Different evolutionary patterns are found within human immunodeficiency virus type 1-infected patients

Concepción Casado,1 Soledad García,2 Carmen Rodríguez,2 Jorge del Romero,2 Gonzalo Bello1 and Cecilio López-Galindez1

1 Centro Nacional de Biología Fundamental, Instituto de Salud Carlos III, Majadahonda, Madrid 28220, Spain
2 Centro de Salud Sandoval, Comunidad Autónoma de Madrid, Madrid 28010, Spain

In order to study the evolution in vivo of human immunodeficiency virus type 1 (HIV-1) in patients with normal clinical evolution, six individuals were selected from a group of 46 patients followed for 1 to 4 years. Patients were selected not by clinical progression characteristics but on the basis of virus genetic variability, as analysed by heteroduplex mobility assay and RNase A mismatch cleavage method. Two patients displayed a homogeneous virus population, two showed very heterogeneous quasispecies and two presented two distinct variants within the virus population. Virus quasispecies were studied by nucleotide sequencing of the C2-fusion domain of the env gene. Virus evolution was approached by analysing the distribution of genetic distances, calculation of divergence and heterogeneity as well as the \( K_a/K_r \) ratio and by the construction of the phylogenetic trees. Three patients displayed the same tree topology, characterized by the presence of independent clades supported by high bootstrap values, whereas this pattern was not present in the other three patients. In the three patients displaying independent clades, a recombination analysis was carried out between distinct subpopulations and recombinant variants were identified. In one patient of this group, different selective pressures were detected in distinct virus clades, measured by their corresponding \( K_a/K_r \) ratios, revealing that different evolutionary forces are occurring at the same time within the same patient. These results show that multiple evolutionary patterns can be found in typical HIV-1-infected patients.

Introduction

Human immunodeficiency virus type 1 (HIV-1) infection can be characterized as a chronic infection because of the continuous virus replication in the host (Ho et al., 1995; Wei et al., 1995), which allows the evolution of HIV-1 populations. It has been estimated that about 300 rounds of replication occur per year in infected patients (Coffin, 1995). Different mechanisms have been invoked to explain HIV-1 evolution in vivo, including recruitment of latently infected CD4+ cells and T cell activation (Cheynier et al., 1994, 1998), positive selection of variants (Larder et al., 1990; Nixon et al., 1988; Wolinsky et al., 1996) and also the effect of negative forces. Moreover, non-selective mechanisms such as random drift (Plikat et al., 1997), founder effects (Cheynier et al., 1994), bottlenecks and Muller’s ratchet (Yuste et al., 1999) and compartmentalization (Wong et al., 1997) are also operating.

The error-prone replication system of HIV-1, the presence of two genomes in the virion, which favours recombination (Jetzt et al., 2000), the high virus turnover and elevated population size result in the presence of multiple variants within patients, described as virus quasispecies (Eigen & Biebricher, 1988; Meyerhans et al., 1989). There are reports of the existence of distinct variants within infected patients, especially in different compartments of the body (Ball et al., 1994; Delwart et al., 1998; Korber et al., 1994; Wong et al., 1997; Zhu et al., 1996), but even in the same organ (Cheynier et al., 1994; Delassus et al., 1992). Variants within infected patients can belong to the same subtype (Sala et al., 1995) but also to distinct subtypes, as shown by the existence of multiple intersubtype recombinants (Sabino et al., 1994) or intergroup M/O recombinants (Takehisa et al., 1999).

The evolution of HIV-1 quasispecies within patients has been explored in several studies that have analysed the degree of genetic variation of HIV-1 in relation to disease progression (Bagnarelli et al., 1999; Delwart et al., 1997; Ganeshan et al., 1997; McDonald et al., 1997; Salvatori et al., 1997;
Shankarappa et al., 1998, 1999; Shioda et al., 1997). However, discordant results were obtained when the contribution of selective pressures, as measured by the \(K_s/K_a\) ratio, to virus evolution was studied in relation to disease progression. The only repeated finding was a positive correlation between progression to AIDS and a homogenization in virus quasi-species (Ganeshan et al., 1997; Liu et al., 1997; McDonald et al., 1997; Shioda et al., 1997; Wolinsky et al., 1996; Yamaguchi & Gojobori, 1997; Zhang et al., 1993). Recently, a general pattern of divergence (considered as the evolution from an initial founder strain) and diversity (the breadth of the virus population at a given time-point) has been described (Shankarappa et al., 1999).

The evolution time of HIV-1 infection has permitted the classification of patients into slow, typical or rapid progressors. In this study, we have analysed virus genomes from a group of 46 patients with typical patterns of clinical evolution. From these patients, six individuals were selected on the basis of virus genetic variability and they were submitted to quasi-species analysis from two to three samples obtained over 1 to 4 years. The patterns of virus evolution obtained were not the same in all the patients selected, pointing to the existence of different virus and host factors contributing to the evolution of HIV-1 quasispecies in vivo.

**Methods**

**Patients.** Forty-six patients from an outpatient clinic (Centro de Salud Sandoval, CAM, Madrid) were followed for 1 to 4 years. Most of the patients were asymptomatic at the beginning of the study. Clinical data have been described previously (Casado et al., 2000a) and data on the patients analysed at the quasispecies level are shown in Table 1.

**Clinical samples and separation of viral nucleic acids.** Peripheral venous blood was collected on EDTA. Plasma was prepared by centrifugation at 3000 g for 10 min at 4 °C and stored at −80 °C until further analysis. Peripheral mononuclear cells (PBMC) were recovered after centrifugation over a Ficoll–Hypaque density gradient, washed twice in PBS and resuspended in 10% DMSO and 90% foetal calf serum for cryopreservation in liquid nitrogen until further testing. Proviral DNA was obtained from 5 × 10⁶ cells by standard phenol extraction (Peruco et al., 1981) and genomic RNA was isolated from 200 μl plasma according to Boom et al. (1990).

