

Genetic differences between human immunodeficiency virus type 1 subpopulations in faeces and serum

Lia van der Hoek,¹ Cees J. A. Sol,² Jolanda Maas,¹ Vladimir V. Lukashov,¹ Carla L. Kuiken³ and Jaap Goudsmit¹

^{1,2}Departments of Human Retrovirology¹ and Medical Microbiology², Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

³Los Alamos National Laboratory, Los Alamos, N. Mex., USA

To study human immunodeficiency virus type 1 (HIV-1) compartmentalization between intestine and blood, paired faecal and serum samples were collected from 204 HIV-1-infected persons. Direct sequencing of the gp120 V3 region obtained from 33 persons showed that faecal and serum sequences could be nearly homologous (0·3% different) or very dissimilar (11·3% different). Individual clones were obtained and sequenced from the faecal and serum samples of 13 persons. In 6 persons the HIV-1 subpopulations in faeces and serum were similar, whereas in 7 persons, distribution of V3 genotypes showed a marked difference. Genetic characterization of the HIV-1 subpopulations showed less heterogeneity in faecal subpopulations than in serum subpopulations in 5 of the 7 subjects. Furthermore, faecal and serum

subpopulations differed predominantly by non-synonymous nucleotide substitutions (in 6 of 7 persons). Comparison of the HIV-1 subpopulations in faeces and serum of these 7 persons, using resampling techniques, revealed a significant difference between faecal and serum subpopulations at an N-linked glycosylation site, C-terminal of the V3 loop (amino acids 331–333). Sequences from faecal subpopulations of all 7 persons contained a glycosylation site at amino acid position 331–333. Four of these 7 harboured serum variants lacking a glycosylation site at this position. The faecal subpopulations in these 4 persons showed limited nonsynonymous substitutions compared to synonymous substitutions, indicating that purifying selection is operational on these subpopulations.

Introduction

Receptive anal sexual intercourse is the main route of human immunodeficiency virus type 1 (HIV-1) infection in homosexual and bisexual males (van Griensven *et al.*, 1987; Moss *et al.*, 1987). For a vaccine to be effective it should protect against virus variants with the ability to enter and replicate in the intestinal mucosa. The intestinal mucosa comprises the largest pool of lymphoid cells, yet to date there is no experimental evidence for the existence of HIV-1 variants with a preference for this tissue.

The HIV-1 population in an infected person often contains a diverse mix of variants. In particular, the external envelope

protein gp120 of HIV-1 exhibits extensive heterogeneity (Saag *et al.*, 1988). The envelope contains critical determinants of cell tropism (O'Brien *et al.*, 1990; Shioda *et al.*, 1991), and variation at the hypervariable regions of gp120 may facilitate evasion of the immune system (Albert *et al.*, 1990; Zwart *et al.*, 1994). In particular the third variable region, V3, has been shown to be involved in cell tropism (Hwang *et al.*, 1991), syncytium formation (de Jong *et al.*, 1992; Fouchier *et al.*, 1992), and immune recognition (Javaherian *et al.*, 1990).

HIV-1 has been detected in a wide variety of lymphoid and nonlymphoid tissues (reviewed by Levy, 1993). Tissue-specific variants can be detected (Epstein *et al.*, 1991; Barnett *et al.*, 1991; Kuiken *et al.*, 1995; Korber *et al.*, 1994; Zhu *et al.*, 1996; Overbaugh *et al.*, 1996), indicating compartmentalization of HIV-1 in different tissues. Most of this work has focused on the central nervous system, but Barnett *et al.* (1991) showed that isolates recovered from the bowel can differ in biological and serological properties from HIV-1 strains recovered from

Author for correspondence: Lia van der Hoek.

Fax +31 20 6916531. e-mail L.VANDERHOEK@AMC.UVA.NL

The sequences presented in this study have been deposited in GenBank, accession numbers AF034392–AF034408.

serum. We have also shown that in some persons HIV-1 subpopulations in intestinal tissue can differ from those in serum (van der Hoek *et al.*, 1996). However, a study by Donaldson *et al.* (1994) showed restricted sequence variability in the HIV-1 V3 loop among different tissues, including the colon.

All previous studies were performed on post mortem samples or intestinal biopsies from a very few individuals. To address whether compartmentalization occurs frequently in intestinal tissue and whether tissue specificity can be found in the V3 sequences obtained from intestinal tissue, a study must enlist a large number of HIV-infected persons. Since ethical considerations make intestinal biopsies difficult to collect, we chose to characterize the V3 sequences in the intestinal mucosa by studying HIV-1 in faeces. This approach is feasible because HIV-1 RNA is frequently present in faeces (van der Hoek *et al.*, 1995) and the HIV-1 sequences in faeces show homology to the HIV-1 sequences in intestinal tissue (van der Hoek *et al.*, 1996).

In this study, paired faecal and serum samples were collected from 204 HIV-infected persons. Of these, 99 persons (49%) were positive for HIV-1 RNA in faeces, whereas 195 persons (96%) were positive for HIV-1 RNA in serum. The V3 region was sequenced directly from 33 paired faecal and serum samples, and the difference between faeces and serum was determined. Individual clones obtained from 13 persons were sequenced, and recently described methods for analysing groups of sequences were employed (Kuiken *et al.*, 1995, 1996) to identify amino acid motifs possibly involved in tissue specificity.

Methods

■ **Subjects and specimens.** Faecal and serum samples were collected on the same day from a cohort of 204 HIV-1-infected persons. Of these subjects, 100 were diagnosed with an AIDS-defining illness; 59 showed AIDS-related symptoms that were not AIDS-defining; and 45 persons were asymptomatic. Of the 13 persons from whom clones were sequenced, 8 were diagnosed with an AIDS-defining illness (subjects 03, 15, 58, 64, 66, 73, 79 and 92); 2 showed AIDS-related symptoms that were not AIDS-defining (subjects 02 and 76); 3 were asymptomatic (subjects 12, 134 and 149). Persons 03 and 02 have been described earlier as persons A and B (van der Hoek *et al.*, 1995); subjects 12 and 15 have been described previously as persons F and N (van der Hoek *et al.*, 1995, 1996). Serum was stored at -70°C and faeces were stored in broth as reported previously (van der Hoek *et al.*, 1995). The search for HIV-1 nucleic acids in various specimens was performed with the informed consent of all subjects.

