MOLECULAR CLONES OF HIV-1 AND USES THEREOF

The present invention relates to the HIV-1 strains MN-ST1 and BA-L which are typical United States HIV-1 isotypes. The present invention relates to DNA segments encoding the envelope protein of MN-ST1 or BA-L, to DNA constructs containing such DNA segments and to host cells transformed with such constructs. The viral isolates and envelope proteins of the present invention are of value for use in vaccines and bioassays for the detection of HIV-1 infection in biological samples, such as blood bank samples.
### FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<table>
<thead>
<tr>
<th>Code</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Austria</td>
</tr>
<tr>
<td>AU</td>
<td>Australia</td>
</tr>
<tr>
<td>BB</td>
<td>Barbados</td>
</tr>
<tr>
<td>BE</td>
<td>Belgium</td>
</tr>
<tr>
<td>BF</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>BG</td>
<td>Bulgaria</td>
</tr>
<tr>
<td>BJ</td>
<td>Benin</td>
</tr>
<tr>
<td>BR</td>
<td>Brazil</td>
</tr>
<tr>
<td>CA</td>
<td>Canada</td>
</tr>
<tr>
<td>CF</td>
<td>Central African Republic</td>
</tr>
<tr>
<td>CG</td>
<td>Congo</td>
</tr>
<tr>
<td>CH</td>
<td>Switzerland</td>
</tr>
<tr>
<td>CI</td>
<td>Côte d'Ivoire</td>
</tr>
<tr>
<td>CM</td>
<td>Cameroon</td>
</tr>
<tr>
<td>CS</td>
<td>Czechoslovakia</td>
</tr>
<tr>
<td>DE*</td>
<td>Germany</td>
</tr>
<tr>
<td>DK</td>
<td>Denmark</td>
</tr>
<tr>
<td>ES</td>
<td>Spain</td>
</tr>
<tr>
<td>FI</td>
<td>Finland</td>
</tr>
<tr>
<td>FR</td>
<td>France</td>
</tr>
<tr>
<td>GA</td>
<td>Gabon</td>
</tr>
<tr>
<td>GB</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>GN</td>
<td>Guinea</td>
</tr>
<tr>
<td>GR</td>
<td>Greece</td>
</tr>
<tr>
<td>HU</td>
<td>Hungary</td>
</tr>
<tr>
<td>IT</td>
<td>Italy</td>
</tr>
<tr>
<td>JP</td>
<td>Japan</td>
</tr>
<tr>
<td>KP</td>
<td>Democratic People's Republic of Korea</td>
</tr>
<tr>
<td>KR</td>
<td>Republic of Korea</td>
</tr>
<tr>
<td>LI</td>
<td>Liechtenstein</td>
</tr>
<tr>
<td>LK</td>
<td>Sri Lanka</td>
</tr>
<tr>
<td>LU</td>
<td>Luxembourg</td>
</tr>
<tr>
<td>MC</td>
<td>Monaco</td>
</tr>
<tr>
<td>MG</td>
<td>Madagascar</td>
</tr>
<tr>
<td>ML</td>
<td>Mali</td>
</tr>
<tr>
<td>MN</td>
<td>Mongolia</td>
</tr>
<tr>
<td>MR</td>
<td>Mauritania</td>
</tr>
<tr>
<td>MW</td>
<td>Malawi</td>
</tr>
<tr>
<td>NL</td>
<td>Netherlands</td>
</tr>
<tr>
<td>NO</td>
<td>Norway</td>
</tr>
<tr>
<td>PL</td>
<td>Poland</td>
</tr>
<tr>
<td>RO</td>
<td>Romania</td>
</tr>
<tr>
<td>SD</td>
<td>Sudan</td>
</tr>
<tr>
<td>SE</td>
<td>Sweden</td>
</tr>
<tr>
<td>SN</td>
<td>Senegal</td>
</tr>
<tr>
<td>SU*</td>
<td>Soviet Union</td>
</tr>
<tr>
<td>TD</td>
<td>Chad</td>
</tr>
<tr>
<td>TG</td>
<td>Togo</td>
</tr>
<tr>
<td>US</td>
<td>United States of America</td>
</tr>
</tbody>
</table>

*Any designation of “SU” has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.*
MOLECULAR CLONES OF HIV-1 AND USES THEREOF

BACKGROUND OF THE INVENTION

HIV-1 has been identified as the etiologic agent of the acquired immunodeficiency syndrome (AIDS) (Barre-Sinoussi et al., Science 220, 868-871, 1983; Popvic et al., Science 224, 497-500, 1984; Gallo et al., Science 224, 500-503, 1984). Infected individuals generally develop antibodies to the virus within several months of exposure (Sarnadharan et al., Science 224, 506-508, 1984), which has made possible the development of immunologically based tests which can identify most blood samples from infected individuals. This is a great advantage in diagnosis, and is vital to maintaining the maximum possible safety of samples from blood banks.

An important aspect of HIV-1 is its genetic variability (Hahn et al., Proc. Natl. Acad. Sci. U.S.A. 82, 4813-4817, 1985). This is particularly evident in the gene for the outer envelope glycoprotein (Starcich et al., Cell 45, 637-648, 1986; Alizon et al., Cell 46, 63-74, 1986; Gurgo et al., Virology 164, 531-536, 1988). Since the outer envelope glycoprotein is on the surface of the virus particle and the infected cell, it is potentially one of the primary targets of the immune system, including the target of neutralizing antibodies and cytotoxic T cells. This variability may also lead to differences in the ability of antigens from different strains of HIV-1 to be recognized by antibodies from a given individual, as well as to differences in the ability of proteins from different strains of virus to elicit an immune response which would be protective against the mixture of virus strains that exists in the at risk populations.

Several biologically active complete molecular clones of various strains of HIV-1 have been obtained and sequenced. These clones, however, seem to represent viral genotypes which are relatively atypical of United States HIV-1 isolates. In addition, several of the translational reading frames for non-structural viral proteins are not complete. Further, viruses derived from these clones do
not grow in macrophages, in contrast to many HIV-1 field isolates and, perhaps, because of this lack of ability to infect macrophages efficiently, these clones do not replicate well in chimpanzees. This latter ability is important for testing candidate vaccines in animal systems. In addition, the ability to infect macrophages is critical in evaluating the possible protective efficacy of elicited immune response since neutralization of infectivity on macrophage may differ from the better studied neutralization on T cells.

Neutralizing antibodies (Robert-Guroff et al., Nature 316, 72-74, 1985; Weiss et al., Nature 316, 69-72, 1985) have been demonstrated in infected individuals, as have cytotoxic T cells responses (Walker et al, Nature 328, 345-348, 1988). Although these do not appear to be protective, it is likely that if they were present prior to infection, they would prevent infection, especially by related strains of virus. This is supported by the finding that macaques can be protected by immunization with inactivated simian immunodeficiency virus (SIV) from infection with the homologous live virus (Murphy-Corb et al., Science 246, 1293-1297, 1989). Chimps also have been passively protected against challenge by live virus by prior administration of neutralizing antibodies to the same virus (Emiri et al., J. Virol. 64, 3674-3678, 1989). One problem, however, is that at least some of the neutralizing antibodies studied depend on recognition of a variable region on the envelope (Matsushita et al., J. Virol. 62, 2107-2114, 1988; Rusche et al., Proc. Natl. Acad. Sci. U.S.A. 85, 3198-3202, 1988; Skinner et al., AIDS Res. Hum. Retroviruses 4, 187-197, 1988) called the V3 region (Starcich et al., Cell 45, 637-648, 1986).

An at least partial solution to the problem of viral heterogeneity is to identify prototypical HIV-1 strains, that is, those that are most similar by DNA sequence data or serologic reactivity to strains present in the population at risk. The inclusion of a limited number of such prototype strains in a polyvalent vaccine
cocktail might then result in elicitation of an immune response protective against most naturally occurring viruses within a given population. Such a mixture should also provide the maximum possible sensitivity in diagnostic tests for antibodies in infected individuals.

Components of highly representative isolates of a geographical area provide the maximum possible sensitivity in diagnostic tests and vaccines. Production of viral proteins from molecular clones by recombinant DNA techniques is the preferred and safest means to provide such proteins. Molecular clones of prototype HIV-1 strains can serve as the material from which such recombinant proteins can be made. The use of recombinant DNA avoids any possibility of the presence of live virus and affords the opportunity of genetically modifying viral gene products. The use of biologically active clones ensures that the gene products are functional and hence, maximizes their potential relevance.

Infectious clones, that is, those which after transfection into recipient cells produce complete virus, are desirable for several reasons. One reason is that the gene products are by definition functional; this maximizes their potential relevance to what is occurring in vivo. A second reason is that genetically altered complete virus is easy to obtain. Consequently, the biological consequences of variability can be easily assessed. For example, the effect of changes in the envelope gene on the ability of the virus to be neutralized by antibody can be easily addressed. Using this technique, a single point mutation in the envelope gene has been shown to confer resistance to neutralizing antibody (Reitz et al., Cell 54, 57-63, 1988). A third reason is that a clonal virus population provides the greatest possible definition for challenge virus in animals receiving candidate vaccines, especially those including components of the same molecularly cloned virus.
SUMMARY OF THE INVENTION

It is an object of the present invention to provide vaccine components for an anti HIV-1 vaccine which would represent a typical United States isolate HIV-1.

It is another object of the present invention to provide diagnostic tests for the detection of HIV-1.

Various other objects and advantages of the present invention will become apparent from the drawings and the following description of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 shows the structure and restriction map of the lambda MN-PH1 clone.

FIGURE 2 shows the restriction map of the MN-PH1 envelope plasmid clone.

FIGURE 3 shows the restriction map and structure of the lambda MN-ST1 clone.

FIGURE 4 shows the structure of the lambda BA-L clone.

FIGURE 5 shows the restriction map of the clone BA-L1.

Detailed Disclosure of the Invention

The present invention relates to the HIV-1 virus strains, MN-ST1 and BA-L, which are more typical of the HIV-1 isolates found in the United States than previously known HIV-1 strains. Local isolates provide better material for vaccine and for the detection of the virus in biological samples, such as blood bank samples.

The present invention relates to DNA segments encoding the env protein of MN-ST1 or BA-L (the DNA sequence given in Figures 5 and 8 being two such examples) and to nucleotide sequences complementary to the segments referenced above as well as to other genes and nucleotide sequences contained in these clones. The present invention also relates to DNA segments encoding a unique portion of the MN-ST1 env protein or the BA-L env protein. (A "unique portion" consists of at least five (or six) amino acids or corresponding at least 15 (or 18) nucleotides.)
The invention further relates to the HIV-1 virus strains MN-ST1 and BA-L themselves. The HIV-1 virus strains of the present invention are biologically active and can easily be isolated by one skilled in the art using known methodologies.

The above-described DNA segments of the present invention can be placed in DNA constructs which are then used in the transformation of host cells for a generation of recombinantly produced viral proteins. DNA constructs of the present invention comprise a DNA segment encoding the env protein and the flanking region of MN-ST1 (or BA-L) or a portion thereof and a vector. The constructs can further comprise a second DNA segment encoding both a rev protein and a rev-responsive region of the env gene operably linked to the first DNA segment encoding the env protein. The rev protein facilitates efficient expression of the env protein in eucaryotic cells. Suitable vectors for use in the present invention include, but are not limited to, pSP72, lambda EMBL3 and SP65gpt.

Host cells to which the present invention relates are stably transformed with the above-described DNA constructs. The cells are transformed under conditions such that the viral protein encoded in the transforming construct is expressed. The host cell can be procaryotic (such as bacterial), lower eucaryotic (such as fungal, including yeast) or higher eucaryotic (such as mammalian). The host cells can be used to generate recombinantly produced MN-ST1 (or BA-L) env protein by culturing the cells in a manner allowing expression of the viral protein encoded in the construct. The recombinantly produced protein is easily isolated from the host cells using standard protein isolation protocols.

Since HIV-1 strains MN-ST1 and BA-L represent relatively typical United States genotypes, non-infectious MN-ST1 or BA-L proteins (for example, the env protein), peptides or unique portions of MN-ST1 or BA-L proteins (for example, a unique portion of the env protein), and even whole inactivated MN-ST1 or BA-L can be used as an
immunogen in mammals, such as primates, to generate antibodies capable of neutralization and T cells capable of killing infected cells. The protein can be isolated from the virus or made recombinantly from a cloned envelope gene. Accordingly, the virus and viral proteins of the present invention are of value as either a vaccine or a component thereof, or an agent in immunotherapeutic treatment of individuals already infected with HIV-1.

As is customary for vaccines, a non-infectious antigenic portion of MN-ST1 or BA-L, for example, the env protein, can be delivered to a mammal in a pharmacologically acceptable carrier. The present invention relates to vaccines comprising non-infectious antigenic portions of either MN-ST1 or BA-L and vaccines comprising non-infectious antigenic portions of both MN-ST1 and BA-L. Vaccines of the present invention can include effective amounts of immunological adjuvants known to enhance an immune response. The viral protein or polypeptide is present in the vaccine in an amount sufficient to induce an immune response against the antigenic protein and thus to protect against HIV-1 infection. Protective antibodies are usually best elicited by a series of 2-3 doses given about 2 to 3 weeks apart. The series can be repeated when circulating antibody concentration in the patient drops.

Virus derived from the infectious HIV-1(MN) clones, MN-ST1, may also be used for reproducible challenge experiments in chimpanzees treated with candidate HIV-1 vaccines or in vitro with human antiserum from individuals treated with candidate vaccines. A candidate vaccine can be administered to a test mammal, such as a chimpanzee prior to or simultaneously with the infectious MN-ST1 virus of the present invention. Effectiveness of the vaccine can be determined by detecting the presence or absence of HIV-1 infection in the test mammals. Side-by-side comparative tests can be run by further administering to a second set of test mammals the virus alone and comparing the number of infections which develop in the two sets of test mammals. Alternatively, candidate
vaccines can be evaluated in humans by administering the vaccine to a patient and then testing the ability of the MN-ST1 virus to infect blood cells from the patient.

The present invention also relates to the detection of HIV-1 virus in a biological sample. For detection of an HIV-1 infection, the presence of the virus, proteins encoded in the viral genome, or antibodies to HIV-1 is determined. Many types of tests, as one skilled in the art will recognize, can be used for detection. Such tests include, but are not limited to, ELISA and RIA.

In one bioassay of the present invention all, or a unique portion, of the env protein is coated on a surface and contacted with the biological sample. The presence of a resulting complex formed between the protein and antibodies specific therefor in the serum can be detected by any of the known methods commonly used in the art, such as, for example, fluorescent antibody spectroscopy or colorimetry.

The following non-limiting examples are given to further demonstrate the present invention without being deemed limitative thereof.

**EXAMPLES**

**MN-PH1 Clone**

The permuted circular unintegrated viral DNA representing the complete HIV-1(MN) genome was cloned by standard techniques (Sambrook et al., 1989, Molecular Cloning. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press) into the Eco RI site of lambda gtWES.lambda B DNA from total DNA of H9 cells producing HIV-1(MN). This clone is designated lambda MN-PH1, and its structure and restriction map are shown in Figure 1. The clone was subcloned into M13mp18 and M13mp19, and the DNA sequence of the entire clone, given in Figure 2, was obtained by the dideoxy chain termination method (Sanger et al., Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467, 1977). The amino acid sequence of the envelope protein (see Table I) was inferred from the DNA sequence. A restriction map of the cloned unintegrated viral DNA (see
Figure 1) was also obtained from the DNA sequence of lambda PH1 and used in conjunction with the inferred amino acid sequence of the viral proteins to subclone the envelope (env) gene into the commercially available plasmid pSP72 (Promega Biological Research Products, Madison, WI), as shown in Figure 2. This plasmid (pMN-PH1env) contains, in addition to the coding regions for the envelope proteins, the coding region for the rev protein (Feinberg et al., Cell 46, 807-817, 1986) and the portion of the env gene which contains the rev-responsive region (Dayton et al., J. Acquir. Immune. Defic. Syndr. 1, 441-452, 1988), since both are necessary for efficient expression of the envelope protein in eucaryotic cells. This plasmid thus contains all the elements required for production of envelope protein following placement into appropriate expression vectors and introduction into recipient cells, all by standard techniques known to molecular biologists.