**Quantification of plasma viraemia.** The HIV-1 RNA copy number in plasma was quantified with the Amplicor HIV Monitor test kit following the manufacturer’s protocols (Roche Diagnostic Systems) (Mulder et al., 1994).

**PCR amplification, cloning and sequencing.** Three initial amounts of proviral DNA, 0.1, 1 and 5 μg, were amplified in a nested PCR. In order to avoid bottlenecking, the DNA concentration used in the study was the second dilution that gave a positive band in the gel. The first amplification covered almost the complete gp160 and, in a second amplification, the C2-fusion peptide region was obtained. The sequences of primers 60EU, 50EBND and 27EU have been described previously (Casado et al., 2000a, b). The sequence of primer 22ED is 5’-AGACAATAATTGTCTGGCCTGTACCGT-3’, numbers correspond to positions in the NL4-3 genome (GenBank accession number M19921). The first-round primers, 60EU and 50EBND, were used with the following cycling conditions: 94 °C for 5 min; 35 cycles of 94 °C for 1 min, 55 °C for 1 min and 72 °C for 2 min; and a final incubation at 72 °C for 10 min. In the second-round PCR, 2 μl of the first-round products was amplified in fresh reaction buffer plus primers 27EU and 22ED with the same cycling conditions, except that the length of the elongation step at 72 °C was 1 min. Titan RT–PCR (Promega) was used to amplify RNA samples. In this method, cDNA synthesis as well as the first-round PCR were performed in one step, without the requirement for the addition of reagents between cDNA synthesis and first-round PCR. Primers and cycling conditions were the same as in nested PCR for PBMC samples, with the addition of a preliminary step of 30 min at 50 °C in the first PCR. PCR products were cloned into a TA vector (Invitrogen) and 15 to 20 clones per clinical sample (clones derived from proviral DNA or genomic RNA) were sequenced directly with primers 27EU and 96ED (5’-AGACAATAATTGTCTGGCCTGTACCGT-3’), spanning a region that includes part of the C2 conserved region, the V3, V4 and V5 hypervariable regions, the CD4”-binding domain and the gp120–gp41 cleavage domain. A physical separation of the different amplification steps was established to avoid cross-contamination, but replicate control amplifications with no template were also included in every PCR experiment to test carry-over contamination. DNA sequencing was performed by using the ABI PRISM Dye Terminator cycle sequencing ready reaction kit (Perkin-Elmer) according to the manufacturer’s instructions. Sequencing reactions were resolved by electrophoresis on a 5% polyacrylamide gel in an ABI PRISM 377 automated sequencer (Perkin-Elmer).

**Heteroduplex mobility assay (HMA).** The first sample from each patient was analysed by HMA by using the second-round PCR products obtained. The PCR was then adjusted to 0-1 M NaCl, 10 mM Tris–HCl, pH 7.8, and 2 mM EDTA, using a 10 × DNA-annealing solution stock. Heteroduplex formation was maximized by denaturing 9 μl PCR products at 94 °C for 2 min and rapid cooling in wet ice. Heteroduplexes were then resolved in a 5% polyacrylamide gel in TBE at 200 V for 5 h (Delwart et al., 1993). Gels were stained with ethidium bromide, illuminated with UV and photographed.

**Phylogenetic analysis of sequences.** Sequence editing and assembling were performed using the program CHROMAS 1.45 (C. McCarthy, Griffith University, Southport, Australia). All alignments of nucleotide sequences were performed with CLUSTAL W (Thompson et al., 1994) and later edited manually. All positions with an alignment gap in at least one sequence were excluded from pairwise comparisons. Phylogenetic reconstructions were generated by using the program packages PHYLP (Felsenstein, 1993) and MEGA (Kumar et al., 1993). Pairwise evolutionary nucleotide distances were estimated with Kimura’s two-parameter model with a transition/transversion ratio of 2. Phylogenetic trees were constructed from the same distance matrices with the neighbour-joining algorithm; robustness of the trees was evaluated by bootstrap analysis on 1000 replicates. Phylogenetic trees were edited using TreeView version 1.5 (Page, 1996). Intrasample (heterogeneity) and intersample (divergence) sequence variations were respectively expressed as the mean distance for all pairwise comparisons between sequences within a sample or from two different samples. Rates of synonymous nucleotide substitutions per synonymous site (\(K_s\)) and non-synonymous substitutions per non-synonymous site (\(K_a\)) were estimated by the method of Nei & Gojobori (1986) as implemented in the MEGA program.

First, a phylogenetic identification of the viruses infecting each patient was carried out. For this analysis, several clones of the virus quasispecies of each individual patient were selected, as well as a sequence of virus MN for use as an outgroup. Phylogenetic trees were established by the neighbour-joining method as well as the maximum-parsimony and
Table 1. Clinical data from patients analysed at the quasispecies level

