US 20020172995A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2002/0172995 A1**

**Shao et al.** (43) **Pub. Date: Nov. 21, 2002**

(54) **ISOLATED HUMAN SECRETED PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND USES THEREOF**

(76) Inventors: **Wei Shao**, Frederick, MD (US); **Jane Ye**, Boyds, MD (US); **Fangcheng Gong**, Germantown, MD (US); **Valentina Di Francesco**, Rockville, MD (US); **Ellen M. Beasley**, Darnestown, MD (US)

Correspondence Address:
**CELERA GENOMICS CORPORATION**
**45 West Gude Dr. C2-4#20**
**Rockville, MD 20850 (US)**

(21) Appl. No.: **09/858,546**

(22) Filed: **May 17, 2001**

**Publication Classification**

(51) **Int. Cl.**[7] ..... **C12Q 1/68**; G01N 33/53; A01K 67/00; C07H 21/04

(52) **U.S. Cl.** ..... **435/69.1**; 435/325; 435/6; 435/7.1; 530/350; 536/23.1; 800/8

(57) **ABSTRACT**

The present invention provides amino acid sequences of peptides that are encoded by genes within the human genome, the secreted peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the secreted peptides, and methods of identifying modulators of the secreted peptides.

```
   1 TTGGTGATCG TGAGTGATGG CAAGAGGATT TAGCCTCGGC ATTAACTTGG
  51 AGCGGAGTGC AGGGGGGCAG TGAAGCGCCC GCCATCTGGC CCGCGCCGCG
 101 CCGGGGGGAT GCCCCGGCTC CCCGACGAGA CGCCGCGAAG CCCACCCGGG
 151 CCGGGGGCTG CCCGGCGCCC GAGCGCGGGT CCTCCCCGGG CCGCCCAGGG
 201 GGGCCAAAAA GTTTGCACTT GTTAGCGGCG ACCTCCCGCT CAGCCCGGGC
 251 GGGCGATGCG GGCGGCGCGG GCGGCCCCCT CCCCCGGCCC GCCTCTCCGG
 301 GACGGCTGCG GGCGGCCCCC CCGGCGGCCG GAGGGCTCCC TGGCCCCGAT
 351 CTGACGGCGG CGGCGGCGGC GGCCACAGCG GCGGGAGCGG CGCGGGGAAG
 401 GAGCAGCGGC TCGCAGCCCT CGGCCCGCGC CCCCACCCAG CGCCAGCCCG
 451 AGGGGGGAGG CGCAGCGCCG GAGGGTGGCG GTCCTCGGCC CTCCCAGGTC
 501 TCCGCGCCGG GAAGCCGCTC CGAGCCGGGA TTGTGCCTAT GATTGGGGGG
 551 CTGTTTCTCA GTGCCTGGCT CTTCATGGTT GGCCAGCACG TGGAGCACTC
 601 TGTTCATCTG TACATATAGA CAAGTCAGAC CCCAGGTACC TGGATGGATT
 651 GAGAGCTGAG ATCTCAGAAA CTTCATAATA AGTTGCCGAT GGCTACCAGC
 701 CAAGGCTGGA GGGTTGTTTT GCCGTTGTGT TGAGCACGTC ACCCATTAAG
 751 AGCCCTTTAA AGACCTGGAT TGATTGGAAG GACAAAAATT AAAAGCAATC
 801 TGATCCAGCC TCATGCAGGA TCCCTGCGGA TTTTCTCCTT ATCCCATTTC
 851 CATCCACTGT CACAATTTGA GAATCTGCCT GATTTGATCA GATTCACCTC
 901 CAGGGGAGGT GTGATACCAG GGTTAGGAGG ACGTGAAGTT ATGGGCAACT
 951 TTCTGATCTG TCCATCAGCA GTCTGAGAAA CGCTGGCTCT GAATTTTCCG
1001 TGTCGGCCTT TTGGAAACAA CAAGTTCCTC GCTGTTTGCA AAGCTTCAGT
1051 GCTCGGGTCC CTGGGACACC CCGGCCACCC TCGCCTGGTA GATGTGGCAT
1101 TTCCATGCTG AGGCCGCGAG TCCCGCCTGA CCCCGTCGCT GCCTCTCCAG
1151 GGCTTCTCTG GGCCGCGCCT CTGCAGACTG CGCAGCCATG CTGCATCTGC
1201 TGGCGCTCTT CCTGCACTGC CTCCCTCTGG CCTCTGGGGA CTATGACATC
1251 TGCAAATCCT GGGTGACCAC AGATGAGGGC CCCACCTGGG AGTTCTACGC
1301 CTGCCAGCCC AAGGTGATGC GCCTGAAGGA CTACGTCAAG GTGAAGGTGG
1351 AGCCCTCAGG CATCACATGT GGAGACCCCC CTGAGAGGTT CTGCTCCCAT
1401 GAGAATCCCT ACCTATGCAG CAACGAGTGT GACGCCTCCA ACCCGGACCT
1451 GGCCCACCCG CCCAGGCTCA TGTTCGACAA GGAGGAGGAG GGCCTGGCCA
1501 CCTACTGGCA GAGCATCACC TGGAGCCGCT ACCCCAGCCC GCTGGAAGCC
1551 AACATCACCC TTTCGTGGAA CAAGACCGTG GAGCTGACCG ACGACGTGGT
1601 GATGACCTTC GAGTACGGCC GGCCCACGGT CATGGTCCTG GAGAAGTCCC
1651 TGGACAACGG GCGCACCTGG CAGCCCTACC AGTTCTACGC CGAGGACTGC
1701 ATGGAGGCCT TCGGTATGTC CGCCCGCCGG GCCCGCGACA TGTCATCCTC
1751 CAGCGCGCAC CGCGTGCTCT GCACCGAGGA GTACTCGCGC TGGGCAGGCT
1801 CCAAGAAGGA GAAGCACGTG CGCTTCGAGG TGCGGGACCG CTTCGCCATC
1851 TTTGCCGGCC CCGACCTGCG CAACATGGAC AACCTCTACA CGCGGCTGGA
1901 GAGCGCCAAG GGCCTCAAGG AGTTCTTCAC CCTCACCGAC CTGCGCATGC
1951 GGCTGCTGCG CCCGGCGCTG GGCGGCACCT ATGTGCAGCG GGAGAACCTC
2001 TACAAGTACT TCTACGCCAT CTCCAACATC GAGGTCATCG GCAGGTGCAA
2051 GTGCAACCTG CATGCCAACC TGTGCTCCAT GCGCGAGGGC AGCCTGCAGT
2101 GCGAGTGCGA GCACAACACC ACCGGCCCCG ACTGCGGCAA GTGCAAGAAG
2151 AATTTCCGCA CCCGGTCCTG GCGGGCCGGC TCCTACCTGC CGCTGCCCCA
2201 TGGCTCTCCC AACGCCTGTG CCGCTGCAGG TTCCTTTGGC AACTGCGAAT
2251 GCTACGGTCA CTCCAACCGC TGCAGCTACA TTGACTTCCT GAATGTGGTG
2301 ACCTGCGTCA GCTGCAAGCA CAACACGCGA GGTCAGCACT GCCAGCACTG
2351 CCGGCTGGGC TACTACCGCA ACGGCTCGGC AGAGCTGGAT GATGAGAACG
2401 TCTGCATTGA GTGTAACTGC AACCAGATAG GCTCCGTGCA CGACCGGTGC
2451 AACGAGACCG GCTTCTGCGA GTGCCGCGAG GGCGCGGCGG GCCCCAAGTG
2501 CGACGACTGC CTCCCCACGC ACTACTGGCG CCAGGGCTGC TACCCCAACG
2551 TGTGCGACGA CGACCAGCTG CTGTGCCAGA ACGGAGGCAC CTGCCTGCAG
2601 AACCAGCGCT GCGCCTGCCC GCGCGGCTAC ACCGGCGTGC GCTGCGAGCA
2651 GCCCCGCTGC GACCCCGCCG ACGATGACGG CGGTCTGGAC TGCGACCGCG
2701 CGCCCGGGGC CGCCCCGCGC CCCGCCACCC TGCTCGGCTG CCTGCTGCTG
2751 CTGGGGCTGG CCGCCCGCCT GGGCCGCTGA GCCCCGCCCG GAGGACGCTC
2801 CCCGCACCCG GAGGCCGGGG GTCCCGGGGT CCCGGGGCGG GGCCGGCGTC
2851 CGAGGCCGGG CGGTGAGAAG GGTGCGGCCC GAGGTGCTCC CAGGTGCTAC
2901 TCAGCAGGGC CCCCCGCCCG GCCCGCGCTC CCGCCCGCAC TGCCCTCCCC
2951 CCGCAGCAGG GGCGCCTTGG GACTCCGGTC CCCGCGCCTG CGATTTGGTT
3001 TCGTTTTTCT TTTGTATTAT CCGCCGCCCA GTTCCTTTTT TGTCTTTCTC
3051 TCTCTCTCTT TTTTTTTTTT TTTTTCTGGC GGTGAGCCAG AGGGTCGGGA
3101 GAAACGCTGC TCGCCCCACA CCCCGTCCTG CCTCCCACCA CACTTACACA
```

FIGURE 1

```
 3151 CACGGGACTG TGGCCGACAC CCCCTGGCCT GTGCCAGGCT CACGGGCGGC
 3201 GGCGGACCCC GACCTCCAGT TGCCTACAAT TCCAGTCGCT GACTTGGTCC
 3251 TGTTTTCTAT TCTTTATTTT TCCTGCAACC CACCAGACCC CAGGCCTCAC
 3301 CGGAGGCCCG GTGACCACGG AACTCACCGT CTGGGGGAGG AGGAGAGAAG
 3351 GAAGGGGTGG GGGGCCTGGA AACTTCGTTC TGTAGAGAAC TATTTTTGTT
 3401 TGTATTCACT GTCCCCTGCA AGGGGGACGG GGCGGGAGCA CTGGTCACCG
 3451 CGGGGGCCGA TGGTGGAGAA TCCGAGGAGT AAAGAGTTTG CTCACTGCTG
 3501 CCTCCAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAA
(SEQ ID NO:1)
```

**FEATURES:**

```
5'UTR:       1 - 1187
Start Codon: 1188
Stop Codon:  2778
3'UTR:       2781
```

**Homologous proteins:**

Top 10 BLAST Hits

| | Score | E |
|---|---|---|
| CRA\|160000026582933 /altid=gi\|9909148 /def=dbj\|BAB12010.1\| (AB0... | 698 | 0.0 |
| CRA\|18000005211756 /altid=gi\|4406683 /def=gb\|AAD20057.1\| (AF131... | 596 | e-169 |
| CRA\|160000026582932 /altid=gi\|9909146 /def=dbj\|BAB12009.1\| (AB0... | 458 | e-127 |
| CRA\|160000026582930 /altid=gi\|9909142 /def=dbj\|BAB12007.1\| (AB0... | 445 | e-123 |
| CRA\|18000005222985 /altid=gi\|7662426 /def=ref\|NP_055732.1\| KIAA... | 444 | e-123 |
| CRA\|160000026582929 /altid=gi\|9909140 /def=dbj\|BAB12006.1\| (AB0... | 443 | e-123 |
| CRA\|160000026582931 /altid=gi\|9909144 /def=dbj\|BAB12008.1\| (AB0... | 442 | e-122 |
| CRA\|160000026582928 /altid=gi\|9909138 /def=dbj\|BAB12005.1\| (AB0... | 441 | e-122 |
| CRA\|18000005101082 /altid=gi\|2394302 /def=gb\|AAB70266.1\| (AF017... | 216 | 1e-54 |
| CRA\|18000004965556 /altid=gi\|2497605 /def=sp\|Q90922\|NET1_CHICK ... | 215 | 2e-54 |

EST:

| | | |
|---|---|---|
| gi\|11284965 /dataset=dbest /taxon=96... | 1281 | 0.0 |
| gi\|12096965 /dataset=dbest /taxon=96... | 1122 | 0.0 |
| gi\|430880 /dataset=dbest /taxon=9606 /... | 317 | 6e-84 |

**EXPRESSION INFORMATION FOR MODULATORY USE:**

```
library source:
gi|11284965| glioblastoma with EGFR amplification
gi|12096965| adrenal cortex carcinoma cell line
gi|430880|
```

