Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

doi:10.1016/S0022-2836(03)00661-2 J. Mol. Biol. (2003) 331, 213–229

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Memory in Retroviral Quasispecies: Experimental Evidence and Theoretical Model for Human Immunodeficiency Virus

Carlos Briones[1], Esteban Domingo[1,2] and Carmen Molina-París[3,4]*

[1]*Centro de Astrobiología (CSIC-INTA), Carretera de Ajalvir, Km. 4, Torrejón de Ardoz, 28850 Madrid, Spain*

[2]*Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM) Cantoblanco, 28049 Madrid Spain*

[3]*Mathematics Institute University of Warwick Coventry CV4 7AL, UK*

[4]*Applied Mathematics Department, University of Leeds, Leeds LS2 9JT, UK*

**Corresponding author*

Viral quasispecies may possess a molecular memory of their past evolutionary history, imprinted on minority components of the mutant spectrum. Here we report experimental evidence and a theoretical model for memory in retroviral quasispecies *in vivo*. Apart from replicative memory associated with quasispecies dynamics, retroviruses may harbour a "cellular" or "anatomical" memory derived from their integrative cycle and the presence of viral reservoirs in body compartments. Three independent sets of data exemplify the two kinds of memory in human immunodeficiency virus type 1 (HIV-1). The data provide evidence of re-emergence of sequences that were hidden in cellular or anatomical compartments for extended periods of infection, and recovery of a quasispecies from pre-existing genomes. We develop a three-component model that incorporates the essential features of the quasispecies dynamics of retroviruses exposed to selective pressures. Significantly, a numerical study based on this model is in agreement with the experimental data, further supporting the existence of both replicative and reservoir memory in retroviral quasispecies.

Keywords: quasispecies; memory; viral reservoirs; retroviruses; human immunodeficiency virus

Introduction

RNA viruses replicate as complex distributions of mutant genomes termed viral quasispecies.[1–3] Quasispecies was developed as a general theory to understand the dynamics of the first replicative molecules in the context of the origin of information and early evolution of life.[4] Quasispecies dynamics, characterized by a continuous process of mutant generation, competition and selection, has also provided an interpretation of the great adaptive potential of pathogenic RNA viruses, and has evidenced the need to use combination antiviral therapies and to design polyvalent antiviral vaccines for the control of viral disease (review by Domingo *et al*.[3]). An understanding of quasispecies dynamics has opened new antiviral strategies through virus entry into error catastrophe, also termed lethal mutagenesis (review by Eigen[5]).

Experiments with the animal pathogen foot-and-mouth disease virus (FMDV) have documented that RNA virus quasispecies may possess a genetic memory of those genomes that were dominant at an earlier phase of their evolution.[6–8] Memory was detected as minority genomes that replicated in the mutant spectrum of two independent evolutionary lineages of FMDV in cell culture.[6,8] The presence of these replicative memory genomes in viral quasispecies may represent a selective advantage for the virus whenever fluctuating selective pressures occur, for example during prolonged chronic infections.[9] However, no direct evidence of quasispecies memory *in vivo* has been presented so far.

HIV-1 is a human retroviral pathogen whose populations participate of quasispecies dynamics[10] (review by Crandall[11]) and that includes two disparate phases in its replication cycle.[12] As

Present address: C. Molina-París, Applied Mathematics Department, University of Leeds, Leeds LS2 9JT, UK.

Abbreviations used: HIV-1, human immunodeficiency virus type 1; FMDV, foot-and-mouth disease virus; RTI, reverse transcriptase inhibitor; PRI, protease inhibitor; HAART, highly active antiretroviral therapy; AZT, zidovudine; ddI, didanosine; ddC, zalcitabine; d4T, stavudine; 3TC, lamivudine; RTV, ritonavir; SQV, saquinavir; NFV, nelfinavir; IDV, indinavir; NVP, nevirapine; HU, hydroxyurea.

E-mail address of the corresponding author: carmen@maths.leeds.ac.uk

integrated proviruses, cDNA copies of HIV-1 RNA replicate as cellular genes, with the high fidelity copying inherent to cellular DNA polymerases endowed with proofreading-repair activities and subjected also to post-replicative repair pathways.[13] Upon expression from integrated sites, viral RNAs are processed to direct the synthesis of multiple proteins to culminate in the formation of retroviral particles whose first replicative step upon entering a new host cell is the error-prone copying of genomic RNA into DNA by the retroviral reverse transcriptase (RT).[12] This replication cycle favours the existence of a second type of genetic memory in HIV-1 which differs from the replicative memory described for FMDV, a picornavirus whose RNA replication cycle does not include a DNA step. Integrated DNA copies of the HIV-1 genome constitute a reservoir of viral sequences that upon cellular activation can contribute new mutant distributions to the replicating pool of the virus.[11,12] Several anatomical compartments (including lymph nodes) and cell types (resting CD4+ lymphocytes, macrophages, and others) are potential reservoirs for HIV-1.[14–16] Viruses in reservoirs can evolve more slowly than those circulating in the plasma of the same patient, and they can maintain for months or years the genotype exhibited when they entered the reservoir. This genotype can be either wild-type or drug-resistant, independent of the genotype currently dominant in the plasma.[17,18] Integrated proviruses are not affected by the selective pressure of reverse transcriptase inhibitors (RTI), whilst their expression, following activation of the host lymphocyte, may be sensitive to protease inhibitors (PRI).[11,12] Also, some anatomical sites may provide a barrier to the antiretroviral pressure, due to limited access of some drugs (i.e. PRI).[17,18] Viruses from such reservoirs could re-emerge and occupy the circulating quasispecies as documented by phylogenetic analyses of consensus sequences.[19] Re-emergence could be a consequence of structured treatment interruptions.[20–22] Figure 1 shows a flux diagram of HIV-1 infection that takes into account the different cellular types that contribute to (and can represent sources of) memory in the circulating HIV-1 quasispecies (quantitative data[23–28]). Here we provide direct experimental evidence of the occurrence of memory in HIV-1 quasispecies, and we present a mathematical model to distinguish replicative memory from transiently non-replicative (or reservoir) memory operative in HIV-1. We discuss the implications of the two types of memory for retrovirus adaptation.

Results

Re-emergence of HIV-1 sequences after prolonged times of infection: non-replicative (or reservoir) memory in HIV-1

To investigate the existence and putative sources of memory in HIV-1 quasispecies, three patients (referred to as case 1, case 2 and case 3, detailed in Materials and Methods) were selected due to their poor response to the therapy and/or the fluctuations observed in their viral load and CD4 + lymphocyte count during their complex treatment histories (Table 1). The possible contribution of non-replicative (reservoir) memory was studied by means of sequence analysis of retrospective viral samples from two patients (cases 1 and 2). Consensus nucleotide sequences were used to deduce the amino acid substitutions present at each codon of the protease (PR) and the first 220 codons of the RT (Table 2). Sequential HIV-1 sequences from case 1 showed a progressive accumulation of resistance mutations after a five year exposure to AZT (in monotherapy from April 1992 to December 1995, and in combination with ddI until May 1997). This accumulation of nucleoside-resistance mutations (Table 2) led to selection of multinucleoside-resistant genomes encoding the insertion T69SSS in sample 5.97. The shift to a HAART combination (d4T + 3TC + RTV) resulted in the absence of this insert-containing genotype in a sample collected only two months later. Interestingly, this sample (7.97) and the following one (3.98) exhibited a mutation pattern in the RT very close to that of the first available sample of this patient (11.94). HIV-1 proviral DNA collected in August 1998 conserved some of the mutations found in the RT-coding region prior to May 1997. The re-emergence of the initial genotype is not evident for the consensus sequence of the PR-coding region, although a shift to a consensus sequence lacking the key substitutions N37S and R41K was observed in sample 7.97 (Table 2). The different behaviour of the RT and PR genes could be due to the absence of a selective pressure acting on the PR, resulting in their independent evolution, or to intra-patient recombination events. Different phylogenetic approaches were used to analyse consecutive consensus sequences in order to evaluate the possibility of a re-emergence of viruses that could have been hidden in cellular or anatomical compartments from November 1994 until July 1997. NJ, UPGMA, MP and ML methods produced the same overall tree topology for consensus sequences of *pol* and *env* regions (Figure 2(a) and (b)). The phylogenetic relationships are highly consistent and show a clear separation between two groups of isolates: 7.95/12.95/9.96/5.97/8.98-DNA (genomes with a progressive accumulation of nucleoside analog-resistance mutations, and the proviral sequence), and 11.94/7.97/3.98 (the ancestral genome and those present after its re-emergence).

Case 2 was a patient with a poor adherence to treatment, who showed large fluctuations in viral load as a response to different HAART regimes during the period of study (Table 1). After the interruption of a regime composed of ddI + d4T + RTV + SQV in September 1997, large fluctuations both in viral load and mutational pattern were recorded in consensus sequences from October 1997 to September 1998 (Tables 1 and 2). The

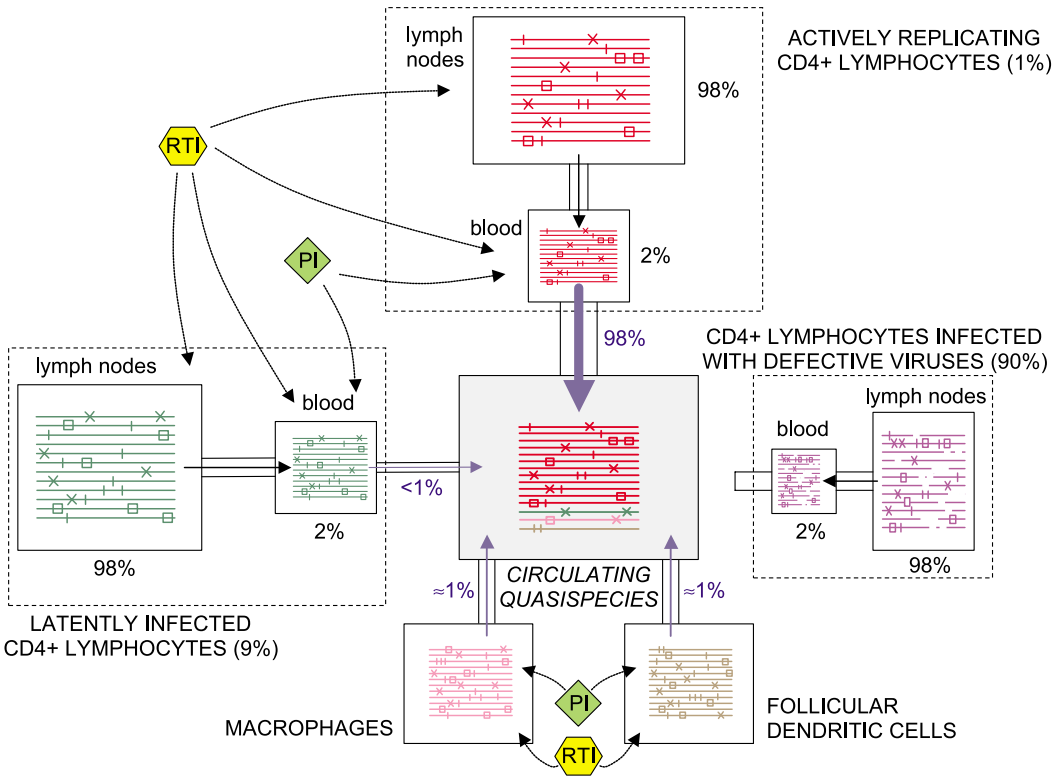

Figure 1. Flux diagram of quasispecies components in a typical HIV-1 infection and the possible sources of genetic memory. The contribution of different infected cells to the circulating HIV-1 quasispecies is as follows: 98% of the circulating virions are produced by actively replicating CD4+ lymphocytes (which constitute 1% of the total number of CD4+ lymphocytes), about 1% of the virions come from macrophages, another 1% from follicular dendritic cells and, finally, less than 1% are produced upon activation of latently infected CD4+ lymphocytes (which represent 9% of the total number of CD4+ lymphocytes). Most (about 90%) of the CD4+ lymphocytes are infected with defective viruses, and they do not contribute to the circulating quasispecies. These three types of CD4+ lymphocytes are similarly distributed between lymph nodes (98%) and the bloodstream (2%) (based on data from Refs. 23–28, and references therein).

