Office de la propriété intellectuelle du Canada
n organisme
..ndustrie Canada
www.opic.gc.ca

Canadian Intellectual Property Office
An Agency of
Industry Canada
www.cipo.gc.ca

(11) (C) **1,341,482**

(21) 617,159

(22) 1985/10/28

(45) 2005/05/10

(52) 195-1.21

(51) Int.Cl. ^{7} C12N 15/10; C07H 21/04

(19) (CA) **CANADIAN PATENT** (12)

(54) Process for Preparing Fragments of Aids-Associated Retroviruses

(72) LUCIW, PAUL A. U.S.A.
DINA, DINO U.S.A.
STEIMER, KATHELYN U.S.A.
PESCADOR, RAY SANCHEZ U.S.A.
GEORGE-NASCIMENTO, CARLOS U.S.A.
PARKES, DEBORAH U.S.A.
HALLEWELL, ROB U.S.A.
BARR, PHILIP J. U.S.A.
TRUETT, MARTHA U.S.A.

(73) CHIRON CORPORATION U.S.A.

(30) (US) U.S.A. (667,501) 1984/10/31
(US) U.S.A. (696,534) 1985/01/30
(US) U.S.A. (773,447) 1985/09/06

(57) 6 Claims

(62) 494,006 1985/10/28

611,159

Abstract

Recombinant hTLR proteins are made using expression vectors containing DNA constructs derived from ARV-2 nucleic acid. These proteins are useful in serological immunoassays to detect antibodies to hTLR.

A

This application is a divisional application of Canadian Patent Application 494,006 filed on 28 October 1985.

Technical Field

This invention is in the field of genetic engineering. More particularly, it relates to recombinant viral proteins associated with lymphadenopathy syndrome and/or acquired immune deficiency syndrome.

Background Art

With the discovery of human T-cell lymphotropic virus-1 (HTLV-I) as an infectious agent in humans, it was established that retroviruses could infect humans and could be the etiological agent of disease. After HTLV-I was established, a second retrovirus of the same family, HTLV-II was found in a hairy cell leukemia established strain. Since that time, other human retroviruses have been isolated which are associated with lymphadenopathy syndrome (LAS) and/or acquired immune deficiency syndrome (AIDS) victims. Various retroviruses have been isolated from individuals with AIDS (sometimes called HTLV-III) or LAS (sometimes called LAV). See for example, Barre-Sinoussi, et al, Science (1983) 220:868-871 and Montagnier, et al, Cold Spring Harbor Symposium (1984), Vilmer, et al, Lancet (1984) 1:753; Popovic, et al, Science (1984) 224:497 and Gallo, et al, Science (1984) 224:500. A comparison of HTLV-III and LAV may be found in Feorino, et al, (1984), supra. See also, Klatzman, et al, Science (1984) 225:59-62; Montagnier, et al, ibid (1984) 63-66, and the references cited therein for a survey of the field. A general discussion of the T-cell leukemia viruses may be found in Marx, Science (1984) 224:475-477; Levy, et al, Science (1984) 225:840-842 report the isolation of ARV (AIDS-associated retroviruses), now commonly referred to as HIV-I retrovirus.

For the purposes of this application, these viruses (HTLV-III, LAV, and ARV) will be generically referred to as human T-cell lymphotropic retrovirus (hTLR). The hTLRs may be shown to be of the same class by being similar in their morphology, serology, reverse transcriptase optima and cytopathology, as identified in the above references. For example, the reverse transcriptase prefers Mg^{+2}, and has a pH optima of about 7.8.

Disclosure of the Invention

hTLR DNA sequences that encode hTLR proteins, recombinant expression vectors containing such sequences, recombinant hTLR polypeptides, and diagnostic assay using such polypeptides are provided.

Specific aspects of the invention of the parent application include:

A recombinant DNA construct comprising a replication system recognized by a unicellular microorganism and a DNA sequence of at least about 21 bP having an open reading frame characterized in that the sequence includes a sequence substantially complementary to a sequence found in the gag, env, or pol region of proviral hTLR DNA, and coding for a polypeptide which is immunologically non-cross-reactive with HTLV-I and HTLV-II, and reactive with hTLR.

A method for cloning DNA specific for hTLR, characterised in that a unicellular microorganism containing the above-described DNA construct is grown, whereby said DNA sequence is replicated.

A method for producing an expression product of hTLR which is characterized by the following steps:

(a) transforming a unicellular microorganism host with a DNA construct having transcriptional and translational initiation and termination regulatory signals functional in said host and a DNA sequence of at least 21 bp having an open reading frame and including a sequence substantially complementary to a sequence found in gag, env, or pol region of proviral hTLR DNA and coding for a polypeptide that is immunologically non-cross-reactive with HTLV-I and HTLV-II and reactive with hTLR and under the regulatory control of said signals; and

(b) growing said host in a nutrient medium, whereby said expression product is produced; and

(c) isolating the expression product from the host or culture medium.

In an immunoassay for detecting antibodies to hTLR in a sample suspected of containing the hTLR antibodies the characterizing step of using at least one antigenic recombinant hTLR or ARV-2 gag, pol, or env polypeptide to bind to the antibodies wherein the antigenic recombinant hTLR polypeptide is immunologically non-cross reactive with HTLV-I and HTLV-II.

A recombinant polypeptide characterized in that the recombinant polypeptide is:

(a) ARV-2 p16 gag;

(b) ARV-2 p25 gag;

(c) ARV-2 env;

(d) ARV-2 p31 pol;

(e) a fusion protein of ARV-2 p16 gag and ARV-2 p25 gag; or

(f) a fusion protein of superoxide dismutase and ARV-2 p31 pol.

An article of manufacture for use in an immunoassay for hTLR antibodies characterized in that the article comprises at least one of the above-described recombinant polypeptides bound to a solid support.

Brief Description of the Drawings

Figure 1 is a restriction map of proviral DNA (ARV-2).

Figure 2 is the nucleotide sequence of ARV-2(9B). The amino acid sequences for the products of the gag, pol, and env genes are indicated. The U3, R, and U5 regions of the LTRs are also designated. The cap site is position +1. A 3 bp inverted repeat at the ends of the LTR, the TATA box at position -29, the sequence complementary to the 3'-end of the $tRNA^{lys}$ at position 183, and the polyadenylation signal at position 9174 are underlined. The overlines indicate the amino sequences determined from virion proteins. The nucleotides at the beginning of each line are numbered, and the amino acids at the end of each line are indicated.

Figure 3 is a flow diagram showing the procedures for making the plasmid pGAG25-10.

Figure 4 is the nucleotide sequence of the p25 gag gene cloned in plasmid pGAG25-10 and the amino acid sequence encoded by that gene.

Figure 5 is the coding strand of the nucleotide sequence cloned in pGAG41-10 for producing the fusion protein p41 gag and the corresponding amino acid.

Figure 6 is a nucleotide sequence coding for ARV-2 p16 gag protein that was cloned into plasmid ptac5 to make an expression plasmid for producing p16 gag protein in bacteria.

Figure 7 is a nucleotide sequence that encodes ARV-2 env protein that was used to prepare plasmid pDPC303.

Figure 8 is a nucleotide sequence that encodes ARV-2 p31 protein and is contained in plasmid pTP31.

Modes for Carrying Out the Invention

The invention of the parent application comprises a recombinant DNA construct comprising a replication system recognized by a unicellular microorganism and a DNA sequence of at least about 21 bp having an open reading frame characterized in that the DNA sequence encodes an antigenic HIV-I amino acid sequence which is immunologically non-cross-reactive with HTLV-I and HTLV-II and is reactive with HIV-I, said amino acid sequence comprising a portion of or all of the amino acid sequence of Figure 2.

The invention of this divisional application relates to a process for the preparation of an isolated DNA polynucleotide comprising a fragment of at least 21 bp from the gag or env region of the ARV-2 sequence of Figure 2, wherein the DNA polynucleotide is chemically synthesised at least in part.

The DNA polynucleotide may be labelled, for exmaple, with an isotopic or non-isotopic label.

The invention of this divisional application further relates to an isolated polynucleotide comprising a fragment of at least 21 bp from the gag or env region of the ARV-2 sequence of Figure 2, wherein said polynucleotide is not greater than 180 bp.

The hTLR DNA sequences, either isolated and cloned from proviral DNA or cDNA or synthesized, may be used for expression of polypeptides which may be a precursor protein subject to further manipulation by cleavage, or a complete mature protein or fragment thereof. The smallest sequence of interest, so as to provide a sequence encoding an amino acid sequence capable of specific binding to a receptor, e.g., an immunoglobulin, will be 21 bp, usually at least 45 bp, exclusive of the initiation codon. The sequence may code for any greater portion of or the complete polypeptide, or may include flanking regions of a precursor polypeptide, so as to include portions of sequences or entire sequences coding for two or more different mature polypeptides. The sequence will usually be less than about 5 kbp, more usually less than about 3 kbp.

Sequences of particular interest having open reading frames (Figure 5) define the structural genes for the gag proteins (p16 and p25), the env protein, and the pol protein (p31). It is to be understood that the above sequences may be spliced to other sequences present in the retrovirus, so that the 5'-end of the sequence may not code for the N-terminal amino acid of the expression product. The splice site may be at the 5'-terminus of the open reading frame or internal to the open reading frame. The initiation codon for the protein may not be the first codon for methionine, but may be the second or third methionine, so that employing the entire sequence indicated above may result in an extended protein. However, for the gag and env genes there will be proteolytic processing in mammalian cells, which processing may include the removal of extra amino acids.

In isolating the different domains the provirus may be digested with restriction endonucleases, the fragments electrophoresed and fragments having the proper size and duplexing with a probe, when available, are isolated, cloned in a cloning vector, and excised from the vector. The fragments may then be manipulated for expression. Superfluous nucleotides may be removed from one or both termini using Bal31 digestion. By restriction mapping convenient restriction sites may be located external or internal to the coding region. Primer repair or *in vitro* mutagenesis may be employed for defining a terminus, for insertions, deletion, point or multiple mutations, or the like, where codons may be changed, either cryptic or changing the amino acid, restriction sites introduced or removed, or the like. Where the gene has been truncated, the lost nucleotides may be replaced using an adaptor. Adaptors are particularly useful for joining coding regions to ensure the proper reading frame.

The env domain of the hTLR genome can be obtained by digestion of the provirus with *Eco*RI and *Kpn*I and purification of a 3300 base pair (bp) fragment, which fragment contains about 400 bp of 5' non-coding and about 200 bp of 3' non-coding region. Three different methionines coded for by the sequence in the 5' end of the open reading frame may serve as translational initiation sites.

Digestion of proviral sequences with *Sac*I and *Eco*RV provides a fragment of about 2300 bp which contains the gag domain and a second small open reading frame towards the 3' end of the gag region. The gag domain is about 1500 bp and codes for a large precursor protein which is processed to yield proteins of about 25,000 (p25), 16,000 (p16) and 12,000 (p12) daltons. Digestion with SacI and BglII may also be used to obtain exclusively the gag domain with p12, p25 and partial p16 regions.

Digestion of the previous with KpnI and SstI provides a fragment containing the portion of the pol domain that encodes p31.

The polypeptides which are expressed by the above DNA sequences may find use in a variety of ways. The polypeptides or immunologically active fragments thereof, may find use as diagnostic reagents, being used in labeled or unlabeled form or immobilized (i.e., bound to a solid surface), as vaccines, in the production of monoclonal antibodies, e.g., inhibiting antibodies, or the like.

The DNA sequences may be joined with other sequences, such as viruses, e.g., vaccinia virus or adenovirus, to be used for vaccination. Particularly, the DNA sequence of the viral antigen may be inserted into the vaccinia virus at a site where it can be expressed, so as to provide an antigen of hTLR recognized as an immunogen by the host. The gag, pol, or env genes or fragments thereof that encode immunogens could be used.

Another alternative is to join the gag, env, or pol regions or portions thereof to HBsAg gene or pre-S HBsAg gene or immunogenic portions thereof, which portion is capable of forming particles in a unicellular microorganism host, e.g., yeast or mammalian cells. Thus, particles are formed which will present the hTLR immunogen to the host in immunogenic form, when the host is vaccinated with assembled particles.

As vaccines, the various forms of the immunogen can be administered in a variety of ways, orally, parenterally, intravenously, intra-arterially, subcutaneously, or the like. Usually, these will be provided in a physiologically acceptable vehicle, generally distilled water, phosphate-buffered saline, physiological saline, or the like. Various adjuvants may be included, such as aluminum hydroxide, and the dosages, number of times of administration and manner of administration determined empirically.

In order to obtain the hTLR sequence, virus can be pelleted from the supernatant of an infected host cell. A 9 kb RNA species is purified by electrophoresis of the viral RNA in low-melting agarose gels, followed by phenol extraction. The purified RNA may then be used as a template with random primers in a reverse transcriptase reaction. The resulting cDNA is then screened for hybridization to polyA+ RNA from infected and uninfected cells. Hybridization occurring from infected, but not uninfected cells, is related to the hTLR.

Genomic DNA from infected cells can be restriction enzyme digested and used to prepare a bacteriophage library. Based upon restriction analysis of the previously obtained fragments of the retrovirus, the viral genome can be partially digested with EcoRI and 9 kb-15 kb DNA fragments isolated and employed to prepare the library. The resulting recombinant phage may be screened using a double-lift screening method employing the viral cDNA probe, followed by further purification, e.g., plaque-purification and propagation in large liquid cultures. From the library, the complete sequence of the virus can be obtained and detected with the previously described probe.

hLTR DNA (either provirus or cDNA) may be cloned in any convenient vector. Constructs can be prepared, either circular or linear, where the hTLR DNA, either the entire hTLR or fragments thereof, may be ligated to a replication system functional in a microorganism host, either prokaryotic or eukaryotic cells (mammalian, yeast, arthropod, plant). Microorganism hosts include E. coli, B. subtilis, P. aeruqenosa, S. cerevisiae, N. crassa, etc. Replication systems may be derived from ColE1, 2 mμ plasmid, λ, SV40, bovine papilloma virus, or the like, that is, both plasmids and viruses. Besides the replication system and the hTLR DNA, the construct will usually also include one or more markers, which allow for selection of transformed or transfected hosts. Markers may include biocide resistance, e.g., resistance to antibiotics, heavy metals, etc., complementation in an auxotrophic host to provide prototrophy, and the like.

For expression, expression vectors will be employed. For expression in microorganisms, the expression vector may differ from the cloning vector in having transcriptional and translational initiation and termination regulatory signal sequences and may or may not include a replication system which is functional in the expression host. The coding sequence is inserted between the initiation and termination regulatory signals so as to be under their regulatory control. Expression vectors may also include the use of regulatable promoters, e.g., temperature-sensitive or inducible by chemicals, or genes which will allow for integration and amplification of the vector and hTLR DNA such as tk, dhfr, metallothionein, or the like.

The expression vector is introduced into an appropriate host where the regulatory signals are functional in such host. The expression host is grown in an appropriate nutrient medium, whereby the desired polypeptide is produced and isolated from cells or from the medium when the polypeptide is secreted.

Where a host is employed in which the hTLR transcriptional and translational regulatory signals are functional, then the hTLR DNA sequence may be manipulated to provide for expression of the desired polypeptide in proper juxtaposition to the regulatory signals.

The polypeptide products can be obtained in substantially pure form, particularly free of debris from human cells, which debris may include such contaminants as proteins, polysaccharides, lipids, nucleic acids, viruses, bacteria, fungi, etc., and combinations thereof. Generally, the polypeptide products will have less than about 0.1, usually less than about 0.01 weight percent, of contaminating materials from the expression host. Depending upon whether the desired polypeptide is produced in the cytoplasm or secreted, the manner of isolation will vary. Where the product is in the cytoplasm, the cells are harvested, lysed, the product extracted and purified, using solvent extraction, chromatography, gel exclusion, electrophoresis, or the like. Where secreted, the desired product will be extracted from the nutrient medium and purified in accordance with the methods described above.

The expression products of the env, gag, and pol genes and immunogenic fragments thereof having immunogenic sites may be used for screening antisera from patients' blood to determine whether antibodies are present which bind to hTLR antigens. One or more of the recombinant antigens may be used in the serological assay. Preferred modes of the assay employ a combination of gag, env, and pol antigens. A combination of p25, p31 and env recombinant antigens is particularly preferred. A wide variety of immunoassay techniques can be employed, involving labeled or unlabeled antigens or immobilized antigens. The label may be fluorescers, radionuclides, enzymes, chemiluminescers, magnetic particles, enzyme substrates, cofactors or inhibitors, ligands, or the like.

