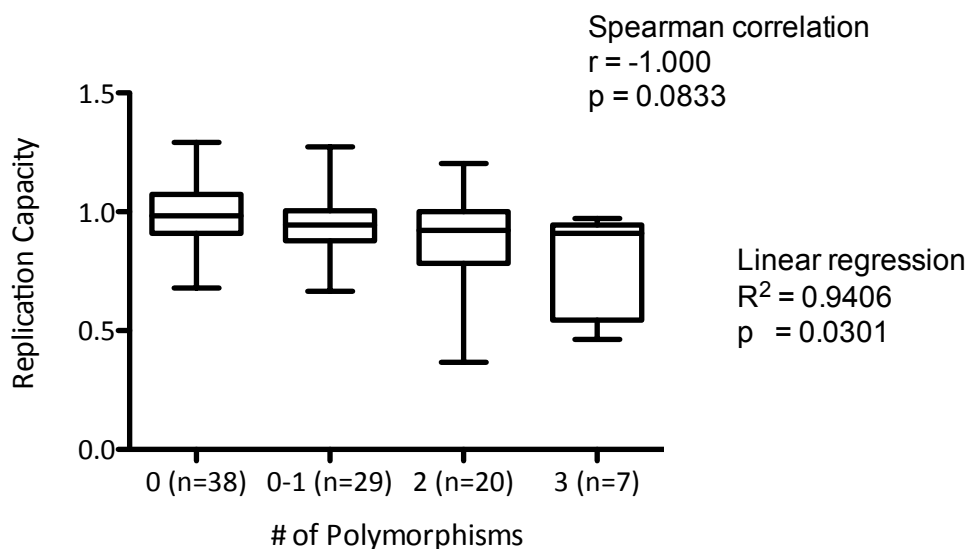


Figure 40: Additive effect of integrase polymorphisms on viral replication capacity. The viral replication capacities (RC) were stratified according to the total number of polymorphisms (S17N, I72V, S119P, and D256E) (all $P < 0.05$) (Table 2) encoded by each patient-derived sequence. An inverse association between viral RC and the number of polymorphisms was observed ($R^2 = 0.9406$, $p = 0.0301$, linear regression; Spearman $R = -1.00$; $P = 0.0833$).

HIV-1 protease robustness determined by in vitro evolution

Comparable vulnerability of the HIV-1 wild type and an artificial mutated protease to single random amino acid mutations