■ **Extraction of nucleic acids from serum and faeces.** Nucleic acids were isolated from serum specimens using protocol Y/SC of Boom *et al.* (1990). Nucleic acids were isolated from faeces–broth mixtures using protocol F/SC, an adaptation of protocol Y/SC for use with faeces (van der Hoek *et al.*, 1995). In each experiment, negative and positive extraction controls were included. As a positive extraction control, 10^5 HIV-1 RNA molecules isolated from a virus culture were used.

■ **RT-PCR and DNA-PCR.** Reverse transcription (RT) and amplification (via PCR) of the V3 coding region were performed using primers

L10 (RT), L9 and 5'V3NOT in the first PCR, and primers 3'V3NOT and 5'KSI in the nested PCR as described previously (van der Hoek *et al.*, 1995). Minus-RT reactions (DNA-PCR) were performed as the RT reaction but without the addition of reverse transcriptase (van der Hoek *et al.*, 1995). PCR products were visualized by ethidium bromide staining and Southern blot hybridization.

■ **Cloning and sequencing.** For cloning, the product of the nested PCR was purified from agarose gel slices (van der Hoek *et al.*, 1995). The purified DNA was eluted in $6\ \mu\text{l}$ H_2O . Purified PCR fragments ($2\ \mu\text{l}$) were cloned using the TA cloning kit (Invitrogen) according to the manufacturer's instructions. Screening for clones containing a plasmid with an insert was performed by PCR (van der Hoek *et al.*, 1996). Double-stranded sequencing was performed using dye-labelled T7 primer, M13 reverse primer or -21M13 primer according to the manufacturers' protocols (dye primers, Perkin Elmer/Applied Biosystems; thermo sequenase fluorescent-labelled primer cycle sequencing kit, Amersham). The products were analysed on an automatic sequencer (Applied Biosystems). In the direct sequences of PCR products that contained more than one V3 variant, nucleotide positions with variation could be observed. Those nucleotide positions showing more than one nucleotide in both plus and minus strand sequences were considered to be ambiguous. For each PCR product, 6 to 16 clones were sequenced.

■ **Reproducibility of amplification from faeces.** To reduce the possibility that differences between faecal and serum sequences were caused by sampling errors, we performed the extraction, amplification and direct sequencing of the PCR products at least twice. The only persons included in further analysis were those in whom multiple direct sequences were largely identical (more than 98% of the nucleotides being identical). Clones were sequenced from each of the two or three extractions.

■ **Sequence analysis.** Sequencing of individual clones allowed analysis of a 303–309 bp fragment (HIV-1_{MN} nucleotide positions 7060–7366; Myers *et al.*, 1996) which included the V3 encoding region of gp120. Nucleotide sequences were aligned manually. All positions with an alignment gap in at least one sequence were excluded from any pairwise sequence comparison. The nucleotide consensus sequences of the clones of a genotype were determined by selecting the nucleotide that most frequently occurred for each nucleotide position. Phylogenetic analysis was carried out using the PHYLIP package (Felsenstein, 1993) and the neighbour-joining method of the MEGA program (Kumar *et al.*, 1993). The nucleotide distance matrix input for the neighbour-joining program was generated by Kimura's two-parameter estimation (Kimura, 1980). Bootstrap resampling (100 replications) was employed to place approximate confidence limits on individual branches. Probability (P) values were determined by using a two-tailed Fisher's exact test. Separation indexes (S-indexes) and the Pearson correlation coefficients were calculated as described by Kuiken *et al.* (1995, 1996). To determine the cut-off value for the Pearson correlation coefficient, sequences were divided in two random groups 100 times, and for all positions the (random) correlation was calculated. For the S-index, a significant difference was assumed when the S-index score was higher than the highest among 10000 random groupings (added with twice the standard deviation). Synonymous and nonsynonymous nucleotide P -distances (d_s and d_n , respectively) were calculated using the MEGA program (Kumar *et al.*, 1993). The percentage of nucleotide difference between direct sequences was determined by counting the number of nucleotides which were different between the two sequences, and also counting the number of positions where ambiguity was seen in one sequence, and only one nucleotide in the other sequence.

Results

Characterization of HIV-1 subpopulations in faeces and serum by direct sequencing

Faeces and serum were collected from 204 persons on the same date. Nucleic acids were isolated from faeces and serum, and the V3-encoding sequence was amplified by nested RT-PCR. Of the 204 subjects, 99 (49%) were positive for HIV-1 RNA in faeces, whereas 195 persons (96%) were positive for HIV-1 RNA in serum. The presence or absence of HIV-1 RNA in faeces was not associated with a clinical category nor with the CD4 count ($P = 0.75$ and 0.65 , respectively; F. Snijders, L. van der Hoek, G.-J. Weverling, J. Goudsmit, A. Tijmens, T. van Gool, C. I. Gallimore, C. J. A. Sol & S. A. Danner, unpublished results).

Direct sequencing of PCR products was employed to gauge the extent by which the HIV-1 population in faeces differs from the HIV-1 population in serum. HIV-1 RNA from faeces was isolated, amplified and sequenced in duplicate. Based on the reproducibility of duplicate sequencing, 33 persons were selected for whom the direct V3 sequence obtained from faeces and serum were compared (see Methods for details of comparisons). The nucleotide differences between their faecal and serum sequences ranged from 0.3% (only 1 of 310 nucleotides) to 11.3% (35 positions of the 310 nucleotides).

V3 genotype distribution in faeces and serum

To characterize the HIV-1 subpopulations genetically, 13 persons were chosen for cloning and sequencing of the V3 region from both faeces and serum. Selection of these subjects was based on their representation of the spectrum of differences found in the direct sequences. The nucleotide difference was < 3% in subjects 64 and 134; 3–4% in subjects 66 and 149; 4–5% in subjects 02 and 03; 5–6% in subject 12; 6–7% in subjects 58 and 79; 7–9% in subjects 15 and 76; and > 9% in subjects 73 and 92. For each of these 13 persons, between 9 and 33 clones were sequenced from both faeces and serum. For faeces, clones were sequenced from two or three independently obtained PCR products. Sequences were subjected to phylogenetic analysis to determine whether different clusters, assigned genotypes, could be identified within each person. The distribution of genotypes in faeces and serum was then determined.

