Biokemistri Vol. 5, No. 1, June, 1995
Printed in Nigeria

0795-8080/95 $3.00 + 0.00
Klobex Academic Publishers

BKR9540/5106

MOLECULAR CHARACTERIZATION OF THE gp120 ENCODING REGION OF A NEW SUBTYPE A STRAIN OF HIV-1 (HIV-1 IbNg) FROM NIGERIA

T. M. Howard[1], O. D. Olaleye[1,2]*, and S. Rasheed[1]

[1]Laboratory of Viral Oncology and AIDS Research, University of Southern California, School of Medicine, 1804 N. Soto St., Los Angeles, CA 90032-3626

[2]Department of Virology, College of Medicine, University College Hospital, Ibadan, Nigeria

(Received April 28, 1995)

ABSTRACT: We have characterized the coding region of the surface protein, gp120, of an HIV-1 strain (HIV-1 IbNg) from Nigeria. This represents the first HIV isolate from this region of Africa. Based on the nucleic acid sequence of the gp120 encoding region as well as the deduced amino acid seqience of this protein, the virus has been identified to cluster with the clade A strains of HIV-1, which have been primarily isolated from regions of Western and Central Africa. Comparison of the deduced amino acid sequences of the HIV-1 IbNg gp120 proteins with those of other A clade strains revealed the conservation of 19 cysteine residues among these viruses as well as the number of potential N-linked glycosylation sites and their relative positions. Since there is very little information on the genomic sequences of HIV-1 strains belonging to clade A, which are predominantly isolated from areas hardest hit by the AIDS epidemic, characterization of the entire genomic sequence of HIV-1 IbNg is important for identifying the molecular structure(s) of the viral gene(s) which may be involved in the pathogenesis of subtype A strains as well as in the development of effective HIV-1 vaccines.

Key Words: AIDS; HIV-1;Genomic sequences; gp120 encoding region.

INTRODUCTION

A major impediment in controlling the spread of HIV-1 infection is the ability of this virus to mutate, allowing it to escape the immune system of the host and survive as a quasi-species. Because of the high mutation rate and genetic diversity of the envelope protein, gp120, the viral env gene has been extensively studied.

Analysis of the envelope encoding sequences of HIV-1 isolates taken from the major centres of the AIDS pandemic has identified at least six distinct HIV-1 genetic clusters or clades, tentatively designated A, B, C, D, E and F, that are approximately equidistant from one another: A - Western and Central Africa, B - North and South America, Europe and Thailand, C - Southern Africa and India, D - Central Africa, E- Thailand, Central Africa and India, and F - Brazil, Romania and Zaire (1). Recently, new strains of HIV-1 have been identified which do not readily cluster in any of the six operationally defined clades, suggesting the existence of additional genetic clusters (1-3). The distribution and number of genetic clusters will therefore change in future as the number of HIV-1 nucleotide sequences of

*To whom correspondence should be addressed at the Department of Virology, College of Medicine, University College Hospital, Ibadan, Nigeria.

isolates collected from different geaofraphical regions of the world increases.

To develop an effective HIV-1 vaccine, it would be important to determine the extent and distribution of HIV variability throughout the world and identify dominant HIV-1 subtype(s) or clade(s) for specific geographic regions. However, the present database contains very few complete HIV envelope sequences and most subtypes have been identified only on the basis of partial envelope sequences (4).

Currently, very little information is available regarding the genetic variability of HIV-1 within different geographical locations of Africa. Of major concern is the paucity of basic epidemiological information from many developing African countries that have been hard hit by the HIV epidemic. In addition, basic immunological, virological and biological data on HIV strains from Africa are scanty, presumably because only a few strains of HIV have been isolated and characterized from this region of the world.

Recently, we reported the results of epidemiologic studies on the seroprevalence and distribution of HIV-1 and HIV-2 infections in Nigeria (5,6). We have also isolated a new strain of HIV-1, designated HIV-1_{IbNg}, from the peripheral blood mononuclear cells of an apparently healthy, 23 year-old Nigerian male. The details of virus isolation and biological characterization of HIV-1_{IbNg} are as described in the previous paper (7). We describe here the complete nucleotide sequence and analysis of the gp120 encoding region of this virus. Evidence is also presented to support the view that this virus clusters with HIV-1 of the A subtype or clade of strains which are predominantly isolated from inhabitants of Central and West Africa. This is the first report of a strain of HIV-1 from Nigeria.

MATERIALS AND METHODS

Origin of HIV-1_{IbNg}:

HIV-1_{IbNg} was isolated from the peripheral blood mononuclear cells (PBMCs) of an apparently healthy, 23 year-old male blood donor from Ibadan, Nigeria. Since a routine pre-transfusion safety screening of the donated blood was found to contain antibodies fo HIV, PBMCs were isolated by Ficoll-Hypaque density gradient centrifugation within 24 hours of collection and frozen in a medium composed of 80% heat-inactivated foetal bovine serum, 10% RPMI-1640 and 10% dimethyl sulphoxide (DMSO) at -70^{o}C. The frozen cells were transported on dry ice to the Laboratory of Viral Oncology and AIDS Research, University of Southern California, Los Angeles, CA.

Virus Propagation:

A mixture of HIV-1_{IbNg} infected cells and culture supernatant from the initial HIV-1_{IbNg} isolation was used to infect approximately 1.2 x 10^{8} PBMCs from a single HIV sero-negative donor pre-stimulated with phytohaemagglutinin (5 µg/ml) (7). Cell cultures were maintained in RPMI-1640 medium supplemented with 10% heat-inactivated foetal bovine serum, 5% interleukin-2 (IL-2), and 50 µg/ml gentamycin. Cultures were maintained at 37^{o}C in an atmosphere of 5% CO_2. Extra PBMCs, 50. x 10^{7}, from sero-negative donors and pre-stimulated with phytohaemagglutinin were added on day 4 and day 6. The co-cultivated cells were finally harvested on day 8.

RNA Isolation:

Total cytoplasmic RNA was isolated from approximately 1.0 x 10^{7} infected PBMCs 8 days post-infection. Cells were pelleted by centrifugation at 300 x g for 10 min at 4^{o}C and then washed free of medium in 20 ml of ice-cold phosphate buffered saline (PBS) [137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4.7H_2O$, 1.4 mM KH_2PO_4, pH 7.3]. The washed cells were again pelleted and then resuspended in 3 ml of an ice-cold buffer containing 10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 1.5 mM $MgCl_2$. Cells were lysed by the addition of 0.1 volume (300 µl) of 5% (v/v) Nonidet P-40 and 5% (v/v) deoxycholic acid (DOC). The cellular lysates were clarified twice by centrifugation at 300 x g for 10 min at 4^{o}C. An equal volume of 2x proteinase-K (PK) buffer, containing 200 mM Tris-HCl, pH 7.5, 300 mM NaCl, 25 mM EDTA and 2%

sodium dodecyl sulphate (SDS) was added along with proteinase K to a final concentration of 200 µg/ml. The lysates were incubated at 37^{O}C for 1 hr and extracted twice with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1 v/v/v) and once with an equal volume of chloroform/isoamyl alcohol (24:1 v/v). Finally, 3M sodium acetate, pH 5.2 (0.1 volume) and 100% ethanol (2.5 volumes) was added to the lysates to precipitate the RNA. RNA was stored at -20^{O}C as an ethanol precipitate.