For those patients for whom the seroconversion date was missing, the infection date is the likely period during which infection occurred; for these patients, the upper and lower limits of
the time since infection are indicated. The numbers of CD4<sup>+</sup> and CD8<sup>+</sup> lymphocytes per ml of blood at the time of bleeding are indicated. Virus load is expressed as the number of vRNA copies per ml of plasma. Clinical status of patients is given according to the CDC classification (Anonymous, 1992). Antiviral treatments given to patients at the time of sample bleeding
are given: AZT, zidovudine; 3TC, lamivudine; Ind, indinavir; ddl, dideanosine; ddC, zalcitabine.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Infection date</th>
<th>Sampling date</th>
<th>Time since infection (years)</th>
<th>Lymphocytes (cells/ml)</th>
<th>Virus load (vRNA copies/ml)</th>
<th>Clinical status</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CD4&lt;sup&gt;+&lt;/sup&gt;</td>
<td>CD8&lt;sup&gt;+&lt;/sup&gt;</td>
<td>CD4/CD8</td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I01</td>
<td>1981–83</td>
<td>1993</td>
<td>10–12</td>
<td>160</td>
<td>1546</td>
<td>0.103</td>
<td>21000</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td></td>
<td>14–16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H16</td>
<td>1982–91</td>
<td>1993</td>
<td>2–11</td>
<td>505</td>
<td>776</td>
<td>0.65</td>
<td>&lt;2000</td>
</tr>
<tr>
<td></td>
<td>1994</td>
<td></td>
<td>2–12</td>
<td>330</td>
<td>802</td>
<td>0.41</td>
<td>29000</td>
</tr>
<tr>
<td>IH03</td>
<td>1979–84</td>
<td>1994</td>
<td>10–14</td>
<td>1148</td>
<td>2712</td>
<td>0.42</td>
<td>3300</td>
</tr>
<tr>
<td></td>
<td>1998</td>
<td></td>
<td>14–19</td>
<td>852</td>
<td>1289</td>
<td>0.66</td>
<td>1100</td>
</tr>
<tr>
<td></td>
<td>... Means</td>
<td></td>
<td>11–4</td>
<td>526</td>
<td>1281</td>
<td>0.45</td>
<td>12933</td>
</tr>
<tr>
<td>Group 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I09</td>
<td>1989–92</td>
<td>1993</td>
<td>1–4</td>
<td>572</td>
<td>922</td>
<td>0.62</td>
<td>51000</td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td></td>
<td>4–7</td>
<td>236</td>
<td>1187</td>
<td>0.20</td>
<td>17000</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td></td>
<td>5–8</td>
<td>308</td>
<td>745</td>
<td>0.41</td>
<td>1800</td>
</tr>
<tr>
<td>H16</td>
<td>1991</td>
<td>1993</td>
<td>2</td>
<td>878</td>
<td>1150</td>
<td>0.76</td>
<td>10000</td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td></td>
<td>5</td>
<td>641</td>
<td>1356</td>
<td>0.47</td>
<td>25000</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td></td>
<td>6</td>
<td>542</td>
<td>940</td>
<td>0.57</td>
<td>57000</td>
</tr>
<tr>
<td>IH07</td>
<td>1987–92</td>
<td>1995</td>
<td>3–8</td>
<td>509</td>
<td>1150</td>
<td>0.44</td>
<td>73000</td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td></td>
<td>4–9</td>
<td>168</td>
<td>1009</td>
<td>0.07</td>
<td>250000</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td></td>
<td>5–10</td>
<td>68</td>
<td>–</td>
<td>–</td>
<td>230000</td>
</tr>
<tr>
<td></td>
<td>... Means</td>
<td></td>
<td>5–2</td>
<td>405</td>
<td>932</td>
<td>0.52</td>
<td>136422</td>
</tr>
</tbody>
</table>
maximum-likelihood methods with similar branching patterns, as implemented in the PHYLIP package. The statistical value of the tree was evaluated by bootstrap analysis of 1000 replicas as described in the previous paragraph. The analysis showed that all isolates formed independent clades in the tree, supported by highly statistically significant bootstrap values (> 95%; data not shown), confirming the lack of cross-contamination. Five of the six viruses formed a group of sequences that included the MN sequence, identifying the set as subtype B. Virus I16 displayed a large genetic distance from this group, as expected for a subtype F virus (Casado et al., 2000a).

**Recombination analysis.** Recombination analysis was performed with the SimPlot program (Ray, 1999). Briefly, this program calculates and plots the percentage identity of the query sequence to a panel of reference sequences in a sliding window, which is moved along the alignment in steps. A 140-nucleotide window and a step size of 20 nucleotides were used. This program identified informative sites and possible recombination breakpoints in the alignments. These breakpoints divided the alignment into segments, which were submitted to phylogenetic tree constructions as described above.

**Statistical analysis.** All of the statistical analyses were performed with Graph Pad PRISM version 2.01. The unpaired t-test was used to compare group means.

**Results**

**Genetic characterization of HIV-1 isolates**

Viruses obtained from a group of 46 asymptomatic patients followed for 1 to 4 years were analysed genetically. They were tested by two genetic screening methods, first by HMA (Fig. 1) and then confirmed by the RNase A mismatch cleavage method (RAMM) (data not shown) (López-Galíndez et al., 1991). A group of samples was selected for further analysis according to the quasispecies pattern from the above screening methods. Three criteria were used in the selection of virus samples: (i) samples with a unique discrete band in the HMA, mirroring a predominant virus population (viruses I01 and I16), (ii) samples displaying a smear of bands in the HMA, reflecting high variation and a very complex quasispecies (I09 and IH07),

---

**Fig. 1.** HMA analysis of the C2–V5 region of the env gene. A group of the first samples of the patients studied was amplified and subjected to heteroduplex analysis as indicated in Methods. M corresponds to φX174 HaeIII as a molecular mass marker. Numbers above each lane identify the sample analysed. Arrows point to samples that were selected for further analysis at the quasispecies level. Samples I01 and I16 displayed a homogeneous HMA pattern, as shown by an unique homoduplex band. Samples IH07 and I09 had a very heterogeneous HMA pattern, whereas samples H16 and IH03 showed patterns corresponding to more than one virus population.