Seven of 13 persons (subjects 02, 03, 12, 15, 73, 79 and 92) showed a significant difference in distribution of genotypes between faecal and serum subpopulations (Fig. 1). The consensus sequences of the genotypes showed a nucleotide distance of over 0.035 from other genotypes that coexisted in the same person (Fig. 2). Based on statistical analysis, the faecal subpopulations clearly differed from the serum subpopulations, and genotypes could be labelled either faeces- or serum-specific (Fig. 2).

Phylogenetic analysis of the genotypes specific for faeces

or serum of the 7 persons did not show a clustering of the faeces-specific sequences (Fig. 3). It did, however, show that except in one case the faecal and serum genotypes of a particular person clustered together: in subject 79 the faecal and serum sequences did not cluster together. The most plausible explanation for this result would be PCR contamination but, since genotypes 79A and 79B were detected in both faeces and serum (although with different frequencies), contamination is unlikely. To exclude that possibility, 15 sequential serum samples were analysed, and both genotype sequences were again found in 14 samples. It is therefore concluded that this person is a specific case in which two distinct populations coexist for a prolonged period of time (L. van der Hoek, V. V. Lukashov, C. J. A. Sol, J. Maas & J. Goudsmit, unpublished results).

For 6 of the 13 persons, the HIV-1 subpopulation in faeces was not significantly different from the HIV-1 subpopulation in serum (subjects 58, 64, 66, 76, 134 and 149). In 4 persons, different genotypes could not be identified (data not shown). For 2 persons more than one genotype was present in faeces and serum, but their difference in distribution in faeces and serum was only borderline significant ($P = 0.050$ for subject 58 and $P = 0.076$ for subject 76). For these 2 persons the main serum genotype was also the main faecal genotype, and the distinction between faeces and serum lay in the detection of minority genotypes in either faeces or serum (data not shown).

Conservation of an N-linked glycosylation site in V3 variants detected in faeces

To detect an intestinal tissue-specific V3 motif, the entire group of faecal clone sequences was compared with the entire group of serum clone sequences of the 7 persons who displayed a significant difference between the subpopulations of faeces and serum. We used recently described methods for analysing groups of sequences in order to identify amino acid positions which distinguish groups: calculation of the S-index for separation, and determination of the Pearson correlation coefficients (Kuiken *et al.*, 1995, 1996). The S-index for separation can identify positions that contain different amino acids in the faecal sequences compared to sequences from serum, as well as positions that are conserved in one group and variable in the other. To evaluate the strength of the association between the amino acid positions and the two groups of sequences, a Pearson correlation coefficient was calculated from a faecal sequence matrix and a serum sequence matrix (Kuiken *et al.*, 1996). The only amino acid positions showing a significant S-index and a Pearson correlation above the cut-off were positions 331 and 333. For position 331 the Pearson correlation coefficient was 0.21 with a cut-off for this position of 0.12; the S-index was 0.27 with a cut-off value of 0.22. For position 333 the Pearson correlation coefficient was 0.24 with a cut-off of 0.13; the S-index was 0.29 with a cut-off of 0.22. These positions form an N-linked glycosylation site (N-X-S,

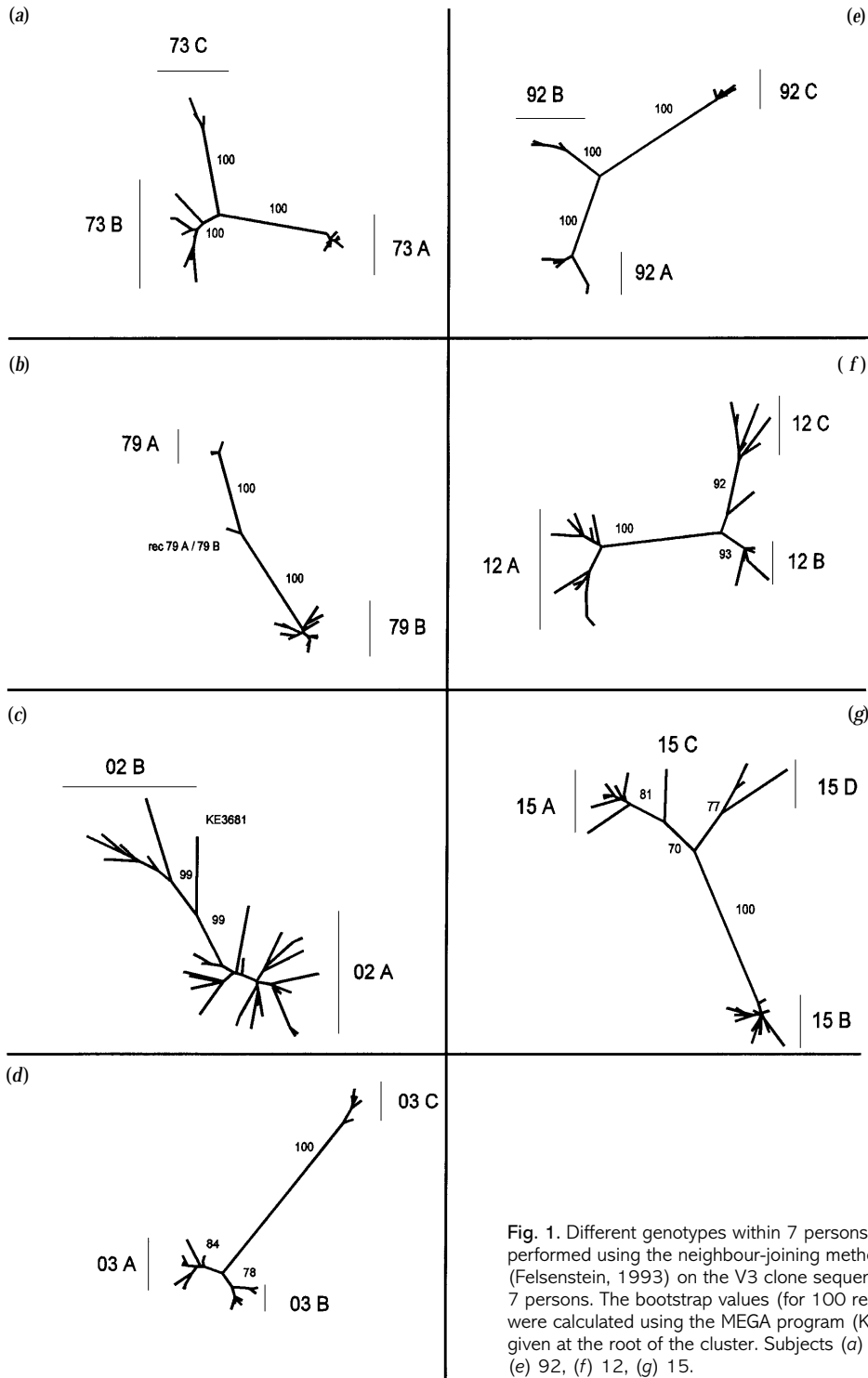


Fig. 1. Different genotypes within 7 persons. Phylogenetic analysis was performed using the neighbour-joining method in the PHYLIP package (Felsenstein, 1993) on the V3 clone sequences from faeces and serum of 7 persons. The bootstrap values (for 100 replications) of the genotypes were calculated using the MEGA program (Kumar *et al.*, 1993) and are given at the root of the cluster. Subjects (a) 73, (b) 79, (c) 02, (d) 03, (e) 92, (f) 12, (g) 15.

where X can be any amino acid except proline), located just C-terminal of the V3 loop.