amino acid substitutions in PR and RT suggested the occurrence of two progressive re-emergence events. Sample 8.98 would be the result of the re-emergence of viruses related to those in sample 10.97 (with some definitory amino acid replacements, such as V32I and V82A in the PR, and K70R and K122E in the RT), and sample 9.98 would result from the re-emergence of sequences from sample 1.98 (with signature substitutions L10I, I62V, G73S, I84V and L90M at the PR, and M41L and L210W at the RT). Most of these amino acid replacements are related to drug resistance to PRI or RTI.[29,30] A phylogenetic analysis of the sequential consensus sequences using MP, NJ, UPGMA and ML indicated isomorphic topologies for *pol* and *env* in all cases (Figure 2(c) and (d)). The phylogenies show a clear clustering of samples 10.97/8.98 distant from the cluster formed by samples 1.98/9.98/1.99, consistent with a double process of re-emergence of genomes hidden in reservoirs.

Replicative memory in HIV-1

Apart from non-replicative memory due to re-emergence from cellular or anatomic compartments, a replicative memory could also exist in retroviruses, analogous to that previously documented in FMDV.[6–9] To evaluate the possible presence of replicative memory in HIV-1, a detailed phylogenetic analysis was performed on sequential clonal sequences from case 3. In this patient, key mutations in HIV-1 related to decreased sensitivity to inhibitors were maintained at all time points, yet dominance of genomes encoding the RT insertion T69SSS was lost within a six month interval (from January to June 1995), probably as a result of the switch from AZT + ddC treatment to ddI monotherapy. The re-introduction of AZT (in combination with 3TC) was associated with selection (from June to August 1995) of a virus whose consensus sequence contained the RT insert (Table 2). MP analysis of 693 bp long sequences of 120 molecular clones spanning the PR-coding region and the first 130 codons of RT for six sequential samples (Table 1) using HIV-1 CAM-1 as the reference sequence showed a separation (statistically supported by bootstrap re-sampling) between sequences harbouring the dipeptide insertion (either Ser-Ser or Ser-Gly) and those without any

Table 1. CD4+ count, viral load and treatment history of the patients included in this study

Case	Sampling date[b]	Sample identification[c]	CD4+ count[a] No.	(%)	Viral load[a] (RNA copies/ml)	Antiretroviral therapy introduced[a]
1	3.4.92		338	26	–	AZT
	15.11.94	11.94	297	31	<500	AZT
	27.7.95	7.95	299	27	6240	AZT
	19.12.95	12.95	208	26	8920	AZT + ddI
	5.9.96	9.96	390	30	<500	AZT + ddI
	13.5.97	5.97	297	27	1610	d4T + 3TC + RTV
	21.7.97	7.97	245	26	5830	d4T + 3TC + SQV
	7.10.97		360	29	<500	d4T + 3TC + SQV
	3.3.98	3.98	416	32	650	d4T + 3TC + SQV
	21.7.98		337	29	<50	d4T + 3TC + SQV + NFV
	11.8.98	8.98-DNA	–	–	–	d4T + 3TC + SQV + NFV
	15.12.98		377	30	<50	d4T + 3TC + IDV
	11.1.99		–	–	–	d4T + NVP + NFV
	12.3.99		494	38	<50	d4T + NVP + NFV
2	12.3.92		–	–	–	AZT
	23.3.95		240	11	–	AZT + ddI
	21.9.95	9.95	286	11	–	AZT + ddI
	9.1.96		250	10	–	AZT + ddI
	3.5.96		176	8	–	AZT + ddC
	21.5.96		–	–	–	AZT + 3TC
	26.6.96		–	–	1,300,000	AZT + d4T + 3TC
	6.8.96	8.96	189	9	120,060	AZT + d4T + 3TC
	8.1.97		168	8	–	d4T + 3TC + IDV
	24.4.97		261	9	156,150	d4T + 3TC + IDV
	19.8.97		281	10	26,520	d4T + 3TC + IDV
	3.9.97		–	–	–	ddI + d4T + RTV + SQV (Stops all treatment: 26.9-17.10)
	17.10.97	10.97	280	10	3730	ddI + d4T + RTV + SQV
	5.1.98	1.98	280	10	35,910	ddI + d4T + RTV + SQV (Bad compliance)
	17.3.98		182	7	68,640	ddI + d4T + RTV + SQV (Bad compliance)
	14.5.98		–	–	49,020	d4T + 3TC + RTV + SQV
	30.7.98		200	7	<500	d4T + 3TC + NVP + RTV + SQV
	13.8.98	8.98	182	7	309	d4T + 3TC + NVP + RTV + SQV (Bad compliance)
	29.9.98	9.98	234	6	28,420	d4T + 3TC + NVP + RTV + SQV (Bad compliance)
	26.10.98		–	–	–	ddI + d4T + NVP + RTV + NFV + HU
	4.1.99	1.99	–	–	61,050	ddI + d4T + NVP + RTV + NFV + HU
3	6.7.89		303	–	–	AZT
	21.9.90		252	–	–	AZT
	25.2.93		134	–	–	AZT
	4.6.93		120	–	–	AZT + ddC
	28.2.94	1-1 to 1-20	102	–	154,190	AZT + ddC
	16.5.94		91	–	–	AZT + ddC
	6.10.94		47	–	–	AZT + ddC
	19.1.95	2-1 to 2-20	35	–	68,630	ddI
	22.6.95	3-1 to 3-20	21	–	156,210	AZT + 3TC
	28.8.95	4-1 to 4-20	30	–	143,300	AZT + 3TC
	26.9.95	5-1 to 5-20	26	–	162,800	AZT + 3TC
	16.5.96	6-1 to 6-20	15	–	50,100	AZT + 3TC
	22.6.96		13	–	–	AZT + 3TC + IDV
	9.96		–	–	–	Exitus

[a] Abbreviations: –, not determined; AZT, zidovudine; ddI, didanosine; ddC, zalcitabine; d4T, stavudine; 3TC, lamivudine; RTV, ritonavir; SQV, saquinavir; NFV, nelfinavir; IDV, indinavir; NVP, nevirapine; HU, hydroxyurea.
[b] Sampling dates are written as day.month.year.
[c] All samples were of viral RNA, except sample 8.98-DNA from case 1, which corresponds to a proviral DNA extracted from PBMCs. Sample identification is written as month.year in cases 1 and 2, and sample-clone number in case 3.

RT insertion (Figure 3). Sample 2, which as sample 1 had a consensus sequence with T69SSS RT (Table 2), contained one clone (2-17) without insertion, which clustered with all clones (except three) from sample 3. Conversely, sample 3, with a consensus sequence without RT insertion (Table 2), harboured three clones (3-3, 3-5 and 3-8) with a T69SSS genotype. These minority sequences from sample 3 clustered with all the insert-containing sequences from samples 4, 5, and 6. Therefore, minority genomes without insertion were present in populations dominated by insertion-coding gen-

Table 2. Amino acid substitutions found in consensus sequences of the protease (A) and reverse transcriptase (B, C) of sequential HIV-1 samples from cases 1 to 3

A. *Protease*

			Amino acid position[a]																									
			10	12	14	15	16	19	32	35	37	38	41	46	57	59	61	62	63	64	65	72	73	77	82	84	90	93
Case	Sample	Antiretroviral therapy[b]	L	T	K	I	G	L	V	E	N	L	R	M	R	Y	Q	I	P	I	E	I	G	V	V	I	L	I
1	11.94	AZT	–	–	(R)	–	–	–	–	–	S	–	K	–	–	–	–	–	**L**	–	–	–	–	–	–	–	–	–
	7.95	AZT	–	–	–	–	–	–	–	–	S	–	K	–	–	–	–	–	**L**	–	–	–	–	(**L**)	–	–	–	–
	12.95	AZT	–	–	–	–	–	–	–	–	S	–	K	–	–	–	–	–	**L**	–	–	–	–	–	–	–	–	–
	9.96	AZT + ddI	–	–	–	–	–	–	–	–	S	–	K	–	–	–	–	–	**L**	–	–	–	–	–	–	–	–	–
	5.97	AZT + ddI	–	–	–	–	–	–	–	–	S	–	K	–	–	–	–	–	**L**	–	–	–	–	–	–	–	–	–
	7.97	d4T + 3TC + RTV	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	**L**	–	–	–	–	–	–	–	–	–
	3.98	d4T + 3TC + SQV	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	**L**	–	–	–	–	–	–	–	–	–
	8.98-DNA	d4T + 3TC + SQV + NFV	–	–	–	–	–	–	–	–	–	S	–	–	K	–	–	–	**L**	–	–	–	–	–	–	–	–	–
2	9.95	AZT + ddI	–	–	–	–	–	–	–	D	–	–	–	–	(K)	–	–	–	–	–	–	V	–	**I**	–	–	–	L
	8.96	AZT + d4T + 3TC	–	–	–	–	–	–	–	D	–	–	–	–	–	–	–	–	–	–	–	V	–	**I**	–	–	–	L
	10.97	ddI + d4T + RTV + SQV (stopped)	–	–	–	–	–	–	**I**	–	–	–	–	**I**	–	–	–	–	–	–	–	V	–	–	**A**	–	–	L
	1.98	ddI + d4T + RTV + SQV	**I**	–	–	–	–	–	–	D	–	–	–	**I**	–	–	–	V	–	–	–	V	**S**	**I**	–	**V**	**M**	L
	8.98	d4T + 3TC + NVP + RTV + SQV	–	–	–	–	–	–	**I**	–	–	–	–	**I**	–	–	–	–	–	–	–	–	–	–	**A**	–	–	L
	9.98	d4T + 3TC + NVP + RTV + SQV (b.c.)	**I**	–	–	–	–	–	–	D	–	–	–	**I**	–	–	–	V	–	–	–	V	**S**	–	–	**V**	**M**	L
	1.99	ddI + d4T + NVP + RTV + NFV + HU	**I**	–	–	–	–	–	–	D	–	–	–	**I**	–	–	–	V	–	–	–	V	**S**	–	–	**V**	**M**	L
3	1 (2.94)	AZT + ddC	–	–	–	–	E	–	–	–	–	–	–	–	–	–	–	V	**E**	L	–	–	–	**I**	–	–	–	–
	2 (1.95)	AZT + ddC	–	–	–	–	E	V	–	–	–	–	–	–	–	–	(K)	V	**C**	–	–	–	–	**I**	–	–	–	–
	3 (6.95)	ddI	–	–	–	(L)	E	–	–	–	–	(P)	–	–	–	–	E	V	**Q**	–	–	–	–	**I**	–	–	–	–
	4 (8.95)	AZT + 3TC	–	K	–	(V)	E	–	–	–	–	–	–	–	K	–	E	V	**Q**	–	D	–	–	**I**	–	–	–	–
	5 (9.95)	AZT + 3TC	–	–	–	–	E	–	–	–	–	–	–	–	–	–	N,D	V	**Q**,**H**	–	(D)	–	–	**I**	–	–	–	–
	6 (5.96)	AZT + 3TC	–	–	–	–	E	–	–	–	–	(P)	–	–	–	(S)	–	V	**Q**	–	–	(V)	–	**I**	–	–	–	–