The mutational robustness of a natural HIV-1 clone, HXB2 was compared to the robustness of an HXB2 mutant, 17a, carrying 4 mutations (I15V, I62V, H69R, and I85V). The 17a mutant was generated in vitro and was chosen for this study because it displayed a good in vitro catalytic efficiency (Figure 41, Table 14). HXB2 and 17a proteases were subjected to random mutagenesis and 185 individual clones carrying one single amino acid substitution were selected; 125 and 60 clones from HXB2 and 17a, respectively. The catalytic efficiency of the 185 HIV-1 single proteases was then determined.

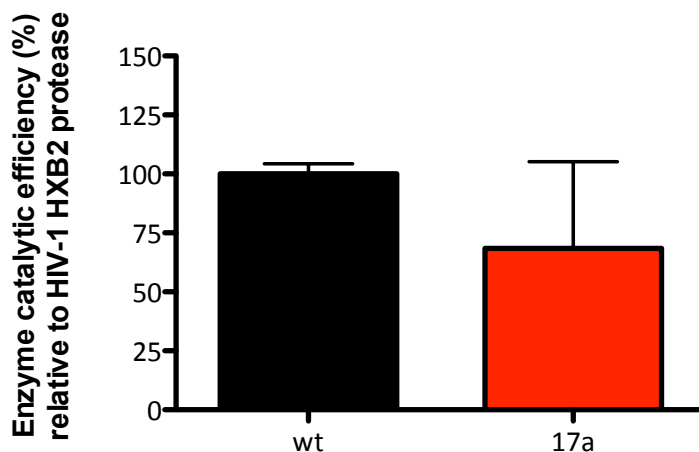
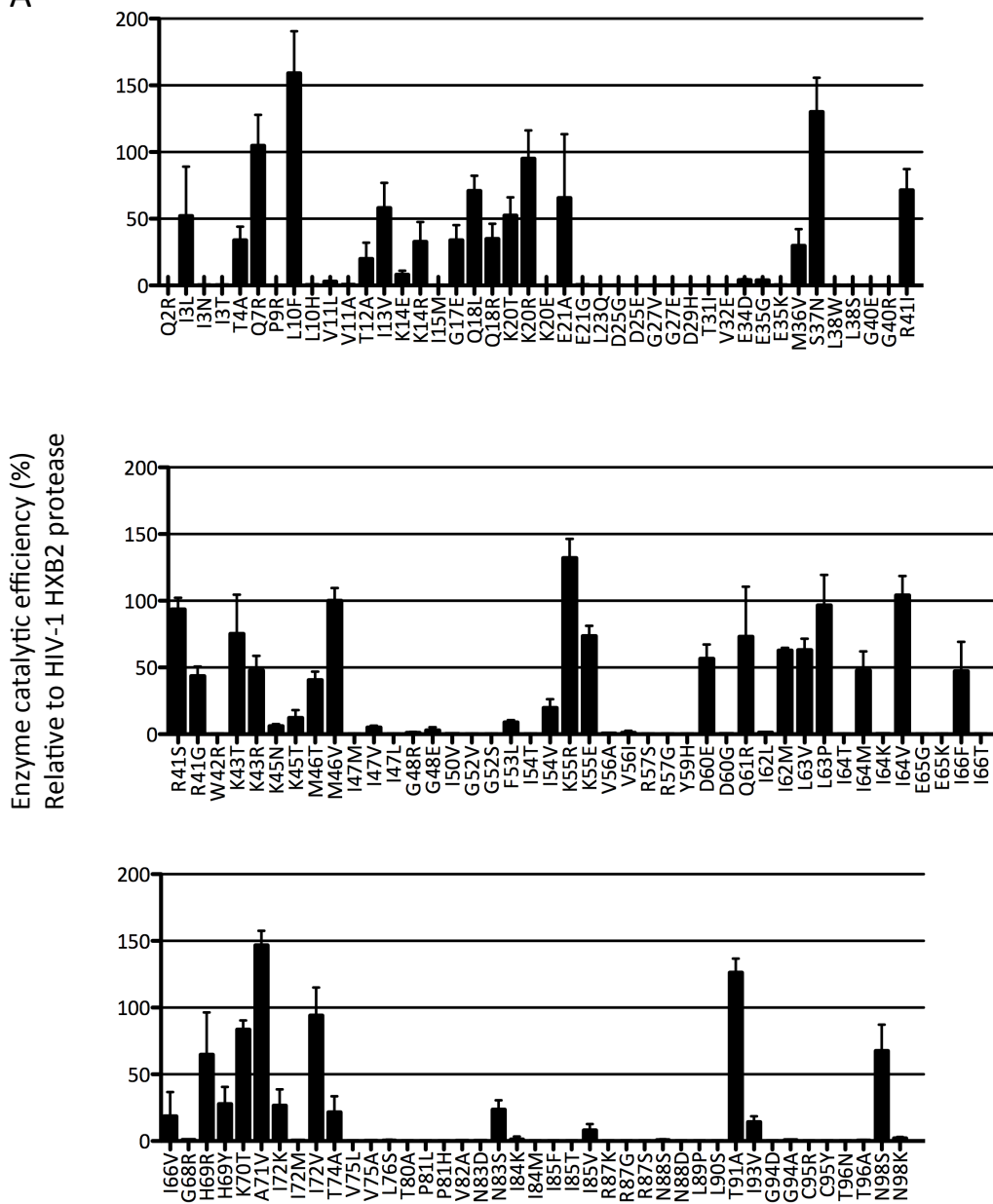