A particularly convenient technique is to bind the antigen to a support such as the surface of an assay tube or well of an assay plate or a strip of material, such as nitrocellulose or nylon, that binds proteins and contact the sample with the immobilized antigen. After washing the support to remove non-specifically bound antisera, labeled antibodies to human Ig are added. The support is then washed again to remove unbound labeled anti-human Ig. The presence of bound analyte is then determined through detection of the label.

ELISA and "dot-blot" assays are particularly useful for screening blood or serum samples for anti-hTLR antibodies. The ELISA assay uses microtiter trays having wells that have been coated with the antigenic hTLR polypeptides(s). The wells are also typically post-coated with a non-antigenic protein to avoid non-specific binding of antibodies in the sample to the well surface. The sample is deposited in the wells and incubated therein for a suitable period under conditions favorable to antigen-antibody binding.

Anti-hTLR antibodies present in the sample will bind to the antigen(s) on the well wall. The sample is then removed and the wells are washed to remove any residual, unbound sample. A reagent containing enzyme labeled antibodies to human immunoglobulin is then deposited in the wells and incubated therein to permit binding between the labeled anti-human Ig antibodies and hTLR antigen-human antibody complexes bound to the well wall. Upon completion of the incubation, the reagent is removed and the wells washed to remove unbound labeled reagent. A substrate reagent is then added to the wells and incubated therein. Enzymatic activity on the substrate is determined visually or spectrophotometrically and is an indication of the presence and amount of anti-hTLR antibody-containing immune complex bound to the well surface.

The "dot-blot" procedure involves using hTLR antigen(s) immobilized on a piece or strip of bibulous support material, such as nitrocellulose filter paper or nylon membrane, rather than antigen-coated microtiter trays. The support will also be treated subsequently with a non-antigenic protein to eliminate non-specific binding of antibody to the support. The antigen-carrying support is dipped into the sample and allowed to incubate therein. Again, any anti-hTLR antibodies in the sample will bind to the antigen(s) immobilized on the support. After a suitable incubation period the support is withdrawn from the sample and dipped repeatedly in wash buffer to remove any unbound sample from the paper. The support is then dipped into the enzyme-labeled antibody to human Ig reagent for a suitable incubation period. Following treatment with the labeled reagent the support is dipped in wash buffer, followed by incubation in the substrate solution. Enzymatic activity, indicating the presence of anti-hTLR antibody-containing complexes on the support, causes color changes on the support which may be detected optically.

Either of these techniques may be modified to employ labels other than enzymes. The reading or detection phases will be altered accordingly.

The antigenic polypeptides of hTLR may also be used as immunogens by themselves or joined to other polypeptides for the production of antisera or monoclonal antibodies which may be used for therapy or diagnosis. The immunoglobulins may be from any mammalian source, e.g., rodent, such as rat or mouse, primate, such as baboon, monkey or human, or the like. For diagnosis, the antibodies can be used in conventional ways to detect hTLR in a clinical sample.

The hTLR DNA sequences may also be labeled with isotopic or non-isotopic labels or markers and be used as DNA probes to detect the presence of native hTLR nucleotide sequences in samples suspected of containing same.

The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL

1. AIDS related virus-2 (ARV-2) purification and preparation of viral RNA.

HUT-78 cells infected with ARV-2 (ATCC Accession No. CRL 8597, deposited on August 7, 1984) were obtained from Dr. Jay Levy, University of California, San Francisco. Cultures were grown for two weeks in RPMI medium with 10% fetal calf serum. Cultures were centrifuged at 2 Krpm for 1 hr at 4°C using a SW-28 rotor. The pellet, containing the virus, was resuspended in 10 mM Tris-HCl, pH 7.5 on ice. The resuspended pellet was treated with 10 µg of DNase

(Boehringer-Mannhein) and was layered onto a linear sucrose gradient (15-50% in 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 20 mM NaCl). The gradient was spun at 34 Krpm for 4 hr at 4°C, in SW-41 rotor. Five 2.5 ml fractions were collected and an aliquot of each was electrophoresed in a 1% agarose, 5 mM methyl mercury hydroxide gel (Bailey and Davidson, Anal Biochem (1976) 70:75-85) to determine which contained the 9 kb viral RNA. The fraction containing the viral RNA was diluted to 10 ml in 10 mM Tris-HCl, pH 7.5, 1 mM EDTA and was centrifuged at 34 Krpm for 2 hr at 4°C. The pellet was resuspended in 20 mM Tris-HCl, pH 7.6, 10 mM EDTA, 0.1% SDS, and 200 µg/ml proteinase K. Incubation was carried out for 15 min at room temperature. The mixture was extracted with phenol and the aqueous phase was made 400 mM NaCl and precipitated with ethanol. The pellet was resuspended in water and stored at -70°C.

To purify the viral RNA from the nucleic acid pellet obtained as described above, a sample was electrophoresed in a low-melting 1% agarose gel containing 5 mM Methyl mercury hydroxide. After electrophoresis, the gel was stained with 0.1% ethidium bromide and nucleic acid bands were visualized under UV light. The region corresponding to 9 kb was cut from the gel and the agarose was melted at 70°C for 2 to 3 min in three volumes of 0.3 M NaCl, 10 mM Tris, pH 7.5, 1 mM EDTA. The mixture was extracted with an equal volume of phenol. The aqueous phase was reextracted with phenol and was precipitated with ethanol. The pellet was washed with cold 95% ethanol, air dried, resuspended in water and stored at -70°C until use. One hundred ml of culture medium yielded 0.5 to 1 µg of purified RNA.

2. Synthesis of labeled homologous viral probe.

A ^{32}P-labeled cDNA was made to the gel purified viral RNA using random primers (calf thymus primers) prepared as described in Maniatis, et al, A Laboratory Manual, Cold Spring Harbor, NY, 1982. The reaction mixture contained 2 µl of 0.5 M $MgCl_2$; 5 µl of 0.1 M dithiothreitol; 2.5 µl each of 10 mM dATP, 10 mM dGTP and 10 mM dTTP; 2.5 µl calf thymus primer ($100A_{260}$/ml); 0.5 µg viral RNA; 5 µl of actinomycin D (200 µg/ml); 10 µl of ^{32}P-dCTP (> 3000 Ci/mmole, 1 mCi/ml) and 1 µl of AMV reverse transcriptase (17 units/µl) in a 50 µl reaction volume. The reaction was incubated for 1 hr at 37°C. The probe was purified away from free nucleotides by gel filtration using a Sephadex® G50 column. The void volume was pooled, NaCl was added to a final concentration of 400 mM and carrier single-stranded DNA to 100 µg/ml, and the cDNA was precipitated with ethanol. The pellet was resuspended in water and incorporated ^{32}P counts were determined.

3. Detection of ARV sequences in polyA+ RNA prepared from infected HUT-78 cells.

PolyA+ RNA was prepared from HUT-78 cells infected with ARV-2, ARV-3 or ARV-4 (three different isolates from three different AIDS patients) and from uninfected HUT-78 cells. The polyA+ RNA was electrophoresed on 1% agarose gels containing 5 mM methyl mercury hydroxide (Bailey and Davidson, supra), was transferred to nitrocellulose filters, and hybridized with the homologous probe prepared as described in Section 2. Hybridizations were carried out in 50% formamide, 3 x SSC at 42°C. Washes were at 50°C in 0.2 x SSC. A 9 kbp band was present in all three samples of infected HUT-78 cells. This band was absent in polyA+ from uninfected cells.

4. Detection of ARV sequences in infected and non-infected HUT-78 cells.

High molecular weight DNA (chromosomal) was prepared from cultures of HUT-78 cells infected with ARV-2 and from non-infected HUT-78 cells following the procedure of Luciw, et al. Molec and Cell Biol (1984) 4:1260-1269. The DNA was digested with restriction enzyme(s), electrophoresed in 1% agarose gels and blotted onto nitrocellulose following the procedure described by Southern, (1975), supra. Blots were hybridized with the ^{32}P-labeled probe (10^6 cpm/blot) in a mixture containing 50% formamide, 3 x SSC, 10 mM Hepes, pH 7.0, 100 µg/ml denatured carrier DNA, 100 µg/ml yeast RNA and 1 x Denhardt's for 36 hr at 42°C. Filters were washed once at room temperature in 2 x SSC and twice at 42°C in 0.2 x SSC, 0.1% SDS. Filters were air dried and exposed to X-Omat® film using an intensifying screen.

The homologous ^{32}P-probe to ARV-2 hybridized specifically to two bands in the DNA from infected cells restricted with SacI. These bands were absent when DNA of non-infected cells was used, indicating that the probe is hybridizing specifically to infected cells presumably to the provirus integrated in the chromosomal DNA. The molecular weight of the bands is approximately 5 kb and 3 kb.

In order to determine if different enzymes would cut the proviral sequence, several other restriction digestions of the cell DNA were carried out using EcoRI, SphI or KpnI or double digestions using two of them. Southern results show specific bands hybridizing when DNA of infected cells is used. Figure 1 shows a schematic map of the positions of restriction enzyme sites in the proviral sequence, and indicates fragment sites.

5. Cloning of proviral ARV-2 DNA.

High molecular weight cell DNA from infected HUT-78 cells was prepared following the procedure of Luciw, et al, supra. The DNA was digested with *Eco*RI, which cuts once in the provirus, centrifuged in a sucrose gradient and fractions corresponding to 8-15 kb were pooled, dialyzed and concentrated by ethanol precipitation. The bacteriophage λ derivative cloning vector, EMBL-4 (Karn, et al, Methods Enzymol (1983) 101:3-19) was digested to completion with a mixture of *Eco*RI, *Bam*HI and *Sal*I restriction enzymes and the DNA then deproteinized by phenol-chloroform extraction, precipitated with cold ethanol and resuspended in ligation buffer. The EMBL-4 phage DNA and *Eco*RI digest of cellular DNA were mixed and ligated and the resultant recombinant phage genomes packaged *in vitro*. After phage infection of λ-sensitive *E. coli* (DP50supF), about 500,000 phage plaques were transferred onto nitrocellulose filters, DNA was fixed and the filters were screened with a homologous ^{32}P-probe prepared as described in Section 2. Eleven recombinant phage out of 500,000 phage annealed in the initial double-lift screening method (Maniatis, et al, Molecular Cloning, A Laboratory Manual, NY, 1982) to viral cDNA probe, and these were further plaque-purified and propagated in large liquid cultures for preparation of recombinant DNA. Plaque-purified phage containing ARV DNA were propagated in liquid culture in *E. coli* DP50supF; phage particles were harvested and banded in CsCl gradients and recombinant phage DNA was prepared by phenol extraction followed by ethanol precipitation (Maniatis, et al, supra). One µg of purified phage DNA was digested with restriction enzymes, electrophoresed on 1% agarose gels, and visualized with ethidium bromide under ultraviolet light. The DNA from these gels was transferred to nitrocellulose and annealed with viral cDNA probe.

One of the 11 phage, designated λ ARV-2(9B), was deposited at the ATCC on 25 January 1985 and given Accession No. 40158. λ ARV-2(9B) contained an insertion of full-length proviral DNA along with flanking cell sequences. Digestion of λ ARV-2(9B) DNA with SacI yielded viral DNA fragments of 3.8 kb and 5.7 kb. EcoRI digestion of λ ARV-2(9B) produced virus containing DNA species at 6.4 kb and 8.0 kb; a double digest of SacI and EcoRI gave viral DNA fragments at 3.8 kb and 5.4 kb. This pattern is consistent with that of a provirus linked to cell DNA.

In addition to λ ARV-2(9B), phage was obtained that (1) possessed the left half of the viral genome from the EcoRI site in viral DNA extending into flanking cell DNA (λARV-2(8A)) and (2) phage that had the right half of the viral genome (λARV-2(7D)) from the EcoRI site in viral DNA extending into flanking cell DNA. Bacteriophages λARV-2(7D) (right) and λARV-2(8A) (left) were deposited at the ATCC on October 26, 1984 and given Accession Nos. 40143 and 40144, respectively.

6. Polymorphism.

To measure the relatedness of independent ARV isolates, restriction enzyme digests of DNA from HUT-78 cells infected with ARV-3 and ARV-4 were analyzed with the probe made from cloned ARV-2 DNA. The _SacI_ digest of ARV-3 DNA was similar to that of ARV-2 whereas the _HindIII_ digests displayed different patterns. The _SacI_ digest and the _PstI_ digest of ARV-4 DNA differed from the corresponding digests of ARV-2 DNA. The intensity of the annealing signals obtained with ARV-3 and ARV-4 samples was much lower (about 10-fold less) than that for ARV-2 DNA probably as a result of the fact that fewer cells were infected in the ARV-3 and ARV-4 cultures. The viral-specific DNA fragments produced by _SacI_ treatment of ARV-3 and ARV-4 DNA totaled 9.0-9.5 kbp, a value similar to that of ARV-2 and in consonance with the RNA genome sizes.

7. Sequencing of proviral DNA.

Fragments or subfragments of ARV-2 DNA from λ phage 9B were prepared and cloned into M13 according to conventional procedures (Maniatis, et al, supra). Sequencing was performed according to Sanger, et al, _Proc Natl Acad Sci USA_ (1977) _74_:5463, using the universal M13 primer or chemically synthesized primers complementary to ARV-2 sequence. The sequence is shown in Figure 2.

8. Amino acid sequence analysis of p25 and p16 gag coded proteins.

ARV-2 was prepared and purified as described in Section 1. The viral proteins were electrophoresed on an acrylamide gel, and the band corresponding to a 24,000 dalton or 16,000 dalton protein was excised from the gel and used for sequencing. Micro-sequence analysis was performed using Applied Biosystems model 470A protein sequencer similar to that described by Hewick, et al, _J Biol Chem_ (1981) _256_:7990-7997.

Phenylthiohydantoin amino acids were identified by HPLC using a Beckman Ultrasphere® ODS column and a trifluoroacetic acid-acetonitrile buffer system as reported by Hawke, et al, Anal Biochem (1982) 120:302-311. Table 1 shows the first 20 amino acids from the amino terminus determined for p25-gag protein and Table 2 shows the first 30 amino acids for p16-gag protein.

TABLE 1: Amino-terminal sequence of p25-gag

Position	Amino acid
1	Pro
2	Ile
3	Val
4	Gln
5	Asn
6	Leu
7	Gln
8	Gly
9	Gln
10	Met
11	Val
12	(His)
13	Gln
14	Ala
15	Ile
16	(Ser)
17	Pro
18	(Arg, Lys)
19	Thr
20	(Leu)

TABLE 2: Amino-terminal sequence of p16-gag

Position	Amino acid
1	(Met)
2	Gln
3	Arg
4	Gly
5	Asn
6	Phe
7	Arg
8	Asn
9	Gln
10	Arg
11	Lys
12	Thr
13	Val
14	Lys
15	--- (Cys)
16	Phe
17	Asn
18	--- (Cys)
19	Gly
20	Lys
21	Glu
22	Gly
23	(His)
24	Ile
25	Ala
26	(Lys)
27	Asn
28	(Gly)
29	(Arg)
30	(Ala, Leu)

The amino acid sequence of Table 1 is predicted from the ARV-2 DNA sequence of Figure 2. Therefore, these results confirm that the indicated gag open reading frame is in fact being translated and identifies the N-termini of p25 and p16.

9. Expression of p25 gag protein of ARV-2 in bacteria.

A. Host-vector system

The p25 gag protein is synthesized by *E. coli* strain D1210 transformed with plasmid pGAG25-10.

Plasmid pGAG25-10 is a pBR322 derivative which contains the sequence coding for p25 gag under transcriptional control of a hybrid tac promoter (De Boer et al, PNAS (1983), 80:21-25) derived from sequences of the trp and the lac UV5 promoters. Expression of p25 gag is induced in bacterial transformants with isopropylthiogalactoside (IPTG).

E. coli D1210, a lac-repressor overproducing strain, carries the $lacI^q$ and $lacY^+$ alleles on the chromosome but otherwise is identical to *E. coli* HB101 (F^- $lacI^+$, $lacO^+$, $lacZ^+$, $lacY^-$, gal^-, pro^-, leu^-, thi^-, end^-, hsm^-, hsr^-, $recA^-$, $rpsL^-$) from which it was derived.

B. Construction of pGAG25-10.

Plasmid pGAG25-10 was constructed by cloning a 699 bp DNA fragment coding for p25 gag into plasmid ptac5, according to the scheme shown in Figure 3. The vector ptac5 is a pBR322 derivative which contains the tac promoter, Shine Delgarno sequences, and a polylinker as a substitution of the original pBR322 sequences comprised between the EcoRI and PvuII restriction sites.