**Fig. 2.** Histogram representation of the distribution of pairwise genetic distances and corresponding phylogenetic trees from individual quasispecies of patients displaying similar genetic distances. Yearly samples of quasispecies from patients with similar genetic distance values were analysed. All of the genetic distances between sequences from the quasispecies were compared pairwise and the results are shown as histograms. The same quasispecies sequence was then analysed by neighbour-joining as indicated in Methods and the resulting tree is shown below the corresponding histogram. Numbers on the branches of the trees correspond to bootstrap values of the branch (for 1000 replicas). Bar, 1% genetic distance. The 1993 sample from patient H16 and the 1996 sample from patient I16 displayed similar mean genetic distances (3.0 and 2.9%) (a). The same was true of the 1993 I01 sample and the 1995 IH07 sample (1.8 and 2.0%) (b).
Different patterns of HIV-1 evolution in vivo

Table 2. Genetic heterogeneity in the env region within each quasispecies analysed

Heterogeneity is the mean ± SEM nucleotide genetic distance as a percentage between all pairwise comparisons of sequences from within each time-point. The range of values is given in parentheses. $K_s$ and $K_w$ respectively represent the non-synonymous and synonymous genetic distances as percentages between all pairwise comparisons of sequences from within each time-point; values are means ± SEM.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sample</th>
<th>Heterogeneity</th>
<th>$K_s$</th>
<th>$K_w$</th>
<th>$K_s/K_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I01</td>
<td>1993</td>
<td>1.8 ± 0.20 (0.0–6.3)</td>
<td>1.6 ± 0.25</td>
<td>2.0 ± 0.53</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td>3.0 ± 0.42 (0.2–6.9)</td>
<td>3.2 ± 0.51</td>
<td>4.8 ± 0.91</td>
<td>0.66 (C: 0.32; D: 0.46)*</td>
</tr>
<tr>
<td>IH03</td>
<td>1994</td>
<td>4.0 ± 0.48 (0.0–6.9)</td>
<td>3.4 ± 0.57</td>
<td>5.3 ± 1.10</td>
<td>0.64 (A: 0.50; B: 0.74)*</td>
</tr>
<tr>
<td></td>
<td>1998</td>
<td>2.4 ± 0.34 (0.0–4.9)</td>
<td>1.8 ± 0.36</td>
<td>4.3 ± 1.02</td>
<td>0.42</td>
</tr>
<tr>
<td>H16</td>
<td>1993</td>
<td>3.0 ± 0.40 (0.0–6.6)</td>
<td>3.2 ± 0.54</td>
<td>2.0 ± 0.69</td>
<td>1.60 (A: 0.46; B: 1.27)*</td>
</tr>
<tr>
<td></td>
<td>1994</td>
<td>2.1 ± 0.29 (0.4–5.8)</td>
<td>2.0 ± 0.38</td>
<td>2.3 ± 0.56</td>
<td>0.90</td>
</tr>
<tr>
<td>I09</td>
<td>1993</td>
<td>4.9 ± 0.53 (0.2–8.0)</td>
<td>4.9 ± 0.67</td>
<td>4.3 ± 0.89</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td>6.0 ± 0.52 (1.1–10.4)</td>
<td>6.0 ± 0.78</td>
<td>3.9 ± 0.68</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td>4.8 ± 0.48 (0.5–10.9)</td>
<td>4.6 ± 0.67</td>
<td>4.4 ± 0.84</td>
<td>1.05</td>
</tr>
<tr>
<td>I16</td>
<td>1993</td>
<td>1.5 ± 0.20 (0.0–5.4)</td>
<td>1.4 ± 0.25</td>
<td>1.6 ± 0.36</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td>2.9 ± 0.41 (0.3–4.8)</td>
<td>2.4 ± 0.46</td>
<td>4.1 ± 0.90</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td>3.3 ± 0.40 (0.6–6.0)</td>
<td>3.0 ± 0.43</td>
<td>3.6 ± 0.74</td>
<td>0.83</td>
</tr>
<tr>
<td>IH07</td>
<td>1995</td>
<td>2.0 ± 0.30 (0.4–4.1)</td>
<td>2.0 ± 0.40</td>
<td>1.9 ± 0.50</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td>4.3 ± 0.49 (0.3–7.0)</td>
<td>4.0 ± 0.62</td>
<td>4.7 ± 0.93</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td>4.1 ± 0.48 (0.8–6.8)</td>
<td>3.8 ± 0.58</td>
<td>4.3 ± 0.82</td>
<td>0.88</td>
</tr>
</tbody>
</table>

$K_s/K_w$ ratios were also calculated from two individual populations.

and (iii) viruses with two or more major separated homoduplex bands in the HMA (H16 and IH03) (see Fig. 1). Of the group of patients analysed, 22% of the patients displayed the first HMA pattern, 37% the complex quasispecies pattern and 41% showed two or more variants in the HMA. Two patients of each group were selected and subjected to genetic analysis by nucleotide sequencing of proviral DNA of 20 clones in the C2-fusion region of the env gene.

Genetic analysis of individual samples from selected patients

In order to analyse the genetic variability of the different samples, we first studied the distribution of genetic heterogeneity within quasispecies from each year. When comparing samples with very similar mean heterogeneity values, the distribution of genetic distances was very different. In isolates H16 (1993 sample) and I16 (1996 sample), with respective mean heterogeneities of 3.0 and 2.9%, two different patterns were observed in the histograms. Patient H16 displayed two peaks in the distribution, whereas a normal distribution was observed in patient I16 (Fig. 2a). The same two patterns were obtained from patients I01 (1993 sample) and IH07 (1995 sample), with heterogeneity values of 1.8 and 2.0% (Fig. 2b). This study was performed by a pairwise analysis of the mean genetic heterogeneity distances displayed in Table 2 and as explained in Methods. Patient IH03 displayed the separate population profile and patient I09 the normal distribution (data not shown). In order to confirm the existence of different patterns in the distribution of genetic distances, a phylogenetic analysis performed by neighbour-joining was carried out with these sequences, as indicated in Methods. The trees obtained are shown in Fig. 2 (a, b) below the corresponding histograms. As can be seen, the trees displayed two very different topologies. In the trees of viruses I01 and H16, two independent groups of sequences were present and the clades were separated in 100% of the bootstrap replicas performed. The two clades defined by the phylogenetic analysis were separated by a genetic distance of more than 3–4% between the nodes. In contrast, in the trees of patients I16 and IH07, the bootstrap values of the internal branches were lower and no independent groups could be defined.