(*continued*)

Table 2 Continued

B. *Reverse transcriptase (aa 1–120)*

Case	Sample	Antiretroviral therapy[b]	Amino acid position (1–120)[a]																						
			20	41	46	49	53	54	57	61	67	68	69	ins.	70	72	83	88	90	100	102	103	106	109	118
			K	M	K	K	E	N	N	F	D	S	T	–	K	R	R	W	V	L	K	K	V	L	V
1	11.94	AZT	–	–	–	R	–	–	–	S	–	–	–	–	–	–	–	–	I	–	R	–	–	–	–
	7.95	AZT	–	–	–	–	–	–	–	–	–	–	–	–	**R**	–	–	–	–	–	–	–	–	–	–
	12.95	AZT	–	–	–	–	–	–	–	–	–	–	–	–	**R**	–	–	–	–	–	–	–	–	–	–
	9.96	AZT + ddI	–	–	–	–	–	–	–	–	**N**	–	–	–	**R**	(G)	–	–	–	(V)	–	–	(**L**)	(R)	–
	5.97	AZT + ddI	(R)	**L**	–	–	–	–	–	–	–	–	**S**	**SS**	**R**	–	–	–	–	–	–	–	–	–	(**F**)
	7.97	d4T + 3TC + RTV	–	–	–	R	–	(K)	–	–	–	–	–	–	–	–	–	–	I	–	(R)	–	–	–	–
	3.98	d4T + 3TC + SQV	–	–	–	R	–	–	–	–	–	–	–	–	–	–	–	–	I	–	–	–	–	–	–
	8.98-DNA	d4T + 3TC + SQV + NFV	–	–	–	–	–	–	–	–	–	–	–	–	(**R**)	–	–	–	–	–	–	–	–	–	–
2	9.95	AZT + ddI	–	–	–	–	–	–	–	–	**N**	–	–	–	**R**	–	–	–	–	–	–	–	–	–	–
	8.96	AZT + d4T + 3TC	–	–	–	–	–	–	–	–	**N**	–	–	–	**R**	–	–	–	–	–	–	–	–	–	–
	10.97	ddI + d4T + RTV + SQV (stopped)	–	–	–	–	–	–	–	–	**N**	–	**D**	–	**R**	–	–	–	–	–	–	–	–	–	–
	1.98	ddI + d4T + RTV + SQV	–	**L**	–	–	–	–	–	–	**N**	–	**D**	–	–	–	–	–	–	–	–	–	–	–	–
	8.98	d4T + 3TC + NVP + RTV + SQV	–	–	–	–	–	–	–	–	**N**	–	**D**	–	**R**	–	–	–	–	–	–	–	–	–	–
	9.98	d4T + 3TC + NVP + RTV + SQV (b.c.)	–	**L**	–	–	–	–	S	–	**N**	–	**D**	–	–	–	–	–	–	–	–	**N**	–	–	**I**
	1.99	ddI + d4T + NVP + RTV + NFV + HU	–	**L**	–	–	–	–	–	–	**N**	–	**D**	–	–	–	–	–	–	–	–	–	–	–	**I**
3	1 (2.94)	AZT + ddC	–	**L**	–	–	–	–	–	–	–	–	**S**	**SS**	–	–	K	–	–	–	–	–	–	–	–
	2 (1.95)	AZT + ddC	–	**L**	(E)	–	(G)	–	–	(S)	–	–	**S**	**SS**	–	–	K	S	–	–	–	–	–	–	–
	3 (6.95)	ddI	–	**L**	–	–	–	–	–	(S)	–	–	**S**	–	–	–	K	–	–	–	–	–	–	–	–
	4 (8.95)	AZT + 3TC	–	**L**	–	–	–	–	–	–	–	(N)	**S**	**SG**	–	–	–	S	–	–	–	–	–	–	–
	5 (9.95)	AZT + 3TC	–	**L**	–	–	–	–	–	–	–	–	**S**	**SG**	–	–	–	S	–	–	–	–	–	–	–
	6 (5.96)	AZT + 3TC	–	**L**	–	–	–	–	–	–	–	–	**S**	**SG**	–	–	–	S	–	–	–	–	–	–	–

(*continued*)

Table 2 Continued

C. *Reverse transcriptase (aa 121–220)*

Case	Sample	Antiretroviral therapy[b]	Amino acid position (121–220)[a]																		
			122	123	125	127	135	162	163	166	169	177	178	184	186	188	196	202	210	211	214	215
			K	D	R	Y	T	S	S	K	E	D	I	M	D	Y	G	I	L	R	L	T
1	11.94	AZT	–	–	–	–	I	–	–	–	D	–	–	–	–	–	–	–	–	K	F	–
	7.95	AZT	E	N	–	–	I	–	–	–	D	–	–	–	–	–	–	–	–	K	F	–
	12.95	AZT	–	–	–	–	I	(C)	–	–	D	–	–	–	–	–	–	–	–	–	F	Y
	9.96	AZT + ddI	–	–	–	(F)	V	C	–	–	–	(E)	(M)	–	–	–	–	–	–	–	F	**Y**
	5.97	AZT + ddI	–	–	(K)	–	V	C	(T)	–	–	(E)	(M)	–	–	–	–	–	–	–	F	**Y**
	7.97	d4T + 3TC + RTV	–	–	–	–	I	–	–	–	–	–	–	–	–	–	–	–	–	K	F	–
	3.98	d4T + 3TC + SQV	–	–	–	–	I	–	–	–	D	–	–	–	–	–	–	–	–	K	F	–
	8.98-DNA	d4T + 3TC + SQV + NFV	–	–	–	–	I	C	–	–	–	–	–	–	–	–	–	–	–	–	F	–
2	9.95	AZT + ddI	–	–	–	–	I	–	–	–	–	–	–	–	–	–	–	–	–	K	F	**Y**
	8.96	AZT + d4T + 3TC	–	–	–	–	I	–	–	–	–	–	–	(**V**)	–	–	–	–	–	K	F	**Y**
	10.97	ddI + d4T + RTV + SQV (stopped)	E	–	–	–	I	–	–	–	–	–	–	**V**	–	–	–	–	–	K	F	**Y**
	1.98	ddI + d4T + RTV + SQV	–	–	–	–	I	–	–	–	–	–	–	–	–	–	–	–	**W**	K	F	**Y**
	8.98	d4T + 3TC + NVP + RTV + SQV	E	–	–	–	I	–	–	–	–	–	–	**V**	–	–	–	–	–	K	F	**Y**
	9.98	d4T + 3TC + NVP + RTV + SQV (b.c.)	–	–	–	–	I	–	–	–	–	–	–	**V**	–	–	–	–	**W**	K	F	**Y**
	1.99	ddI + d4T + NVP + RTV + NFV + HU	–	–	–	–	I	–	–	–	–	–	–	–	–	**L**	–	–	**W**	K	F	**Y**
3	1 (2.94)	AZT + ddC	–	–	–	F	I	–	–	R	–	N	–	–	–	–	E	–	–	–	F	**Y**
	2 (1.95)	AZT + ddC	–	–	–	F	I	–	–	R	–	–	–	–	Y	–	E	V	–	–	F	**Y**
	3 (6.95)	ddI	–	–	–	F	I	–	–	R	–	–	–	–	–	–	E	–	–	–	F	**Y**
	4 (8.95)	AZT + 3TC	–	–	–	–	I	–	–	R	–	E	–	–	–	–	E	–	–	–	F	**Y**
	5 (9.95)	AZT + 3TC	–	–	–	F	I	–	–	R	–	E	–	–	–	–	E	–	–	–	F	**Y**
	6 (5.96)	AZT + 3TC	–	–	–	F	I	–	–	R	–	E	–	–	–	–	E	–	–	–	F	**Y**

Sample identification, treatment history and clinical events of the patients are summarized in Table 1.

[a] The single letter amino acid (aa) code is used. Dash means no change relative to the reference HIV-1 sequence shown at the top (HIV-1 CAM-1 strain, GenBank accession number D10112).[50] Amino acids in parenthesis indicate that the residue is found in a mixture with the amino acid of the reference sequence shown at the top in a proportion of about 50% (range 40–60% according to the sequence peak pattern). Data from cases 1 and 3 are adapted from Briones *et al.*[51] Amino acid substitutions at positions relevant to resistance to PRI or RTI have been highlighted with bold-face letters.[29,30]

[b] Current therapy at the sampling date is indicated. Abbreviation: b.c., bad compliance. Dates of antiretroviral therapy introduction and drugs abbreviations are shown in Table 1.