The 699 bp DNA fragment codes for the complete p25 gag protein (amino acid residues 139 to 369 as numbered in Figure 2), the only difference being that a methionine was added as the first amino acid in pGAG25-10 to allow for translational initiation. This change, as well as other changes in nucleotide sequence as indicated below, was achieved by using chemical synthesis of parts of the DNA fragment. The DNA fragment also includes two stop codons at the 3' end of the sequence.

Figure 4 shows the nucleotide sequence cloned in pGAG25-10 and the amino acid sequence derived from it. DNA sequences that are not underlined in the figure were derived directly from the ARV-2(9B) cDNA. All other sequences were chemically synthesized or derived from vector ptac5. Changes were introduced in this DNA sequence, with respect to the original cDNA, to create or delete restriction sites, to add a methionine prior to the proline (first residue of p25) or to include stop codons after the last codon of p25 gag. However, as previously indicated, all changes in the DNA sequence, except those in the first codon, do not alter the amino acid sequence of p25 gag.

C. Preparation of D1210 (pGAG25-10) strain and characterization of p25 gag protein expressed by transformants.

E. coli D1210 cells are made competent for transformation following a standard protocol (Cohen et al, PNAS (1972) 69:2110). Transformation is performed as indicated in the protocol with 25-50 ng of pGAG25-10. The transformation mix is plated on agar plates made in L-broth containing 100 μg/ml ampicillin. Plates are incubated for 12 hr at 37°C.

Single ampicillin resistant colonies are transferred into 1 ml L-broth containing 100 µg/ml ampicillin and grown at 37°C. Expression of p25 gag protein is induced by adding 10 µl of 100 mM IPTG (Sigma) to a final concentration of 1 mM followed by incubation at 37°C for 2 hr.

Cells from 1 ml of induced cultures are pelleted and resuspended in 100 µl Laemmli sample buffer. After 3 cycles of boiling and freezing, portions of resultant lysates are analyzed on standard denaturing acrylamide gels. Proteins are visualized by staining with Coomassie blue.

The extent of expression is initially determined by appearance of new protein bands for induced candidate samples compared with control. Proteins of molecular weights expected for the genes expressed comprised 2%-5% of total cell protein in the highest expressing recombinants as determined by visual inspection with reference to a standard protein of known amount.

Authenticity of the expressed proteins is determined by standard Western transfer of proteins to nitrocellulose and analysis with appropriate human or rabbit immune sera or mouse monoclonal antibodies (see E.4.a. below) or by ELISA assays of soluble E. coli proteins using human immune sera from AIDS patients (see E.4.b. below).

D. Fermentation process.

D.1. Preparation of transformant master seed stock.

Transformant cells from a culture expressing high levels (3%) of p25 gag are streaked onto an L-broth plate containing 100 µg/ml ampicillin and the plate is incubated overnight at 37°C. A single colony is inoculated into 10 ml of L-broth, 100 µg/ml ampicillin and grown overnight at 37°C. An aliquot is used to verify plasmid structure by restriction mapping with SalI and PstI. A second aliquot is used to induce expression of p25 gag and the rest of the culture is made 15% glycerol by adding 1/4 volume of 75% sterile glycerol. Glycerol cell stocks are aliquoted in 1 ml and quickly frozen in liquid nitrogen or dry-ice ethanol bath. These master seed stocks are stored at -70°C.

D.2. Master plate/single colonies and overnight cultures.

The master seed stock is scraped with a sterile applicator which is used to streak an L-broth plate containing 100 µg/ml ampicillin. Single colonies from this plate are used to inoculate 20-50 ml of L-broth/amp, which is incubated at 37°C overnight.

D.3. Fermentor inoculum.

An aliquot of the overnight culture is used to inoculate larger volumes (1-6 liters) of L-broth/amp. Cells are incubated at 37°C overnight and reach an $O.D._{650}$ of approximately 5 prior to use as inoculum for the fermenter run.

D.4. Fermentation and harvest.

Fermenters (capacity: 16 liters) containing 10 l of L-broth and 1 ml of antifoam are inoculated with 100-500 ml from the inoculum culture. Cells are grown at 37°C to an O.D. of about 1. Expression of p25 gag is induced by addition of 100 ml of an IPTG solution (100 mM) to yield a 1 mM final concentration in the fermenter. Cells are grown for 3 additional hours and subsequently harvested using continuous flow centrifugation. At this step cells may be frozen and kept at -20°C until purification of p25 gag proceeds. Alternatively, 250 l fermenters are inoculated with 1-5 l from the inoculum culture. Growth, induction, and harvest are as indicated before.

E. Purification and characterization of p25 gag.

E.1. Cell breakage.

Frozen *E. coli* cells are thawed and suspended in 2.5 volumes of lysis buffer (0.1M sodium phosphate (NaPi), pH 7.5, 1 mM EDTA, 0.1 M NaCl). Cells are broken in a non-continuous system using a 300 ml glass unit of a Dyno Mill at 3000 rpm and 140 ml of acid-washed glass beads for 15 min. The jacketed chamber is kept cool by a -20°C ethylene glycol solution. Broken cells are centrifuged at 27,000 x g for 25 minutes to remove debris and glass beads. The supernatant is recovered and kept at 4°C.

E.2. Selective protein precipitation.

The cell extract is made 30% $(NH_4)_2SO_4$ by slowly adding the ammonium sulfate at 4°C. The extract is stirred for 10 min after the final concentration is achieved, followed by contrifugation at 27,000 x g for 20 min. The pellet is resuspended in 1 M NaCl, 1 mM EDTA, 1% Triton® X-100, and 5% SDS, and then boiled for 5 min.

E.3. Gel filtration.

The fraction obtained by selective precipitation is submitted to gel filtration using a G50 Sephadex® column equilibrated in 0.03 M NaPi, pH 6.8.

Chromatography is developed in the same solution. Fractions are collected and absorbance at 280 nm is determined. Protein-containing fractions are pooled and characterized by protein gel electrophoresis, Western analysis, and ELISA.

E.4. Characterization of recombinant p25 gag.

a. Protein gel electrophoresis. SDS-polyacrylamide gel analysis (10%-20% gradient gels) of proteins from pGAG25-containing cells and control cells indicated that varying levels of a protein of a molecular weight of about 25,000 were specifically induced in cells containing p25 gag expression plasmids after derepression of the tacI promoter with IPTG. Identity of the p25 gag gene product was confirmed by both an enzyme-linked immunosorbent assay (ELISA, see E.4.c.) and Western immunoblot analysis (see E.4.b.) using both AIDS patient serum and a monoclonal antibody to viral p25 gag.

b. Western analysis. Samples were electrophoresed under denaturing conditions on a 10%-20% polyacrylamide gradient gel. Samples were electroblotted onto nitrocellulose. The nitrocellulose paper was washed with a 1:250 dilution of AIDS patient reference serum (EW5111, obtained from P. Feorino, Centers for Disease Control, Atlanta, Georgia) and then with a 1:500 dilution of HRP-conjugated goat antiserum to human immunoglobulin (Cappel, No. 3201-0081). Alternatively, the nitrocellulose was washed with undiluted culture supernatant from 76C, a murine monoclonal antibody to ARV-2 p25 gag, and then with a 1:500 dilution of HRP-conjugated goat antiserum to mouse immunoglobulin (TAGO, No. 6450). The substrate for immunoblots was HRP color development reagent containing 4-chloro-1-naphthol.

The p25 gag protein reacted with both AIDS patient reference serum and with the monoclonal antibody, while it shows no reactivity with the non-immune serum.

c. ELISA. p25 gag was purified from bacterial extracts as previously described. The reactivity of sera with the purified protein was assayed by coating wells of microtiter plates with 0.25 µg/ml, adding dilutions of test sera (positive reference serum EW5111 of human negative serum), followed by a 1:1000 dilution of HRP-conjugated goat antiserum to human immunoglobulin. p25 gag protein reacted with the positive serum with a midpoint of titration curve of approximately 1:800. There was no reactivity with serum from a normal individual.

10. Comparison of recombinant p25 gag protein and natural p25 gag protein in ELISA.

The reactivity of purified recombinant p25 gag to various sera was compared to that of natural p25 gag protein purified by preparative polyacrylamide gel electrophoresis in an ELISA assay. For control, assays were also made using disrupted gradient purified virus (5 µg/ml).

PVC microtiter plates were incubated for 2 hr at 37°C with 10 µg/ml (50 µl/well in 0.1 M sodium borate, pH 9.0) of the Ig fraction of ascites from murine anti-p25 gag monoclonal antibody 76C. The plates were washed with PBS and the wells were filled with 10% normal goat serum in PBS. Following a 30 min incubation at room temperature, the plates were washed with normal saline containing 0.05% Triton X-100 (ST) and dilutions of the test ARV protein (50 µl/well in ST with 10% goat serum [STGS]) were added to the wells. The-plates were incubated for 2 hr at 37°C, washed with ST, and then incubated for 1 hr at 37°C with 50 µl/well of rabbit antiserum raised against disrupted ARV (1:1000 dilution in STGS). The wells were washed, incubated for 1 hr with 50 µl of a 1:1500 dilution in STGS of HRP-conjugated goat antiserum to rabbit immunoglobulin, washed, and then the wells received 50 µl/well of substrate solution (150 µg/ml 2,2'-azino-di-[3-ethylbenzthiazolene sulfonic acid], 0.001% H_2O_2, 0.1 M citrate pH 4). The reaction was stopped after incubation for 30 min at 37°C by the addition of 50 µl/well of 10% SDS. The absorbance was read on a Flow Titertech ELISA reader at 414 nm. Samples were assayed in duplicate beginning at a dilution of 1:10 and by serial 2-fold dilutions thereafter.

The table below summarizes the results of assays on 8 AIDS sera that scored positive in the assay with disrupted virus and 6 normal sera that were negative in the disrupted virus assay.

SERUM NUMBER	ELISA ASSAY TITER[a]		
	Disrupted Virus	Recomb. p25 gag	Viral p25 gag
Group I: Sera Scoring As Positive in Virus ELISA[b]			
1	51,200	3,125	3,125
5	12,800	25	25
6	12,800	625	625
7	12,800	3,125	3,125
8	25,600	15,625	15,625
9	12,800	625	625
13	800	125	125
18	3,200	625	625
GroupII: Sera Scoring Negative in Virus ELISA[b]			
15	–[c]	–	–
16	–	–	–
19	–	–	–
21	–	–	–
26	–	–	–
33	–	–	–

---

a. Reported as the reciprocal of the serum dilution that gave a signal equivalent to 50% of the maximum.

b. Results were confirmed by immunofluorescence and immunoblotting as described previously.

c. No detectable signal at a 1:25 serum dilution.

These results show that p25 gag purified from bacteria behaves identically to similarly purified p25 gag from AIDS virus in an ELISA of the eight AIDS patient sera. The results of the ELISA show that there is a wide variation in the levels of anti-p25 gag antibodies and suggests that antibodies to some virus-encoded proteins may not be detected using conventional virus-based assay systems.

11. Expression of p41 gag protein of ARV-2 in bacteria.

A fusion protein of the p25 gag and p16 gag proteins of ARV-2, designated p41 gag, was synthesized in E. coli strain D1210 transformed with plasmid pGAG41-10. pGAG41-10 was constructed from plasmid pGAG25-10 as shown in Figure 3 by inserting an SphI-HpaI fragment from the ARV-2 genome containing the sequences from the C-terminal p16 gag portion of the p53 gag precursor polyprotein and part of the p25 gag protein between the SphI and BamHI sites of pGAG25-10. The coding strand of the DNA sequence cloned in pGAG41-10 is shown in Figure 5. Transformation and induction of expression were effected by the procedures described above. The cells were treated and the p41 gag protein was visualized on Coomassie-stained gel as described above. The approximate molecular weight of the observed protein was 41,000 daltons. The protein reacted with AIDS sera and monoclonal antibody to p25 gag in Western and ELISA analyses carried out as above.

12. Expression of p16 gag protein of ARV-2 in bacteria.

The sequence shown in Figure 6 and coding for the p16 gag protein was chemically synthesized using yeast-preferred codons. The blunt-end SalI fragment (381 bp) was cloned into PvuII-SalI digested and gel-isolated ptac5 (see 9 and 11 above). The resulting plasmid was used to transform D1210 cells, as in 9 above. Expression was induced with IPTG, and proteins were analyzed by polyacrylamide gel electrophoresis and Western analysis. A band of about 16,000 daltons was induced by IPTG in the transformed cells. This protein showed reactivity in Western blots with immune sera from AIDS patients. No reactivity was observed with sera from normal individuals.

A recombinant gag protein was also expressed in Cos (mammalian) cells.

13. Production of ARV-2 env protein by yeast.

A. Host-vector system.

A partial env protein is synthesized by *S. cerevisiae* 2150-2-3 transformed with plasmid pDPC303. Plasmid pDPC303 is a yeast expression vector which contains the sequence coding for 2/3 of the env protein as well as pBR322 sequences including the amp^R gene and 2-micron sequences including the yeast leu 2-04 gene. Expression of env is under regulation of the yeast pyruvate kinase promoter and terminator sequences. Yeast strain *S. cerevisiae* 2150-2-3 has the following genotype: Mat a, ade 1, leu 2-112, cir°. This strain was obtained from Dr. Leland Hartwell, University of Washington.

B. Construction of pDPC303, a yeast expression vector for env protein.

Plasmid pDPC303 contains an "expression cassette" (described below) for env cloned into the *Bam*HI site of vector pC1/1. Vector pC1/1 contains pBR322 and 2 micron sequences including the amp^R and yeast leu 2-04 markers. It was derived from pJDB219d (Beggs, Nature (1978), 275: 104) by replacing the pMB9 region with pBR322 sequences.

The "expression cassette" for env consists of the following sequences fused together in this order (5' to 3'): yeast pyruvate kinase (PYK) promoter, env cDNA, and PYK terminator. The PYK promoter and terminator regions were derived from PYK cDNA isolated as described in Burke, et al, *J Biol Chem* (1983) 258:2193-2201.

The env fragment cloned into the expression cassette was derived from ARV-2 cDNA and comprises a 1395 bp cDNA fragment which codes for env amino acid residues coded by nt 5857 to nt 7251 (Figure 2). In addition, there are 5 extra codons fused in reading frame in the 5' end, the first codon corresponding to a methionine, and 4 extra codons fused in reading frame at the 3' end followed by a stop codon. The extra codons were incorporated to facilitate cloning procedures exclusively.

Figure 7 shows the coding strand of the nucleotide sequence cloned in pDPC303 and the amino acid sequence derived from it. DNA sequences that are not underlined in the figure were derived directly from the ARV-2 (9B) cDNA described above. All other sequences were either chemically synthesized or derived from the PYK vector.

C. Preparation of 2150 (pDPC303) strain

Yeast cells S. cerevisiae 2150-2-3 (Mat a, ade 1, leu 2-04, cir°) were transformed as described by Hinnen et al (PNAS (1978) 75:1929-1933) and plated onto leu- selective plates. Single colonies were inoculated into leu- selective media and grown to saturation. Cells were harvested and the env protein was purified and characterized as described below.

D. Purification and characterization of env protein.

D.1. Cell breakage.

Frozen S. cerevisiae 2150-2-3 (pDPC303) are thawed and suspended in 1 volume of lysis buffer (1 μg/ml pepstatin, 0.001 M PMSF, 0.001 M EDTA, 0.15 M NaCl, 0.05 M Tris-HCl pH 8.0), and 1 volume of acid-washed glass beads are added. Cells are broken in a non-continuous system using a 300 ml glass unit of Dyno Mill at 3000 rpm for 10 min. The jacket is kept cool by a -20°C ethylene glycol solution. Glass beads are decanted by letting the mixture set for 3 minutes on ice. The cell extract is recovered and centrifuged at 18,000 rpm (39,200 x g) for 35 min. The supernatant is discarded and the precipitate (pellet 1) is further treated as indicated below.

D.2. Selective SDS extraction of insoluble material.

Pellet 1 is resuspended in 4 volumes of Tris-HCl buffer (0.01 M Tris-HCl, pH 8.0, 0.01 M NaCl, 0.001 M PMSF, 1 µg/ml pepstatin, 0.001 M EDTA, 0.1% SDS) and extracted for 2 hr at 4°C with agitation. The solution is centrifuged at 6,300 x g for 15 min. The insoluble fraction (pellet 2) is resuspended in 4 volumes (360 ml) of PBS (per liter: 0.2 g KCl, 0.2 g KH_2PO_4, 8.0 g NaCl, 2.9 g $Na_2HPO_4 \cdot 12H_2O$), 0.1% SDS, 0.001 M EDTA, 0.001 M PMSF, 1 µg/ml pepstatin, and centrifuged at 6,300 x g for 15 min. The pellet (pellet 3) is suspended in 4 volumes of PBS, 0.2% SDS, 0.001 M EDTA, 0.001 M PMSF, 1 µg/ml pepstatin and is extracted for 12 hr at 4°C with agitation on a tube rocker. The solution is centrifuged at 6,300 x g for 15 min. The soluble fraction is recovered for further purification as indicated below. (The pellet can be reextracted by resuspending it in 4 volumes of 2.3% SDS, 5% ß-mercaptoethanol, and boiling for 5 min. After boiling, the solution is centrifuged at 6,300 x g for 15 min. The soluble fraction is recovered for further purification.)