As shown in Fig. 2, the glycosylation site at position 331 is conserved in all faecal sequences, whereas in serum of 4 persons sequences were detected which lacked the glycosylation site at position 331. For 3 persons these serum variants,

without the 331 glycosylation site, contained an N-linked glycosylation site at position 333–335 (N-X-T).

To evaluate whether the glycosylation site at position 331 is frequently seen in the group of 33 persons, all direct sequences from faeces and serum that did not have ambiguous nucleotides at this site were examined for the presence of a

CB	amino acid position: 270 VVIRSENFDTNAKTIIVQLNESVEINCRPNHNTRKSIHIGPGRAFYTTGIEIGDIROAHCNISRAKUNNTLKQIVKKLREQ...FGN.KTIVFNQSSGGDPEIVM	specificity	no of clones		p-value
			Faeces ^a	Serum	
C79A:	- I - - - D - - - SN - - - - - KA - K - - - S - - - - - A - - - T - - - - - D - - - - - I - - - - - K - - - - A - K - - - - -	faeces	16/17	4/9	< 0.01
recA/B	- - - - - N - - - - - R - PIN - T - I - S - - - - P - - - - K - - - A - D - - - - - D - - - - - I - - - - - K - - - - A - K - - - - -	^a	1/17	0/9	
C79B:	- - - - - N - - - - - K - PIN - T - I - S - - - - P - - - - K - - - A - D - - - - - ILNGT - - - D - - - - - I - - - - - K - - - - A - K - - - - -	serum	0/17	5/9	< 0.01
C92A:	- - - - - L - - - - - I - - - - - KDP - - - - - G - - - A - WGS - - A - D - - - - - Y - - - N - - - D - - R - V - E - - KGK - - - KN - - - I - - H - - - - -	faeces	13/13	0/16	< 0.0001
C92B:	- I - - - L - - - I - - - KDP - - - S - - - RG - - A - - - AA - D - - - Y - - - S - - - GKV - E - - K - K - - - KN - - - I - - H - - - - -	serum	0/13	5/16	< 0.05
C92C:	- I - - - L - - - I - - - KDP - V - - - S - - - - - A - - - A - D - - - - T - N - TT - D - G - V - - - - LQ - - - - I - - H - - - - -	serum	0/13	11/16	< 0.0001
C03A:	- - - - - K - SN - R - - - - K - P - K - I - V - - - - PL - - - - - A - - - D - - - N - - - - Q - A - Y - - Q - A - - - A - - - K - - R - G - - - - - T - - - - -	faeces	18/33	0/9	< 0.01
C03B:	- - - - - K - SN - - - - - K - P - K - I - V - - - - R - - - A - S - A - S - - - - Q - A - Y - - Q - A - - - A - - - K - - R - G - - - - - T - - - - -	faeces	15/33	1/9	> 0.1
C03C:	- - - - - A - - - N - - - - - K - - - D - I - - - - R - - - - QSI - A - - - N - - - - ILDKT - - - A - E - - - - - KIN - - - - - T - - - - -	serum	0/33	8/9	< 0.0001
C02A:	- - - - - N - - - - - KKA - - - - - A - - - - V - - - - Y - - - I - - - E - - T - - - - - P - - - - -	faeces	24/24	8/15	< 0.001
C02B:	- - - - - N - - - - - KA - V - - - - G - - - - Q - - - - Y - V - KD - D - E - - T - - - - - K - - - - H - - - - -	serum	0/24	6/15	< 0.001
ke3681	- - - - - N - - - - - KKA - - - - - G - - - - A - - - - Y - TLNGTE - E - - T - - - - - S - - - - H - - - - -	^b	0/24	1/15	
C12A:	- - - K - L - N - - - - V - - - - S - - - - PF - - K - - A - D - - - - TE - - - K - - T - - - - - KD - - - I - - H - - - - -	faeces	16/27 ^c	1/9 ^c	< 0.05
C12B:	- - - L - - - - - q - V - - - - LS - - - - PL - - - V - A - D - - - - L - TE - - - G - T - - - - - K - - - - KH - - - - -	serum	6/27 ^c	6/9 ^c	< 0.05
C15A:	- - - A - N - - - V - - - D - - - G - - - R - P - - K - L - - - - L - - - D - RH - AI - - K - - - - K - - - I - - P - - - - E - - - - -	faeces	10/23 ^c	1/12	0.05
C15B:	- - - A - N - - - V - - - G - - - R - SL - - VY - - - Q - - - - R - AG - K - - - - E - - TH - - - - - A - - - - -	serum	8/23 ^c	11/12	< 0.005
C73A:	I - - - - SN - - - S - - - - QPI - - I - - S - - - V - - K - L - - D - - - - GTD - - - - E - - - - V - - - I - - - - S - - - - -	faeces	16/20	0/11	< 0.0001
C73B:	I - - - - SN - - - Q - - - - G - - - - V - - - Y - - - GAD - K - - N - VAI - S - - - K - - - P - - - - S - - - - -	serum	4/20	7/11	< 0.05
C73C:	I - - - - SN - - - Q - - - - G - - - - RDR - - L - - - - G - - - D - - - V - I - SK - - - K - - - P - - - - L - - - - -	serum	0/20	4/11	0.01