**MN-ST1 Clone**

The infectious molecular clone, lambda MN-ST1, was obtained by cloning integrated provirus from DNA purified from peripheral blood lymphocytes infected with HIV-1(MN) and maintained in culture for a short time (one month). The integrated proviral DNA was partially digested with the restriction enzyme Sau3A under conditions which gave a maximum yield of DNA fragments of from 15-20 kilobases (kb). This was cloned into the compatible BamHI site of lambda EMBL3, as shown in Figure 3. Figure 3 also shows the restriction map of clone lambda MN-ST1. The DNA sequence of the entire clone, given in Table II, was obtained by the dideoxy chain termination method (Sanger et al., Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467, 1977). The amino acid sequence was predicted from the DNA sequence (see Table II). This clone can be transfected into recipient cells by standard techniques. After transfection, the cloned proviral DNA is expressed into biologically active virus particles, which can be used as a source for virus stocks. The proviral DNA whose
restriction map is shown in Figure 2, was removed from the lambda phage vector by digestion with BamHI and inserted into a plasmid, SP65gpt (Feinberg et al., Cell 46, 807-817, 1986). This plasmid, pMN-ST1, contains an SV40 origin of replication. Consequently, transfection into COS-1 cells (Gluzman, Y. Cell 23, 175-182, 1981), which produce a SV40 gene product which interacts with the cognate origin of replication, results in a transient high plasmid copy number with a concomitant production of large amount of replication competent, infectious virus (Feinberg et al., Cell 46, 807-817, 1986). This provides a convenient source of genetically homogeneous virus, as well as a way to introduce desired mutations using standard methods.

The envelope gene was excised from the lambda phage clone and cloned into a plasmid as described above for lambda MN-PH1. This clone (pMN-ST1env), is similar to pMN-PH1env, described above, except that it derives from a biologically active cloned provirus. Like pMN-PH1env, it can be placed in a suitable vector and host to produce the envelope protein of HIV-1(MN) by well known techniques.

**BA-L Clone**

A Hind III fragment of unintegrated viral DNA representing the HIV-1(BA-L) genome was cloned by standard techniques into lambda phage Charon 28 DNA from total DNA of peripheral blood macrophages infected with and producing HIV-1(BA-L). A positive clone was selected by hybridization using a radiolabelled probe for the HIV-1 envelope. This clone, designated lambda BA-L1, was found to contain the entire gene for the envelope protein. Its structure is given in Figure 4. The insert was transferred into a plasmid (pBluescript, Stratagene, LaJolla, CA) and the DNA sequence of the env gene was determined (see Table III). This clone is designated pBA-L1.

The amino acid sequence of the envelope protein, shown in Table III, was inferred from the DNA sequence. A restriction map was also obtained from the DNA sequence of BA-L1 (shown in Figure 5) in order to determine the
appropriate restriction enzyme sites for cloning the \textit{env} gene into suitable expression vectors. An Eco RI-HindIII fragment of 0.4 Kb and a 2.8 Kb HindIII-XbaI fragment when cloned together constitute the entire \textit{env} gene. This plasmid contains, in addition to the coding regions for the envelope proteins, the coding region for the \textit{rev} protein and the portion of the \textit{env} protein which contains the \textit{rev}-responsive region. Both are necessary for efficient expression of the envelope protein in eucaryotic cells (Feinberg et al., Cell 46, 807-817, 1986; Dayton et al., J. Acquir. Immune. Defic. Syndr. 1, 441-452). This plasmid thus contains all the HIV-1 genetic elements required for production of envelope protein following placement into appropriate expression vectors and introduction into recipient cells, all by standard techniques well known in the art.

\textbf{Statement of Deposit}

The lambda MN-ST1 clone and the BA-L plasmid clone were deposited at the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland 20852, U.S.A., on September 13, 1990, under the terms of the Budapest Treaty. The lambda MN-ST1 clone has been assigned the ATCC accession number ATCC 40889 and the BA-L plasmid clone has been assigned the ATCC accession number ATCC 40890.

\* \* \* \* \* \* \*

All publications mentioned hereinabove are hereby incorporated by reference.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be appreciated by one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention.
**TABLE I**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGGAGGCTT AATTACCTCC CAACAGAAGAC AAGATATCCCT TGATCTGTGG ATCTTACACAAAA 60</td>
<td></td>
</tr>
<tr>
<td>CACAAGGCTA CTCCCTGAT TACGAGAAC AAGCACCCAG GCGAGGGATC AGATATCCAC 120</td>
<td></td>
</tr>
<tr>
<td>TGACCTTTGG ATGGTGCTAC AGCTTAGTC CAGTTGAGCC AGAGAAGTTA GAAGAGGCA 180</td>
<td></td>
</tr>
<tr>
<td>ACAAAGAGA GAACACCGAC TCTTATACCT CAGTGAGCTT GCATGGAAGT GATGACCCCGG 240</td>
<td></td>
</tr>
<tr>
<td>AGAGAGAAGT GTAGTATGGG AGGTTTGACA GCGGCTTACG ATTTTACAC ATGGCCCGAG 300</td>
<td></td>
</tr>
<tr>
<td>AGCTGATACC GGAGTACTTC AAGAACTGCT GACATCGAGC TTGCTACAAG GCCATTTCGG 360</td>
<td></td>
</tr>
<tr>
<td>CGGCGGACTT TCCAGGAGGG CTTGCGCTGG GCGGACTGG GAGGTCGGCA GCCCTCAGAT 420</td>
<td></td>
</tr>
<tr>
<td>CCTCATATA AGCAGCTGCT TTTGCCTGC TCTGGGGTCTC TCTGGTTAGA CCAGATCTGAA 480</td>
<td></td>
</tr>
<tr>
<td>GCTGGGGAGC TCTGTCGGCTA ACTAGGAAAC CCACACTTCA ATGGCTTACACT 540</td>
<td></td>
</tr>
<tr>
<td>TGATGTCGTC AAGTAGTGGG TGCCGGCTCG TAGTGAGCTA GTGTGATACG GAGATCCTCCT 600</td>
<td></td>
</tr>
<tr>
<td>AGATCCTTTTT AGCGAGTTGG GAAATACTCT AGCAGTGGCC CCGAAACAGG GACTTGGAA 660</td>
<td></td>
</tr>
<tr>
<td>CGAAGAAAAA ACCAGAGCTC TCTCGAGGCA GAGACTGGCT TGCCTAAGCG CGGCGGGCA 720</td>
<td></td>
</tr>
<tr>
<td>GAGGCGGAGG CGGGCGGCTG GTGAGTACGC CAAAAATCT TACGTACCG GAGGCTGAGG 780</td>
<td></td>
</tr>
<tr>
<td>GAGAGAGATG CTTTCGAGAG GCTGTGTTATT AAGCGGGGGA GAATAGATAC GATGGAAAAA 840</td>
<td></td>
</tr>
<tr>
<td>CAATGCTTA AAGCCAGGG AAGAAAGAAA ATATAATA AAAAAAAAAAT TAGTTGCGAG 900</td>
<td></td>
</tr>
<tr>
<td>CAGGAGCTA GAAAGATCG CAGTCATCC TTGGCTCTTTA GAAACCATGCG AAGGCCTGAT 960</td>
<td></td>
</tr>
<tr>
<td>ACAAAATCTG GGCAGGCTAC AACCATCCTC TCAGACAGGA TCAGAAGAAC TTAAATGATT 1020</td>
<td></td>
</tr>
<tr>
<td>ATATAATACA GTGCAACCC CATCTTTATG GCACTAAAG ATAGAGATAA AAGACACCA 1080</td>
<td></td>
</tr>
<tr>
<td>GGAAGCTTTA GAGAAATAG AGGAAGAGCA AAACAAAGTG AAGAAGAAA CACAGCAAGC 1140</td>
<td></td>
</tr>
<tr>
<td>AGCACGCTAC ACAGGAAACA GAGGAAACAG CAGCCGACTG AGCCAAATTT ACCCATAGT 1200</td>
<td></td>
</tr>
<tr>
<td>GCAGAACATC GAGGGCGAAA TGTTGACACT GCCGTTATCAC CTAGAAGCTT TAAATGAGT 1260</td>
<td></td>
</tr>
<tr>
<td>GGTAAAGATG TAGTGAGAGA AGGCTTTCAG CCCAGAAGT ATACCACTGT TTGTACGATT 1320</td>
<td></td>
</tr>
<tr>
<td>ATCAAGAGGA GCCACCCCGCA AAGATTTCTC CACATGCATC AACAGCTGGG GGGACATCA 1380</td>
<td></td>
</tr>
<tr>
<td>AGCAACCATG CAAAGTTCAA AAGAGACCTAT CAATGAGGAA GCTCAGAATG GAGTATAATT 1440</td>
<td></td>
</tr>
<tr>
<td>GCATCCGATG CAGCGAGGC CAATTCATACC AGGCAGGATG AGGAAACCAA GGGAGTAA 1500</td>
<td></td>
</tr>
<tr>
<td>CATAGCAAGA ACTACTAGTA CCCTTCAGGA ACAAAATAGGA TGGAATCACAA ATAAATCACC 1560</td>
<td></td>
</tr>
<tr>
<td>TATCCAGTA GGAAGAATCT ATAAAGATG GATAACCTGT GGATTAAATA AAAAAATGA 1620</td>
<td></td>
</tr>
<tr>
<td>GATGTATAGC CTTTCCAGCA TTCTGGAATG AAGCAGGAAG CAAAGAGGAC CCTTATAGAGA 1680</td>
<td></td>
</tr>
<tr>
<td>CTATGTAGAC GCTGGCTATA AAAACTCTAGG AGGCAGGCAA GCTCCAGAG AGGAAAATAA 1740</td>
<td></td>
</tr>
<tr>
<td>CCGACGACA AAAAACTTTGG TAGTTCCAAA TGCGAACCA GATTGTAAGA CTATTATATT 1800</td>
<td></td>
</tr>
<tr>
<td>AGCATGGGGA CGAGGACTA CACTAGAAG AAGAGATGCAA GCGATTCAGG GAGTGAGGAG 1860</td>
<td></td>
</tr>
<tr>
<td>ACCTGGGTAT AAGCAAGAG GGTGGCCCGCA AGCGATGAGC CAAGTACCAA ATTCAGCTAC 1920</td>
<td></td>
</tr>
</tbody>
</table>
CATATGATG CAGAGGCGCA ATTGTAGGA TAAAGAGAGGATTATCAAGT GCCGTTCAATTG 1980
TGGCAAGGAA GGGGAGTATG CCAAAATTGG CAGGGGCTC AAGAAGAGGG CGCTTTGGAAG 2040
ATGTCAGAAC GAGGACCCAA AAATGAAGA TTGACTCTCGA AGACAGCTGA ATTATTTAGG 2100
GAAGATCTGG CCTCTCCTGA AAGGAAAGGG GAAATTTTC CAGAGCAGAA CAGAGCCGAC 2160
AGCCCACCAGAAGAGACTCTGC AGAGGGGCAA ACAACTCTCC ACTAGAACCGA 2220
GGAGAAGAAAG AAGGAGAGGAG TACAGAAGG GTAAGTGCTC CTTCTACCTCTT TAGTCTTCCC TCAAATCACT 2280
CTTGGGCAAAC GACCCATTGT CAAATTAAG ATAGGGGGCC AACTAAAGGA AGCTCATTAT 2340
GATACCGAGG CAGATGTAC AGTATTAGGA GAAATGAAATT TGCCCAAGAGG AGTGAACCA 2400
AAAATGATAA GGGGAATTGG AGCTTTATAC AAAGTAAGGA GCTATGATGCA GATACCATCA 2460
GGAAATCTGTG GACATAAAGC TATAGATACG GTATTAGTAG GACCTACACC GTGCTAACAC 2520
ATTGGAAAGA ATCTGGTGAC TCAAGTGCTCC TGCACTTTAA AATTTCCTCAT TAGTCTCATT 2580
GAACACTGTA CAGATAAAAT AAGGCAGAGA ATGGAGGGCC CAAAAGTTAA ACAAAGGCAG 2640
TTGACAGAGAG AAAATTAAGA AGAATTTAAG TCAAGGAGGA AAGAGGAGGA 2700
AAAATTCAGC AATGAGGGCC TGAAGATCC TACAATACG TAGATTGGCC AACAAGGAAA CACCCAGGAT 2760
AAAGACAGTA CTAAATGGAG AAAAAATTGA GATTTCAAGG AACTTAATGG AAAAAACTCA 2820
GACTTCTGGG AAGTTCAATT AGGAATACCA CATCTCGGAG GGTTAAAAAA GAAAATACCA 2880
GTAACACAGG AGACAAATAA AGAATTTAGA GAGGAAATGGTAA GAGAGACGACG 2940
AAAGTATGACTG CATTATAATCTT ACCAGTATTAA ACAAATGAGAA CACAGGAGAT TAGTATACG 3000
TACAAATGGCC TTAAGAGCGA ATGGAGAGGA TCCAGACGAA TATTCAAGG TAGCATGACA 3060
AAAATCTTAG AGCCTTTTTAG AGAAAAAAT CACGACATAG TTATCTATCA ATACATAGGAT 3120
GATTTGTATG TAGATCTGCA CTTAGAAATCC CAGACAGCAT GAGGGAAAAT AGAGGAACGT 3180
AGACAGGACTG TTGGAAGGTT GGGATTTAC CACAGCAAGAAAAACACTCA AAGAAAGACT 3240
CCATTCTTTT GGATGGTTAT TGAACTCCAT CCTGATAAAT GGAAGATCAA GCTATACTG 3300
CTACCAAGAA AGGACAGCTG GACTTGCAAT GACATAACAG AGTTAGGGG AAAATTTAAT 3360
TGGCAAGGTC AGATTTGCCG AGGGATTAAA GTAAGCAATTT TAGTAAACT CTTAGAGGA 3420
ACCAAAAGCAG AACAAGGTTA AATACACACT AGAAGAAGAG CAGAGCTGAC ACTGGCGAGA 3480
AACGAGGAAA TTCTAAAAGAG ACCAGACTAT GGAGTGATT ATGGACCCAT AAAAAACCTCA 3540
ATACAGAAGG TACAAAGGGA GAGGCGAGGC AAATTGACAT ATCAAATTTA TCAAGAGCCA 3600
TTTTAAAAAT TGAAGAGCAA CGAATTAGCA AAAGATGAGGG GGTCCTCACAC TTAAGGACTA 3660
AAACAATGAA CGAGGGAGAT GCAAAAAATA GCCAAGAAAA GCACTACATAG ATGGGGGAAG 3720
ACTCTTAAT TTAGACTACC CATAAAAAA GAAACATGGA AAACAATTGGA GACAGGTAT 3780
AGCTAGCCAT CCGAGATCCC TGAGTGAGGG GTTGCTAATA CCCCTCCCTT AGTGAATTTA 3840
TGGTGAGCTG TAGAAGAAAG ACCCATAGTGA GGTGCAAGAA CTTCTCTATGT AGATGGGGGA 3900
GCTAACAGGG AAGACTAAAA AGGAAAGGCA GGATAGTTAA CTAACAGGG AAGACAAAG 3960
GTGTCTCCCA AACTGACAC AACAATACAG AAGACTGACT TACAAGCAAT TCATCTAGCT 4020
TTGCAAGATT CAAGGGTTAG AGTAAAAACT GTAACAGACT CACAATATGC ATTAGGAAATC 4080
ATTCAGACCA AACAGATTA AAGTGAATCA GAGTTGATCA GTCAAATAAT AGAGCAGTTA 4140
ATAAAAAAGG AAAAGGCTCTA TCTGCGAGTG GTACACAGAC ACAAGGAAAT TGGAGGAAT 4200
GAACAGTAG ATAAATTAGT CAGTGTCTGA ATCAGGGAAG TACTATTTTT ATAGCGGAATA 4260
GATAAGGGCC CGAAGAGCCA TGAGAAATAT CACGATATTT CAGAGGCAAT GCCTAATGGC 4320
TTTAACCTAC CACCTTATGT AGCAAAAGAA ATAGTGACCA GCTGTGATAA ATGTCAGCTA 4380
AAAGGAGAAG CCAATCACAT GCAAGTACAG CTAAGATCCAG GAATATGCGA ACTAGATTGT 4440
ACACATTAGT AAGGAAAGGT TATCTCTGTA GCGTTTCATG TAGGCGATGG ATACATAGAA 4500
GCAGAAGTTA TTTCCAGCAGA GCAGGGCAGA GACAGCAGAT ACTTTCTCTT AAAATTAGCA 4560
GGAGATCGCC CAGTAAAAAC AATACATACA GACAATGGCC CCAATTCAC CAGTACTACG 4620
GTAAAGGGCC CCGTTTGTGT GACGGGAAATC AAGCAGGAAAT TTGGGATTTGC CTACAAATCCC 4680
CAAAAGTCAG GAGATAATGA ATCTATGAA GTAGGAAAAG ATAGAAATTT AGAGAGGTA 4740
AGMAGTCCAG GTGAACACTCT TAAAGAGACAA GTCAAATAGG CAGTTATCTC CCACAAATTTT 4800
AAAGAAGAGG CACGGGTTGG CCAGCCGTCAG AACAGGAAAA AATAGATTCG CATAATAGCA 4860
ACMGACATTAC AAAATCAAGA ACTACAAAAA CAAATACCAA AAATTCAAA TTTTCGAGGT 4920
TATTAAGGGG ACCAGACGAGG AAGGTTTGGT CAGAAGGCGA CAAAGCTCTT CTGGAAAGTG 4980
GAAGGGGGACT ACTTAATACG AGATTAATAA GACTAAGAG TATGCCAAG AAAGAACAGCA 5040
AAGGTGCAATG GGGATTGAGG AAAACAGACG CACGCTGATG ATGATGCGG AACGAGACAG 5100
GATGAGGATT AGAACATGGG AAGTTTGTG TAAACACATT ATGTATATTT CAAAGAAGGAC 5160
TAAAGGACGG TTTTATAGAC ATCAGTACTGA AAGCAGTCTAG CCAAGAATAC GTTACAGAAT 5220
ACGATCCAGG CTAAGGGAAG TCATGAGGTT AATACAACA TATGGGGTCT TGCTACAAGG 5280
AGAAGAGAGC TGGCATTTAG GTCAAGGAGT CTGCATAGAA TGCGAGAAGA AGAGATACTA 5340
CACACAGGTG GACCGTGACC TGGCAGGCAA CTAATTTCG ATCCATTACT TTTGATTGT 5400
TTCAGACTCT CCGTACAGAA AGCCCTATTT TGAGAGACAGA TTTTGTGACTT 5460
TCAGAGAGCA CATAACAGGG TAGACCTCCTC AGATCTACGT GCATACACG CATTAATAC 5520
ACCAAAGAAG ATAAAGCGAC CTTTGCCTAG GTTTAAGAAA CTGACAGCGG ATAGATGGA 5580
CAAGGCCTAC AAGGACACGG GCAGAGACAGG GAGCGCATAC ATCAATGGGC ACTAGACCTT 5640
TTAGAGGAGC TTAAGATAGG AGCTGTTAGA CATTTTCCTT GAGATGCGCT CCGATGGGTA 5700
GGCGCACTAC TCTTTGGAAG TTATGGGGAT ACTTGGGCGG GAGTGGAGGC CATAAATAG 5760
ATTCTACAC AACTGCTTGT TATTCTATTTG AGAATTGGGT GTGACATAG CAGAAATAGG 5820
ATTATTCAGC AGAGGAGAGC AGAAGATGGA GCCAGTAGAT CCTGAGCTAG AGCCCTGGA 5880
GCATCCAGGA AGTGAGCCTTGA AGACTGCTTG TACCATCTGC TATTGTAAAA GTGTGCTCTT 5940
TCATGGCCAA GTTGGTTTCA CAAAAAGGC CTTAGGCATC TCCTATGGCA GGAAGAAGCG 6000
GAGACAGGCA CGAGAGCTC CTGAGACGAG TCAGACTCAT CAGTTTCTC TACCAAAGCG 6060
GTAAGTGAATACTAATGCGAACCTTACTGAGTGATAATGCGATCATTAGTAC 6120
AGGAACTAATA CCAATAGGTTG TGTGATCCAT AGTATTCAATA ATATAGGAA GAAATAAGG 6180
ACAAAGAAAA ATAGACAGGTTAATTGATAGAATATACGGGAA AGAGCAGAG ACAGTGCCA 6239