D.3. Selective precipitation and gel filtration.

The soluble fraction is concentrated by precipitation with 30% ammonium sulfate at 4°C. The pellet (pellet 4) is resuspended in 2.3% SDS, 5%

β-mercaptoethanol, and chromatographed on an ACA 34* (LKB Products) gel filtration column. The column is equilibrated with PBS, 0.1% SDS, at room temperature. Chromatography is developed in the same solution with a flow rate of 0.3 ml/min. Five ml fractions are collected, pooled and characterized by protein gel electrophoresis, Western analysis, and ELISA. If needed, pooled fractions are concentrated by vacuum dialysis on Spectrapor #2* (MW cutoff 12-14K).

D.4. Characterization of recombinant env.

SDS polyacrylamide gel analysis (12% acrylamide gels) showed that a new 55,000 dalton protein was being synthesized in yeast cells transformed with the env-containing vector. The 55,000 dalton protein is absent from cells transformed with control plasmid (vector without env insert). The identity of env was confirmed by both ELISA (see 9.E.4.c) and Western analysis using AIDS patient serum. In both assays the 55,000 dalton protein showed immunoreactivity. No reactivity was obtained with serum from a normal individual.

Recombinant env was also expressed in mammalian (Cos) cells.

14. Expression of p31 pol protein of ARV-2 in bacteria.

A. Host vector system

The C-terminal region of the polymerase gene (p31 pol) is synthesized by E. coli strain D1210 transformed with plasmid pTP31.2. Plasmid pTP31.2 is a pBR322 derivation which contains the sequence coding for p31 under transcriptional control of the hybrid tac

---

* trade-mark

promoter (described in 9.A). Expression of p31 is induced in bacterial transformants by IPTG.

B. Construction of pTP31.2.

B.1. Construction of M13 template 01100484.

A 5.2 kb DNA fragment was isolated from a KpnI digest of ARV-2 (9b) containing the 3' end of the pol gene, orf-1, env and the 5' end of orf-2, that had been run on a 1% low melting point agarose (Sea-Pack) gel and extracted with phenol at 65°C, precipitated with 100% ethanol and resuspended in TE. Eight µl of this material were further digested with SstI for 1 hr at 37°C in a final volume of 10 µl. After heat inactivation of the enzyme, 1.25 µl of this digest were ligated to 20 ng of M13mp19 previously cut with KpnI and SstI, in the presence of ATP and in a final volume of 20 µl. The reaction was allowed to proceed for 2 hr at room temperature. Five µl of this mixture were used to transform competent *E. coli* JM101. Clear plaques were grown and single-stranded DNA was prepared as described in Messing and Vieira, Gene (1982) 19:269-276.

B.2. In Vitro mutagenesis of 01100484.

The DNA sequence in 01100484 was altered by site specific mutagenesis to generate a restriction site recognized by NcoI (CCATGG). An oligodeoxynucleotide that substitutes the A for a C at position 4299 (Figure 2) and changes a T for an A at position 4305 (Figure 2) was synthesized using solid phase phosphoramidite chemistry. Both of these changes are silent in terms of the amino acid sequence, and the second one was introduced to decrease the stability of the heterologous molecules. The oligomer was named ARV-216 and has the sequence: 5'-TTAAAATCACTTGCCATGGCTCTCCAATTACTG and corresponds to the non-coding strand since the M13 derivative template 01100484 is single-stranded and contains the coding strand. The 5' phosphorylated oligomer was annealed to the 01100484 M13 template at 55°C in the presence of 5' dephosphorylated M13 sequence-ing primer, 50 mM Tris-HCl pH 8, 20 mM KCl, 7 mM $MgCl_2$ and 0.1 mM EDTA. The polymerization reaction was done in 100 µl containing 50 ng/µl DNA duplex, 150 µM dNTPs, 1 mM ATP, 33 mM Tris-acetate pH 7.8, 66 mM potassium acetate, 10 mM magnesium acetate, 5 mM DTT, 12.5 units of T4 polymerase, 100 µg/ml T4 gene 32 protein and 5 units of T4 DNA ligase. The reaction was incubate-ed at 30°C for 30 min and was stopped by the addition of EDTA and SDS (10 mM and 0.2% respectively, final concentration). Competent JM101 E. coli cells were transformed with 1, 2, and 4 µl of a 1:10 dilution of the polymerization product and plated into YT plates. Plaques were lifted by adsorption to nitrocellulose filters and denatured in 0.2 N NaOH, 1.5 M NaCl, followed by neutralization in 0.5 M Tris-HCl pH 7.3, 3 M NaCl and equilibrated in 6 x SSC. The filters were blotted dry, baked at 80°C for 2 hr and preannealed at 37°C in 0.2% SDS, 10 x Denhardt's, 6 x SSC. After 1 hr, 7.5 million CPM of labeled ARV-216 were added to the filters and incubated for 2 additional hr at 37°C. The filters were washed in 6 x SSC at 42°C for 20 min, blot-dried and used to expose film at -70°C for 1 hr using an intensifying screen. Strong hybridizing plaques were grown and single-stranded DNA was prepared from them and used as templates for sequencing. Sequencing showed that template 01021785 contains the

NcoI site as well as the second substitution mentioned above.

A second oligomer was synthesized to insert sites for SalI and EcoRI immediately after the termination codon of the pol gene (position 5101, Figure 2). This oligomer was called ARV-248 and has the sequence: 5'-GGTGTTTTACTAAAGAATTCCGTCGACTAATCCTCATCC. Using the template 01020785, site specific mutagenesis was carried out as described above except that the filter wash after the hybridization was done at 65°C. As above, 8 strong hybridizing plaques were grown and single-stranded DNA was sequenced. The sequence of template 01031985 shows that it contains the restriction sites for NcoI, SalI, and EcoRI as intended.

B.3. Isolation of NcoI-EcoRI and NcoI-SalI DNA fragments that contain p31.

Replicative form (RF) of the 01031985 template was prepared by growing 6 clear plaques, each in 1.5 ml of 2 x YT at 37°C for 5 hr. Double-stranded DNA was obtained as described by Maniatis, et al, Molecular Cloning, a Laboratory Manual, Cold Spring Harbor, 1982, pooled and resuspended in 100 µl final volume. Ten µl of RF were digested with NcoI and EcoRI in a final volume of 20 µl. This fragment was used for direct p31 expression in bacteria. An additional 20 µl of RF were cut with NcoI and SalI in 40 µl. This fragment was used for p31 expression in yeast. The samples were run on a 1% low melting point agarose (Sea-Pack) gel and the DNAs were visualized by fluorescence with ethidium bromide. The 800 bp bands were cut and the DNAs were extracted from the gel as mentioned above and resuspended in 10 µl of TE. The fragments were called ARV248NR#2 and ARV248NL, respectively.

B.4. Cloning of p31 into plot7.

The vector plot7 (3 µg) (Hallewell, et al, Nucl Acid Res (1985) 13, No. 6, pp. 2017-2034) was cut with NcoI and EcoRI in 40 µl final volume and the enzymes were heat-inactivated after 3 hr. Two µl of this digest were mixed with 2 µl of ARV248NR#2 and ligated in 20 µl in the presence of ATP and T4 DNA ligase at 14°C overnight, and 10 µl of this mixture were used to transform competent D1210 cells. Colonies resistant to 2 mM IPTG and 100 µg/ml ampicillin were selected and supercoiled DNA was extracted from each of them. The DNAs were then restricted with NcoI and EcoRI and analyzed by agarose gel electrophoresis. Clones with the appropriate 800 bp insert were selected for further use. They are designated pRSP248 numbers 3 and 4.

B.5. Construction of pTP31.

The NcoI site introduced into 01100485 is 52 bp downstream from the putative start of p31. Three oligomers were synthesized as above that code for the first 18 amino acids of p31 and generate a cohesive NcoI end at the 3' end of the molecule. The 5' end of the molecule has been extended beyond the initiation codon to include a ribosome binding site. The oligomers that were synthesized have the sequences:

```
     ARV-221.-
           CCCC     C  C
5'AGGXAACAGAAAAATGATAGATAAGGCACAAGAA
           TTTT     T

     ARV-222.-
5'GAACATGAGAAATATCACAGTAATTGGAGAGC

     ARV-223.-
3'CGTGTTCTTCTTGTACTCTTTATAGTGTCATTAACCTCTCGGTAC
```

One hundred fifty picomoles each of dephosphorylated ARV-211, phosphorylated ARV-222 and ARV 223 were ligated to 20 µg of pRSP248 previously cut with NcoI, at 14°C for 18 hr in a final volume of 62 µl. After phenol extraction and ethanol precipitation, the DNA was resuspended in 40 µl H_2O and incubated with 15 units of Klenow fragment in the presence of 0.5 mM dNTPs for 1 hr at room temperature. The sample was phenol extracted, ethanol precipitated, resuspended in 40 µl H_2O, and digested with EcoRI. The DNA was then run on a low melting point agarose gel and the fragment of about 820 bp was extracted as described above and resuspended in a final volume of 20 µl of H_2O. After phosphorylating the ends, 5 µl of the sample were incubated for 18 hr at 14°C with 150 ng of plot7 that had been cut with PvuII and EcoRI and its ends dephosphorylated, in the presence of T4 DNA ligase, ATP and in a final volume of 31 µl. Five µl of ligation product were used to transform RR1deltaM15. Clones resistant to 100 µg/ml of ampicillin were selected and supercoiled DNA was extracted from them. The DNAs were digested with NcoI and EcoRI and resolved on a 1% agarose gel. Colonies with the appropriate size insert were obtained and named pTP31. The p31 sequence contained in the insert is shown in Figure 8. Underlined sequences were chemically synthesized. Others were derived from DNA.

C. Screening of transformants for specific proteins that react with AIDS sera.

Bacterial transformants containing either the vector alone, or the vector with the p31 DNA (pTP31.2) were grown in L-broth with 0.02% ampicillin to an OD_{650} of 0.5. Cultures were induced by the addition of IPTG to a final concentration of 2 mM and grown for 3 more hr. Bacteria from 1 ml cultures were pelleted and resuspended in 200 µl of gel sample buffer. The cells were disrupted by three cycles of freezing and thawing, boiled, and the extracts loaded onto 12.5% polyacrylamide-SDS minigels. Proteins were electrophoresed and transferred to nitrocellulose by electroblotting. The nitrocellulose filters were reacted with serum EW5111 (diluted 1:100; positive reference serum from the CDC that reacts strongly with viral p31), horse radish peroxidase-conjugated goat anti-human IgG and HRP substrate. A prominent band at ~30,000 d and several lower molecular weight species were seen in gels of extracts from transformants with the p31 DNA, but not in extracts from bacteria transformed with the vector alone.

D. <u>Demonstration that the polypeptide from the C-terminal region of the pol gene is analogous to the viral p31 protein</u>.

Lysozyme-NP40* extracts were prepared from bacteria transformed with pTP31.2 or vector alone. Five ml cultures were grown, the cells pelleted and resuspended in 1 ml of 50 mM Tris-HCl pH 8, 0.5 mM EDTA, 1 mg/ml lysozyme and incubated at 0°C for 15 min NaCl, $MgCl_2$, and NP40* were added to final concentrations of 0.4, 5 mM and 0.5% respectively, mixed and incubated with DNAse I (100 µg/ml) at 0°C for 30 min. When EW5111 serum (diluted 1:100) was preincubated with a 1:10 dilution of the cell extracts from bacteria transformed with pTP31.2, prior to reaction with a virus blot, the viral p31 band was completely eliminated, while reactivity with other viral proteins remained unaffected. In contrast, extracts from bacteria

* trade-mark

transformed with the vector alone did not absorb out the p31 reactive antibodies. The viral p31 protein is thus the product of the C-terminal or endonuclease region of the pol gene of ARV-2.

15. Expression of p31 pd protein in yeast.

A. Construction of a yeast vector p31/GAP-ADH2: Cloning of p31 into pAB24.

The ARV248NL fragment was cloned into pBS100 previously cut with NcoI and SalI. pBS100 is a bacterial vector derived from pAB12 with a BamHI cassette consisting of the GAP-ADH2 promoter, an ARV-env gene as an NcoI-SalI fragment, and the GAP terminator. The BamHI cassette from two positive clones of pBS100/p31/GAP-ADH2 was cloned into pAB24, a yeast vector with both ura and leu selection capabilities. Both orientations of the cassette in this vector were screened for and used to transform the yeast strain AB110 (Mat a, ura 3-52, leu 2-04, or both leu 2-3 and leu 2-112, pep 4-3, his 4-580, cir°). These cells were plated in both ura- and leu- plates. Also, ura- cells were plated onto leu- plates.

B. Induction of p31 expression.

Three different induction procedures were done: 1.- Ura- colonies patched on ura- plates were induced for 24 hr in YEP/1% glucose. Both a Western and a polyacrylamide gel were run on these samples. Both results were negative. 2.- Colonies from ura- plates patched on leu- plates were induced in either leu-/3% ethanol or YEP/1% glucose for 24 hr. A Western and a polyacrylamide gel were run on these samples and the results were also negative. 3.- Colonies from leu-plates patched on leu- plates were induced in either leu-/3% ethanol or YEP/1% glucose for 24 hr. The polyacrylamide gel showed a negative result. No Western was run on these samples.

16. Expression of superoxide dismutase (SOD)-p31 fusion protein in yeast.

A. Construction of pC1/1-pSP31-GAP-ADH2 derivative.

For the construction of a gene for a fused protein SOD-p31 to be expressed in yeast, a plasmid (pSI4/39-2) was used. This plasmid contains the SOD gene fused to the proinsulin gene under the regulation of the ADH-2/GAP promoter. The proinsulin gene is located between EcoRI and SalI restriction sites. To substitute the proinsulin gene with the ARV248NL fragment, two oligomers designated ARV-300 and ARV-301, respectively, were synthesized using phosphoramidite chemistry. The sequences generate cohesive ends for EcoRI and NcoI on each side of the molecule when the two oligomers are annealed. ARV-300 and ARV-301 have the sequences:

```
ARV-300    5' AATTCAGGTGTTGGAGC
                   GTCCACAACCTCGGTAC 5'   ARV-301
```

Two µg of pSI4/39-2 linearized with EcoRI were ligated to 100 picomoles each of phosphorylated ARV-300 and dephosphorylated ARV-301 in the presence of ATP and T4 DNA ligase in a final volume of 35 µl. The reaction was carried out at 14°C for 18 hr. The DNA was further digested with SalI and the fragments were resolved on a 1% low melting point agarose gel and a fragment containing the vector plus the SOD gene (~6.5

kb) was purified as described above and resuspended in 50 µl of TE. Five µl of this preparation were ligated to 5 µl of ARV248NL in 20 µl final volume for 18 hr at 14°C and 5 µl used to transform competent HB101 cells. The resultant plasmid was called pSP31. Twenty µg of this plasmid were digested with BamHI and a fragment of about 2900 bp was isolated by gel electrophoresis, resuspended in TE and ligated to pC1/1 previously cut with BamHI. This DNA was used to transform HB101 and transformants with the BamHI cassette were obtained. Yeast strains 2150, PO17, and AB110 were transformed with this pC1/1-pSP31-GAP-ADH2 derivative, both short and long orientations. The strain 2150 gave no transformants. All other transformants were patched on leu- plates.

B. Induction of pC1/1-pSP31-GAP-ADH2.

Three different kinds of inductions were tried: 1.- PO17 colonies were induced in either a 10 ml culture of YEP/1% glucose or a leu-/3% ethanol culture for 24 hr. The yeast pellets were analyzed by both polyacrylamide gels and Westerns and even though the Coomassie-stained gel showed a negative result, the Western did light up a band of the correct molecular weight with both induction methods. 2.- PO17 colonies were induced in a 30 ml culture of YEP/1% ethanol for 48 hr. Aliquots were analyzed by PAGE at various time points during the induction. The Coomassie-stained gel shows a band in the correct molecular weight range (47-50 kd) that appears after 14 hr in YEP/1% ethanol and reaches a maximum intensity at 24 hr of induction. The Western result correlates well with the Coomassie-stained gel, showing strong bands at 24 and 48 hr. 3.- AB110 colonies were induced in either leu-/3% ethanol or

YEP/1% glucose for 24 hr. PAGE and Westerns were run and the results were negative for the PAGE and positive for the Western, in both induction methods.