FIGURE 1

```
   1 MLHLLALFLH CLPLASGDYD ICKSWVTTDE GPTWEFYACQ PKVMRLKDYV
  51 KVKVEPSGIT CGDPPERFCS HENPYLCSNE CDASNPDLAH PPRLMFDKEE
 101 EGLATYWQSI TWSRYPSPLE ANITLSWNKT VELTDDVVMT FEYGRPTVMV
 151 LEKSLDNGRT WQPYQFYAED CMEAFGMSAR RARDMSSSSA HRVLCTEEYS
 201 RWAGSKKEKH VRFEVRDRFA IFAGPDLRNM DNLYTRLESA KGLKEFFTLT
 251 DLRMRLLRPA LGGTYVQREN LYKYFYAISN IEVIGRCKCN LHANLCSMRE
 301 GSLQCECEHN TTGPDCGKCK KNFRTRSWRA GSYLPLPHGS PNACAAAGSF
 351 GNCECYGHSN RCSYIDFLNV VTCVSCKHNT RGQHCQHCRL GYYRNGSAEL
 401 DDENVCIECN CNQIGSVHDR CNETGFCECR EGAAGPKCDD CLPTHYWRQG
 451 CYPNVCDDDQ LLCQNGGTCL QNQRCACPRG YTGVRCEQPR CDPADDDGGL
 501 DCDRAPGAAP RPATLLGCLL LLGLAARLGR
(SEQ ID NO:2)

FEATURES:
Functional domains and key regions:
[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site
Number of matches: 5
        1       122-125  NITL
        2       128-131  NKTV
        3       310-313  NTTG
        4       395-398  NGSA
        5       422-425  NETG

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site
Number of matches: 7
        1       178-180  SAR
        2       205-207  SKK
        3       239-241  SAK
        4       297-299  SMR
        5       327-329  SWR
        6       359-361  SNR
        7       375-377  SCK

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site
Number of matches: 12
        1         27-30  TTDE
        2         60-63  TCGD
        3         84-87  SNPD
        4       117-120  SPLE
        5       205-208  SKKE
        6       235-238  TRLE
        7       248-251  TLTD
        8       279-282  SNIE
        9       297-300  SMRE
        10      312-315  TGPD
        11      363-366  SYID
        12      416-419  SVHD

[4] PDOC00007 PS00007 TYR_PHOSPHO_SITE
Tyrosine kinase phosphorylation site
                228-234  RNMDNLY

[5] PDOC00008 PS00008 MYRISTYL
N-myristoylation site
Number of matches: 6
        1         58-63  GITCGD
        2       301-306  GSLQCE
        3       339-344  GSPNAC
        4       348-353  GSFGNC
        5       450-455  GCYPNV
        6       498-503  GGLDCD

[6] PDOC00021 PS00022 EGF_1
EGF-like domain signature 1
Number of matches: 3
        1       305-316  CECEHNTTGPDC
        2       427-438  CECREGAAGPKC
        3       475-486  CACPRGYTGVRC
```

FIGURE 2

```
[7] PDOC00021 PS01186 EGF_2
EGF-like domain signature 2
               475-486  CACPRGYTGVRC

[8] PDOC00961 PS01248 LAMININ_TYPE_EGF
Laminin-type EGF-like (LE) domain signature
Number of matches: 2
       1       373-409  CVSCKHNTRGQHCQHCRLGYYRNGSAELDDENVCIEC
       2       427-456  CECREGAAGPKCDDCLPTHYWRQGCYPNVC
```

**SignalP results:**

```
# Measure  Position  Value  Cutoff  Conclusion
  max. C    18       1.000   0.37   YES
  max. Y    18       0.943   0.34   YES
  max. S     4       0.977   0.88   YES
  mean S     1-17    0.942   0.48   YES
# Most likely cleavage site between pos. 17 and 18: ASG-DY
```

**Membrane spanning structure and domains:**

```
 Helix Begin   End   Score Certainity
     1   509   529   1.087 Certain
```

**BLAST Alignment to Top Hits:**

```
>CRA|160000026582933 /altid=gi|9909148 /def=dbj|BAB12010.1| (AB038667)
           Netrin-G1a [Mus musculus] /org=Mus musculus /taxon=10090
           /dataset=nraa /length=539
         Length = 539

 Score =  698 bits (1781), Expect = 0.0
 Identities = 312/515 (60%), Positives = 382/515 (73%), Gaps = 3/515 (0%)
 Frame = +3

Query: 1236 GDYDICKSWVTTDEGPTWEFYACQPKVMRLKDYVKVKVEPSGITCGDPPERFCSHENPYL 1415
          G YD+CKS + T+EG  W++ ACQP+   +  Y+KVK++P  ITCGDPPE FC+  NPY+
Sbjct: 28   GHYDVCKSLIYTEEGKVWDYTACQPESTDMTKYLKVKLDPPDITCGDPPESFCAMGNPYM 87

Query: 1416 CSNECDASNPDLAHPPRLMFDKEEEGLATYWQSITWSRYPSPLEANITLSWNKTVELTDD 1595
          C+NECDAS P+LAHPP LMFD E    +T+WQS TW  YP PL+ NITLSW+KT+ELTD+
Sbjct: 88   CNNECDASTPELAHPPELMFDFEGRHPSTFWQSATWKEYPKPLQVNITLSWSKTIELTDN 147

Query: 1596 VVMTFEYGRPTVMVLEKSLDNGRTWQPYQFYAEDCMEAFGMSARRARDMSSSSAHRVLCT 1775
          +V+TFE GRP  M+LEKSLD GRTWQPYQ+YA DC+ AF M  +  +D+S  +   ++CT
Sbjct: 148  IVITFESGRPDQMILEKSLDYGRTWQPYQYYATDCLHAFHMDPKSVKDLSQHTVLEIICT 207

Query: 1776 EEYSRWAGSKKEKHVRFEVRDRFAIFAGPDLRNMDNLYTRLESAKGLKEFFTLTDLRMRL 1955
          EEYS    S   K + FE++DRFA FAGP LRNM +LY +L++ K L++FFT+TDLR+RL
Sbjct: 208  EEYST-GYSTNSKIIHFEIKDRFAFFAGPRLRNMASLYGQLDTTKKLRDFFTVTDLRIRL 266

Query: 1956 LRPALGGTYVQRENLYKYFYAISNIEVIGRCKCNLHANLCSMREGSLQCECEHNTTGPDC 2135
          LRPA+G  +V   +L +YFYAIS+I+V GRCKCNLHA  C      L CECEHNTTGPDC
Sbjct: 267  LRPAVGEIFVDELHLARYFYAISDIKVRGRCKCNLHATSCLYDNSKLTCECEHNTTGPDC 326

Query: 2136 GKCKKNFRTRSWRAGSYLPLPHGSPNACAAA-GSFGNCECYGHSNRCSYIDFLNVVTCVS 2312
          GKCKKN++ R W  GSYLP+P G+ N C  +  S GNCEC+GHSNRCSYID LN V CVS
Sbjct: 327  GKCKKNYQGRPWSPGSYLPIPKGTANTCIPSISSIGNCECFGHSNRCSYIDLLNTVICVS 386

Query: 2313 CKHNTRGQHCQHCRLGYYRNGSAELDDENVCIECNCNQIGSVHDRCNETGFCECREGAAG 2492
          CKHNTRGQHC+ CRLGY+RN SA+LDDENVCIEC CN +GS+HDRCN +GFCEC+ G  G
Sbjct: 387  CKHNTRGQHCELCRLGYFRNASAQLDDENVCIECYCNPLGSIHDRCNGSGFCECKTGTTG 446

Query: 2493 PKCDDCLPTHYWRQGCYPNVCDDDQLLCQNGGTCLQNQRCACPRGYTGVRCEQPRCDPAD 2672
          PKCD+CLP + W  GC PNVCD++ L CQNGGTC  N RCACP  YTG+ CE+ RC+ A
Sbjct: 447  PKCDECLPGNSWYYGCQPNVCDNELLHCQNGGTCQNNVRCACPDAYTGILCEKLRCEEA- 505

Query: 2673 DDGGLDCDRAPGAAPR--PATLLGCLLLLGLAARL 2771
            G    +   GA PR  PA LL   +LLG A  L
Sbjct: 506  --GSCGSESGQGAPPRGSPALLL-LTMLLGTAGPL 537 (SEQ ID NO:4)
```

FIGURE 2

```
>CRA|18000005211756 /altid=gi|4406683 /def=gb|AAD20057.1| (AF131842)
           Unknown [Homo sapiens] /org=Homo sapiens /taxon=9606
           /dataset=nraa /length=260
         Length = 260

 Score =  596 bits (1519), Expect = e-169
 Identities = 260/260 (100%), Positives = 260/260 (100%)
 Frame = +3

Query: 1998 LYKYFYAISNIEVIGRCKCNLHANLCSMREGSLQCECEHNTTGPDCGKCKKNFRTRSWRA 2177
          LYKYFYAISNIEVIGRCKCNLHANLCSMREGSLQCECEHNTTGPDCGKCKKNFRTRSWRA
Sbjct: 1    LYKYFYAISNIEVIGRCKCNLHANLCSMREGSLQCECEHNTTGPDCGKCKKNFRTRSWRA 60

Query: 2178 GSYLPLPHGSPNACAAAGSFGNCECYGHSNRCSYIDFLNVVTCVSCKHNTRGQHCQHCRL 2357
          GSYLPLPHGSPNACAAAGSFGNCECYGHSNRCSYIDFLNVVTCVSCKHNTRGQHCQHCRL
Sbjct: 61   GSYLPLPHGSPNACAAAGSFGNCECYGHSNRCSYIDFLNVVTCVSCKHNTRGQHCQHCRL 120

Query: 2358 GYYRNGSAELDDENVCIECNCNQIGSVHDRCNETGFCECREGAAGPKCDDCLPTHYWRQG 2537
          GYYRNGSAELDDENVCIECNCNQIGSVHDRCNETGFCECREGAAGPKCDDCLPTHYWRQG
Sbjct: 121  GYYRNGSAELDDENVCIECNCNQIGSVHDRCNETGFCECREGAAGPKCDDCLPTHYWRQG 180

Query: 2538 CYPNVCDDDQLLCQNGGTCLQNQRCACPRGYTGVRCEQPRCDPADDDGGLDCDRAPGAAP 2717
          CYPNVCDDDQLLCQNGGTCLQNQRCACPRGYTGVRCEQPRCDPADDDGGLDCDRAPGAAP
Sbjct: 181  CYPNVCDDDQLLCQNGGTCLQNQRCACPRGYTGVRCEQPRCDPADDDGGLDCDRAPGAAP 240

Query: 2718 RPATLLGCLLLLGLAARLGR 2777
          RPATLLGCLLLLGLAARLGR
Sbjct: 241  RPATLLGCLLLLGLAARLGR 260 (SEQ ID NO:5)
```

**HMM results:**

| Model | Description | Score | E-value | N |
|---|---|---|---|---|
| PF00053 | Laminin EGF-like (Domains III and V) | 72.6 | 2.8e-19 | 4 |
| PF00055 | Laminin N-terminal (Domain VI) | 49.6 | 4.1e-14 | 3 |
| PF02012 | BNR repeat | 17.3 | 0.0031 | 2 |
| PF00008 | EGF-like domain | 15.1 | 0.015 | 4 |
| CE00234 | E00234 Nicein | 7.3 | 0.49 | 1 |
| PF01414 | Delta serrate ligand | 6.3 | 1.5 | 2 |
| PF00059 | Lectin C-type domain | 3.3 | 3.5 | 1 |

Parsed for domains:

| Model | Domain | seq-f | seq-t | | hmm-f | hmm-t | | score | E-value |
|---|---|---|---|---|---|---|---|---|---|
| PF02012 | 1/2 | 25 | 36 | .. | 1 | 12 | [] | 7.3 | 3.5 |
| PF00055 | 1/3 | 60 | 69 | .. | 21 | 31 | .. | -1.2 | 1.3e+02 |
| PF00059 | 1/1 | 67 | 77 | .. | 101 | 111 | .] | 3.3 | 3.5 |
| PF00008 | 1/4 | 69 | 81 | .. | 1 | 13 | [. | -3.4 | 2.5e+03 |
| PF02012 | 2/2 | 152 | 163 | .. | 1 | 12 | [] | 10.0 | 0.53 |
| PF00055 | 2/3 | 137 | 201 | .. | 113 | 181 | .. | 40.8 | 1.9e-11 |
| PF00055 | 3/3 | 268 | 285 | .. | 247 | 264 | .] | 9.5 | 0.066 |
| PF00008 | 2/4 | 289 | 316 | .. | 1 | 45 | [] | 7.0 | 2.9 |
| PF00053 | 1/4 | 287 | 344 | .. | 1 | 59 | [] | 33.4 | 4.8e-08 |
| PF00053 | 2/4 | 353 | 406 | .. | 1 | 59 | [] | 26.9 | 3.6e-06 |
| PF01414 | 1/2 | 418 | 438 | .. | 47 | 67 | .] | 8.6 | 0.29 |
| PF00008 | 3/4 | 420 | 438 | .. | 12 | 45 | .] | 5.9 | 5.7 |
| CE00234 | 1/1 | 373 | 443 | .. | 1 | 70 | [. | 7.3 | 0.49 |
| PF00053 | 3/4 | 409 | 451 | .. | 1 | 59 | [] | 41.0 | 3.2e-10 |
| PF00008 | 4/4 | 456 | 486 | .. | 1 | 45 | [] | 27.1 | 6.2e-06 |
| PF01414 | 2/2 | 476 | 486 | .. | 57 | 67 | .] | 1.0 | 61 |
| PF00053 | 4/4 | 456 | 488 | .. | 14 | 42 | .. | -5.4 | 6.1e+03 |