ATG AGA GTG AAG GGG ATC AGG AGG AAT TAT CAC TGG TGG GGA TGG Met Arg Val Lys Gly Ile Arg Arg Asn Tyr Gln His Trp Trp Gly Trp 6287
1 5 10 15
GCC AGG ATG CTC ATT GGG TTA TTA ATG ATC TGT AGT GCT ACA GAA AAA Gly Thr Met Leu Leu Gly Leu Met Ile Cys Ser Ala Thr Glu Lys 6335
20 25 30
TTG TGG GTC ACA GTC TAT TAT TGG GTA CCT GTG TGG AAA GAA GCA ACC Leu Trp Val Thr Val Tyr Gly Val Pro Val Pro Trp Lys Ala Thr 6383
35 40 45
ACC ACT CTA TTT GTG CCA TCA GAT GCT AAA GCA TAT GAT ACA GAG GTA Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val 6431
50 55 60
CAT AAT GGT TGG GCC ACA CAA GCC TGT GTA CCC ACA GAC CCC AAC CCA His Asn Val Trp Ala Thr Gln Ala Cys Val Pro Thr Asp Pro Asn Pro 6479
65 70 75 80
CAG GAA GTA GAA TGG GTA AAT GTG ACA GAA AAT TTT AAC ATG TGG AAA Gln Glu Val Glu Leu Val Asn Val Thr Glu Asp Ile Ser Leu Trp Asp 6527
85 90 95
AAT AAC ATG GTA GAA CAG ATG CAT GAG GAT GTA ATC AGT TTA TGG GAT Asn Asn Met Val Glu Glu Met His Glu Asp Ile Thr Ser Leu Trp Asp 6575
100 105 110
CAG AGC CTA AAG CCA TGT GTA AAA TTA ACC CCA CTC TGT GCT ACT TTA Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu 6623
115 120 125
AAT TGC ACT GAT TTG AGG AAT ACT ACT AAT ACC AAT AGT ACT GCT Anl Thr Asp Leu Arg Asn Thr Thr Asn Thr Asn Asn Ser Thr Ala 6671
130 135 140
AAT AAC AAT AGT AAT AGC GAG GGA ACA ATA AAG GGA GGA AAA GAT AAA Asn Asn Asn Ser Ser Glu Gly Thr Ile Lys Gly Gly Glu Met Lys 6719
145 150 155 160
AAC TGC TCT TTC AAT ATC ACC ACA AGT AGA GAT AAG ATG CAG AAA Asn Cys Ser Phe Asn Ile Thr Ser Asp Arg Lys Glu Met Lys 6767
165 170 175
GAA TAT GCA CTT CTT TAT AAA CTT GAT GTA TCA ATA GAT GAT GTA Glu Tyr Ala Leu Leu Tyr Lys Leu Asp Ile Val Ser Ile Asp Asn Asp 6815
180 185 190
AGT ACC AGC TAT AGG TTG ATA AGT TGT AAT ACC TCA GTC ATT ACA CAA Ser Thr Ser Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln 6863
195 200 205
GCT TGT CCA AAG ATA TTC TTT GAG CCA ATT CCC ATA CAC TAT TGT GCC Ala Cys Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala 6911
210 215 220
CCG GCT GGT TTT GCG ATT CTA AAA TGT AAC GAT AAA AAG TTC AGT GGA
Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Lys Lys Phe Ser Gly
225 230 235 240

AAA GGA TCA TGT AAA AAT GTC AGC ACA GTA CAA TGT ACA CAT GGA ATT
Lys Gly Ser Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile
245 250 255

AGG CGA GTA TCA ACT CAA CTG CTG TTA AAT GCC AGT GTA GCA GAA
Arg Pro Val Val Ser Thr Gln Leu Leu Leu Gly Ser Leu Ala Glu
260 265 270

GAA GAG GTA GTA ATT AGA TCT GAG AAT TTC ACT GAT AAT GCT AAA ACC
Glu Glu Val Val Ile Arg Ser Glu Asn Phe Thr Asp Asn Ala Lys Thr
275 280 285

ATC ATA GTA CAT CGT AAT GAA TCT GTA CAA ATT AAT TGT ACA AGA CCC
Ile Ile Val His Leu Asn Glu Ser Val Gln Ile Asn Cys Thr Arg Pro
290 295 300

AAC TAC AAT AAA AGA AAA AGG ATA CAT GGA CCC GCA GGG AGA GCA TTT
Asn Tyr Asn Lys Arg Arg Ile Lys Gly Pro Gly Arg Ala Phe
305 310 315 320

TAT ACA ACA AAA AAT ATA ATA GGA ACT ATA AGA CAA GCA CAT TGT AAC
Tyr Thr Thr Thr Asn Lys Ile Gly Thr Ile Gly Arg Glu Ala His Cys Asn
325 330 335

ATT AGT AGA GCA AAA TGG AAT GAC ACT TTA AGA CAG ATA GCT TTT ACC AAA
Ile Ser Arg Ala Lys Trp Asn Asp Thr Leu Arg Glu Ile Val Ser Lys
340 345 350

TAA AAA GAA CAA TTT AAG AAT AAA ACA ATA GTC TTT AAT CAA TCC TCA
Leu Lys Glu Glu Phe Lys Asn Lys Thr Ile Val Phe Asn Glu Ser Ser
355 360 365

GGA GGG GAC CCA GAA ATT GTA ATG CAT TAT TTT AAT TGT GGA GGG GAA
Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu
370 375 380

TTT TAC TAC TGT AAT ACA TCA CCA CTG TTT AAT AGT ACT TGG AAT GGT
Phe Phe Tyr Cys Asn Thr Ser Pro Leu Phe Asn Ser Thr Trp Asn Gly
385 390 395 400

AAT AAT ACT TGG AAT AAT ACT ACA GGG TCA AAT AAC AAT ATC ACA CTT
Asn Asn Thr Trp Asn Asn Thr Gly Ser Asn Asn Ile Thr Leu
405 410 415

CAA TGG AAA ATA AAA CAA ATT ATA AAC ATG TGG CAG GAA GTA GGA AAA
Gln Cys Lys Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Lys
420 425 430

GCA ATG TAT GCC CCT CCC ATT GTA GCA CAA ATT AGA TGT TCA TCA AAT
Ala Met Tyr Ala Pro Pro Ile Glu Gly Glu Ile Arg Cys Ser Ser Asn
435 440 445

ATT ACA GGG CTA CTA ACA AGA GAT GGT AAG GAC AGC ACC GAC AGC
Ile Thr Gly Leu Leu Thr Arg Asp Gly Gly Lys Asp Thr Asp Thr
450 455 460

AAC GAC ACC GAG ATC TCC AGA CCT GGA GGA GGA GAT AGG AGG GAC AAT
Asn Asp Thr Glu Ile Phe Arg Pro Gly Gly Asp Met Arg Asp Asn
465 470 475 480
TGG AGA ACT GAA TTA TAT AAA TAT AAA GTA GTA ACA ATT GAA CCA TTA
Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Thr Ile Glu Pro Leu
485  490  495

GGA GTA GCA CCC ACC AAG GCA AAG AGA AGA GTG GTG CAG AGA GAA AAA
Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys
500  505  510

AGA GCA GCC ATA GGA GCT CTG TCC GCT GCC TAA GCA GCA GCA GCA
Arg Ala Ala Ile Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly
515  520  525

AGC ACT ATG GCC GCA GCG TCA GTG ACG CTG ACG GTA CAG GGC AGA CTA
Ser Thr Met Gly Ala Ala Ser Val Thr Leu Thr Val Gln Ala Arg CTA
530  535  540

TTA TTG TCT GTT ATA GTG CAA CAG CAG CAA AAT TTG CAG ACC GCT ATT
Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Arg Ala Ile
545  550  555  560

GAG GCC CAA CAG CAT ATG TTG CAA CTC ACA GTC TGG GCC ATT AAG CAG
Glu Ala Glu Gln His Met Leu Glu Leu Thr Val Tyr Ile Lys Gln
565  570  575

CTC CAG GCA AGA GTG CTG GTG GAA AGA AGA TAC CTA AAG GAT CAA CAG
Leu Gln Ala Arg Val Leu Val Glu Arg Tyr Leu Asp Gln Gln
580  585  590

CAG CTG GGC TTT TGG GTT TCG TCT GCA AAA CTC ATT GCC ACC ACT ACT
Leu Leu Gly Phe Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr
595  600  605

GTG CCT TGG AAT GCT AGT TGG AGT AAT AAA TCT CTG GAT GAT ATT TGG
Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu Asp Asp Ile Trp
610  615  620

AAT AAC ATG ACC TGG ATG CAG TGG GAA AGA GAA ATT GAC CAT TAC ACA
Asn Asn Met Thr Trp Met Glu Trp Glu Arg Glu Asp Tyr Thr
625  630  635  640

AGC TTA ATA TAC TCA TTA CAA AAA TCG CAA ACC CAA CAA GAA AAG
Ser Leu Ile Tyr Ser Leu Leu Glu Lys Ser Thr Gln Gln Glu Lys
645  650  655

AAT GAA CAA GAA TTA TGG GAA TTG GAT AAA TGG GCA AGT TTG TGG AAT
Asn Glu Gln Glu Leu Leu Leu Leu Asp Lys Trp Ala Ser Leu Trp Asn
660  665  670

TGG TTT GAC ATA ACA AAT TGG CTG TGG TAT ATA AAA ATA TTC ATT ATA
Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met
675  680  685

ATA GTA GGA GCC TGG GTA GGT TTA AGA AGA AGA GTT TTT GCT GTA CTT TCT
Ile Val Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu Ser
690  695  700

ATA GTG AAT AGA GTT AGG CAG GGA TAC TCA CCA TGG TCG TTG CAG ACC
Ile Val Asn Arg Val Arg Glu Gly Tyr Ser Pro Leu Ser Leu Gln Thr
705  710  715  720

GCC CCC CCA GCT CCG AGG GGA CCC GAC AGG CCC GAA GGA ATC GAA GAA
Arg Pro Pro Val Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu
725  730  735
GAA GGT GGA GAG AGA GAC AGA ACA TCC GGT CGA TTA GTG CAT GGA 8495
Glu Gly Gly Glu Arg Arg Arg Thr Ser Gly Arg Leu Val His Gly
740  
745  750

TTC TTA GCA ATT ATC TGG GTC GAC CTT CAG GGG AGC CTG CTC CTC TCC AGC 8543
Phe Leu Ala Ile Ile Trp Val Arg Arg Leu Ser Leu Phe Leu Phe Ser
755  760  765