17. Purification and characterization of SOD-p31 from bacteria or yeast

Frozen bacteria (yeast) cells are thawed at room temperature and suspended in 1.5 volumes of lysis buffer (20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 1 mM PMSF, for bacteria; 50 mM Tris-Cl, pH 8.0, 2 mM EDTA, 1 mM PMSF for yeast), and mixed with 1 volume of acid-washed glass beads.

Cells are broken for 15 min in a non-continuous mode using the glass chamber of a Dynomill unit at 3,000 rpm, connected to a -20°C cooling unit. Glass beads are decanted for 2-3 min on ice, the cell lysate is removed. The decanted glass beads are washed twice with 30 ml of lysis buffer at 4°C. The cell lysate is centrifuged at 39,000 x g for 30 min.

The pellet obtained from the above centrifugation is washed once with lysis buffer, after vortexing and suspending it at 4°C (same centrifugation as above). The washed pellet is treated with 0.2% SDS (for bacteria) and 0.1% SDS (for yeast) in lysis buffer and is agitated by rocking at 4°C for 10 min. The lysate is centrifuged at 39,000 x g for 30 min. The pellet is boiled in sample buffer (67.5 mM Tris-Cl, pH 7.0, 5% ß-mercaptoethanol, 2.3% SDS) for 10 min and centrifuged for 10 min at 39,000 x g. The supernatant is recovered and further centrifuged at 100,000 x g for 60 min (60 Ti rotor). This step is replaced by a 0.45 µm filtration when yeast is used. The supernatant from the above centrifugation is loaded (maximum 50 mg of protein) on a gel filtration column (2.5 x 90 cm, ACA

34* LKB) with a flow rate of 0.3-0.4 ml/min, equilibrated with phosphate-buffered saline (PBS), 0.1% SDS. The fractions containing SOD-p31 are pooled and concentrated either by vacuum dialysis or using a Diaflo® YM5 Amicon membrane at 40 psi. The protein is stored at -20°C as concentrated solution.

Gel electrophoresis analysis shows that the SOD-p31 protein migrates having a molecular weight of about 46 kd and is over 90% pure.

18. ELISA for antibodies to hTLR using recombinant ARV-2 polypeptides

Stock solutions of purified p25 gag protein (1.25 mg/ml in 20 mM sodium phosphate, 0.1% SDS, pH 7.2), purified env protein (2 mg/ml in 20 mM sodium phosphate, 0.1% SDS, pH 7.2), and purified SOD-p31 fusion protein (2 mg/ml in 20 mM sodium phosphate, 0.1% SDS, pH 7.2) were prepared.

For coating microtiter plates (Dynatech Immulon® I), 1 part each of the stock solutions of p25 gag, env, and SOD-p31 were added to 997 parts of borate coating buffer (0.05 M borate, pH 9.0). One hundred microliters of the coating solution was added to each well, and the plates were covered and incubated 2 hr at 37°C or 12 hr at 4°C. The coating solution was then aspirated from the wells and the plates washed 4 x with wash solution (0.137 M 0.8% NaCl, 0.05% Triton X-100).

Serum samples were diluted 1:100 in dilution solution (0.1% casein, 1 mM EDTA, 1% Triton X-100, 0.5 M NaCl, 0.01% thimerosal, pH 7.5) with yeast protein (strain AB103.1) extract (1:40 dilution, approximately 2 mg protein per ml in PBS containing 1% Triton X-100, 2 mM PMSF, 0.01% thimerosal) and *E. coli* protein extract (1:40 dilution, approximately 1 mg protein per ml in PBS

---

* trade-mark

containing 1% Triton X-100, 2 mM PMSF, 0.01% thimerosal) added to the dilution solution. Extraction procedures were similar to those described in 13 and 14 above but using non-recombinant strains. One hundred microliters of diluted serum was added to each well and incubated 30 min at 37°C. The plates were then washed 6 x with wash solution.

Goat anti-human Ig labeled with horseradish peroxidase (Cappel) diluted 1:8000 in dilution solution without added yeast and E. coli extracts were added at 100 µl/well to the plates and incubated 30 min at 37°C. The plates were then washed 6 x with wash solution. Substrate solution (10 ml citrate buffer, 10.5 g citric acid/liter dH_2O, pH to 4.0 with 6 M NaOH), 0.1 ml ABTS (15 mg/ml 2,2'-azino-di-(3-ethylbenzthiazolene sulfonic acid) in dH_2O) and 3.33 µl H_2O_2) at 100 µl/well was then added to the plates and the plates wrapped in foil and incubated at 37°C for 30 min. The reaction was then stopped by adding 50 µl/well of 10% SDS. Readings were made with a Dynatech ELISA reader set for dual wavelength reading: absorbance wavelength of 1 (410 nm) and reference wavelength of 4.

Results

The following sera were tested:

A. 89 consecutive blood donors from the Kansas City Blood Bank ("normal blood donors"): log nos. 1001-1081, 1085-1092.

B. 52 sera from patients with lymphadenopathy syndrome (LAD) or AIDS or sexual partners of persons with LAD or AIDS (referred to as "contacts")--all obtained from UCSF AIDS Serum Bank panel: log nos. 4601-4652.

The positive/negative cut-off used was 5 x (average background signal - signal with diluent alone) and was determined to be 0.195. Thus, sera with signals below 0.195 were rated (-); those above were rated (+). Each sample was also evaluated by the commercially available ABBOTT HTLV III EIA kit (Abbott Labs) and by Western analysis.

Tests on the normal blood donor samples indicated all except one were negative in the invention ELISA. This normal serum scored negative in the ABBOTT HTLV III EIA test, but was actually positive, as confirmed by Western analysis.

The results of the tests on the 52 sera from LAD and AIDS patients and contacts are tabulated below:

Serum No.	Diagnosis	ABBOTT EIA	Invention ELISA		Western
4601	Contacts	+	1.89	+	+
02	Contacts	–	0.04	–	–
03	Contacts	+	1.44	+	+
04	Contacts	+	1.92	+	+
05	Contacts	–	0.04	–	–
06	Contacts	+	>2	+	+
07	Contacts	+	1.37	+	+
08	Contacts	+	1.60	+	+
09	Contacts	+	>2	+	+
10	Contacts	+	>2	+	+
11	Contacts	+	1.94	+	+
12	Contacts	+	>2	+	+
13	Contacts	+	>2	+	+
14	Contacts	+	>2	+	+
15	Contacts	+	1.97	+	+
16	AIDS	+	0.61	+	+
17	AIDS	+	>2	+	+
18	AIDS	+	>2	+	+
19	AIDS	+	1.58	+	+
20	AIDS	+	1.58	+	+
21	AIDS	+	0.76	+	+
22	AIDS	+	1.74	+	+
23	LAD	+	1.26	+	+
24	LAD	+	>2	+	+
25	AIDS	+	1.04	+	+
26	AIDS	+	1.24	+	+
27	AIDS	+	1.40	+	+
28	AIDS	–	0.07	–	–
29	LAD	+	1.93	+	+
30	Contacts	+	1.96	+	+
31	AIDS	+	1.76	+	+
32	AIDS	+	0.90	+	+
33	AIDS	+	1.69	+	+
34	LAD	+	1.09	+	+
35	AIDS	+	1.54	+	+
36	AIDS	+	1.22	+	+
37	AIDS	+	1.96	+	+
38	AIDS	–	>2	+	+
39	LAD	+	1.85	+	+
40	LAD	+	>2	+	+
41	LAD	+	0.84	+	+
42	LAD	+	1.59	+	+
43	LAD	+	1.71	+	+
44	AIDS	+	1.40	+	+
45	LAD	+	>2	+	+
46	AIDS	+	1.38	+	+

Serum No.	Diagnosis	ABBOTT EIA	Invention ELISA		Western
47	AIDS	+	1.29	+	±
48	LAD	+	1.93	+	+
49	LAD	+/-	0.48	+	+
50	LAD	-	0.04	-	-
51	LAD	-	0.07	-	-
52	LAD	+	1.92	+	+

The above results show that the invention ELISA using recombinant ARV proteins is at least as good as the ABBOTT HTLV III EIA test or Western analysis.

In the invention ELISA reported in this example the yeast and bacterial extracts were added to the serum to bind serum antibodies to yeast and bacteria to prevent such antibodies from binding to the recombinant ARV-2 proteins. Both yeast and bacterial extracts were required since the recombinant polypeptides included polypeptides expressed in yeast and polypeptides expressed in bacteria. If all the polypeptides were expressed in the same type of organism, only one extract would be needed. For instance, if a p25 gag polypeptide expressed in yeast was substituted for the bacterially produced p25 gag polypeptide of the example, only yeast extract would be added to the serum samples.

19. Dot-blot assay for antibodies to hTLR using recombinant ARV-2 polypeptides.

Nitrocellular strips (0.5 x 5 cm) are spotted with 50 ng polypeptide in PBS (spotting volume 2 µl). After spotting the strips are dried at room temperature for 1 hr or more. The strips are then post-coated in a 5% solution of Carnation non-fat dry milk in PBS, 0.01% Thimerosal, for 15-60 min at room temperature. Each test solution sample is diluted 1:50 in 0.5 ml of the post-coating solution in a test tube. A post-coated strip is then placed in the tube and incubated in the sample with rocking at 37°C for 1 hr. The strip is then removed from the tube and washed with post-coating solution. The strip is then incubated for 15 min at room temperature in goat anti-human Ig reagent labeled with horse radish peroxidase diluted 1:500 in post-coating solution. After incubation in the labeled antibody, the strip is washed serially with PBS, 1% Triton, and distilled water. The strips are developed by incubating them in substrate solution (see 23 above) for 15 min at room temperature.

Positive samples will cause a visually perceptible color change at the spotting site. Normal (negative) sera sample yield no color change or give a faint signal that is discernible from a positive signal. Competition assays may be run on sera giving faint signals to verify that they are negative. In the competition assay, polypeptide (10-25 µg/ml) is added to the test sample and incubated from 1 hr at 37°C before the strip is incubated in the sample. With authentic positive sera the signal is completely blocked by the added polypeptide, whereas with normal (negative) sera there is no change in signal.

Samples of organisms that express the above-described ARV-2 p25 gag and ARV-2 env polypeptides and the fusion protein of ARV-2 p31 and SOD were deposited at the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland under the provisions of the Budapest Treaty. The accession numbers and dates of these deposits are listed below.

Expression Product	ATCC Accession No.	Deposit Date
ARV-2 p25 gag	53246	27 August 1985
ARV-2 env	20769	27 August 1985
ARV-2 p31/SOD	20768	27 August 1985

CLAIMS:

1. A process for the preparation of an isolated DNA polynucleotide comprising a fragment of at least 21 bp from the gag or env region of the ARV-2 sequence of Figure 2, wherein the DNA polynucleotide is chemically synthesised at least in part.

2. The process of claim 1, wherein the DNA polynucleotide comprises a fragment of at least 21 bp from the gag region of the ARV-2 sequence of Figure 2.

3. The process of claim 1, wherein the DNA polynucleotide comprises a fragment of at least 21 bp from the env region of the ARV-2 sequence of Figure 2.

4. The process of any one of claims 1 to 3, wherein the DNA polynucleotide is labelled.

5. The process of claim 4, wherein the DNA polynucleotide is labelled with an isotopic label.

6. The process of claim 4, wherein the DNA polynucleotide is labelled with a non-isotopic label.

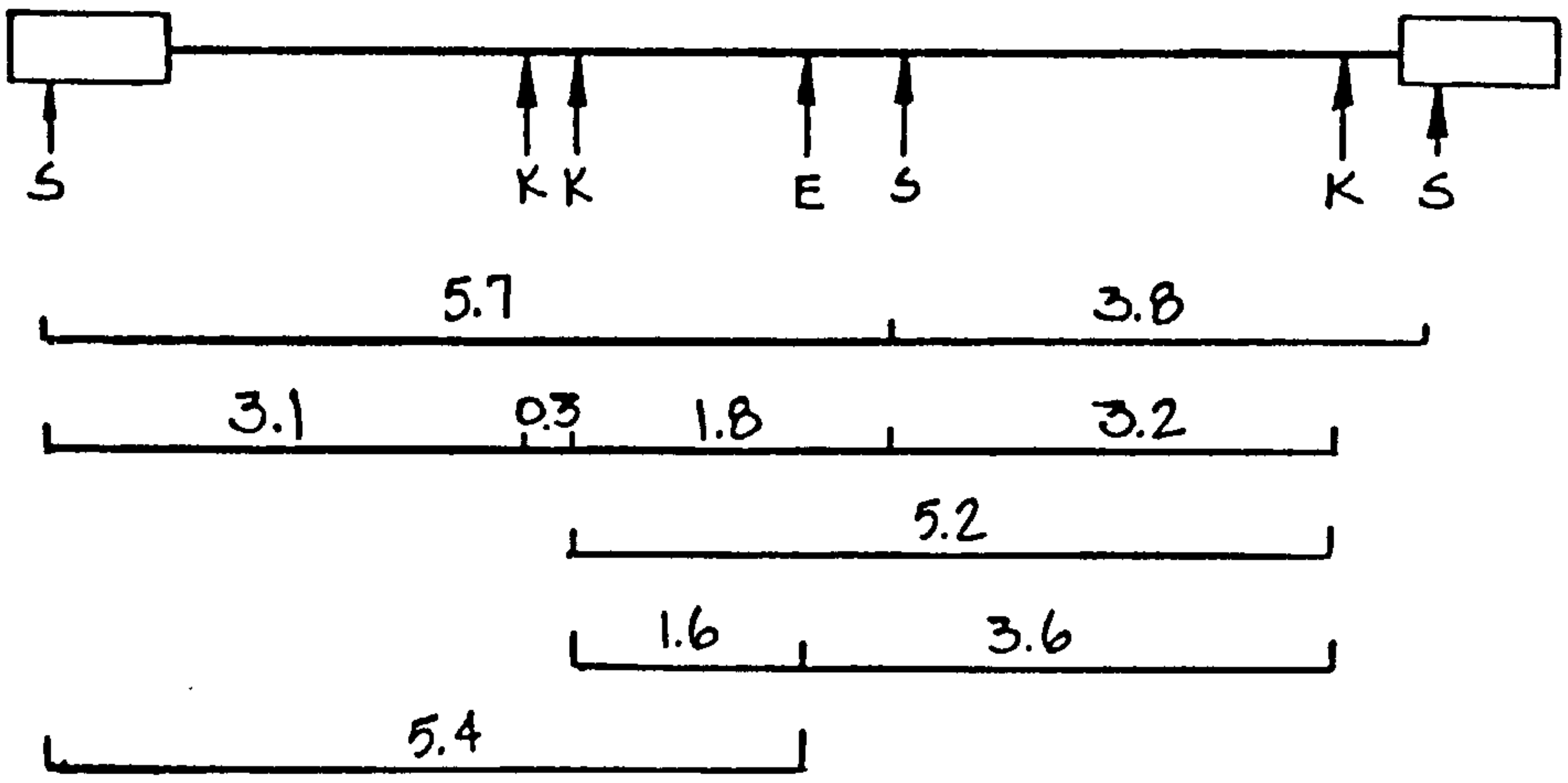

FIG. 1

```
     U3→
-453 CTGGAAGGGCTAATTTGGTCCCAAAGAAGACAAGAGATCCTTGATCTGTGGATCTACCAC
     ACACAAGGCTACTTCCCTGATTGGCAGAATTACACACCAGGGCCAGGGATCAGATATCCA
-333 CTGACCTTTGGATGGTGCTTCAAGCTAGTACCAGTTGAGCCAGAGAAGGTAGAAGAGGCC
     AATGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCTGCATGGGATGGAGGACGCG  L
-214 GAGAAAGAAGTGTTAGTGTGGAGGTTTGACAGCAAACTAGCATTTCATCACATGGCCCGA
     GAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCG T
 -93 CTGGGGACTTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCAGATG
                                      ←U3 R→
     CTGCATATAAGCAGCTGCTTTTTGCCTGTACTG GGTCTCTCTGGTTAGACCAGATCTGAG R
  28 CCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTT
           ←R U5→
     GAGTGCTTCA AGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCA
                                       ←U5
 148 GACCCTTTTAGTCAGTGTGGAAAAATCTCTAGCAG TGGCGCCCGAACAGGGACGCGAAA
     GCGAAAGTAGAACCAGAGGAGCTCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACAG

 268 CAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAATTTTTGACTAGCGGAGGCTAGAAG
              MetGlyAlaArgAlaSerValLeuSerGlyGlyGluLeuAspLysTrpGlu     17
     GAGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAATGGGAA
     LysIleArgLeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTrpAla
 388 AAAATTCGGTTAAGGCCAGGGGGAAAGAAAAAATATAAGTTAAAACATATAGTATGGGCA
     SerArgGluLeuGluArgPheAlaValAsnProGlyLeuLeuGluThrSerGluGlyCys     57
     AGCAGGGAGCTAGAACGATTCGCAGTCAATCCTGGCCTGTTAGAAACATCAGAAGGCTGC
     ArgGlnIleGluGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSer
 508 AGACAAATATTGGGACAGCTACAGCCATCCCTTCAGACAGGATCAGAAGAACTTAGATCA
     LeuTyrAsnThrValAlaThrLeuTyrCysValHisGlnArgIleAspValLysAspThr     97
     TTATATAATACAGTAGCAACCCTCTATTGTGTACATCAAAGGATAGATGTAAAAGACACC
     LysGluAlaLeuGluLysIleGluGluGluGlnAsnLysSerLysLysLysAlaGlnGln
 628 AAGGAAGCTTTAGAGAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAGGCACAGCAA
     AlaAlaAlaAlaAlaGlyThrGlyAsnSerSerGlnValSerGlnAsnTyrProIleVal    137
     GCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTCAGCCAAAATTACCCTATAGTG
     GlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrp
 748 CAGAACCTACAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGG
     ValLysValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeu    177
     GTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTA
     SerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGln
 868 TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAA
```