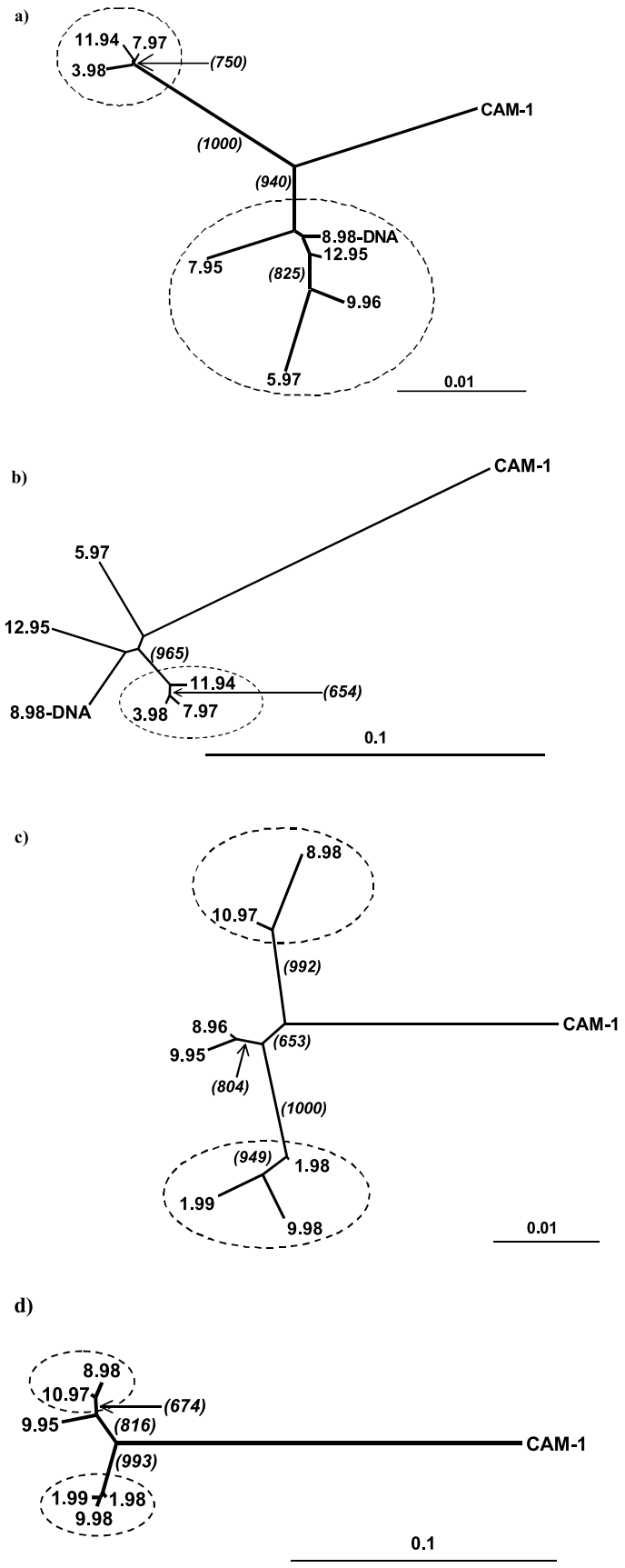

Figure 2. Neighbor-Joining trees of *pol* ((a) case 1; (c) case 2) and *env* ((b) case 1; (d) case 2) sequences. Amplification of the *env* region was not possible for samples 7.95 and 9.96 in case 1 (b) and for sample 8.96 in case 2 (d). Bootstrap values (1000 replicas) higher than 0.65 per unit are indicated in parenthesis. Clearly defined groups of sequences have been encircled. The bar represents 0.01 ((a) and (c)) or 0.1 ((b) and (d)) substitutions per nucleotide. Genetic distances can be calculated from the branch lengths; exact values of them will be provided upon request. The same topology was found using other phylogenetic approaches such as UPGMA, maximum parsimony and maximum likelihood (data not shown).

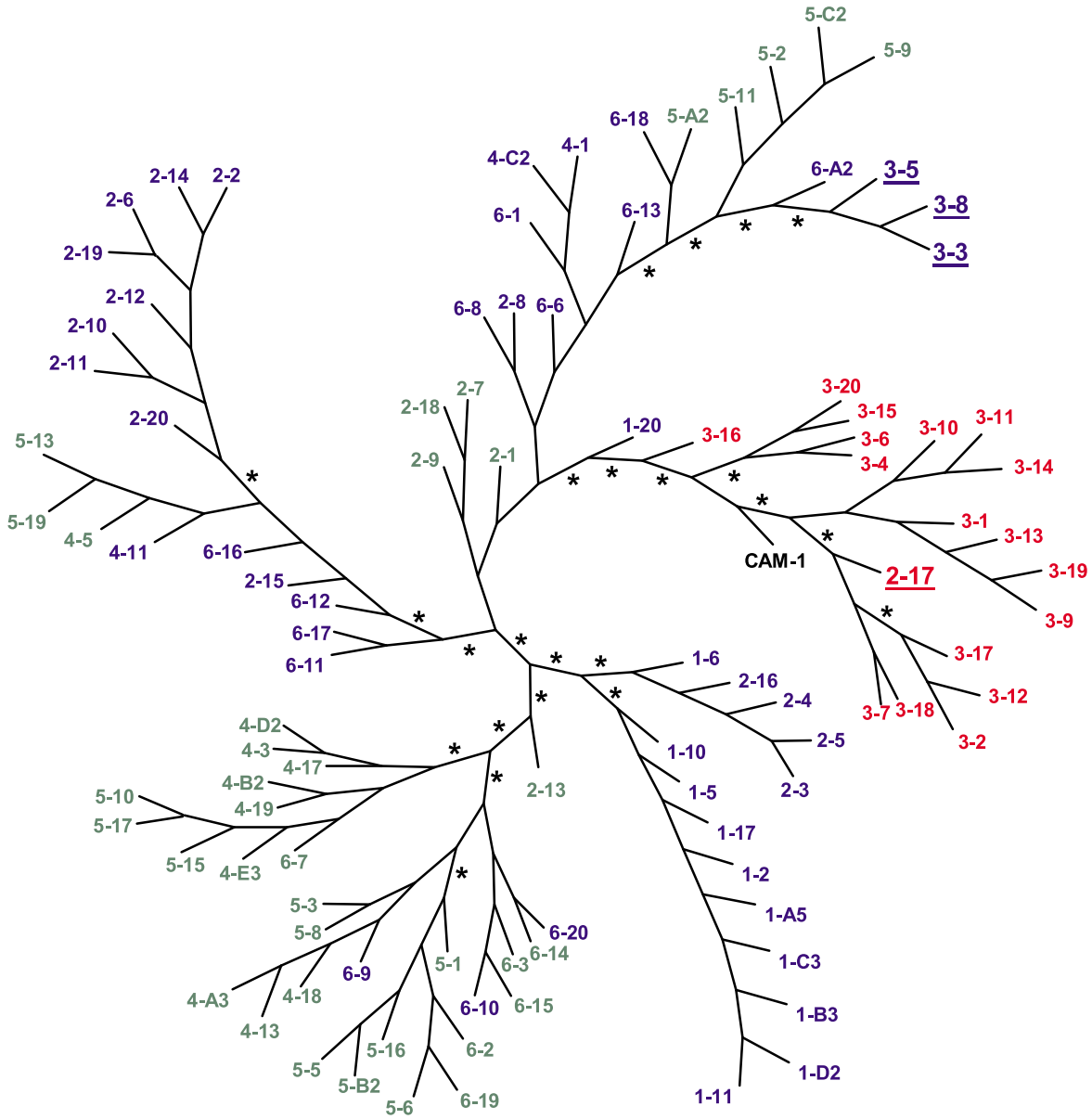

Figure 3. Maximum parsimony clustering of the 120 sequences of molecular clones from six sequential samples from case 3, using HIV-1 CAM-1 as reference sequence.[50] Unique sequences are identified as sample-clone number (see Table 1). Groups of clones with identical sequences are labelled as sample-letter-number: the letter is consecutive for each group in one sample and the number indicates how many identical clones belong to that group. Sample 1 (February 1994) showed four groups of identical sequences: 1-A5 (clones 1-1, 1-4, 1-7, 1-14 and 1-16), 1-B3 (1-3, 1-9 and 1-12), 1-C3 (1-8, 1-18 and 1-19) and 1-D2 (1-13 and 1-15). Samples 2 (January 1995) and 3 (June 1995) had all their clonal sequences different from each other. Sample 4 (August 1995) showed five groups of identical sequences: 4-A3 (4-2, 4-6 and 4-20), 4-B2 (4-4 and 4-12), 4-C2 (4-7 and 4-16), 4-D2 (4-8 and 4-9) and 4-E3 (4-10, 4-14 and 4-15). Sample 5 (September 1995) had three groups of identical sequences: 5-A2 (5-4 and 5-12), 5-B2 (5-7 and 5-20) and 5-C2 (5-14 and 5-18). Sample 6 (May 1996) showed one group of identical sequences: 6-A2 (6-4 and 6-5). Therefore, the 120 clones were represented by 100 different sequences in the clustering. Five defective clones detected in sample 5 were 5-2, 5-9, 5-11 and those in group 5-C2. Colour code: blue, sequences with the RT insertion T69SSS; green, sequences with the RT insertion T69SSG; red, sequences without RT insertions. Asterisks indicate clusters for which a bootstrap analysis (1000 replicas) gave a significance higher than 0.98 per unit. Minority sequences in samples 2 (2-17) and 3 (3-3, 3-5 and 3-8) have been underlined.

omes, and, conversely, minority insertion-coding genomes were present in populations dominated by viruses devoid of the insertion. These results suggest the presence of a replicative memory in HIV-1 quasispecies.

Quantitative model of quasispecies dynamics involving replicative and non-replicative memory

The results of sequential analyses of HIV-1

sequences from three patients (Tables 1 and 2, and Figures 2 and 3) suggest that both a replicative memory of the type described previously for FMDV,[6,8] together with a non-replicative (or reservoir) memory are in operation during HIV-1 replication *in vivo*. In order to describe quantitatively these two types of memory, we made use of a mathematical model based on the theory of quasispecies[4,31,32] involving a three-component system. Although some very detailed models of the dynamics of HIV-1 infection (considering the immune response of the host) have been developed[23,28,33–43] we have focused here on viral quasispecies dynamics with regard to the emergence of replicative and non-replicative memory.