TAC CAC CAG AGA GAC TTA CTC TTT ATT GCA GCC GCG AGG ATT GTC GAA CTT 8591
Tyr His His Arg Asp Leu Leu Leu Ile Ala Ala Arg Ile Val Glu Leu
770  775  780

CTG GGA CGC AGG GGG TGG GAA GTC CTC AAA TAT TGG TGG AAT CTC CTA 8639
Leu Gly Arg Arg Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn Leu Leu
785  790  795  800

CAG TAT TGG AGT CAG GAA CTA AAG AGT AGT GCT GCT AGT TTG GCT ATT 8687
Gln Tyr Trp Ser Gln Glu Leu Lys Ser Ser Ala Val Ser Leu Leu Asn
805  810  815

GCC ACA GCT ATA GCA GTA GCT GAG GGG ACA AGG GTG ATA GAA GTA 8735
Ala Thr Ala Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val
820  825  830

CTG CAA AGA GCT GGT AGA GCT ATT CTC CAC ATA CCT ACA AGA ATA AGA 8783
Leu Gln Arg Ala Gly Arg Ala Leu His Pro Thr Arg Ile Arg
835  840  845

CAG GCC TGG GAA AGG GCT TGT CTA TAAGATGGGT GCCAATGCT GAACACGCTT 8837
Gln Gly Leu Glu Arg Ala Leu Leu
850  855

GACTGGATTG CCTACTGTTAA GGGAAAGAAAT GAGACAGACT GAACACGCTG AGCTAGCGAC 8897

AGATGGGGTG GGGAGCCCTG CCCACAGGTG GAAAGAACAT GGAGCATTG CAGTGACAA 8957

TACAGCAGCT ACCAAGTCTG ATGTGCGCTG GCTAGAAGCG CAAGAGGGAG AGAAGTGGG 9017

TTTCCCCAGT AAACCTCAGG TACCTTTTAG ACCAATGACT TACAGAGCAG CTTTGGATCT 9077

TAGCCACCTTT TTAAGAGAAAA AGGGGGGGCT GGATGGGTTA ATTACTCCCA AAAAAAGACA 9137

AGACATCTTT GATCTGTGGG TCTACTCCAC ACAAGGCTAC TTTCTGAGTT GCGAGACTA 9197

CACACGCGG CGAGGAGTCG GATACTCAG ACTCTTGGGA TGGGCTTACA GAGTAGTACC 9257

AGTGTAGCCA GACAACTGAG TAAAAGGGAA AACAAGTCTG CTTGTGACCA GGCAGGCGAC 9317

TAGAAAGCAG CATGGATTGCA TGAGCCCCAG GAGAAAGTGT TACAGTGGAA GCTCTGAGAC 9377

CACCTAGCAT TCCAGATGA TGGGCGAGAG CTGCAATCAGG AGTACTACAA GAACCTGCTG 9437

CATGAGCTTA TCTACAAGGG ACTCTCGGCT GGAGACTGCC CAGGAGGAGG TGGGCTTGGC 9497

GGGACCGGG AGTGGCGGGC CCTAGAATGG TCGATATAG CAGTGTGGG TCCGCTCTTT 9557

TGGGTCTCTC TGCGCTGACC AGACGCGGAG CTGGAGGCTG TCTCGCTTAA GGCGAGACCC 9617

ACTGCTTCAA CCTCAAATAA CCTGCCCTTG AGTGCGTCGA GTGTGCTGCT CCCGCTCGT 9677

ATGTGACTCT GCTGACCTAGA CACTCCCTAG ATCTCTTAAG CAGTGGTGGG AAATCTCTAG 9737

CA 9739

Met Arg Val Lys Gly Ile Arg Arg Asn Tyr Gln His Trp Trp Gly Trp
1  5  10  15
<table>
<thead>
<tr>
<th>Amino Acids</th>
<th>Positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly Thr Met Leu Leu Gly Leu Leu Met Ile Cys Ser Ala Thr Glu Lys</td>
<td>20 25 30</td>
</tr>
<tr>
<td>Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr</td>
<td>35 40 45</td>
</tr>
<tr>
<td>Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val</td>
<td>50 55 60</td>
</tr>
<tr>
<td>His Asn Val Trp Ala Thr Gln Ala Cys Val Pro Thr Asp Pro Asn Pro</td>
<td>65 70 75 80</td>
</tr>
<tr>
<td>Gln Glu Val Glu Leu Val Asn Val Thr Glu Asn Phe Asn Met Trp Lys</td>
<td>85 90 95</td>
</tr>
<tr>
<td>Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Thr Asp</td>
<td>100 105 110</td>
</tr>
<tr>
<td>Gln Ser Leu Tyr Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu</td>
<td>115 120 125</td>
</tr>
<tr>
<td>Asn Cys Thr Asp Leu Arg Asn Thr Thr Asn Thr Asn Asn Ser Thr Ala</td>
<td>130 135 140</td>
</tr>
<tr>
<td>Asn Asn Asn Ser Asn Ser Glu Gly Thr Ile Lys Gly Gly Glu Met Lys</td>
<td>145 150 155 160</td>
</tr>
<tr>
<td>Asn Cys Ser Phe Asn Ile Thr Thr Ser Ile Arg Asp Lys Met Gln Lys</td>
<td>165 170 175</td>
</tr>
<tr>
<td>Glu Tyr Ala Leu Leu Tyr Lys Leu Asp Ile Val Ser Ile Asp Asn Asp</td>
<td>180 185 190</td>
</tr>
<tr>
<td>Ser Thr Ser Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln</td>
<td>195 200 205</td>
</tr>
<tr>
<td>Ala Cys Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala</td>
<td>210 215 220</td>
</tr>
<tr>
<td>Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Ser Gly</td>
<td>225 230 235 240</td>
</tr>
<tr>
<td>Lys Gly Ser Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile</td>
<td>245 250 255</td>
</tr>
<tr>
<td>Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu</td>
<td>260 265 270</td>
</tr>
<tr>
<td>Glu Glu Val Val Ile Arg Ser Glu Asn Phe Thr Asp Asn Ala Lys Thr</td>
<td>275 280 285</td>
</tr>
<tr>
<td>Ile Ile Val His Leu Asn Glu Ser Val Gln Ile Asn Cys Thr Arg Pro</td>
<td>290 295 300</td>
</tr>
<tr>
<td>Asn Tyr Asn Lys Arg Lys Arg Ile His Ile Gly Pro Gly Arg Ala Phe</td>
<td>305 310 315 320</td>
</tr>
<tr>
<td>Tyr Thr Thr Lys Asn Ile Ile Gly Thr Ile Arg Gln Ala His Cys Asn</td>
<td>325 330 335</td>
</tr>
<tr>
<td>Ile Ser Arg Ala Lys Trp Asn Asp Thr Leu Arg Gln Ile Val Ser Lys</td>
<td>340 345 350</td>
</tr>
<tr>
<td>Leu Lys Glu Gln Phe Lys Asn Lys Thr Ile Val Phe Asn Gln Ser Ser</td>
<td>355 360 365</td>
</tr>
</tbody>
</table>
Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu 370 375 380

Phe Phe Tyr Cys Asn Thr Ser Pro Leu Phe Asn Ser Thr Trp Asn Gly 385 390 395 400

Asn Asn Thr Trp Asn Asn Thr Gly Ser Asn Asn Asn Ile Thr Leu 405 410 415

Gln Cys Lys Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Lys 420 425 430

Ala Met Tyr Ala Pro Pro Ile Glu Gly Gln Ile Arg Cys Ser Ser Asn 435 440 445

Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Lys Asp Thr Asp Thr 450 455 460

Asn Asp Thr Glu Ile Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn 465 470 475 480

Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Thr Ile Glu Pro Leu 485 490 495

Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys 500 505 510

Arg Ala Ala Ile Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly 515 520 525

Ser Thr Met Gly Ala Ala Ser Val Thr Leu Thr Val Gln Ala Arg Leu 530 535 540

Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Arg Ala Ile 545 550 555 560

Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Gly Ile Lys Gln 565 570 575

Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr Lys Asp Gln Gln 580 585 590

Leu Leu Gly Phe Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Thr 595 600 605

Val Pro Trp Asn Ala Ser Trp Ser Asn Ser Leu Asp Asp Ile Trp 610 615 620

Asn Asn Met Thr Trp Met Gln Trp Glu Arg Glu Ile Asp Asn Tyr Thr 625 630 635 640

Ser Leu Ile Tyr Ser Leu Leu Glu Lys Ser Gln Thr Gln Gln Glu Lys 645 650 655

Asn Glu Gln Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn 660 665 670

Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met 675 680 685

Ile Val Gly Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu Ser 690 695 700

Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Leu Gln Thr 705 710 715 720
Arg Pro Pro Val Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu
725 730 735
Glu Gly Gly Glu Arg Asp Arg Asp Thr Ser Gly Arg Leu Val His Gly
740 745 750
Phe Leu Ala Ile Ile Trp Val Asp Leu Arg Ser Leu Phe Leu Phe Ser
755 760 765
Tyr His His Arg Asp Leu Leu Leu Ile Ala Ala Arg Ile Val Glu Leu
770 775 780
Leu Gly Arg Arg Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn Leu Leu
785 790 795 800
Gln Tyr Trp Ser Gln Glu Leu Lys Ser Ser Ala Val Ser Leu Leu Asn
805 810 815
Ala Thr Ala Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu Val
820 825 830
Leu Gln Arg Ala Gly Arg Ala Ile Leu His Ile Pro Thr Arg Ile Arg
835 840 845
Gln Gly Leu Glu Arg Ala Leu Leu
850 855
<table>
<thead>
<tr>
<th>Table II</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGGATGGGTCT AATTATCCCA CAAAGAGACA AGACATCCTT GATCTGGTGG TCTACCAAC</td>
</tr>
<tr>
<td>ACAAGGCTAC TTCTCCTAGT GGCAGAAGTA CAACACACGG CCAGGGATCA GATATCCACT</td>
</tr>
<tr>
<td>GACCTTTGGA TGCTGGCTCTA AGCTAGTACC AGTTGACCCA GAGAGATAG AACAGCCCAA</td>
</tr>
<tr>
<td>TAAAGGAGGA AACAATGCTG TCTTACACCC TATGAGCGAG CATGGGATAG AGATGCCCGA</td>
</tr>
<tr>
<td>GAGAAGATGC TTAATGGGTA AGCTGGAGAC CCACCTAGCA TTTACCAATT ATGCCCGGAG</td>
</tr>
<tr>
<td>GCTGCACTCC GAGTACTACA AGAAGCTGCG ACATCTGAGT ATCTACAGGG GACCTTCCGC</td>
</tr>
<tr>
<td>TGGGCGATTG CACGCCTGGTT GCGCCCGCTGG CCGGACGGGagna CGGCTGATGG</td>
</tr>
<tr>
<td>CTGCCATATAA GCACGGCTGT TCTCCTGGTA CTGGGGTCTCT CTGGTTAGAC CAGATCTGAG</td>
</tr>
<tr>
<td>CCTGGGGAGCT CTTTGGTCTAA CTAAGGGAAAC CACTCTTCAA GCCTCAATAA AGCTTG GCCGT</td>
</tr>
<tr>
<td>GAGTGGCTTCA AGTGTGGGT GCCCCCTGCTA TATGAGCTCTC TTGGATGCAG AGATCCCTCA</td>
</tr>
<tr>
<td>GATCTTTTTA GCAGTGTTGG AAAATTCTTA GCAGTGCGCG CCAGACAGGG ACTTGAAGGC</td>
</tr>
<tr>
<td>GAAAGAGAAAA CGAGAGAGCG TGTCCCGGAG CAGCATCTGG CTGGTCTGAG CGGCGCACGGC</td>
</tr>
<tr>
<td>AAGAGCCGAG GGGCGCCGAC GGGTGGATAC GCCAAATTC TTGAATGGCG GAGCCTAGAA</td>
</tr>
<tr>
<td>GGAGAGACAT GGGTGAGGAG GCGCTGGATG TAAGCGGGGG AGAGTTGGAT GTGGGAGAAA</td>
</tr>
<tr>
<td>AAATTCGTTT AGGGCCAGAG GCCGGAGAAA AATATATAAT AAAACACTGA GTATGCGGCA</td>
</tr>
<tr>
<td>GCAGGGACCT AGAGAGATTC GCAGTCCATC TCTGCCGTTA AGGAAATCTCA GAAGGGCTCTA</td>
</tr>
<tr>
<td>GAAATACTA GGGACAGCTCA CAACAGCTCC TTCAAGACAGG ATCAGAAGAA CTTAAATCAT</td>
</tr>
<tr>
<td>TATATAATAC AGTGACCAACC CTCTATTGTG TGCACTACCAAAA GATAGAGATA AAAGACACCA</td>
</tr>
<tr>
<td>AGGAACCTTT AGAGAAAAAA GAGGAAGGCC AAAACAAAAA TAAGAAGAAA GCACAGCAC</td>
</tr>
<tr>
<td>CAGTAGCTGA CACAGGAAAC AGAGGAACCA GCAGCCAGGT CAGCAGTAAAT TACCCCTAGT</td>
</tr>
<tr>
<td>TGGAAAACAT TGAGGGAAAAGAGCTACATG AGGCCCATATC ACCTAGAAC TTTAAATCAT</td>
</tr>
<tr>
<td>GGTAAAGGT AGTGAAGAGG AAGGCTTCTCA GCCTCGAGAT AATACCTGAT TTTTCGAGCT</td>
</tr>
<tr>
<td>TATCCAGGAG AGGGCCAAAA CAAAGGATTC AAGCATCTGTT AACAACGATG GGGGAGCATG</td>
</tr>
<tr>
<td>AAGCGGAGGC GGAAATGTGA AAAGAGACCA TCAATAGAGGA AGCTGCAGAA TGGTAGATAG</td>
</tr>
<tr>
<td>TGCACTCAGT GCAATGCGGG CCTATTGCAC CAGGCCGAGAT GAGAGAACC GGAGGAGAGT</td>
</tr>
<tr>
<td>ACAATGCAGG AACTACTAGT ACCCTTCAGG ACAAATAAGG ATGGATGACA AATAAACCCAC</td>
</tr>
<tr>
<td>CTATCCGGT AGAGAAATCT TATAAAGAGT GGAATACCTT GGATTAATAT AAAATATGTAA</td>
</tr>
<tr>
<td>GAGATGTATAG CCTTCCGAGC ATTCGGACCA TAAAGCAGAG ACCAAAGAGA CCTTTAAGAG</td>
</tr>
<tr>
<td>ACTATGAGA CCGGGTTCTAT AAAAACTCTAA GAGCAGGAGCA AGCTCGACG GATGAAAATAA</td>
</tr>
<tr>
<td>ATGATAGAGC AGAAACCTTG TTGTCCCAA ATCCGGAACCC AGATGTAAG ACTATTTTAA</td>
</tr>
<tr>
<td>AAGCATTGGG ACCAGCAGC ATACATAGGG AAATGTAGAC AGCATCTGAC GGACTGCGAG</td>
</tr>
<tr>
<td>GACCTGGTCGA TAAGAGCAAGA GATTTCAGCGG AACAGATGAG CCAGAATCAACA AATTCAGCTA</td>
</tr>
<tr>
<td>1920</td>
</tr>
</tbody>
</table>
CCATAATGAT GCAGAGGAGC AATTTTAGA ATCAAAGAAA GATTATCACG TGCTTCAATT 1980
GTGGCAAGAA AGGGCACTAC GCCAAAATTAT GCAGGGCCC TGGAAAAAG GCCGTGTGGGA 2040
AATGTGGAAA GGAAGGACAC CAAATGAAG ATTGTACTGA GAGACAGGCT AATTTTTTAG 2100
GGGAGATCG CCTCTCCTGC AAGGGAAGGC AGGGAAATT TCTCAGAGCA GAACAGAGGC 2160
AACAGCCCCA CCAGAAGAGA GCTCTAGGGT TGGGGAAAGG ACAACAACCT CCTATCAGAA 2220
GCCAGAGAG AGCCAGAGA GGATAGAAGC GACCTCTGTAT CCTTATGCTT CTCTCACAATC 2280
ACTCTTGGGC AAGCAACCAC TTGTCAATAA AAGATAGGGG GCAACACTAA GGAAGCTCTA 2340
TTAGATACAG GACGGATGTA TACAGTTATTA GAAGAAATGA ATTTGCCAGG AGAGATGGAA 2400
CCAAAAATGA TAGGGGAAT TGGAGGTTT ATCAAAATGA GACATGATTG TCAGATRACC 2460
ATAGAAATCT TGTTGACATAA AGCTATAGGT AGCATATTAG TAGACCTCAG ACCTGTCAC 2520
ATAATTGGAA GAAATCCTGT GACTCAGTTT GGGTGCACTT TAAATTTCGC CATTAGTCT 2580
ATTGAACCTG TACATGTTAA ATTTAACCCA GGAATGAGAT GCGAAAAAGTT TAAACATGG 2640
CCATTGCAG AREGAAAAAG AAAGCAATTA ATAGAAAATT GTAGAAGAAT GGAAGAGGAA 2700
GGGAAATTT CAARAAATTTTC GCCTAGAATA ACCATAACCA CTCCAGATTG TCGCAATAAG 2760
AAAAAGACA GTCTAATAGT GAGAAATTA ATGATTTTCA GAGAACGATG TAGAAAATCA 2820
CAAGACTTCT GGGAGATCCA ATTAGGAATA CCCATCTGCG CAGGGTTAAA AAAGAAAAAA 2880
TCAGTAAACAG TACTGGATGT GGGTGGATCG TATTTTTCAG TTCCCTTGA TAAGACCTTC 2940
AGGAGCTATA TGCGATTAC TACTTGATGT ATGAGAAATGA ATACACACGG GATTTGAAT 3000
CAGTACAATGT TGCTTTCCAG GGGATGAAAA GGAATCAGG CAATATCTCA AAGTACGATG 3060
ACAAAAATCT TAGACCCCTT TGAAGAACAA AATCCAGACA TAGTTACCTA TCAGATACAG 3120
GATGATTGTG ATGTAGGATC TGACCTTGAAG ATAGGGCACG ATAGGAACAA ATAGAAGGG 3180
CTGAGACCGAC ATCTGTGGAG GTGGGGATTT ACCACACACG ACAAAAAACT CGAAGAAGAA 3240
CTCTCATTCC TTGGGATGAGTT TTAATGACTC CATCGGTATA AATGGACAGT ACAGGCTATA 3300
GTGGCTGCCAG AAAAGAGCAG CTGGACTGTC AATGACATAC AGAAGTTGTT GGGAAAATTG 3360
AATGGGCACA GTCAAATTTA GCGAGGATT AAAGTAAGGC AATTATGTAAT ACTCCTTAGA 3420
GGACCCAAAG CACTAAGAGA AGTATTACCA CTAACAGAGA AAGCAGACCT AGAAGCTGCA 3480
GAAAACAGGG AAATCTCAA AGAACCAGTA CATGGAGTGT ATTATGACCC ATCAAAAGAC 3540
TTAATAGCAG AAGTACAGAA GCAGGGCACA GGCCTAATGA CATATCAATTT TTAATCAAGG 3600
CCATTAAAAA ATCTCGAACG AGCCAAATAT GCAAGAATGA GGGGGGCACA CACTAATGAT 3660
GTAAAAACAT TAACAGAGGC AGTGCAAAAA ATAGCACCAG AAGAGATAGT AATATGGGA 3720
AAGACTTCTA AATTTTAGACT ACCAATCAAA AGAAAGAACG GGGAAACATG GTGGACAGG 3780
TATTGGCAAG CCACTGTGAT CTCTGAGTTG GAGTTTGTCA ATACCCTTCC CTAGTGAGAA 3840
TTATGGTACC AGTTAGAGGA AGAAGCCATA GTAGGACAG AAGACTTCTA TGATGATGGA 3900
GCAGCTAAAG GGAGACTAAG AAAAGAAAAC GCAGATGATG TTACTACAG AGGAAGACAA 3960
AACGTTGTCT CCCTAACTGA CACACAATAA CAGAAGAATG AGTTACAAGG AATTCATCTA 4020
GCTTTGCAAG ATTCAGGGTT AGAAGTAAAC ATAGTAACAG ACTCACAATA TGCAATTAGGA 4080
ATCATTCAAG CACACCCAGA AAAAGGTAAG TCAGAGTTAG TCAGTCCAAT AATTAGACCA 4140
TTAAATAAAAA AGGAAAGCTT CTATCTGGCAG TGGGTACCCAG CACACAAGGG AATTGGAGGA 4200
AATGAAACAG TAGATAAATTT AGTCAGGCTG GGAATCAGGA AATGACTATT TTTGATAGGA 4260
ATAGATAGCC CCAAGGAAGA CCATGAGAAAT TATCAGAAGA ATGGCTAGCT 4320
GACTTTAACC TACACCTATG ATGAGCCEAA GAAATAGTGC CCAAGCGTGAA TAAATAGTGC 4380
CTAAAAGGGAG AAGCCCATCCA TGCAAGAATGA ATGCAGTGTG GTAGCAGTCTG GTGGAATCATG 4440
TGTCACATT TAGGAAGAAA AGTTATGGTA ATGAAAGTGC ATGGAATCTG GCAACTGAGAT 4500
GAACCGAGAG TATTTCCGCAG AGAGCAGGAG CAGAGACAGA CATACTTCTT CTAAAGAATAA 4560
GCAAGGAGAT GCCAGTCAA AACAATACAC AGAACAATTG GCCCAATTTT CACCAGCATTG 4620
ACGTTTAAAG CGCGCTTGTT GTGGCGCGGG ATCAAGCGAG GATTTGGGATA TCCCTACATAT 4680
CCCCAAAGTC AAGAGTAAAT AGAATCTTAG AATAAAAAAT TAAAGAAAAAT TATAGGCAAG 4740
GTAGAGATCT AGGCTGRCAGA CTCTAAGAGA GCAGTACAAA GAGGCTAATTT CATTCCACAT 4800
TTTAAAGAGA AAGGGGGAAT TGCGGGTCAC AGTCAGCGGG AAAGAATAAG AGACATATAA 4860
GCAACAGACA TACAAACTAA AGAACTACAG AAAAAACAGA CAAAATAACT TAAATTTCCG 4920
GTTTATTACA GGGACAGCGG AGATTTACCTT TGGAACGGGC CACCAAGACCT TCTCACGAA 4980
GGTTGAAAGGA CAGTATGATAT ACAAGTATAAT AGTGCACATAA AAGTACTGCCC GAAGAAAAA 5040
GCAAAGATCA TTAGGGATTA TGGAACAGAG ATGGCAAGTG ATGATTGTAG GCAAGATGAA 5100
CAGGATGAGG ATTAGAACAT GGAAAGTTT AGTAAAACAC CATATGTATA TTTCAAAAGA 5160
AGCTAAAGGG TGTTTTTATA GACATCAGTA TGAAAGCAGC CATCCAGAAGA AATGACCTGA 5220
AGTACACATC CCATCAAGGGG ATGCTAGGTT GGTAAATACAC ACAATTAGGG TGCAGTATAG 5280
AGGAGAAAGA GACTGCCCATT TACGCAGGGG AGTCTCCATA GAAATGGGGA AAAAAAGGATA 5340
TACGCACAAA GTAGACCCCTG ACCTAGACGA CCACTAATT CATTCTGCCT CTTTTGATTG 5400
TTTTTCGAC CTCTCCATATA GAAAGGCAGA ATTAGGACAT AGAGTGTAGC TATTTCTGGTA 5460
ATTTCRACCA GCATTACACG AGGTAGGAAT CATCACAGTA TTGGAATCTG CAGCATTATT 5520
AACACACAAA AAGATAAAGCC CACCTTTGGC TACTTGTTAG AAACGTGACAG AGGTAGATG 5580
GAACAGCCC CAAAGACCCG AGGGCCAGAG AGGGAGCCAG ACAATCAGTG GGGAGTAGAG 5640
CTTTTAGAGG AGCTTAAGAAA TGAGACTGTTT AGACATTTTC CCATTTGTAG GCTCCAGCAG 5700
TCTTAGGCCAC ATATCTATAG AACTTATGGG GATACCTGGG CAGGAGTGGG AGCCCTAAATA 5760
AGAATTATAC AACAAGCTGT GTTTATCTAT TACAGAATTTT GGTTTCGACA TAGCAGAAATA 5820
GGCAATTTC GACAGAGGAG AGCAAGAATA GAGGCACAGTA CAGCTAGACT TGAAGCCCTG 5880
GAGCATCCAG GGAAGTCCGC TCTAGACTCCT GTGTATTAGTG AAAAACTGGTG 5940
CTTCCATTGC CAAGTTGGT TTCAAAAAAA AGCCCTAGGC ATCTCCTATG GCAGGAAGAA 6000
GGGAGAGAG CGACGAGAG CTCTGGAAG CAGTCAGACT CATCAAGTTT CTCTACAAAA 6060
GCAGTAACT GTAAGGTTAAG TGCAGCTTTT AGTTAAAGCA GCATTAGTAG CATTTAGTAG 6120
AGCAGGAATA ATAGCAATAG TTGAGTGCAC CATCATATTC ATAGAATATA GGAAGATAAG 6180
AAAGACAAGA AAAATAGACA GGGTAAATTG CAGAAAAAGC GAAGAGAGC AGAAGACCTG 6240
CA ATG AGA GTG AAG GGG ATC AGG AGG AAT TAT CAG CAC TGG TGG GGA 6287
Met Arg Val Lys Gly Ile Arg Arg Asn Tyr Gln His Trp Trp Gly
1 5 10 15
TGG GGC ACG ATG CTC CTT GGG TTA TTA ATG ATC TGT AGT GCT ACA GAA 6335
Trp Gly Thr Met Leu Leu Gly Leu Met Ile Cys Ser Ala Thr Glu
20 25 30
AAA TTG TGG GTC ACA GTC TAT TAT GGG GTC ATA TGG AAA GAA GCA 6383
Lys Leu Trp Val Thr Val Tyr Val Gly Val Pro Val Trp Lys Glu Ala
35 40 45
ACC ACC ACT CTA TTT TGT GCA TCA GAT GCT AAA GCA TAT TAT GCA GAG 6431
Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu
50 55 60
GTA CAT AAT GTG GCC ACA CAT GCC TGT GTC CCC ACA GAC CCC AAC 6479
Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn
65 70 75
CCA CAA GAA GTA GAA TTG GTA AAT GTG ACA GAA AAT TTT AAC ATG TGG 6527
Pro Gln Glu Val Glu Leu Val Asn Thr Glu Asn Phe Asn Met Trp
80 85 90 95
AAA AAT AAC ATG GTA GAA CAG ATG CAT GAG GAT ATA ATG AGT TTA TGG 6575
Lys Asn Asn Met Val Glu Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp
100 105 110
GAT CAA AGG AAG CCA TGG GTA AAA TTA ACC CCA CTC TGT GTT ACT 6623
Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr
115 120 125
TTC AAT TGC ACT GAT TTG AGG AAT ACT ACT AAT ACC AAT AGT ACT 6671
Leu Asn Cys Thr Asp Leu Arg Asn Thr Asn Thr Asn Ser Thr
130 135 140
GCT AAT AAC AAT AGT AAT AGC GAG GGA ACA ATA ARG GGA GGA GAA ATG 6719
Ala Asn Asn Asn Ser Asn Ser Glu Gly Thr Ile Lys Gly Gly Glu Met
145 150 155
AAA AAC TGC TCT TAC ATC ACC ACA AGC ATA AAG GAT AAG ATG CAG 6767
Lys Asn Cys Ser Phe Asn Thr Thr Ser Ile Arg Asp Lys Met Glu
160 165 170 175
AAA GAA TAT GCA CTT CTT TAT AAA CTT GAT ATA GTG TCA ATA AAT AAT 6815
Lys Glu Tyr Ala Leu Tyr Lys Leu Asp Val Ser Ile Asn Asn
180 185 190
GAT AGT ACC AGG TAT AGG TGG ATA AGT TGT AAT ACC TCA GTC ATT ACA 6863
Asp Ser Thr Ser Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr
195 200 205
CAA GCT TGT CCA AAG ATA TCC TTC TTT GAG CCA ATT CCC ATA CAC TAT TGT 6911
Gln Ala Cys Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys
210 215 220
GCC CCG GCT GTT GCG ATT CTA AAG TGT AAC CAT AAA AAG TTC AGT
  Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Lys Phe Ser
  225 230
GGA AAA GGA TCA TGT AAA AAT GTC AGC ACA GTA CAA TGT ACA CAT GGA
  Gly Lys Gly Ser Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly
  240 245 250 255
ATT AGG CCA GTA GTA TCA ACT CAA CTG TTA AAT GGC AGT CTA GCA
  Ile Arg Pro Val Val Ser Thr Gln Leu Leu Asn Gly Ser Leu Ala
  260 265 270
GAA GAA GAG GTA GTA ATT AGA TCT GAG AAT TTC AAT GAT GCT AAA
  Glu Glu Glu Val Val Ile Arg Ser Glu Asn Phe Asp Asn Ala Lys
  275 280 285
ACC ATC ATA GTA CAT CTG AAT GAA TCT GTA CAA ATT AAT TGT ACA AGA
  Thr Ile Ile Val His Leu Asn Glu Ser Val Gln Ile Asn Cys Thr Arg
  290 295 300
CCC AAC TAC AAT AAA AGA AAA AGG ATA CAT ATA GGA CCA GGG AGA GCA
  Pro Asn Tyr Asn Lys Arg Lys Arg Ile His Ile Gly Pro Gly Arg Ala
  305 310 315
TTT TAT ACA ACA AAA AAT ATA ATA GGA ACT ATA AGA CAA GCA CAT TGT
  Phe Tyr Thr Lys Asn Ile Ile Gly Thr Ile Arg Gln Ala His Cys
  320 325 330 335
AAC ATT AGT AGA GCA AAA TGG AAT GAC ACT TTA AGA CAG ATA GTT AGC
  Asn Ile Ser Arg Ala Lys Trp Asn Thr Leu Arg Gln Ile Val Ser
  340 345 350
AAA TTA AAA GAA CAA TTT AAG AAT AAA ACA ATA GTC TTT ATT CAA TCC
  Lys Leu Lys Glu Glu Phe Lys Asn Thr Ile Val Phe Asn Gly Ser
  355 360 365
TCA GGA GGG GAC CCA GAA ATT GTA ATG CAC AGT ATT AAT TGT GGA GGG
  Ser Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly
  370 375 380
GAA TTT TAC TAG TAT ACG GTA CCA CTG TTT ATT AGT ACT TGG AAT
  Glu Phe Phe Tyr Cys Asn Thr Ser Pro Leu Phe Asn Ser Thr Trp Asn
  385 390 395
GTT AAT AAT ACT TGG AAT AAT ACT ACA GGG TCA AAT AAC AAT ATC ACA
  Gly Asn Thr Trp Asn Thr Asp Asn Asn Ile Thr
  400 405 410 415
CTT CAA TGC AAA ATA AAA CAA ATT ATA AAC ATG TGG CAG GAA GTA GGA
  Leu Gln Cys Lys Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly
  420 425 430
AAA GCA ATA TAT GCC CCT CCC ATT GAA GGA CAA ATT AGA TGT TCA TCA
  Lys Ala Ile Tyr Ala Pro Pro Ile Glu Gly Gln Ile Arg Cys Ser Ser
  435 440 445
AAT ATT ACA GGG CTA CTA ATT ACA AGA GAT GGT GGG ACT GAC ACG GAC
  Asn Ile Thr Gly Leu Leu Thr Arg Gly Gly Lys Asp Thr Asp
  450 455 460
ACG AAC GAC ACC GAG ATC TCC AGA CCT GGA GGA GGA GAT ATG AGG GAC
  Thr Asn Asp Thr Glu Ile Phe Arg Pro Gly Gly Gly Asp Met Arg Asp
  465 470 475
AAT TGG AGA AGT GAA TTA TAT AAA TAT AAA GTA GTA ACA ATT GAA CCA 7727
Aan Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Thr Ile Glu Pro 465 490 495

