Human immunodeficiency virus type 1 gp120 envelope characteristics associated with disease progression differ in family members infected with genetically similar viruses

Elly Baan,1 Renée M. van der Sluis,1 Margreet E. Bakker,1 Vincent Bekker,2 Dasja Pajkrt,2 Suzanne Jurriaans,3 Taco W. Kuijpers,2 Ben Berkhout,1 Katja C. Wolthers,3 William A. Paxton1 and Georgios Pollakis1

The human immunodeficiency virus type 1 (HIV-1) envelope protein provides the primary contact between the virus and host, and is the main target of the adaptive humoral immune response. The length of gp120 variable loops and the number of N-linked glycosylation events are key determinants for virus infectivity and immune escape, while the V3 loop overall positive charge is known to affect co-receptor tropism. We selected two families in which both parents and two children had been infected with HIV-1 for nearly 10 years, but who demonstrated variable parameters of disease progression. We analysed the gp120 envelope sequence and compared individuals that progressed to those that did not in order to decipher evolutionary alterations that are associated with disease progression when individuals are infected with genetically related virus strains. The analysis of the V3-positive charge demonstrated an association between higher V3-positive charges with disease progression. The ratio between the amino acid length and the number of potential N-linked glycosylation sites was also shown to be associated with disease progression with the healthier family members having a lower ratio. In conclusion in individuals initially infected with genetically linked virus strains the V3-positive charges and N-linked glycosylation are associated with HIV-1 disease progression and follow varied evolutionary paths for individuals with varied disease progression.

INTRODUCTION

Large differences have been described amongst humans in their susceptibility to infection with human immunodeficiency virus type 1 (HIV-1) as well as rates of disease progression. After infection some individuals progress slowly in their disease course (slow progressors/long-term non-progressors) (Buchbinder et al., 1994; Cao et al., 1995; Easterbrook, 1994; Lévy, 1993) some progress rapidly and develop disease in as little as 2–5 years (fast progressors) (Anzala et al., 1995), whilst a small minority show no progression over 20 years of infection (elite controllers) (Deeks & Walker, 2007). Many viral as well as host factors have been associated with this variation in viral transmission and disease progression (Morgan et al., 2002). One important host factor associated with both HIV-1 transmission and disease progression is the 32 bp deletion in the CCR5 co-receptor (CCR5Δ32) for HIV-1 (Liu et al., 1996). Individuals homozygous for the deletion are highly protected against infection, whilst individuals heterozygous for the deletion are not protected from infection, but disease progression is attenuated once infected (Huang et al., 1996). Other host factors involved with the chemokine/chemokine receptor (axis) have been associated with altered rates of HIV-1 disease progression (Paxton & Kang, 1998; Singh et al., 2003). Numerous allelic polymorphisms in genes involved in the adaptive immune recognition by T-cells,
mainly through human leukocyte antigen (HLA) restriction have shown to influence disease course, presumably through the elimination of infected cells and control of viral replication (Altfeld et al., 2003; Carrington & O’Brien, 2003; Carrington et al., 1999; Hendel et al., 1999). More recently, a number of genetic factors linked to the innate immune response were shown to be associated with both risk of transmission and disease progression (Koizumi et al., 2007).

The viral factor frequently associated with disease progression is the viral set-point (following the period of acute infection), with higher viral loads associated with faster progression to disease (Mellors et al., 1996). Additionally, HIV-1 lacking the nef protein through a genetic deletion in the gene has been associated with decreased pathogenicity and longer survival (Birch et al., 2001). However, the majority of research linking viral variation with HIV-1 transmission and disease progression has focused mainly on the gp120 envelope gene. One of the key characteristics to HIV-1 infection is the preferential transmission of viruses utilizing the CCR5 co-receptor (R5) (Margolis & Shattock, 2006) and the infrequent transmission of CXCR4 using (X4) or dual-tropic viruses (R5/X4) (Michael et al., 1998; Sheppard et al., 2002). This restriction has been attributed to a number of factors including cell-type of infection and a better immune control of the X4 viruses. It has also been well documented that a switch from R5 to X4 viruses during disease is linked with increased HIV-1 RNA concentrations, reduced CD4+ cell counts and faster progression (Tersmette et al., 1989). What precise mechanisms control this switch in co-receptor activity is still poorly understood, but many studies have addressed this by comparing envelope sequences. The most obvious genetic alterations occur in the variable regions of the envelope. These alter both charge and potentially N-linked glycosylation patterns (Cho et al., 1998; Morikita et al., 1997; Pollakis et al., 2001; Shieh et al., 2000). These alterations in N-linked glycosylation influence the extent to which effective neutralizing antibodies can bind to HIV-1 and thus prevent infection (Sagar et al., 2006; Wei et al., 2003; Wu et al., 1995). In all likelihood the diversity within the gp120 variable regions are contributing to both co-receptor activity and neutralizing antibody escape providing for the evolution observed during disease progression.