We consider a population with three different viral components $j \in \{1,2,3\}$ and denote by $n_j(t)$ the number of viruses of component j at time t, and by k_j the rate of reproduction of component j (the viruses of type j reproduce at a constant rate, given by k_j offspring per parent per unit time). We introduce the quality factor q_j ($q_j \in [0,1]$) : the fraction of faithful copies of j; the rest, $1-q_j$, are mutants from j to other components k. The constant rate of mutation from j to k is termed μ_{kj}. We assume μ_{kj} to be a constant $\mu_{kj} = \mu$, with μ in the range from 10^{-5} to 10^{-3}. The differential equation satisfied by $n_j(t)$, assuming a constant population size and no degradation, is:[4,31,32]

$$\frac{dn_j(t)}{dt} = (k_j q_j - \bar{k}(t))n_j(t) + \sum_{k\neq j} \mu_{jk} n_k(t) \quad (1)$$

where $\sum_{j=1}^{3} n_j(t) = \kappa$ is the total population size, $\bar{k}(t) = \sum_{j=1}^{3} k_j n_j(t)/\kappa$, and the rates of reproduction, quality factors, and mutation rates are not independent quantities, but are related by the identity:

$$k_j(1-q_j) = \sum_{k\neq j} \mu_{kj} \quad (2)$$

In the case of three viral species, and from equality (2), we have:

$$\frac{dn_1(t)}{dt} = (k_1 - 2\mu - \bar{k}(t))n_1(t) + \mu n_2(t) + \mu n_3(t) \quad (3)$$

$$\frac{dn_2(t)}{dt} = (k_2 - 2\mu - \bar{k}(t))n_2(t) + \mu n_1(t) + \mu n_3(t) \quad (4)$$

$$\frac{dn_3(t)}{dt} = (k_3 - 2\mu - \bar{k}(t))n_3(t) + \mu n_1(t) + \mu n_2(t) \quad (5)$$

where:

$$\bar{k}(t) = \frac{k_1 n_1(t) + k_2 n_2(t) + k_3 n_3(t)}{n_1(t) + n_2(t) + n_3(t)}$$

In the case of only two species contributing to the quasispecies, the equations are slightly different, and condition (2) implies:

$$\frac{dn_1(t)}{dt} = \left[k_1 - \mu - \frac{k_1 n_1(t) + k_2 n_2(t)}{n_1(t) + n_2(t)}\right] n_1(t) + \mu n_2(t) \quad (6)$$

$$\frac{dn_2(t)}{dt} = \left[k_2 - \mu - \frac{k_1 n_1(t) + k_2 n_2(t)}{n_1(t) + n_2(t)}\right] n_2(t) + \mu n_1(t) \quad (7)$$

The system of equation (1) can be written in a more compact form: $\dot{n} = Mn - k1n$, where the dot means a time derivative, and we have introduced the matrix M. We are interested in the steady-state quasispecies solution.[31] A solution is provided by diagonalizing the matrix M, which means finding a diagonal matrix λ and a non-singular matrix O, such that $\lambda = OMO^{-1}$ and $M = O^{-1}\lambda O$. If the quasispecies is made up of N different components, the late time limit of $n_j(t)$, which we denote by $\bar{n}_j$, is given by:

$$\lim_{t\to+\infty} \frac{n_j(t)}{\sum_{i=1}^{N} n_i(t)} \equiv \frac{\bar{n}_j}{\sum_{i=1}^{N} \bar{n}_i} = \frac{O_{j1}^{-1}}{\sum_{i=1}^{N} O_{i1}^{-1}} \quad (8)$$

where we have assumed that component 1 of the quasispecies is fitter than the rest $N-1$ ($k_1 > k_j > 0$). The columns of the matrix O^{-1} correspond to the components of the eigenvectors of M. Therefore, the computation of the relative fractions (given by equation (8)) requires only the largest eigenvalue λ_1 and its corresponding eigenvector $\hat{v}_1$, which is the first column of the matrix O^{-1}.

In the case $N = 2$, the eigenvalue equation is the polynomial:

$$\lambda^2 - \lambda(m_{11} + m_{22}) + m_{11}m_{22} - m_{12}m_{21} = 0 \quad (9)$$

with $m_{11} = k_1 - \mu$, $m_{12} = \mu$, $m_{21} = \mu$, and $m_{22} = k_2 - \mu$.

The relative fractions of the quasispecies are given by (as by Ruíz-Jarabo *et al.*[8]):

$$\frac{\bar{n}_1}{\bar{n}_1 + \bar{n}_2} = \frac{\mu}{(\lambda_1 - m_{11} + \mu)} \quad (10)$$

$$\frac{\bar{n}_2}{\bar{n}_1 + \bar{n}_2} = \frac{(\lambda_1 - m_{11})}{(\lambda_1 - m_{11} + \mu)} \quad (11)$$

In the case $N = 3$, the matrix M that needs to be diagonalized is:

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} = \begin{pmatrix} k_1 - 2\mu & \mu & \mu \\ \mu & k_2 - 2\mu & \mu \\ \mu & \mu & k_3 - 2\mu \end{pmatrix} \quad (12)$$

The corresponding eigenvalue equation is:

$$\lambda^3 + A_2\lambda^2 + A_1\lambda + A_0 = 0 \tag{13}$$

where the coefficients are given by:

$$A_0 = -3\mu^2(k_1 + k_2 + k_3) + 2\mu(k_1k_2 + k_2k_3 + k_1k_3) - k_1k_2k_3 \tag{14}$$

$$A_1 = 9\mu^2 - 4\mu(k_1 + k_2 + k_3) + (k_1k_2 + k_2k_3 + k_1k_3) \tag{15}$$

$$A_2 = 6\mu - (k_1 + k_2 + k_3) \tag{16}$$

The solution for the largest eigenvalue, λ_1, is given by:

$$\lambda_1 = 2\sqrt{-Q}\cos\left(\frac{\theta}{3}\right) \tag{17}$$

with $Q = (3A_1 - A_2^2)/9$ and $\cos\theta = -R/\sqrt{-(Q^3)}$, where $R = (9A_2A_1 - 27A_0 - 2A_2^3)/54$. In terms of this eigenvalue the relative fractions of the steady-state quasispecies are given by:

$$\frac{\bar{n}_1}{\sum_{i=1}^{3}\bar{n}_i} = \frac{\mu}{\lambda_1 - m_{11} + \mu} \tag{18}$$

$$\frac{\bar{n}_2}{\sum_{i=1}^{3}\bar{n}_i} = \frac{\mu}{\lambda_1 - m_{22} + \mu} \tag{19}$$

$$\frac{\bar{n}_3}{\sum_{i=1}^{3}\bar{n}_i} = \frac{(\lambda_1 - m_{11})(\lambda_1 - m_{22}) - \mu^2}{(\lambda_1 - m_{11} + \mu)(\lambda_1 - m_{22} + \mu)} \tag{20}$$

A numerical approach allowed solving the system of differential equations (equations (6) and (7) in the case $N = 2$, and equations (3)–(5) in the case $N = 3$) given initial and intermediate conditions. We start at time t_0 with only viral components 1 and 2 present, at arbitrary initial conditions $n_1(t_0)$ and $n_2(t_0)$, respectively (the late time limit does not depend on the initial conditions). Rates of reproduction are $k_1 = 1.0 > k_2$ (all values of k_i are relative to the highest one, which is always considered as $k = 1.0$). From time t_0 to time t_1 we let the system evolve, and assume that t_1 is sufficiently late so that $n_1(t_1) = \bar{n}_1$ and $n_2(t_1) = \bar{n}_2$ are the quasispecies values for the corresponding choice of parameters k_1 and k_2, respectively. The evolution from t_1 to t_2 is characterized by $\bar{n}_1$ and $\bar{n}_2$. At a later time t_2 the system is exposed to a selective pressure, with component 1 being affected and component 2 not affected. In the particular case of dynamics of HIV-1 *in vivo* such selective pressure can be identified with an antiviral treatment, so that 1 is sensitive to the treatment, whereas 2 is resistant. This corresponds to the situation $k_2 = 1.0 > k_1$. After a short period δ, that is, at $t_2 + \delta$, the selective pressure is withdrawn, expecting to recover $k_1 = 1.0 > k_2$, with k_2 at $t_2 + \delta$ being larger than the initial k_2 value at t_0, which means that component 2 has become fitter. With these new values for k_1 and k_2 we let the system evolve to a sufficient late time t_3. At t_3, $n_1(t_3)$ and $n_2(t_3)$ have reached their new quasispecies values (as the value of the constants has changed). In particular, $\bar{n}_1(t_3) < \bar{n}_1(t_1)$ and $\bar{n}_2(t_3) > \bar{n}_2(t_1)$. The value $\bar{n}_2(t_3)/\kappa$ corresponds to the level of "molecular" or replicative memory of component 2. The evolution from t_3 to t_4 is characterized by these quasispecies values. At a later time t_4 we apply a second selective pressure (a different antiviral treatment), which favours a new component 3 to move from a reservoir to the bloodstream (thus the circulating quasispecies includes now three components). We assume that component 1 is sensitive to the treatment, components 2 and 3 are resistant, and 3 is fitter than 2 (this is the reason why component 3 has moved from a reservoir compartment to the bloodstream), so $k_3 = 1.0 > k_2 > k_1$. If this second treatment is maintained after t_4, component 3 will remain the majority in the quasispecies, and will be recognized as the consensus sequence. This corresponds to the situation reported for cases 1 and 2, where the re-emergence of ancestral "component 3" genomes was detected as the consensus sequence, following a long dominance of "component 1" genomes. If we discontinue this second treatment at a time $t_4 + \delta$, the system will recover the fitness ratio $k_1 = 1.0 > k_3 > k_2$ and, at a later time t_5, sufficient for the system to reach its steady-state, we will obtain new quasispecies values for this set of parameters: $\bar{n}_1(t_5)$, $\bar{n}_2(t_5)$ and $\bar{n}_3(t_5)$. The quasispecies level $\bar{n}_3(t_5)/\kappa$ would correspond to the contribution of non-replicative (or reservoir) components to the quasispecies memory.

We have run three numerical computations corresponding to three different values of the mutation rate, namely $\mu = 10^{-3}$, 10^{-4} and 10^{-5}. Figure 4 depicts one example of the time evolution of the system for $\mu = 10^{-3}$. Therefore, the theoretical treatment of quasispecies dynamics permits modelling of two types of memory during HIV-1 replication, in agreement with experimental analyses of HIV-1 genome evolution in patients undergoing antiretroviral therapy.