FIGURE 2

```
   1 GAGTTTTAAT ACCCTGGCTA TCAGCCCCCC TTGGTTCCTG AGGACTCTTA
  51 AAAGAAAATA AAGCACATTG ATTCTATTTG TTTCTGGGAG CTGCAGTTTC
 101 TTAATAATAT CAGGTGAAGA TAAATTTTCC ACGGAGAAAA CGATCCTCCG
 151 GGATGCAGCT TCTTACTCTG AAAATTTCCC TGCCGACTCC TCACTCTCTG
 201 CGCTCCTCCT CGTTATCCGG GGACTCCTGC CTCTCTTCCC CCTTCTCTTT
 251 TTTCTTTTTG GCAGAACCCG CCTGCAATAT TCGTGTGCTG AGCTCGTAAT
 301 TCCCCCTGCG ATGCCAGCAA CGCCCAATTG ATTGACTAGT TGTAAACACA
 351 TTTTTCCCCT GGCAGATTTT GTTGTTGTTA GGGTTTTTAA AATTTATTTA
 401 TTTTCCAGGG AATGCGTGGC ATTTAAACCA ACAGGACTGC AATTAATAGA
 451 TTTGCGAGTT GCGCCGCGCG CGCCGCTCGC CCCAGCCTCC CGGCCTCCGG
 501 GCCTCGCTGC CTCCCCGCGC CCGGCGGCGT CCAGCGCCCT GCAAGCCCCG
 551 AGCAGCCGCG GGTCCTGCAG CTGAAGGAAG GTTGCAGCTG CGCCCTCCTT
 601 GCAAGCCGCA GCCCGGCGTC CTGGTTGTCC CAGCAGCCAG GAGATCCCTA
 651 CCTGTTAGTG AACAGTTAGG AGTCGACTGC TGGAAGAATT AATTAGGAAC
 701 GTGCTGTGCT CTGGGCAGCG CGAGCTCGGG TAGAGGCATC CAAACCTTTG
 751 CCGGCGGCGC TATTTTATTT TTACTACATT TTCTCAGGTT GCAAAAATAG
 801 ACACCGGGCA CGTTCTTTCT TAGAGTTTTC TAGCAAGGAG CGCCTTCAAG
 851 GCCAGGCAGG CTCTGTAACA GGTTCCCCTT TAAACAGCCA GAGGTGAGAC
 901 GGGGAAAATG GTCCTGGCTG GGTTCTCGTT CATCTCCATC AGCAGTCCTT
 951 CACCCAGAGA GAGGGGCAGG GGTCGCCCTA ACTCAGATGA ATGAGTCCCA
1001 TGCCTGGAGC CCTGGGGCCC TGGCTGGGGG CTGCTCCGAG CCTGAGGTGC
1051 TCAGGGCGCT CAGGGCAGCA AGTGTCCGCC ACTTCGGTTT GTCATTTTTG
1101 GCAGGAGCGT TTTTCTGTCT GGGTGGAGAA TGGAGTTTCA CGGAAACACA
1151 GTTAACTCTT CAGGGGCCTT GCAAGTACAG GAGGTGAAGA GGATGTCAGG
1201 GGAGAGCCAG GTCCAGACTG GACATTTGGG GTGGTTTGGG AAATCAAATG
1251 CAATCATCGA AGACATATTA ACCAGAATAA TTAATCATGC AGGCACTTTT
1301 TTACTGCAGT AACCTTTGCC CTATTGGCCA ATATTTTTGG CCAGAATCCC
1351 ATGCTGGCTG GACTTGGATT CTCCGGGTGA CGTATCCAGT GTCTGGAACA
1401 CACCACAGTA CTGCAGTCGT GTTATTTCCC AATGTAACAT TCATGTAACT
1451 GGTATCTATT TTGATATAAT ATATTAATTA TATCTATTTT GATTTTAAAT
1501 AATTAACAGA AGCTTAAAAT AGCACAGCAA TAATCGTAAT TGTAACCCAT
1551 TTATAGACCT ACCCTCTAAG AATGATTTGC AAAATGCTGG CCTTGATCAG
1601 AAAAATCTGA ACTCACAAAG CATTGTTACC TCTTTGGCAG TCTTCAATAT
1651 CCCTAGTTTC TTACAGTTAA AAAAAATTAA ATTTGCCATT TCAGATTGTG
1701 CCTATGATTG GGGGGCTGTT TCTCAGTGCC TGGCTCTTCA TGGTTGGCCA
1751 GCACGTGGAG CACTCTGTTC ATCTGTACAT ATAGGTATGG GTCCATCTGC
1801 ACCTATAGTT ACATGCTCAT CTTTGCCTTT AACAACTTTG ACATTCTGAC
1851 TTGACAGTCA TGGTATTTTA AAGCAACCAT TAAATCTTGG CTCCTGGGGA
1901 TGCTTTTGAA AGTCTCTGAC CCCCAAGCCG GGGCACTTCT GCTGAACTAA
1951 CACTCCCATA AATGAGAAAA AAATGCATCA CCTTTTAAAT AACATGCCCG
2001 ATCTCCAAAT GTGCAACTCT GATATTAATA AAAATAACCC AGTTTTCTCG
2051 GGAGACTTTG CTAATGCAGC CTCATTTTTT GCACATTTTG CCGGAGAGCT
2101 TTTGTTCTTA TACGTCTCTA TTCTCCCCCT CTTTAATTTG TTGCAGGTGT
2151 TGTTGCTAAT GAGCTCTCTC TCTCTCCTCC CCTCTTACAA TGAAAGACAA
2201 GTCAGACCCC AGGTACCTGG ATGGATTGAG AGCTGAGATC TCAGAAACTT
2251 CATAATAAGT TGCCGATGGC TACCAGCCAA GGTCAGCTGG GCCCCATTAG
2301 TGCCGGCCCC CACCAAAGCA GAACCAAATA GCCTTCTCCC CAGTGAACAC
2351 CTCAGTAGCT TTTCATTCTA GTCCAGTCAC ACAGCTGTTG CCACCTTAAG
2401 CTCATATGAA AGAAATCTCT TTTATTGGTC TGAGAACCCA AGTCCAGTCC
2451 CAAAGAGGGA ACATGTTTCC AGCTAACATG CCCACCTCCT GATTTTATTT
2501 TATCTTTACA ACGCAGGCTG GAGGGTTGTT TTGCCGTTGT GTTGAGCACG
2551 TCACCCATTA AGAGCCCTTT AAAGACCTGG ATTGATTGGA AGGACAAAAA
2601 TTAAAAGCAA TCTGATCCAG CCTCATGCAG GATCCCTGCG GATTTTCTCC
2651 TTATCCCATT TCCATCCACT GTCACAATTT GAGAATCTGC CTGATTTGAT
2701 CAGATTCACC TCCAGGGGAG GTGTGATACC AGGGTTAGGA GGACGTGAAG
2751 TTATGGGCAA CTTTCTGATC TGTCCATCAG CAGTCTGAGA AACGCTGGCT
2801 CTGAATTTTC CGTGTCGGCC TTTTGGAAAC AACAAGTTCC TCGCTGTTTG
2851 CAAAGCTTCA GTGCTCGGGT CCCTGGGACA CCCCGGCCAC CCTCGCCTGG
2901 TAGATGTGGC ATTTCCATGC TGAGGCCGCG AGTCCCGCCT GACCCCGTCG
2951 CTGCCTCTCC AGGGCTTCTC TGGGCCGCGC CTCTGCAGAC TGCGCAGCCA
3001 TGCTGCATCT GCTGGCGCTC TTCCTGCACT GCCTCCCTCT GGCCTCTGGG
3051 GACTATGACA TCTGCAAATC CTGGGTGACC ACAGATGAGG GCCCCACCTG
3101 GGAGTTCTAC GCCTGCCAGC CCAAGGTGAT GCGCCTGAAG GACTACGTCA
3151 AGGTGAAGGT GGAGCCCTCA GGCATCACAT GTGGAGACCC CCCTGAGAGG
3201 TTCTGCTCCC ATGTAAGTCC ACTTACTGCT CTTTTGTTTG CCCAGGCCAA
3251 GTGGGAGGAG GTCTGGAGGA GATCCTTGGG GTGGACGAGC AGAGCTGGAG
3301 GACTGAGATG CAAGGCTGAC TTTCCTGCCT CTTGACCAGG TCTGGACCAG
3351 ACCTGGACCA GGTCTTTGTC CCACCTTGGA AATGTGTCAA AACAGAGGAC
3401 ACCCTGAGGA CACTGGGGTC ATGTGACATT GTTCTCTGGG GTAGGGGCAT
3451 TCTCGGCCAG CTGGCCACTA GTTCAGTTCC CTCGGGAAGC CTATAGTATT
3501 CAGCTCCGCA GCCTTCAGGC TAAGCCCCAC CTCCTGTGTA GGAAGTCAGC
```