We have identified two families within which the siblings demonstrate major differences in HIV-1 RNA loads, CD4+ cell counts and disease progression (as monitored through initiation of therapy) and who have in all likelihood been infected with viruses from a similar source. An increased insight in the genetic make-up of the transmitted HIV-1 strains can provide information on viral as well as host factors that can be associated with HIV-1 disease progression.

RESULTS

Subjects and their disease description

Both families were from central Africa but of different ethnicity. From family 1 the father’s samples were not available for analysis. The mother (M1) when first seen at the clinic possessed a high HIV-1 RNA load (log$_{10}$4.91 virus copies ml$^{-1}$) (Fig. 1a) and low CD4+ cell count (20 cells ml$^{-1}$) (Fig. 1c). She was diagnosed with a WHO B3 classification and anti-retroviral therapy (ART) was initiated in March 2000 (IDV/3TC/d4T) and switched regime in August of the same year (NVP/3TC/AZT). This therapy subsequently failed due to non-compliance and a new regime was initiated, resulting again in reduced viral loads after which viral blips were observed in 2003 and 2005 (Fig. 1). The mother experienced a slow but sustained increase in CD4+ cell counts during her treatment period. Her first child (C1A), was considered to be a slow progressor (log$_{10}$2.06 virus copies ml$^{-1}$) and 780 CD4+ cells ml$^{-3}$ at study entry. C1A was asymptomatic (WHO classification A) with no requirement for treatment. The subsequent viral load rose steadily over the following years, while CD4+ cell counts remained steady (Fig. 1a and c). The second child (C1B) was born 5 years after C1A and was diagnosed with a WHO classification B HIV-1 infection when first seen at the clinic. This child’s baseline viral load was moderate (log$_{10}$4.26 virus copies ml$^{-1}$) with high CD4+ cell counts (680 cells ml$^{-3}$) and the patient was started on combination ART with nelfinavir/d4T/3TC, after which the viral load became undetectable with occasional blips (Fig. 1a). C1B demonstrated a sustained drop in CD4+ cell counts over the years despite successful therapy (Fig. 1c) and also demonstrated signs of progressing.

In family 2, the father (F2) presented with a high viral load (log$_{10}$5.74 virus copies ml$^{-1}$) and low CD4+ cell counts (220 cells ml$^{-3}$) (Fig. 1b and d) and received successful ART. There were no records from his first wife (M2a) who presumably died of AIDS. The second wife (M2b) had low viral load (log$_{10}$2.91 virus copies ml$^{-1}$) with high CD4+ cell counts (580 cells ml$^{-3}$) and was considered a non-progressor. The first child (C2a) was born in 1991 to M2a (first wife) and was first seen in the clinic with a moderate viral load (log$_{10}$4.59 virus copies ml$^{-1}$) and high CD4+ cell counts (690 cells ml$^{-3}$) (WHO classification A2). No treatment was initiated and the viral load burden remained stable during the entire observation period (Fig. 1b). The second child (C2b) born to M2b 2 years after C2a had a similar viral load as C2a (log$_{10}$4.64 virus copies ml$^{-1}$), but with a much lower CD4+ cell counts (60 cells ml$^{-3}$) when first seen at the clinic. C2b received ART (Kaletra/AZT/3TC) that successfully reduced viral loads throughout follow up (Fig. 2b).

All the children were HLA typed for their class A, B and C alleles, but we did not observe any association with those alleles known to be involved with risk of transmission or alteration in disease progression (Carrington & O’Brien, 2003; Carrington et al., 1999) (data not shown). HLA data were not available for M2b who is clearly controlling viraemia (Fig. 1b).