First Strand cDNA Synthesis:

Approximately 15 µg of cytoplasmic RNA was recovered from the ethanol precipitate by centrifugation. The RNA pellet was resuspended in 20 µl water, denatured at 65^{O}C for 3 min and quenched on ice. The first strand cDNA was made using avian myeloblastosis virus (AMV) reverse transcriptase and random hexanucleotides as primers. The resuspended RNA was added to the reverse transcription mixture containing 50 mM Tris-HCl, pH 8.5, 8 mM $MgCl_2$, 30 mM KCl, 1 mM DTT, 80 Units of human placental ribonuclease inhibitor (Boehringer Mannheim), 5 mM of each dNTP and 50 Units AMV reverse transcriptase (Boehringer Mannheim) in a final volume of 50 µl. The mixture was incubated at 42^{O}C for 2 hours. After the first strand cDNA synthesis, 450 µl of 10 mM Tris-HCl, pH 7.4/ 1 mM EDTA (TE) was added and the cDNAs were stored at -20^{O}C.

Oligodeoxynucleotide Synthesis:

All oligodeoxynucleotide primers used for PCR and DNA sequencing were synthesized in-house using a Pharmacia Gene Assembler Plus. After synthesis, all oligonucleotides were desalted (NAP-10 columns, Pharmacia), eluted in 1.0 ml of water, and quantitated by spectrophotometry prior to use.

Gene Amplification by PCR:

Since little was known *a priori* about the genomic sequence of HIV-1_{IbNg}, amplification primers were chosen, based on a comparison of the genomic sequences of known HIV-1 strains, from highly conserved regions of the HIV-1 genome (4). Primer pairs were chosen so that the predicted size of the amplification products would be kept below 4.0 kilobases. Two of the five amplification products contained env sequences analysed in this report. The amplification mixtures contained Taq DNA polymerase buffer (10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, 50 mM KCl and 100 µg/ml gelatin), 2.5 µM of each primer, 200 µM of each dNTP, 5 - 10 µl of diluted first strand cDNA and 2.5 Units of Taq DNA polymerase (Boehringer Mannheim) in a final volume of 50 µl. The cycling parameters of each amplification differed according to the primer pairs used and the sizes of the predicted amplicons. The primer pairs were SK-32 (sense strand primer, 5'-ACC/TGC/CAC/CTG/TAG/TAG-3', nt, no. 4308-4325) and ENV-6661 (anti-sense strand primer,
5'-ACT/AGT/TAG/GTT/GCT/ACT/G-3', nt. no. 6661-6643), which were predicted to produce an amplicon of 2,354 bp in size, and ENV-6588 (sense strand primer, 5'-TAA/CCC/CAC/TCT/GTG/TTA-3', nt. no. 6588-6605) and SK-69 (anti-sense strand primer,
5'-CCA/GAC/TGT/GAG/TTG/CAA/CAG-3',
nt. no 7936-7916), which were predicted to produce an amplicon of 1,349 bp in size. The numbering system used was according to tha of the HIV-1_{HXB2} strain (GenBank acc. no. K03455). The conditions used for amplifications consisted of either 30 cycles of 94^{O}C for 1 min, 42^{O}C for 2 min and 72^{O}C for 4 min (SK-32 and ENV-6662 primers) or 30 cycles of 94^{O}C for 1 min, 48^{O}C for 2 min, 72^{O}C for 2.5 min (ENV-6588 and SK-69). Samples were subjected to 72^{O}C for 10 min to ensure complete extensions of all amplicons.

Construction of Plasmids:

Each amplicon was subjected to agarose gel electrophoresis and appropriately sized amplification products were isolated, purified and inserted into pBlueScriptII(SK)+ (Stratagene Cloning Systems, La Jolla, CA). The resultant

constructs were named according to the regions of the HIV-1_{IbNg} genome contained in their respective inserts, e.g. pIbNg pol/env, and pIbNg env.

DNA Sequencing:

Supercoiled, plasmid DNA was sequenced by the dideoxy chain termination method in which products were labelled with (α-^{35}S)-dATP. Sequence Version 2.0 (United States Biochemical) was used for all sequencing reactions. The protocol used for primer/template annealing was a modification of that recommended by the manufacturer (United States Biochemical, 1993). Briefly, plasmid DNA was purified prior to sequencing by spin column chromatography (CHROMA SPIN-400, Clontech Laboratories, Inc., Palo Alto, CA). Approximately, 7-10 µg of plasmid DNA was denatured at 37^{o}C for 30 min in 0.2 M NaOH/200 µM EDTA. The denatured plasmid DNA was then placed on ice. The mixture was neutralized with 0.1 volume of 3M sodium acetate, pH 5.2, and the DNA was precipitated in 3 volumes of absolute ethanol. DNA was recovered by centrifugation after incubation at -80^{o}C for 30 min and pelleted DNA was washed once with 70% ethanol. For sequencing, the DNA pellet was resuspended in 2 µl 5X Sequenase buffer, 7 µl water and 1 µl of 1 pmole/µl primer. The nucleic acid mixture was incubated at 65^{o}C for 2 min and then immediately transferred to 37^{o}C for 30 min to allow the primer to anneal to the template. Samples were placed on ice and the manufacturer's protocol was followed for the sequencing reactions (8). The reaction products were separated under denaturing conditions using gels of 6% acrylamide (19:1) and 7M urea. A glycerol tolerant buffer was used in all gels and running buffers (United State Biochemical, 1993). Sequencing gels were fixed in 15% methanol/5% glacial acetic acid and dried under vacuum at 80^{o}C. Finally, Kodak XAR-5 film was exposed overnight at room temperature to the dried gels.