Figure 41: Catalytic efficiency of the mutated HIV-1 protease 17a. The catalytic efficiency of the mutated protease was compared with HXB2 wt protease (100%). Twenty independent replicates were performed for each sample. Error bars correspond to standard deviations.

The 125 single substitutions were distributed through the protein and affected 73 different residues in HXB2 (Fig 42, A) and 52 residues in 17a (Fig 42, B) out of 99 HIV-1 protease amino acid residues. Both groups shared identical substitutions in 26 residues and 13 amino acid positions were affected by different substitutions. To determine the enzymatic activity of the different identified single-mutant proteases, a phage lambda-based genetic screen was used. This genetic screen is based on the phage lambda regulatory circuit in which the viral repressor *cl* is specifically cleaved to initiate the lysogenic to lytic switch (Sices & Kristie, 1998). Introduction of an HIV-1 protease in a wild-type phage will cleave a mutant *cl* repressor containing a specific HIV-1 protease cleavage site, allowing the phage to go into the lytic replication cycle. As we have previously demonstrated, the *cl* repressor cleavage is directly proportional to the protease catalytic efficiency (Cabana *et al.*, 2002; Martinez *et al.*, 2000; Martínez & Clotet, 2003; Parera *et al.*, 2004).

A



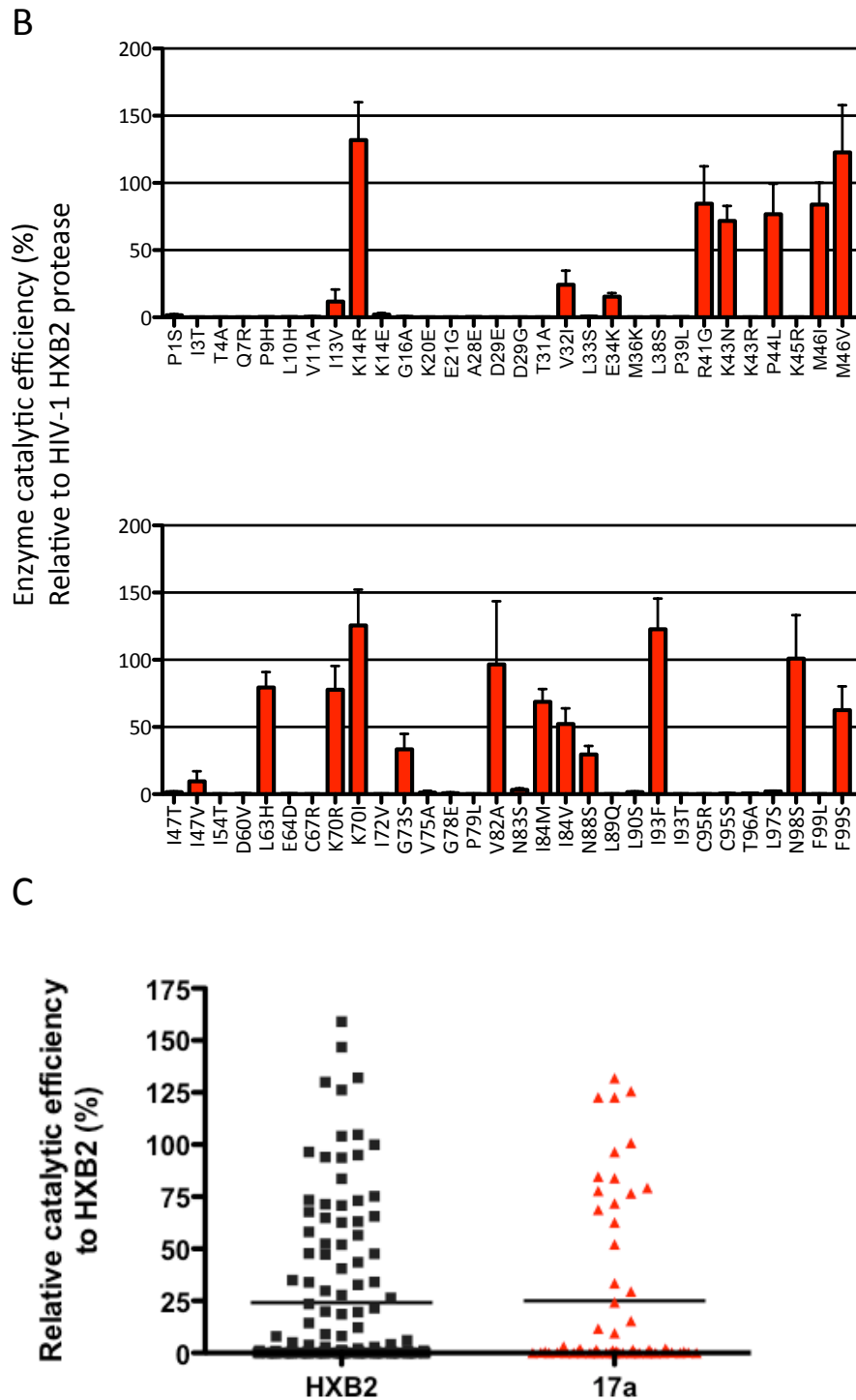


Figure 42: Comparative catalytic efficiency of 185 HIV-1 protease mutants carrying unique single amino acid substitutions. The catalytic efficiency of the single protease variants was compared with HXB2 protease (100%). Three independent replicates were performed for each sample. Error bars correspond to standard deviations. (A) Panel that shows the 125 single-mutants obtained from HXB2 protease. (B) Panel that shows the 60 single-mutants obtained from the 17a protease. (C) Comparison of both HIV-1 one-single variant mutants relative catalytic efficiency.

Results

This HIV-1 protease genetic screen model has to be seen as a complement to the classical biochemical approach for monitoring protease proteolytic activity. The enzymatic activities of the different single-variant proteases analysed in the study were related to the activity of the wild type HXB2 protease (100%)(Fig 42). Although the mean catalytic efficiency of the HXB2 proteases, 24.13 ± 3.39 (mean \pm SE), was lower compared to the 17a proteases, 25.02 ± 5.24 (mean \pm SE), this difference was not statistically significant ($p = 0.8842$, unpaired t-test)(Fig 42, C). Mutations were classified depending on their catalytic efficiency (Table 12, Fig 43). In the group of variants obtained from the HXB2 clone, sixty-eight (54%) mutations were lethal ($p < 0.05$, t-test), forty (32%) were deleterious ($p < 0.05$, t-test), thirteen (10%) had no significant effect and four (3%) were beneficial ($p < 0.05$, t-test). Similarly, in the single-variant proteases obtained from the 17a clone, twenty-eight (47%) of the mutations were lethal ($p < 0.05$, t-test), fifteen (25%) were deleterious ($p < 0.05$, t-test), fourteen (23%) had no significant effect and three (5%) were beneficial ($p < 0.05$, t-test compared to positive control). Single-variant proteases obtained from the 17a clone showed a relative lower amount of deleterious (25%) and lethal mutations (47%) compared to the single-variant proteases from the wild type protease (32%, 54% respectively), as well as a relative higher amount of neutral (23%) and beneficial mutations (5%) compared to HXB2 proteases, 10% and 3%, respectively. Nevertheless, these differences were not statistically significant ($p = 0.0731$, Chi-square test). Moreover, the average catalytic efficiency of the different type of mutations (lethal, neutral and beneficial) was similar for variants from both templates. Although 17a deleterious proteases displayed a lower average catalytic efficiency compared to HXB2 deleterious proteases (Table 12), this difference is not relevant since analysed proteases are independent.

Table 12: Summary of the enzymatic activities of the 125 and 60 single-variant mutant proteases obtained from HXB2 and 17a clone, respectively.

