



Recombination Breakpoints Detected in C-Terminal Protease and N-Terminal RT Regions of HIV-1 Unique Recombinant Forms do not Influence Viral Replication Capability

Andrea Galli, Dr.
 Dipartimento di Scienze Cliniche "Luigi Sacco"
 Via G.B. Grassi, 74
 20157, Milano, ITALY
 Tel: +39-02-503.19775
 Fax: +39-02-503.19768
 galli_andrea@msn.com

Galli Andrea¹, Lai Alessia¹, Corvasce Stefano¹, Saladini Francesco², Riva Chiara¹, Dehò Lorenzo¹, Romano Laura², Galli Massimo¹, Zazzi Maurizio², Balotta Claudia¹

Department of Clinical Sciences "L.Sacco", University of Milan, Milan, Italy and Department of Molecular Biology, University of Siena, Siena, Italy

Background

It has been lately recognized that HIV-1 evolution, mainly due to the high viral replication rate and the lack of proof-reading activity of Reverse Transcriptase (RT), is further complicated by the ability of this enzyme to switch RNA template during transcription. The strand transfer activity has led to the generation of 20 Circulating Recombinant Forms (CRFs) and a growing number of Unique Recombinant Forms (URFs). Recombination seems to be a random process, yet some hot-spots have been identified in the HIV-1 genome. The identification of similar hot-spots in the *pol* region could be important for the understanding of fitness alteration in recombinant virions and for the development of an effective vaccine.

Methods

- We evaluated 38 URF sequences, encompassing 987 bases starting from position 1 of Protease gene (PR), obtained from 4 different molecular epidemiological studies and recorded into our own database. Specifically 31 URFs were from the ARCA (Antiretroviral Resistance Cohort Analysis) Genotype-Response Initiative, 5 from the European SPREAD (Strategy to Control SPREAD of HIV Resistance) Study, 1 from the European CASCADE (Concerted Action on SeroConversion to AIDS and Death in Europe) Study and 1 from a cohort of transsexual subjects followed at our Department.
- Sequences were obtained with home-made procedures or commercial kits and aligned using BioEdit 7.0.5 with the standard HIV-1 subtype reference alignment available on Los Alamos National Laboratory website, together with A3 (DDIJ579, DDJ360, DDJ369) and CRF09 (SN1795, SN7808, DE4057) reference strains. Bootscan analysis for recombination points was performed with Simplot software version 3.5.1, using a 300nt-wide window, sliding along the sequence in 10nt steps, with 100 bootstrap replicates in order to achieve reliable results in the phylogenetic trees. The same analyses were performed on a dataset of 50 URF sequences selected from the GenBank DataBase (GBDB) as an external control (Table 1).
- Since bootscanning data could be questionable for the first and last 150nt, in order to have reliable results beginning at base 1 of the original samples, we repeated the analysis using longer sequences (where available) encompassing 150 additional nucleotides up and downstream of the studied region. This control was performed on all 50 samples from GBDB and on 7 samples from our own database.
- To assess the significance of distributions, a Dispersion Test was applied.

Results

- A preliminary evaluation of the frequency of URFs with similar subtype composition in our and in GBDB URFs is given in Table 1. Despite the wide geographical distribution of URFs, the most representative forms of GBDB (A/G and B/F, but not A/D) were present also in our dataset.

Subtypes involved in URFs	# our URFs	# GBDB URFs
A/B	0	4
A/C	0	1
A/D	0	1
A/E	1	2
A/G	3	3
A/G/K	4	0
A/H	0	1
A/I	0	1
A/J	3	0
A/K	2	3
B/C	1	0
B/C/F	1	0
B/D	1	0
B/E	1	0
B/F	12	0
B/G	2	1
B/H	0	1
C/D	0	1

Results (continued)

- As indicated in Figure 1, a graphical representation of the observed recombination patterns hinted at a non-random location of recombination points.

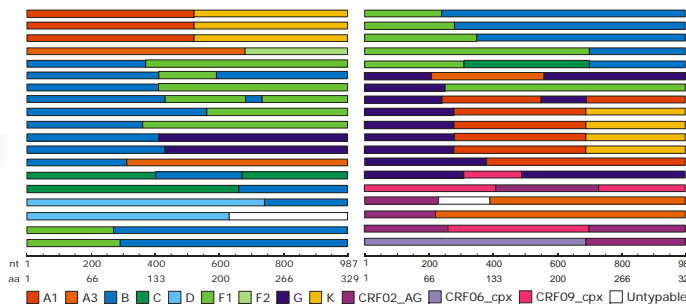


Figure 1. Recombination patterns of analyzed URFs

Twelve different subtypes and CRFs were found to be involved in the recombination patterns observed in our field's URFs. Recombination points are roughly spread all over the sequence. Nonetheless, a clustering can still be noticed around positions 400 and 700. A similar pattern was observed analyzing the samples from GenBank (data not shown).