Sequence Analysis:

GELENTER, GELMERGE and GELASSEMBLE (Sequence Analysis Software Packages, version 7.3, Genetics Computer Group, Inc., Madison, WI) were used to maintain and assemble the raw sequencing data into a final HIV-1_{IbNg} env consensus sequence (9). PILEUP (10) (Sequence Analysis Software Package, version 7.3, Genetics Computer Group, Inc., Madison, WI) was used to align the HIV-1_{IbNg} env consensus sequence at both the genomic DNA and the deduced amino acid levels with the envelope encoding regions of other known strains of HIV-1 that cluster within the A clade. These include HIV-1_{SF1703} (GenBank acc. no. M66533) (11), HIV-1_{U455} (GenBank acc. no. M62320) (12), HIV-1_{Z321} (GenBank acc. no. M15896) (13), and HIV-1_{Ug06} (GenBank acc. no. M98503) (14).

Phylogenetic Tree Construction:

The HIV-1 phylogenetic tree was created using the maximum parsimony algorithm PAUP, Phylogenetic Analysis Using Parsimony, version 2.4.2 (David L. Swofford), at the Los Alamos National Laboratory, Los Alamos, NM. The consensus nucleotide sequence of the envelope open reading frame of HIV-1_{IbNg} along with the following HIV-1 and SIV strains were used to create this tree. HIV-1_{SF1703} (GenBank acc. no M66533), HIV-1_{U455} (GenBank acc. no M62320), (12), HIV-1_{Z321} (GenBank acc. no M15896), (13), HIV-1_{JRCSF} (GenBank acc. no M38429), HIV-1_{SF2} (GenBank acc. no L07422), HIV-1_{MN} (GenBank acc. no M17449), HIV-1_{JY1} (GenBank acc. no J03653), HIV-1_{ELI} (GenBank acc. no K03454), HIV-1_{Z2Z6} (GenBank acc. no M22639), HIV-1_{TN243} (GenBank acc. no L03703), HIV-1_{NOF} (GenBank acc. no L07426), HIV-1_{U747}, HIV-1_{AN170} GenBank acc. no M31171), (15), HIV-$1_{MVP6180}$, and $SIV_{CFZGAB.}$ Specifically, the coding region of the env gene which spanned from the beginning of the V1 domain of gp120 down to the first 76 amino acids of gp41 were used in the original alignment of the nucleotide sequences.

RESULTS

Nucleotide and deduced amino acid sequence of the HIV-1_{IbNg} gp120 protein:

Fig. 1 shows the nucleotide sequence (1,506 base pairs) of the gp120 coding region of the env gene of HIV-1_{IbNg} along with the predicted amino acid sequence (502 amino acids) of its gp120 protein. The gp120 coding region of HIV-1_{IbNg} was contained in two different amplicons which overlapped in sequence. The region encoding the 5' end of gp120 was contained within a 2.3 kb amplification product which was used to construct the plasmid pIbNg pol/env. The region encoding the 3' end of gp120 as well as the 5' end of gp41 was contained within a 1.4 kb amplification product which was used to construct the plasmid pIbNg env. The nucleotide sequence is a consensus listing the most frequent nucleotide found at each position which was derived from the env sequences of 8 clones of pIbNg pol/env and 6 clones of pIbNg env. Since both G and T nucleotide residues were found with equal frequency at nucleotide position no. 394 of this gene, no amino acid residue could be assigned to position no. 132 of the deduced gp120 protein. However, either a valine or a phenylalanine residue would be predicted to be located at this position.

Identification of HIV-1_{IbNg} clade:

The predicted amino acid sequence of the V3 loop region of the HIV-1_{IbNg} gp120 protein was aligned with the concensus amino acid sequences associated with each of five HIV-1 subtypes (A, B, C, D and E) (4). This alighment was based on the positions of conserved cycteine residues of the respective gp120 proteins and the highly conserved GPGQ motif located at the tip of their V3 loops (Fig. 2). A comparison of the V3 loops in this manner revealed that HIV-1_{IbNg} clustered most closely with the A clade of viruses, a group geographically associated with West and Central Africa. Alignment of the 35 amino acid residues comprising the V3 loop of HIV-1_{IbNg} gp120 protein and the A clade consensus gp120 proteins indicated that a glycine residue at position # 11 of the consensus is substituted by a serine residue in HIV-1_{IbNg}. Comparison of amino acid residues surrounding the V3 loop regions of these viruses revealed significant differences which placed HIV-1_{IbNg} in a distant category from other HIV-1 clades. Alignment of the V3 loops of HIV-1_{IbNg} and the other four clades (B, C, D and E) of HIV-1 indicated differences of four or more amino acid residues. Moreover, regions surrounding the V3 loops of these viruses showed more differences than similarities.

To analyze in more detail the clustering of HIV-1_{IbNg} with the clade A of viruses, a phylogenetic tree based on the nucleic acid sequences of the envelope coding regions of different HIV-1 strains was constructed. In this analysis, the nucleic acid sequences of the entire HIV-1_{IbNg} gp120 coding region, the "test" sequence, was aligned with well characterized "reference" or "probe" sequences representing each of the five clades. The results of this analysis also indicated that HIV-1_{IbNg} clustered most closely with the A clade of viruses (Fig. 3).

Comparison of the deduced HIV-1_{IbNg} gp120 amino acid sequences with other members of the HIV-1 A clade:

To analyze the similarities and differences between the gp120 proteins of HIV-1_{IbNg} and other HIV-1 strains belonging to the A clade, the predicted amino acid residues of these proteins were aligned. Results of this alignment are shown in Fig. 4. This diagram shows the five hypervariable regions (V1 - V5) of the predicted gp120 proteins of each strain as well as the invariant 19 cysteine residues and the positions of potential N-glycosylation sites (Asn-Xaa-(Ser or Thr) of these proteins. The majority of amino acid residue differences between the predicted gp120 proteins of these strains was located within or near the five hypervariable regions (V1 - V5, especially the V4 region. In addition, the sequences of the N-termini of these proteins were also fairly diverse. However, the number and positions of the cysteine residues, as well as the number, 21 - 24, and positions of the potential N-glycosylation sites (Asn-Xaa-(Ser or Thr)), were highly conserved among these five strains.

```
    M  R  V  M  G  I  Q  K  N  Y  P  L  L  W  R  W  G  T  N  I
1   ATG AGA GTG ATG GGG ATA CAG AAG AAT TaT CCA CTC TTA TGG AGA TGG GGT ACG AAT ATA

    F  W  I  M  I  I  C  N  A  E  Q  L  W  V  T  V  Y  Y  G  V
61  TTT TGG ATA ATG ATA ATT TGT AAT GCT GAA CAA TTG TGG GTC ACG GTC TAC TAT GGGGTA

    P  V  W  K  T  A  E  T  T  L  F  C  A  S  D  A  K  A  Y  D
121 CCT GTG TGG AAG ACG GCA GAG ACC ACC CTA TTT TGT GCA TCA GAT GTC AAA GCA TAT GAT

    T  E  V  H  N  V  W  A  T  H  A  C  V  P  T  D  P  N  P  Q
181 ACA GAA GTA CAT AAT GTC TGG GCT ACA CAT GCC TGT GTA CCC ACA GAC CCT AAC CCA CAA

    E  I  H  L  E  N  V  T  E  K  F  N  M  W  K  N  N  M  V  E
241 GAA ATA CAT TTG GAa AAT GTA ACA GAA AAG TTT AAC ATG TGG AAA AAT AAC ATG GTA GAA

    Q  M  H  E  D  I  I  S  L  W  D  Q  S  L  K  P  C  V  K  L
301 CAG ATG CAT GAA GAT ATA ATT AGT CTA TGG GAC CAA AGC CTA AAG CCA TGT GTA AAG TTA

    T  P  L  C  V  T  L  D  C  H  N  X  N  N  S  Y  S  N  S  S
361 ACC CCT CTC TGC GTT ACT TTA GAT TGT cAT AAT XTC AAC AAC AGC taC AGC AAC AGT AGC

    N  T  L  S  D  M  N  G  E  I  K  N  D  S  F  N  I  T  T  E
421 AAC CTA ACT AGT GaC ATG AAt GgG GAA ATA AAA AAC TGC TCT TTC AAT ATA ACC ACA GAA

    V  R  D  K  K  K  K  M  H  A  L  F  Y  R  L  D  V  V  Q  I
481 GTA AGA GAT AAG AAA AAG AAG ATG cAT GCA CTT TTT TAT AGA CTT GAT GTA GTA CAA ATT

    N  E  N  N  G  S  Q  Y  R  L  I  N  C  N  T  S  A  I  T  Q
541 AAT gAA AgT AAT GgT AGT CAG TAT AGG TTA ATA AAT TGT AAT ACC TCA GCC ATT ACA CAG

    A  C  P  K  V  S  F  E  P  I  P  I  H  Y  C  A  P  A  G  F
601 GCT TGT CCA AAG GTA TCC TTT GAG CCA ATT CCC ATA CAT TAT TGT GCc CCA GCT GGT TTT

    A  I  L  K  C  K  D  K  G  F  N  G  T  G  P  C  K  N  V  S
661 GCG ATT CTA AAG TGT AAG GAT AAA gGG TTC AAT gGA ACA GGG CCA TGC AAA AAT GTC AGC

    T  V  Q  C  T  H  G  I  K  F  V  V  S  T  Q  L  L  L  N  G
721 ACA GTA CAA TGC ACA CAT GGA ATC AAG CCA GTA GTA TCA ATC CAA CTG CTG TTA AAT GGC

    S  L  A  E  G  E  V  V  I  R  S  E  N  I  T  N  N  A  K  T
781 AgT CTA GCA GAA GGA GAA GTA GTG ATC AGA TCT GAA AAT ATC ACA AAC AAT GCC AAA AcC

    I  I  V  Q  L  A  N  P  V  K  I  N  C  T  R  P  N  N  N  T
841 ATA ATA GTA CAG TTG GCT AAT CCT GTA AAA ATT AAT TGT ACC AGA CCT AAC AAC AAT ACA

    R  K  G  V  H  I  G  P  G  Q  A  F  Y  A  T  G  D  I  I  G
901 AGA AAA GGT GTA CAT ATA GGA CCA GGG CAA GCA TTC TAT GCA ACA GGT GAC ATA ATA GGG

    D  I  R  Q  A  H  C  N  V  S  K  T  E  W  N  K  T  L  H  Q
961 GAT ATA AGA CAA GCA CAT TGT AAT GTC AGT AAA ACA GAA TGG AAT AAA acT Tta cat caG

    V  V  T  Q  L  K  T  Y  F  K  N  T  T  I  I  F  A  N  P  L
1021 gta gTt aca cAa TTA AAG ACG TAC TTT AAG AAC ACC ACA ATA ATC TTT GCT AAC CCC TTA

    G  G  D  V  E  I  T  T  H  S  F  N  C  G  G  E  F  F  Y  C
1081 GGA GGG GAT GTA GAA ATT ACA ACA CAT AGT TTT AAT TGT GGA GGA GAA TTT TTC TAT TGC

    N  T  S  K  L  F  N  S  T  W  D  N  S  N  S  T  A  N  H  T
1141 AAT ACA TCA AAA CTG TTT AAT ACG ACT TgG GaT AAT AGC AAT AGC ACT GCC AAC CAC ACA

    G  S  N  D  T  I  T  L  Q  C  R  I  K  Q  I  V  N  M  W  Q
1201 GGG TCA AAT GAC ACT ATA ACT CTC CAA TGC AGA ATA AAG CAA AAT GTA AAT ATG TGG CAG

    K  V  G  Q  A  M  Y  A  P  P  I  Q  G  I  I  R  C  D  S  N
1261 AAA GTA GGA CAA GCA AtG TAT GCC CCT CCC ATC CAA GGA ATA ATA AGA TGT GAT TCA AAC

    I  T  G  L  L  L  T  R  D  G  G  N  N  S  T  N  E  T  F  R
1321 ATT ACA GGA CTA CTA TTA ACA AGA GAT GGA GGG AAT AAT AGT ACA AAT GAA ACC TTC AGG

    P  G  G  G  D  M  R  D  N  W  R  S  E  L  Y  K  Y  K  V  V
1381 CCT GGA GGA GGA GAT ATG AGG GAC AAT TGG AGA AGT GAA TTA TAT AAG TAT AAA GTA GTA

    K  I  E  P  L  G  V  A  P  T  R  A  K  R  R  V  V  E  R  E
1441 AAA ATT GAA CCA CTA GGT GTA GCA CCC ACC
```

Fig. 1: HIV-1_{IbNg} genomic nucleotide sequences comprising the coding region for the gp120 protein. The first 1,506 nucleotide sequences encoding the HIV-1_{IbNg} gp120 protein is shown along with the predicted amino acid sequence for this protein. Consensus nucleotide sequence of the HIV-1_{IbNg} gp120 coding region (bottom). Capital letters represent total agreement between the various clones at that nucleotide position; small letters represent the predominant nucleotide at that position between the clones. Also shown is the deduced amino acid sequence, based on the consensus nucleotide sequence of the gp120 protein (top).