Mutations	Lethal		Deleterious		Neutral		Beneficial	
	wt	17a	wt	17a	wt	17a	wt	17a
Absolut Number	68	28	40	15	13	14	4	3
Relative Number (%)	54	47	32	25	10	23	3	5
Average catalytic efficiency (%)	0 ± 0	0 ± 0	32 ± 9	5 ± 3	88 ± 25	74 ± 20	141 ± 17	127 ± 26

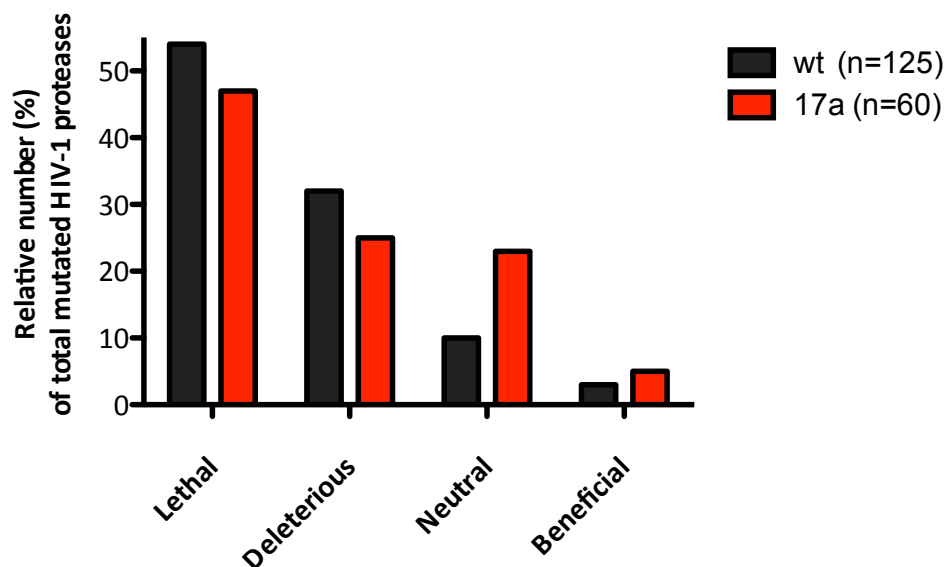


Figure 43: Comparison of the one single amino acid substitution effect on the wt and 17a HIV-1 proteases catalytic efficiency.

A considerable amount of lethal or highly deleterious mutants were distributed among conserved residues critical for the structure and proper function of the protease (Fig 42, 44) (Ceccherini-Silberstein *et al.*, 2004; Ishima, 2001; Loeb *et al.*, 1989; Rao *et al.*, 1991). Of the 185 mutant proteases, 16 carried a mutation on the catalytic site (residues 21-32), 25 were mutated on the flap situated on top of the catalytic site (residues 44-56), 23 proteases carried a mutation on the substrate binding site (residues 78-88), and 27 were mutated whether on the amino or carboxyl terminal residues (residues 1-9, 94-99 respectively) which are involved in the protease dimer stabilization. In short, 91 out of the 185 single-mutant proteases carried a mutation in a conserved region of the enzyme. Nevertheless, lethal or deleterious mutations were also found outside these preserved coding regions (Fig 44). Residues located outside the above 5 critical regions in which we found lethal or highly deleterious mutants (V11, G16, G17, Q18, E34, E35, M36, L38, G40, Y59, D60, I62, I66, G68, G73, T74, V75, L76, L89, and L90) were absent or rarely mutated in drug-naïve infected individuals (Ceccherini-Silberstein *et al.*, 2004; Wu *et al.*, 2003), confirming the possible *in vivo* lethal or deleterious effect caused by the former mutations. Interestingly, when comparing identical amino acid substitutions shared by both groups, different outcomes were observed. Some mutations located at the same position had a very different effect depending on the amino substitution type and on the single-variant template, HXB2 or 17a clone (Table 13). For example,

amino acid substitution phenylalanine (F) for leucine (L) at position 10 (L10F) turned out to have a beneficial effect in the wild type protease, whereas histidine substitution for leucine (L10H) had a lethal effect on both wild type and 17a proteases. Consistent with this result, L10F has been described to be associated with resistance to all PIs when present with other mutations (Robert W Shafer, 2001). It may be argued that our assay does not fully reflect the requirements for functionality of HIV-1 protease in vivo. Nevertheless, site-directed mutagenesis of residue L10 has demonstrated that viruses carrying this substitution could be as infectious as the wild-type virus (Mammano *et al.*, 2000). Similarly, mutations on position 14 had a lethal effect on both groups when lysine (K) was substituted by a glutamic acid (E). On the contrary, when K was substituted by an arginine (R), it turned out to be deleterious in the wild type protease but beneficial in the 17a protease. Neutral or beneficial mutations K20R, K43T, I64V, I93F were located in very well known polymorphic positions in the virus isolated from infected individuals (Ceccherini-Silberstein *et al.*, 2004; Wu *et al.*, 2003) that have also been described to be associated with resistance to different PIs (BA, 2012; Ceccherini-Silberstein *et al.*, 2004; Robert W Shafer, 2001). A 3-dimensional image representing the crystallographic HIV-1 protease structure showed that most of the neutral or beneficial mutations were located at the peripheral areas of the enzyme and almost entirely positioned in surface loops far from the active site and substrate-binding regions.

Results

Table 13: Comparison of single-variant proteases from HXB2 or 17a carrying amino acid substitutions in the same position with different mutational effects.

AA position	AA substitution	Mean Catalytic Efficiency	
		HXB2	17a
L10	F	159	-
	H	0	0
K14	E	8	2
	R	33	132
K20	E	0	0
	R	95	-
	T	52	-
K43	R	48	0
	T	75	-
	N	-	72
I64	K	0	-
	M	48	-
	T	0	-
	V	104	-
	D	-	0
H69	R	65	-
	Y	28	-
I72	K	27	-
	M	0	-
	V	94	0
I93	F	-	123
	T	-	0
	V	14	-

NB: Dash line indicates that this substitution is not present in the corresponding library.