FIG. 2A

```
      AlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgVal  217  G
      GCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCAGAATGGGATAGAGTG

      HisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAsp
 988  CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGGGGAAGTGAC
                                                                         A
      IleAlaGlyThrThrSerThrLeuGlnGluGlbIleGlyTrpMetThrAsnAsnProPro  257
      ATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCT
                                                                         G
      IleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArg
1108  ATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGA

      MetTyrSerProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAsp  297
      ATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGAT

      TyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsn
1228  TATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCTTCACAGGATGTAAAAAAT

      TrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLys  337
      TGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAA

      AlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGly
1348  GCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGA

      ProGlyHisLysAlaArgValLeuAlaGluAlaMetSerGlnValThrAsnProAlaAsn  377
      CCCGGCCATAAAGCAAGAGTTTTGGCTGAAGCCATGAGCCAAGTAACAAATCCAGCTAAC

      IleMetMetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCys
1468  ATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGACTGTTAAGTGTTTCAATTGT

      GlyLysGluGlyHisIleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpArg  417
      GGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAGA

      CysGlyArgGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGly
                                                           .PhePheArgG
1588  TGTGGAAGGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGG

      LysIleTrpProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluPro  457
      luAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAla  23
      AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA

      ThrAlaProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGlnLys
     AsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnserLeuSerGluA       P
1708  ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG

      GlnGluProIleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsn  497
      laGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGln  63
      CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC       O

      AspProSerSerGlnOC
     ArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyA
1828  GACCCCTCGTCACAATAAGGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAG
                                                                         L
      laAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLysMetIle 103
      CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG

     GlyGlyIelGlyGlyPheIleLysValArgGlnTyrAspGlnIleProValGluIleCysG
1948  GGGGAATTGGAGGTTTTATCAAAGTAAGACAGTACGATCAGATACCTGTAGAAATCTGTG
```

FIG. 2B

```
      lyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArg 143
      GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA

     AsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrValP
2068 ATCTGTTGACTCAGATTGGTTGTACTTTAAATTTCCCCATTAGTCCTATTGAAACTGTAC

      roValLysLeuLysProGlyMetaspGlyProLysValLysGlnTrpProLeuThrGlu 183
      CAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAGCAATGGCCATTGACAGAAG

     GluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerL
2188 AAAAAATAAAAGCATTAGTAGAGATATGTACAGAAATGGAAAAGGAAGGGAAAATTTCAA

      ysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 223
      AAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCTATAAAGAAAAAAGACAGTA

     ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpG
2308 CTAAATGGAGAAAACTAGTAGATTTCAGAGAACTTAATAAAAGAACTCAAGACTTCTGGG

      luValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVal 263
      AAGTTCAGTTAGGAATACCACACCCCGCAGGGTTAAAAAAGAAAAAATCAGTAACAGTAT

     LeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLysTyrThrA
2428 TGGATGTGGGTGATGCATACTTTTCAGTTCCCTTAGATAAAGACTTTAGAAAGTATACTG

      laPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 303
      CATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGC

     LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuG
2548 TGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTAG

      luProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyr 343
      AGCCTTTTAGAAAACAGAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATG

     ValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisL
2668 TAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAACTGAGACAGCATC

      euLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 383
      TGTTGAGGTGGGGATTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTTT

     TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMetLeuProGluL
2788 GGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAATGCTGCCAGAAA

      ysAspSerTrpThrValAsnAspIleGlnLysLeuValglyLysLeuAsnTrpAlaSer 423
      AAGACAGCTGGACTGTCAATGACATACAGAAGTTAGTGGGAAAATTGAATTGGGCAAGTC

     GlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThrLysAlaL
2908 AGATTTATGCAGGGATTAAAGTAAAGCAGTTATGTAAACTCCTTAGAGGAACCAAAGCAC

      euThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 463  P
      TAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGGGAGA

     IleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeuValAlaGluI
3028 TTCTAAAAGAACCAGTACATGAAGTATATTATGACCCATCAAAAGACTTAGTAGCAGAAA

      leGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 503  O
      TACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAATC

     LeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuT
3148 TGAAAACAGGAAAGTATGCAAGGATGAGGGGTGCCCACACTAATGATGTAAAACAGTTAA

      hrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLysIleProLys 543  L
      CAGAGGCAGTGCAAAAAGTATCCACAGAAAGCATAGTAATATGGGGAAAGATTCCTAAAT
```

FIG. 2C

```
     PheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAlaT
3268 TTAAACTACCCATACAAAAGGAAACATGGGAAGCATGGTGGATGGAGTATTGGCAAGCTA

      hrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 583
      CCTGGATTCCTGAGTGGGAGTTTGTCAATACCCCTCCCTTAGTGAAATTATGGTACCAGT

     LeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgG
3388 TAGAGAAAGAACCCATAGTAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCTAATAGGG

      luThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValValSer 623
      AGACTAAATTAGGAAAAGCAGGATATGTTACTGACAGAGGAAGACAAAAAGTTGTCTCCA

     IleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspS
3508 TAGCTGACACAACAAATCAGAAGACTGAATTACAAGCAATTCATCTAGCTTTGCAGGATT

      erGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla 663
      CGGGATTAGAAGTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCAC

     GlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIleLysLysG
3628 AACCAGATAAGAGTGAATCAGAGTTAGTCAGTCAAATAATAGAGCAGTTAATAAAAAAGG

      luLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal 703
      AAAAGGTCTACCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAG

     AspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIleAspLysAlaG
3748 ATAAATTAGTCAGTGCTGGAATCAGGAAAGTACTATTTTTGAATGGAATAGATAAGGCCC

      lnGluGluHisGluLysTyrHisSerAsnTrpArgAlaMetAlaSerAspPheAsnLeu 743
      AAGAAGAACATGAGAAATATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTGC

     ProProValValAlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGluA
3868 CACCTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAAG

      laMetHisGlyGlnValAspCysSerProGlyIleTrpGlnLeuAspCysThrHisLeu 783
      CCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAACTAGATTGTACACATCTAG

     GluGlyLysIleIleLeuValAlaValHisValAlaSerGlyTyrIleGluAlaGluValI
3988 AAGGAAAAATTATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAGCAGAAGTTA

      leProAlaGluThrGlyGlnGluThrAlaTyrPheLeuLeuLysLeuAlaGlyArgTrp 823
      TTCCAGCAGAGACAGGGCAGGAAACAGCATATTTTCTCTTAAAATTAGCAGGAAGATGGC

     ProValLysThrIleHisThrAspAsnGlySerAsnPheThrSerThrThrValLysAlaA
4108 CAGTAAAAACAATACATACAGACAATGGCAGCAATTTCACCAGTACTACGGTTAAGGCCG

      laCysTrpTrpAlaGlyIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGln 863
      CCTGTTGGTGGGCAGGGATCAAGCAGGAATTTGGCATTCCCTACAATCCCCAAAGTCAAG

     GlyValValGluSerMetAsnAsnGluLeuLysLysIleIleGlyGlnValArgAspGlnA
4228 GAGTAGTAGAATCTATGAATAATGAATTAAAGAAAATTATAGGACAGGTAAGAGATCAGG

      laGluHisLeuLysThrAlaValGlnMetAlaValPheIleHisAsnPheLysArgLys 903
      CTGAACACCTTAAGACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAG

     GlyGlyIelGlyGlyTyrSerAlaGlyGluArgIleValAspIleIleAlaThrAspIleG
4348 GGGGGATTGGGGGATACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATAC

      lnThrLysGluLeuGlnLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArg 943
      AAACTAAAGAACTACAAAAGCAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGG
```

FIG. 2D

```
      AspAsnLysAspProLeuTrpLysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAlaV
4468 ACAACAAAGATCCCCTTTGGAAAGGACCAGCAAAGCTTCTCTGGAAAGGTGAAGGGGCAG

     alValIleGlnAspAsnSerAspIleLysValValProArgArgLysAlaLysIleIle 983
     TAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAAGAAGAAAAGCAAAAATCATTA

    ArgAspTyrGlyLysGlnMetAlaGlyAspAspCysValAlaSerArgGlnAspGluAsnA
4588 GGGATTATGGAAAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATT

     M
     AGAACATGGAAAAGTTTAGTAAAACACCATATGTATATTTCAAAGAAAGCTAAAGGATGG

4708 TTTTATAGACATCACTATGAAAGTACTCATCCAAGAGTAAGTTCAGAAGTACACATCCCC

     CTAGGGGATGCTAAATTGGTAATAACAACATATTGGGGTCTGCATACAGGAGAAAGAGAA

4828 TGGCATTTGGGCCAGGGAGTCGCCATAGAATGGAGGAAAAAGAAATATAGCACACAAGTA

     GACCCTGGCCTAGCAGACCAACTAATTCATCTGCATTATTTTGATTGTTTTTCAGAATCT

4948 GCTATAAAAAATGCCATATTAGGATATAGAGTTAGTCCTAGGTGTGAATATCAAGCAGGA

     CATAACAAGGTAGGATCTCTACAATACTTGGCACTAGCAGCATTAATAACACCAAAAAAG

5068 ACAAAGCCACCTTTGCCTAGTGTTAAGAAACTGACAGAGGATAGATGGAACAAGCCCCAG

     AAGACCAAGGGCCACAGAGGGAGCCATACAATGAATGGACACTAGAGCTTTTAGAGGAGC

5188 TTAAGAGAGAAGCTGTTAGACATTTTCCTAGGCCATGGCTCCATAGCTTAGGACAATATA

     TCTATGAAACTTATGGGGATACTTGGGCAGGAGTGGAAGCCATAATAAGAATTCTGCAAC

5308 AACTGCTGTTTATTCATTTCAGAATTGGGTGTCAACATAGCAGAATAGGCATTATTCAAC

     AGAGGAGAGCAAGAAGAAATGGAGCCAGTAGATCCTAATCTAGAGCCCTGGAAGCATCCA

5428 GGAAGTCAGCCTAGGACTGCTTGTAACAATTGCTATTGTAAAAAGTGTTGCTTTCATTGC

     TACGCGTGTTTCACAAGAAAAGGCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAG

5548 CGACGAAGAGCTCCTCAGGACAGTCAGACTCATCAAGCTTCTCTATCAAAGCAGTAAGTA

     GTAAATGTAATGCAATCTTTACAAATATTAGCAATAGTATCATTAGTAGTAGTAGCAATA

5668 ATAGCAATAGTTGTGTGGACCATAGTACTCATAGAATATAGGAAAATATTAAGACAAAGA

                                                        MetLysVal   3
     AAATAGACAGATTAATTGATAGAATAAGAGAAAAAGCAGAAGACAGTGGCAATGAAAGTG

     LysGlyThrArgArgAsnTyrGlnHisLeuTrpArgTrpGlyThrLeuLeuLeuGlyMet
5788 AAGGGGACCAGGAGGAATTATCAGCACTTGTGGAGATGGGGCACCTTGCTCCTTGGGATG

     LeuMetIleCysSerAlaThrGluLysLeuTrpValThrValTyrTyrGlyValProVal  43
     TTGATGATCTGTAGTGCTACAGAAAAATTGTGGGTCACAGTTTATTATGGAGTACCTGTG

     TrpLysGluAlaThrThrThrLeuPheCysAlaSerAspAlaArgAlaTyrAspThrGlu
5908 TGGAAAGAAGCAACTACCACTCTATTTTGTGCATCAGATGCTAGAGCATATGATACAGAG

     ValHisAsnValTrpAlaThrHisAlaCysValProThrAspProAsnProGlnGluVal  83
     GTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTA
```

FIG. 2E

```
     ValLeuGlyAsnValThrGluAsnPheAsnMetTrpLysAsnAsnMetValGluGlnMet
6028 GTATTGGGAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACATGGTAGAACAGATG

     GlnGluAspIleIleSerLeuTrpAspGlnSerLeuLysProCysValLysLeuThrPro 123
     CAGGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCA

     LeuCysValThrLeuAsnCysThrAspLeuGlyLysAlaThrAsnThrAsnSerSerAsn
6148 CTCTGTGTTACTTTAAATTGCACTGATTTGGGGAAGGCTACTAATACCAATAGTAGTAAT

     TrpLysGluGluIleLysGlyGluIleLysAsnCysSerPheAsnIleThrThrSerIle 163
     TGGAAAGAAGAAATAAAAGGAGAAATAAAAAACTGCTCTTTCAATATCACCACAAGCATA

     ArgAspLysIleGlnLysGluAsnAlaLeuPheArgAsnLeuAspValValProIleAsp
6268 AGAGATAAGATTCAGAAAGAAAATGCACTTTTTCGTAACCTTGATGTAGTACCAATAGAT

     AsnAlaSerThrThrThrAsnTyrThrAsnTyrArgLeuIleHisCysAsnArgSerVal 203
     AATGCTAGTACTACTACCAACTATACCAACTATAGGTTGATACATTGTAACAGATCAGTC

     IleThrGlnAlaCysProLysValSerPheGluProIleProIleHisTyrCysThrPro
6388 ATTACACAGGCCTGTCCAAAGGTATCATTTGAGCCAATTCCCATACATTATTGTACCCCG

     AlaGlyPheAlaIleLeuLysCysAsnAsnLysThrPheAsnGlyLysGlyProCysThr 243  E
     GCTGGTTTTGCGATTCTAAAGTGTAATAATAAAACGTTCAATGGAAAAGGACCATGTACA

     AsnValSerThrValGlnCysThrHisGlyIleArgProIleValSerThrGlnLeuLeu
6508 AATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAATAGTGTCAACTCAACTGCTG

     LeuAsnGlySerLeuAlaGluGluGluValValIleArgSerAspAsnPheThrAsnAsn 283  N
     TTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGACAATTTCACGAACAAT

     AlaLysThrIleIleValGlnLeuAsnGluSerValAlaIleAsnCysThrArgProAsn
6628 GCTAAAACCATAATAGTACAGCTGAATGAATCTGTAGCAATTAACTGTACAAGACCCAAC

     AsnAsnThrArgLysSerIleTyrIleGlyProGlyArgAlaPheHisThrThrGlyArg 323  V
     AACAATACAAGAAAAAGTATCTATATAGGACCAGGGAGAGCATTTCATACAACAGGAAGA

     IleIleGlyAspIleArgLysAlaHisCysAsnIleSerArgAlaGlnTrpAsnAsnThr
6748 ATAATAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGAGCACAATGGAATAACACT

     LeuGluGlnIleValLysLysLeuArgGluGlnPheGlyAsnAsnLysThrIleValPhe 363
     TTAGAACAGATAGTTAAAAAATTAAGAGAACAGTTTGGGAATAATAAAACAATAGTCTTT

     AsnGlnSerSerGlyGlyAspProGluIleValMetHisSerPheAsnCysArgGlyGlu
6868 AATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTAGAGGGGAA

     PhePheTyrCysAsnThrThrGlnLeuPheAsnAsnThrTrpArgLeuAsnHisThrGlu 403
     TTTTTCTACTGTAATACAACACAACTGTTTAATAATACATGGAGGTTAAATCACACTGAA

     GlyThrLysGlyAsnAspThrIleIleLeuProCysArgIleLysGlnIleIleAsnMet
6988 GGAACTAAAGGAAATGACACAATCATACTCCCATGTAGAATAAAACAAATTATAAACATG

     TrpGlnGluValGlyLysAlaMetTyrAlaProProIleGlyGlyGlnIleSerCysSer 443
     TGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGGAGGACAAATTAGTTGTTCA

     SerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyThrAsnValThrAsnAspThr
7108 TCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTACAAATGTAACTAATGACACC

     GluValPheArgProGlyGlyGlyAspMetArgAspAsnTrpArgSerGluLeuTyrLys 483
     GAGGTCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAA
```