Fig. 2. Deduced amino acid sequences of the V3 genotypes. Consensus sequences of the genotypes were aligned against the consensus sequence of HIV-1 subtype B, and the distribution of the genotypes in faeces and serum is given at the end of the sequence. The last column shows probabilities (*P*) for the association between the detection of a genotype in either faeces or serum. Values of *P* were determined by using the two-tailed Fisher's exact test. The letter C before a sequence indicates that it is a consensus sequence calculated from a group of sequences that clustered together phylogenetically (shown in Fig. 1). ^a, For faeces, clones were sequenced from two or three PCR products (described in Methods). ^a Sequence of clone recA/B, a clone that is a recombinant of genotype A and B sequences, most probably an artefact produced by RT-PCR. ^b Sequence of clone Ke3681, an individual sequence that clustered near the O2B sequences (Fig. 1). ^c For subjects 12 and 15, other sequences were also detected in faeces and serum that did not belong to genotype A or B (van der Hoek *et al.*, 1996) but, being a minority, they were not assigned as faeces- or serum-specific. Dashes indicate identity with the reference sequence; dots indicate deletions. Amino acid numbering in gp120 is indicated at the top of the figure.

glycosylation site at position 331 or 333. For serum, of the 22 direct sequences that could be analysed, 20 contained a V3 sequence with an *N*-linked glycosylation site at position 331 (data not shown). The other 2 serum samples harboured V3 sequences with a glycosylation site at position 333. Of the 22 faecal samples, again 20 V3 sequences had a glycosylation site at position 331, whereas a glycosylation site at position 333 was seen in two. Those faecal samples were from the same 2 persons whose serum sequences showed the 333 glycosylation site.

Synonymous and nonsynonymous substitutions

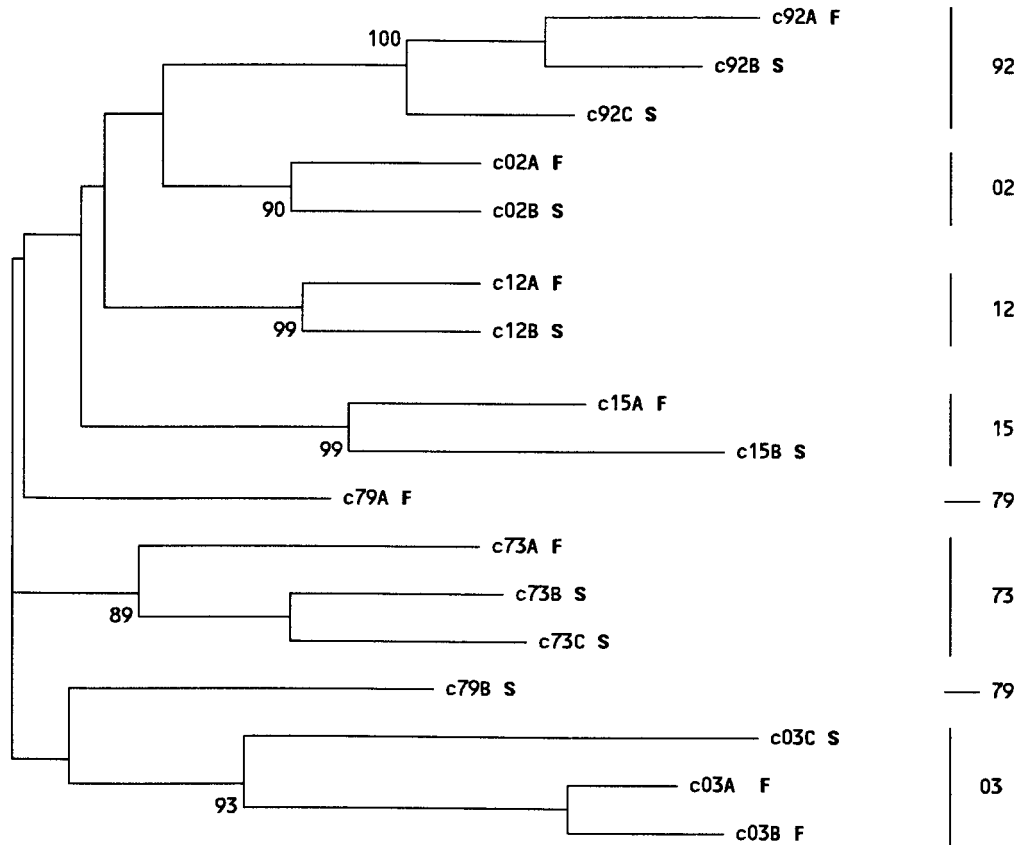
The nature of the substitutions that occur in a population supplies insight into the evolutionary forces operating on that population. We therefore determined the mean proportion of synonymous (d_s) and nonsynonymous (d_a) substitutions both within and between the faecal and serum subpopulations for the group of 7 persons showing a clear difference between faecal and serum subpopulations. The intrasample d_s and d_a values in faeces and serum facilitate comparison of the heterogeneity of HIV-1 within these samples. In the majority (5 of 7 persons), the HIV-1 population in faeces showed less heterogeneity than its corresponding serum sample (Table 1), in both synonymous and nonsynonymous substitutions.

The ratio of synonymous and nonsynonymous substitutions is commonly used to analyse selective forces which

drive the evolution of nucleotide sequences. A bias for either synonymous or nonsynonymous substitutions points to purifying selection or positive selection, respectively. The d_s/d_a ratio was determined both within and between the subpopulations of faeces and serum. The d_s/d_a ratio in serum showed a median of 0.96 (ranging from 0.72–1.45). For faeces, the median d_s/d_a ratio was 2.06 (range 0.42–5.49), indicating less selection for nonsynonymous changes in faecal subpopulations. In 4 persons, the ratio between d_s and d_a was > 1 (subjects 79, 92, 03 and 02; Table 1). In contrast this was seen in serum subpopulations of only 1 of these 4 persons (subject 03). These 4 persons were those in whom serum variants lacking a glycosylation site at position 331 were detected. Comparison of the d_s/d_a ratio between the faecal and serum subpopulations showed that in 6 of 7 individuals, the ratio was < 1 (Table 1). This result suggests that selection for certain nonsynonymous changes occurred between these subpopulations.