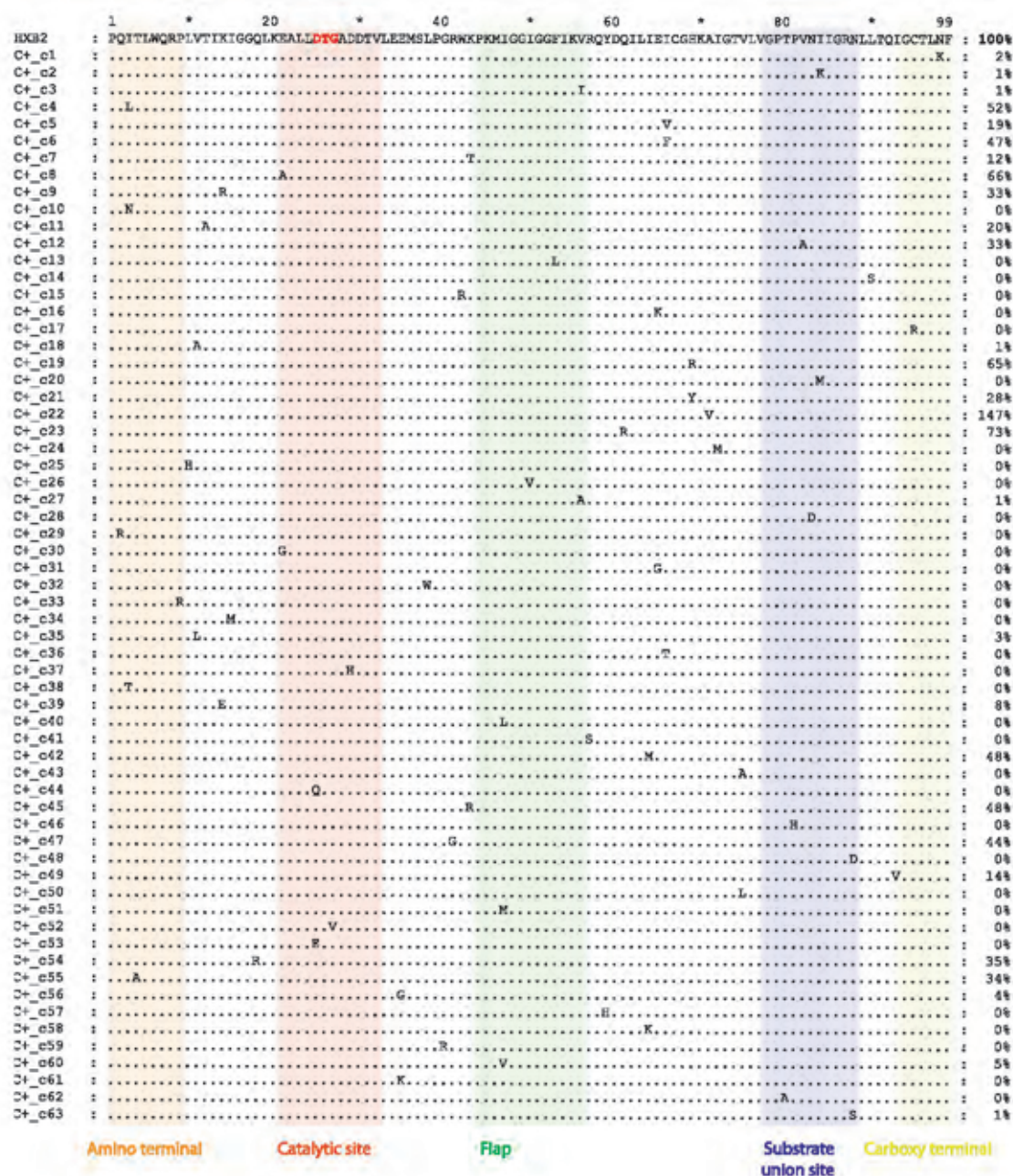
AA means amino acid.

In contrast, some neutral and beneficial mutations were located in the previously described conserved regions of the protease. Particularly, a neutral mutation, E21A, was located in the active site; 3 neutral and 1 beneficial mutations (P44L, M46V and M46I) were located in the flap region; 2 neutral mutations, V82A and I84M, were located in the substrate union site; and 2 neutral mutations, I2L and N98S, were located in the amino and carboxyl terminal regions. Interestingly, the beneficial mutations M46I is located in a very well known polymorphic position in the virus isolated from infected individuals (Ceccherini-Silberstein *et al.*, 2004; Wu *et al.*, 2003), which has been associated with resistance to protease inhibitors such as Atazanavir, Ritonavir, Indinavir, Nelfinavir and Tripanavir (Johnson *et al.*, 2011). The M46V mutation, with a beneficial effect in this study, has also been reported to be