FIG. 2F

```
     TyrLysValIleLysIleGluProLeuGlyIleAlaProThrLysAlaLysArgArgVal
7228 TATAAAGTAATAAAAATTGAACCATTAGGAATAGCACCCACCAAGGCAAAGAGAAGAGTG

     ValGlnArgGluLysArgAlaValGlyIleValGlyAlaMetPheLeuGlyPheLeuGly 523
     GTGCAGAGAGAAAAAAGAGCAGTGGGAATAGTAGGAGCTATGTTCCTTGGGTTCTTGGGA

     AlaAlaGlySerThrMetGlyAlaValSerLeuThrLeuThrValGlnAlaArgGlnLeu
7348 GCAGCAGGAAGCACTATGGGCGCAGTGTCATTGACGCTGACGGTACAGGCCAGACAATTA

     LeuSerGlyIleValGlnGlnGlnAsnAsnLeuLeuArgAlaIleGluAlaGlnGlnHis 563
     TTGTCTGGTATAGTGCAACAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAACAT

     LeuLeuGlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgValLeuAlaValGlu
7468 CTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAGTCCTGGCTGTGGAA

     ArgTyrLeuArgAspGlnGlnLeuLeuGlyIleTrpGlyCysSerGlyLysLeuIleCys 603
     AGATACCTAAGGGATCAACAGCTCCTAGGGATTTGGGGTTGCTCTGGAAAACTCATTTGC

     ThrThrAlaValProTrpAsnAlaSerTrpSerAsnLysSerLeuGluAspIleTrpAsp
7588 ACCACTGCTGTGCCTTGGAATGCTAGTTGGAGTAATAAATCTCTGGAAGACATTTGGGAT

     AsnMetThrTrpMetGlnTrpGluArgGluIleAspAsnTyrThrAsnThrIleTyrThr 643
     AACATGACCTGGATGCAGTGGGAAAGAGAAATTGACAATTACACAAACACAATATACACC

     LeuLeuGluGluSerGlnAsnGlnGlnGluLysAsnGluGlnGluLeuLeuGluLeuAsp
7708 TTACTTGAAGAATCGCAGAACCAACAAGAAAAGAATGAACAAGAATTATTAGAATTGGAT

     LysTrpAlaSerLeuTrpAsnTrpPheSerIleThrAsnTrpLeuTrpTyrIleLysIle 683
     AAGTGGGCAAGTTTGTGGAATTGGTTTAGCATAACAAACTGGCTGTGGTATATAAAGATA

     PheIleMetIleValGlyGlyLeuValGlyLeuArgIleValPheAlaValLeuSerIle     E
7828 TTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAGAATAGTTTTTGCTGTGCTTTCTATA

     ValAsnArgValArgGlnGlyTyrSerProLeuSerPheGlnThrArgLeuProValPro 723
     GTGAATAGAGTTAGGCAGGGATACTCACCATTGTCATTTCAGACCCGCCTCCCAGTCCCG     N

     ArgGlyProAspArgProAspGlyIleGluGluGluGlyGlyGluArgAspArgAspArg
7948 AGGGGACCCGACAGGCCCGACGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGA

     SerValArgLeuValAspGlyPheLeuAlaLeuIleTrpGluAspLeuArgSerLeuCys 763 V
     TCCGTTCGATTAGTGGATGGATTCTTAGCACTTATCTGGGAAGATCTGCGGAGCCTGTGC

     LeuPheSerTyrArgArgLeuArgAspLeuLeuLeuIleAlaAlaArgThrValGluIle
8068 CTCTTCAGCTACCGCCGCTTGAGAGACTTACTCTTGATTGCAGCGAGGACTGTGGAAATT

     LeuGlyHisArgGlyTrpGluAlaLeuLysTyrTrpTrpSerLeuLeuGlnTyrTrpIle 803
     CTGGGGCACAGGGGGTGGGAAGCCCTCAAATATTGGTGGAGTCTCCTGCAGTATTGGATT

     GlnGluLeuLysAsnSerAlaValSerTrpLeuAsnAlaThrAlaIleAlaValThrGlu
8188 CAGGAACTAAAGAATAGTGCTGTTAGCTGGCTCAACGCCACAGCTATAGCAGTAACTGAG

     GlyThrAspArgValIleGluValAlaGlnArgAlaTyrArgAlaIleLeuHisIleHis 843
     GGGACAGATAGGGTTATAGAAGTAGCACAAAGAGCTTATAGAGCTATTCTCCACATACAT

     ArgArgIleArgGlnGlyLeuGluArgLeuLeuLeuOC
8308 AGAAGAATTAGACAGGGCTTGGAAAGGCTTTTGCTATAAGATGGGTGGCAAGTGGTCAAA

     ACGTAGTATGGGTGGATGGTCTGCTATAAGGGAAAGAATGAGACGAGCTGAGCCACGAGC
```

FIG. 2G

```
8428 TGAGCCAGCAGCAGATGGGGTGGGAGCAGTATCTCGAGACCTGGAAAAACATGGAGCAAT
     CACAAGTAGCAATACAGCAGCTACTAATGCTGATTGTGCCTGGCTAGAAGCACAAGAGGA
8548 GGAAGAGGTGGGTTTTCCAGTCAGACCTCAGGTACCTTTAAGACCAATGACTTACAAGGC
                                              U3──►
     AGCTTTAGATATTAGCCACTTTTTAAAAGAAAAGGGGGGA CTGGAAGGGCTAATTTGGT     L
8667 CCCAAAGAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTCCCTG
     ATTGGCAGAATTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCT     T
8787 TCAAGCTAGTACCAGTTGAGCCAGAGAAGGTAGAAGAGGCCAATGAAGGAGAGAACAACA
     GCTTGTTACACCCTATGAGCCTGCATGGGATGGAGGACGCGGAGAAAGAAGTGTTAGTGT     R
8907 GGAGGTTTGACAGCAAACTAGCATTTCATCACATGGCCCGAGAGCTGCATCCGGAGTACT
     ACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGA
9027 GGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCAGATGCTGCATATAAGCAGCTGC
          ◄──U3 R──►
     TTTTTGCCTGTACTG GGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGC
                                                     ◄─R U5─►
9146 TAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTTCA AGTAGT
     GTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGT
                   ◄──U5
9265 GTGGAAAAATCTCTAGCAG
```

FIG. 2H

FIG. 3

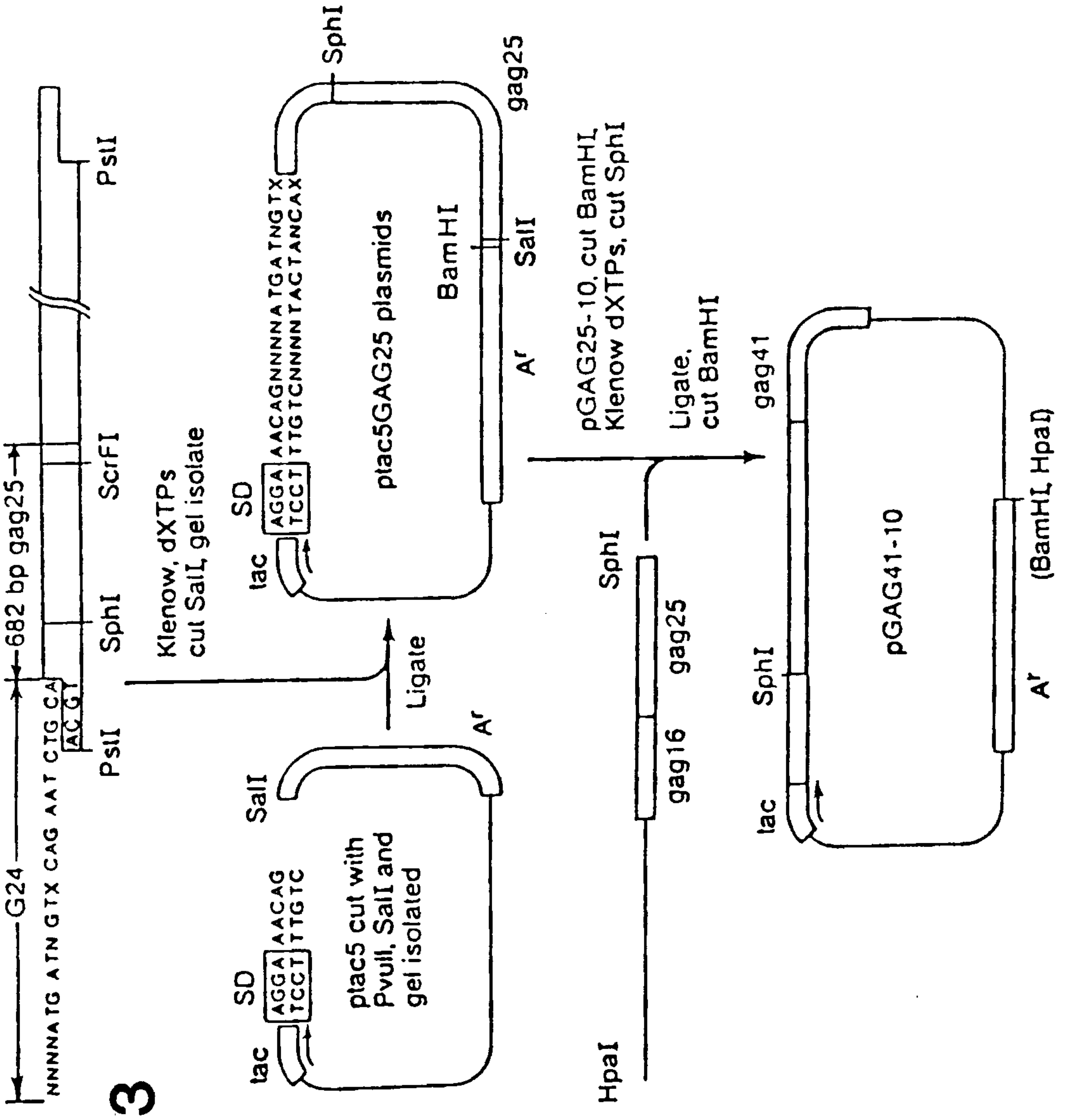
G24
682 bp gag25
NNNNATG ATN GTX CAG AAT CTG CA
ACGT
PstI
SphI
ScrFI
SalI
PstI
Klenow, dXTPs
cut SalI, gel isolate
Ligate
tac
SD
AGGA
TCCT
AACAG
TTGTC
ptac5 cut with
PvuII, SalI and
gel isolated
SalI
A^r
AACAGNNNNATGATNGTX
TTGTCNNNNTACTANCAX
ptac5GAG25 plasmids
SphI
gag25
BamHI
SalI
pGAG25-10, cut BamHI,
Klenow dXTPs, cut SphI
Ligate,
cut BamHI
HpaI
gag16
gag25
SphI
pGAG41-10
gag41
(BamHI, HpaI)

FIG. 4

```
                                                                                        *
                                                               ptac 5 Promotor        MetProIleVal  141
                                                                                      ATGCCTATCGTA

     GlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeu  181
748  CAGAATCTGCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGGGTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTA

     SerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgVal  221
868  TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCAGAATGGGATAGAGTG

     HisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProPro  261
988  CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGGGGAAGTGACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCT

     IleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAsp  301
1108 ATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGAT

     TyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsnTrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLys  341
1228 TATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCTTCACAGGATGTAAAAAATTGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAA

     AlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeu Stop Stop   ptac 5
1348 GCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGACCCGGGCATAAAGCAAGAGTTTTGTGATAG
```

```
                                                            MetIleVal 141
                          ptac 5 Promotor                   ATGATCGTA
     GlnAsnLeuGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrp
 748 CAGAATCTGCAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGG

     ValLysValValGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeu 181
     GTAAAAGTAGTAGAAGAAAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTA    G

     SerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGln
 868 TCAGAAGGAGCCACCCCACAAGATTTAAACACCATGCTAAACACAGTGGGGGGACATCAA

     AlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAlaAlaGluTrpAspArgVal 221
     GCAGCCATGCAAATGTTAAAAGAGACTATCAATGAGGAAGCTGCAGAATGGGATAGAGTG

     HisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAsp-
 988 CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGGGGAAGTGAC    A

     IleAlaGlyThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrAsnAsnProPro 261
     ATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCT

     IleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArg
1108 ATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGA
                                                                     G
     MetTyrSerProThrSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAsp 301
     ATGTATAGCCCTACCAGCATTCTGGACATAAGACAAGGACCAAAGGAACCCTTTAGAGAT

     TyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnAspValLysAsn
1228 TATGTAGACCGGTTCTATAAAACTCTAAGAGCCGAACAAGCTTCACAGGATGTAAAAAAT

     TrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLys 341
     TGGATGACAGAAACCTTGTTGGTCCAAAATGCAAACCCAGATTGTAAGACTATTTTAAAA

     AlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGly
1348 GCATTGGGACCAGCAGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGA

     ProGlyHisLysAlaArgValLeuAlaGluAlaMetSerGlnValThrAsnProAlaAsn 381
     CCCGGCCATAAAGCAAGAGTTTTGGCTGAAGCCATGAGCCAAGTAACAAATCCAGCTAAC

     IleMetMetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCys
1468 ATAATGATGCAGAGAGGCAATTTTAGGAACCAAAGAAAGACTGTTAAGTGTTTCAATTGT

     GlyLysGluGlyHisIleAlaLysAsnCysArgAlaProArgLysLysGlyCysTrpArg 421
     GGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAGA

     CysGlyArgGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGly
                                                          PhePheArgG
1588 TGTGGAAGGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGG

     LysIleTrpProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluPro 461
     luAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAla 23
     AAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCA

     ThrAlaProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGlnLys
   AsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSerLeuSerGluA
1708 ACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAG

     GlnGluProIleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsn 501
     laGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGln 63
     CAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAAC
```

FIG. 5A

```
      AspProSerSerGlnOC
     ArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyA
1828 GACCCCTCGTCACAATAAGGATAGGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAG

      laAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLysMetIle 103
      CAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGATAG

     GlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProValGluIleCysG
1948 GGGGAATTGGAGGTTTTATCAAAGTAAGACAGTACGATCAGATACCTGTAGAAATCTGTG

      lyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArg 143
      GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA

     AsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrValP
2068 ATCTGTTGACTCAGATTGGTTGTACTTTAAATTTCCCCATTAGTCCTATTGAAACTGTAC

      roValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 183
      CAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTTAAGCAATGGCCATTGACAGAAG

     GluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerL
2188 AAAAAATAAAAGCATTAGTAGAGATATGTACAGAAATGGAAAAGGAAGGGAAAATTTCAA

      ysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 223
      AAATTGGGCCTGAAAATCCATACAATACTCCAGTATTTGCTATAAAGAAAAAAGACAGTA

     ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpG
2308 CTAAATGGAGAAAACTAGTAGATTTCAGAGAACTTAATAAAAGAACTCAAGACTTCTGGG

      luValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrVal 263
      AAGTTCAGTTAGGAATACCACACCCCGCAGGGTTAAAAAAGAAAAAATCAGTAACAGTAT

     LeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLysTyrThrA
2428 TGGATGTGGGTGATGCATACTTTTCAGTTCCCTTAGATAAAGACTTTAGAAAGTATACTG

      laPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 303
      CATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGC

     LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuG
2548 TGCCACAGGGATGGAAAGGATCACCAGCAATATTCCAAAGTAGCATGACAAAAATCTTAG

      luProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyr 343
      AGCCTTTTAGAAAACAGAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATG

     ValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisL
2668 TAGGATCTGACTTAGAAATAGGGCAGCATAGAACAAAAATAGAGGAACTGAGACAGCATC

      euLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 383
      TGTTGAGGTGGGGATTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTTT

     TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMetLeuProGluL
2788 GGATGGGTTATGAACTCCATCCTGATAAATGGACAGTACAGCCTATAATGCTGCCAGAAA

      ysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 423
      AAGACAGCTGGACTGTCAATGACATACAGAAGTTAGTGGGAAAATTGAATTGGGCAAGTC

     GlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThrLysAlaL
2908 AGATTTATGCAGGGATTAAAGTAAAGCAGTTATGTAAACTCCTTAGAGGAACCAAAGCAC

      euThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 463  P
      TAACAGAAGTAATACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAAAACAGGGAGA     O
                                                                       L
```