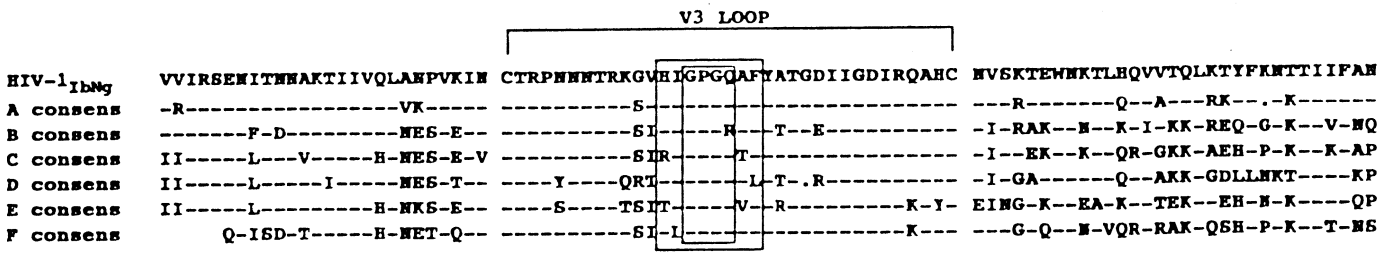


Fig. 2: Alignment of deduced amino acid sequences of the HIV-1_{IbNg} gp120 V3 loop and the surrounding region (top) compared to the consensus sequence of each of the five operationally defined HIV-1 clades. A dash (-) represents concurrence with the HIV-1_{IbNg} sequence while a period (.) indicates a deletion. Brackets along with the offset indicate the position of the V3 loop region. The inner and outer boxes indicate the tetrapeptides and octapeptides, respectively, believed to be located at the tip of the V3 loop of the gp120 protein.

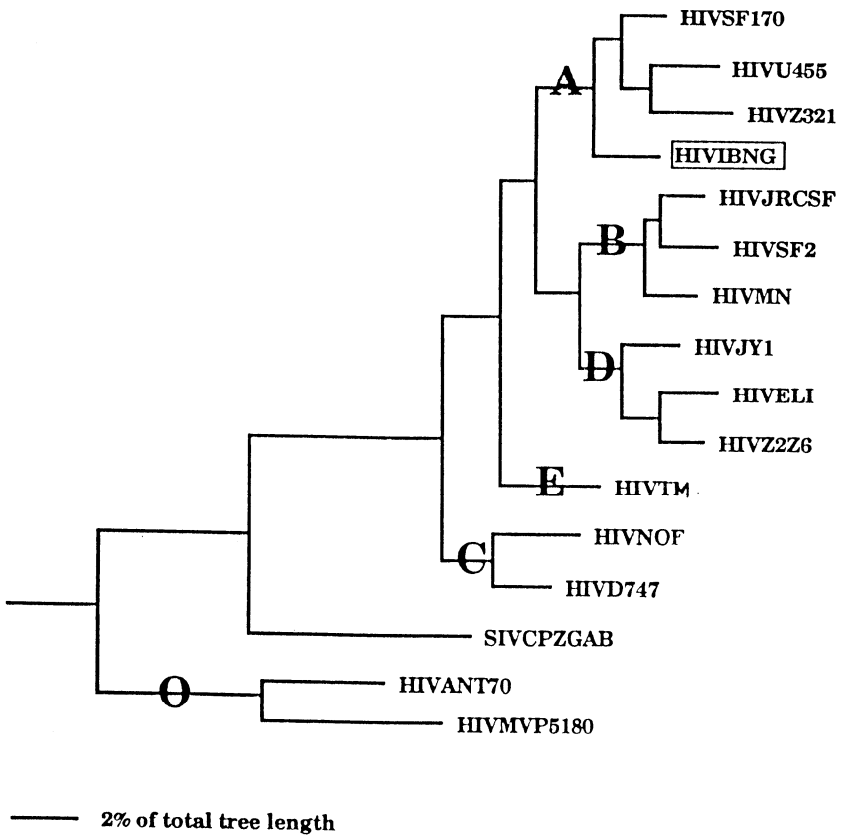


Fig. 3: Diagram of HIV-1 envelope tree. HIV-1_{IbNg} clusters with the clade A strains of HIV-1. Shown are the five operationally defined clades of HIV-1, labelled A - E, along with the names of the HIV-1 and SIV strains used in this analysis. The portion of the envelope gene used to construct this tree spanned the region from the beginning of the V1 lop down to the first 76 amino acids of gp41.