Phylogenetic analysis and genetic diversity

We chose to compare gp120 region sequences of viruses isolated from all study individuals. Phylogenetic inference
utilizing the Kimura two-parameter/neighbour-joining model showed that the members of family 1 were infected with an HIV-1 subtype C virus equidistant within the topology of the overall subtype C HIV-1 phylogeny (Fig. 2). Family 2 harboured the CRF-01 recombinant form of HIV-1 that is more common in South-east Asia and infrequent in central Africa. We therefore determined that all infected members within each family carried a virus of common origin. This analysis also confirmed that M2a and M2b had to have been infected with a virus similar to F2, since C2a and C2b were also infected with closely related F2 strains, despite having different mothers.

Greater diversity was observed amongst the viral envelope sequences in the children compared with the parents when the study began (i.e. arrival at the clinic) (Fig. 3a). However, no differences in sequence diversity were observed between the study subjects at this time based on requirement for ART initiation or not (Fig. 3b). The virus diversity did not correlate with viral load values (Fig. 3c), indicating that when the study began there was no association between intra-patient virus diversity and disease status.

**V3 loop amino acid analysis**

We analysed the gp120 sequences to identify associations with clinical outcome in those patients with varied disease status, but who were infected with related virus strains. Initially, we studied the overall predicted positive charges of the V3 region for each family member. In family 1, M1 harboured mainly +4 V3 charges at baseline and prior to receiving therapy. Child C1A had HIV viruses with lower V3 charges, mainly +2, which gradually increased over time, whilst C1B had viruses of a +3 charge at initial analysis. In family 2, the V3 charges of the father’s HIV viruses ranged from +3 to +5 with the majority being +4. Mother M2b harboured viruses with lower V3 charges with the majority being +2 at the first time point analysed with attenuated charges of +1 a year later, coinciding with a lower viral load without therapy. When analysing the children, C2a had mainly viruses with +2 charges with some being +3. The other child C2b possessed viruses with higher charges ranging from +3 to +6 and this was the child receiving ART. For this child, we also observed that other V3 alterations were common, such as loss of a potential N-linked glycosylation site (PNGS) at the stem of the V3 loop (Fig. 4a), which had previously been associated...
with a switch in the R5 to X4 phenotype. Furthermore, virus strains from C2b showed a gain in positive amino acid charge at position 11 (E→S→K), and a loss of negative charges in positions 25 and 29 that are also associated with a switch in co-receptor usage. The existing algorithms for predicting co-receptor usage have not been applied here since they are CRF-01 strains. When the WebPSSM (http://indra.mullins.microbiol.washington.edu/pssm/) bioinformatic tool was applied to viral sequences obtained from family 1 (infected with subtype C) all viruses were predicted to utilize the CCR5 co-receptor.

In general, those individuals who did not receive ART tended to possess virus variants with the lowest positive V3 charges. With high V3 charge, the individual tended to possess higher HIV-1 viral loads or lower CD4+ cell counts. Furthermore, higher V3-positive charges were observed in family members that were in need of ART in comparison to those that did not (Fig. 4b, $P<0.0001$). The difference remained significant when C2b sequences, suspected of CXC4 co-receptor usage, were excluded from the analysis ($P<0.0001$). Interestingly C1A, who initially harbourd viruses with low viral loads, but which were shown to steadily increase over time, demonstrated a corresponding increase to the V3 charge. M2b who presented with very low viral loads and which dropped over time in the absence of therapy showed a decrease in the V3 loop-positive charge.

**Amino acid length and N-linked glycosylation analysis**

We wished to identify whether alterations in the amino acid lengths of the various g120 envelope regions or alterations to the number of PNGS would be associated with the disease status of the two family members. We analysed the V1V4 sequences isolated and observed a marked difference between the two families when the amino acid lengths were plotted against the number of PNGS. Overall there was a strong positive correlation ($P<0.0001$) between the amino acid length and number of PNGS (Fig. 5a). However, when the sequences were
re-analysed, taking into account who received therapy and who did not, we found differences between family members requiring ART. A strong positive correlation between amino acid length and number of PNGS was observed in those not requiring therapy ($P<0.0001/r^2=0.7681$), while the correlation was poor for those who were in need of ART ($P=0.0044/r^2=0.0952$) (Fig. 5a). We calculated and compared the ratio of amino acid length to number of PNGS (aa/PNGS) between family members requiring ART and those not and found those requiring ART possessed a significantly higher ratio (Fig. 5b). When the analysis was performed including only the HIV-1 virus sequences obtained from children at baseline, equally significant differences were found between C1B and C2b (fast disease progressors) and C1A and C2a (slow progressors) ($P<0.0001$) (Fig. 5c). We conclude from the above that differences in the ratio between the length and the number of PNGS associate with disease progression.