Discussion

The presence of memory genomes in viral quasispecies, documented with designed experiments employing FMDV,[6,8,9] is an expected consequence of quasispecies dynamics.[9,31,44–46] For retroviruses such as HIV-1, the level of complexity of the replicating quasispecies at any one time is expected to be greater than for non-retroviral riboviruses because proviral sequences (maintained with the minimal occurence of point mutations associated with cellular genes)[13] can be activated to perturb

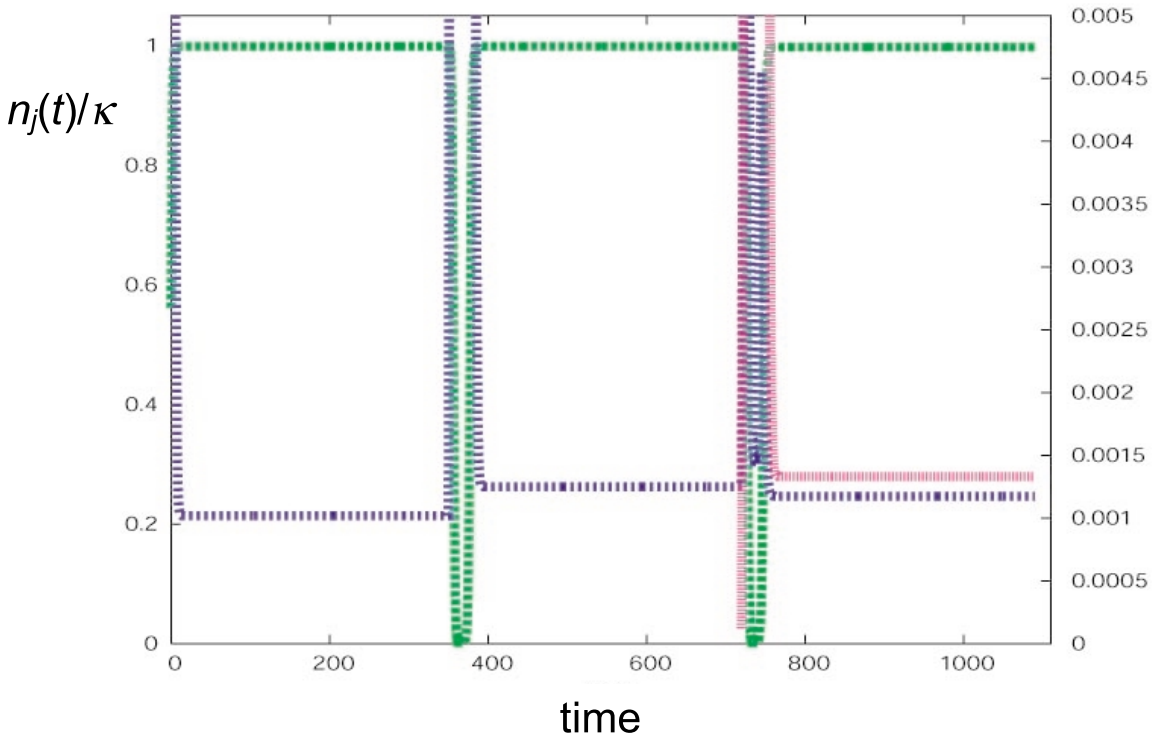

Figure 4. Dynamics of three-component viral quasispecies. The plot gives relative fractions of viral components ($n_j(t)/\kappa$) as a function of time for $\mu = 10^{-3}$. The left y axis corresponds to $n_1(t)/\kappa$ (green lines), and the right y axis corresponds to both $n_2(t)/\kappa$ (blue lines) and $n_3(t)/\kappa$ (pink lines). Notice the different scales on the left and right y axes. Values at the right y axis decrease by a power of 10 each time μ decreases by a power of 10 (simulations with $\mu = 10^{-4}$ and $\mu = 10^{-5}$, not shown). The parameters chosen for this simulation are as follows: at time $t_0 = 0$, $n_1(t_0) = 500$, $n_2(t_0) = 500$, $k_1 = 1.0$ and $k_2 = 0.02$; at time $t_1 = 250$ the system has reached its steady-state (corresponding to the values of k_j at t_0); at time $t_2 = 350$, the two-component system is exposed to a selective pressure (antiviral treatment) and $k_1 = 0.01$ and $k_2 = 1.0$; at time $t_2 + \delta = 370$ the selective pressure is switched off and $k_1 = 1.0$ and $k_2 = 0.2$; at time $t_3 = 620$ the system has reached its steady-state (corresponding to the values of k_j at $t_2 + \delta$), and the value of $n_2(t_3)/\kappa$ corresponds to the level of replicative memory of component 2 (which has a level of 1.02×10^{-3} in this example, with $\mu = 10^{-3}$); at time $t_4 = 720$ the system is exposed to a second selective pressure (a different antiviral treatment) that forces a new component 3 to move from a reservoir to the bloodstream (the initial condition for component 3 was chosen as $n_3(t_4) = 10^{-4}[n_1(t_4) + n_2(t_4)]$), and $k_1 = 0.03$, $k_2 = 0.3$, $k_3 = 1.0$; at time $t_4 + \delta = 740$ the selective pressure is switched off and $k_1 = 1.0$, $k_2 = 0.15$, and $k_3 = 0.25$; at time $t_5 = 990$ the system has reached its steady-state (corresponding to the values of k_j at $t_4 + \delta$), and the value of $n_3(t_5)/\kappa$ reflects the contribution of non-replicative (or reservoir) components to the quasispecies memory (which has a level of 1.17×10^{-3} in this simulation, with $\mu = 10^{-3}$).

the dynamics of replicating retroviral particles. Evidence of re-emergence of ancestral viral sequences was previously obtained for HIV-1,[19,21,47] hepatitis C virus[48] and hepatitis B virus[49] (review by Domingo *et al.*[3]).

Phylogenetic methods have been extensively applied to the molecular epidemiology of viruses, including the origin of HIV, its spread to different geographical areas, and the impact of viral mutation, recombination and host interactions on HIV-1 population dynamics.[50–56] In the present study we have used such procedures to provide evidence for the presence of memory in HIV-1 quasispecies from three patients subjected to successions of well defined selective antiviral pressures. Case 1 documented the re-emergence of viral genomes that could have been hidden in cellular or anatomical body compartments for 32 months. These memory viruses lacked all the nucleoside resistance mutations, including the insertion T69SSS, that had progressively accumulated in the circulating quasispecies. The introduction of a potent HAART regime (d4T + 3 TC + RTV) would have eliminated most of the circulating viruses while ancestral, probably fitter, viruses would have moved from the reservoirs to the bloodstream. A previous clonal analysis[51] did not provide evidence that minority members of the replicating quasispecies survived the treatment, although such minority could have been below the detection level of the analysis. Case 2 revealed that re-emergence of viral genomes can occur in two consecutive rounds, as a result of treatment interruption or incomplete adherence. In case 3, MP analysis of molecular clones indicated that minority genomes without an RT insertion in sample 2 (5% of the total quasispecies) became majority (85%) in sample 3, probably due

to their higher fitness after six months in the absence of AZT + ddC therapy and in the presence of ddI. Minority genomes that maintained the T69SSS genotype in sample 3 (15% of the total quasispecies) were likely selected by the re-introduction of AZT (together with 3TC) and restored a quasispecies dominated by viruses with the RT insertion in sample 4 (collected two months after sample 3; Table 1). The inserts in RT were probably maintained during the rest of the infection due to the multi-nucleoside resistance they confer to HIV-1.[57–60] Therefore, minority components of the mutant spectra of a HIV-1 quasispecies can be maintained in memory as a record of the previous evolutionary history, and can become dominant when the same (or a similar) selective pressure that originated them is reintroduced. The case analysed here has shown the presence of minority memory genomes in samples 2 and 3 that then originated the majority of mutant spectrum in samples 3 and 4, respectively.

These experimental results strongly suggest the participation of memory genomes in successive dominances of different mutant distributions. The presence of memory HIV-1 genomes, harbouring constellations of mutations associated with resistance to antiretroviral inhibitors different from those mutations present in dominant genomes, confer the virus a selective advantage to respond more rapidly whenever a change in treatment is implemented. Probably, inhibitor-based environmental fluctuations favoured the identification of the two kinds of memory genomes in the cases studied. However, it must be considered that other environmental fluctuations (either internal, due to physiological alterations, or external, such as treatment interruptions) may also favour memory genomes as selective reservoirs to gain dominance in the replicating pool of viruses.

The mathematical model developed here contemplates both the replicative and non-replicative (or reservoir) memory. It is based on the theory of quasispecies, and considers a three-component model in which the third component comes from a latent reservoir after activation by a biological stimulus (opportunistic pathogen or other). The distinction between the contributions of both memory sources is not obvious, because the irruption of a third (non-replicative) component from the latent reservoir may resemble the rise of a replicative minority genome to participate in the multi-component dynamics of the circulating HIV-1 quasispecies. The numerical solution of the time evolution of the system of equations (involving the three components) has shown that both replicative and non-replicative memory produce minority genomes in the steady- state, which to a first approximation (due to the small value of μ when compared to k_j) are proportional to μ (for $\mu = 10^{-3}$, they are 1.02×10^{-3} and 1.17×10^{-3}, respectively; Figure 4). Nevertheless, the mathematical model is an approximation without a dynamical mechanism for k_j, which could be relevant for the quasispecies dynamics *in vivo*.

Despite the fact that reservoir memory has its most immediate consequences in the replicative dynamics of retroviruses, it must be pointed out that its implications may extend to other RNA viruses whenever multiple replication sites occur within an organism. Some RNA viruses appear to be highly monotropic, such as human hepatitis viruses, which replicate mainly in hepatocytes (reviews of viral systems see Domingo[56]). However, other RNA viruses (several picornaviruses, flaviviruses, coronaviruses, etc.) may target different organs and cell subsets providing a complex mosaic of quasispecies distributions. Our model will apply whenever slower replicating mutant distributions can increase their replication rate (for example, as a result of immune-escape or local physiological alterations that increase permissivity to viral replication) and enter into competition with other mutant distributions. Additional studies on quasispecies memory of non-retroviral RNA viruses *in vivo* are needed before applicability of the model presented here can be assessed.

These results with HIV-1 emphasize that mutant spectra in evolving quasispecies may not only be representations of mutant distributions generated as a result of mutational pressure and relative fitness of the mutants generated, but may also include some history-dependent sub-structuring in the composition of genomes. From a practical point of view, this encourages the development of analytical procedures to determine at any one time not only the dominant genomic sequences but also minority genomes that may reflect the evolutionary history of the virus and that may provide the basis for rapid adaptation.[61] Microarray-based technologies offer the potential to detect minority sequences that are in the range of 0.1–1% in a genome population.[62–64] In the case of HIV-1 this can have a predictive value regarding treatment efficacy.

Materials and Methods

Virus samples

HIV-1 samples were obtained from three patients undergoing highly active antiretroviral therapy (HAART) at Hospital Carlos III in Madrid, Spain. All of them were infected by B subtype strains of HIV-1, and they were not linked epidemiologically. The patient in case 1 was an intravenous drug user born in 1962 who became HIV-1 seropositive during the period 1980–1984. The patient in case 2 was born in 1958 and was infected by homosexual intercourse at some undetermined time before 1992. The patient in case 3 was a haemophiliac man born in 1970, infected at the age of 12 after receiving contaminated blood products.

Sequential retrospective samples from these three patients were collected, and plasma viremia was quantified using the Quantiplex bDNA assay[65] (Bayer, Barcelona, Spain), with detection limits of 500 and 50

HIV-RNA copies/ml (versions 2.0 and 3.0, respectively). The CD4 + count was measured by flow cytometry (Coulter, Madrid, Spain).