FIGURE 3

```
3551 ATTCTGGGCG AGTGAGCAAG ATGCTACCTG CAACATGATA CTGTAAGCTC
3601 CCTCTGTTCA TCCTTTCTGT GGTGCAACCT CTTAGCCCAC TCAATCCAAT
3651 CCAGCAGACG GTACCATGAA GCTAGAGCAT GCCAAGCACT GAGCTGGCCC
3701 TTTGCATGGG GAGGTTTGAC GGAGGCTCAG AGGGGATTCA CGCAGGATGA
3751 AGTTGGGGGT TAAGCTGGGA CAGACAGTGC AGCTTGGGTC CCCTTCCACT
3801 TCCATTCCTA CATGCAATGA TGGCAGCCCC TGGGTAAGTC GGGGGAAGGC
3851 AGACATTCAG GCGTTGTCCC TTGCCTCCCT AGCTAGAGAA GGAGGGGTCC
3901 TAGGGGCACG GAGTACTTCA GTACTCAAGT AATGTTAACA GCAACAACAA
3951 CAGCAGAGAT GGTTGCGTGC GAGCTGTCTG AGTGTGTGGG CCGAGCCTGA
4001 TGCCATTGCG AGGAGTCCTC CCCTCAGCTA GCCAGGTGGA TTCTGTGTTG
4051 TATAAAGCAG GAGCGTCAGG GGAGGGCTCT GGCCCAGTGA CGTCTGTGGG
4101 CTTCTGTTCT GTCATCTGCA AAATGGGCAC ACTAAGAGCA CACACTCCTA
4151 GGGTCATTGT GAGGAGTTTG TGGTTTAATT AACGTAGGTA AAGTGTTCGG
4201 AATAGGACCT GGCACAGAGT AAGTGCACGC AGATGTTAGC CGTTGTCATT
4251 CTGGTCATAC AGGTGGGGTA ACTGAGGCAG CCCAGGGGTG ACGGGTAAAG
4301 GCATCTGGCC AAGGTCACAC TCCAGGAGGT GGCGAAGCTG TGATTCCAGC
4351 TTAGAGTGGC TCCAATGTCT CTGAGCTAAG CTGCTTCCCA CTGGGCAGTG
4401 CTCCGGAGGC CGTCCCTGGC AGGGCAGGGC AGAGCTGGGT AGGGCAGCCA
4451 GGCTGCAGAA GCTCACAGGA GGGGCTTGAT GCCATCCCCC AGGCAGCTGG
4501 TACCTCTGCG TGTCCTTGGA GGAGCCTCCA GGTCTCTGGG TTTGTGGTGG
4551 GGCTGACCCG GTGCCCCCAC CTCAGAGTCC TGAGGACTTG ATCTCATGGG
4601 CCGGCTCTGC TCACATCACA GGGCAGTCAG CCTGAGAAGG TAGCTTCTTA
4651 CTCAGGCTTG TCAGTGGTGA TGAGGCTGTC ACTGTGGGTG GTGGCTGGGC
4701 CAGGGCCAGC TGGGGAGAGA GAGAGGGAGG GAGAGAGGGA AGGAGAAGGC
4751 GGCACGGAGC CAGGAGCTGG GGCTGGATAG TCTGTGGCCA TAACTGCCCC
4801 GGGGACCGCA GGGCCGAGCA AGGGGGCTGG GCTCTGGAAG CCAGGAGGAA
4851 GGCCAGGATA GGGGCTGGTA CTCAGTCCAC ATCTCAAAGC CGGTGGGAGG
4901 GTTCTCCACA CGCTCTCGGG CACGGTCAAC CTCTGTCTCT CGTATTAGAG
4951 TCTCGACTGT ATTTTCTCTC TTAAATATTA ATCACTGCCT TAACGTGCTT
5001 GGGGGAGCAG CTAAATCATA ATTCTAGGAC CAGCTTTGGG TCGAGGGCTT
5051 GAGGTGGGGA GATGACCCTC AGAGTCAAGT CCGAGGCCCT CTCCTTGCCA
5101 AAGCTGTCTA GGGTGTCATT GGCCCTGGAC CTCTGCCCTG CCCCCACCCT
5151 CAGACAGAGA ACCCAGTGCA GTGGGCGGCT GTTCTGGGGA GGTGGTCACC
5201 CCTCCCAGTC CCAGTGCCGG CAGAGCCTCA TCCCAGGCAG CCAGCCTCAA
5251 GCCCTGGGGT CTAGAAGAGT GCTCTCTCCC ATCCCCATCC CCTGCTCCTC
5301 TCCTGGGCAG ACAGGTGGGG AAGGCAGGGG AGAAAGAACA GTCCCTCCAC
5351 AATCTCCCAC ATGGGCAAGT CCTCGGCGTC TCGCCACCTG TGTGATGGAC
5401 TTAAATATTT CATCATGGGC TGCCATCCAG CTCTGCTTTG ATTACAAATG
5451 TGTGTCCGAT GAGGACGGGG AGGCCGCCGT GGCAGGTGGA CGGCAGCCTT
5501 TTGCAGGGCT GGCTTTTGGA GGGCTGGCTT TGGAGGGCTG GCTTTTGGAG
5551 GGGTGGTTTT TGGAGGGGGT GGTTACTTCT AGATAGATCT GGGTTCAAAC
5601 CCTGATGCCA CAGTTTATTG TGGCTTCCCC TGGGAGGGTC CGGGCAGTAG
5651 GATGCTCTGA GGTAGGTCCT ACTCTAATCT CATTTTTGGG AATAGAAAGT
5701 GTTGCCTAAA TCCTGCCAAC ATCACTCCGT TGAGAAGGCA GGCTGGGAAC
5751 TGCCTGTAGC CAAGGCCCAT GCCCACCCAC CAGGCGGCAC AGCCTCCCCT
5801 CTCTCCACTT CTGACCTCTC TGAGTTTCAC TGTCTTCCTC TGCCCAGTAG
5851 AAACCATAAT GTAGGAGTGT CACGGTGCTG GCGGGACGGC ACAGCCCGGC
5901 CGTGCTGCCG TGGTACGTGC GGACACATAG TAGGTACTCA GTGTGCAATA
5951 CTAGTCACCA GCTTATCACT GTTGTGATTA TGGCCTCGGT GATTTTCCTG
6001 GCTTTGCTTC TTCCCCGGGC TCCTTTGCCT CTCCCCAGCC TTGGGGACGC
6051 CACGCCTTGT ATTGCTAGTC TCTGGCAGCA TGTGGAAGAT GCAGGCTGGG
6101 GGAGCTTCCT CGCCTGCAGC CCCAGCAGCT GTTGTTAACC AGCCACCAGG
6151 GGGCGCCCTA ATGGCGGCCA GATGCCCACT GCCCCCTTAG CCTTGCCATC
6201 GCCCATGGGT CCTTGCTCCC CTTCGCCCCT TCAGCGGCAG GTGCTACGTG
6251 TGTCCCAACA TCCAGCCCAC GGGAGCGCAA GCCCGGCTGC ACTCCGAGGG
6301 TACAGGAGGG ACCCAGGGGC GGACGGCCTT TTCAGATGCG GGGTGCAGAC
6351 CCTCTCTCTT TCCCAGGCTT CCTGTCCTTC AGCAAGGCCA CCTGAGAACT
6401 GAATTGTAAA TTCCAGCCTT GCAAGGAAGG GGAGGAGGCT GAAAGACAAG
6451 ATAAATGAAT AAATAAAAAC CTATTGGCTC TAAATGCACA ATGAGAATTA
6501 ATAGGGATGA GCTCTCAACA GAGGATTCTT GGGCAAGGAC AAAATTGCAA
6551 TGAGGGCTGT GGGAGAAGAG AGGCGGCCCC CACCCACATT CCCAGGGCCT
6601 GCCCCTGAGG TCAGCCCAGC AAGGCTGCAG GCCGAGAAGG AGGCAGGAAG
6651 GGCGACGGGC AAAGGCATCT CAGCCCCTCA CAGGGCAACT GCTGCTTGCA
6701 GGACCCTTGG AGATGGGGAG GGCGTGACTG GCATTGGGAG GTGCCCCTTA
6751 GCGCTCTGCA AGACACGTCT CCCGCCAGCC ACCTGTGGTC CCCGTGAGAG
6801 GAAGGGGGAA ACCCTTCCTT TGTGCCTTGG AGACAAAAAC TAGCCACCCG
6851 AGACTCAGGG GTGAGCTGAC GAGCAGGTGA GAGAGAGAGA GAGAACTACT
6901 GCCCAGGCAG CAGGGTGCCA AGGGAATCCT GCCACTCCCA CTCAGAGCAA
6951 TCGCCTCAGT GCAGCAGCGC TGACCTCACC TTGATAGAGG TGCAGAATCC
7001 CAGACCCTGC CCCACAACTG CCGAGCCAGA CTCTGCCTTT TAACAGGAAG
7051 TCTCTGGAGC GCTGGGGTGG GGGTTGGTCC AGGATGGTGG ATGAAGCCTG
```

FIGURE 3

```
 7101 CCGTTGACGA AGGGGCCAGA GGTCTGTGGG TCCCATGGGG TCTGTGCAGG
 7151 TTAAGAGGAG AGGCATTGCT GAGGGGAGAC AGGAGCCAAC AGGAGAGCAC
 7201 TAGGGGGTCC TGGGAATGGA GGGATGGAGC TGGGAGCCGA GATTCCCAGA
 7251 TGCGGGCCTT GCTTGCCAAG GCGGGGGGTC TGTGCTTCTG AGCAGGAGCC
 7301 AGTGGAAGGT TCCAGAGAGA GGCAGCACTA CCTCCTGACA AGAGCTACAG
 7351 GCCTAGAATC TGAAGCCCTG AGCTCACATC CCAGTGTTGT CACTTGGGGG
 7401 CTGGGTGGTC TGAGGTGAGT CATTCAAATT CTCCAAGCCT CAGTTTCCCT
 7451 ATCTGTAAAA GGGGTATGAT CTTCACTGCC TTGCCCACCT CACCGACTTC
 7501 CTCGGAGAGG AAACGCCTCT CACAACATAA CAGAAGGTGC TTGGCAGAGA
 7551 CTGTAAGTTC TTTTCTAGTG TGAGGCTTGA ATATTATTGT TTATCCAATC
 7601 TTGATAAGCT TGACATTTCT ACATTGACAT GGGTGAACAT TCAAATTTCA
 7651 GCCATGTTGG AGTCTCCAGC ATACAGAGGC ACCTTGGCTT TGGTGGACAG
 7701 GAAAGCCAAG GGCAGGCTGT GAACGTGCAT CTTTGAACAC ACATCTTACT
 7751 CAGTTAGAAC CTGCTTTAGC TTCTCTGTGC CTCAGTTTAC CCAGCTGCAG
 7801 AGCAAATGGT TCCATTCCTC CCCCAGGGGG TCTGCTGGGC TGTTGAATGC
 7851 AGACAGCAGG CAGGTCCTGG AGTCCTTGAG CCACTCCTCA CTGTGACATT
 7901 GAACACCACA TGCAGAATCA AGGTCACAGA CACTCTGCCT CCTGCCAGAC
 7951 CTTGAGGATC ATGAAATGTT TCCATCCAAC AATGCAGAGC TGACCGGCCC
 8001 GATGGCTCTG GCCTCCTTCC CCCTCTGTGT TTCTGGCCCC AGCCTCCCTC
 8051 CTCTCGCCGT CACAACCCTC CAGTCGGCAG ATGTTTCCAG GTATGACCTC
 8101 ATGCCAGGCA CAGTTGTGAT GGGAAAGTGA GATGACTAGA GATAAATTAA
 8151 GTGAAAAACA AGAGAAAAAC ACTACAGACA TTGAAGACAA TACCCAGCTA
 8201 ACCACAAATG GTGACACCGG GGCCTGATAT GATAGGAGCT GGGGAAGGAG
 8251 GAGGGCGTGC CCCTCCCCAT GAACTCTCCG GGCGCTCCAG CCTTGGCCCG
 8301 TGCTGTCTCC TCCACCCGAA TGCCTGCCCC CTCCTTCCTT TCTTTTCAAG
 8351 GCTTCCGCTT CTACCTGGAT AACCCTGGCT CCTTTTTCAA GGTTCAGGTT
 8401 GGACCCTACC TCCTCCTCCA GGAGGCCTTC TCTGACTCCT TTGCCCAGTG
 8451 GGGTGAGGAG GCCCTCATTT GGTCATCATC CTGCTTACTG ACTGGCCTAT
 8501 CAGTGAAGAG TTCTGGCCCT AGGAATGGAG AGTTTAGAAT CCCGTGCTGT
 8551 GTGATCTCTG ATGTGAGACC TAACCTCTCT GAACCTTAGT TTCTCAATGG
 8601 GAGCTATTAA TAGAACGTGC ATAGGAGACT GTCATGAGCA CGGTGCTTGG
 8651 CACATAGTAG TTGCTCACTG CGTGCTGGCT GTTGGGACCA ATTCTGTAAC
 8701 TTTAGATGTC TGCTTTTCCC CTGGACTGTG AGCTCCTCAA AGGCAGGGCT
 8751 GTCTCGCTAT CTTTGCATCC CCAGCACCTC CCTCTGGACT TGATCAGTGC
 8801 CCAGTAAGCA TTTCCTGAAT GAATGGATGC ATGGGAGGAA GCATGGTGTG
 8851 ATCCAGGAAC GCTTCTCAGA GGAGGCAGGA TCAGGGCTGT ATTTTTTTTT
 8901 TTTTTTTTTT TTTTTTGGAG TCAAAGTCTC GCTCTGTCGC CCAGGCTGGA
 8951 GTGCAGCGGC GCAATCTCGG CTCGCTGCAA GCTCTGCCTC CTGGGTTCAC
 9001 GCCATTCTCC TGCCTCAGCC TCCCCAGTAG CTGGGACTAC AGGCACCTGC
 9051 CACCACGCCA GGCTAATTTT TTGTATTTTT AGTAGAGACG GGGTTTCACT
 9101 GTGTTAGCCA GGATGGTCTC GATCTCCTGA CCTCGTGATC CACCCACCTC
 9151 AGCCTCCCAA AGTGCTGGGA TTACAGGAGT GAGCCACTGC ACTTGGCCAG
 9201 GGCTGAGTCT TGAAGGAAAA ACTGGGGTTT GGGTCAGGAC AGAGGAGACC
 9251 CTGGAAGCCC CTGCTTCTCT CCACTGCAGT CCCTGTTCTG TGGGATTTGC
 9301 GATTGGATGA AGCCGGGAGG TTTGCACAAC TCTGTCCTTA AGTCAGTTGC
 9351 AAGTGACTTC GGCACCTGAG CTGCACCAGC CGTTAAAGCC ACTCAGTCTC
 9401 TTGAAATGCC CGAGGCAGGG CCCAGCCTAG GACAAGAATA GTTCTGTGAA
 9451 ATGACATCTT GTTGCACAGT GAAGTCTCCC TCCTGGGCAG TAGACAATGA
 9501 GAAGACCGAG GCCCGGGGCC CAGGGAGTGA GACCCTTGCT TCTGACTTCC
 9551 CTTGAGGGAA TGAGGTTGGG TCCAGACACC CCGTGGAAGG CAGGCAGCTG
 9601 TGTGAAAGGG CCCAGACGGG ACATCTTTCC AAAGAATGTC AGAGACTTAG
 9651 AGACCCCCAG ACCTTTCCGG TTTGCGCATC CCCACCTTCC CAGGCTGTCT
 9701 TCCTCTATGC TTCCTAACTC TGATGTTTAA TCCATTTCCC TTTTTCTCAT
 9751 TTACTGTGGG TATAATGACA AGCTGCCTCC AATCCCACCT GCGATGGGGC
 9801 AGGCAGTGGA CGGATGGACA GACGAACGGA CAGACAGGCA GGCCGCACCA
 9851 TGCTGCGGAT GAGACGGATG GATGGACAGA CGGACAGACA GGCAGGAGCA
 9901 CCATGCTGCG GATGAGATGG ATGGACGGAC GAACGGAGAG GCAGGCAGGT
 9951 CGAACCATAC TGCGGATGAG ACGGACGGAC GGACAGACAG GCAGGCCACA
10001 CCATGCTGCG GATGAGATGG ATGGATGGAC AGACGGACAG ACAGGCAGGC
10051 CGCACCATGC TGCGGATGAG ATGGACGGAC GGACGGACAG ATGGACAGAC
10101 AGGCAGGAGC ACCATGCTGC GGATGAGATG GATGGACAGA CGAACGGACA
10151 GGCAGGCAGG TCGAACCATG CTACGGATGA GACGGACGGA TGGACGGACA
10201 GGCAGGCAGG CCACACCATG CTGCGGATGA GATGGACAGA CGGACAGAAC
10251 GGACAGATAG GCAGGCCGCA CCATGCTGCG GATGAGATGG ACAGACGGAC
10301 AGACGGACAG ACAGGCAGGC CACACCATGC TGCGGATGAG ATGGACAGAC
10351 GGACAGACGG ACAGATAGGC AGGCCGCACC ATGCTGCGGA TGAGATGGAC
10401 AGACGGACAG ATAGGCAGGC CGCACCATGC TGCGGATGAG ACGGACGGAC
10451 GGACAGGCAG GTCGAACCAT GCTGCAGATT AGACGGACAG ATGGACGGAC
10501 GGACAGACAG GCAGGCCGCA CCATGCTGCG GATGAGACGG ACGGAGACAG
10551 ACGGACAGAC AGCAGGTCGA ACCATGCTGC AGATTAGACG GACAGATGGA
10601 CGGACGGACA GACAGGCAGG CCGCACCATG CTGTGGATGA GATGGATGGA
```