TTA GGA GTA GCA CCC ACC AGA GCA AAG AGA AGA GTA GTG CAG AGA GAA 7775
Leu Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Glu Arg Glu 500 505 510

AAA AGA GCA GCG ATA GGA GCT CGT TTC CTT GGC TTC TTA GCA GCA GCA 7823
Lys Arg Ala Ile Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala 515 520 525

GGA AGC ACT ATG GGC GCA GCG TCA GTG AGC CTG ACG GTA CAG GCC AGA 7871
Gly Ser Thr Met Gly Ala Ala Ser Val Thr Leu Thr Val Glu Ala Arg 530 535 540

CTA TTA TTG TCT GGT ATA GTG GCA CAG CAG AAC AAT TGG CTG AGG GCC 7919
Leu Leu Ser Gly Val Gln Gln Gln Asn Asn Leu Leu Arg Ala 545 550 555

ATT GAG GCG CAA CAT ATG TTG CAA CTC ACA GTC TGG GGC ATC AAG 7967
Ile Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys 560 565 570 575

CAG CTC CAG AGA ATC CTG GCT GTG GAA AGA TAC CTA AAG GAT CAA 8015
Gln Leu Gln Ala Arg Ile Val Glu Arg Tyr Leu Lys Asp Gln 580 585 590

CAG CTC CTG GGG ATT TGG GGT TGC TCT GGA AAA CTC ATT TGC ACC ACT 8063
Gln Leu Leu Gly Ile Trp Gly Cys Ser Lys Lys Ser Cys Thr Thr 595 600 605

ACT GTG CCT TGG AAT GCT AGT TGG AGT AAT AAA TCT CTG GAT GAT ATT 8111
Thr Val Pro Trp Asn Ala Ser Trp Ser Asn Lys Ser Leu Asp Asp Ile 610 615 620

TGG AAT AAC ATG ACC TGG ATG CAG TGG GAA AGA GAA ATT GAC AAT TAC 8159
Trp Asn Asn Met Thr Trp Met Gln Trp Glu Arg Ile Asp Asn Tyr 625 630 635

ACA AGC TTA ATA TAC TCA TTA CTA GAA AAA TCG CAA ACC CAA CAA GAA 8207
Thr Ser Leu Ile Tyr Ser Leu Leu Glu Lys Ser Gln Thr Gln Gln Glu 640 645 650 655

ATG AAT GAA CAA GAA TTA TTG GAA TTG GAT AAA TGG GCA AGT TTG TGG 8255
Met Asn Glu Gln Leu Gln Leu Gln Asp Leu Asp Ala Ser Leu Trp 660 665 670

AAT TGG TTT GAC ATA ACA AAT TGG CTG TGG TAT ATA AAA ATA TAC ATA 8303
Asn Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile 675 680 685

ATG ATA GTA GGA GGC GTG GTA GTT AGA ATA GTT TTT GCT GTA CCT 8351
Met Ile Val Gly Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu 690 695 700

TCT ATA GTG AAT AGA GTG AGG CAG GGA TCA TCC TTA TCG TGG TCG CAG 8399
Ser Ile Val Asn Arg Val Arg Glu Gly Tyr Ser Pro Leu Ser Leu Gln 705 710 715

ACC CCC CCC CCA GTT AGG GGA CCC GAC AGG CCC GAA GGA ATC GAA 8447
Thr Arg Pro Pro Val Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu 720 725 730 735
GAA GAA GGT GGA GAG AGA GAC ACA TCC GGT CGA TTA GTG CAT  8495
Glu Glu Gly Gly Glu Arg Asp Arg Asp Thr Ser Gly Arg Leu Val His 1 740 745 750

GGA TTC TTA GCA ATT ATC TGG GTC GAC CTG CGG AGC CTG TTC CTC TTC TCT  8543
Gly Phe Leu Ala Ile Ile Trp Val Asp Arg Ser Leu Phe Leu Phe 1 755 760 765

AGC TAC CAC CAC TTG AGA GAC TTA CTC TTG ATT GCA GCG ARG ATT GTG  8591
Ser Tyr His His Leu Arg Asp Arg Leu Leu Ile Ala Arg Ile Val 1 770 775 780

GAA CTT CTG GGA CGC AGG GGG TGG GAA GTC CTC AAA TAT TGG GAT AAT  8639
Glu Leu Leu Gly Arg Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn 1 795 798

CTC CTA CAG TAG TGG AGT CAG GAA CTA AAG AGT AGT GCT GTT AGG TTG  8687
Leu Leu Gin Tyr Trp Ser Gin Glu Leu Lys Ser Ser Ala Val Ser Leu 1 800 805 810 815

CTT AAT GCC ACA GAT ATA GCA GTA GCT GAG GGC ACA GAT AGG GTT ATA  8735
Leu Asn Ala Thr Asp Ile Ala Val Ala Gly Thr Asp Arg Val Ile 1 820 825 830

GAA GTA CTG CAA AGA GCT GGT AGA GCT ATT CTC CAC ATA CCA ACA AGA  8783
Glu Val Leu Gin Arg Gly Arg Gly Met Ala Thr Ile Ala Val His Thr Arg 1 835 840 845

ATA AGA CAG GCC TTG GAA AGG CCT TTG CTA TAAGATGGG GGCAAGATGC  8833
Ile Arg Gin Gly Leu Glu Arg Arg Ala Leu Leu 1 850 855

CAAAAGGTTG GACTGGATGG CCTACTGTTAA GGGAAAAAT GAGACGACT GAAACGCTG 8893

AGCCACGACG AGATGGGGTG GGAACGCTCAT CCCGAACCTT GGAAAAACAT GGACGACTCA 8953

CAAGTAGCAA TATAGCAGCT ACCAATGCTG ATGTGTGGGTGC AGTGGACTACCA GAGAGGAGG 9013

AGGAAATGGG TTGTGCCAGTC AGACCTCGAG TACCTTTTATG ACCAAAGACT TACCAAGGAGG 9073

CCTTTAGCTC TAGGCACTTT TTAAAAAGAAA AGGGGGGGACT GGATGGGTTA ATTACTCCCC 9133

AAAAGAGACA AGACATCTTT GATCGTGGGT TCTACCAACA ACAAGGCTCT TTTTTGATT 9193

GGCGAGACTA CACACCAGG CGAGGATGAC GATATCCATT GACCTTTGGA TGGTGTCTTC 9253

AGCTTACTAC AGTGGAGCCT GAGAAGATA GAGAGGCTAA TAAAGGAGG AACAGGTGCT 9313

TTTACACCC TATGAGCAG CATGGTGGG ATGACCCGGG GAGAGGCGT TTATGTGGGA 9373

AGTCTGACAG CCACTTACTA TTTACACATT ATGCCCCGGAAG CTCGATGCC GAGTCTACA 9433

AGAATACTGTG ACATCGAGCT ATCTACAGG GACCTTTCCG TGGGACTTT CCAGGGACTT 9493

GTGGGCTGGG CGGGACCGGG GAGTGCGGAG CCCTACAGGT CTGTATATGA GCAGCGCTT 9553

TCTGCTTGTA CGGGGCTGCT CGGGGCTGCT CCGGGACGCT CCCGCGCTAA 9613

CTAGGAAAC ACCATCGTTAA GCCTCAATAA AGCTTGCATT GAGTGGACTA AGTACTGTT 9673

GGCCGCTGGT TAGTGGACTC TGTTAGCTAG AGATACCTCA GATCCTTTCA GGCGGAGG 9733

AAAATCTTCTA GCA 9746

Met Arg Val Lys Gly Ile Arg Arg Asn Tyr Gln His Trp Trp Gly Trp 1  5 10 15
<table>
<thead>
<tr>
<th>Number</th>
<th>Amino Acid(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Gly Thr Met Leu Leu Gly Leu Leu Met Ile Cys Ser Ala Thr Glu Lys</td>
</tr>
<tr>
<td>25</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr</td>
</tr>
<tr>
<td>40</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val</td>
</tr>
<tr>
<td>55</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro</td>
</tr>
<tr>
<td>70</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>Gln Glu Val Glu Leu Val Asn Val Thr Glu Asn Phe Asn Met Trp Lys</td>
</tr>
<tr>
<td>90</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp</td>
</tr>
<tr>
<td>105</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu</td>
</tr>
<tr>
<td>120</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>Asn Cys Thr Asp Leu Arg Asn Thr Thr Asn Asn Asn Ser Thr Ala</td>
</tr>
<tr>
<td>135</td>
<td></td>
</tr>
<tr>
<td>140</td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>Asn Asn Asn Ser Asn Glu Gly Thr Thr Lys Gly Gly Gly Glu Met Lys</td>
</tr>
<tr>
<td>150</td>
<td></td>
</tr>
<tr>
<td>155</td>
<td></td>
</tr>
<tr>
<td>160</td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>Asn Cys Ser Phe Asn Ile Thr Ser Ile Arg Asp Lys Met Gln Lys</td>
</tr>
<tr>
<td>170</td>
<td></td>
</tr>
<tr>
<td>175</td>
<td></td>
</tr>
<tr>
<td>180</td>
<td>Glu Tyr Ala Leu Leu Tyr Lys Leu Asp Ile Val Ser Ile Asn Asn Asp</td>
</tr>
<tr>
<td>185</td>
<td></td>
</tr>
<tr>
<td>190</td>
<td></td>
</tr>
<tr>
<td>195</td>
<td>Ser Thr Ser Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln</td>
</tr>
<tr>
<td>200</td>
<td></td>
</tr>
<tr>
<td>205</td>
<td></td>
</tr>
<tr>
<td>210</td>
<td>Ala Cys Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala</td>
</tr>
<tr>
<td>215</td>
<td></td>
</tr>
<tr>
<td>220</td>
<td></td>
</tr>
<tr>
<td>225</td>
<td>Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Phe Ser Gly</td>
</tr>
<tr>
<td>230</td>
<td></td>
</tr>
<tr>
<td>235</td>
<td></td>
</tr>
<tr>
<td>240</td>
<td></td>
</tr>
<tr>
<td>245</td>
<td>Lys Gly Ser Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile</td>
</tr>
<tr>
<td>250</td>
<td></td>
</tr>
<tr>
<td>255</td>
<td></td>
</tr>
<tr>
<td>260</td>
<td>Arg Pro Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu</td>
</tr>
<tr>
<td>265</td>
<td></td>
</tr>
<tr>
<td>270</td>
<td></td>
</tr>
<tr>
<td>275</td>
<td>Glu Glu Val Val Ile Arg Ser Glu Asn Phe Asn Asp Ala Lys Thr</td>
</tr>
<tr>
<td>280</td>
<td></td>
</tr>
<tr>
<td>285</td>
<td></td>
</tr>
<tr>
<td>290</td>
<td>Ile Ile Val His Leu Asn Glu Ser Val Gln Ile Asn Cys Thr Arg Pro</td>
</tr>
<tr>
<td>295</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td></td>
</tr>
<tr>
<td>305</td>
<td>Asn Tyr Asn Lys Arg Lys Arg Ile His Ile Gly Pro Gly Arg Ala Phe</td>
</tr>
<tr>
<td>310</td>
<td></td>
</tr>
<tr>
<td>315</td>
<td></td>
</tr>
<tr>
<td>320</td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>Tyr Thr Thr Lys Asn Ile Ile Gly Thr Ile Arg Gln Ala His Cys Asn</td>
</tr>
<tr>
<td>330</td>
<td></td>
</tr>
<tr>
<td>335</td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>Ile Ser Arg Ala Lys Trp Asn Asp Thr Leu Arg Gln Ile Val Ser Lys</td>
</tr>
<tr>
<td>345</td>
<td></td>
</tr>
<tr>
<td>350</td>
<td></td>
</tr>
<tr>
<td>355</td>
<td>Leu Lys Glu Gln Phe Lys Asn Lys Thr Ile Val Phe Asn Gln Ser Ser</td>
</tr>
</tbody>
</table>
Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu
Gly 370 375 380
Phe Phe Tyr Cys Asn Thr Ser Pro Leu Phe Asn Ser Thr Trp Asn Gly
Thr 385 390 395 400
Asn Asn Thr Trp Asn Asn Thr Gly Ser Asn Asn Asn Ile Thr Leu
Thr 405 410 415
Gln Cys Lys Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Lys
Gln 420 425 430
Ala Ile Tyr Ala Pro Pro Ile Glu Gly Gln Ile Arg Cys Ser Ser Asn
Gln 435 440 445
Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Lys Asp Thr Asp Thr
Thr 450 455 460
Asn Asp Thr Glu Ile Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn
Gln 465 470 475 480
Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Thr Ile Glu Pro Leu
Thr 485 490 495
Gly Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys
Gln 500 505 510
Arg Ala Ala Ile Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly
Gln 515 520 525
Ser Thr Met Gly Ala Ala Ser Val Thr Leu Thr Val Gln Ala Arg Leu
Ser 530 535 540
Leu Leu Ser Gly Ile Val Gln Gln Asn Asn Leu Arg Ala Ile
Ser 545 550 555 560
Glu Ala Gln Gln His Met Leu Gln Leu Thr Val Trp Gly Ile Lys Gln
Lys 565 570 575
Leu Gln Ala Arg Ile Leu Ala Val Gln Arg Tyr Leu Lys Asp Gln Gln
Val 580 585 590
Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Thr
Val 595 600 605
Val Pro Trp Asn Ala Ser Trp Ser Asn Ser Leu Asp Asp Ile Trp
Ala 610 615 620
Asn Asn Met Thr Trp Met Gln Trp Glu Arg Glu Ile Asp Asn Tyr Thr
Glu 625 630 635 640
Ser Leu Ile Tyr Ser Leu Leu Lys Ser Gln Thr Gln Gln Glu Met
Thr 645 650 655
Asn Glu Gln Glu Leu Leu Gln Leu Asp Lys Trp Ala Ser Leu Trp Asn
Val 660 665 670
Trp Phe Asp Ile Thr Asn Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met
Pro 675 680 685
Ile Val Gly Gly Leu Val Gly Leu Arg Ile Val Phe Ala Val Leu Ser
Thr 690 695 700
Ile Val Asn Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Leu Gln Thr
Glu 705 710 715 720
<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>725</td>
<td>730</td>
<td>Arg Pro Pro Val Pro Arg Gly Pro Asp Arg Pro Glu Gly Ile Glu Glu</td>
</tr>
<tr>
<td>740</td>
<td>745</td>
<td>Glu Gly Gly Glu Arg Asp Arg Asp Thr Ser Gly Arg Leu Val His Gly</td>
</tr>
<tr>
<td>755</td>
<td>760</td>
<td>Phe Leu Ala Ile Ile Trp Val Asp Leu Arg Ser Leu Phe Leu Phe Ser</td>
</tr>
<tr>
<td>770</td>
<td>775</td>
<td>Tyr His His Leu Arg Asp Leu Leu Ile Ala Ala Arg Ile Val Glu</td>
</tr>
<tr>
<td>785</td>
<td>790</td>
<td>Leu Leu Gly Arg Arg Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn Leu</td>
</tr>
<tr>
<td>805</td>
<td>810</td>
<td>Leu Gln Tyr Trp Ser Gln Glu Leu Lys Ser Ser Ala Val Ser Leu Leu</td>
</tr>
<tr>
<td>820</td>
<td>825</td>
<td>Asn Ala Thr Asp Ile Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu</td>
</tr>
<tr>
<td>835</td>
<td>840</td>
<td>Val Leu Gln Arg Ala Gly Arg Ala Ile Leu His Ile Pro Thr Arg Ile</td>
</tr>
<tr>
<td>850</td>
<td>855</td>
<td>Arg Gln Gly Leu Glu Arg Ala Leu Leu</td>
</tr>
<tr>
<td>Residue</td>
<td>Sequence</td>
<td>Met</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>-----</td>
</tr>
<tr>
<td>ACG</td>
<td>GAG ATC AGG</td>
<td>704</td>
</tr>
<tr>
<td>THR</td>
<td>GLU ILE ARG</td>
<td></td>
</tr>
<tr>
<td>ILE</td>
<td>GLY SER TYR</td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>GLN HIS TRP</td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>TRP GLY ILE</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>Arg Val</td>
<td></td>
</tr>
</tbody>
</table>