HIV-1 sequence evolution determined by phylogenetic analysis of the env gene

In order to study the evolution of virus sequences within patients, a global phylogenetic analysis was carried out, analysing all the samples of each patient together. The study was performed by the neighbour-joining method as indicated in Methods and the resulting trees are displayed in Fig. 3. Samples from patients I01, IH03 and H16 displayed the same tree topology, characterized by the presence of independent clades of sequences with highly significant bootstrap values (> 94%). The other topology, shown by isolates I16, IH07 and I09, was defined by lower statistical bootstrap support of internal tree branches and by the presence of long external
Fig. 3. Phyllogenetic trees derived by the neighbour-joining method corresponding to all quasispecies analysed in each patient. All trees were constructed as indicated in Methods and were drawn at the same scale, allowing direct comparison of genetic distances. The different symbols identify the sequences from each year. Symbols with numbers identify sequences cited in the text. Underlined sequences indicate hypermutated sequences. Numbers on the trees represent bootstrap values (for 1000 replicas). Circles around a group of sequences distinguish clades that were analysed independently. NSI indicates a group of sequences that showed a non-syncytium-inducing V3 genotype and SI represents a syncytium-inducing genotype. Sequences in boxes correspond to recombinant genomes (see Fig. 4). Bars, 1% genetic variation.
obtained from virus quasispecies calculations. mutated sequences were excluded from further studies and (1995 sample). These virus sequences corresponded to hyper-long branches, like sequences 09, 10 and 17 of patient IH03 ancestor (Fig. 3). Some of the trees also displayed a few very branches. Moreover, in the quasispecies of viruses from patients I01, IH03 and H16, sequences found at one time-point were never mixed with sequences of quasispecies from the subsequent years and it was impossible to establish relationships between sequences from the different years. In contrast, in viruses from patients I16, IH07 and I09, sequences obtained at one time-point were mixed in the same branches of the trees with sequences of the later samples (Fig. 3). Also, some clones of patients I16, IH07 and I09 could be considered as the origin of sequences obtained in later years, because of their positions in the tree. For example, in patient I09, 1996 sequences 12, 13, 15, 26 and 28 and 1997 sequences 05, 13 and 20 were phylogenetically related and shared the same common ancestor (Fig. 3). Some of the trees also displayed a few very long branches, like sequences 09, 10 and 17 of patient IH03 (1994 sample) and sequences 44, 18 and 24 in patient IH07 (1995 sample). These virus sequences corresponded to hypermutated sequences (Varatian et al., 1991). All the hypermutated sequences were excluded from further studies and calculations.

Quantification and analysis of the genetic variation obtained from virus quasispecies

The genetic heterogeneity of virus quasispecies was calculated, and the mean values ranged from homogeneous samples in patient I16, with 1.5% (1993 sample) (Myers et al., 1995), to very heterogeneous viruses, with a value of 60%, in patient I09 (1996 sample), as shown in Table 2. Patients IH03 and H16 showed a significant decrease in heterogeneity in the last sample. This pattern of genetic variation was different from that displayed by all the other patients analysed. Mean virus divergence increased in every patient with time (Table 3), but the rate was very heterogeneous, varying from 0.32% per year in patient I09 to 2.50% per year in patient IH07 in the period 1995–1996; the most frequent value was between 0.64 and 0.75% in different samples of patients I09, I16, I01, IH03 and H16.

In order to analyse the existence of different evolutionary forces operating within and between patients, the contributions of synonymous and non-synonymous mutations to virus divergence and heterogeneity were calculated. This was carried out in each quasispecies according to the method of Nei & Gojobori (1986) and the results are displayed in Tables 2 and 3. In the first three patients, the \( K_s/K_a \) ratio was always below 1, except in patient H16 (1993 sample). Virus subpopulations were separated in statistically independent clades in these viruses (see Fig. 3 and previous paragraph). \( K_a/K_s \) ratio calculations were carried out for some of the individual clades. The values obtained were lower than the global values and lower than 1. However, in the case of patient H16 (1993 sample), we obtained an unexpected result. Population A (see Fig. 3) had a \( K_a/K_s \) value of 0.46, whereas population B had a value of 1.27 (Table 2). This result indicated that the intensity of selection was greater in one of the HIV-1 subpopulations than in the other within the same patient over the same period of time. \( K_s/K_a \) ratios in patients I09, I16 and IH07 were higher than those in the first three subjects (Table 2), indicating the presence of higher positive selective forces in this group of patients. This trend was more evident when the \( K_a/K_s \) ratio was calculated on the divergence values (see Table 3).