Fig. 3. Mean nucleotide variation of the gp120 envelope fraction of the virus strains: (a) comparison of the parents’ strains, open circles, and their children, closed circles, (b) comparison of family members requiring therapy with those that did not. The analysis was performed by using the neighbour-joining algorithm and the Kimura-2-parameter method of the MEGA software package (*$P<0.05$). (c) Correlation of the plasma viral load to the mean nucleotide variation: $d_K$ is the overall diversity by the Kimura-2-parameter method, $d_{ns}$ is the non-synonymous nucleotide variation and $d_s$ is the synonymous nucleotide variation.

Fig. 4. V3-positive charge analysis: (a) representative V3 loop virus sequences found in three time points analysed of the child C2a. The loss of the PNGS at position 9, changes at the GPGQ motif, the gain of a positive charge (K) and the loss of negative charges (D), shown in bold, are indicating the presence of X4 strains. The column at the right indicates the V3-positive charge. The numbers below indicate: 1, the PNGS; 2, the GPGQ crown-motif; 3 and 4 the negative charge positions. (b) V3-positive charge comparison between family members requiring therapy with those that did not. Family members in both families requiring therapy had significantly lower V3-positive charges as demonstrated by the Mann–Whitney test.
In addition to the HIV-1 gp120 sequence analysis of the virus strains isolated from the study participants we extracted and analysed gp120 sequences from the Los Alamos HIV-1 sequence database. We compared the sequences obtained during early infection to those obtained late. Our assumption being that early sequences were collected from individuals that had not yet progressed in their disease, whilst late sequences were from individuals who had. We calculated the V1V4 amino acid length to number of PNGS and compared the two groups (Fig. 5d). Early sequences showed a lower ratio when compared with sequences collected later in disease from individuals expected to have progressed in their disease course (P<0.0001). Furthermore, comparison of early sequences with those obtained from individuals whom had AIDS also revealed a difference, although not as pronounced (P=0.001) as seen with late patients, probably due to the loss of immune control hence selection pressure occurring in AIDS. We also studied long-term non-progressors (LTNPs) and found a difference in comparison to sequences from late patients (P=0.001), but no difference with viruses found early in disease. This analysis supports the findings obtained from the sequence comparisons from the two families studied here, indicating that the aa/PNGS ratio is associated with disease progression.

**DISCUSSION**

Here, we describe individuals infected with genetically related HIV-1 strains and who have progressed differently in their disease course. We studied members of two families with variant ethnicities, although both from sub-Saharan Africa, one which are infected with different HIV-1 subtypes which are genetically distinct (C and CRF01-AE). Through comparing viral gp120 envelope sequences from progressors and non-progressors in both families we found similar findings for both subtypes.

Numerous reasons have been proposed to explain why disease progression within HIV-1-infected individuals is so varied. Host genetic factors have been associated with variant disease course, such as co-receptor availability.
(Paxton et al., 1999) or HLA allelic variation (O’Brien & Nelson, 2004). A few studies have associated virus characteristics with slower progression such as a defective nef gene (Birch et al., 2001) or decreased HIV-1 replication capacity (Jere et al., 2010). Nevertheless, few studies have investigated genotypic differences of viruses circulating in individuals with varied progression. The vast sequence diversity of the HIV-1 renders such studies difficult and often there is no apparent consistency in the trends observed, with a lack of clearly identifiable patterns of amino acid sequence changes among patients (O’Brien & Nelson, 2004). Here, the individuals studied were infected with genetically related virus strains rendering such comparisons possible despite the small number of study subjects included. The fact that the study participants were originally infected with genetically related viruses means that relationships can be drawn between subsequent viral evolution and markers of infection and course of disease progression in sibling pairs.