RNA extraction, reverse transcription, PCR amplification and nucleotide sequencing

Viral RNA was extracted from plasma samples by a silica-based method.[66] HIV-1 RNA was reverse-transcribed by AMV RT (Promega, Madison, WI) using either the primer NE135 (reverse 5′-CTTACTAACTTCT GTATGTCATTGACAGTCCAGCT-3′, the 5′ residue corresponds to position 3336 of HIV-1 CAM-1 genome[50]) for the *pol* gene, or the primer ED12m (reverse 5′-AGT GCTTCCTGCTGCTCCCAAGAACCCAA-3′, the 5′ residue corresponds to position 7812 of HIV-1 CAM-1) for the *env* gene. When appropriate, proviral DNA was extracted from peripheral blood mononuclear cells (PBMCs) by lysis with non-ionic detergents (Tween 20 and NP40) and proteinase K.

Nested PCR was used to amplify two fragments of the HIV-1 genome: (a) a region of 1086 bp of the *pol* gene including 18 bp preceding the start of the PR-coding region, the entire PR-coding region and the first 220 codons of the RT; (b) a region of 368 bp of the *env* gene including the hypervariable V3 loop of the gp120 protein. External primers used to amplify the *pol* region were PRF1 (forward 5′-CAGCCCCACCAGAAGA GAGC-3′,5′ at position 2158 of HIV-1 CAM-1) and NE135, and the inner primers were PRF2 (forward 5′-CAGAAGAAAGCTTCAGGTTTGGG-3′,5′ at position 2167 of HIV-1 CAM-1) and RT4 (reverse 5′-AGTTCATA ACCCATCCAAAG-3′, 5′ at position 3252 of HIV-1 CAM-1). Amplification of the *env* region was carried out with the external primers JA19d (forward 5′-CACAGTA CAATGTACACATG-3′, 5′ at position 6957 of HIV-1 CAM-1) and ED12m, and the inner ones 6601d (forward 5′-AATGGCAGTCTAGCAGAAG-3′, 5′ at position 7012 of HIV-1 CAM-1) and ED33 (reverse 5′-TTACAGTAGA AAAATTCCCCTC-3′, 5′ at position 7379 of HIV-1 CAM-1). All PCR reactions were performed using the high-fidelity, thermostable *Pfu* DNA polymerase (Promega, Madison, WI) according to the manufacturer's protocol.

Consensus nucleotide sequences were determined in both strands of purified, PCR-amplified DNA using the ABI PRISM BigDye Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA). Sequences of the *pol* fragment were obtained with the forward primers PRF2 and 55F (5′-CAAAAATTGGGCCTGAAAATCC-3′,5′ at position 2694 of HIV-1 CAM-1), and the reverse ones 51R (5′-ATTGTATGGATTTTCAGGCCC-3′, 5′ at position 2722 of HIV-1 CAM-1) and RT4. Sequences of the *env* region were obtained with the primers 6601d and ED33. Sequence alignments and assemblies into 975–981 bp (*pol*) and 306–315 bp (*env*) contigs were performed by means of the Lasergene v5.0 package (DNASTAR Inc, Madison, WI). Nucleotide sequences have been deposited in the GenBank database with accession numbers AY188555 to AY188581. Nucleotide sequencing of 120 molecular clones of six sequential samples extracted from case 3 (see Table 1) were described elsewhere,[51] and GenBank accession numbers of these sequences are AF168235 to AF168354.

Phylogenetic methods

Multiple alignments of consensus sequences and molecular clones were carried out with the program CLUSTAL W,[67] using HIV-1 CAM-1 (GenBank accession number D10112)[50] as the reference sequence. Evolutionary distances were estimated by the program DNADIST from the PHYLIP v3.5 package,[68] using the Kimura 2-parameter correction method that weighs transitions and transversions at 2:1. Tree topology was inferred by the Neighbor-Joining method (NJ)[69] using the program NEIGHBOR from PHYLIP v3.5. The topology of the trees was also determined by a different distance-based cladistic method, the "Unweighted Pair Group Method With Arithmetic Mean" (UPGMA)[70] using NEIGHBOR from PHYLIP v3.5.

Alternative procedures were used for measuring genetic relatedness among samples. Maximum parsimony (MP) methods (program DNAPARS from PHYLIP v3.5 package) were used to construct the tree topology that required the least number of sequence changes from one node to another.[71–73] Bootstrap re-sampling (1000 data sets) of the multiple alignment was used to test the statistical robustness of the trees obtained by NJ, UPGMA and MP methods.[74] Maximum likelihood (ML) trees were generated by the program PUZZLE,[75] using the Tamura-Nei substitution model[76] and the Gamma distributed rates with eight parameters (TN-8Γ) as heterogeneity model. The TN-8Γ model has been applied to phylogenetic analysis of human retroviruses.[52,77]

Numerical methods

A code written in C was developed to solve numerically the system of differential equations of the mathematical model. It incorporates a Runge–Kutta solver of second order[78,79] and run on a Linux workstation.

Acknowledgements

We thank V. Soriano for providing sequential HIV-1 samples, and A. Corral and G. Gómez-Mariano for technical assistance. We are indebted to A. Mas for the discussions on re-emergence of sequences in quasispecies, and to J. Pérez-Mercader and F. Morán for valuable suggestions. Work at CBMSO was supported by grants BMC 2001-1823-C02-01, CAM 08.2/0046/2000, and Fundación Ramón Areces. Work at CAB was supported by the European Union, Instituto Nacional de Técnica Aerospacial and Comunidad de Madrid. The experiments described in this paper are at the origin of a project with Genetrix S.L., supported by a PROFIT grant (reference FIT 010000-2002-38) from Ministerio de Ciencia y Tecnología.

References

1. Eigen, M. & Biebricher, C. K. (1988). Sequence space and quasispecies distribution. In *RNA Genetics* (Domingo, E., Ahlquist, P. & Holland, J. J., eds), vol. 3, CRC Press, Boca Raton, FL.
2. Eigen, M. (1996). On the nature of virus quasispecies. *Trends Microbiol.* **4**, 216–218.
3. Domingo, E., Biebricher, C., Holland, J. J. & Eigen, M.

(2001). *Quasispecies and RNA Virus Evolution: Principles and Consequences*, Landes, Austin, TX.
4. Eigen, M. (1971). Self-organization of matter and evolution of biological macromolecules. *Naturwissenschaften*, **58**, 465–523.
5. Eigen, M. (2002). Error catastrophe and antiviral strategy. *Proc. Natl Acad. Sci. USA*, **99**, 13374–13376.
6. Ruíz-Jarabo, C. M., Arias, A., Baranowski, E., Escarmís, C. & Domingo, E. (2000). Memory in viral quasispecies. *J. Virol.* **74**, 3543–3547.
7. Arias, A., Lázaro, E., Escarmís, C. & Domingo, E. (2001). Molecular intermediates of fitness gain of an RNA virus: characterization of a mutant spectrum by biological and molecular cloning. *J. Gen. Virol.* **82**, 1049–1060.
8. Ruíz-Jarabo, C. M., Arias, A., Molina-París, C., Briones, C., Baranowski, E., Escarmís, C. & Domingo, E. (2002). Duration and fitness dependence of quasispecies memory. *J. Mol. Biol.* **315**, 285–296.
9. Domingo, E. (2000). Viruses at the edge of adaptation. *Virology*, **270**, 251–253.
10. Meyerhans, A., Cheynier, R., Albert, J., Seth, M., Kwok, S., Sninsky, J. *et al.* (1989). Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. *Cell*, **58**, 901–910.
11. Crandall, K. A. (1999). *The Evolution of HIV*, The Johns Hopkins University Press, Baltimore, MD.
12. Goff, S. P. (2001). Intracellular trafficking of retroviral genomes during the early phase of infection: viral exploitation of cellular pathways. *J. Gene Med.* **3**, 517–528.
13. Friedberg, E. C., Walker, G. C. & Siede, W. (1995). *DNA Repair and Mutagenesis*, American Society of Microbiology, Washington.
14. Pantaleo, G., Graziosi, C. & Butini, L. (1991). Lymphoid organs function as major reservoirs for human immunodeficiency virus. *Proc. Natl Acad. Sci. USA*, **88**, 9838–9842.
15. Finzi, D., Hermankova, M. & Pierson, T. (1997). Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science*, **278**, 1295–1300.
16. Blankson, J. N., Persaud, D. & Siciliano, R. F. (2002). The challenge of viral reservoirs in HIV-1 infection. *Annu. Rev. Med.* **53**, 557–593.
17. Devereux, H. L., Burke, A., Lee, C. A. & Johnson, M. A. (2002). *In vivo* HIV-1 compartmentalisation: drug resistance-associated mutation distribution. *J. Med. Virol.* **66**, 8–12.
18. Ruff, C. T., Ray, S. C., Kwon, P., Zinn, R., Pendleton, A., Hutton, N. *et al.* (2002). Persistence of wild-type virus and lack of temporal structure in the latent reservoir for human immunodeficiency virus type 1 in pediatric patients with extensive antiretroviral exposure. *J. Virol.* **76**, 9481–9492.
19. Karlsson, A. C., Gaines, H., Sällberg, M., Lindbäck, S. & Sönnerborg, A. (1999). Reappearance of founder virus sequence in human immunodeficiency virus type 1-infected patients. *J. Virol.* **73**, 6191–6196.
20. Miller, V. (2001). Structured treatment interruptions in antiretroviral management of HIV-1. *Curr. Opin. Infect. Dis.* **14**, 29–37.
21. Kijak, G. H., Simon, V., Balfe, P., Vanderhoeven, J., Pampuro, S. E., Zala, C. *et al.* (2002). Origin of human immunodeficiency virus type 1 quasispecies emerging after antiretroviral treatment interruption in patients with therapeutic failure. *J. Virol.* **76**, 7000–7009.
22. Deeks, S. G., Grant, R. M., Wrin, T., Paxinos, E. E., Liegler, T., Hoh, R. *et al.* (2003). Persistence of drug-resistant HIV-1 after a structured treatment interruption and its impact on treatment response. *AIDS*, **17**, 361–370.
23. Perelson, A., Neumann, A. & Markowitz, M. (1996). HIV-1 dynamics *in vivo*: virion clearance rate, infected cell life-span, and viral generation time. *Science*, **271**, 1582–1586.
24. Perelson, A., Essunger, P. & Cao, Y. (1997). Decay characteristics of HIV-1 infected compartments during combination therapy. *Nature*, **387**, 188–190.
25. Ho, D. D. (1997). Dynamics of HIV-1 replication *in vivo*. *J. Clin. Invest.* **11**, 2565–2567.
26. Markowitz, M. H. (1997). HIV pathogenesis and viral dynamics. *Antiviral Ther.* **2**, 7–17.
27. Cavert, W., Notermans, D. & Staskus, K. (1997). Kinetics of response in lymphoid tissues to antiretroviral therapy of HIV-1 infection. *Science*, **276**, 960–964.
28. Wodarz, D. & Nowak, M. A. (1999). Dynamics of HIV Pathogenesis and Treatment. In *Origin and Evolution of Viruses* (Domingo, E., Webster, R. G. & Holland, J. J., eds), Academic Press, San Diego, CA.
29. Schinazi, R., Larder, B. & Mellors, J. (1999). Mutations in retroviral genes associated with drug resistance: 1999–2000 update. *Int. Antiviral News*, **7**, 46–49.
30. Yeni, P. G., Hammer, S. M., Carpenter, C. C., Cooper, D. A., Fischl, M. A., Gatell, J. M. *et al.* (2002). Antiretroviral treatment for adult HIV infection in 2002: updated recommendations of the International AIDS Society—USA Panel. *J. Am. Med. Assoc.*, **288**, 222–235.
31. Batschelet, E., Domingo, E. & Weissmann, C. (1976). The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene*, **1**, 27–32.
32. Rowe, G. (1994). *Theoretical Models in Biology: The Origin of Life, the Immune System, and the Brain*, Clarendon, Oxford, UK.
33. Bonhoeffer, S., May, R. M., Shaw, G. M. & Nowak, M. A. (1997). Virus dynamics and drug therapy. *Proc. Natl Acad. Sci. USA*, **94**, 6971–6976.
34. Wein, L. M., D'Amato, R. M. & Perelson, A. S. (1998). Mathematical analysis of antiretroviral therapy aimed at HIV-1 eradication or maintenance of low viral loads. *J. Theor. Biol.* **192**, 81–98.
35. Ribeiro, R. M. & Bonhoeffer, S. (1998). The frequency of resistant mutant virus before antiviral therapy. *AIDS*, **12**, 461–465.
36. Mittler, J. E., Sulzer, B., Neumann, A. U. & Perelson, A. S. (1998). Influence of delayed viral production on viral dynamics in HIV-1 infected patients. *Math. Biosci.* **152**, 143–163.
37. Ribeiro, R. M. & Bonhoeffer, S. (1999). A stochastic model for primary HIV infection: optimal timing of therapy. *AIDS*, **13**, 351–357.
38. Rouzine, I. M. & Coffin, J. M. (1999). Linkage disequilibrium test implies a large effective population number for HIV *in vivo*. *Proc. Natl Acad. Sci. USA*, **96**, 10758–10763.
39. Rouzine, I. M. & Coffin, J. M. (1999). Search for the mechanism of genetic variation in the pro gene of human immunodeficiency virus. *J. Virol.* **73**, 8167–8178.
40. Nelson, P. W., Murray, J. D. & Perelson, A. S. (2000). A model of HIV-1 pathogenesis that includes an intracellular delay. *Math. Biosci.* **163**, 201–215.
41. Ribeiro, R. M. & Bonhoeffer, S. (2000). Production of