FIGURE 3

```
10651 TGGACAGACG GACAGACAGG CAGGCCGCAC CATGCTGCGG ATGAGAACTT
10701 GGGCTTCTGG AGGGAGGAGA TGGGGCCCGG GGGCATCCCG CACTTCTGGG
10751 ATGTGGGACT TGGGATAAGT CCCTTGTGAC CCTGAGCCTT GGTTTTCTCA
10801 TCTGAAACTG GGCATGGCGC TGGACACAAC CTCGAAGGAC GTGTGTACAA
10851 ATAAGACGAG ACCAGGCGTG TGATGACCTC AGCTGGGAGC CAGCACAAAA
10901 GGAATGCTCA AAAAAAGGGC CGGGTGCACG GTGGCTCAAG CCTGTAATCC
10951 CAGCTCTGTG AGAGGCTGAG GTGGGCGGAT CACCTGAGGT CAGGAGTTCA
11001 AGAACAGCCT GGCCAACATG GGGAAACCCC GTCTCTACTA AAAATACAAA
11051 AATTAGCCGG GTGTGGTGGC GCGCGCCTGT AATCCCAGCT ACTTGGGAGT
11101 CTGAGGCAAG AGAATCACTT GAACCTGGGA GGTGGGGTTT GTAGTGAACT
11151 GAGATCGTCC CACTGCACTC CAGCCTGGGC GACAGAGCAA GACTCTGTCT
11201 CAAAAAAAAA AAAGTGTATT TTTATTTTTA TCCTTTTTAA TTCTAGAATT
11251 TAGCTTGAGG GACAGAAGAG GACCCCATAG GCCAAACCCA CAGCCAGAGG
11301 CACAGGCTGT GGGCTCAGAA GTGGGTCTTG CAGGATGGGA AGGGTCAGGA
11351 GAGTGAGGAT GTGGGCAGAA GGAATGGTTC GTGCAGAGAC GCAGGGAAGG
11401 GGGCCAGGTG ATTCGAGGGG AAAGTGCGTG GGTGACAGAG AGGAGACAGT
11451 CCACTCCCCC TGCCCACCTC ATCCAAGCCC CTGTAGGTCT GTTACTGTGC
11501 ATCTGACCGG TGAATATTCT GAGACTTCTC AGAGCCCACT GAGTGTAGGA
11551 GCTGGGGTTC AGCCTTCCTG TGTCTGGCTC CTGACCGCTC GCTAGGGTTA
11601 GGAAGGATTA GGCCACGGGC TCTGAAGGAG CAAGAGGGGC AGGAGGGCAA
11651 TTGAGGGGCA ATTGAGAGGA ACCCAGAACA TGGAAGCCCT GTGCCGTGGG
11701 GCTGGTCCAG AGCTCACCAG GCTGGACCAC GTGGTTGCTG AGCCATGGCC
11751 CCTGACCGGG GCTGACCTGG CCAGAGTCCC TGTGGCCAGC ACTGATGCAG
11801 GGCTCCTTCC TAGAGGGGCC GGGCCATGAG GAACGGGAGA AACGGCAGAT
11851 GATGCGGGAA CCGGTCTGTT CGGCTTTGGT TTGCAGGATC CGATTTGTTT
11901 TTCATCAGCA GCAGATTTGC TTAAGTATAT GAAAATGTGT TTCTAATTCC
11951 CCGAGCACAC ACCAACTGCT GGCGGGGGAG GGAGCAGTGC ATAGGAGCAG
12001 AGTGAATGCC ACCGGGAGTC AGAGTGCTAG GCCCTGGCTG CTGAGAGAGC
12051 GAGAATACGC CCCCAGCCTC AGTTTCCCCA ACTGAGCAGC CGGGGAAGAT
12101 TTGGCTAGAT TAACCAGTTC ATTCAATGTT CCCTGCTGAT TGCCAGGTAC
12151 ATTCTGGGAG TTTAGGGAAA TCCAGATTGG TCAGAGACAA AACCACACAA
12201 AACAGTGGAC TCCAGTGCAG ACAGAGGGGT CCTAGATGTA TACCCCGGGC
12251 TCAGCATAGC AATAATCATT TTAAAAAAGA TTTTAAAACA TTTTTAAAAC
12301 TCAGGTGAAG TTCACATAAC ATAAAATTAA CCAGTTAAAC AACGTTTTGG
12351 TGGGTGCAGT GGCTCACACC TATAACCCCA GCACTTTGGG AGGCCGAGGC
12401 AGGAGGATCA CTTGAGGCCA AGAGTTTGAG ACCAGCCTGG GCAATGTAGT
12451 GAGACCCCAT CTCTCAAAAA AAAAAAAGTC TTATTGTGTC TAACATAAAA
12501 CTTGCCTTTT AAACTATTTT ACAATATACA ATTCAGTACA TTCACAATGT
12551 TGTGCAACCA GCATCTCTAC TTAGTCCCAA CACGTTTCCA TCGCCCCAAT
12601 AGAAAACCCT GCACCCGTTA GTTACTCCCC ATCTCCTTCC CCTGCCCCTG
12651 AAAACCACGC GTCTACTTTT TGTCTCCATG AATTTAGCTA TGCTAGACAT
12701 TTCATACGAA TGGAATCAGA CAATATGAGG CTCTTTGTGA TGGCCTTCCT
12751 TCACTGGCAA AATGTTTCCA AGGTTTGTCC ACATTGTCGC ATGACTCAGT
12801 GCTTCATTCC TGTTTATGGC TGCATAATAT GCCATCCTGT GGACACACCA
12851 TATTTTGTGT ATCCGTTTCC TAACTGATGG ACATTTGAGC TGCTTCTGCT
12901 TTCTGGCTAT TAGGAGTGAT ACTGCTGTGG ACATTTGGGT CTCAGTTTTT
12951 GCATGTGTGT ATGTCTTCAT TTCTCTTGGC TGTCTACCTA AAAGTGGAGT
13001 TTCTGGGTCA CAAGGTAATT CTATGTGTAA CTTTTTGGGG AGCCACCAAA
13051 CTGTTTTCTA CAGGTGCTGC ACCTCTTACG TTCCCACCAG CAATGTACGA
13101 GAATGCCAGT TTCTCCGAAT CCTTGTCAAC ACTTGTTATT TTCTGGTTTT
13151 GTTTTGTCTT ATTAGGATGA GCCTAGTGGG TGTGGGGCAG TATCCCATTA
13201 TGGTCTTGAT TTGCATTTCC CCGATAGCTA ATGATGTCAG TGTGCTTCTT
13251 AGTCATTTTT TTGTTTTTGT TTTTGTTGTG TGTTGTTTTG AGACAGAGTC
13301 TCATTCTGTC ACCTGGGCTG GAGTGCAGTG TTGCGATCTT GGCTCACTGC
13351 AACCTTCACC TCCTGGGTTC AAATCATTCC TGCCTCAGCT CCCAAGTAGC
13401 TGGGATTACA GGTACACACC ACCACACCCA CCTAATTTTT GTGTTTTTAG
13451 TAGAGACAGG GTTTCACCAT GTTGCCCAGG CTGGTCTCGA ACTCCTGGCC
13501 TCAAGCGATC TGCCCGCCTG GGCCTCCCGA AGTGCTGGGA CCACAGGCGT
13551 GAGCCACCAC GCCCAGCCTA TTTTAAATTT AATGAACTCC AATGTGTGTA
13601 TTTTTTTCTT TTGTTGCTTG TGCTTTTGGT GTCATATCTA AGAAACCACT
13651 GCTAAATCCA AGGTCAGCAG TATTTACCCC CATATTTTCT TCTAAGACTT
13701 TTATAGTTTT AGCTCTTATA TCTAGGTCTT TGATCCATTT TGAGTTAATT
13751 TTTGTATCTG GTGTAAGGGA AAAGGTCTAT CTTTATTCTT TTGCATGTGG
13801 AGATCCAGTT TCCCCAACAC TATTTGTTGA AGAGCCTATT CTTCCCCCAC
13851 TAAATGTTCT TGGCAACCTT GTCGAAAATC AATTGAGCAT AATCTATGCA
13901 CTTACTTCTG GACTCTCAAA TCTCTGGGTT TTTTTGTTTG TTTGTTTGTT
13951 TGTTTTGGAG TCAGAGTCTT GCTCTGTCAC CCAGGCTGGA GTGCAGTGGT
14001 GCGATCTCAG CTCACTGCAA CCTCCCCCTC CCGGGTTCAA GCGATTCTCC
14051 TGCCTCAGCC TCCCAAGCAG CTGGGATTAC AGGCACTAGC CAGCACGCCC
14101 AGCTAATTTT TGTATTTTTA GTAGAGATGG GGTTTCACTA TTTGGCCAGG
14151 CTGATCTCGA ACTCCTGGAC TCAAGTGATC TGCCCACCTC GGCCTCCCAA
```