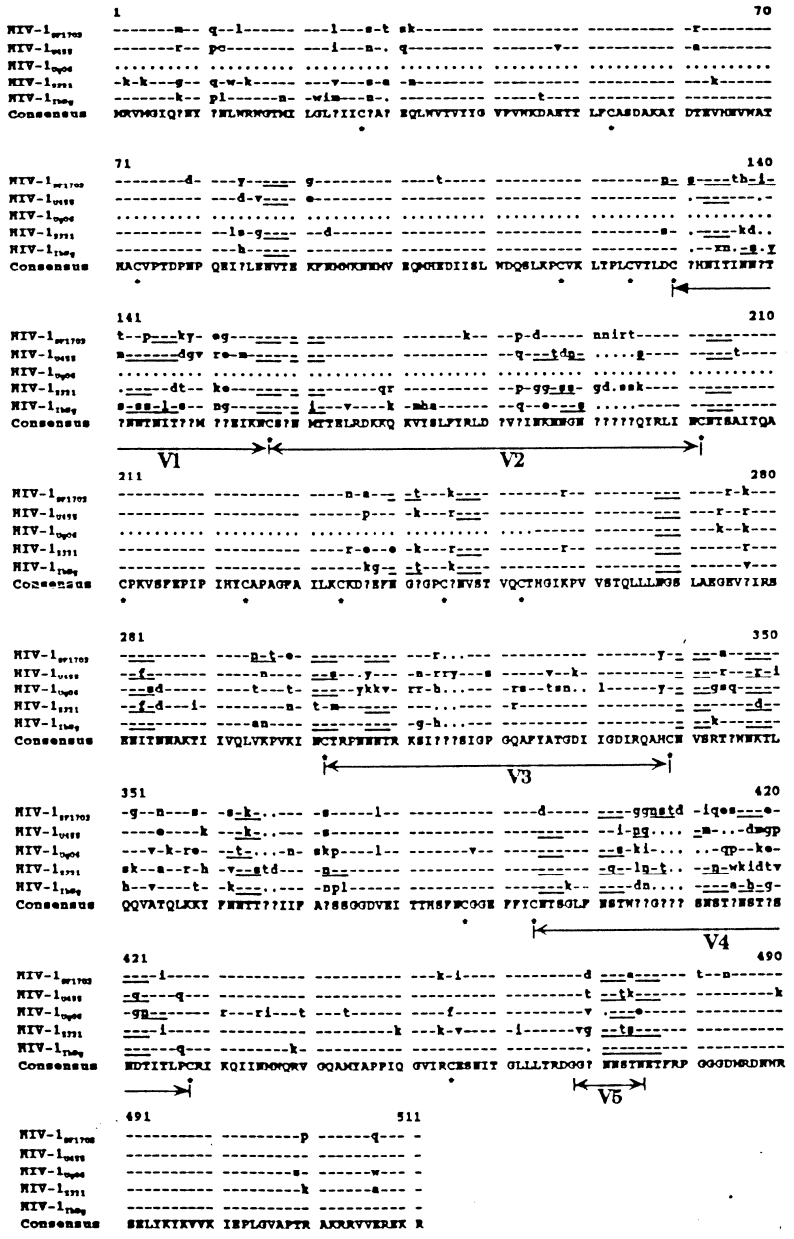


Fig. 4: Alignment of the deduced amino acid sequence of the HIV-1$_{IbNg}$ gp120 protein with other HIV-1 clade A strains. This alignment was constructed using the PILEUP program of the Genetics Computer Group package. The gp120 consensus sequence of the clade A strains, shown at the bottom of the diagram, was constructed from the five strains of HIV-1 identified at the left side of the diagram. A dash (-) represents concurrence with the clade A consensus sequence while a period (.) indicates a deletion. The invariant cysteine residues and potential N-limked glycosylation sites are denoted either with an asterisk (*) or are underscored (_), respectively. The positions of the five hypervariable regions (V1 - V5) within the gp120 protein as well as the putative CD4 protein-binding domain are also shown.

DISCUSSION

We have obtained the complete nucleotide sequence of the surface or envelope protein (gp120) encoding region of a new strain of HIV-1 (HIV-1_{IbNg}) from Nigeria. This represents the first HIV-1 isolate from this region of Africa. Based on the nucleic acid sequence of the gp120 encoding region, we determined that this virus clusters with the clade A strains of HIV-1, which have been primarily isolated from regions of West and Central Africa (4). This conclusion was also corroborated when the deduced amino acid sequences of the third hyper-variable regions, V3, of different HIV-1 gp120 proteins were analyzed.

The gp120 protein of HIV-1 has been shown to be an important determinant of viral pathogenicity. Our attention was again drawn to the gp120 gene of HIV-1_{IbNg} because of the unique biologic characteristics of this strain (7) The person, from whom this virus was isolated, was asymptomatic and apparently healthy at the time of blood donation. However, he went on to develop AIDS within 6 months of the initial blood donation and died after 4 months. In the previous paper (7), we have shown that HIV-1_{IbNg} replicates well and is cytopathic in established cell lines such as the IL-2 dependent T-cell line CEM,, the IL-2-independent T-cell line H9, and the monocytic cell line U937, as well as primary PHA-stimulated PBMC and freshly isolated monocyte/macrophage. In general, HIV-1 isolates are restricted in their ability to infect macrophage or T lymphoid cell lines (16-21). Strains of HIV-1 capable of infecting macrophages are unable to infect lymphoid cell lines and *vice versa* (22). Further, HIV-1 isolates that have been recovered from patients with advanced HIV disease have been shown to be more cytopathic, display a wider host range *in vitro*, and replicate with faster kinetics than isolates obtained from the same individual in a prior healthy state (23,24). Although the functional significance of HIV-1_{IbNg} envelope sequences is unclear, it is possible that the broad host range and the ability of HIV-1_{IbNg} to infect macrophage, T- and monocytoid cell lines is associated with conformational changes in the gp120 proteins of this virus.

Much attention has been focused on analyzing the third hypervariable region of the gp120 protein because of its direct or indirect involvement in immunogenicity and/or pathogenesis. This region has been shown to be a determinant of cell tropism (17, 20, 21, 25), and is required for viral infectivity (26) and viral fusion (27). In addition, epitopes for cytotoxic T-cells (28,29) and helper T-cells (29,30) as well as the principle neutralizing domain (PND) to which neutralizing antibodies bind (27, 31-35) have all been shown to be located within this domain.

The V3 loop is a 35 to 37 amino acid residue structure bounded by a pair of disulphide-bonded cysteine residues (36). This loop is thought to form two anti-parallel B turns with a short C-terminal, a helix (37). The amino acid sequence of this loop varies substantially among different strains of HIV-1 (38). However, it should be noted that the residue flanking the V3 loop of strains belonging to the same clade display more sequence heterogeneity than those located within their V3 loops. Analysis of the deduced amino acid sequences of the V3 region of the HIV-1_{IbNg} gp120 protein revealed that it comprised 35 amino acid residues with only one change from the published consensus sequence of clade A viruses (38). This single amino acid difference, which may be associated with conformational changes in this protein, is a conservative substitution from a serine residue to a glycine. It is possible, therefore, that some of the biological properties of HIV-1_{IbNg} are influenced by the conformational changes of the V3 loop of its gp120 proteins.

Situated near the centre of the gp120 V3 loop of HIV-1_{IbNg} is the tetrapeptide GPGQ. The two most common peptide motifs found at this position are GPGQ, which is the most common motif found in HIV-1 clades A, C, D and E, and GPGR which is the most common motif of the B clade strains (37). Both of these motifs are believed to form a B-turn between two B-sheet side regions (33). This region of the loop may have a role in viral fusion

(26), and it may contain a functional protease cleavage site (39,40).

Analysis of the deduced amino acid sequences of the gp120 proteins of other A clade strains revealed the strict conservation of 19 cysteine residues. Each of the five hyperactive regions, which are all believed to exist as looped structures, are flanked by these residues. Most of these cysteines have been shown to be involved in intra-chain bonding and are therefore assumed to be involved in maintaining the tertiary conformation necessary for the proper functioning of the HIV-1 surface proteins (36).

The role of HIV-1 envelope protein modification by N-linked glycosylation in the pathogenicity of HIV-1 has not been exhaustively addressed. However, the virulence of influenza A virus has ben shown to be affected by the presence or absence of just one N-linked glycosylation site in its haemagglutinin protein (38). The HIV-1_{IbNg} gp120 protein has 26 potential N-linked glycosylation sites. The proteolytic precursor of the gp120 protein (gp160) of HIV-1_{RF} is predicted to be one of the 10 most extensively glycosylated proteins when approximately 18,000 protein sequences in the SWISS-PROT library were analyzed for the number of potential N-linked glycosylation sites (38). A comparison of the potential N-linked glycosylation sites, also known as sequons (38,40), of the deduced HIV-1_{IbNg} gp120 protein with other gp120 proteins of HIV-1 A clade strains revealed that both the number of potential sites and their relative positions are fairly conserved within the clade. A majority of these sites cluster within or immediately surrounding the five hypervariable regions, especially the V1, V4 and V5 domains. It has been conjectured that glycosylation may play a role in the stabilization of the disulphide bonds of these structures (38). Lastly, the host cell in which HIV is replicating may also play a role in the extent to which the gp120 protein is glycosylated as well as the carbohydrate side-chain composition of this molecule (epigenetical variability).