### Table 3. Mean divergence in virus quasispecies in relation to the initial population analysed

Divergence is the mean±SEM nucleotide genetic distance as a percentage between all pairwise comparisons of sequences between different time-points. The ranges of values are given in parentheses. \( K_s \) and \( K_a \) represent the same as in Table 2, but were calculated between different time-points. Values are means±SEM.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sample</th>
<th>Divergence</th>
<th>( K_s )</th>
<th>( K_a )</th>
<th>( K_s/K_a )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I01</td>
<td>1993</td>
<td>1.8±0.20 (0.0–6.3)</td>
<td>1.6±0.25</td>
<td>2.0±0.53</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td>4.4±0.50 (3.2–7.1)</td>
<td>3.1±0.50</td>
<td>6.4±1.50</td>
<td>0.48</td>
</tr>
<tr>
<td>IH03</td>
<td>1994</td>
<td>4.0±0.48 (0.0–6.9)</td>
<td>3.4±0.57</td>
<td>5.3±1.10</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>1998</td>
<td>7.0±0.78 (5.6–8.3)</td>
<td>5.9±0.93</td>
<td>9.4±1.72</td>
<td>0.62</td>
</tr>
<tr>
<td>H16</td>
<td>1993</td>
<td>3.0±0.40 (0.0–6.6)</td>
<td>3.2±0.54</td>
<td>2.0±0.69</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>1994</td>
<td>4.0±0.49 (2.0–5.5)</td>
<td>4.2±0.71</td>
<td>3.1±0.77</td>
<td>1.35</td>
</tr>
<tr>
<td>I09</td>
<td>1993</td>
<td>4.9±0.53 (0.2–8.0)</td>
<td>4.9±0.67</td>
<td>4.3±0.89</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td>5.9±0.59 (2.9–10.3)</td>
<td>5.9±0.73</td>
<td>4.1±0.73</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td>5.7±0.52 (3.5–10.2)</td>
<td>5.6±0.72</td>
<td>4.1±0.78</td>
<td>1.37</td>
</tr>
<tr>
<td>I16</td>
<td>1993</td>
<td>1.5±0.20 (0.0–5.4)</td>
<td>1.4±0.25</td>
<td>1.6±0.36</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td>2.7±0.40 (1.1–4.8)</td>
<td>2.3±0.40</td>
<td>3.4±0.80</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td>3.6±0.50 (0.8–5.8)</td>
<td>3.5±0.53</td>
<td>3.2±0.79</td>
<td>1.09</td>
</tr>
<tr>
<td>IH07</td>
<td>1995</td>
<td>2.0±0.30 (0.4–4.1)</td>
<td>2.0±0.40</td>
<td>1.9±0.50</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td>4.5±0.56 (2.9–6.3)</td>
<td>4.7±0.69</td>
<td>3.4±0.56</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>1997</td>
<td>4.8±0.60 (2.9–6.8)</td>
<td>4.8±0.69</td>
<td>4.2±0.88</td>
<td>1.14</td>
</tr>
</tbody>
</table>
Recombination analysis

Taking advantage of the existence of different clades with high bootstrap values in the quasispecies of patients I01, IH03 and H16 (see Fig. 3), we analysed the contribution of genetic recombination to the evolution of virus quasispecies. The study was performed with the help of the SimPlot program, as indicated in Methods. For this study, we selected separated sequences from patients IH01, IH03 and H16, which are circled in Fig. 3. For example, in isolate H16, the two populations found in the 1993 sample were named A and B, and those obtained in 1994 were designated C and D. We carried out SimPlot analysis of the sequences of groups C and D against sequences from groups A and B. One outgroup sequence from another isolate was used in the analysis and the results obtained with clade D are shown in Fig. 4. The sequence of group D showed more similarity to sequence A at the 5' end, whereas it was closer to sequence B in the middle part of the sequence. Finally, at the 3' end, all clades were homologous (Fig. 4a). Fig. 4(b) shows a bootscan analysis, allowing the identification of a recombination break-point at position 250 in relation to the 5' end of the fragment. Sequences were separated according to these fragments and were analysed by neighbour-joining (Fig. 4 c, d). The sequence of set D clustered with sequence A in 89% of the trees in the 5' region and clustered together with sequence B in the 3' region in 91% of the trees. The 210 nucleotides at the 3' end of the 660 nucleotide fragment were very similar between the A and B groups of sequences. The genetic distance between population D and the two parental populations, A and B, was lower than the distance between populations A and B. Similar analysis with sequences from clade C did not reveal any recombination event in those virus subpopulations (data not shown). Similarly, recombination events were detected in patients IH03 and I01. However, in these patients, only individual recombinant sequences were detected (boxed in Fig. 3), like sequence 16 from 1994 in patient IH03, which was a recombinant between sequences A and B with a breaking point at position 483 (data not shown). Sequence 15 from 1993 in patient I01 was a recombinant between population A and B and the breaking point is at position 180 (data not shown). In patients I09, IH07 and I16, we could not perform the recombination analysis as it
Fig. 5. RNA and DNA quasispecies corresponding to different samples of patients I09 and H16. Sequences obtained from serum samples taken on the same date as the proviral DNA sequences included in Fig. 3 were derived as explained in Methods. The same symbols used in Fig. 3 have been used for proviral DNAs. Trees were derived by the neighbour-joining method. The circled sequences indicate groups of sequences that have been analysed in recombination analysis and the boxes indicate recombinant sequences. Bars, 1% genetic distance.
was impossible to identify statistically independent groups of sequences within the virus quasispecies.

**Comparison of sequences of virus RNA (vRNA) from sera and provirus DNA**

In order to analyse further the evolution of the virus quasispecies within patients, vRNA was studied at different time-points for two of the infected individuals, H16 and I09, as explained in Methods (Fig. 5). In patient I09, proviral sequences from 1993 clones 17 and 28 were intermingled among numerous 1993 vRNA sequences, indicating active replication of this provirus population. These 1993 vRNA sequences resulted in provirus forms found in the 1990 samples. vRNA population C of the 1996 sample was very interesting, because it confirmed that replicating populations were phylogenetically close to the future provirus DNA, such as sequences 13, 20 and 05 from 1997. The same could be applied to vRNA sequences in branch B, which was rooted between sequences from 1993, 1996 and 1997.