FIGURE 3

```
14201 AGTGCTGGGA TTACAGGCGT GAGCCACAGC GCCTGGCCTC AAATCTATTC
14251 CTCTGAAGCA TCAAGCATTC TATGTGCACT ACTTCATGAA ACCCTCCTGG
14301 ATATTCTGCA CTGTAGAAAC GATTACTCTC CTGTTGTGCC CATTTTATAG
14351 ATGAGGAAAC TGAGACTCCA AAACTGAGTG AAGTCAAGGC TCAAACTCAG
14401 ATCCCAGTCA TTTGATGACT AGGCCACAGT GAGGCCTGAG GAGGGGAAAA
14451 ATCCCAATGG TTACCCTCCC CTTCCCCTCC CCACCCTCAT TTTCTTCTCC
14501 CTCTTTCAGG CTGGGATGTG GACTTGGATT CTCAGAGCAG GGTCCTTGGA
14551 AGGAGATGCT GTGACTTCTC TCTGGCCTCC AAATACCTCC TCAGCCTCCA
14601 GTCCACCTCC GTCCCTCTCC CACGCAGCCA GGCACTGTTC TGTCCTCTTC
14651 CTTGTCCCAC AGTCAGTGCT TGCATGTAGC AGGTACTTAA TAAATGCTGA
14701 AGATAATTAT CCATCATTTC AAATAGAGAC ACACAACTTA GAAGGCATGC
14751 TGGGATTGTC TAAGGCCAGA AAAACCCCAA TGTCGATAAG CATGTTACAG
14801 TGAAATTGAC TGCGCCCAGG AAAGGGGACC CCAGAAGCAG GTGGCTGGTG
14851 TCCCCCTACC CTGCCCCAGG CCCCGAGTTC CCCAATCCAC CACTAGGAAG
14901 TCCTGGGCTC CTGTGAAGAC AATATAAAAC CACTGATTAG GCCAAGTGTG
14951 GTGGCTCACA CCTGTAAATC CTAGCACTCT GGGAGGCTGA GGCGGGCGGA
15001 TTGTCTGAGC TCAGGAGTTT GAAACCAGCC AGGGTGACAT GGTGAAACCC
15051 CATCTCTACT AAAAATACAA AAAAAAAAAA AAAAAAAATT AGCTGGGTGT
15101 GGTGGTGCAC ACCTGTAGTC CCAGCTACTC GGGAGACTGA GGCAGAAGAA
15151 TTTCTTGAAC CTGGGAGGCG GAGGCAGAGG TTGCAGTGAG CCGAGACTGT
15201 GCCACTGCAC TCCAGCCTGC ACAACAGAGT GAGACTCGGT CTCAAAAAAA
15251 AAAAAAAAAA AAAAACTACT GATGAGGCAC ATCCCCCCCT CTCATTTCCT
15301 ATGAAGGAGA AACTGAGGCC CAGAGGGTTG GAGTGACTTC CTTGAGCCCC
15351 CCATGAGGAG CTTCAGACCC TGGAGGCTCC ACCCCAGGCC AAGGGCTCTC
15401 CCAGAGGTAG ACTGGAGCCA TGAGGACAGG GGCCCTCCCC AACCAGGTCT
15451 CTGTCCATCT ACACGTGCCC TGGATCTGAC TTCACGTGAT GGCATCTGGT
15501 GGGGGACACA GGATGCCTGC CCGGATGCCA CCTGCAGCCA GTGGGGGCCG
15551 GAGCTGCCTC TTCAGGGTCA GTGAGGGTGA TACATCTACT TCCCAGCCTG
15601 CTTAGGTGAG CTCCCGCCTA TGTGTCACTA CTGGTGACTG GCATGGCTCA
15651 GAGCCAGATC TTGGGGGCCC TGAGGGGATC AAGAGCGTCC CCTAAGCCCA
15701 CCTGCCAGCT GCGGTCTTCT CTGTGGTGGC AGCATCACAG AAAGTGGACA
15751 GAAAGAGTGC TCTGTGCCAG GAGGGCAAGG CCGGGTAGGA TGGTGGCTGG
15801 AATGCTGGCG ATCCCAGCAA TGCCGGCGAT CATGGTGCTG GGTTTTGGTG
15851 GTGTGCTGGA CGCCTGGGAG CCTCATGAGT GAGAGACTGG GGCACACGTG
15901 CTTCCGTAGT GCCATGCACC GGTGGCAATT CAGAGAAAGA CGCTGTGCAA
15951 AGCACCCCAT GTGTGCAGCT TTTTGCCCTC TCGTAACAGG ACGGAGCCAG
16001 GTCAGAGTGC AGATGAGGAG AGGAAGGTGC AGGGAGGTGG AGATGCTGAC
16051 CCAAGTTTGC ACAGCCAAAA CTAGGATCGG TCTCCAGGGC CTCCGTCACT
16101 GTCCTGTCCT GCCTTCTGTC ACACAGGAGT TCGAATGGTC GTTCTGAAAT
16151 TGAGAGCTAG CGGGGCTGGG ATCTCACTGG GCGGCCACAG AGGGGTCCCC
16201 TGACCTCTTG GGGTCTCGTT GGCAGGAGGG AATTGTATTG GAATATCCAG
16251 GTGTGTGGAT CCCTGTGAAT CTAACCCTGG AGTGTTCCAG AACTGCCCAC
16301 CCTGTGGAAA GGGACTCAGG CCTGTCTTCA AGGACCTGGC ATCCTTCTGT
16351 CCCAGGGCAG TTTGTCTTGG GTCTCTCAGG GACCGTTTGG GCCTCTTCAG
16401 CCCCTCATTC CACTTCCCTC CTGCTGCCCA AGTCATTCGT CCACTTGACT
16451 CCAAGAGTCG GCTGGGGAAA TAAAAGGAAA TGAAACACGA CCAGGCATTT
16501 TCCCTTGGCC GAAGCAGAAG TCTGCTGTCG GGCAAAAGGT GAAGAAGAGA
16551 CCAATGAGAG ATGAGCCCAC GGTGCTCCTG CCCTCCGCCA AGGCAGGCCA
16601 TCCTCTGCTG CCAGCCTGCA ACAGGGCAGT GTCCTTCTGG GAGGTGTCCC
16651 TCCCTCTGGG GGATCAAGAG ATGGCCAAAA GCAGGTGGCA GCAAGTGGAG
16701 AAGGCTGTTC ATCCAGAACG CACCTTGTCT CTGCCCCTGT CCCCACCCAG
16751 GCAACATCCA AAACCTTTGC CCACAGTTCC GGGGCTGGCA CCGTCCTGGG
16801 GCTCAGCTCC TAGGGACGGG GCTCCCCCAG GCACTGGCTG CCAGGAACTG
16851 GGTGGCCCCG GGCAAGTCTC TTCCCATTTC GGGGTATAGA CTTCCTGCCT
16901 GTAAAATGAG GGGGTCTGCA GGTCAACCTC AGAGTCCCAC TGTACCCCCA
16951 GATTCTGCTT CAGGGAGACG GAGAGAGAGA GAAAGAGAAA GAACGATAGA
17001 GAGATGCAAT AACCTCCCAG CATCCAGGAA GACCCAGAGG GGAGAAATGC
17051 AGGGAACCTA CCCAGAAAAC CCTGGAGCGG GAGCTTCTCA CTTTTAATGG
17101 TCATGGCCCC ACTTGAGAAT CCATGGTGCT CTTCCTAGAA CCACGCATGT
17151 GCACACGTGT GTGCAAACAC TGGGCTCATG CACAGGCACA CACACACACA
17201 TATAAGGTTG CAAACACTTT CAGGGACTTC CCAGATTTCT CTGAGTCCAT
17251 CCGTGGTCAC TTTCGGTCAA TCATCTGGCC TGAGCAGGTT CTGCCTTCTG
17301 GGGGCTCTTC TACCCTCAGG GAAATCAGGG TTTGGTTCCC TGTAATTGTC
17351 TGGTCCAATT GTCTGAGGAC TTTCTCTTTT TTTTGAGACA GGGTCTCACT
17401 CTGTCACCCA GGCTGGAGTG CAGTGAAGCA GTCTTGGCTC ACTGCAGCCT
17451 CGACCTCCTG GGCTCCAGTG ATCCTCCCAC CTCAGCCTCC TGAGCACCTG
17501 GGATCACAGG CATGCACCAC CATGCCTGGC TAATTTTTGT ATTTTTGTAG
17551 AGATGGGGTT TTGCCATGTT GCCCAGGCTG GTCTTAAACT CCTGGGCTCA
17601 CTTTCTTTTT TTTTTTTTTT TGAGACGGAG TCTTGCTCTG TCACCCAGGC
17651 TGGAGGGCAG TGGCATGATC TTGGCTCACT GCAACCTCCA CCTCCCAAGT
17701 TCAAGCAATC CTCCTTCCTC AGCCTCCCAA GTAGCTGGGA TTACAGGCAC
```

FIGURE 3

```
17751 CTGCCACCAT GCCTGGCTAA TTTTTGTATT TTTAGTAGAG ACAAAGTTTC
17801 ACCATGTTGG CCAGGCTGAT CTCCTGACCT TAAGTGATCC GCCCACCTCA
17851 GCCTCCCAAA GTGTCGGGAT TACAGGCGTG AGCCACTGTG CCTGGCCATA
17901 GCCAGACTTT CTTGATTCTA TATCCTTCTC CTCAGAGCAG AAACATCGAG
17951 CATTTGTTGA GTGCCTCATG TATACCAAGC CCTTAACCTA AGCTATAGCT
18001 CATTGAACTC TCACAGAAGT CTTAAGGTAG AGCTTGTATT TAGATCCGTT
18051 TTGAATATGA GGAATCTCAG GTTCAGAGAA TTTAAGCCAC TTGCCGAAGG
18101 CCACACAGCT TCTAAGTAGA GGAGGCTGGC ACCTCCAGCC TGGGCCGCCC
18151 GGCCCGGCAT CCAGGTTCCT AAACAGGCTG CTCAGCTGAC ACGAGTCGCT
18201 CTGGATCTCA GAAAGCGCCT GATAGACGGA GGCGGCTGTC ATCTGTGTGT
18251 GTGTGTGCGT GCGTGCAGCC ATGTGTGCTG GTGAGCATAC TGTCCCTGTC
18301 AGCCTCTCCT CCCCCAGCAC ACCCCGGCAG CCCAGAGAAG GGAGGGCCCG
18351 GAGGAGTGAC GGTGTTCCCC ACCCCCTGCC CTTTGAGACA CAATGGAGTC
18401 CGCTAATCCA GTTACTTGAT AATTCACTTA TTTCATGTCT ATTTGGCAGC
18451 GAGCGTGCTC CCACGCACCA GCTCTCGGGA AGGCGAGATG GCTTTGCCTG
18501 GAGGAACCTG ATTGTTTTTC TGGGGAAGGA GTGGGGGAAA AAATTGCACC
18551 CAACAATGGA CAATAATGGG CCTAAAAATA GAGGGTGGAG GGTGCAGGGG
18601 GTGGAGGAGT GTGCTGTCTG CCAAGGGAGG GCTCCAGGCC TGTCTGCTTG
18651 GCACGGGGCA GCCCTACCCT CCTGCCCAGT TCCCCTCCCC TGCACTGGGT
18701 TGGCCGCCTC TCAGCCTAGA GGAGGGGCAC TGGAAGGAGG AGGCCCAAGT
18751 GGGTGGGGGG CTGGGTGGCC TTCCTTGCTG TCTCTGCCCG CTCCAAAATG
18801 GAAAATTGTC CCACCCAGGG GGTCCTGGAG GCAGCAGCCA CGACCTTGGG
18851 TGGACGCTGC GCCTCATCAG CCCTGACTAG CCGTGATGCC CAGGACCTTC
18901 CCCCAAGGGG CTCAGGCATC AGCTGAGAAC TGCAGCCTTG GGTACAGAGT
18951 ACGGGTTGTC TCCCAGCAGG AAAGGCAGGT TTCAGGTTTT GTGGCTCTTT
19001 TCCATCTCCC AACACTTGGG GCAGTCTTCT CGAAGGCCTC AAGCCCAGCG
19051 GGCAGCTATG ACCCCACCAG GAGCGGAGCG GGCAGGGACC AGGCTGCCCT
19101 CTAAGCCACT CGGCTGGCTC TCAGCCGGGG TGCACACTGG ACTTGCCTGG
19151 GAGCTTTTCA TTCCCCCCCT GCCGCAGCTG CCCCCCAGAC CAGCTTCAGC
19201 AGCCTCTCTG GGCGGCCCAG CAGGAGGAGG CATTAAACCT CCCCAGGTGG
19251 TCCCAGTGCA CAGCCAAGTT TGAGAAGCAC CGATTGAAAC CTCTCCCAGG
19301 CCTGCCCTGG AGCCCTTCCA GCCTGAAGCA TCTTGTCGTC TTAAAACTGA
19351 AAGACCAGGA GGAAGAGAAT TCCATGGCCT GCCTCGGCTC TCTCGGAGCC
19401 TCTCTCACAT CTGAGCTGCA GGTGCTCCAT CCTCCTTCTG GCTTCCTGGG
19451 TGCCGAGGGG TGCCAGCTCT CCAGGCTTGG GAGAGGGCCA CTTAAGCCCT
19501 CACACTTTGT TCCCAGGCTC TTCACCTGTC TTCCTGGAAG GAGGGGCCG
19551 GCCAGCATTA GGGCTGTCAC GGGCGCTGCT TAATGTCAAG CTGCCCATCT
19601 GGCTCCTGGC CTCCCTTTGG CCTTCTCTCC TGCGCTCCCC ACCAAGCTCC
19651 TGGCTCAGCA GCGTGCATGC GTTAACCCAT TGCCCCCCTG CAGTGTTTTG
19701 TGTGTCCAGC CTGGCCCTTT GCTCAGTCGA CCTAGAGCAC CATCCTCCCA
19751 GACTAGTCGA GTGTCCCTCC ACCTGTCCTG AGTCCAGATG AAATCCCACC
19801 TCCCCCAGGA AGCCTTCTGA CTGCCCCAGC CCGTCACCTC CAGGGCTTGT
19851 CATCTGTGCC ACTCATGGGG ACCAGGACAC AGGTGACTTC TCTGGTGGAC
19901 ACAGCAGAAC GGTCAACATT CCCAAAAGGG AGCAAATTGC CCGAGTCACC
19951 AGAAGTGTGA CCTTGAGCAA GCATCTGGCT CAGGGGCTCT TGGCTTCCCC
20001 ACCTGTAGGT AAAATAACAG GAACAGTGTC ATCGTGTGGG GGCCCTTCCC
20051 TGGACCACCT GGACCAGCCT CTCAAACCTG GCCACACATC TGAGTCACCT
20101 GCTGAGGGTT TTGGTTAGTT GGTTGGTTAA TTACTCGATT GGTTAGTTTG
20151 TTGGTTTGTT TGTTAGCTTT CAATTTGGAG TAGAATTTCT GACCTTAGGG
20201 CCCAGGAATG TTGCCACACC CCTCCCGTCC CCAGGTATTG TCATGTTTTC
20251 ATGGGTGGCC AAATCTAAGA GCTGCTTCTC TGGGGCACGA AGGATGTTCA
20301 CAAATAGTTG ATGAATAAGT GAATGAATAA ATCAATGAAA CTTACCAGCC
20351 CAGCCTCACT ACTCGCACCC ACCCCCAACG ACCAGCCAGG GTTCATCCAC
20401 AGAGGGGTGT ACCTGTCCAG GTGTCCCCAG GTGTGGGCAG ACCCAGTAAC
20451 TTTACTCTTT CATCGGCCCC ACCGCCTCTT AACTCCTCAG AGACCAGCAG
20501 GAAGAAACCC TCGGAGGTCG CAGCTTCTGG CTGTTCTCAG GGGCAGGCCC
20551 CGTCCATCGG GTGCTGTGTC TACTCCTAAG ACCTGGTTCT GAGTATGGAA
20601 CACCTGGAGA GGGAAGGGGC CGAGGAGGGG GAGTCACTCG GCTGTGTCAG
20651 GCTCCGCCCC TGCCTTCCTG AAGCACACAG TGGGGAGGGG ACACACCCGT
20701 CATTAACCGA AGAAGCCACT GGGGAAAACT GTGACCCAGT GCTCCCTTGG
20751 GACTGGGGGG CAGTGGCCAG GGGTGTTTTC CCTGAGGAAA AGAAATTTAA
20801 GCAGACACCT GCCAAAGGCT GGAGGGAGAG CTGTAGACAG AAGATGGCTC
20851 AACCTGAAAG CTCCGCGGGG TGGAGGGGGC ATCCAAAGGG CGGGAGAGAC
20901 TGGCCAGTAG AAAACGAGGC CAGAAGCCGG ACATGGTGGC TCACGCCTGT
20951 AATCCCAGCA CTTTGGGAGG CCGAGGCGGG TGGATCACCT GAGGTCAGGA
21001 GTTCGAGACC AGCCTGACCA ACATGGGGAA ACCCAGTCTC TACTAAAAAT
21051 ACAAAATTAG CCAGGTGTGG AGGCACATCC GGTAATCCCA GCTACTTGGG
21101 AGGCTGAGGT AGGAGAATCA CTTGAACCCG GGAGGCCGAG GTTGCAGTGA
21151 GCCAAGATCG TACCAGTGCA CTCCAGCCTG GACAACAAGA GCGAAACTCT
21201 GTCTCAAAGA AAACAAGGCT GGAAAGACAA GGGAGAGAGG CAGAGCTGGT
21251 GGCAGAGCCA GGCCCAGGGT TTGGACTTAA GAAGGGGAAG GCGCTGGGGC
```