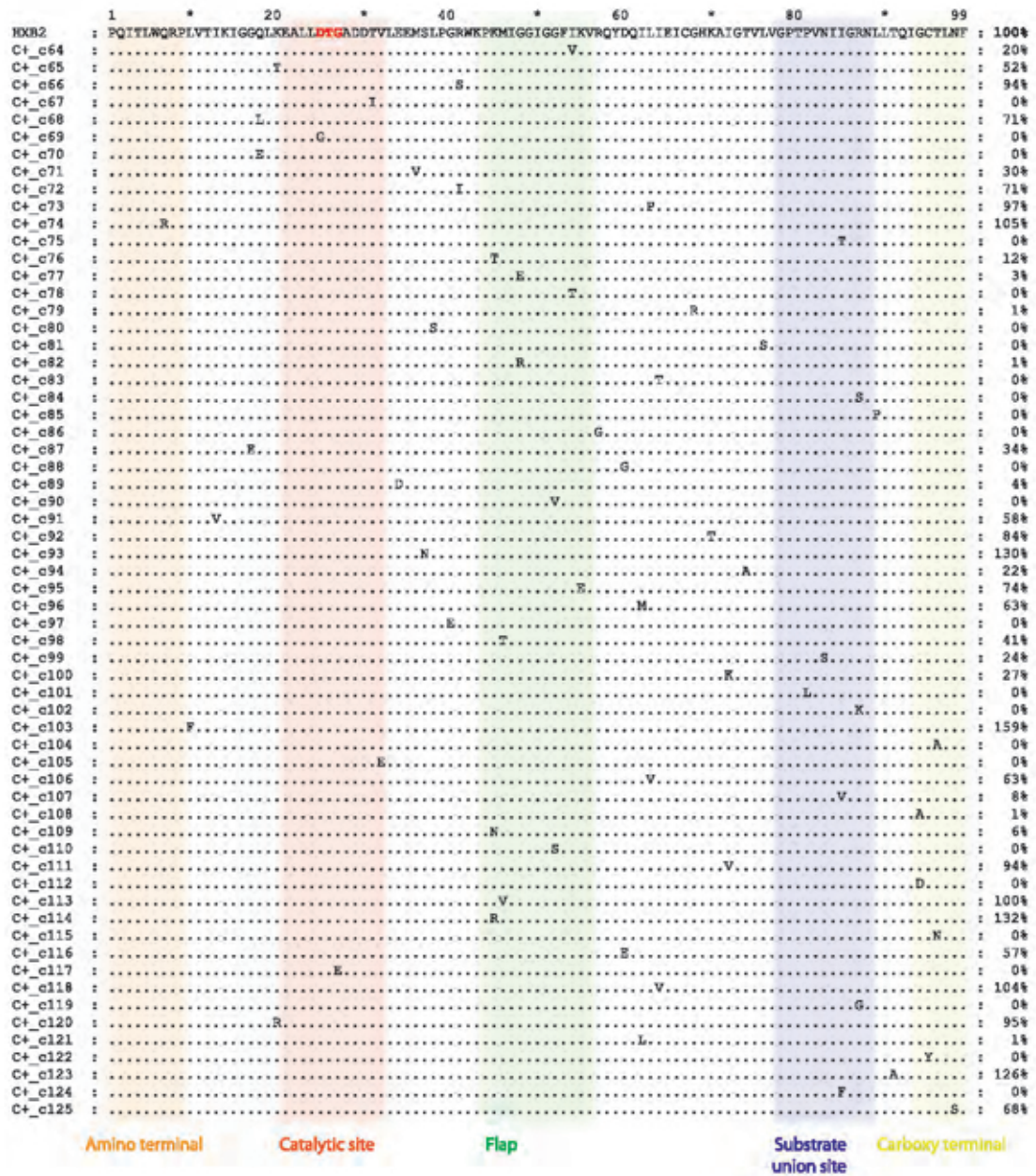
Results

associated with PI treatment (Wu *et al.*, 2003). Besides, the neutral amino acid mutation V82A is also located in a well known polymorphic position in the virus isolated from infected individuals, and has been associated with resistance to protease inhibitors such as Indinavir, Ritonavir and Lopinavir (Robert W Shafer, 2001). The mutation V82A was shown to have a positive impact on virologic response to ritonavir boosted darunavir in 2 independent data sets (Descamps *et al.*, 2009; Johnson *et al.*, 2011).

Taken altogether, these results show that the mutated 17a protease is as vulnerable as the wild type protease to the addition of single random amino acid mutations.