In sample H16, which displayed two independent virus populations in the 1993 proviral DNA, two vRNA populations were detected in 1993, designated E and F. Neither of these populations was coincident with the 1993 provirus populations. Population F was very close to the 1994 proviral DNA population C, and it is possible to consider this vRNA group as the origin of the main subpopulation of DNA and RNA sequences found in the 1994 sample. Recombination analysis similar to that shown in Fig. 4 and mentioned in the previous paragraph showed that the 1993 vRNA population E was a recombinant sequence between proviral DNA A sequences and vRNA population F. Sequence 01 from 1993 vRNA (boxed) was a recombinant and was present as a unique sequence. By its location in the tree and by the similar breaking point, this sequence could be considered as the origin of the proviral recombinant clade D of 1994. The two DNA clades A and B of the 1993 sample and the recombinant population D of the 1994 sample were not detected as replicating vRNA. In summary, we have observed that vRNA populations at one time-point are the origins of proviral DNA sequences detected in later samples (Simmonds et al., 1991) and, as the RNA input used for the amplification was between 1360 and 4080 copies of vRNA per ml, these results show that only a minority of the proviral DNAs are replicating.

**Discussion**

In this ex vivo study, we have analysed HIV-1 evolution in six patients selected by their patterns of virus genetic variability from a group of 46 patients with typical clinical progression. The analysis has been carried out by nucleotide sequencing of proviral DNA from circulating PBMCs and serum vRNA obtained over a 1 to 4 year period. Although HIV-1 recovered from PBMCs corresponds to only a minor proportion (< 2%) of the total virus replication in primary and secondary lymph organs, it could be considered a random sample of the global virus population within patients. Over this period of time, it can be estimated than HIV-1 has undergone between around 300 and 1200 replication cycles (Coffin, 1995). Phylogenetic analysis of nucleotide sequences of samples from different years revealed the existence of different topologies in the trees obtained by the neighbour-joining method. One topology was characterized by the presence of independent virus clades in the quasispecies supported by highly significant bootstrap values (see patients I01, IH03 and H16 in Fig. 3). In this group of isolates, virus sequences obtained in one year were not mixed within the virus quasispecies of subsequent years and it was not possible to derive relationships between sequences of different years or to establish a temporal evolution of the sequences. The existence of clustering in virus populations with different sampling times has been found previously in slow-progressing children from paediatric cases (Ganeshan et al., 1997).

In contrast, the other tree topology could be characterized by the presence of only one clade of virus sequences. Sequences were not separated by years and, moreover, a temporal relationship could be established between sequences from one year and sequences from subsequent years (see patients I09, IH07 and I16 in Fig. 3). In these patients, a few clades showed high bootstrap values. However, these clades correspond to non-synctial (NSI) and synctial (SI) virus populations like the ones identified in patients IH07 and I09 (see later and Fig. 3).

The differences in the evolutionary characteristics found within the group of patients analysed could be attributed to a sampling problem, because two samples were analysed in the first patients and three in the last ones (see Methods and Table 1). However, the individual analysis of each of the 15 different quasispecies studied (as shown in Fig. 2) revealed the same tree topology in the sample corresponding to a single year as in the tree constructed with all the samples of the patient, with the exception of the 1993 sample of patient I16 and the 1997 sample of patient I09.

The divergence and heterogeneity values obtained from patients I09, I16 and IH07 could be associated with the consistent pattern described by Shankarappa et al. (1999), particularly in patients IH07 and I09, where we also detected the existence of the NSI to SI switch that precedes a decrease in virus heterogeneity. The pattern observed in patients I01, IH03 and H16 was different from this model. The detection of this different model could be related to the criteria used in our study for the selection of the patients based on the pattern of genetic variability of the virus.

Several points deserve attention regarding the recombination analysis carried out in patients H16, IH03 and I01. One relates to the evolutionary potential of this mechanism. In patient H16, the appearance of recombinant variants has resulted in a reduction of the heterogeneity of the quasispecies and, from a phylogenetic point of view, this recombinant
variant was filling the evolutionary space between the two initial viruses (Lukashov & Goudsmit, 1997). In this patient, recombination has produced a new variant that displays an evolutionary jump larger than evolution mediated by point mutation (Chao, 1994; Domingo & Holland, 1994; Lai, 1992). The contribution of the recombinant forms to the overall evolution was limited, either because they do not replicate (for example, proviral population D in H16) or because they represent minor variants within the RNA or DNA quasispecies (see patients I01 and IH03). Only RNA subpopulation F of the 1993 sample in H16 did replicate. These results could reflect the lower fitness of recombinant forms in relation to their parental strains.

We have detected episodes of hypermutation in the virus quasispecies, as described previously (Meyerhans et al., 1994; Vartanian et al., 1994). These sequences are the result of sporadic events, but these hypermutated genomes could also contribute to the generation of new variants, such as genomes 10 and 17 of the 1994 sample in patient IH03, which could be the origin of sequence 09, possibly through genetic recombination (Fig. 3). In this case, recombination could result in the rescue of deleterious mutations present in one of the two virion genomes by incorporating information from the intact template, resulting in genetic diversification but also acting as a repair mechanism (Jetzt et al., 2000).

