

DOCTORAL THESIS 2013

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

Elena Capel Malo

Hypothesis and Objectives

Hypothesis

HIV sequence conservation decreases over time, in contrast HIV keeps spreading and infecting new individuals. This work aims to decipher whether HIV genetic diversification is affecting the viral replication capacity. In particular, the effect of genetic diversification will be analysed in two enzymes critical for the viral replication cycle, the viral encoded protease and integrase.

Objectives

- ▶ Compare the sequence diversification of different HIV-1 genes: gag, protease and integrase over time.

- ▶ Investigate the impact of HIV-1 protease diversification on the ex vivo viral replication capacity (RC) and on HIV-1 protease mutational robustness.

- ▶ Investigate the impact of HIV-1 integrase diversification on the ex vivo viral replication capacity (RC).

- ▶ Explore the HIV-1 protease capacity to tolerate one amino acid substitution after a round of random mutagenesis.

Materials and Methods

Ex vivo studies

Patients

All the samples from HIV-1 infected patients used in this work have been obtained from the HIV Unit at “Germans Trias i Pujol” University Hospital in Badalona. All the samples had their plasma separated from peripheral blood mononuclear cells (PBMC) by the Sample Processing and Conservation Service from the IrsiCaixa Institute. Shortly, blood was first centrifuged at 1900 rpm during 10 minutes and supernatant was then centrifuged at 2600 rpm during 10 minutes. Plasma was stored at -80°C. PBMCs were isolated using a Ficoll gradient (Ficoll-Hipaque, Pharmacia), washed with phosphate buffered saline (PBS) and cryopreserved in liquid nitrogen with 90% Fetal Bovine Serum (FBS) and 10% Dimethyl Sulfoxide (DMSO).

HIV-1 RNA was quantified from the obtained plasma with the AMPLICOR HIV Monitor Test (Roche Diagnostics). And CD4+ T cell number from all patients was determined by flow cytometry.

For studying HIV protease, integrase and gag genes evolution we selected 150 samples from 150 HIV-1 infected patients at our HIV clinic. All patients were infected with HIV-1 subtype B. One hundred thirty nine samples were obtained from naïve patients and 11 samples were collected from protease inhibitor (PI) treated patients who were failing therapy.

Ex vivo viral replication capacity (VRC) determinations and mutational robustness experiments were performed on 33 samples in the protease study, whereas in the integrase study 94 samples were analysed.

Recovery and analysis of HIV-1 protease, gag and integrase sequences

RNA extraction and amplification was performed as described previously (Fernandez *et al.*, 2007; Ibáñez *et al.*, 2000; Parera *et al.*, 2004). Briefly, HIV-1 RNA was extracted from 140 µl of plasma and amplified by PCR using a nested primer set. After the viral RNA was isolated with the QIAamp Viral RNA Mini Kit (QIAGEN), 5 µl of resuspended RNA was reverse transcribed and PCR amplified using the SuperScript III One-Step RT-PCR System with Platinum *Taq* DNA Polymerase (Invitrogen) and 10 pmol of specific oligonucleotides to amplify the protease region: 5'prot1 (sense) (5'-

AGGCTAATTTTTAGGGAAGATCTGGCCTCC-3', Annex II, the position of the HIV-1 reference clone HXB2 were 2078 to 2109) and 3'prot1 (antisense) (5'-GCAAATACTGGAGTATTGTATGGATTTTCAGG-3', the position of the HIV-1 reference clone HXB2 were 2703 to 2734). A second amplification (nested PCR) was carried out with oligonucleotides 5'prot2 (sense) (5'-TCAGAGCAGACCAGAGCCAACAGCCCCA-3', HXB2 residues 2136 to 2163) and 3'prot2 (antisense) (5'-AATGCTTTTATTTTTCTTCTGTCAATGGCC-3', HXB2 residues 2620 to 2650). The amplified proteases were sequenced with the flanking oligonucleotides 5'prot2 and 3'prot2 using the Big Dye v3.1 kit and the 3100 DNA sequencing system (Applied Biosystems).

The HIV-1 gag coding region was PCR amplified as described previously (Miura *et al.*, 2008a). Briefly, 5 µl of extracted RNA were used for the reaction, which was performed using the SuperScript III One-Step RT-PCR System with Platinum *Taq* DNA Polymerase (Invitrogen) and 10 pmol of oligonucleotides: 5'gag1 (sense) (5'-AAATCTCTAGCAGTGGCGCCCGAACAG-3', HXB2 residues 623 to 649) and 3'gag1 (antisense) (5'-TAACCCTGCGGGATGTGGTATTCC-3', HXB2 residues 2826 to 2849). A second amplification (nested PCR) was carried out with 5'gag2 (sense) (5'-GACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAATTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGG-3', HXB2 residues 695 to 794) and 3'gag2 (antisense) (5'-GGCCCAATTTTTGAAATTTTCTTCTTTCCATTTCTGTACAAATTTCTACTAATGCTTTTATTTTTCTTCTGTCAATGGCCATTGTTAACTTTTG-3', HXB2 residues 2605 to 2704). Gag sequencing was performed with the Big Dye v3.1 kit and the 3100 DNA sequencer (Applied Biosystems) using oligonucleotides s5'gag2 (5'-GGGCGGCGACTGGTGAGT-3', HXB2 residues 732 to 749), s3'gag2 (5'-CTTTATTGTGACGAGGGTTCG-3', HXB2 residues 2274 to 2294), 5protRv (5'-AGGGGTCGTTGCCAAAGAGTG-3', HXB2 residues 2281 to 2261), and 1Bwt (5'-AGGTGGATTATTTGTCATCCATCCTATTTGTTCTGAAGG-3', HXB2 residues 1515 to 1554).

The HIV-1 integrase coding region was PCR amplified with specific oligonucleotides. Briefly, 5 µl of extracted RNA were used for the reaction, which was performed using the SuperScript III One-Step RT-PCR System with Platinum *Taq* DNA Polymerase (Invitrogen) and 10 pmol of oligonucleotides: 5i1 (sense) (5'-GGAGGAAATGAACAAGTAGAT-3', HXB2 residues 4174 to 4194) and 3i1

(antisense) (5'- GGGATGTGTA CTTCTGAAC -3', HXB2 residues 5193 to 5211). A second amplification (nested PCR) was carried out with 5I2 (sense) (5'-TTTTTAGATGGAATAGATAAGG-3', HXB2 residues 4228 to 4249) and 3I2 (antisense) (5'-TAATCCTCATCCTGTCTAC-3', HXB2 residues 5075 to 5093). Integrase sequencing was performed with the Big Dye v3.1 kit and the 3100 DNA sequencing system (Applied Biosystems) using the flanking oligonucleotides 5I1 and 3I1, and the oligonucleotides 5Irv_int (5'- TATGCTGTTTCCTGCCCTGT-3', HXB2 residues 4504 to 4523), and 3IFw_int (5'- CAGGGACAGCAGAAATCCAC -3', HXB2 residues 4908 to 4927).