FIGURE 3

```
21301 TCTGAGGAGG AGCTGGAGAC AAAGGGAAGG GTGTCTGGCT GCAGGCTGAG
21351 AATGGGCTGG GGGTGGGGAG GCTAGAGGGG TCGGTGGAGG CCCCAACCAA
21401 GGCGGGGCAG GGGAGAACAG AGTGAGCCTT TAGGAGAAAC CACTGCTGGG
21451 GACCAGGCTT CACTCTCAGC CCATCCGGAA ACCTCTTTCA GCTGATGCTT
21501 CCCCCGACCC CCTTCCCCAC CCTGGGCCTC TGTTGGAAGC TGGCTGAGGC
21551 CCTGTTGATC AGCAAGAAAG CCCAGGGCAG CTCTCAGAGA AGAGGAGAGG
21601 GGGCCCAGAA AAGGCCCCCA GGATCTGGGG AGGGGATCCG AGGAGAGGCA
21651 GCTACCAAGC GCCCCAGCCA GGGGGGCCTG TCCCTCACCC CACCCCGGCA
21701 CTGAAACCCT CACAGCCACT TTTTTCCTCC CTCCGTGATA AAATATTCAT
21751 GGCGGCAGAG TGGGCCCTCT TTGGGAAGGC TGCCTGGGTC TAGCTTATGC
21801 TCTGCACAAG CTTTTAAAGA GCAGGGCGCT GTTCCTACTC TCTAAGCATT
21851 TTCTAAGTCC TGAATCAATA ATGCACTTTC CTGGGCTTCT CCGGATGTAG
21901 CCCTCTTCCT CTCGGTGCCT TTCCCCCGCC CGCCCCCTTA TCTTTCTTTC
21951 CCTTTTCTTC TCTCCCGTTC TTTCATCCTT CCATCTCTCC TTTCATTTAC
22001 TTTTTAAAAA AGTATGAAAG TGTTGAGCTG TTTGGGTGGC CAGTGAAGCC
22051 CTGAGTAGGG AGTGGGCAGG AAGGGAGGCG CCAGACTGAG CCCCTGTGTG
22101 TGCAGGGAGG AGGAGGAGGA GTGGGAGGAG GAGGAGGAGT GGGAGGAGGA
22151 GGAGGAGTGG GAGGAGGAGG AGGAGGGAAA GGAGGAGGAG CAGCAGCGGC
22201 TGAGCGCTCC CGCTGGCCCT GCTAGGGAAG TGTTTGAGGA TCACTGAGCT
22251 CCTGGTGTGG GGAAGGAGGA GGGCTTAGCC TCACCCGGCC TCCCTCTCTC
22301 CTTTTTCTAA TCAATTAGAA AGTGTTTACA GCATAGCCAG AGAAAAATAG
22351 GGAACCTGGG ACCAAGAAAA AATGCAAAGC ACCGGCCAAA TTTCAGCCCC
22401 ACACTCGAAG GAGGGAGCAG TGGGGTTTCA CCTATCTGCC TTCTGTGGTA
22451 ATGAAACCCC TGTCGCTAGA GGTATGCAAG AGAAGGGGAG CCTTACCCTG
22501 TCTGAGACAG ACGCCCATGT GTCTGGTCCA TTCTGTCAGT CGCCTGTGGG
22551 GTGCCCCAGG GATGCACGGG CACTCTTCAA CTGGGATAGA ATTTCCTGCC
22601 CCAAACATTC CTGGAAATCT GGCTGTGGGA AGAATCCACA TATGCCCAGG
22651 GCAAAGCAGA ATGTGTCCTT TAAGAAAACA ATAATACATT TTTAAGTTCC
22701 TGGAGAGATT AACCCTTGTC TAGCCAGAGC CATGGCAATG CCTCCCCGCC
22751 CACCACACTC TGGTGGTTCG GCTGACGGAG GAGATCAGTC ATTCAGGGGT
22801 CTGCGGTCCT GATGAGCAGT GGGTGCCCAC ACCAGGCCTG GCATTTCATC
22851 CTTGCTTTCT GACCTTGGCT TCCCAGTTGA CCCTCTCCCG GGCAGCTCGT
22901 CCATCAGGGC AGCCCAATGC CCTCAGGTCC TCCGAAAGGA TCTCAGGGTG
22951 TTCTGTGGGG GCAACCCGAA TTGGTGTAAG AAGACTAAGC AGTCGATCTG
23001 CTGGAACAGC ATCCCCAAAG CGGAGCGAAG CCCGCGGATG CCCACCGCCT
23051 CTCCCCCAGG CAGCGTCCTA CCTGGATAGA ACTGCCTGGA GCCACTGCAG
23101 AGGGTCCTCG CTCAGTTAGG GAATGTTTGT CATATACCGC TGTGTGCAAA
23151 CAGCTGTTGG GAGTGTGGCG CAAAGGTGGG TAAGGCCCCT GCTCTCCCAG
23201 AGTTCACACT CACAGAAGGT TCTGGAAGGA GGAACACTGT GGGCAGGGTT
23251 GAAAGGCCTA AAGTGCTCCC TTTCCTCCCA AATAATGCGG GGTGAGGGGC
23301 GGTGAGGAGA GCCGCTCTGA GCAACCAAGG AACTGAGATG CATTTTCTGG
23351 GTCTCCTTTT GAGCCGAGGC AGGTTCGAGA GGCAGCCAGA GACTCTGGGT
23401 TCAAGGTGGA CCTGTGCCCA GGCCATGCCC ACTGTGGCCC CCCTGGGGGA
23451 GGAGCAGGGG CGGTCGCCGT GGCTTTGGGA GGCTCATTGC TGGGACAAGC
23501 GAGTCCCTGG GGAGGCAGCG CTTGGAGGCT CGCTTGCCTG CCCCTGCCTG
23551 AGTAATTGCT TGGAGCTGGG AGGAAAATTG CTCCAACCAG AAAACAAAAC
23601 AGAAAAGCCG CCTTGGCCAG CTGCAGCTCC AGCCCTAAAA TGCCAGGTTG
23651 GTTTACGCTG ATTCACGAGC GGGGAGGGTG ACCTTGCTGT CTGTTGTCCA
23701 GGGCCTGTGC ACGAAGAGAA TCTGGAAAGG GAAGGAGAGA GACACCTGCA
23751 CGCTGGGGAA GGAATTAGCA GCACAGAGAG CAAGAGGGAC AGCGATCAAT
23801 GAAACCATAG AAGGAGAATG AGAAACACAC ACACAGAGAG CGAGAGGGAG
23851 CAAGAGAGAG AGAGAGAGAG AGAGAGGGAC AGAAAACGAG AGGGAGGGAG
23901 GGAGGGAGAG CTCAGAGAGT TAGAGACCGT CAGGGCCGCT AGAATTAGAA
23951 TCAGCTCTGA ACAGAATCTC CGTTTCCGCT TTGTTAATAA TTTATTCCCT
24001 CTGCAACTTT TCTTACCAAT AAATAGGAAG TAATCTGTTA AGGAGAATTC
24051 CCCTAGCACC CCGGCTTTCT CCCTGGAGTC AGGGGAGGAG GATGTGTCTC
24101 TGTGCCCTTC CTCCCTAGCA GCATGGGGGC CTGAGGAACA CGCAGAACTT
24151 CAGACTTTAG GATGTCAGGG TCAGAGGCGG ACAGCCCACT CCTGCCCGGT
24201 CATTTTGTGA ACGGGGAAAC CAAGGCACAG ATAGGGCAAG GCCCTGGCCA
24251 AGGTCACACA TGGTGTTAGG GGCAGTCCCC TGAGTCCTAA TTCCATGGCC
24301 CCACGGGTCA GGGCACCTAT TGATTTATGC ACCTGCCCAA GCCATAGGGT
24351 TTCCCCCGAA ATGGCAGAGG CCACATCCAA GGAGGAGGGT GGGGCTAGCT
24401 CGGCTGCCTT TCCTTGCCTT CCCCCACGAT TGCTTCCCCC GTGCTCGAGT
24451 CCTGGCCCTC TACCTGGGCA CCCACACCCA GGGCCTCTCC TGGGCAGCCT
24501 CCAGCCTTCC ACCTTGTATG CGGCAGCAGC CTCCCGTCCT GGTGAGGCTG
24551 AGGGGCTGAG GATGAGAAGG GTTCCGTTGG CAAATCAGCA ACAGCAGTCA
24601 AGAGACGTGC CGCCTGCCTC CCCGTGGAAC CCGAGTCTGC GGGAGCACAG
24651 TGCGGCCCAG GCAACAGCGT CCTTTCCCTT TTGGGTGAAG GGCACCATTT
24701 CCCAATTTGT CTCAGGGCCC AGCTCAGTGG GCCATCCCCT GGCTTCTTAT
24751 CCCACCTCAG CTGCTGCCGA GCCGCATGAC CCTGCGACAT TGCTCAGCCT
24801 CTCTGAGTCT CGGTTTCCTG AGGATCGCAC TCTCCAGGAT CCCTGGGAGC
```