Discussion

Transmission of HIV during receptive anal sexual intercourse is restricted to viruses with the ability to enter the intestinal mucosa. The variants able to replicate in the intestinal mucosa are therefore of particular interest. In this study, we examined the HIV-1 subpopulations in faeces and serum to



Scale: each – is approximately equal to the distance of 0.001744

Fig. 3. Phylogenetic analysis of the faeces-specific (F) and serum-specific (S) genotypes. Phylogenetic analysis was performed using the neighbour-joining method (Kumar *et al.*, 1993). Codes refer to the genotype sequences shown in Fig. 2. The bootstrap values of the genotypes are given at the root of the cluster.

determine whether compartmentalization between intestine and blood occurs. Our findings indicate that (1) the HIV-1 subpopulations in faeces and serum can differ significantly within one individual; (2) in general, faecal and serum V3 sequences differ most profoundly at nonsynonymous nucleotide positions; (3) the V3 sequences specific for either faeces or serum show host-specific clustering rather than tissue-specific clustering; (4) in most persons, the heterogeneity of the HIV-1 subpopulation in faeces is lower than the corresponding serum sample; (5) some faecal subpopulations show a bias for synonymous substitutions in comparison with nonsynonymous substitutions; and (6) an *N*-linked glycosylation site at amino acid position 331 is conserved in faecal subpopulations.

The difference between faecal and serum subpopulations may be caused by selective forces operative in different compartments. Since the type of substitutions supplies information about the selective forces, we determined the proportion of synonymous (d_s) and nonsynonymous (d_a) substitutions among subpopulations in faeces and serum, as well as their ratio (d_s/d_a). A bias for synonymous substitutions

suggests selection based on replication fitness, whereas a bias for nonsynonymous substitutions suggests selection to promote diversification of the population. In our study, the d_s/d_a ratio in serum showed a median of 0.96, which is typically observed for the V3 region in serum and is considered to be a reflection of a strong immune pressure on this region (Lukashov *et al.*, 1995). Four persons showed a bias for synonymous substitutions in faeces, resulting in a d_s/d_a ratio > 1 . This finding implies that in these subpopulations, nonsynonymous substitutions are mostly deleterious, and selective forces driving the outgrowth of variants with amino acid changes are absent, whereas a selective force on replication fitness is operational (purifying selection). Of most interest is that these 4 persons are those carrying serum variants lacking the glycosylation site at position 331. So, for those persons harbouring mixtures of variants with and without the glycosylation site at position 331, purifying selection is operational in faeces on those variants with a glycosylation site at position 331.

Purifying selection in intestinal tissue may be the result of a limited choice in types of target cells. Several studies have shown that HIV is most frequently detected in mononuclear

Table 1. Mean proportion of synonymous (d_s) and nonsynonymous (d_a) substitutions

Person	Sample	Mean p-distance value*		
		d_s	d_a	d_s/d_a
79	F†	0.0294 ± 0.025	0.0133 ± 0.020	2.21
	S†	0.0459 ± 0.028	0.0637 ± 0.050	0.72
	F/S‡	0.0512 ± 0.015	0.0642 ± 0.047	0.80
92	F	0.0266 ± 0.023	0.0049 ± 0.004	5.49
	S	0.0277 ± 0.023	0.0314 ± 0.031	0.88
	F/S	0.0615 ± 0.012	0.0714 ± 0.015	0.86
03	F	0.0426 ± 0.024	0.0238 ± 0.018	1.79
	S	0.0526 ± 0.038	0.0378 ± 0.028	1.39
	F/S	0.1099 ± 0.025	0.1254 ± 0.033	0.88
02	F	0.0464 ± 0.023	0.0255 ± 0.010	2.06
	S	0.0537 ± 0.030	0.0561 ± 0.024	0.96
	F/S	0.0571 ± 0.029	0.0483 ± 0.024	1.18
12	F	0.0354 ± 0.027	0.0469 ± 0.031	0.75
	S	0.0342 ± 0.025	0.0236 ± 0.023	1.45
	F/S	0.0381 ± 0.024	0.0482 ± 0.026	0.79
15	F	0.0375 ± 0.022	0.0677 ± 0.040	0.55
	S	0.0151 ± 0.010	0.0215 ± 0.036	0.70
	F/S	0.0326 ± 0.024	0.0685 ± 0.024	0.48
73	F	0.0196 ± 0.017	0.0462 ± 0.055	0.42
	S	0.0612 ± 0.029	0.0494 ± 0.034	1.24
	F/S	0.0643 ± 0.020	0.1062 ± 0.034	0.61

* The proportions of synonymous (d_s) and nonsynonymous (d_a) substitutions and the standard deviation were calculated using the p-distance as implemented in the MEGA program.

† Every sequence within a sample was compared to every other sequence within that sample.

‡ Every sequence of a faeces sample was compared to every sequence of the serum sample.

cells in the lamina propria of the intestinal mucosa, either T cells or macrophages (Ullrich *et al.*, 1989; Heise *et al.*, 1991; Jarry *et al.*, 1990). HIV has also been detected in the epithelial layer (Nelson *et al.*, 1988; Kotler *et al.*, 1991), but whether it infects epithelial cells on a large scale remains questionable. In addition, HIV-infected dendritic reticulum cells in germinal centres of mucosal lymphoid follicles have been described (Jarry *et al.*, 1990). In rhesus macaques infected with simian immunodeficiency virus (SIV), SIV-infected T lymphocytes and macrophages could be detected in the gut-associated lymphoid tissue (Heise *et al.*, 1994). SIV-infected macrophages were seen mainly during the primary and terminal stages of SIV infection. Most of our group of 7 persons whose faecal and serum subpopulations differed were diagnosed with AIDS ($n = 5$), and the cells infected in their intestinal mucosa may have been predominantly macrophages. Two of the group harboured V3 variants in serum that carried SI characteristics (genotypes 15B and 73C), whereas all genotypes specific for faeces contained NSI or macrophage-tropic features (de Jong *et al.*, 1992). However, as SI features were seen in only 2 persons,

whereas in the others the variants from faeces and serum all carry NSI characteristics, no conclusions can be drawn.

A limitation of immune pressure on the virus could also result in purifying selection. Both *in vivo* and *in vitro* studies have shown that HIV-1 replication under low immune pressure results in a low rate of nonsynonymous nucleotide substitutions (Nara *et al.*, 1990; Lukashov *et al.*, 1995). Local intestinal immune pressure is impaired in HIV-infected persons. The lamina propria shows decreased levels of CD4-positive cells and increased levels of CD8-positive cells (reviewed by Riecken *et al.*, 1990). A decrease in CD4-positive cells can be even more pronounced in the lamina propria of intestinal mucosa than the decrease seen in blood (Schneider *et al.*, 1995; Snijders *et al.*, 1996). In addition, the mucosal humoral immune system may be compromised in HIV-infected patients (Janoff *et al.*, 1994). Mucosal secretions show impaired levels of IgA, especially IgA₂, and IgA plasma cell density is decreased (Kotler *et al.*, 1987). Interestingly, comparison of IgA and IgG antibodies directed to V3 found IgA antibodies much less capable of neutralizing HIV-1 than the IgG antibodies in the same person (Kozlowski *et al.*, 1994). The local intestinal immune pressure can be strongly impaired in HIV-infected persons, and immune pressure on HIV in the intestinal mucosa could be sufficiently impaired that diversifying mutations to escape from immune pressure are not seen.