Complete genomic sequences of HIV-1 strains belonging to clade A are unknown. Characterization of the entire genomic sequence of HIV-1_{IbNg} would, therefore, be important in identifying molecular structure(s) of the viral gene(s) involved in the pathogenesis of subtype A strains. These studies would also lead to a better understanding of strategies for the development of vaccines and other interventions that can then be used in regions of the world where HIV-1 strains of the A subtype are endemic, such as West and Central Africa.

ACKNOWLEDGEMENTS: We would like to thank Gerald Myers and Kersti MacInnes of the Los Alamos National Laboratory, Los Alamos, NM, for their assistance in generating the HIV envelope phylogenetic tree, and Zhijuan Sheng of the Laboratory of Viral Oncology and AIDS Research, University of Southern California, Los Angeles, CA, for growing virus infected cells. This research was partially supported by NIAID-sponsored postdoctoral fellowship to T.M.H (T32-A107383) and the Fogarty International Fellowship of the National Institutes of Health, U.S.A. to D.O.O. (TW-04439) to work at the Laboratory of Viral Oncology and AIDS Research at U.S.C.

REFERENCES

1. Myers, G. (1993) Assimilating HIV sequences. AIDS Res. Hum. Retroviruses 9, 697 - 702.
2. Louwagie, J. F. M.; Peeters, M. TP. B.; Sanders-Buell, E.; Eddy, G.; van der Groen, G.; Fransen, K.; Gershy-Damet, G.; DeLeys, R. and Burke, D. (1993) Comparison of gag genes from seventy international HIV-1 isolates provides evidence for multiple genetic subtypes. AIDS 7, 769 - 780.
3. Orlof, G.; Kalish, M.; Chiphangwi, J.; Potts, K.; Ou, C. Y.; Schochetman, G.; Dallabetta, G.; Saah, A. and Miotti, P. (1993) V3 loops of HIV-1 specimens from pregnant women in Malawi uniformly lack a potential N-linked glycosylation site. AIDS Res. Hum. Retroviruses 9, 705 - 706.
4. Myers, G.; Korber, B.; Berzofsky, J.; Smith, R. and Pavlakis, G. (1992) "Human Retroviruses and AIDS 1992: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences", Los Alamos National Laboratory, Los Alamos, NM.
5. Olaleye, O.; Bernstein, L. Ekweozor, C.; Zhijuan, S.; Omilabu, S.; Li, X.; Sullivan, H. J. and Rasheed, S. (1993) Prevalence of

human immunodeficiency virus types 1 and 2 infections in Nigeria. J. Infect. Dis. 167, 710 - 714.
6. Olaleye, D. O.; Ekweozor, C. C.; Zheng, Z. and Rasheed, S. (1994) Evidence of serological cross-reactivities with human immunodeficiency virus types 1 and 2 and human T-lymphotropic virus I and II in sera of pregnant women in Ibadan, Nigeria. Int. J. Epidem. (in press).
7. Olaleye, D. O.; Zheng, Z.; Howard, T. and Rasheed, S. (1995) Isolation and characterization of a new immunodeficiency virus type 1 strain from Nigeria. Biokemistri 5(1),
8. United States Biochemical (1993) Step-by-step protocols for DNA sequencing with sequenase version 2.0 T7 DNA polymerase, 7th Edition.
9. Staden, R. (1980) A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Res. 8, 3673 - 3694.
10. Feng, D. F. and Doolittle, R. (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. 25, 351 - 360.
11. Cheng-Mayer, C.; Homsy, J.; Evans, L. and Levy, J. (1988) Identification of human immunodeficiency virus subtypes with distinct patterns of sensitivity to serum neutralization. Proc. Natl. Acad. Sci. (USA) 85, 2815 - 2819.
12. Oram, J.; Downing, R.; Roff, M.; Clegg, J.; Serwadda, D. and Carswell, J. (1990) Nucleotide sequence of a Ugandan HIV-1 provirus reveals genetic diversity from other HIV-1 isolates. AIDS Res. Hum. Retroviruses 6, 1073 - 1078.
13. Srivivasan, A.; York, D.; Butler, D. (Jr.); Jannoun-Nasr, R.; Getchell, J.; McCormick, J.; Ou, C. Y.; Myers, G.; Smith, T.; Chen, E.; Flaggs, G.; Berman, P.; Schochetman, G. and Kalyanaraman, S. (1989) Molecular characterization of HIV-1 isolated from a serum collected in 1976: Nucleotide sequence comparison to recent isolates and generation of hybrid HIV. AIDS Res. Hum. Retroviruses 5, 121 - 129.
14. Albert, J.; Franzen, L.; Jansson, M.; Scarlatti, G.; Kataaha, P.; Katabira, E.; Mubiro, F.; Rydaker, M.; Rossi, P.; Petterson, U. and Wigzell, H. (1992) Ugandan HIV-1 V3 loop sequences closely related to the US/European consensus. Virology 190, 674 - 681.
15. Deleys, R.; Vanderborght, B.; Haesevelde, M.; Heyndick, L.; van Geel, A.; Wauters, C.; Bernaerts, R.; Saman, E.; Nijs, P.; Willems, B.; Taelman, H.; van der Groen, G.; Piot, P.; Tersmette, T.; Huisman, J. and van Heuverswyn, H. (1990) Isolation and partial characterization of an usual human immunodeficiency retrovirus from two persons of West-Central African origin. J. Virol. 64, 1207 - 1216.
16. Chesebro, B.; Wehrly, K.; Nishio, J. and Perryman, S. (1992) Macrophagetropic human immunodeficiency virus isolates from different patients exhibit unusual V3 sequence homogeneity in comparison with T-celltropic isolates: definition of critical amino acids involved in cell tropism. J. Virol. 66, 6547 - 6554.
17. Hwang, S.; Boyle, T.; Lyerly, H. and Cullen, B. (1991) Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science 253, 71 - 73.
18. O'Brien, W.; Koyanagi, Y.; Namazie, A.; Zhao, J. D.; Diagne, A.; Idler, K.; Zack, J. and Chen, I. (1990) HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. Nature (London) 348, 69 - 73.
19. Schuitmaker, H.; Koostra, N.; deGoede, R.; deWolff, F.; miedema, F. and Tersmette, M. (1991) Monocytotropic human immunodeficiency virus type 1 (HIV-1) variants detectable in all stages of HIV-1 infection lack T-cell line tropism and syncytium-inducing ability in primary T-cell culture. J. Virol. 65, 356 - 363.
20. Shioda, T.; Levy, J. and Cheng-Mayer, C. (1991) Macrophage and T-cell line tropisms of HIV-1 are determined by specific regions of the envelope gp120 gene. Nature (London) 349, 167 - 169.
21. Westervelt, P.; Gendelman, H. and Ratner, L. (1991) Identification of a determinant within the human immunodeficiency virus 1 surface envelope glycoprotein critical for productive infection of primary monocytes. Proc. Natl. Acad. Sci. (USA) 88, 3097 - 3101.
22. Gu, R.; Westervelt, P. and Ratner, L. (1993) Role of HIV-1 envelope V3 loop cleavage in cell tropism. AIDS Res. Hum. Retroviruses 9, 1007 - 1015.
23. Cheng-Mayer, C.; Seto, D.; Tateno, M. and Levy, J. (1988) Biologic features of HIV-1 that correlate with virulence in the host. Science 240, 80 - 82.
24. Cheng-Mayer, C.; Shioda, T. and Levy, J. (1991) Host range, replicative and cytopathic properties of human immunodeficiency virus type 1 are determined by very few amino acid changes in tat and gp120. J. Virol. 65, 6931 - 6941.
25. Ivanoff, L.; Looney, D.; McDanal, C.; Morris, J.; Wong-Staal, F.; Langlois, A.; Petteway, S. and Matthews, T. (1991) Alteration of HIV-1 infectivity and neutralization by a single amino acid replacement in the V3