Results



Results

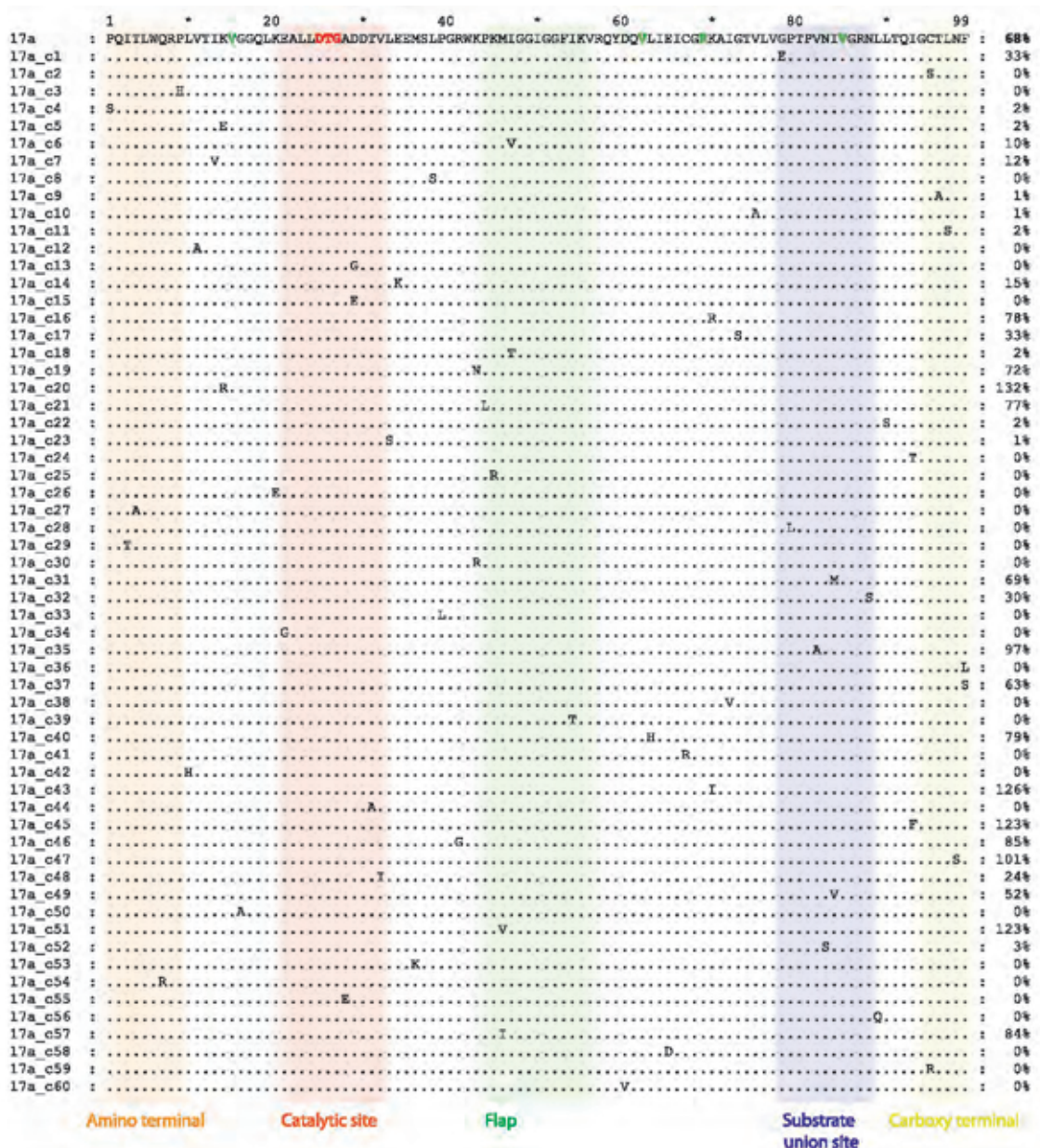


Figure 44: Amino acid sequence alignments of the HIV-1 protease single variants obtained from the wild type and 17a clones. Amino acid changes are indicated relative to the HIV-1 HXB2 sequence obtained from Los Alamos HIV Sequence Database in the case of the wild type single-variants, and to the 17a sequence in the case of the 17a single-variants. The two first blocks show the 125 clones obtained from the wild type protease, whereas the third block shows the 60 clones obtained from the 17a protease. On the left side of the sequences, their relative catalytic efficiency is shown. Dots indicate amino acid sequence identity. Red bold cases indicate the protease catalytic triad formed by Asp 25, Thr 26, and Gly 27; and green bold cases in the 17a sequence indicate the 4 amino acid differences of this protease with the wild type protease (I15V, I62V, H69R, and I85V). Color boxes indicate the protease conserved regions: orange, amino terminal; red, catalytic site; green, flap; blue, substrate union site; and yellow, carboxyl terminal.

Results

In order to confirm that the enzyme activities determined with the genetic screen were affecting, to a similar extent, the ex vivo replication capacities (fitness) (Miura *et al.*, 2008b), we introduced six different protease variants from the 17a mutant library, namely, 17a-K14R, 17a-D29G, 17a-E34K, 17a-M46V, 17a-K70I and 17a-I84V, in the HIV-1 infectious clone pHXB- Δ Pro (Maschera *et al.*, 1995). These proteases, which included the HIV-1 and 17a wild type template sequences, were chosen because they displayed heterogeneous (high, intermediate or low) catalytic efficiencies in the genetic screen. The in vitro generated mutated proteases were recombined with the protease-deleted HXB2 infectious clone. The protease region of the chimeric viruses was then sequenced from the resultant chimeric viral stocks and compared to the original source plasma. No differences at the nucleotide or amino acid level, respectively, were observed between in vitro generated proteases and chimeric viruses. Thus, suggesting that chimeric viruses represented the dominant form of mutated viruses and selection during culture was unlikely to affect the overall analysis.

The recombinant virus obtained from 17a clone 13, carrying the D29G substitution, did not grow sufficiently, giving very low titers even after amplification of the viral stock. Therefore, the ex vivo replication capacity assay could not be performed for this variant. To quantitatively compare the ex vivo viral RC of the resulting 7 chimeric viruses, CEM-GFP cells were infected and viral growth was measured by quantifying the GFP production on days 3 to 7 after infection (Fig 45).