resistant HIV mutants during antiretroviral therapy. *Proc. Natl Acad. Sci USA*, **97**, 7681–7686.
42. Stafford, M. A., Corey, L., Cao, Y., Daar, E. S., Ho, D. D. & Perelson, A. S. (2000). Modeling plasma virus concentration during primary HIV infection. *J. Theor. Biol.* **203**, 285–301.
43. Rouzine, I. M., Rodrigo, A. & Coffin, J. M. (2001). Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiol. Mol. Biol. Rev.* **65**, 151–185.
44. Eigen, M. & Schuster, P. (1977). The hypercycle. A principle of natural self-organization. Part A: emergence of the hypercycle. *Naturwissenschaften*, **64**, 541–565.
45. Biebricher, C. K., Eigen, M. & Gardiner, W. C., Jr (1985). Kinetics of RNA replication: competition and selection among self-replicating RNA species. *Biochemistry*, **24**, 6550–6560.
46. Biebricher, C. K. (1999). Mutation, competition and selection as measured with small RNA molecules. In *Origin and Evolution of Viruses* (Domingo, E., Webster, R. G. & Holland, J. J., eds), Academic Press, San Diego, CA.
47. Borrow, P., Lewicki, H., Wei, X., Horwitz, M. S., Peffer, N., Meyers, H. *et al.* (1997). Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nature Med.* **3**, 205–211.
48. Wyatt, C. A., Andrus, L., Brotman, B., Huang, F., Lee, D. H. & Prince, A. M. (1998). Immunity in chimpanzees chronically infected with hepatitis C virus: role of minor quasispecies in reinfection. *J. Virol.* **72**, 1725–1730.
49. Lau, G. K., Tsiang, M., Hou, J., Yuen, S., Carman, W. F. *et al.* (2000). Combination therapy with lamivudine and famciclovir for chronic hepatitis B-infected Chinese patients: a viral dynamics study. *Hepatology*, **32**, 394–399.
50. Kuiken, C. L., Foley, B., Hahn, B., Korber, B., McCutchan, F., Marx, P. A., *et al.* (2000). Editors of *Human Retroviruses and AIDS: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*, Los Alamos National Laboratory, Los Alamos, NM.
51. Briones, C., Mas, A., Gómez-Mariano, G., Altisent, C., Menéndez-Arias, L., Soriano, V. & Domingo, E. (2000). Dynamics of dominance of a dipeptide insertion in reverse transcriptase of HIV-1 from patients subjected to a prolonged therapy. *Virus Res.* **66**, 13–26.
52. Leitner, T., Kumar, S. & Albert, J. (1997). Tempo and mode of nucleotide substitutions in *gag* and *env* gene fragments in human immunodeficiency virus type 1 populations with a known transmission history. *J. Virol.* **71**, 4761–4770.
53. Li, W.-H., Tanimura, M. & Sharp, P. M. (1988). Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol. Biol. Evol.* **5**, 313–330.
54. Ou, C. Y., Ciesielski, C. A., Myers, G., Bandea, C. I., Luo, C.-C., Korber, B. T. M. *et al.* (1992). Molecular epidemiology of HIV transmission in a dental practice. *Science*, **256**, 1165–1171.
55. Myers, G., MacInnes, K. & Korber, B. (1992). The emergence of simian/human immunodeficiency viruses. *AIDS Res. Hum. Retroviruses*, **8**, 373–386.
56. Domingo, E., Webster, R. G. & Holland, J. J. (1999). Editors of *Origin and Evolution of Viruses*, Academic Press, San Diego, CA.
57. De Antoni, A., Foli, A., Lisziewicz, J. & Lori, F. (1997). Mutations in the *pol* gene of human immunodeficiency virus type 1 in infected patients receiving didanosine and hydroxyurea combination therapy. *J. Infect. Dis.* **176**, 899–903.
58. Winters, M. A., Coolley, K. L., Girard, Y. A., Levee, D. L., Hamdan, H., Shafer, R. W. *et al.* (1998). A 6-basepair insert in the reverse transcriptase gene of human immunodeficiency virus type 1 confers resistance to multiple nucleoside inhibitors. *J. Clin. Invest.* **102**, 1769–1775.
59. Mas, A., Parera, M., Briones, C., Soriano, V., Martínez, M. A., Domingo, E. & Menéndez-Arias, L. (2000). Role of a dipeptide insertion between codons 69 and 70 of HIV-1 reverse transcriptase in the mechanism of AZT resistance. *EMBO J.* **19**, 5752–5761.
60. Lukashof, V. V., Huismans, R., Jebbink, M. F., Danner, S. A., de Boer, R. J. & Goudsmit, J. (2001). Selection by AZT and rapid replacement in the absence of drugs of HIV type 1 resistant to multiple nucleoside analogs. *AIDS Res. Hum. Retroviruses*, **17**, 807–818.
61. Domingo, E., Ruíz-Jarabo, C. M., Arias, A., Molina-París, C., Briones, C., Baranowski, E. & Escarmís, C. (2003). Detection and biological implications of genetic memory in viral quasispecies. In *Proceedings of the International Congress on Cardiomyopathies and heart failure* (Matsumory, A., ed.), Kluwer Academic Publishers, London, UK, in the press.
62. Gerry, N. P., Witowski, N. E., Day, J., Hammer, R. P., Barany, G. & Barany, F. (1999). Universal DNA microarray method for multiplex detection of low abundance point mutations. *J. Mol. Biol.* **292**, 251–262.
63. Fei, Z. & Smith, L. M. (2000). Analysis of single nucleotide polymorphisms by primer extension and matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **14**, 950–959.
64. Amexis, G., Oeth, P., Abel, K., Ivshina, A., Pelloquin, F., Cantor, C. R. *et al.* (2001). Quantitative mutant analysis of viral quasispecies by chip-based matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Proc. Natl Acad. Sci. USA*, **98**, 12097–12102.
65. Kern, D., Collins, M., Fultz, T., Detmer, J., Hamren, S., Peterkin, J. J. *et al.* (1996). An enhanced-sensitivity branched-DNA assay for quantification of human immunodeficiency virus type 1 RNA in plasma. *J. Clin. Microbiol.* **34**, 3196–3202.
66. Boom, R., Sol, C. J., Salimans, M. M., Jansen, C. L., Wertheimvan Dillen, P. M. & van der Noordaa, J. (1990). Rapid and simple methods for purification of nucleic acids. *J. Clin. Microbiol.* **28**, 495–503.
67. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
68. Felsenstein, J. (1991). PHYLIP (Phylogeny Inference Package) version 3.5. University of Washington, Seattle.
69. Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
70. Swofford, D. L. & Olsen, G. J. (1990). Phylogeny reconstruction. In *Molecular Systematics* (Hillis, D. M. & Moritz, C., eds), Sinauer Associates, Sunderland, UK.

71. Sober, E. (1988). *Reconstructing the Past: Parsimony, Evolution and Inference*, MIT Press, Cambridge, MA.
72. Stewart, C. B. (1993). The powers and pitfalls of parsimony. *Nature*, **361**, 603–607.
73. Louwagie, J., McCutchan, F. E., Peeters, M., Brennan, T. P., Sanders-Buell, E., Eddy, G. A. *et al*. (1993). Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes. *AIDS*, **7**, 769–780.
74. Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 738–791.
75. Strimmer, K. & von Haeseler, A. (1997). Likelihood mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl Acad. Sci. USA*, **94**, 6815–6819.
76. Tamura, K. & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **10**, 512–526.
77. Salemi, M., Lewis, M., Egan, J. F., Hall, W. W., Desmyter, J. & Vandamme, A. M. (1999). Different population dynamics of human T cell lymphotropic virus type II in intravenous drug users compared with endemically infected tribes. *Proc. Natl Acad. Sci. USA*, **96**, 13253–13258.
78. Harbison, S. P. & Steele, G. L. (1995). *C, A Reference Manual*, Prentice-Hall, Englewood Cliffs, NJ.
79. Press, W. H. (1988). *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK.

Edited by J. Karn

(Received 25 February 2003; received in revised form 22 April 2003; accepted 20 May 2003)