In order to explore the existence of clinical implications of the HIV-1 evolution observed in the patients, different clinical parameters were evaluated: the numbers of CD4+ and CD8+ cells as well as the CD4+/CD8+ ratio, virus load, antiretroviral treatment and clinical status. The results are shown in Table 1; there were no differences between patients in the numbers of CD4+ and CD8+ cells or the ratio. Neither were there clear differences in pathogenesis or treatment between the patients displaying different evolutionary patterns. Of the first three patients, one was progressing to disease and was put on antiviral therapy (patient I01) and another was evolving to clinical manifestation (patient H16). Of the last three patients, two patients were progressing to AIDS (I09 and IH07) and one of them undertook therapy. In these two patients, two populations were detected in the intermediate samples that displayed sequences associated with different phenotypes (NSI and SI). In the last samples of the two patients, the SI sequences were predominant, confirming the phenotypic switch (see Fig. 3). The two NSI and SI clades were separated by high bootstrap values in the phylogenetic trees.

A positive correlation was found in two other parameters, however. One was the time from HIV-1 seroconversion and the other the virus RNA load. In patients I01, IH03 and H16, the time between infection and sampling was longer than in patients I16, IH07 and I09. In general, the time from infection was not very precise because of the lack of seroconversion dates for some patients. For the probable infection date, we have adopted the median between the beginning of risk practices and the first serological confirmation of infection. For example, patient H16 began to have frequent unprotected sex in 1982, and the first positive serology was obtained in 1991, so we took 1986 as the probable infection date. With this criterion, the mean time since infection was 11.4 years for the first three patients and 5.2 for the last three. Virus load was also different between the groups. For the first three patients shown in Table 1, a mean of 12933 viral RNA copies per ml was calculated, in contrast with the value of 136422 copies per ml for the second group. The mean for the first three patients is more than 10-fold lower than that for the last three individuals, although this difference is not statistically meaningful because of the wide dispersion of the data. The differences found in the pattern of virus evolution could be attributed to the size of the virus populations existing in the patients, because small virus populations tend to promote variation, as shown in vitro by Sánchez-Palomino et al. (1993). However, samples from patients with the same virus load, like sample 1993 of patient H16 and sample 1996 of patient I16 (see Table 1), exhibited distinct phylogenetic patterns of evolution (see Fig. 3).

It is not clear whether the two tree topologies and other phylogenetic characteristics were permanent over longer periods of virus evolution, or whether they reflect a punctuated virus evolution event. It is not possible to discriminate between different hypotheses regarding the evolution of HIV-1 observed in vivo because of the limited sampling and the short time of this study. It is feasible that all of the different HIV-1 evolutionary events detected in the study occur in every infected patient but only the existence of distinct and separated clades allowed its observation, like the discovery of recombinant forms or the presence of different $K_e/K_r$ ratios within the same patient. On the other hand, it is conceivable that the virus evolution observed in patients I01, H16 and IH03 occurs only in a subgroup of patients. It is interesting to note that, in two studies, one with children and another with adults, different phylogenetic patterns of virus evolution were found and, in both reports, a longer clinical progression time was also found among viruses displaying a clustering with time (Ganeshan et al., 1997; Wolinsky et al., 1996).

The mode of virus evolution observed in the first three patients studied, characterized by statistically distinct virus subpopulations in phylogenetic trees, diverse peaks in the distribution of genetic distances, different $K_e/K_r$ ratios in the same patient and lower virus loads, is compatible with contained and compartmentalized virus replication. The progressive replication of HIV-1 in germinal centres of lymph nodes is found in asymptomatic patients and it is followed by the destruction of the architecture of the lymph nodes in late stages of the disease (Pantaleo et al., 1993, 1994). There have been descriptions of differences in virus subpopulations in different germinal centres of the same organ as a consequence of founder effects (Cheynier et al., 1994). These compartmentalized virus populations could be under different immunological pressures, as we detected for patient H16 (see Table 2). The presence of different variants within patients could be the...
consequence of the generation of different variants during virus replication or of infection by more than one variant, either at the same time or by reinfection (Sabino et al., 1994; Wang et al., 2000; Zhu et al., 1995). In contrast, in the mode of virus evolution displayed by patients I09, I16 and IH07, it was impossible to define individual populations, which probably indicates that virus replication was not contained in different lymph organs, perhaps as a consequence of the progressive loss of architectural integrity associated with the progression to AIDS (Pantaleo et al., 1993, 1995). Although low virus loads have been associated with defects in virus genes in some patients (Kestler et al., 1991; Kirchhoff et al., 1995; Mariani et al., 1996), the different patterns of virus evolution observed in the patients analysed in this study seem to be a consequence of host-related factors, as shown in humans (Wolinsky et al., 1996) and macaques (Pelletier et al., 1995).

In summary, analysis of HIV-1 quasispecies from six patients with normal clinical progression patterns, selected for their different genetic variability profiles, has allowed the detection of different patterns of HIV-1 evolution in vivo. The different patterns were characterized by different virus phylogenetic trees and by the modes of genetic distance distribution. The pattern displaying independent clades allowed the detection of recombination episodes, or the recognition of different $K_e/K_s$ ratios between distinct virus subpopulations within patients. The results of this study show that HIV-1 evolution in vivo is not homogeneous and different evolutionary patterns can be found in infected patients.

E. Domingo and A. Telenti and two anonymous reviewers are thanked for the critical reading of the manuscript and for helpful suggestions and A. Meyerhans, N. Villanueva and I. Olivares are thanked for discussions. Stuart Ray is thanked for providing the SimPlot program and for his help with early versions. This work was supported by FIS grants numbers 97/0119 and 00/0268.

References


Different patterns of HIV-1 evolution in vivo


Lukashov, V. V. & Goudsmit, J. (1997). Evolution of the human immunodeficiency virus type 1 subtype-specific V3 domain is confined to a sequence space with a fixed distance to the subtype consensus. Journal of Virology 71, 6332–6338.


Received 25 April 2001; Accepted 4 July 2001