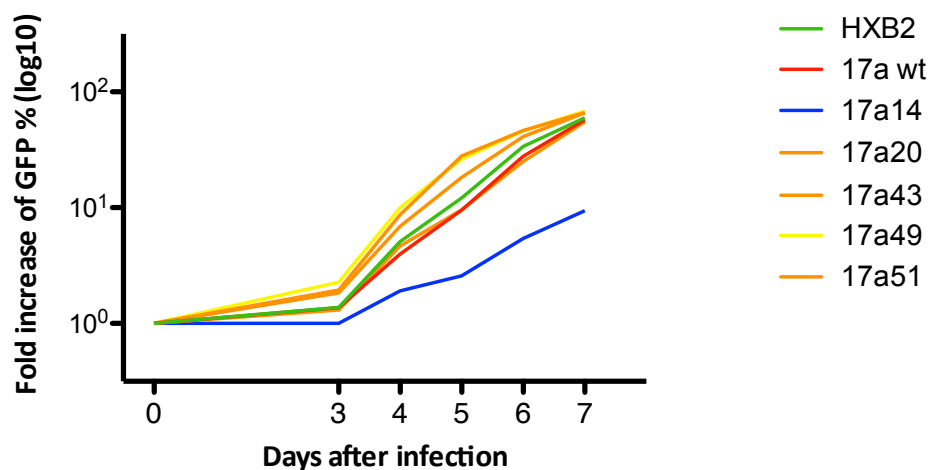


Figure 45: Kinetics of the viral replication capacity assay. Recombinant viruses are represented by different color lines. Green and red lines indicate the HIV-1 wild type (HXB2, positive control) or 17a protease recombinant viruses, respectively. Color lines represent the single-variant recombinant viruses with a low (blue), intermediate (yellow) or high (orange) proteolytic activity, respectively.

Results

Viral RC was calculated as the slope of the natural log of percent GFP-expressing cells between days 3 and 7 after infection (Table 14). The viral RC of the recombinant viruses was assayed in three independent experiments, and the mean viral RC values were calculated.

Table 14. Comparison of the in vitro catalytic efficiencies and ex vivo viral replication capacity of the HIV-1 mutated proteases.

Sample	Mutation	Phage λ	VRC assay
		Genetic screen	
		Mean \pm SD	Mean \pm SD
HXB2		100 \pm 0	100 \pm 2
17a wt		68 \pm 36	100 \pm 3
17a c13	D29G	0 \pm 0	-
17a c14	E34K	15 \pm 3	59 \pm 4
17a c20	K14R	132 \pm 28	92 \pm 2
17a c43	K70I	123 \pm 22	95 \pm 1
17a c49	I84V	50 \pm 20	88 \pm 1
17a c51	M46V	124 \pm 37	97 \pm 4

Although some differences in the percentages of the enzyme activity and fitness reduction were found between these two assays, the observed HIV-1 growth was proportional to the observed enzymatic activity of the corresponding protease (Fig 46). Thus, suggesting that the protease genetic screen system used here can be seen as a complement to the classical biochemical approach for monitoring protease proteolytic activity.

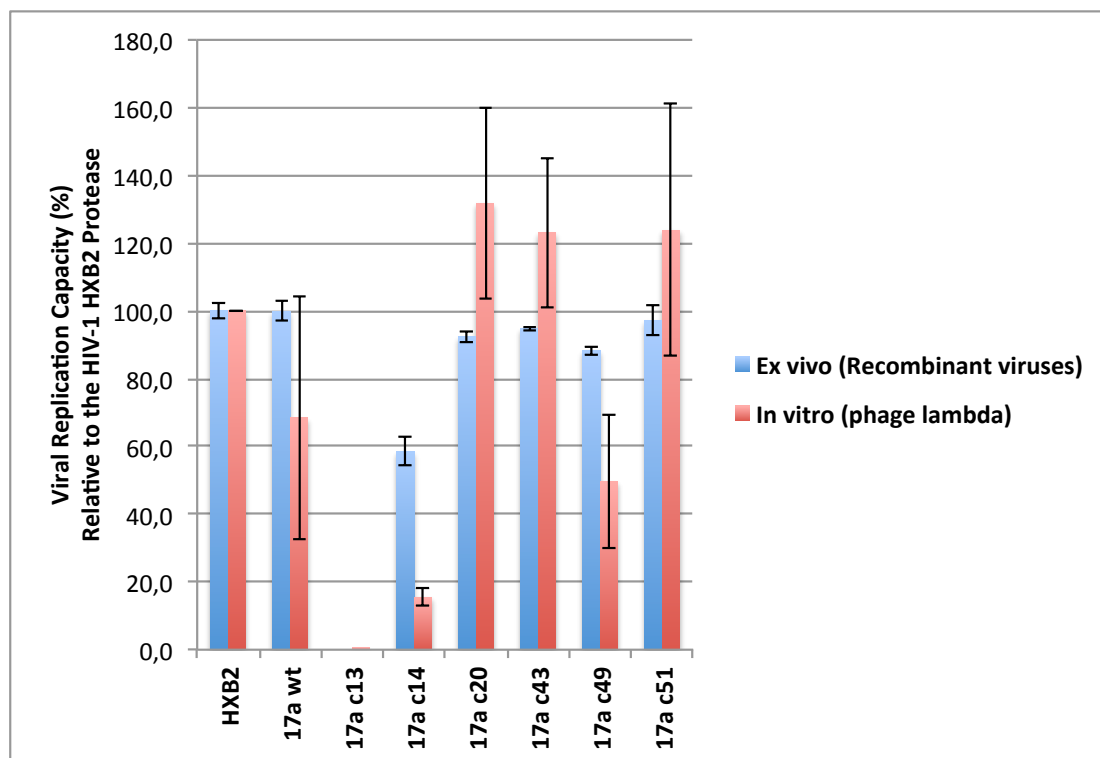


Figure 46: Proportionality between HIV-1 ex vivo replication capacity and in vitro catalytic efficiency. The catalytic efficiencies of eight protease variants (red bars), HXB2, 17a wild type, 17a-K14R, 17a-D29G, 17a-E34K, 17a-M46V, 17a-K70I and 17a-I84V, are compared to the replication capacities of HIV-1 infectious clones (blue bars) carrying these seven protease variants. The catalytic efficiencies and the replication capacities of the seven proteases are represented as percentages relative to that of the HIV-1 HXB2 protease (100%).