- We then plotted the recombination points according to their position and frequency, in order to compare their layout to that of the regions encoding for the different protein domains. Counting the number of recombination points located in each region and normalizing according to their different lengths, we could observe that almost no recombination points are located in the PR N-terminal region (aa 1-62) ($p < 0.0001$). Furthermore, recombination points seem to cluster in two specific regions: the PR C-terminal region (aa 63-99) and the RT fingers domain (aa 1-85 and 118-155).
- Analyzing the URFs from GBDB analogous results were obtained. In particular, both clusters previously found in the RT fingers region were present, while the peak located in the PR C-terminal end, albeit still present, was slid upstream. These data supports our findings, showing an overall layout of breakpoints comparable with that observed in our samples and suggesting the presence of preferential recombination points in the *pol* region (Figure 2).
- When considering 5' and 3' extended sequences, our results indicate that, of the 57 samples for which longer sequences were available, only 4 showed additional breakpoints. Two recombination points were located before nucleotide 150 and 4 were found beyond nucleotide 900. Given the low number of such breakpoints and the extent of the corresponding domain regions, these data do not significantly alter our findings, particularly concerning the PR N-terminal region.

Table 1. Subtypes involved in analysed URFs

Recombinants composition is shown here with absolute number of samples and relative frequencies (in %) for each group of sequences. Our URFs are drawn in red and those from GBDB in blue.

Discussion

- Our data indicate that recombination events involving at least the first 150 amino acids of RT seem not to interfere with viral replication capability since all samples were obtained from patients with an active infection. No conclusions can be drawn regarding viral fitness or infectiveness since studied URFs could have been either generated in the individual host or acquired through infection.
- The absence of breakpoints involving the first 62 amino acids of PR would suggest that HIV-1 variants with such recombinations are unable to replicate efficiently. This could be caused by differences in the encoded proteic structures that cannot coexist or by sequence features favouring the introduction of mutations via slippage synthesis during the recombination process, resulting in an inactive enzyme. Alternatively, this region could lack the structural and/or sequence elements necessary for efficient recombination to occur. Noteworthy, this region includes the entire PR flap domain and the active site, two regions of utmost importance for enzyme activity. On the contrary, the high number of recombination points found in the C-terminal end of PR seems not to interfere with its function, even if this region is deputed to substrate recognition and binding.
- Two clusters of breakpoints found in the RT fingers domain could indicate the presence of recombination hot-spots caused by underlying RNA sequence or structural features, such as RNA hairpins or stop sites. These elements could, in turn, interfere with RT activity and favour the strand switch process. Interestingly, both regions have important functional elements downstream of the putative hot-spots. Given their relevance in the RT activity these features could be more conserved in their genetic sequence throughout HIV-1 subtypes, resulting in an enhanced interaction of templates during reverse transcription and production of a higher number of recombinants.
- Further studies are needed to verify our hypotheses, particularly regarding the identification of hot-spots and the molecular mechanism leading to their appearance. The understanding of RT recombination activity is crucial for the development of a vaccine, the design of new classes of drugs and a better understanding of HIV-1 evolution

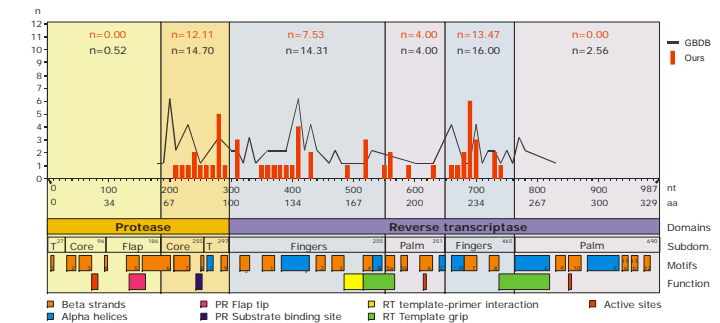


Figure 2. Recombination points aligned with protein domains.

In the upper part of the diagram recombination points detected in our and in GBDB URFs are displayed. Peaks indicate the number of breakpoints located at each position. In the lower portion the different protein domains and subdomains are shown. Relevant functional regions are also depicted.

The 6 vertical frames correspond to different protein regions, to highlight the relations between breakpoints cluster and proteic structural elements. Normalized number of recombination points is indicated at the top of each frame.