Sequence alignment and editing was performed with the Sequencer software program, version 4.1 (GeneCodes). Nucleotide and amino acid p-distances were calculated using the MEGA 5 software package (Tamura *et al.*, 2011), as well as the nucleotide composition of the different HIV-1 genes (protease, gag and integrase). Protein nucleotide sequence trees were drawn by the neighbour-joining method. The HIV-1 protease and integrase subtype B ancestral sequences were obtained from the Los Alamos HIV sequence database (<http://www.hiv.lanl.gov>). Amino acid entropy was calculated using the ENTROPY-TWO software program (<http://www.hiv.lanl.gov>). ENTROPY-TWO compares two sets of aligned DNA or protein sequences and uses the Shannon Entropy as a measure of the variation in the alignments. To determine possible selective pressures, the proportion of synonymous substitutions to potential synonymous sites and the proportion of non-synonymous substitutions to potential non-synonymous sites were calculated using the SNAP software program (<http://www.hiv.lanl.gov>). To estimate codon-specific selection pressures, we used the fixed effects likelihood (FEL) method, which directly estimates non-synonymous and synonymous substitution rates at each site (Pond & Frost, 2005). The HKY85 nucleotide substitution bias model was used with FEL.

Identification of integrase polymorphisms: statistical methods

Polymorphic sites were estimated performing a maximum likelihood (ML) analysis of natural selection codon-by-codon of the integrase sequences. For each codon, estimates of the numbers of inferred synonymous (s) and nonsynonymous (n) substitutions were presented along with the numbers of sites that were estimated to be synonymous (S) and nonsynonymous (N). These estimates are produced using the joint Maximum Likelihood reconstructions of ancestral states under a Muse-Gaut model (Muse & Gaut, 1994) of codon substitution and General Time Reversible

model (Nei & Kumar, 2000) of nucleotide substitution. For estimating ML values, a tree topology was automatically computed. The test statistic $dN - dS$ is used for detecting codons that have undergone positive selection, where dS is the number of synonymous substitutions per site (s/S) and dN is the number of nonsynonymous substitutions per site (n/N). A positive value for the test statistic indicates an overabundance of nonsynonymous substitutions. In this case, the probability of rejecting the null hypothesis of neutral evolution (P-value) is calculated (Pond & Frost, 2005; Suzuki & Gojobori, 1999). Values of P less than 0.05 are considered significant at a 5% level and are highlighted. Normalized $dN - dS$ for the test statistic is obtained using the total number of substitutions in the tree (measured in expected substitutions per site). It is useful for making comparisons across data sets. Maximum Likelihood computations of dN and dS were conducted using HyPhy software package (Pond *et al.*, 2005). The analysis involved 94 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 288 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011).

Generation of an integrase-deleted pNL4-3 plasmid

For generating the integrase recombinant viruses, a plasmid pNL4-3 Δ integrase was developed by inserting unique BstEII restriction sites at the 5' and 3' ends of the integrase gene (pNL4-3 nucleotide positions 3716 and 4350, HXB2 nucleotide positions 4337 and 4971, respectively) by site directed mutagenesis of the wild-type (wt) full-length HIV-1 molecular clone pNL4-3 (Adachi *et al.*, 1986) using overlap extension PCR. First, two fragments of pNL4-3 were amplified using specific primers containing BstEII restriction sites upstream the integrase gene and downstream the integrase gene (Fig 15, Annex II). The plasmid pNL4-3 has two unique restriction sites for the enzymes AgeI and EcoRI at nucleotide positions 2863 and 5121, respectively. The upstream region of the integrase gene was PCR amplified with Platinum *Taq* DNA Polymerase (Invitrogen) and 10 pmol of oligonucleotides: AgeIFw (sense) (5'- CTGGCAGAAAACAGGGAG -3', pNL4-3 residues 2834 to 2851) and BstEIIrV (antisense) (5'- TACCATTCTTTTGCTACTAC -3', pNL4-3 residues 3698 to 3718) (bold case indicates site mutation to introduce the BstEII restriction site; underline indicates BstEII restriction site). The downstream region of the integrase gene was PCR amplified with Platinum *Taq* DNA Polymerase (Invitrogen) and 10 pmol of oligonucleotides:

BstEIIFw (sense)

(5'- GTAGTAGCAAAGAAATGGTAACCGTAGTAATACAAGATAATAG -3', pNL4-3

residues 4349 to 4372) and EcoRIRv (antisense) (5'- GGATAAACAGCAGTTGTTGC -3', pNL4-3 residues 5128 to 5147) (bold case indicates site mutation to introduce the BstEII restriction site; underline indicates BstEII restriction site). Each of the fragments was gel purified and recombined by overlapping PCR using the external primers AgeIFw and EcoRIRv. The resulting PCR product, called pNL4-3.AgeIEcoRI and the pNL4-3 wt plasmid were then digested with both AgeI and EcoRI (New England Biolabs). In order to confirm the insertion of the BstEII restriction site, the PCR pNL4-3.AgeIEcoRI was further digested with BstEII (Roche). Finally, both PCR pNL4-3.AgeIEcoRI and pNL4-3 wt plasmid digested fragments were gel purified and ligated with T4 DNA ligase for 1 hour at 25°C (Fermentas, Thermo Scientific) to make the pNL4-3Δintegrase. The resulting ligation product was transformed into OmniMAX™ Competent Cells (Invitrogen) following the manufacturer's instructions. The transformation mix was plated in selective growth-plates containing tetracycline and ampicillin, and incubated at 37°C overnight. Eight clones were isolated and grown overnight with LB and ampicillin at 37°C. The next day, plasmid DNA was purified using the QIAprep Spin Miniprep Kit (QIAGEN) following standard procedures. Further digestion of the purified plasmid with AgeI and EcoRI confirmed the correct deletion of the integrase gene.