FIG. 5B

```
     IleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeuValAlaGluI
3028 TTCTAAAAGAACCAGTACATGAAGTATATTATGACCCATCAAAAGACTTAGTAGCAGAAA

      leGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 503
      TACAGAAGCAGGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTAAAAATC

     LeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuT
3148 TGAAAACAGGAAAGTATGCAAGGATGAGGGGTGCCCACACTAATGATGTAAAACAGTT

      hrGluAlaValGluLysValSerThrGluSerIleValIleTrpGlyLysIleProLys 543
         ptac 5
```

FIG. 5C

FIG. 6

ARV GAG p16 - synthetic Parts A and B

```
5'                         3'
        MetGlnArgGlyAsnPheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLys
    TATTATGCAAAGAGGTAACTTCAGGAATCAAAGAAAGACCGTTAAGTGTTTCAACTGTGGTAAG
    ATAATACGTTTCTCCATTGAAGTCCTTAGTTTCTTTCTGGCAATTCACAAAGTTGACACCATTC
3'                          ^ 5'
        10 mnl1, 23 hinf1,

        GluGlyHisIleAlaLysAsnCysArgAlaProArgLysLysAlaCysTrpArgCysGly
   63   GAAGGTCACATCGCTAAGAACTGTAGAGCTCCAAGAAAGAAGGCTTGTTGGAGATGTGGT
        CTTCCAGTGTAGCGATTCTTGACATCTCGAGGTTCTTTCTTCCGAACAACCTCTACACCA
                     ^           ^^
        76 dde1, 88 ban2 hgiA hgiJ11 sac1 sduI, 89 alu1,

        ArgGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIle
  123   AGAGAAGGTCACCAAATGAAGGACTGTACCGAAAGACAAGCTAACTTCTTGGGTAAGATC
        TCTCTTCCAGTGGTTTACTTCCTGACATGGCTTTCTGTTCGATTGAAGAACCCATTCTAG
              ^ ^                ^            ^                ^^
        129 bstE2, 131 hph, 148 rsaI, 161 alu1, 178 bgl11 xho2, 179
        sau3a,

        TrpProSerTyrLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAla
  183   TGGCCATCTTACAAGGGTAGACCAGGTAACTTCTTGCAATCCAGACCAGAACCAACCGCT
        ACCGGTAGAATGTTCCCATCTGGTCCATTGAAGAACGTTAGGTCTGGTCTTGGTTGGCGA
        ^^              ^    ^
        183 bal1 cfr1 hae1, 184 hae111, 199 acc1, 204 apy1 ecor11 sc
        rF1,

        ProProGluGluSerPheArgPheGlyGluGluLysThrThrProSerGlnLysGlnGlu
  243   CCACCTGAAGAAAGTTTCAGGTTCGGTGAAGAAAAGACCACCCCATCTCAAAAGCAAGAA
        GGTGGACTTCTTTCAAAGTCCAAGCCACTTCTTTTCTGGTGGGGTAGAGTTTTCGTTCTT
              ^                 ^  ^
        249 mbo11, 267 hph, 270 mbo11,

        ProIleAspLysGluLeuTyrProLeuThrSerLeuArgSerLeuPheGlyAsnAspPro
  303   CCAATCGACAAGGAATTGTACCCATTGACCTCTTTGAGATCCTTGTTCGGTAACGATCCC
        GGTTAGCTGTTCCTTAACATGGGTAACTGGAGAAACTCTAGGAACAAGCCATTGCTAGGG
            ^            ^          ^        ^^               ^   ^^
        307 taq1, 320 rsaI, 331 mnl1, 339 xho2, 340 sau3a, 357 sau3a
        , 361 mnl1, 362 ava1 xho1,

        SerSerGlnOP AM  |
  363   TCGAGCCAATGATAG
        AGCTCGGTTACTATCAGCT
          ^             ^
        363 taq1, 377 acc1 hind11 sal1
```

```
                   PYK Promoter                          MetSer
                                                         ATGTCT
     ArgIleAspCysSerAlaThrGluLysLeuTrpValThrValTyrTyrGlyValProVal 51
     AGAATCGAT GTAGTGCTACAGAAAAATTGTGGGTCACAGTTTATTATGGAGTACCTGTG

     TrpLysGluAlaThrThrThrLeuPheCysAlaSerAspAlaArgAlaTyrAspThrGlu
5908 TGGAAAGAAGCAACTACCACTCTATTTTGTGCATCAGATGCTAGAGCATATGATACAGAG

     ValHisAsnValTrpAlaThrHisAlaCysValProThrAspProAsnProGlnGluVal 91
     GTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCACAAGAAGTA

     ValLeuGlyAsnValThrGluAsnPheAsnMetTrpLysAsnAsnMetValGluGlnMet
6028 GTATTGGGAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACATGGTAGAACAGATG

     GlnGluAspIleIleSerLeuTrpAspGlnSerLeuLysProCysValLysLeuThrPro 131
     CAGGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAATTAACCCCA

     LeuCysValThrLeuAsnCysThrAspLeuGlyLysAlaTheAsnThrAsnSerSerAsn
6148 CTCTGTGTTACTTTAAATTGCACTGATTTGGGGAAGGCTACTAATACCAATAGTAGTAAT

     TrpLysGluGluIleLysGlyGluIleLysAsnCysSerPheAsnIleThrThrSerIle 171
     TGGAAAGAAGAAATAAAAGGAGAAATAAAAAACTGCTCTTTCAATATCACCACAAGCATA

     ArgAspLysIleGlnLysGluAsnAlaLeuPheArgAsnLeuAspValValProIleAsp
6268 AGAGATAAGATTCAGAAAGAAAATGCACTTTTTCGTAACCTTGATGTAGTACCAATAGAT

     AsnAlaSerThrThrThrAsnTyrThrAsnTyrArgLeuIleHisCysAsnArgSerVal 211
     AATGCTAGTACTACTACCAACTATACCAACTATAGGTTGATACATTGTAACAGATCAGTC

     IleThrGlnAlaCysProLysValSerPheGluProIleProIleHisTyrCysThrPro
6388 ATTACACAGGCCTGTCCAAAGGTATCATTTGAGCCAATTCCCATACATTATTGTACCCCG

     AlaGlyPheAlaIleLeuLysCysAsnAsnLysThrPheAsnGlyLysGlyProCysThr 251
     GCTGGTTTTGCGATTCTAAAGTGTAATAATAAAACGTTCAATGGAAAAGGACCATGTACA
```

FIG. 7A

```
     AsnValSerThrValGlnCysThrHisGlyIleArgProIleValSerThrGlnLeuLeu
6508 AATGTCAGCACAGTACAATGTACACATGGAATTAGGCCAATAGTGTCAACTCAACTGCTG

     LeuAsnGlySerLeuAlaGluGluGluValValIleArgSerAspAsnPheThrAsnAsn 291
     TTAAATGGCAGTCTAGCAGAAGAAGAGGTAGTAATTAGATCTGACAATTTCACGAACAAT

     AlaLysThrIleIleValGlnLeuAsnGluSerValAlaIleAsnCysThrArgProAsn
6628 GCTAAAACCATAATAGTACAGCTGAATGAATCTGTAGCAATTAACTGTACAAGACCCAAC

     AsnAsnThrArgLysSerIleTyrIleGlyProGlyArgAlaPheHisThrThrGlyArg 331
     AACAATACAAGAAAAAGTATCTATATAGGACCAGGGAGAGCATTTCATACAACAGGAAGA

     IleIleGlyAspIleArgLysAlaHisCysAsnIleSerArgAlaGlnTrpAsnAsnThr
6748 ATAATAGGAGATATAAGAAAAGCACATTGTAACATTAGTAGAGCACAATGGAATAACACT

     LeuGluGlnIleValLysLysLeuArgGluGlnPheGlyAsnAsnLysThrIleValPhe 371
     TTAGAACAGATAGTTAAAAAATTAAGAGAACAGTTTGGGAATAATAAAACAATAGTCTTT

     AsnGlnSerSerGlyGlyAspProGluIleValMetHisSerPheAsnCysArgGlyGlu
6868 AATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTAGAGGGGAA
                                                                      E
     PhePheTyrCysAsnThrThrGlnLeuPheAsnAsnThrTrpArgLeuAsnHisThrGlu 411
     TTTTTCTACTGTAATACAACACAACTGTTTAATAATACATGGAGGTTAAATCACACTGAA

     GlyThrLysGlyAsnAspThrIleIleLeuProCysArgIleLysGlnIleIleAsnMet
6988 GGAACTAAAGGAAATGACACAATCATACTCCCATGTAGAATAAAACAAATTATAAACATG
                                                                      N
     TrpGlnGluValGlyLysAlaMetTyrAlaProProIleGlyGlyGlnIleSerCysSer 451
     TGGCAGGAAGTAGGAAAAGCAATGTATGCCCCTCCCATTGGAGGACAAATTAGTTGTTCA

     SerAsnIleThrGlyLeuLeuLeuThrArgAspGlyGlyThrAsnValThrAsnAspThr
7108 TCAAATATTACAGGGCTGCTATTAACAAGAGATGGTGGTACAAATGTAACTAATGACACC
                                                                      V
     GluValPheArgProGlyGlyGlyAspMetArgAspAsnTrpArgSerGluLeuTyrLys 491
     GAGGTCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAA

     TyrLysValIleLysIleGluProAsnSerValSer                 PYK Terminator
7228 TATAAAGTAATAAAAATTGAACCAAATTCGGTATCTTGA
```

FIG. 7B

FIG. 8-1

```
Nucleotide
Positions
Relative to
Figure 2            MetIleAspLysAlaGlnGluGluHisGluLysTyrHisSerAsnTrp
              1 AGGXAACAG::::ATGAT:GA:AAGGCACAAGAAGAACATGAGAAATATCACAGTAATTGG
                TCCXTTGTC::::TACTA:CT:TTCCGTGTTCTTCTTGTACTCTTTATAGTGTCATTAACC

                32 mbo11, 38 nla111,

                ArgAlaMetAlaSerAspPheAsnLeuProProValValAlaLysGluIleValAlaSer
3820         62 AGAGCCATGGCTAGTGATTTTAACCTGCCACCTGTAGTAGCAAAAGAAATAGTAGCCAGC
                TCTCGGTACCGATCACTAAAATTGGACGGTGGACATCATCGTTTTCTTTATCATCGGTCG

                66 nco1, 67 nla111, 118 nspBII pvu11, 119 alu1,

                CysAspLysCysGlnLeuLysGlyGluAlaMetHisGlyGlnValAspCysSerProGly
3880        122 TGTGATAAATGTCAGCTAAAAGGAGAAGCCATGCATGGACAAGTAGACTGTAGTCCAGGA
                ACACTATTTACAGTCGATTTTCCTCTTCGGTACGTACCTGTTCATCTGACATCAGGTCCT

                135 alu1, 151 nla111, 152 nsi1 ava3, 155 nla111, 164 acc1, 1
                76 apy1 bstXI ecor11 scrF1,

                IleTrpGlnLeuAspCysThrHisLeuGluGlyLysIleIleLeuValAlaValHisVal
3940        182 ATATGGCAACTAGATTGTACACATCTAGAAGGAAAAATTATCCTGGTAGCAGTTCATGTA
                TATACCGTTGATCTAACATGTGTAGATCTTCCTTTTTAATAGGACCATCGTCAAGTACAT

                198 rsaI, 205 xba1, 223 apy1 ecor11 scrF1, 236 nla111,

                AlaSerGlyTyrIleGluAlaGluValIleProAlaGluThrGlyGlnGluThrAlaTyr
4000        242 GCCAGTGGATATATAGAAGCAGAAGTTATTCCAGCAGAGACAGGGCAGGAAACAGCATAT
                CGGTCACCTATATATCTTCGTCTTCAATAAGGTCGTCTCTGTCCCGTCCTTTGTCGTATA

                263 xmn1,

                PheLeuLeuLysLeuAlaGlyArgTrpProValLysThrIleHisThrAspAsnGlySer
4060        302 TTTCTCTTAAAATTAGCAGGAAGATGGCCAGTAAAAACAATACATACAGACAATGGCAGC
                AAAGAGAATTTTAATCGTCCTTCTACCGGTCATTTTTGTTATGTATGTCTGTTACCGTCG

                321 mbo11, 326 bal1 cfr1 hae1, 327 hae111, 357 bbv fnu4h1,

                AsnPheThrSerThrThrValLysAlaAlaCysTrpTrpAlaGlyIleLysGlnGluPhe
4120        362 AATTTCACCAGTACTACGGTTAAGGCCGCCTGTTGGTGGGCAGGGATCAAGCAGGAATTT
                TTAAAGTGGTCATGATGCCAATTCCGGCGGACAACCACCCGTCCCTAGTTCGTCCTTAAA

                366 hph, 371 sca1, 372 rsaI, 385 hae111, 386 fnu4h1 nsb11, 4
                05 binI, 406 dpn1 sau3a,

                GlyIleProTyrAsnProGlnSerGlnGlyValValGluSerMetAsnAsnGluLeuLys
4180        422 GGCATTCCCTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAATGAATTAAAG
                CCGTAAGGGATGTTAGGGGTTTCAGTTCCTCATCATCTTAGATACTTATTACTTAATTTC

                423 bsm1, 458 hinf1,

                LysIleIleGlyGlnValArgAspGlnAlaGluHisLeuLysThrAlaValGlnMetAla
4240        482 AAAATTATAGGACAGGTAAGAGATCAGGCTGAACACCTTAAGACAGCAGTACAAATGGCA
                TTTTAATATCCTGTCCATTCTCTAGTCCGACTTGTGGAATTCTGTCGTCATGTTTACCGT

                503 dpn1 sau3a, 518 af111, 530 rsaI,

                ValPheIleHisAsnPheLysArgLysGlyGlyIleGlyGlyTyrSerAlaGlyGluArg
4300        542 GTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGGGGATACAGTGCAGGGGAAAGA
                CATAAGTAGGTGTTAAAATTTTCTTTTCCCCCCTAACCCCCTATGTCACGTCCCCTTTCT

                547 fok1, 557 aha111,

                IleValAspIleIleAlaThrAspIleGlnThrLysGluLeuGlnLysGlnIleThrLys
4360        602 ATAGTAGACATAATAGCAACAGACATACAAACTAAAGAACTACAAAAGCAAATTACAAAA
                TATCATCTGTATTATCGTTGTCTGTATGTTTGATTTCTTGATGTTTTCGTTTAATGTTTT

                605 acc1,

                IleGlnAsnPheArgValTyrTyrArgAspAsnLysAspProLeuTrpLysGlyProAla
4420        662 ATTCAAAATTTTCGGGTTTATTACAGGGACAACAAAGATCCCCTTTGGAAAGGACCAGCA
                TAAGTTTTAAAAGCCCAAATAATGTCCCTGTTGTTTCTAGGGGAAACCTTTCCTGGTCGT

                697 xho2, 698 dpn1 sau3a, 713 asu1 ava2,
```

```
4480     722   LysLeuLeuTrpLysGlyGluGlyAlaValValIleGlnAspAsnSerAspIleLysVal
               AAGCTTCTCTGGAAAGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTA
               TTCGAAGAGACCTTTCCACTTCCCCGTCATCATTATGTTCTATTATCACTGTATTTTCAT

               722 hind111, 723 alu1, 737 hph,

4540     782   ValProArgArgLysAlaLysIleIleArgAspTyrGlyLysGlnMetAlaGlyAspAsp
               GTGCCAAGAAGAAAAGCAAAAATCATTAGGGATTATGGAAAACAGATGGCAGGTGATGAT
               CACGGTTCTTCTTTTCGTTTTTAGTAATCCCTAATACCTTTTGTCTACCGTCCACTACTA

               789 mbo11, 833 hph,

4600     842   CysValAlaSerArgGlnAspGluAspAM
               TGTGTGGCAAGTAGACAGGATGAGGATTAGTCGACGGAATTCTTTAGTAAAACACC
               ACACACCGTTCATCTGTCCTACTCCTAATCAGCTGCCTTAAGAAATCATTTTGTGG

               852 acc1, 859 fok1, 863 mnl1, 871 acc1 hind11 sal1, 872 taq1
               , 878 ecor1,
```

FIG. 8-2