The overall V3 loop-positive charge has been linked with a co-receptor switch and this has been associated with faster disease progression. Our study is small in size, but we have identified increased V3-positive charges in the viruses circulating in the members that have progressed in their disease and for both families. We could not determine whether viruses with higher V3-positive charge have switched co-receptor usage phenotypes from R5 to X4, but with the exception of child C2b this is unlikely. In family 1, the members are infected with HIV-1 subtype C, which is characterized by a very low frequency of R5 to X4 phenotype switching (Abebe et al., 1999; Shankarappa et al., 1998) and from the study of large numbers of isolates such charge increases have not been related with a switch in co-receptor usage (Pollakis et al., 2001). In family 2, the members harbour a CRF01-AE virus, which has been more associated with co-receptor switching potential, but with the exception of child C2b all members had V3-positive charges of +4 or less without loss of the V3 stem glycosylation, again highly indicative of CCR5 utilization (Pollakis et al., 2001). In conclusion, our data suggest that V3 charge increase is associated with faster disease progression and is observed in both families. It should be noted that it is not possible to state whether the increase in overall V3-positive charge leads to disease progression or whether disease progression attenuates the immune defences allowing the virus to evolve towards higher charges.

It has previously been described that the length of the V1V2 region can vary during the course of HIV-1 infection, influencing the ability of the virus to infect CCR5-expressing cells or alter the potential for antibody directed neutralization (Sagar et al., 2006). In this study, we have demonstrated an association between disease progression and the length of the envelope protein together with the number of PNGS. The amino acid length or the number of PNGS varied between the study subjects and showed no association with disease progression. It could be that these parameters are strain dependent or that they evolve differently in separate hosts, depending on the potency of the immune responses mounted. When we calculated the aa/PNGS ratio we identified differences for subjects progressing versus not progressing in their disease course. This indicates that V1V2 length and glycosylation are interdependent parameters each contributing to the plasticity of the virus, potentially providing advantages for evading host restrictions. Again, the fact that individuals from one family were all infected with virus strains of common origin indicates that the host’s environment can drive viral evolution, which results in variant disease outcomes. In a previous study, identical twin brothers infected with HIV-1 at the same time showed parallel progression (Draenert et al., 2006). Other studies, investigating fraternal and identical twins, have shown that HIV-1-infected individuals can be born with very similar viral loads with subsequent envelope sequence divergence being dependent on quasi-species stability (Biggar et al., 2003; Sagar et al., 2006). In our study, the two sets of family members have variant compositions with mother M1 progressing, whilst M2b controlled the disease. Among the two sets of siblings, C1A and C1B demonstrated discordant viral loads at baseline, while C2a and C2b showed comparable plasma loads. In both families the second child progressed which has been reported in other studies (Draenert et al., 2006; Hutto et al., 1996). Most viral genes have been associated with varied disease progression and in a previous study of slow progressing sibling’s modifications in viral rev has been postulated to be associated with disease progression (Biggar et al., 2003; Tzitzivaclos et al., 2009).

Here, we have identified that HIV-1 originating from the same gene pool can evolve in very different ways within individuals that are close relatives. These individuals have a varied disease course, however, it is unclear whether subtle differences within the incoming virus strain, the host genetic background or the time of infection determine the final clinical outcome. What is unclear from this study, but would be pertinent to answer, is which variant immune responses or host factors can drive the virus to evolve in the specific directions described. Both the variable loop amino acid length and glycan numbers have been shown to play a role in HIV-1 immunity (Chaillon et al., 2011; Liu et al., 2011; Rusert et al., 2011), a role that may be lost in the late stages of the disease when immune responses are impaired (Borggren et al., 2011; Sagar et al., 2006).

The comparison of gp120 sequences from progressing versus non-progressing individuals demonstrated that the ratio between the amino acid length and the number of PNGS differs in such individuals. The analysis of gp120 sequences from the Los Alamos database supported this finding although in the latter analysis we made the assumption that sequences collected late or early corresponded to individuals that had subsequently progressed or not. The gp120 variable domains can tolerate up to 35% divergence (Hutto et al., 1996; Tzitzivacos et al., 2009).
especially within the variable domains with dramatic insertions, deletions and varied glycosylation patterns identified (Wei et al., 2003) and these changes have also been associated with neutralization resistance against autologous plasma (Gaschen et al., 2002; Rong et al., 2007). Here, we show that in terms of disease progression there is a strong interplay between the envelope characteristics described (amino acid length and number of PNGS) and in all likelihood represent variations in host-induced immunity. The better understanding of these differences, even within closely related individuals (such as siblings), needs to be deciphered to aid the development of future HIV-1 vaccines that need to induce broad protective immunity covering many viral variants in unrelated hosts.