loop domain. AIDS Res. Hum. Retroviruses 7, 595 - 603.
26. Ivanoff, L.; Dubay, J.; Morris, J.; Roberts, S.; Gutshall, L.; Sternberg, E.; Hunter, E.; Matthews, T. and Petteway, S. (Jr.) (1992) V3 loop region of the HIV-1 gp120 envelope protein is essential for virus infectivity. Virology 187, 423 - 432.
27. Rusche, J.; Javaherian, K.; McDanal, C.; Petro, J.; Lynn, D.; Grimaila, R.; Langlois, A.; Gallo, R.; Arthur, L.; Fischinger, P.; Bolognesi, D.; Putney, S. and Matthews, T. (1988) Antibodies that inhibit fusion of human immunodeficiency virus-infected cells binds a 24-amino acid sequence of the viral envelope, gp120. Proc. Natl. Acad. Sci. (USA) 85, 3198 - 3202.
28. Takahashi, H.; Cohen, J.; Hosmalin, A.; Cease, K.; Houghton, R.; Cornette, J.; DeLisi, C.; Moss, B.; Germain, R. and Berzofsky, J. (1988) An immunodominant epitope of the human immunodeficiency virus envelope glycoprotein gp160 recognized by class I major histocompatibility complex molecule-restricted murine cytotoxic T lymphocytes. Proc. Natl. Acad. Sci. (USA) 85, 3105 - 3109.
29. Takahashi, H.; Merli, S.; Putney, S; Houghton, R.; Cornette, J.; Delisi, C.; Moss, B.; Germain, R. and Berzofsky, J. (1989) A single amino acid interchange yields reciprocal CTL specificities for HIV-1 gp160. Science 246, 118 - 212.
30. Palker, T.; Matthews, T.; Langlois, A.; Tanner, M.; Martin, M.; Scearce, R.; Kim, J.; Berzofsky, J.; Bolognesi, D. and Haynes, B. (1989) Polyvalent human immunodeficiency virus synthetic immunogen comprised of envelope gp120 T helper cell sites and B cell neutralization epitopes. J. Immunol. 142, 3612 - 3619.
31. Goudsmit, J.; Debouck, C.; Meloen, R.; Smit, L.; Bakker, M.; Asher, D.; Wolff, A.; Gibbs, C. and Gajdusek, C. (1988) Human immunodeficiency virus type 1 neutralizing epitope with conserved architecture alicits early type-specific antibodies in experimentally infected chimpanzees. Proc. Natl. Acad. Sci. (USA) 85, 4478 - 4482.
32. Ho, D.; Sarngadharan, M.; Hirsch, M.; Schooley, R.; Rota, T.; Kennedy, R.; Chanh, T. and Sato, V. (1987) Human immunodeficiency virus neutralizing antibodies recognize several conserved domains on the envelope glycoproteins. J. Virol. 61, 2024 - 2028.
33. Javaherian, K.; Langlois, A.; McDanal, C.; Ross, K.; Eckler, L.; Jellis, C.; Profy, A.; Rusche, J.; Bolognesi, D.; Putney, S. and Matthews, T. (1989) Principle neutralizing domains of the human immunodeficiency virus type 1 envelope protein. Proc. Natl. Acad. Sci. (USA) 86, 8768 - 8772.
34. Kenealy, W.; Matthews, T.; Ganfield, M.; Langlois, A.; Waselefsky, D. and Petteway, S. (1989) Antibodies from human immunodeficiency virus-infected individuals bind to a short amino acid sequence that elicits neutralizing antibodies in animals. AIDS Res. Hum. Retroviruses 5, 173 - 182.
35. Palker, R.; Clark, M.; Langlois, A.; Matthews, T.; Weinhold, K.; Randall, R.; Bolognesi, D. and Haynes, B. (1988) Type-specific neutralization of human immunodeficiency virus with antibodies to env-encoded synthetic peptides. Proc. Natl. Acad. Sci. (USA) 85, 1932 - 1936.
36. Leonard, C.; Spellman, M.; Riddle, L.; Harris, R.; Thomas, J. and Gregory, T. (1990) Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type-1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. J. Biol. Chem. 265, 10373 - 10381.
37. LaRosa, G.; Davide, J.; Weinhold, K.; Waterbury, J.; Profy, A.; Lewis, J.; Langlois, A.; Dressman, G.; Boswell, R.; Shadduck, P.; Holley, L.; Karplus, M.; Bolognesi, D.; Matthews, T.; Emini, E. nd Putney, S. (1989) Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. Science 249, 932 - 935.
38. Myers, G. and Lenroot, R. (1992) HIV glycosylation: What does it portend? AIDS Res. Hum. Retroviruses 8, 1459 - 1460.
39. Clements, G.; Price-Jones, M.; Stephens, P.; Sutton, C.; Schiltz, T.; Clapham, P.; McKeating, J.; McClure, M.; Thomson, S.; Marsh, M.; Kay, J.; Weiss, R. and Moore, J. (1991) The V3 loops of the HIV-1 and HIV-2 surface glycoproteins contain proteolytic cleavage sites: a possible function in viral fusion? AIDS Res. Hum. Retroviruses 7, 3 - 16.
40. Hattori, T.; Koito, A.; Takatsuki, K.; Kido, H. and Katunuma, N. (1989) Involvement of tryptase-related cellular protease(s) in human immunodeficiency virus type-1 infection. FEBS Letts. 248, 48 - 52.
41. Myers, G. Korber, B.; Wain-Hobson, S.; Smith, R. and Pavlakis, G. (1993) Human retroviruses and AIDS 1993: A compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, NM.