| CTC     | CTT GGG ATA    | 752 | 35  |
| LEU     | TTA ATG ATC    |     |     |
| TYG     | TGT AAT GCT    |     |     |
| GLU     | GAA AAA TCG    |     |     |
| LYS     | GAA TGG GTC    |     |     |
| LEU     | MET ILE ASN    |     |     |
| CYS     | ALA GLU ILE    |     |     |
| GLU     | TRP VAL       |     |     |

| ACA     | GTC TAT GGG    | 800 | 50  |
| THR     | GTA CCT GTG    |     |     |
| VAL     | AAA GAG GCA    |     |     |
| GCA     | ACC ACT CTA    |     |     |
| THR     | VAL GLU TYR    |     |     |
| VAL     | VAL PRO VAL    |     |     |
| GLU     | GLU ALA THR    |     |     |
| LEU     | THR LEU       |     |     |

| TTT     | TGT GCA TCA    | 848 | 65  |
| PHE     | GAT CGT AAA    |     |     |
| CYS     | GCA TAT GAT    |     |     |
| ALA     | ACA GAG GTA    |     |     |
| CAT     | CAT AAT GTT    |     |     |
| ASP     | ARG LYS ALA    |     |     |
| ASP     | TYR THR GLU    |     |     |
| ARG     | THR HIS ASN    |     |     |
| THR     | VAL GLU       |     |     |

| TGG     | GCC ACA CAT    | 896 | 80  |
| TRP     | GCC CCC ACC    |     |     |
| ALA     | GAC CCC ACG    |     |     |
| THR     | VAL HIS ALA    |     |     |
| ASP     | PRO THR ASP    |     |     |
| ASP     | PRO GLN GLU    |     |     |
| GLN     | VAL TRP       |     |     |
| LEU     | GLU ASN VAL    |     |     |
| THR     | LEU ASN       |     |     |

| GAA     | TTG AAA AAT    | 944 | 95  |
| GLU     | GTG ACA GAA    |     |     |
| LEU     | AAT TTT ACG    |     |     |
| THR     | TGG AAA AAT    |     |     |
| GLU     | ACG ATG       |     |     |
| LEU     | VAL THR GLU    |     |     |
| ASN     | MET TRP       |     |     |
| ASN     | ASN PHE       |     |     |
| ASN     | VAL TRP       |     |     |
| ASN     | ARG ASN       |     |     |
| ASN     | THR SER       |     |     |

| GTA     | GAA CAT GAG    | 992 | 115 |
| GLU     | GAT GAT CAA    |     |     |
| GLU     | GAG AAC CTG    |     |     |
| MET     | HIS GLU ASL    |     |     |
| ARG     | ILE SER LEU    |     |     |
| TRP     | ASP GLN SER    |     |     |
| LEU     | ALA THR       |     |     |

| AAG     | CCA TGT GTA    | 1040| 130 |
| LYS     | AAA TAA ACC    |     |     |
| GLU     | CCC TGT GTT    |     |     |
| PRO     | VAL LYS LEU    |     |     |
| THR     | PRO LEU CYS    |     |     |
| VAL     | VAL LEU ASN    |     |     |
| THR     | ASN CYS THR    |     |     |
| THR     | LEU ASN       |     |     |

| GAT     | TTG AAT ACT    | 1088| 145 |
| ASP     | AAT GGT AAT    |     |     |
| Asp     | GAC ACT AAT    |     |     |
| ACT     | AAT ACC ACT    |     |     |
| ACT     | AGT ASL       |     |     |
| THR     | ASN THR       |     |     |
| THR     | THR SER       |     |     |
AGC AGG GGA ATG GTG GGG GGA GAA ATG AAA AAT TGC TCT TCC AAT
Ser Arg Gly Met Val Gly Gly Gly Glu Met Lys Asn Cys Ser Phe Asn
150 155 160

ATC ACC ACA AAC ATA AGA GGT AAG GTG CAG AAA GAA TAT GCA CTG TTT
Ile Thr Thr Asn Ile Arg Gly Lys Val Glu Lys Glu Tyr Ala Leu Phe
165 170 175

TAT AAA CTT GAT ATA GCA CCA ATA GAT AAT AAG GTT AAT AGA TAT
Tyr Lys Leu Asp Ile Ala Pro Ile Asp Asn Ser Asn Arg Tyr
180 185 190 195

AGG TTG ATA AGT TGT AAC ACC TCA GTC ATT ACA CAG GCC TGT CCA AAG
Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gin Ala Cys Pro Lys
200 205 210

GTA TCC TTG GAG CCA ATT CCC ATA CAT TAT GTC TGT GCC CCG GCT TTT
Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe
215 220 225

GCC ATT CTA AAG TGT AAA GAT AAG AAG TTC AAT GGA AAA GGA CCA TGT
Ala Ile Leu Lys Cys Lys Asp Lys Phe Asn Gly Lys Gly Pro Cys
230 235 240

ACA AAT GCT AGC ACA GTA CAA TGT ACA CAT GAA ATT AGG CCA GTA GTA
Thr Asn Val Ser Thr Val Gin Cys Thr His Gly Ile Arg Pro Val Val
245 250 255

TCA ACT CAA CTG CTG TTA AAT GGC ACT GTA GCA GAA GAC GTA GAA
Ser Thr Gin Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Val Val
260 265 270 275

ATT AGA TCC GCC AAT TTC GCG GAC AAT GCT AAG AAT GTA ATA GAG
Ile Arg Ser Ala Asp Phe Ala Asp Asn Ala Lys Val Ile Ile Gin
280 285 290

CTG AAT GAA TCT GTA GAA ATT AAT TGT ACA AGA CCC AAC AAG AAT ACA
Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn Asn Thr
295 300 305

AGA AAA AGT ATA CAT ATA GGA CCA GCC AGA GCA TTT TAT ACA ACA GGA
Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe Tyr Thr Thr Gly
310 315 320

GAA ATA ATA GGA AGA ATA CAA GCC CCA GTA ATT AGT AGA GCA
Glu Ile Ile Gly Asp Ile Arg Gin Ala His Cys Asn Leu Ser Arg Ala
325 330 335

AAA TGG AAT GAC ACT TTA AAT AAG ATA GTT ATA AAA TTA GAA AGA CAA
Lys Trp Asp Thr Leu Asn Lys Ile Val Ile Lys Leu Arg Glu Gin
340 345 350 355

TTT GGG AAT AAA ACA ATA GTT TTT AAG CAC TCC TCA GGA GGG GAC CCA
Phe Gly Asn Lys Thr Ile Val Phe Lys His Ser Gly Gly Asp Pro
360 365 370

GAA ATT GTG ACG CAC AGT TTT AAT TGT GGA GGG GAA TTT TTT TAC TGT
Glu Ile Val Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys
375 380 385

AAT TCA ACA CAA CTG TTT AAT AGT ACT TGG AAT GTT ACT GAA GAG TCA
Asn Ser Thr Gin Leu Phe Asn Ser Thr Trp Asn Val Thr Glu Glu Ser
390 395 400
AAT AAC ACT GTA GAA AAT AAC ACA ATC ACA CTC CCA TGG AGA ATA AAA 405 1904
Asn Asn Thr Val Glu Asn Asn Thr Ile Thr Leu Pro Cys Arg Ile Lys 410 415
CAA ATT ATA AAC ATG TGG CAG GAA GTA GGA AGA ATG TAT GCC CCT 1952
Gln Ile Ile Asn Met Trp Glu Val Gly Arg Ala Met Tyr Ala Pro 420 425 430 435
CCC ATC AGA GGA CAA ATT AGA GTG TCA TCA AAT ATT ACA GGG CTG CTA 2000
Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu 440 445 450
TTA ACA AGA GAT GGT GGT CCT GAG GAC AAC AAG ACC GAG GTC TTC AGA 2052
Leu Thr Arg Asp Gly Gly Pro Glu Asp Asn Lys Thr Glu Val Phe Arg 455 460 465
CCT GGA GGA GAT ATG AGG GAT AAT TGG AGA AGT GAA TTA TAT AAA 2096
Pro Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys 470 475 480

TAT AAA GTA GTA AAA ATT GAA CCA TTA GGA GTC GCA CCC ACC AAG GCA 2144
Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala 485 490 495
AAG AGA AGA GTG GTG CAG AGA AAA AGA GCA GTG GGA ATA GGA GCT 2192
Lys Arg Arg Val Val Glu Arg Lys Arg Ala Val Gly Ile Gly Ala 500 505 510 515
GTG TTC CTT GGG TTC TTG GGA GCA GCA GGA AGC ACT ATG GGC GCA GCG 2240
Val Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala 520 525 530
GCA ATG AGC CTT AGC GTA CAG GCC AGA CTA TTA TTG TCT GAT GAT 2288
Ala Met Thr Leu Thr Val Glu Ala Arg Leu Leu Ser Gly Ile Val 535 540 545
CAA CAG CAG AAC AAT CTG CTG AGG GCT ATT GAG GCG CAA CAG CAT CTG 2336
Gln Glu Glu Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu 550 555 560
TTG CAA CTC AGA GCT TGG GGC ATC AAG CAC CTC CAG GCA AGA GTC CTG 2384
Leu Glu Leu Thr Val Trp Gly Ile Lys Gln Leu Glu Ala Arg Val Leu 565 570 575
GCT GTG GAA AGA TAC CTA AGG GAT CAA CAG CTC CTG GGG ATT TGG GCT 2432
Ala Val Glu Arg Tyr Leu Arg Asp Glu Leu Gly Ile Trp Gly 580 585 590 595
TGC TCT GAA AAA CTC ATC TGC ACC ACT GCT GTG CTG TGG AAT GCT AGT 2480
Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro Trp Asn Ala Ser 600 605 610
TGG AGT AAT AAA TCT CTG AAT AAG ATT TGG GAT AAC ATG ACC TTA 2528
Trp Ser Asn Lys Ser Leu Asn Lys Ile Trp Asp Asn Met Thr Trp Ile 615 620 625
GAG TGG GAC AGA GAA ATT AAC AAT TAC ACA AGA ATA ATA TAC AGC TTA 2576
Glu Trp Asp Arg Glu Ile Asn Tyr Thr Ser Ile Ile Tyr Ser Leu 630 635 640
ATT GAA GAA TGG CAG AAG CAA GAA GAA AAG AAT GAA CAA GAA TTA TTA 2624
Ile Glu Glu Ser Glu Asn Glu Glu Gly Lys Asn Glu Glu Leu Leu 645 650 655
GAA TTA GAT AAA TGG GCA AGT TTG TGG AAT TGG TTT GAC ATA ACA AAA Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp Phe Asp Ile Thr Lys 660 665 670 675
TGG CTG TGG TAT ATA AAA ATA TTC ATA ATG ATA GTA GGA GGC TTG ATA Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met Ile Val Gly Gly Leu Ile 680 685 690
GQT TTA AGA ATA GTT TTT TCT GTA CTT TCT ATA GTG AAT AGA GTT AGG Gly Leu Arg Ile Val Phe Ser Val Leu Ser Ile Val Asn Arg Val Arg 695 700 705
CAG GGA TAC TCA CCA TTA TCG TTT CAG ACC CCA CTC CCA TCC TTG AGG Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr His Leu Pro Ser Ser Arg 710 715 720
GGA CCC GAC AGG CCC GGA ATG GAA GAA GAA GGT GGA GAG AGA GAC Gly Pro Asp Arg Ser Gly Ile Glu Glu Gly Gly Glu Arg Asp 725 730 735
AGA GAC AGA TCC GGT CCA TTA GTG AAT GAC TCA TTG TTG GCT TTT ATC TTG Arg Asp Arg Ser Gly Pro Leu Val Asn Gly Phe Leu Ala Ile Trp 740 745 750 755
GTC GAT CTG CGG AGC TCT TTC TTT TCT TCT AAC TAC CAG CGC TTG AGA GAC Val Asp Leu Arg Ser Leu Phe Leu Phe Ser Tyr His Arg Leu Arg Asp 760 765 770
TTA CTC TTG ATT GTG ATG AGG ATT GTG GAA CTT CTG GGA CTA GCA GGG Leu Leu Leu Ile Val Met Arg Ile Val Glu Leu Leu Gly Leu Ala Gly 775 780 785
GGG TGG GAA GTT CTC AAA TAT TGG TGG AAT CTC CTA CAG TAT TGG AGT Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn Leu Leu Gln Tyr Trp Ser 790 795 800
CAG GGA CTA AAG AAT AGT GCT GTT AGC TTG CTC AAT GCC ACA GCT GTA Gln Gly Leu Arg Asn Ser Ala Val Ser Leu Leu Asn Ala Thr Ala Val 805 810 815
GCA GTA GCT GAA GGG ACA GAT AGG GTT ATA GAA GTA TTA CAG AGA GCT Ala Val Ala Gly Thr Asp Arg Val Ile Glu Val Leu Gln Arg Ala 820 825 830 835
GTT AGA GCT ATT CTC CAC ATA CCT AGA AGA ATA AGA CAG GGC TTG GAA Val Arg Ala Ile Leu His Ile Pro Arg Arg Arg Gly Gly Leu Glu 840 845 850
AGG GCT TTG CTA TAAGATGGGT GCCAAGTGCT CAAACGGTAG TATAGTCGTA Arg Ala Leu Leu 855
TGCCCTGCTG TAAGGAAAAG AATGAGAAGA ACTGAGCCAG CAGCAGATGG AGTAGGACCA 3312
GTATCTAGAG ACCTGGAAAA ACATGGGACA ATCACAAGTA GCCATACAGC AGCTAACAAAT 3372
GCTGATTGTC CCTGGCGCTAGA ACGCAACAGGG GATGAAAGAG TGGGTCTTTCC AGTGAAGCCT 3432
CAGGTACCCCT TAAGACCAAT GACTGCGAGT CGACGTAGTAT ATCTTGGCAGA CTTTTTTAAG 3492
AAAAAGGGG GACTGAAAGG GCTAATTCCAC TCCACAAAAA GACAGGATAT CTTGATTG 3552
TGGGCTACCC ACACAACAGG CTACTTCCCT GATGGCGAGA ACTACACACC AGGGCCAGGG 3612
ACACAGATCC CACTGACCTT TGGATGGTGC TCAAGCTAG TACCAGTGA GCCGAGAGAAG 3672
35