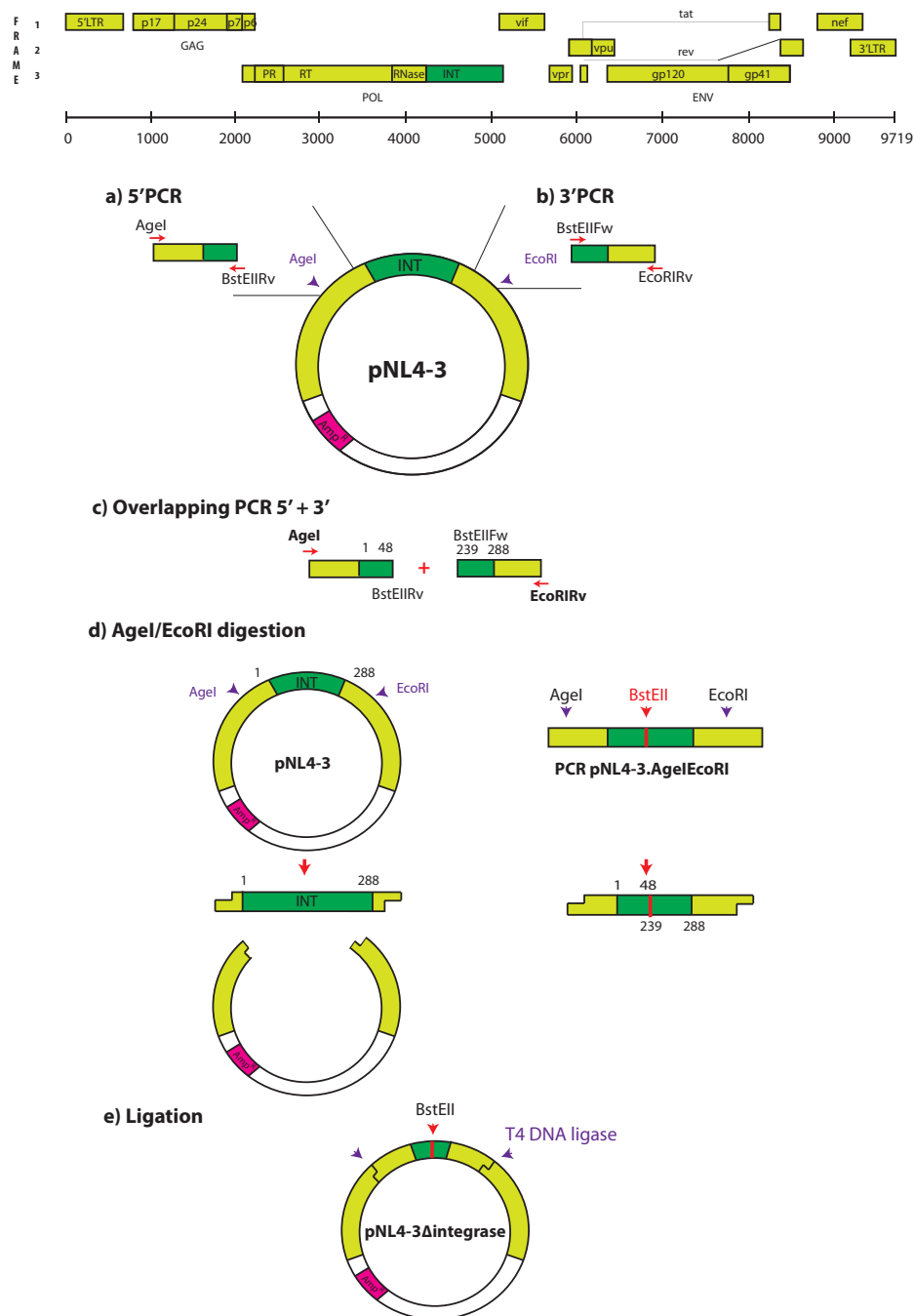


Figure 15: Construction of the integrase-deleted pNL4-3 clone by overlapping PCR. A schematic diagram of HIV-1 HXB2 genome is shown at the top of the figure. A schematic diagram of the pNL4-3 plasmid is shown below, indicating the unique *AgeI* and *EcoRI* restriction sites, the integrase fragment (dark green), the HIV genome (light green) and the ampicillin resistance (pink). a) Schematic representation of PCR of the upstream integrase region using primers *AgeI*Fw and *BstEII*Rv, that amplify the pNL4-3 region containing the *AgeI* restriction site and integrase amino acids 1 to 48. b) Schematic representation of PCR of the downstream integrase region using primers *AgeI*Fw and *BstEII*Rv, that amplify the pNL4-3 region containing the *EcoRI* restriction site and integrase amino acids 239 to 288. c) Overlapping PCR (5'+3') uses PCR fragments generated in a) and b) and oligonucleotides *AgeI*Fw and *EcoRI*Rv in order to obtain the deleted integrase region with the *AgeI* and *EcoRI* restriction sites. d) Digestion of pNL4-3 wt plasmid and of the c) PCR with *AgeI* and *EcoRI*. e) Ligation of the digestion products to obtain the pNL4-3Δintegrase with T4 DNA ligase.

Generation of protease and integrase recombinant viruses

Protease recombinant viruses were generated as described previously (Betancor *et al.*, 2010; Kusic *et al.*, 2011). Briefly, full-length PCR products were purified with the QIAquick PCR Purification Kit (QIAGEN), and 150 ng were cotransfected into 5×10^6 MT-4 cells (Annex III) with 1 μ g of protease-deleted HXB2 clone that had been linearized with BstEII (Roche) at 60°C for 2h for the protease samples (Maschera *et al.*, 1995) in 250 μ l of R10 Plus medium (RPMI medium with 10% fetal calf serum) by electroporation (exponential protocol, 250 V, 975 μ F) using a Bio-Rad GenePulser II instrument in single cuvettes. Cells were transferred to T25 flasks in 10 ml of R10 Plus medium.