Serum- or faeces-specific sequences are usually not exclusively present in serum or faeces but are unevenly distributed, predominating in one of the two. This finding suggests migration of variants from blood to intestinal tissue, or vice versa, rather than independent evolution of the variants in the intestinal tissue. Similar conclusions have been drawn regarding HIV subpopulations studied in blood and the central nervous system (Korber *et al.*, 1994). It may be that distinctions between faecal and serum subpopulations are largely a reflection of time-related factors.

For 6 persons in the group of 13, we found no evidence that the subpopulations in faeces and serum differed. Donaldson *et al.* (1994) also showed that for 2 persons harbouring an HIV-1 population with restricted sequence variability in the V3 loop, the subpopulation obtained from colon was not significantly different from the subpopulations in other tissues or peripheral blood nucleated cells.

In conclusion, the subpopulations in faeces and serum are distinct in some persons, whereas in others they are similar. In persons whose subpopulations are distinct, conservation of a glycosylation site at position 331 is seen in faecal sequences. Whether conservation of the 331 glycosylation site is based on selection due to tropism or to the local immune pressure remains an interesting subject for further studies.

We especially thank F. Snijders for collecting faecal and serum specimens and clinical data; A. van Strien for preparing buffers, chemicals and extraction tubes for nucleic acid extraction; M. T. E. Cornelissen for enlightening discussions and Lucy D. Phillips for editorial review.

References

- Albert, J., Abrahamsson, B., Nagy, K., Aurelius, E., Gaines, H., Nyström, G. & Fenyö, E.-M. (1990). Rapid development of isolate-specific neutralizing antibodies after primary HIV-1 infection and consequent emergence of virus variants which resist neutralization by autologous sera. *AIDS* **4**, 107–112.
- Barnett, S. W., Barboza, A., Wilcox, C. M., Forsmark, C. E. & Levy, J. A. (1991). Characterization of human immunodeficiency virus type 1 strains recovered from the bowel of infected individuals. *Virology* **182**, 802–809.
- Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. L., Wertheim-van Dillen, P. M. E. & van der Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology* **28**, 495–503.
- de Jong, J. J., Goudsmit, J., Keulen, W., Klaver, B., Krone, W., Tersmette, M. & de Ronde, A. (1992). Human immunodeficiency virus type 1 clones chimeric for the envelope V3 domain differ in syncytium formation and replication capacity. *Journal of Virology* **66**, 757–765.
- Donaldson, Y. K., Bell, J. E., Holmes, E. C., Hughes, E. S., Brown, H. K. & Simmonds, P. (1994). *In vivo* distribution and cytopathology of variants of human immunodeficiency virus type 1 showing restricted sequence variability in the V3 loop. *Journal of Virology* **68**, 5991–6005.
- Epstein, L. G., Kuiken, C., Blumberg, B. M., Hartman, S., Sharer, L. R., Clement, M. & Goudsmit, J. (1991). HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies. *Virology* **180**, 583–590.
- Felsenstein, J. (1993). PHYLIP Manual version 3.5.
- Fouchier, R. A. M., Groenink, M., Kootstra, N. A., Tersmette, M., Huisman, H. G., Miedema, F. & Schuitemaker, H. (1992). Phenotype associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *Journal of Virology* **66**, 3183–3187.
- Heise, C., Dandekar, S., Kumar, P., Duplantier, R., Donovan, R. M. & Halsted, C. H. (1991). Human immunodeficiency virus infection of enterocytes and mononuclear cells in human jejunal mucosa. *Gastroenterology* **100**, 1521–1527.
- Heise, C., Miller, C. J., Lackner, A. & Dandekar, S. (1994). Primary acute simian immunodeficiency virus infection of intestinal lymphoid tissue is associated with gastrointestinal dysfunction. *Journal of Infectious Diseases* **169**, 1116–1120.
- Hwang, S. S., Boyle, T. J., Lyerly, H. K. & Cullen, B. R. (1991). Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. *Science* **253**, 71–74.
- Janoff, E. N., Jackson, S., Wahl, S. M., Thomas, K., Peterman, J. H. & Smith, P. D. (1994). Intestinal mucosal immunoglobulins during human immunodeficiency virus type 1 infection. *Journal of Infectious Diseases* **170**, 299–307.
- Jarry, A., Cortez, A., René, E., Muzeau, F. & Brousse, N. (1990). Infected cells and immune cells in the gastrointestinal tract of AIDS patients. An immunohistochemical study of 127 cases. *Histopathology* **16**, 133–140.
- Javaherian, K., Langlois, A. J., Larosa, G. J., Profy, A. T., Bolognesi, D. P., Herlihy, W. C., Putney, S. D. & Matthews, T. J. (1990). Broadly neutralizing antibodies elicited by the hypervariable neutralizing determinant of HIV-1. *Science* **250**, 1590–1593.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111–120.
- Korber, B. T. M., Kunstman, K. J., Patterson, B. K., Furtado, M., McEvilly, M. M., Levy, R. & Wolinsky, S. M. (1994). Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. *Journal of Virology* **68**, 7467–7481.
- Kotler, D. P., Scholes, J. V. & Tierney, A. R. (1987). Intestinal plasma cell alterations in the acquired immunodeficiency syndrome. *Digestive Diseases and Sciences* **32**, 129–138.
- Kotler, D. P., Reka, S., Borcich, A. & Cronin, W. J. (1991). Detection, localization and quantitation of HIV-associated antigens in intestinal biopsies from patients with HIV. *American Journal of Pathology* **139**, 823–830.
- Kozlowski, P. A., Chen, D., Eldridge, J. H. & Jackson, S. (1994). Contrasting IgA and IgG neutralization capacities and responses to HIV type 1 gp120 V3 loop in HIV-infected individuals. *AIDS Research and Human Retroviruses* **10**, 813–822.
- Kuiken, C. L., Goudsmit, J., Weiller, G. F., Armstrong, J. S., Hartman, S., Portegies, P., Dekker, J. & Cornelissen, M. (1995). Differences in human immunodeficiency virus type 1 V3 sequences from patients with and without AIDS dementia complex. *Journal of General Virology* **76**, 175–180.
- Kuiken, C. L., Cornelissen, M. T. E., Zorgrager, F., Hartman, S., Gibbs, A. J. & Goudsmit, J. (1996). Consistent risk group associated differences in human immunodeficiency virus type 1 *vpr*, *vpu* and V3 sequences despite independent evolution. *Journal of General Virology* **77**, 783–792.
- Kumar, S., Tamura, K. & Wei, M. (1993). Molecular evolutionary genetics analysis (MEGA), version 1.0. Institute of Molecular Evolutionary Genetics, Pennsylvania State University.
- Levy, J. A. (1993). Pathogenesis of human immunodeficiency virus infection. *Microbiological Reviews* **57**, 183–289.
- Lukashov, V. V., Kuiken, C. L. & Goudsmit, J. (1995). Intrahost human immunodeficiency virus type 1 evolution is related to length of the immunocompetent period. *Journal of Virology* **69**, 6911–6916.
- Moss, A. R., Osmond, D., Bacchetti, P., Chermann, J.-C., Barre-Sinoussi, F. & Carlson, J. (1987). Risk factors for AIDS and HIV seropositivity in homosexual men. *American Journal of Epidemiology* **125**, 1035–1047.
- Myers, G., Korber, B., Foley, B., Jeang, K.-T., Mellors, J. W. & Wainhobson, S. (1996). *Human Retroviruses and AIDS*. Los Alamos, N. Mex.: Los Alamos National Laboratory.
- Nara, P. L., Smit, L., Dunlop, N., Hatch, W., Merges, M., Waters, D., Kelliher, J., Gallo, R. C., Fischinger, P. J. & Goudsmit, J. (1990). Emergence of viruses resistant to neutralization by V3-specific antibodies in experimental human immunodeficiency virus type 1 IIIB infection of chimpanzees. *Journal of Virology* **64**, 3779–3791.
- Nelson, J. A., Reynolds-Kohler, C., Margaretten, W., Wiley, C. A., Reese, C. E. & Levy, J. A. (1988). Human immunodeficiency virus detected in bowel epithelium from patients with gastrointestinal symptoms. *Lancet* **i**, 259–262.
- O'Brien, W. A., Koyanagi, Y., Namazie, A., Zhoa, J., Diagne, A., Idler, K., Zack, J. A. & Chen, I. S. Y. (1990). HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. *Nature* **348**, 69–73.
- Overbaugh, J., Anderson, R. J., Ndinya-Acola, J. O. & Kreiss, J. K. (1996). Distinct but related human immunodeficiency virus type 1 variant populations in genital secretions and blood. *AIDS Research and Human Retroviruses* **12**, 107–115.
- Riecken, E. O., Zeitz, M. & Ullrich, R. (1990). Non-opportunistic causes of diarrhoea in HIV infection. In *Gastroenterological Aspects of AIDS*,

Baillière's Clinical Gastroenterology, pp. 385–403. Edited by B. G. Gazzard. London: Baillière Tindall.

Saag, M. S., Hahn, B. H., Gibbons, J., Li, Y., Parks, E. S., Parks, W. P. & Shaw, G. M. (1988). Extensive variation of human immunodeficiency virus type 1 *in vivo*. *Nature* **334**, 440–444.

Schneider, T., Jahn, H.-U., Schmidt, W., Riecken, E.-O., Zeitz, M., Ullrich, R. & Berlin Diarrhoea/Wasting Syndrome Study Group (1995). Loss of CD4 T lymphocytes in patients infected with human immunodeficiency virus type 1 is more pronounced in the duodenal mucosa than in the peripheral blood. *Gut* **37**, 524–529.

Shioda, T., Levy, J. A. & Cheng-Mayer, C. (1991). Macrophage and T cell-line tropisms of HIV-1 are determined by specific regions of the envelope gp120 gene. *Nature* **349**, 167–169.

Snijders, F., Meenan, J., van den Blink, B., van Deventer, S. J. H. & ten Kate, F. J. W. (1996). Duodenal intraepithelial and lamina propria T-lymphocytes in human immunodeficiency virus-infected patients with and without diarrhoea. *Scandinavian Journal of Gastroenterology* **31**, 1176–1181.

Ullrich, R., Zeitz, M., Heise, W., l'Age, M., Höffken, G. & Riecken, E. O. (1989). Small intestinal structure and function in patients infected with human immunodeficiency virus (HIV): evidence for HIV-induced enteropathy. *Annals of Internal Medicine* **111**, 15–21.

van der Hoek, L., Boom, R., Goudsmit, J., Snijders, F. & Sol, C. J. A. (1995). Isolation of human immunodeficiency virus type 1 (HIV-1) RNA

from feces by a simple method and difference between HIV-1 subpopulations in feces and serum. *Journal of Clinical Microbiology* **33**, 581–588.

van der Hoek, L., Sol, C. J. A., Snijders, F., Bartelsman, J. F. W., Boom, R. & Goudsmit, J. (1996). Human immunodeficiency virus type 1 RNA populations in faeces with higher homology to intestinal populations than to blood populations. *Journal of General Virology* **77**, 2415–2425.

van Griensven, G. J. P., Tielman, R. A. P., Goudsmit, J., van der Noordaa, J., de Wolf, F., de Vroome, E. M. M. & Coutinho, R. A. (1987). Risk factors and prevalence of HIV antibodies in homosexual men in The Netherlands. *American Journal of Epidemiology* **125**, 1048–1057.

Zhu, T., Wang, N., Carr, A., Nam, D. S., Moor-Jankowski, R., Cooper, D. A. & Ho, D. D. (1996). Genetic characterization of human immunodeficiency virus in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission. *Journal of Virology* **70**, 3098–3107.

Zwart, G., Back, N. K. T., Ramautarsing, C., Valk, M., van der Hoek, L. & Goudsmit, J. (1994). Frequent and early HIV-1 neutralizing capacity in sera from Dutch HIV-1 seroconverters is related to antibody reactivity to peptides from the gp120 V3 domain. *AIDS Research and Human Retroviruses* **10**, 245–251.

Received 14 July 1997; Accepted 1 October 1997