GTGAGAGGG CCAATGAAGG AGAGAACAC TGCTGTCA CACCTATAGG CCTGCATGGG 3732
ATGGATGACC CGGAGAAGGA AGTGTAGCA TGGAAGTGTG ACAGCAAGCCT AGCATTCCAT 3792
CAGCTGGCCC GAGAA 3807

Met Arg Val Thr Glu Ile Arg Lys Ser Tyr Gln His Trp Trp Arg Trp
1  5 10 15
Gly Ile Met Leu Leu Gly Ile Leu Met Ile Cys Asn Ala Glu Glu Lys
20 25 30
Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Glu Ala Thr
35 40 45
Thr Thr Leu Phe Cys Ala Ser Asp Arg Lys Ala Tyr Asp Thr Glu Val
50 55 60
His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
65 70 75 80
Gln Glu Val Glu Leu Asn Val Thr Glu Asn Phe Asn Met Trp Lys
85 90 95
Asn Asn Met Val Glu Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
100 105 110
Gln Ser Leu Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Thr Leu
115 120 125
Asn Cys Thr Asp Leu Arg Asn Ala Thr Asn Gly Asn Asp Thr Asn Thr
130 135 140
Thr Ser Ser Ser Arg Gly Met Val Gly Gly Gly Glu Met Lys Asn Cys
145 150 155 160
Ser Phe Asn Ile Thr Thr Asn Ile Arg Gly Lys Val Gln Lys Glu Tyr
165 170 175
Ala Leu Phe Tyr Lys Leu Asp Ile Ala Pro Ile Asp Asn Asn Ser Asn
180 185 190
Asn Arg Tyr Arg Leu Ile Ser Cys Asn Thr Ser Val Ile Thr Gln Ala
195 200 205
Cys Pro Lys Val Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro
210 215 220
Ala Gly Phe Ala Ile Leu Lys Cys Lys Asp Lys Phe Asn Gly Lys
225 230 235 240
Gly Pro Cys Thr Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg
245 250 255
Pro Val Val Ser Thr Gln Leu Leu Asn Gly Ser Leu Ala Glu Glu
260 265 270
Glu Val Val Ile Arg Ser Ala Asn Phe Ala Asp Asn Ala Lys Val Ile
275 280 285
Ile Val Gln Leu Asn Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn
290 295 300
Asn Asn Thr Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe Tyr
305 310 315 320
Thr Thr Gly Glu Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Leu 325 330 335
Ser Arg Ala Lys Trp Asn Asp Thr Leu Asn Lys Ile Val Ile Lys Leu 340 345 350
Arg Glu Gln Phe Gly Asn Lys Thr Ile Val Phe Lys His Ser Ser Gly 355 360 365
Gly Asp Pro Glu Ile Val Thr His Ser Phe Asn Cys Gly Gly Glu Phe 370 375 380
Phe Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp Asn Val Thr 385 390 395 400
Glu Glu Ser Asn Thr Val Glu Asn Thr Ile Thr Leu Pro Cys 405 410 415
Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Arg Ala Met 420 425 430
Tyr Ala Pro Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr 435 440 445
Gly Leu Leu Thr Arg Asp Gly Gly Pro Glu Asp Asn Lys Thr Glu 450 455 460
Val Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu 465 470 475 480
Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro 485 490 495
Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg Ala Val Gly 500 505 510
Ile Gly Ala Val Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met 515 520 525
Gly Ala Ala Ala Met Thr Leu Thr Val Gln Ala Arg Leu Leu Leu Ser 530 535 540
Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln 545 550 555 560
Gln His Leu Leu Glu Leu Thr Val Trp Gly Ile Lys Glu Leu Glu Ala 565 570 575
Arg Val Leu Ala Val Glu Arg Tyr Leu Arg Asp Gln Glu Leu Leu Gly 580 585 590
Ile Trp Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Ala Val Pro Trp 595 600 605
Asn Ala Ser Trp Ser Asn Lys Ser Leu Asn Lys Ile Trp Asp Asn Met 610 615 620
Thr Trp Ile Glu Trp Asp Arg Glu Ile Asn Asn Tyr Thr Ser Ile Ile 625 630 635 640
Tyr Ser Leu Ile Glu Glu Ser Gln Asn Gln Glu Lys Asn Glu Gln 645 650 655
Glu Leu Leu Glu Leu Asp Lys Trp Ala Ser Leu Trp Asn Trp Phe Asp 660 665 670
Ile Thr Lys Trp Leu Trp Tyr Ile Lys Ile Phe Ile Met Ile Val Gly
675  680  685
Gly Leu Ile Gly Leu Arg Ile Val Phe Ser Val Leu Ser Ile Val Asn
690  695  700
Arg Val Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr His Leu Pro
705  710  715  720
Ser Ser Arg Gly Pro Asp Arg Pro Gly Gly Ile Glu Glu Gly Gly Gly
725  730  735
Glu Arg Asp Arg Asp Arg Ser Gly Pro Leu Val Asn Gly Phe Leu Ala
740  745  750
Leu Ile Trp Val Asp Leu Arg Ser Leu Phe Leu Phe Ser Tyr His Arg
755  760  765
Leu Arg Asp Leu Leu Ile Val Met Arg Ile Val Glu Leu Gly
770  775  780
Leu Ala Gly Gly Trp Glu Val Leu Lys Tyr Trp Trp Asn Leu Leu Gln
785  790  795  800
Tyr Trp Ser Gln Glu Leu Lys Asn Ser Ala Val Ser Leu Leu Asn Ala
805  810  815
Thr Ala Val Ala Val Ala Glu Gly Thr Asp Arg Val Ile Glu Glu Val Leu
820  825  830
Gln Arg Ala Val Arg Ala Ile Leu His Ile Pro Arg Arg Ile Arg Gln
835  840  845
Gly Leu Glu Arg Ala Leu Leu
850  855
WHAT IS CLAIMED IS:

1. A substantially pure preparation of a molecular clone capable of yielding after transfection into recipient cells active cultures of the Human Immunodeficiency Virus Type 1 (HIV-1) virus strain MN-ST1, having the identifying characteristics of ATCC 40889.

2. A substantially pure preparation of DNA containing the envelope and rev coding sequences of the (HIV-1) virus strain BA-L, having the identifying characteristics of ATCC 40890.

3. A DNA segment encoding an envelope (env) protein of MN-ST1.

4. The DNA segment according to claim 3 having the sequence given in Table III.

5. A DNA segment encoding an env protein of BA-L.

6. A DNA segment according to claim 5 having the sequence given in Table III.

7. A purified MN-ST1 env protein.

8. The protein according to claim 7 having the sequence given in Table II.


10. The protein according to claim 9 having the sequence given in Table III.

11. A DNA construct comprising:
   i) the DNA segment according to claim 3; and
   ii) a vector.

12. The DNA construct according to claim 11 further comprising a DNA segment encoding a rev protein and a rev-responsive region.

13. A DNA construct comprising:
   i) the DNA segment according to claim 5; and
   ii) a vector.

14. The DNA construct according to claim 13 further comprising a DNA segment encoding a rev protein and a rev-responsive region.
15. A recombinantly produced MN-ST1 env protein.
17. A host cell stably transformed with said recombinant DNA construct according to claim 11 or claim 13, in a manner allowing expression of said viral protein encoded in said recombinant DNA molecule.
18. A method of producing a recombinant HIV-1 virus strain MN-ST1 protein comprising culturing said host cells according to claim 17, in a manner allowing expression of said viral protein and isolating said viral protein.
19. A vaccine for mammals against HIV-1 infection comprising a non-infectious antigenic portion of said MN-ST1 virus strain according to claim 1, in an amount sufficient to induce immunization against said infection, and a pharmaceutically acceptable carrier.
20. A vaccine for mammals against HIV-infection comprising a non-infectious antigenic portion of said BA-L virus strain according to claim 2 in an amount sufficient to induce immunization against said infection, and a pharmaceutically acceptable carrier.
21. The vaccine according to claim 19 or claim 20 which further comprises an adjuvant.
22. A vaccine for mammals against HIV-1 infection comprising at least 5 amino acids of a MN-ST1 virus strain env protein, in an amount sufficient to induce immunization against said infection, and a pharmaceutically acceptable carrier.
23. A vaccine for mammals against HIV-1 infection comprising at least 5 amino acids of a BA-L virus strain env protein, in an amount sufficient to induce immunization against said infection, and a pharmaceutically acceptable carrier.
24. The vaccine according to claim 22 or 23 wherein said protein is a recombinantly produced protein.
25. A method of testing candidate vaccines against HIV-1 infection comprising administering said vaccine and the MN-ST1 virus strain according to claim 1,
to a test mammal and detecting the presence or absence of said infection.

26. A method of screening drugs for their ability to effect HIV-1 activity comprising contacting host cells according to claim 17, with said drug under conditions such that said activity of said virus can be effected.

27. A bioassay for the detection of HIV-1 in a biological sample comprising the steps of:

   i) coating a surface with at least 5 amino acids of a env protein from MN-ST1 or BA-L virus;

   ii) contacting said coated surface with said sample; and

   iii) detecting the presence or absence of a complex formed between said protein and antibodies specific thereto present in said sample.
<table>
<thead>
<tr>
<th>BAFINAL</th>
<th>0</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>2500</th>
<th>3000</th>
<th>3500</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCOT22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAE</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HHA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIND</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HINF1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPH</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KPN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ksp632</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAM</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBO</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MST</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PFL M1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 5B**

**SUBSTITUTE SHEET**
<table>
<thead>
<tr>
<th>0</th>
<th>500</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
<th>2500</th>
<th>3000</th>
<th>3500</th>
</tr>
</thead>
<tbody>
<tr>
<td>PVU 2</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>RRU 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SAC 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SAU 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SAU 3A</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SCA 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SFA N1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SNA 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SPE 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>SSP 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>STU 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>TAQ 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>TTH111</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>XBA 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>XCM 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>XHO 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>XHO 2</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>XMN 1</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

FIG. 5C

SUBSTITUTE SHEET
INTERNATIONAL SEARCH REPORT

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) 5

According to International Patent Classification (IPC) or to both National Classification and IPC:

IPC(5): C07H 15/12; C12N 3/10; 7/02, 7/04, 15/49;
C07K 3/12, 13/00, 17/00; C12Q 1/78; A61K 39/10: G01N 33/53

II. FIELDS SEARCHED

Classification System

Minimum Documentation Searched 7

<table>
<thead>
<tr>
<th>Classification System</th>
<th>Classification Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S.</td>
<td>435/7.1, 235.1, 236, 240.1; 530/350; 536/27; 424/88</td>
</tr>
</tbody>
</table>

Documentation Searched other than Minimum Documentation
to the extent that such Documents are Included in the Fields Searched 8

DIALOG DATABASES: BIOSIS PREVIEWS 1985+, MEDLINE 1975+, NTIS, AIDSLINE, CA SEARCH, BIOTECHNOLOGY ABSTRACTS 1982+

III. DOCUMENTS CONSIDERED TO BE RELEVANT 9

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of Document,10 with indication, where appropriate, of the relevant passages 12</th>
<th>Relevant to Claim No 13</th>
</tr>
</thead>
</table>

* Special categories of cited documents 12

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on prior composition or which is cited to establish the publication date of another publication or other special reason (as specified)
- "G" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

IV. CERTIFICATION

Date of the Actual Completion of the International Search
15 JANUARY 1991

Date of Mailing of this International Search Report
30 JANUARY 1992

International Searching Authority
ISA/US

Signature of Authorized Officer
JOHNNY F. RAILLEY II

Form PCT/ISA/210 (second sheet) (Rev.11-87)
|---|---|---|

**V. OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claim numbers , because they relate to subject matter not required to be searched by this Authority, namely:

2. Claim numbers , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

   Claims 25 and 26 are so vague and indefinite as to prevent a meaningful and thorough search.

3. Claim numbers , because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

**VI. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING**

This International Searching Authority found multiple inventions in this international application as follows:

See attachment

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

**Remark on Protest**

- The additional search fees were accompanied by applicant’s protest.
- No protest accompanied the payment of additional search fees.
Attachment to Form PCT/ISA/210, Part VI
Continuation of OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Group I: Claims 1 and 25, drawn to a first product, cloned HIV-1 strain MN-ST1, and the first appearing use of the product, a method of testing vaccines against HIV-1 using strain MN-ST1.

Group II: Claims 2, 5, 6, 13 and 14, drawn to a second product, HIV-1 strain BA-L env and rev coding sequences, DNA segments encoding the env gene, and vector constructs containing these sequences.

Group III: Claims 3, 4, 11 and 12, drawn to a third product, DNA encoding strain MN-ST1 env gene and vectors containing this env gene.

Group IV: Claim 17 (first species), drawn to a fourth product, host cells stably transformed with recombinant construct of claim 11.

Group V: Claim 17 (second species), drawn to a fifth product, host cells stably transformed with recombinant construct of claim 13.

Group VI: Claim 18 (first species), drawn to a method of use of the fourth product, host cells transformed with the recombinant construct of claim 11.

Group VII: Claim 18 (second species), drawn to a method of use of the fifth product, host cells transformed with the recombinant construct of claim 13.

Group VIII: Claims 7, 8 and 15, drawn to a sixth product, HIV-1 strain MN-ST1 env protein.

Group IX: Claims 9, 10 and 16, drawn to a seventh product, HIV-1 strain BA-L env protein.

Group X: Claims 19 and 21 (first species), drawn to an eighth product, vaccines using MN-ST1.

Group XI: Claims 20 and 21 (second species), drawn to a ninth product, vaccines using BA-L.
Attachment to Form PCT/ISA/210, Part VI
Continuation of OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

Group XII: Claims 22 and 24, drawn to a tenth product, vaccines using at least 5 amino acids of the env protein of MN-ST1.

Group XIII: Claim 23, drawn to an eleventh product, vaccines using at least 5 amino acids of the env protein of BA-L.

Group XIV: Claim 26, drawn to a twelfth product, a method of screening for drugs affecting HIV-1 activity.

Group XV: Claim 27, drawn to a thirteenth product, a bioassay to detect HIV-1 in biological samples.

The claims of Group I are drawn to a first product and a first specific method of use of the first product. Groups II-XV are drawn to separate products and methods of use of the products. PCT Rules 13.1 and 13.2 do not provide for multiple products and methods within a single general inventive concept. Note also 37 CFR § 1.475.
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of Document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to Claim No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>J. Virology, Vol. 63, No. 3, issued March 1989, H. Hadzopoulou-Cladaras et al, &quot;The rev (trs/art) Protein of Human Immunodeficiency Virus Type 1 Affects Viral mRNA and Protein Expression via a cis-Acting Sequence in the env Region,&quot; pages 1265-1274. See entire article.</td>
<td>2, 7, 9, 12, 14, 15, 16, 20, 21</td>
</tr>
<tr>
<td>Category</td>
<td>Citation of Document, with indication, where appropriate, of the relevant passages</td>
<td>Relevant to Claim No</td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Y</td>
<td>J. Virology, Vol. 64, No. 9, issued September 1990, P. J. Dillon et al, &quot;Function of the Human Immunodeficiency Virus Types 1 and 2 Rev Proteins Is Dependent on Their Ability To Interact with a Structured Region Present in env Gene mRNA,&quot; pages 4428-4437. See entire article.</td>
<td>2, 7, 9, 12, 14, 15, 16, 20, 21</td>
</tr>
<tr>
<td>Y</td>
<td>J. Virology, Vol. 64, No. 5, issued May 1990, A. Aldovini et al, &quot;Mutations of RNA and Protein Sequences Involved in Human Immunodeficiency Virus Type 1 Packaging Result in Production of Noninfectious Virus,&quot; pages 1920-1926. See entire article.</td>
<td>19</td>
</tr>
</tbody>
</table>