METHODS

Study subjects. We studied HIV-1 infection in two families where the members (parents and children) have been infected for 10 years or more and some members had progressed, while others had not. Both families are of African origin and in each family two infants were HIV-1 infected, presumably with related virus strains, but with variant viral loads and CD4⁺ cell counts (Fig. 1). In both families, the youngest child had progressed and required ART upon their visit to the clinic. Time of seroconversion is not known for the parents and it is assumed that all children were vertically infected by the mother either in-utero or intra-partum. Blood samples were collected between 2001 and 2007 with consent obtained from participating parents and on behalf of their infants with all experiments conducted under medical ethical review relevant to the Amsterdam Cohorts Studies.

Viral RNA amplification and sequence analysis. For all individuals, longitudinal plasma samples were available with time points as indicated (Fig. 1). The viral genomic RNA was isolated as described previously (Boom et al., 1990). A region of the gp120 env was reverse-transcribed utilizing SuperScript III Reverse Transcriptase (Invitrogen) and amplified under standard PCR conditions with AmpliTaq DNA Polymerase (Applied Biosystems). The outer primers corresponding to the HXB2-positions 6540–6562 and 7834–7863 and the inner primers corresponding to the HXB2-positions 6540–6562 and 7789–7810 were used for the amplification. The PCR-amplion was cloned into the TOPO-TA vector (Invitrogen) and bacterial colonies carrying the HIV-1 env region were screened and sequenced (between 10 and 16 per time point). Nucleotide sequences were aligned with respect to the predicted amino acid sequence of the corresponding reference alignment extracted from the Los Alamos HIV database (http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html).

Phylogenetic analysis was performed using MEGA version 3.0 (Kumar et al., 2004) and confirmed by the DnADIST, NEIGHBOR and DRAWTREE options of the PHYLIP software package (http://evolution.genetics.Washington.edu/phylip.html). The distance matrix was generated by Kimura’s two-parameter estimation (Kumar et al., 2004). The number of potential N-linked glycosylation NX(S/T) positions was determined using the GLYCOSITE (http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html) provided through the Los Alamos site (Zhang et al., 2004). The WebPSMM bioinformatic tool was applied to predict HIV-1 co-receptor usage patterns (Jensen et al., 2003). Sequences of the V1V4 region of the gp120 envelope protein were extracted from the Los Alamos HIV-1 sequence database (http://www.hiv.lanl.gov/components/sequence/HIV/search/search.html) for comparative analysis. We defined four groups according to timing of infection, early (n=227) or late (n=191) and progression, namely AIDS (n=176) or LTNP (n=143) as defined in the patient information section of the website.

CD4⁺ T-cell count and viral load analyses. All subjects were tested for HIV-1 antibodies using the HIVSPOT assay (Genelabs Diagnostics) and Vironostika ELISA (Organon). Results were confirmed by Western blot analysis (HIV Blot 2.2; Genelabs Diagnostics). Blood was drawn at 6 month intervals. CD4⁺ T-lymphocyte count was carried with standard fluorescent activated cell sorting (FACS), using commercially available fluorescent-labelled antibodies (Becton Dickinson Immunocytometry). HIV-RNA was measured with the bDNA 3.0 assay (Bayer Diagnostics).

Statistical analyses. Descriptive statistics were performed utilizing the Prism package software (GraphPad software Inc., version 4.0) (http://www.graphpad.com/scientific-software/prism.htm). The tests performed were two tailed and chosen upon the outcome of the normality test and are indicated in the figure legends.

REFERENCES


Sheppard, H. W., Celum, C., Michael, N. L., O'Brien, S., Dean, M., Carrington, M., Dondero, D. & Buchbinder, S. P. (2002). HIV-1 infection in individuals with the CCR5-Delta32/Delta32 genotype:


