Characterization and Selection of HIV-1 Subtype C Isolates for Use in Vaccine Development

CAROLYN WILLIAMSON, LYNN MORRIS, MAUREEN F. MAUGHAN, LI-HUA PING, SERGEY A. DRYGA, ROBIN THOMAS, ELIZABETH A. REAP, TONIE CILLIERS, JOANNE VAN HARMELEN, ALVARO PASCUAL, GITA RAMJEE, GLENDA GRAY, ROBERT JOHNSTON, SALIM ABDOOL KARIM, and RONALD SWANSTROM

ABSTRACT

HIV-1 genetic diversity among circulating strains presents a major challenge for HIV-1 vaccine development, particularly for developing countries where less sequence information is available. To identify representative viruses for inclusion in candidate vaccines targeted for South Africa, we applied an efficient sequence survey strategy to samples from recently and chronically infected persons residing in potential vaccine trial sites. All 111 sequences were subtype C, including 30 partial gag, 26 partial pol, 27 V2–V3 env, and 28 V5–partial gp41 sequences. Of the 10 viruses cultured from recently infected individuals, 9 were R5 and 1 was R5X4. Two isolates, Du151 and Du422, collected within 2 months of infection, were selected as vaccine strains on the basis of their amino acid similarity to a derived South African consensus sequence. The selection of recently transmitted R5 isolates for vaccine design may provide an advantage in a subtype C R5-dominant epidemic. The full-length Du422 gag and Du151 pol and env genes were cloned into the Venezuelan equine encephalitis (VEE) replicon particle (VRP) expression system. Du422 Gag protein expressed from the VRP accumulated to a high level and was immunogenic as demonstrated by cytotoxic T lymphocyte responses in mice vaccinated with gag-VRPs. Optimization of codon use for VRP expression in human cells did not enhance expression of the gag gene. The cloned Du151 env gene encoded a functional protein as demonstrated by fusion of VRP-infected cells with cells expressing CD4 and CCR5. Genes identified in this study have been incorporated into the VEE VRP candidate vaccines targeted for clinical trial in South Africa.

INTRODUCTION

There is intense interest in the development of an effective vaccine able to curtail the continuing expansion of the HIV-1 epidemic. The number of people infected worldwide is estimated to be nearly 42 million, with more than 90% living in developing countries, 10% of whom are in South Africa. In South Africa it is estimated that approximately one in four pregnant women attending public antenatal clinics is infected. HIV-1 subtype C has emerged as the most significant subtype globally and is associated with more than 50% of infections. It is estimated that this subtype accounts for more than 95% of infections in the southern African region, where it is spread heterosexually. It is also the most common subtype in India.

Although the significance of genetic variation in vaccine efficacy is unresolved, there is some indication that it may be important. Although individuals vaccinated with a subtype B-based HIV-1 vaccine can elicit cross-subtype cytotoxic T
lymphocyte (CTL) reactivity\textsuperscript{11} and CTLs from individuals infected with one subtype can lyse target cells infected with a different subtype.\textsuperscript{12–15} This cross-reactive response may be diminished when compared with intrasubtype responses. In a study of HIV-1-resistant sex workers in Nairobi, Kenya, Rowland-Jones \textit{et al.}\textsuperscript{16} found that there was preferential CTL recognition of subtype A or D peptides, compared with B peptides, and, similarly, preferential CTL recognition was demonstrated in infected individuals in Thailand when target cells were infected with matched subtypes.\textsuperscript{15} In addition, comparison of well-characterised CTL epitopes shows a greater conservation within a subtype compared with between subtypes.\textsuperscript{17} Until more is known about the effect of diversity on vaccine efficacy, it is prudent to develop vaccines to match the viruses circulating in the target population.

Of the many vaccine approaches, the ones furthest in clinical development are the recombinant canarypox vector vaccines, either individually or in combination with the gp120 protein subunit. A promising novel approach is the Venezuelan equine encephalitis virus (VEE) replicon particle (VRP) vaccine. This VRP delivers self-replicating, noninfectious RNA to the lymphoid tissue where the foreign gene is expressed. This approach has been shown to elicit both CTL and antibody responses and was able to reduce virus load and disease progression in an SIV challenge experiment.\textsuperscript{18}

This study aimed to identify representative, recently transmitted HIV-1 subtype C isolates for use in the development of vaccines targeted for South Africa and the southern African region, and to provide an efficient strategy for surveying sequence diversity among circulating viruses. Genes from these isolates were cloned into VRP expression vectors and tested for immunogenicity and function. Phase I vaccine trials of VRP vaccine expressing the selected subtype C Gag are planned in the United States and South Africa. This study is unique in that vaccine strains were selected to match the circulating viruses in proposed phase III trial sites in an effort to minimize the potential impact of viral genetic diversity on vaccine efficacy. While a consensus sequence provides the advantage of being, on average, closer to the circulating virus, it is a predicted sequence and may not be a functional gene. This may be disadvantageous for vaccines aimed at eliciting antibody responses. Utilizing genes from viable isolates also has the advantage in future testing of immune responses in vaccinees as it enables cross-reactive immunogenicity studies of homologous and heterologous strains.

MATERIALS AND METHODS

Patients

HIV-1-infected individuals, recruited from three provinces in South Africa, were classified into two groups according to the timing of infection.

\textit{Acute infections.} Fifteen blood samples were obtained from female sex workers from five truck stops located along the road between Johannesburg and Durban (Du prefix; Table 1). At the time of this study, women were being tested monthly for sexually transmitted infections as part of their participation in a...
phase III HIV-1 prevention microbicide trial.\textsuperscript{19,20} HIV-1 infection was assessed by enzyme-linked immunosorbent assay (ELISA) (Abbott, Chicago, IL), and confirmed with a Vironostika HIV Uniform II micro-ELISA 4 system (Omnimed, Madison, WI). Viral load was measured by a Bayer HIV-1 RNA 3.0 assay (bDNA) (Bayer Diagnostics, Puteaux Cedex, France).

Prevalent, asymptomatic infections. Blood samples were collected from individuals who were clinically well, the majority of whom had CD4\textsuperscript{+} cell counts greater than 500 cells/\mu L. Seven samples were collected from women attending an antenatal clinic in Johannesburg (GG prefix), and nine samples were collected from men attending a sexually transmitted disease (STD) clinic at West Driefontein gold mine, Carltonville (RB prefix). Two additional samples from outpatient clinics in Cape Town (CT prefix) were included in this group.

Virus isolation and characterization

HIV-1 was isolated from peripheral blood mononuclear cells (PBMCs), using standard coculture techniques with mitogen-activated donor PBMCs.\textsuperscript{21} Isolates able to replicate in MT-2 cells as evidenced by the formation of syncytia were termed syncytium inducing (SI), whereas those that showed no syncytia within 6 days were termed nonsyncytium inducing (NSI). For coreceptor analysis, U87.CD4 cells stably transfected with either CCR5 or CXCR4 were used for viral infection (NIH Reference and Reagents Program, DAIDS, NIAID, NIH).\textsuperscript{22} Isolates that produced syncytia and increasing concentrations of p24 antigen were considered positive for virus growth as previously described.\textsuperscript{21}

RT-PCR of blood plasma

RNA was isolated from plasma (QIAamp viral RNA kit; Qiagen, Heidelberg, Germany) and cDNA to HIV-1 was isolated from peripheral blood mononuclear cells (PBMCs), using standard coculture techniques with mitogen-activated donor PBMCs.\textsuperscript{21} RT-PCR was performed as follows: one cycle at 94°C for 5 min; followed by 35 cycles of 94°C for 45 sec, 55°C for 45 sec, 72°C for 60 sec; then a final extension step of 72°C for 5 min. For nonnested PCR, 3 \mu L of PCR mix (1× Expand HF buffer with 1.5 mM MgCl\textsubscript{2}, 15 pmol of the downstream primer, 2 U of Expand High Fidelity enzyme mix [Boehringer, Mannheim, Germany]) was added to the room temperature reaction mixture (20 \mu L). PCR was carried out as follows: one cycle at 94°C for 2 min and 45 sec; then 40 cycles at 94°C for 45 sec, 53°C for 45 sec, and 68°C for 1 min (after the first 10 cycles, 1 min was added to the 68°C extension cycle for the next 10 cycles, 2 min was added to the next 10 cycles, and 3 min was added for the last 10 cycles). The final extension step lasted for 7 min. For samples that showed poor amplification, the PCR product was reamplified after gel purification.

Sequencing and sequence analysis

The amplified DNA fragments were purified with a QIAquick PCR purification kit (Qiagen) and both strands were sequenced, using inner PCR primers and an ABI-Prism 377 DNA sequencer (Applied Biosystems, Foster City, CA). All sequences have been deposited in GenBank (AF543896–AF544010).

The nucleic acid sequences (\textit{gag} [313 contiguous codons, 939 nucleotides], \textit{pol} [278 contiguous codons, 834 nucleotides], and \textit{env} [229 codons in two noncontiguous segments, 687 nucleotides]) were compared phylogenetically with a subtype C consensus sequence (Consensus C; http://hiv-web.lanl.gov), published subtype C sequences from the southern Africa region (http://hiv-web.lanl.gov), and the South African consensus sequence (ConsensusZA) defining the most common nucleotide at each position in the sequence generated in this current work. Sequences were aligned with DNAMAN (version 4.0; Lynnon Biosoft, Vaudreuil-Dorion, Canada) with final manual adjustment. Neighbor-joining phylogenetic trees were generated by PHYLIP (Phylogeny Inference Package, version 3.56c; J. Felsenstein, Department of Genetics, University of Washington) using corrected distance matrices generated by Dnadist and the maximum likelihood model to incorporate different rates of transition and transversion (1.5). Trees were generated by Neighbor and drawn with TreeView (version 1.6.1; R. Page homepage, http://taxonomy.zoology.gla.ac.uk/rod/rod.html).

Amino acid sequences were derived and used to determine a South African consensus sequence. Sequences were aligned with Pileup in the Genetics Computer Group (GCG) (Wisconsin Package version 10.0). Amino acid sequence analysis was performed with corrected distances, using the distances determined by Distances.

Cloning of full-length genes

Full-length \textit{env} and \textit{pol} genes were generated by nested PCR from whole cell DNA purified from PBMC cultures infected with the Du151 isolate (low passage). DNA was extracted from pellet cells with a QIAamp DNAeasy tissue kit (Qiagen). The
following primers were used: outer env (5'-GAATTACGGAAAGCAGAAGAC-3'), positions 6195–6218; 5'-CACTATGCTGCTTTTGACCCTAC-3', positions 8807–8803); inner env (5'-ATGAGAGTGTGGGGATACAG-3', positions 6225–6245; 5'-TTATTTAGCCAGCTCTCAGC-3', positions 8785–8795); outer pol (5'-ACCAATTACAGCTGACTGACG-3', positions 2050–2073; 5'-GATTAGCTCTCTGAAACATACATATGGTG-3', positions 5119–5148); inner pol (5'-TTTGGGAACTAGTATTTCCTCTC-3', positions 2058–2108; 5'-AAATCTTCATGCTGCTACCTGCCCACAC-3', positions 5068–5094).

PCR was carried out with the Expand Long Template PCR system (Boehringer) in a total volume of 50 μl containing 1× Expand PCR buffer 3, 2.5 μl of 10 mM dNTPs, 3 μl of 10 μM outer primers, 5 μl of DNA template, and 0.75 μl of the enzyme mix. PCR cycling was carried out as follows: one cycle at 94°C for 20 sec; 40 cycles at 94°C for 35 sec, 55°C for 45 sec, and 68°C for 2 min (after the first 10 cycles, 1 min was added to the 68°C extension step for the next 10 cycles, 2 min was added to the next 10 cycles, and 3 min was added for the last 10 cycles); the final extension time was 7 min. An aliquot of 1 μl of the first PCR was used for the second PCR, using the same PCR program but with the inner primers.

PCR products were purified with a QIAquick PCR purification kit (Qiagen) and were cloned into pT7 Blue after end conversion (Novagen, Madison, WI). Clones with inserts were identified by colony PCR and one clone was selected for each gene and sequenced in its entirety.

The full-length gag gene was generated by RT-PCR directly from RNA purified from plasma. The following primers were used for the outer nested PCR, with the outer reverse primer used for generating cDNA: o-gag-f (5'-AAATCTCTAGCAGTGCGC-3', positions 623–640) and o-gag-r (5'-ACCTCATCATCTGCTCTATC-3', positions 2390–2408). PCR cycles were as follows: 94°C for 5 min; then 35 cycles of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 1 min; then 72°C for 5 min. The inner PCR primers were as follows: p17-eco-f (5'-GAATTACGGAAAGCAGAAGAC-3', positions 790–810) and I-gag-sal (5'-GTCGACAAATCTGCTCATCTGCTCTATC-3', positions 2329–2350). PCR cycles were as follows: 94°C for 5 min; then 5 cycles of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 1 min; then 35 cycles of 94°C for 30 sec and 72°C for 1 min; then 72°C for 1 min. The PCR product was cloned into the pGEM-T Easy vector (Promega, Madison, WI) and positive clones were identified by restriction enzyme analysis. The full-length gene was sequenced in its entirety.

The full-length gag and env genes were cloned into the VEE replicon plasmid pERK, which is a derivative of pV2R21 with additional restriction enzyme sites added (details available on request). The initial cloning step involved the PCR amplification of the full length genes to introduce a downstream flanking PacI site. Cloning of the gag gene included the incorporation of a mutation in the second gag codon to change it from a Gly codon to an Ala codon to ablate the encoded myristoylation signal.

The Du422 gag gene was also synthesized to allow the incorporation of codons representing those most frequently used in human cells (hu-gag; Qiagen Operon, Alameda, CA). This gene was cloned into pERK in the same manner as the viral genes.

Expression of full-length gag and env genes

BHK cells and VERO cells were infected with Du422 gag-VRPs and hu-gag-VRPs at a multiplicity of infection of 3.5 infectious units per cell for 18 hr at 37°C. The cells were rinsed with phosphate-buffered saline (PBS) and resuspended in 0.5 ml of gel loading buffer containing sodium dodecyl sulfate (SDS) and 2-mercaptoethanol (Invitrogen, Carlsbad, CA). A cell lysate derived from approximately 2000 cells was loaded onto a 4–12% gradient acrylamide gel (Invitrogen) and the proteins were resolved by electrophoresis. Proteins were either stained in the gel with Coomassie Brilliant Blue or transferred onto a polyvinylidene difluoride (PVDF) membrane and probed with serum from an HIV-positive patient followed by goat anti-human–alkaline phosphatase (AP) conjugate. The membrane was developed with NBT/BCIP (Bio-Rad, Hercules, CA).

U87.CD4-CR5 and U87.CD4-CXCR4 cells22 and BHK cells were used to examine the expression and fusogenicity of Du151 Env protein expressed from the VRPs. Cells were infected with Du151 env-VRPs or green fluorescent protein (GFP) VRPs as a control at a multiplicity of infection of 3 infectious units per cell and examined for syncytia 18 hr postinfection. For Western blot analysis, the cells were rinsed with PBS and resuspended in 0.1 ml of lysis buffer (0.05 M Tris-HCl [pH 8.0], 0.1 M NaCl, 0.001 M EDTA, 0.5% Nonidet P-40 [NP-40], 1 μM aprotinin, 1 μM leupeptin, 1 μM pepstatin, and 0.2 mM phenylmethylsulfonyl fluoride). The equivalent of 5×10^4 cells was mixed with 5 μl of gel loading buffer and analyzed as described above.

Gag immunogenicity was tested with 6- to 8-week-old female BALB/c mice, obtained from Charles River Laboratories (Wilmington, MA). Groups of mice were inoculated subcutaneously in both rear footpads with a total of 10^6 infectious units of gag-VRPs or PBS. Booster immunizations were performed 33 days later at the same dose and site. At least 1 week after the booster immunization splenocytes were harvested. Effector spleen cells were stimulated in vitro with syngeneic spleen cells infected with recombinant vaccinia virus expressing the HIV-1 gag gene (subtype C, 96ZM651.8) or vaccinia virus alone. The syngeneic stimulator cells were prepared by infection with either the gag-expressing vaccinia virus or vaccinia virus alone at a multiplicity of infection of 5 for 1 hr. Excess virus was removed and the cells were incubated for an additional 4 hr. The cells were washed again and then UV irradiated in the presence of psoralen. The infected spleen cells were then cultured with splenic lymphocytes from experimental animals at a ratio of 1:3 for 6 days. The cells were then purified through Lympholyte-M (Cedarlane, Hornby, ON) and chromium release assays were performed. MHC-matched P815 mastocytoma target cells (American Type Culture Collection [ATCC], Manassas, VA) were infected with the gag-expressing vaccinia virus or vaccinia virus alone, incubated overnight, labeled with 51Cr for 1 hr, washed, and then incubated with effector cells for 4 hr. The percent specific lysis, measured as released 51Cr, was calculated as follows: [(experimental release − spontaneous release)/(maximum release − spontaneous release)] × 100.
RESULTS

Patient populations and isolates used to determine HIV-1 diversity in South Africa

To characterize currently circulating viruses, this study targeted recent seroconvertors (Du prefix) and asymptomatic individuals with CD4\(^+\) cell counts > 500 cells/\(\mu\)l (GG, RB, and CT prefix). Samples from the sex workers (Du cohort, \(n = 15\)) working along the main trucking route between Durban and Johannesburg were likely to reflect isolates from multiple geographical origins as the truck drivers, who are the main clients, travel between provinces of South Africa and to other countries in southern Africa. To further ensure that isolates were representative of the genetic diversity within the country, samples were also collected from a second potential vaccine trial site, Chris Hani Baragwanath Hospital, Soweto (GG, \(n = 7\)), as well as from West Driefontein gold mine, Carltonville (RB, \(n = 9\)), and outpatient clinics in Cape Town (CT, \(n = 2\)).

Most women in the Du cohort were identified within 1 year of infection, although two had seroconverted 14–16 months before sample collection (Du174 and Du179) (Table 1). The CD4\(^+\) cell counts ranged from 367 to 841 cells/\(\mu\)l, and a wide range of viral loads was detected with high viral loads possibly reflecting viremia before the containment of viral replication. The median CD4\(^+\) cell count for the asymptomatic group was 579 cells/\(\mu\)l, with all individuals having a CD4\(^+\) cell count above 500 cells/\(\mu\)l, except for GG1, GG6, and GG10 who had CD4\(^+\) cell counts between 383 and 423 cells/\(\mu\)l. CTSC1 and CTSC2 has CD4\(^+\) cell counts of 271 and 390 cells/\(\mu\)l, respectively.

Virus isolation was attempted using blood samples from all 15 women in the Du cohort. Isolates were obtained from six women, using the baseline sample, and an additional four isolates were obtained with later samples (Table 1). Nine of the 10 isolates were nonsyncytium inducing (NSI) in an MT-2 assay and replicated in the U87.CD4.CCR5 cells, indicating that they were R5 isolates. One isolate, Du179, induced syncytia (SI) when grown in MT-2 cells and also grew in both U87.CD4.CCR5 cells and U87.CD4.CXCR4 cells, indicating that it is a dual-tropic virus (RSX4). Biological clones of Du179 confirmed that it is a dual-tropic isolate and not a mixture of R5 and X4 variants (data not shown).

Phylogenetic analysis of segments of gag, pol, and env

A sequence-sampling strategy was devised in which two segments each of the gag, pol, and env genes were amplified, with the lengths of the amplicons selected to enable generation of the entire sequence in a single sequencing run. The two overlapping segments for gag and pol gave continuous sequence, 939 nucleotides and 834 nucleotides in length, respectively, whereas for env the two segments were discontinuous to avoid inclusion of the variable length regions V1, V2, V4, and V5, yielding 687 nucleotides in total. Sampling over the length of the genome permitted assessment of potential intersubtype recombination. All sequences were phylogenetically compared with reference subtypes\(^25\) and were identified as subtype C in all regions analyzed, indicating a relative absence of intersubtype recombinants in this geographical region (results not shown).

To investigate the phylogenetic relationship among southern African viruses these sequences were then compared with all available subtype C sequences from the southern African region, as well as with the reference subtype C sequences (BR025 and ET2220) (Fig. 1). A number of Botswanan clusters, previously identified,\(^9\) were maintained in this analysis. Only limited internal structure was observed in all phylogenetic trees. However, a subcluster of South African sequences (Du422, Du151, Du281, Du467, Du457, GG3, RB28, Du123, Du368, and RB12) was largely preserved between the gag and env gene segments although this relationship was not supported by high bootstrap values. A significant relationship was detected between RB14 and RB18, and was supported by high bootstrap values in all regions analyzed, indicating epidemiological linkage.

The average nucleotide distance between the gag, pol, and env gene segments of subtype C viruses representing the southern African region was 7.8, 6.0, and 9.9%, respectively. This did not differ markedly from nucleotide distances between gene segments representing the viruses from South Africa only, which were 7.1, 5.1, and 11.3% for gag, pol, and env, respectively.

Selection of representative subtype C HIV-1 genes for vaccine development

To identify representative genes for candidate vaccines, each sequence was translated and compared with all the other amino acid sequences from this study; in addition, a consensus sequence was generated (South African consensus, SAcon) (Fig. 2). In Gag, sequences were between 90.5 and 98.7% identical to the South African consensus sequence, with Du422 being closest to the consensus and the one chosen as the vaccine candidate. The amino acid sequence of the cloned full-length Du422 gag gene was 99.4% identical to the original sequence fragment derived from viral RNA in plasma. The remaining South African viruses were found to be between 90.5 and 96.7% identical to the Du422 gag clone. For 83% of the viruses, the Du422 clone sequence fell within the upper quartile, and in the upper half for all of them when each sequence was compared individually with all of the other sequences. On average, each isolate was 0.7% more similar to the South African consensus sequence than to the Du422 clone, except for one isolate (CTSC1), which was more similar to the clone.

As expected, the inferred Pol protein sequences showed the highest level of sequence identity among the three genes examined. Du151 was chosen as being closest to the South African consensus, at 98.9%. The Du151 Pol amino acid sequence inferred from the cloned pol gene shared between 96.8 and 93% sequence identity with all but one of the other viruses. When each sequence was compared with all the other viral sequences the Du151 clone was in the upper quartile 48% of the time and in the upper half 72% of the time. When each sequence was compared with the South African consensus sequence and the Du151 pol clone, there was an average of 1.6% greater similarity to the consensus sequence than to the clone sequence.

The inferred Env amino acid sequence displayed the greatest variability. Nine viral sequences shared between 92 and 95% identity with the South African consensus sequence. From this group we chose Du151 from the recent seroconverter cohort as...
this sample also gave a virus isolate that grew well in culture. The cloned sequence ranked in the upper quartile 54% of the time and in the upper half 77% of the time when each sequence was compared with all of the other viral sequences. All but one of the sequences were more similar to the South African consensus sequence than to the cloned sequence (average of 3.6%).

A comparison of the V3 amino acid sequences from the Du cohort showed Duv79 to have an amino acid deletion at position 22 and an overall positive charge of +1.6, which is consistent with its SI phenotype. The remainder of the V3 amino acid sequences had charges of between +1.3 and +1.5, consistent with CCR5 usage.
The full-length Du422 gag and Du151 env genes were cloned in the VEE replicon vector pERK. The second codon of gag was modified from Gly to Ala to ablate the myristoylation signal present at the N terminus of the Gag protein. This mutation causes Gag to be retained in the cell.

26–28 In parallel, the entire gag gene was chemically synthesized, altering the codons to represent optimized human codon usage (hu-gag). VEE replicon particles (VRPs) were generated as previously described18 and used to infect BHK cells and VERO cells, and Gag expression was monitored after gel electrophoresis by both staining for protein mass and by Western blot analysis. Large amounts of p55 Gag protein accumulated in the cells (Fig. 3). In addition, the optimization of codon usage did not affect the levels of expression. Equivalent levels of expression were seen in both cell types.

The gag-VRP was also tested for immunogenicity. A strong Gag-specific CTL response was detectable in cells cultured from the spleens of gag-VRP-immunized mice but not from the control mice immunized with PBS (Fig. 3C). Spleen cells from both gag-VRP- and PBS-immunized mice cultured with syngeneic cells expressing the vaccinia virus vector alone showed no Gag-specific CTL response. gag-VRP-immunized mice developed a significant antibody response with an average reciprocal anti-Gag antibody titer, as measured by ELISA, in excess of $10^4$ (data not shown).

Env expression and function were assessed by using Du151 env-VRPs to infect BHK cells, U87.CD4-CXCR4 cells, U87.CD4-CCR5 cells, or mixtures of BHK cells with either of the other two cell types. Expression was assessed by Western blot analysis with antisera from subject Du151 to probe the blot. All three cell types expressed the full-length env gene product gp160 (Fig. 3A). Unexpectedly, the cells expressing the cognate coreceptor accumulated more processed env proteins than those expressing the xenic coreceptor (Fig. 3B). These results are consistent with the coreceptor usage of the Du151 isolate (CCR5) and demonstrate that the cloned env gene encodes a functional protein.

**DISCUSSION**

Within an HIV-1 subtype, viruses are more closely related to each other than to those belonging to other subtypes, and consequently there are a greater number of shared amino acids and potential epitopes. It is therefore relevant to develop vaccines based on viral sequences circulating in the target population in an attempt to maximize the number of potentially cross-reactive epitopes between the vaccine and the challenge virus. Two isolates were selected as the source of genes for inclusion into vaccines: Du422 for gag, and Du151 for pol and env (Figs. 1 and 2). These genes have been cloned, and the gag and env genes have been used to express full-length viral proteins (Fig. 3). Optimization of codon use for expression in human cells did not enhance expression of the gag gene in the robust VRP system. This suggests that improvements seen with codon optimization in other systems may be related to nuclear expression and/or export of the mRNA to the cytoplasm. Thus unaltered native HIV-1 genes may function well for expression using cytoplasmic expression vectors such as VEE and poxviruses. Du151 env and pol genes have also been codon optimized and are available for use. Genes from these vaccines strains are presently being included in a number of different vaccine approaches including VEE replicon particles (VRPs),29 oligomeric proteins,30 adeno-associated virus vectors,31 DNA vaccines, the MVA (modified vaccinia Ankara) vaccine vector, and the BCG vaccine vector.32 Of these, VRPs represent the vaccine candidate furthest along in development, and phase I trials to assess the safety of VRPs expressing the Du422 gag gene are planned for South Africa and the United States. In an...
FIG. 2. Pairwise distance comparisons among South African HIV-1 amino acid sequences. The percent amino acid identity (calculated by Distances, GCG) was determined for each gene segment against each of the other South African viral sequences generated in this study: (A) Gag; (B) Pol; (C) Env. The total range of values is shown by the vertical bar, the box shows the inner quartile range, and the horizontal bar shows the median value. In addition, the sequence for each virus generated from the bulk sequence of plasma viral RNA was compared with the same region from the cloned Du422 gag gene (A), the cloned Du151 pol gene (B), or the cloned Du151 env gene (C), and that comparison is indicated by the solid circle. Finally, the bulk sequence for each virus was compared with the South African consensus sequence, generated as the most common amino acid at each position from among the sequences determined in this study. That comparison is shown in the open circle. The order of viral sequence comparisons has been arbitrarily placed in decreasing order of similarity to the cloned gene sequence.
FIG. 3. Expression of cloned *env* and *gag* genes, using VRPs. (A) Cells were infected with Du151 *env*-VRPs. At 18 hr postinfection the cells were lysed and the lysate was run in a denaturing polyacrylamide gel. Proteins were transferred out of the gel and onto a filter, and the filter was probed with serum from subject Du151 in a Western analysis. Lane 1, uninfected U87.CD4-CXCR4 cells; lane 2, uninfected U87.CD4-CCR5 cells; lane 3, infection of a mixed culture of U87.CD4-CXCR4 cells and BHK cells (mixtures were used as a positive control in case the U87 cells were refractory to infection with VRPs, which did not turn out to be the case); lane 4, infected U87.CD4-CXCR4 cells; lane 5, infected BHK cells; lane 6, infection of a mixture of BHK cells and U87.CD4-CCR5 cells; lane 7, infected U87.CD4-CCR5 cells. The position and molecular weight of markers run in the same gel are shown on the right, and the inferred positions of gp160 and gp120 are shown on the left. (B) BHK and VERO cells were infected with Du422 *gag*-VRPs or with Du422 hu-*gag*-VRPs. As a control, another set of cells was infected with VRPs expressing green fluorescent protein (GFP). Eighteen to 20 hr after infection, the cells were harvested, lysed, and run in a denaturing polyacrylamide gel. Lanes 1–6: The gel was stained with Coomassie Brilliant Blue. Lanes 7–12: Proteins in the gel were transferred to a filter and probed in a Western analysis with serum from an HIV-1-infected person. The positions of Gag and GFP are shown, as are the positions of molecular weight markers run in the same gel. The heavy band seen at 72 kDa in lane 3 is BSA. Lanes 1, 4, 7, and 10: Cells infected with *gag*-VRPs. Lanes 2, 5, 8, and 11: Cells infected with hu-*gag*-VRPs. Lanes 3, 6, 9, and 12: Cells infected with GFP-VRPs. (C) Mice were immunized (prime and boost) with Du422 *gag*-VRPs and spleen cells were harvested. Effector cells were stimulated in culture by the addition of cells infected with vaccinia virus expressing HIV-1 *gag*. The presence of antigen-specific effector cells (E) was demonstrated by mixing the effector cells with MHC-matched, 51Cr-loaded target cells (T) infected with a control vaccinia virus (squares, vector) or vaccinia virus expressing *gag* (circles). The ratio of effector to target cells was varied and the amount of cell lysis of the target cells by the effector cells was measured by the release of the label from the target cells. One representative experiment of three is shown. Two spleens were pooled in each experiment.
extension of this current investigation, the Du179 env gene has also been included in potential vaccines based on the presence of broadly cross-reacting neutralizing antibodies in this subject. A similar strategy has been used to identify candidate vaccine genes for subtype B HIV-1 (L.-H. Ping and R. Swanstrom, unpublished observation).

An alternative approach to using representative viral sequences would be to construct a gene based on a consensus sequence, although such a sequence may not necessarily be a functional gene. The viral sequences chosen from our sequence survey are similar to the consensus sequence (98.7, 98.9, and 95% similarity in Gag, Pol, and Env, respectively) and performed well when compared with a consensus sequence in cross-sectional distance comparisons (Fig. 2). Still, there is an inherent sampling limitation in selecting viral sequences for inclusion in a vaccine because of incomplete knowledge of all circulating viruses. In addition, as it will be some time before any vaccine is implemented on a large scale, further genetic drift, the introduction of other subtypes, and broadening of the diversity are likely trends that will ultimately impact the quality of the match between the vaccine sequence and the target epidemic. In a separate analysis of all available subtype C full-length sequences, Gaschen et al. found that the median percent similarity between C clade sequences and the C consensus sequence was 95.4% in p24. This was not markedly different from Du422 p24, which was 94.8% similar to C clade sequences. A comparison of C clade gp160 sequences with the C consensus demonstrated 87.5% similarity, in contrast to 81.3% similarity when Du422 gp160 was compared with clade C gp160s. Although the env gene, as the region with the greatest diversity, would have the most potential benefit from creating a consensus sequence, it is likely that manipulations of Env to improve its antigenicity (e.g., see Reitter et al. and Earl et al.) will have greatest impact on the performance of an Env-based vaccine.

The sequence survey strategy was designed to improve the efficiency of data acquisition. The six amplicons were each designed to approximate the length of a DNA segment that can be reliably sequenced. The major replicative genes are represented in the sequence, spanning across the genome to allow the assessment of intersubtype recombinants. The length-variable regions of env were avoided so that bulk sequence could be used without loss of register. Representative candidate genes from currently circulating viruses were identified and characterized and have been included in candidate vaccines. Given the unknown impact sequence diversity will have on vaccine efficacy, there is a parallel need to maximize the magnitude of the immune response to any vaccine so as to reduce the potential impact of this diversity.

ACKNOWLEDGMENTS

This study was funded by the International AIDS Vaccine Initiative; the Poliomyelitis Research Foundation and THRIP, South Africa; the South African AIDS Vaccine Initiative; and by an award from the National Institutes of Health (P01-AI46023). We acknowledge the support of the UNC Centre for AIDS Research (P30-A150410). We thank Ron Ballard, Ye Hun, and Gary Maartens for blood samples; Jeff Frelinger for assistance with CTL work; Clive Gray for assistance with specimen processing; Mary Phoswa for virus isolation; and Anita Schwegmann for assistance with PCR. We also thank Francine McCutchan for useful discussions; Wolfgang Resch for assistance in preparing the figures; and Ian Caley, Pamela Norberg, and Tameta Clark for assistance in the analysis of expression studies. Alvaro Pascual was the recipient of a fellowship (PR98 0028531309) from the Ministerio de Educacion y Cultura (Spain). Lynn Morris is a recipient of a Wellcome Trust International Senior Research Fellowship in Biomedical Sciences.
REFERENCES


30. Moore JP, Binley J, Sanders R, Schuelke N, Maddon P, and Ol-


Address reprint requests to:
Carolyn Williamson
Division of Medical Virology
University of Cape Town, Observatory
Cape Town, South Africa 7925

E-mail: cwilliam@curie.uct.ac.za
This article has been cited by:


13. Sonia Resik, Philippe Lemer, Li-Hua Ping, Vivian Kouri, Jose Joanes, Jorge Pérez, Anne-Mieke Vandamme, Ronald Swanstrom. 2007. Limitations to Contact Tracing and Phylogenetic Analysis in Establishing HIV Type 1 Transmission Networks in Cuba. Limitations to Contact Tracing And Phylogenetic Analysis in Establishing HIV Type 1 Transmission Networks in Cuba. *AIDS Research and Human Retroviruses* 23:3, 347-356. [Abstract] [PDF] [PDF Plus]
14. Mia Coetzer, Tonie Cilliers, Maria Papathanasopoulos, Gita Ramjee, Salim Abdool Karim, Carolyn Williamson, Lynn Morris. 2007. Longitudinal Analysis of HIV Type 1 Subtype C Envelope Sequences from South Africa. Longitudinal Analysis of HIV Type 1 Subtype C Envelope Sequences from South Africa. *AIDS Research and Human Retroviruses* 23:2, 316-321. [Abstract] [PDF] [PDF Plus]


18. Eleanor Cave, Marc S. Weinberg, Tonie Cilliers, Sergio Carmona, Lynn Morris, Patrick Arbuthnot. 2006. Silencing of HIV-1 Subtype C Primary Isolates by Expressed Small Hairpin RNAs Targeted to gag Silencing of HIV-1 Subtype C Primary Isolates by Expressed Small Hairpin RNAs Targeted to gag. *AIDS Research and Human Retroviruses* 22:5, 401-410. [Abstract] [PDF] [PDF Plus]


21. Tonic Cilliers, Trudy Patience, Candice Pillay, Maria Papathanasopoulos, Lynn Morris. 2004. Sensitivity of HIV Type 1 Subtype C Isolates to the Entry Inhibitor T-20. *AIDS Research and Human Retroviruses* 20:5, 477-482. [Abstract] [PDF] [PDF Plus]


23. Matthew E. Harris, Shlomo Maayan, Bohye Kim, Michael Zeira, Guido Ferrari, Deborah L. Birx, Francine E. McCutchan. 2003. A Cluster of HIV Type 1 Subtype C Sequences from Ethiopia, Observed in Full Genome Analysis, Is Not Sustained in Subgenomic Regions. A Cluster of HIV Type 1 Subtype C Sequences from Ethiopia, Observed in Full Genome Analysis, Is Not Sustained in Subgenomic Regions. *AIDS Research and Human Retroviruses* 19:12, 1125-1133. [Abstract] [PDF] [PDF Plus]