Integrase recombinant viruses were generated as described above. Briefly, full-length integrase PCR products were purified with the QIAquick PCR Purification Kit (QIAGEN), and 150 ng were co-transfected into 5×10^6 MT-4 cells with 1 μ g of the integrase-deleted pNL4-3 clone, described before, that had been linearized with BstEII (Roche) at 60°C for 2h in 250 μ l of R10 Plus medium (RPMI medium with 10% fetal calf serum) by electroporation (exponential protocol, 250 V, 975 μ F) using a Bio-Rad GenePulser II instrument in single cuvettes. Cells were transferred to T25 flasks in 10 ml of R10 Plus medium. Cell culture supernatants were harvested when the HIV-1 p24 antigen concentration surpassed 1 μ g/ml and were stored at -80°C as viral stocks for subsequent RC experiments.

Construction of random HIV-1 protease mutation libraries

An error-prone PCR mutagenesis protocol was performed. PCR was carried out in 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM MgCl₂, 0.5 mM MnCl₂, 0.2 μ M of each oligonucleotide, 2.5 units Taq polymerase (Promega), and biased deoxynucleoside triphosphate concentrations (240 μ M deoxycytidine triphosphate/1 mM deoxythymidine triphosphate/240 μ M deoxyadenosine triphosphate/1 mM deoxyguanosine triphosphate, Promega), as described previously (Parera *et al.*, 2007). The cycling parameters were as follows: 50 \times (95°C, 30 s; 50°C, 30 s; 72°C, 10 min). Input DNA was 1 ng of the corresponding HIV-1 protease PCR DNA per 100 μ l reaction. The following PCR oligonucleotides were used: 5' prot 2 and 3' prot 2. The resulting PCR products were purified with the QIAquick PCR Purification Kit (Qiagen) and 150 ng were recombined with 1 μ g of protease-deleted HXB2 clone that had been linearized with BstEII as described above. To compute the mutant frequencies generated by the mutagenic PCR, PCR products were cloned into the pGEM-T Easy

vector (Promega), transformed using One Shot® Omni-MAX 2 T1 (Invitrogen), and subsequently sequenced using the ABI Prism Dye Terminator Cycle Sequencing Kit (Applied Biosystems) as described previously (Nevot *et al.*, 2011). The composition of the library was determined by nucleotide sequencing of the encoded HIV-1 protease gene in individual plasmid colonies.

Replication capacity assays

Protease RC assays were performed in triplicate. Cell culture supernatants were harvested on days 3, 5, and 7 after transfection. Under these culture conditions (cells cultured in R10 medium and incubated in a humidified atmosphere ($\geq 95\%$ RH) at 37°C and $5\% \text{CO}_2$), a robust and clear exponential curve can be plotted before it plateaued 6 to 7 days post-transfection. The slope of the natural log of per cent of p24 antigen production was calculated between days 0 and 7. To discard that the exponential increase of p24 in the supernatant was simply the accumulation of p24 from the transfected MT-4 cells without new infection, infectivity titers were also determined (data not shown). The natural log was used for the slope calculation, which is appropriate for an exponential growth curve. The RC of each variant was compared with that of the virus with the HXB2 recombinant protease (100%). Figure 16 shows a schematic representation of the generation of the HIV-1 protease recombinant viruses and the replication capacity assay.

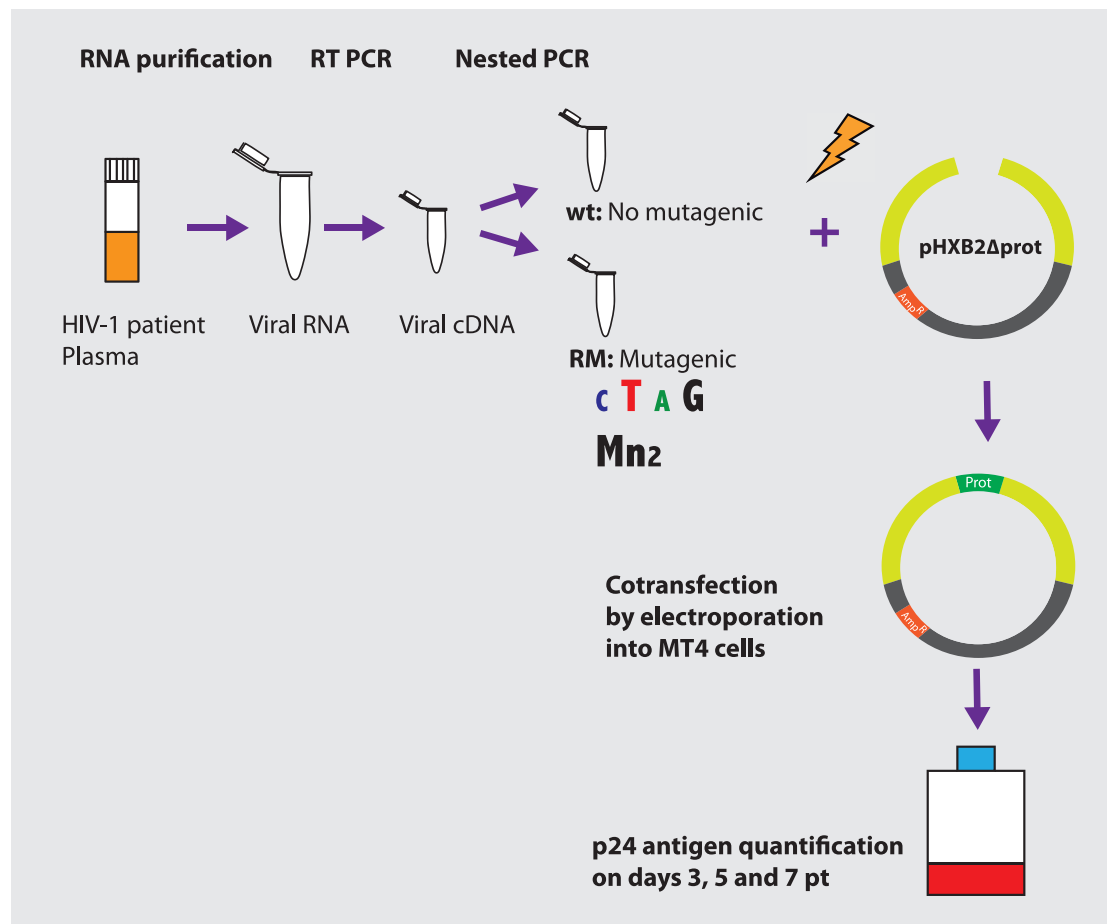


Figure 16: Schematic representation of the recombinant HIV-1 protease virus generation and RC assay.

Patients' integrase and in vitro generated mutated proteases RC assays were performed in triplicate. Cell culture supernatants were harvested when the HIV-1 p24 antigen concentration surpassed 1 $\mu\text{g}/\text{ml}$ and were stored at -80°C as viral stocks for subsequent RC experiments. Progeny virus was titrated in CEM-GFP cells (Annex III), a Tat-driven GFP reporter T-cell line that expresses GFP when infected by HIV-1 (Gervaix *et al.*, 1997), following the Reed and Muench method (Pannecouque *et al.*, 2008; Reed & Muench, 1938), where the 50% cell culture infective dose (CCID₅₀) value is calculated. Briefly, the virus stocks produced were ten-fold serial diluted in 96-well round-bottom plates. First, 25 μl of the viral stock were diluted in 100 μl of RPMI 10% FBS medium, resulting in a five-fold dilution. Last column was not infected in order to serve as a negative control. Then, 50 μl of exponentially growing CEM-GFP cells, prepared at 6×10^5 cells per ml, were added to each well, as well as 50 μl RPMI 10% FBS medium, reaching a ten-fold dilution in a final volume of 200 μl . Each plate contained 6 replicates of the virus stock. The plates were incubated in a humidified atmosphere ($\geq 95\%$ RH) at 37°C and 5% CO_2 for 5 days.

After the 5 days of incubation, the cells were examined microscopically for eventual HIV-induced cytopathogenic effect observed as green cells due to GFP expression. A well was scored positive if any trace of CPE was observed. The 50% cell culture infective dose (CCID₅₀) value was calculated using the Reed and Muench method. The calculation is as follows: $M = \text{inv log}\{x_1 + [(x_2 - x_1)((y_1 - 50)/(y_1 - y_2))]\}$, where y_1 = percent of wells scored positive closest to, but higher than, 50% at a certain virus dilution, y_2 = percent of wells scored positive closest to, but lower than, 50% at a certain virus dilution, x_1 = the log (dilution of the virus where y_1 was observed), x_2 = the log (dilution of the virus where y_2 was observed) and M = dilution of virus stock for 1 CCID₅₀.

Virus RC assays were performed as described previously (Brockman *et al.*, 2007; Miura *et al.*, 2008b; Schneidewind *et al.*, 2007) at a low multiplicity of infection (MOI) of 0.01, in order to minimize potential recombination events. Moreover the MOI of 0.01 was chosen based on initial experiments with this cell line and the recombinant viruses showing that, at this MOI, a robust and clear exponential curve could be plotted before reaching a plateau level of 30 to 40% at 7 to 8 days post-infection, which represents the saturation level in this system. An aliquot of 250 μ l for each culture was harvested 3, 4, 5, 6, and 7 days after infection and fixed in 1% formaldehyde in PBS. GFP expression was monitored by flow cytometry (FACSCalibur; BD Biosciences, San Jose, CA). The slope of the natural log of percent GFP-expressing cells was calculated between days 3 and 7 after infection. The natural log was used for the slope calculation, which is appropriate for an exponential growth curve. The RC of each variant was related with that of the virus with a positive control HIV-1 recombinant integrase or the HXB2 recombinant protease (100%) respectively. Figure 17 shows a schematic representation of the generation of the HIV-1 integrase recombinant viruses and the RC assay.

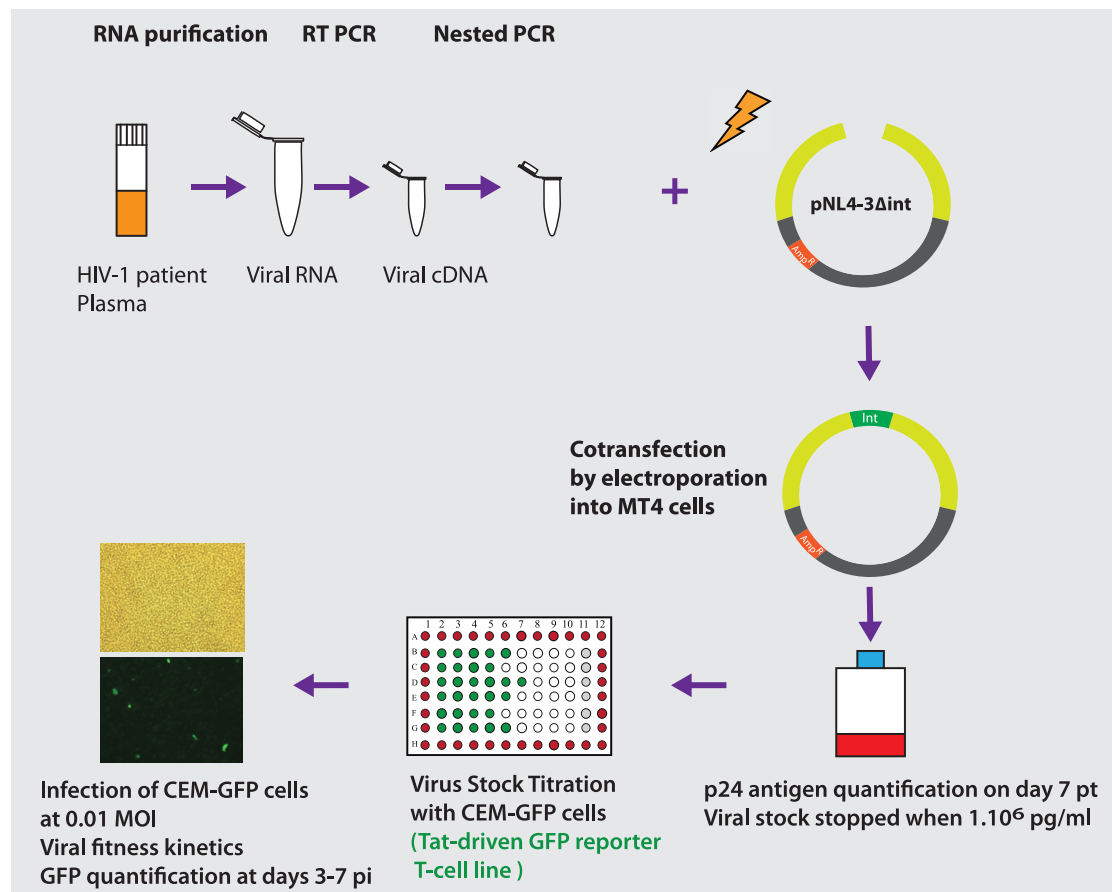


Figure 17: Schematic representation of the recombinant HIV-1 integrase virus generation and RC assay.

Re-sequencing of recombinant viral stocks

For all bulk (quasispecies containing) recombinant viruses, HIV-1 RNA was extracted from viral culture supernatant (QIAamp viral RNA kit, QIAGEN) on day 7 after transfection, amplified and sequenced using the same flanking oligonucleotides (5'prot2 and 3'prot2 for the protease recombinant viruses or 5I1 and 3I1 for the integrase recombinant viruses), sequenced using the Big Dye v3.1 kit and the 3100 DNA sequencer (Applied Biosystems), and compared to the original plasma HIV-1 RNA sequences (Annex IV). Overlapping regions of protease-deleted HXB2 clone with patient protease sequences and of integrase-deleted pNL4-3 clone with patient integrase sequences were verified, in order to confirm that the sequences from recombinant viruses corresponded to the corresponding plasma sequence. Nucleotide alignments were performed using Sequencer and maximum-likelihood phylogenetic trees were generated using MEGA 5 software.

Statistical analysis

Student's T-test was used to compare differences in RC between groups (e.g. early/recent infected; wild type/randomly mutated). The unpaired t test, linear regression analysis and Spearman's rank correlation coefficient of covariates were performed using GraphPad Prism version 4.00 for Windows (San Diego, CA, USA). The relationship between virus RC and the clinical parameters (CD4 and pVL) was assessed using Spearman's correlation from the GraphPad Prism version 4.00 for Windows as well.

Nucleotide sequence accession numbers

The GenBank/EMBL/DDBJ accession numbers for the HIV-1 protease and gag sequences reported in this work are JQ846091 to JQ846240. And the GenBank/EMBL/DDBJ accession numbers for the HIV-1 integrase early/recent sequences are KC238570 to KC238663.

In vitro studies

Construction of Random HIV-1 Protease Mutation Libraries in Lambda Phage

An error-prone polymerase chain reaction (PCR) mutagenesis protocol that generates complementary mutational spectra was performed. PCR was carried out in 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM MgCl₂, 0.5 mM MnCl₂, 0.2 μM of each oligonucleotide, 2.5 units Taq polymerase (Promega), and biased deoxynucleoside triphosphate concentrations (300 μM deoxycytidine triphosphate/1 mM deoxythymidine triphosphate/300 μM deoxyadenosine triphosphate/1 mM deoxyguanosine triphosphate, Promega), as described previously (Parera *et al.*, 2006; Vartanian *et al.*, 1996). The cycling parameters were as follows: 50 × (95°C, 30 s; 50°C, 30 s; 72°C, 10 min). Long elongation times were used to favor elongation after mismatches. Transition metal ions such as manganese (Mn²⁺) and cobalt (Co²⁺) may decrease the fidelity of DNA synthesis including PCR (Beckman *et al.*, 1985; MF *et al.*, 1983). Input DNA was 1 ng of a pBluescript SK plasmid containing a wild type HIV-1 protease DNA (HXB2 strain) or an HIV-1 protease mutated clone (17a) DNA obtained from a previous study (Parera *et al.*, 2006) per 100 μl reaction. The following PCR oligonucleotides were used: T3proL (5'-AATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGA ACTAGTGGATCCCCGGGCTGCAGGAATTCT**TCCTTTAACTTCCCTCAG**-3', Annex II, bold case indicates the residues of HIV-1 reference clone HXB2 in the position 2241 -

2258; underline indicates the EcoRI restriction site) and T7XHO (5'-TAATACGACTCACTATAGGGCGAATTGGGTACCGGGCCCCCCTCGAGTCAA**AGGCCATCCATTCTGGC**-3', bold case indicate the residues of HIV-1 reference clone HXB2 in the position 2588 - 2604; underline indicates the XhoI restriction site).

The resulting PCR products were purified with the QIAquick PCR purification kit (QIAGEN) and eluted in 35 µl of Buffer EB (10 mM Tris·Cl, pH 8.5), 25 µl of which were sequentially digested with EcoRI and XhoI restriction enzymes (FastDigest® Restriction Enzymes from Fermentas), 1 Unit of each with 1X FastDigest® Buffer during 15 and 5 minutes respectively at 37°C. Digested PCR fragments were run on a gel electrophoresis, then purified with the QIAquick Gel extraction kit (QIAGEN) and eluted in 50 µl of Buffer EB (10 mM Tris·Cl, pH 8.5) in order to discard the undesired PCR fragments and restriction enzymes. Digestion was ligated to lambda DNA (Uni-ZAP® XR Vector Kit, Stratagene) following the manufacturer's instructions. The ligations were packaged (Uni-ZAP® XR Gigapack® Cloning Kit, Stratagene), titrated, and amplified according to standard procedures, thus obtaining our recombinant phages.

The RecA⁻ *Escherichia coli* (*E. Coli*) host strain XL1-Blue MRF' cells (Annex III) were infected in order to obtain individual clones of the phage derived from packaging supernatants. These XL1-Blue MRF' cells were first grown during 15 hours at 30°C in the presence of 0.2% maltose and were centrifuged at 4500 rpm for 10 minutes at 4°C, in order to be infected. Cell pellet was resuspended with 10 mM MgSO₄ at an optical density at 600 nm (OD₆₀₀) of 2,0 per ml. 10 µl of a 10⁻¹ dilution of the packaging supernatant, containing the phage particles, were used to infect 200 µl of these XL1-Blue MRF' cells resuspended in 10 mM MgSO₄. They were incubated for 15 minutes at 37°C, plated into Bacto-Agar plates (Annex I) with tetracycline at 12.5 µg/ml and 3ml of Top-Agar (Annex I), and incubated for 6 hours at 37°C. At least 100 phages were recovered from each packaging of the different samples. Each lytic plaque was isolated and eluted in 100 µl SM (Annex I).

The composition of the library was determined by nucleotide sequencing of the encoded HIV-1 protease gene included in the individual phage colonies. Phage DNA from individual colonies was PCR amplified with the flanking oligonucleotides T3 (5'-AATTAACCCTCACTAAAGGG-3') and T7 (5'-TCGAGGTCGACGGTATC-3') of the phage taking 1 µl of phage supernatant as a template. The following concentrations were

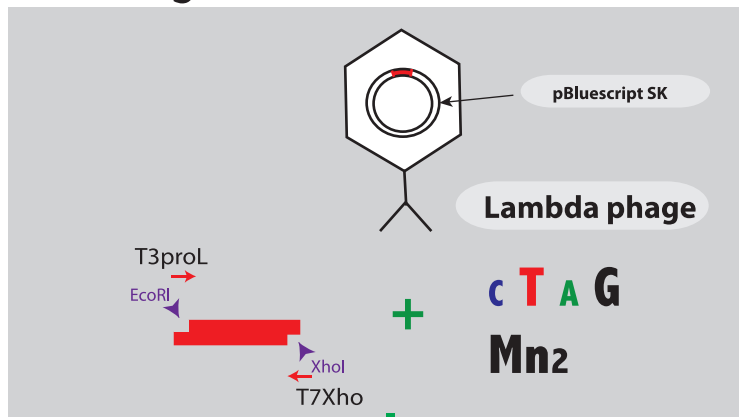
used: 1X PCR Buffer, 10 pmol from each primer, 5 mM MgCl₂, 200 μM dNTPs i 2U Taq DNA Polymerase (Promega). The cycling parameters were as follows: 2 minutes at 94°C followed by 35 cycles of 30 seconds at 94°C, 30 seconds at 55°C, 40 seconds at 72°C and a final extension of 7 minutes at 72°C. Resulting PCR products were purified with the enzyme “Exosap-IT clean-up” (Amersham Biosciences). Finally, phage DNA was sequenced with the flanking oligonucleotides T3 and T7 using the ABI Prism Dye Terminator Cycle Sequencing Kit (Applied Biosystems). Sequence alignment and editing was performed with the Sequencer version 4.1 (GeneCodes) software program. Figure 18 shows a schematic representation of the construction of the lambda phage library.

To quantify the mutant frequency generated by PCR mutagenesis, 367 individual protease clones from both HXB2 and 17a protease libraries were sequenced (Table 5). The mutagenic PCR used in this study had a mean mutation frequency per nucleotide of $9,41 \times 10^{-3} \pm 2,56 \times 10^{-3}$ (Table 5). Of the 125 mutants carrying unique single amino acid substitutions from the wild type HIV-1 protease analysed in this study, 97 were obtained from the previous cited study (Parera *et al.*, 2006), and 28 were obtained from the generation of the randomly mutated library as explained above.

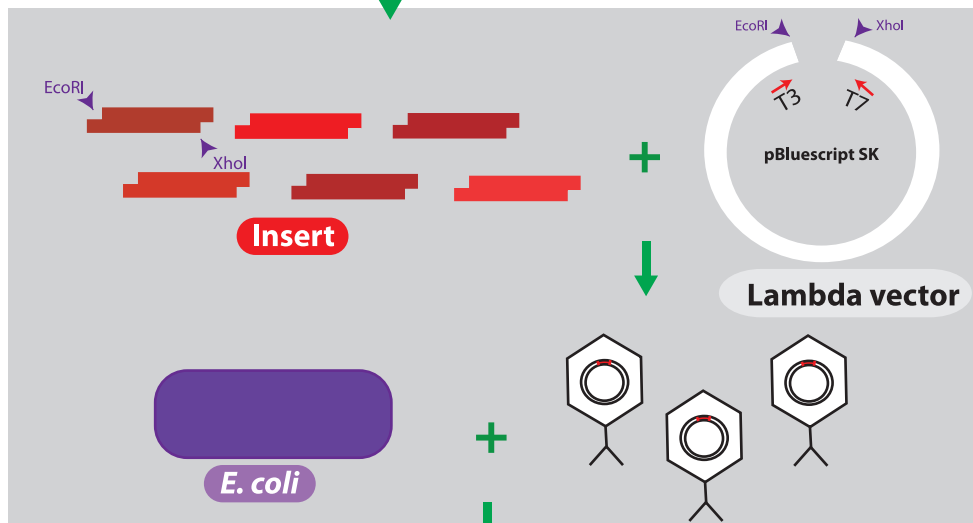
Table 5: Mean nucleotide mutation frequencies with the hypermutagenesis protocol HM300.

	Mean	SD	# Sequences	# Sequences with 1 AA substitution
C+	$1,03 \times 10^{-2}$	$2,07 \times 10^{-3}$	80	28
17a	$8,10 \times 10^{-3}$	$1,60 \times 10^{-3}$	287	60
Total	$9,41 \times 10^{-3}$	$2,56 \times 10^{-3}$	367	88

A. Mutagenic PCR



B. Clonation



C. Sequencing

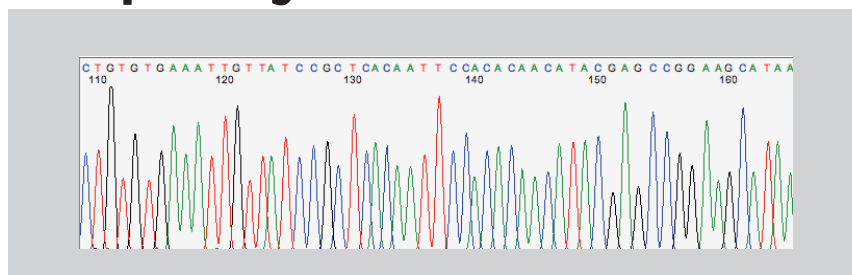


Figure 18: Schematic representation of the construction of a lambda phage library. (A) An error-prone PCR using biased dNTPs and manganese is performed taking DNA from a pBluescript containing the template HIV-1 protease. (B) Resulting PCR products are purified (not shown) and sequentially digested with EcoRI and XhoI restriction enzymes in order to be cloned in the lambda DNA (Uni-ZAP® XR Vector Kit, Stratagene) and then packaged. XL1 *E. coli* cells are infected with the obtained lambda phages in order to isolate individual plaques. (C) Finally, isolated individual phage colonies are sequenced in order to determine the library composition.

Production of lambda-phage stocks

To produce the phage stocks with the different proteases, XL1-Blue MRF' cells were grown with 0.2% maltose for 15 hours at 30°C, and then were centrifuged at 4500 rpm during 10 minutes at 4°C and were resuspended at an OD₆₀₀ of 2.0 with 10 mM MgSO₄. 200 µl of XL1-Blue MRF' cells were infected with 10 µl of the bacteriophages previously eluted in SM. They were incubated for 15 minutes at 37°C, plated with tetracycline at 12.5 µg/ml and 3ml of Top-Agar, and incubated for 6 hours at 37°C.

Titration of lambda-phage stocks

Once the different stocks were obtained, the amount of phages per µl of solution was measured. In order to determine the titer of each stock, the same cells and conditions described above were used, but this time 10 µl of the 10⁻⁵ and 10⁻⁶ dilutions were used to infect 200 µl of XL1-Blue MRF' cells. Serial dilutions 1/10 were done with SM solution, a buffered solution containing NaCl, MgSO₄, and gelatine (Annex I).

Determination of Protease Enzymatic Activities

The catalytic efficiencies of the different HIV-1 proteases were calculated using a bacteriophage lambda (λ)-based genetic screening, as previously described (Martinez *et al.*, 2000; Sices & Kristie, 1998) (Fig 19). Expression of the phage-encoded repressor (cI) results in repression of the bacteriophage's lytic functions. Induction from the lysogenic state is initiated by specific recA-mediated cleavage of the repressor, reducing the steady-state levels. The subsequent expression of cro results in the complete repression of cI expression, the induction of the phage's replicative functions, and the ultimate lysis of the host cell (Fig 19). A recombinant cI repressor containing a specific proteolytic processing site of the HIV-1 protease is expressed in *E. coli* and infected with a λ-cDNA expression library. The expression of a random cDNA has no effect on the function of the recombinant cI repressor (Fig 19 left), which prohibits the expression of the phage lytic replication functions. If, however, a phage encodes the appropriate specific protease, HIV-1 protease in this case (Fig 19 right), the target repressor is cleaved and the phage enters the lytic replication cycle, resulting in selective isolation of the appropriate phage. Thus, when infecting cells that contain this mutated repressor with a phage that expresses an HIV-1 protease, the enzyme activity can be estimated according to the number of plaques produced.

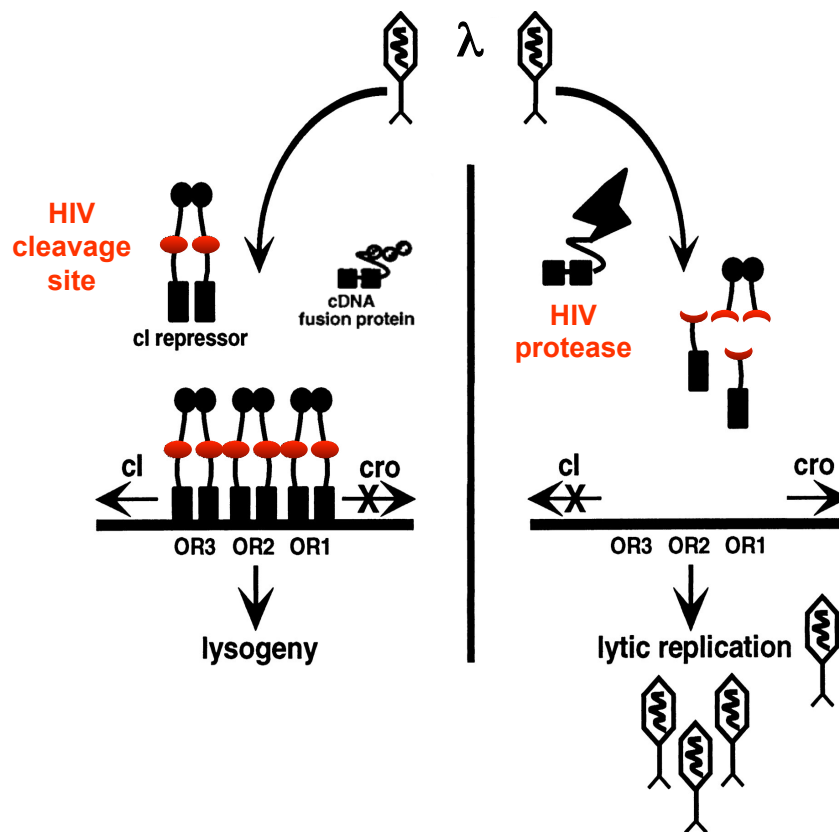


Figure 19: A genetic system for the isolation and characterization of HIV-1 proteases. (Left) The infecting phage encodes a random β -gal cDNA fusion protein that has no effect on the steady-state levels of the target repressor containing the specific proteolytic processing site of the HIV-1 (MA-CA), shown in red. (Right) The infecting phage encodes the HIV-1 protease, which is the appropriate site-specific protease that cleaves the recombinant repressor present in the *E. coli* p2X-cl.HIV/pcl.HIV-cro cells, ultimately resulting in lytic replication. Modified from (Sices & Kristie, 1998).

Briefly, *E. coli* JM109 cells (Annex III) containing the plasmid p2X-cl.HIV were transformed with plasmid pcl.HIV-cro. The resulting cells were grown in the presence of 0.2% maltose, harvested by centrifugation, and suspended to an OD₆₀₀ of 2.0 per ml in 10 mM MgSO₄. Cells (200 μ l) were infected with 5×10^7 pfu of phages containing the different HIV-1 proteases. After 15 min at 37°C, the cells were washed with 1 ml of 10 mM MgSO₄, harvested by centrifugation, and suspended in 1 ml of Luria broth (LB) medium containing 12.5 μ g of tetracycline, 0.2% maltose, 10 mM MgSO₄, and 0.1 mM isopropyl-beta-D-1-thiogalactopyranoside (IPTG). The cell cultures were incubated at 37°C for 2h30 and harvested by centrifugation. An additional cycle of selective growth was done by suspending the infected cells with a fresh aliquot (200 μ l) of JM109 pcl.HIV-cro cells. After 2 selective growth cycles, the titer of the resulting phage was determined by co-plating the cultures with 200 μ l of *E. coli* XL-1 Blue cells (OD₆₀₀ = 2.0/ml in 10 mM MgSO₄) on LB plates using 3 ml top

agar containing 12.5 µg of tetracycline/ml, 0.2% maltose, and 0.1 mM IPTG. After incubation at 37°C for 6 h, the plaques were counted for growth scores. The catalytic efficiency of every mutant was calculated as the mean \pm standard error of at least 3 independent replicates. The different activities were compared to the HXB2 reference clone activity. In addition, as a negative control, a construct carrying an inverted protease was used.