FIGURE 3

```
24851 GTGGGAGGTG GGGTTGGGGC ACACAGGGCG CCCAGCACAG GGCCGAGGTG
24901 GAAGACATGC TCCCTAACGG CGGGGCCTGC TGTTTGCTGA AGCACCAGGC
24951 CAGACAGTGG CCATGAATGT GCTCCCAGCA TCCATCACCC ATGAGCTGGC
25001 ACCACCGAGG CACTTGCCAT GGTGCACCTG GCATCATTCC TATGACAACC
25051 CTGTGAAGCC AGTGCTAGTA ACCTCATTGA GCGTTCATTC ATTCTCCGAA
25101 GATTTCCCGA GTCCCTGAGG AGGGCCGGGG GCTGGGGCTG GAGTGGGGAC
25151 AGGATCAGAT GTGGTCGCTG CCCGCATGAA GCCTCCCCTC CAACAGAGAA
25201 GCTGAGGCTC TCGGGCAGGA GAAAGATCTT TTCCTCACCC ATTCTATGTT
25251 AGTGGCTGAG GGCCCATCAT AACAGACAAA TTAATACCAC CAAAGCATAC
25301 CAGTGGATTT AATATAAGTT TTATGTGAAA AAGGCTTTAA GCCTTTCTTT
25351 TTTTTTTTTT TTTTTTTTTT TTTTAGACAG GGTCTCACTC TGTCACTCAG
25401 GCTGGAGTGC AGTGGCACAG TTACGGCTCA CTACAGCCTC GACCTCCTGG
25451 GCCCAAGGGA TCTTCCTATC TCAGCCACCC AAGTAGCTGG GACCACTGGT
25501 GTGTGCCACC ATGCCCGGCT AGTTTTCTTT TTTGTTTTTT GAGGTTTTTT
25551 TCTGTAGAGA TGGCATCTCC CTGTGTTGCC TGGCCTCATG GGAGCTTTCA
25601 TAAGGAATGA AGACCCAAAA CATTGGTGAA CATCTATTTT GTATGCTAGG
25651 TTTAATGGAG AAATAGTCAT GGAGAAGTAC GATTGGCTTA AAAAAAAGTA
25701 TCATCTCCTG GTGATAAACT GGCGGGAATT TTGCAAGACC TGTGTGTCCA
25751 GGTCCCTCTC TGTGACCCTG CATCTTTGGA GATGAGAATG TTCCTTCCTC
25801 CGGGCATTGG GAGGGCACCT CTCGAATGAG CCTCATGTCC TGCTTCAGGG
25851 AAGAAGGGCA GGGGAAGGTC AAAGAGTAAC CTTCCGCTTC TGTGGTTTTC
25901 TCAAATCCCT TCAGCTTAAA AAAAATTATT TTTTTGAGAC GGAGTCTCAC
25951 TCTGTCACCC AGGCTGGAGT GCAGTGGTGT GATCTCAGCT CACTGCAACC
26001 TCTGCCTCCC AGGTTCAAAT GNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
26051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
26101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
26151 NGGGAGGCCG AGGCGGGCGG CTCATGAGGT CAGGAGATCG AGACCATCCT
26201 GGCTAACACG GTGAAACCCC GTCTCTACTA AAAATACAAA AACTTAGCTG
26251 GGCGTGGTGG CGGGCACCTG TAATCCCAGC TACTTGGGAG GCTGAGGCAG
26301 GAGAATGGTG TGAACCCGGG AGGCAGAGCT TGCAGTGACC CGAGATAGCG
26351 CCACTGCACT CCAGACTGGG CGAAAGAGAG AGACTCCATC TCAAAAAAAA
26401 AAAAAAAAAA AAAAAAAAAA AGAAGGTTTT AGGGTAGCAT GTCCTGGGCT
26451 TCATTGGCCT CAAGGAAATT GGGGGACTCG CCCAAGCCTG TGGAGCTGGG
26501 GGAGCTGCTG AGGAAGTGAG CCAGTGCATC CTTCCACAAA CCACAGAGCC
26551 CTGCCCGGGA CACCCTGGCG ATTTCATCCC AACTCTGAGA TTTCATCCCA
26601 GCGCAGGCTT GGGTGCAAGG CCAGCTGCAT GACGTTGTCT GTTCTTCTTC
26651 CTGAGACTTG GTGACCATTC CAGTGACCGC CCCTCCACGG CCTCCTTAGT
26701 CAGGGGCCTG GAAGTTCAAA TGGCTGGGCT TCCCACAGGC ACATTTACCT
26751 CCATTCTTGC TAACAGTTCA CTTTCCCATT CAATTATGTT TTTTGCCTGC
26801 AGCTTGCCTA TGATGTTTAC AACCTGGCCT CTGACTTTGA CTGTACCCTT
26851 TGCACAGAAA ATAACATAAA AGGAAAAGCA TTTAAGTGCC CATGACGAGG
26901 CTTGAAAAGC ATGGCATGAA ATGGTTTCTC CATAGAATAT TTCATGCCAG
26951 GAGCTCAGGC TTGGCATCTG TGTAGGAGGC TCCTCCTCAC CCGGTTCTCT
27001 GGTGCTATGG GCCTGGCTGG CTGCCTTCCC TTCCCTCTGC CCACCCTCCT
27051 TTGACTTCTA GCCACACTGC CTGACCTTTC AGGTGACTTG CCTATGTGTG
27101 TCCAGAGAGA TAGGAAAGCT GTAGACATGA TGGGCTTGGT TTCCCCAAGA
27151 TTCCTCAAGG TTGGGTCCCG TGGAGTCCAG GGGATGGGTA AGTGATGTGC
27201 GGCCATGGGT ACTTAGTGTC TTATCTGAGA CGTGGAGCTG ACTGTAGCAC
27251 CTGCCTCCTG TGGGTTGCTG AGAGGGTGCA GGGAGGAGCG GCTGTGAAGT
27301 GCCCAACCCA GCCCGGGCAC AGGGCTGGCA CTCAGTGAAT GTTACAATCA
27351 TCACACTCTT CTGAGTCAGC CGTTCCCGGG ACAGTCCACG CCATGAAGTT
27401 CCAGGGTTTC TCATGCACAA GCCTGGTGGT CTCAGCCTCA TCCCTTCCTC
27451 CTGTGGAAGG TTGCTGGGAG TGGAGTGTCC CTGAGTTAAT ACGAAGCTGC
27501 TGTTTCAAAA TAATCGCTCC CGTTTGGAGA CACTTCCTAG GGATTCTAGG
27551 TAATGTGAGA CACAGACCAT CTCACATGGC AATCAATAAG AAAATAGAGA
27601 CTCAGAGAAG TCAAGTGACT TTGTCAAGGT CACAGAGCCT CAAAGGAGAG
27651 AGCTGAGAGG GAACTCAGGG CTTTCAGACG CCAGGGCCCA AAACTGTATG
27701 GAAAATGGGT ATGTTAACTG CATCTTTCTT TTTTTTTTTC TTTTAGACAG
27751 AGTCTTGCTC TGTCGCCCAG GCTGGAGTGC TGGAGTGCAG TGGCGCGATC
27801 TCGGCTCACT GCAAGCTCCA CCTCCTGGGT TCACGCCATT CTCCTGCCTC
27851 AGCTTCCTGA GTAGCTGGGA CTACAGGCGC CCGCCGCTAC ACCCGGCTAA
27901 TTTTTTGTAT TTTTTAGTAG AGACGGGGTT TCACTGTGTT AGCCAGGATG
27951 GTCTCGATCT CCTTACCTCA TATCCACCCG CCTCAGCCTC CCAGAGTGCT
28001 TGGATTACGG GCGTGAGCCA CCGCGCCCGG CCATTACCGC ATCTTTCTAG
28051 AGAAAATCCC AAGACTCTTT TTAAAAATCA GCGGTATGAT TTTTGTTGTT
28101 GTTTTAATTT TCATGAAATA TTTAAAGAGG CAGCTACTAC TTCTGATACT
28151 ATCAAAGGGC GGCTTTGGAG CCATGCTGAA AGGCTAGAGG TGTGCCTAAC
28201 AGTCTCTCCC TATATTAGGC CACTTTGTTC TGACTGCTGT TTTTGTGATT
28251 AGTTGATCTG CTGTCCTGGA GATGGGTGGA AACGCGCAGA CCCAGAGGAG
28301 GCTGGCAAAC CTTCACCCCC ACTCAGAGGA GACACTCCCA GTCTCAGCCC
28351 CACCCCCATG ACATTATCAG CTGTCAGATG CTGACTGGGG ACTGGGGTGG
```

FIGURE 3

```
28401 GGGGACTGGG GTAAACTGGT CTTCTAATAC CCCGGTGGGT CCCCAGAATT
28451 CCCCCAGTGA AGAATAGAAG GTCCAGGTGC AATGGCTCAC GCCTGTAATC
28501 CCAGCACTTT GGGAGGCCGA GGTGGGCGGA TCACCTGAGG TCAGGAGTTC
28551 GAGACCAGCC TTGCCAACAT GGTGAAACCC TGTCTCTACT AAAAATACAA
28601 AAATTAGCCA GGCGTGGTGG CGCACGCCTG TAGTCCCAGC TACCTGGGAG
28651 ACTGAGGCAG GAGTATCGCT TGAACCCAGG AAGTGGAGGT TGCAGTGAGC
28701 CGAGATTGTG CCACTGCATT CCAGCCTGGG CCATGGAGCA AGATTCTGTC
28751 TCAAAAAACA AAACAAAACA AAACAAGAAT AGAAGGAAAG TGAGAGGAAG
28801 CAATACTGTG TTTGTAGAGG AAAAGGAAGG TCAGGTCAGC GCCAACAGTG
28851 CCTTTGCTGG GATGCTCCAC ACTCCTGACC AGCTCAGTGG AATTCTGAGA
28901 CCTTGGGGAG CAAAATCAGA GGGGGACAAG GAGAAAGACA GAAAGAGAAA
28951 GAGGGAGAGA GAAAATTGTG TTGGCCGCGG CTCGGGATTT ATTTATTGTA
29001 CTTGCTGTTG ACTGAGCTCG CTCCACAGCA GGCGAGGGGC CTGTAAACAC
29051 ACAGGCTGAT TCATTAGTTT CTGACCATCT GCTTCCCGGG CCGGGGCGGG
29101 GGCCGGAGTG CCCCATAGGT TGGAGCAGCC AGCTTAGAGC CCCCATCTCC
29151 AGCCAGAGCC AGCTGTGTCA GTCTTCCCAG TGACAGACAG GCCGGGTGCA
29201 ACTGGGGGCT TTGAGGGATA CCCTGTCCCC CGCCACAATG CCCCCACCCC
29251 ATAGAACCCA TCAGGGGCCT GGGCAGGGCA GGGCTAAGTA AGGCAGGAAG
29301 ATGAAGAGAG ACCAGGGAAA AGGTGAAGTG TGGGTGGCTC AGCAGTTGGG
29351 GATGACGGGA ATACTATCAA CACTCAATAG GAGATATGTC AATATCACTG
29401 TCATTTTATT AATAGCTTAT TAATATGATG ATAATAATAA TAAGTATAAT
29451 TGAGTGAGTG GTTACTGAGG TCCAGGCCCA ATATGCAGGC CTCACTACAC
29501 TGTCCCCAAT AAGCACATTA ACAACTCTAT GATGGATGAT GAGAGAACTG
29551 GGTTCAAAGA TTGATTCAAG TCAGGAGGGC CTCCTGGAGG AGGCACATGT
29601 GGGAGGGCTC TGAAAACCAG GAGGACAGAC AGTGGCCAAA GGAGCACAGG
29651 CTCTGAGGTG GGAGGGGACG AAGTGGCCTG GGAAATGGCA CAGCTTCAGT
29701 GCTGGGTTGA TAAGAGCTTT CTGGGTAGGT GACGCAGTGT GACGCCACTT
29751 GGTGCCAGCG TACTGAGAAA CCAGGCTGGT GGAGGAGACC CCAGAGACAT
29801 CCTCTGGGGA AATAGGATGA AAAACTCCCA ATTCGTCTCT ACCCCTCTGA
29851 GCCATTCCTC CCGGCCTGTT CCCTCTTTGC TCACCTGGCA TGAGGCTTGT
29901 CTTAGTCAGC TCCGGCCACC ATAACAAAGT AGCACACGAG ACCCAGGTGG
29951 CTTAAACAAC AAAAATGTCT TTCCTTGTTC TGGAGGCTGA AAGTCCAAGA
30001 TGAAGTTGTT GGCAGGTTTG GTTTCTGGAG AGGCCTCTCT CCCTGGCTTG
30051 GAGCCAGCCG CCTTCTCACT GTGCCTTCAA ACGGCCTTTT CTCTGTGCAC
30101 ATGGGCCCCA GCGTCTCATC CGGCAGCCAA ACTTCTTCTT CTTCCAAGGC
30151 CACCAGTCAG ATTGGACTAG GGCCAACCTG AGGGTCCCAT TTAACCTCAA
30201 TCACCTCTTG AAAGAGAATC TGTCTCCAAA TACAGTCCCA TTCTCAGATA
30251 CTAGGGTTAG GGCTTCAACA CAGAAATCTG GGGGACACCA TTCAGCCCCT
30301 AGGAAGTCCC CAGGGGCAGC AGAAGCAGCA CTGGGACTTC AGGAAGGGAG
30351 TGATGGCTGG GCTGGGGGAA GGGGGGCATG TGCCTGGGTG CCAGAGCCCA
30401 GGTGCAAGTT CTTGGAGCAT ATGCCCATCA CTCCTCCTCC CGAGCGGCAC
30451 AGACAGAGCC CATTCTGCAT CCAGGCTCAC AGCACCCTTC CCGGCTGCTG
30501 GCTCCCTCAC GGGTCAAGCC TGCCTCCTCC AGCCCAGCAC CTGCTACCCA
30551 CTCAGGCCAT TGGGCCATCC CCACAGCCTC TCAGTCCCTC ACCATTCCAG
30601 GCAACAGCAC AGGACAGGGT GTGTCTTGCC CAGAGTCACC CGGCCAGTGA
30651 GTGAGGAAGC TGGGAACCCC CCCGCCCCAT GTCCCTCTGC TGTCCTGGCA
30701 TGGTGGGCTT GCCAGTGGTG TGGATATCTC TGATGATAAC AGCAGACTCC
30751 TCCTCAGAGG TGCCATCTGC TCTCGAGAAG GCTGGGGCCA GTCCTACCCT
30801 GGCCCGGGCC TCAGTCCCCC CATCTGCTTA AGGAGATATT GGACTCTATG
30851 GTGAGCTCCT GCCACCCTGA GGTTGCATGG GGGCCTGAGG ACAGAGCTAT
30901 TTTGGTCAAA GAGGGATTTG TGGAACCTGC CTCAATGAAC CGAGGGTCAG
30951 CCCTGCCACC GAGGAGCCCC AGGCAGCCAG GGGACCTGTC CTTCACCCTA
31001 GAGGATAGAA GCCCAAAGGC TGCGGCTGCT GCAGTGGCAC CTGGTGGGGG
31051 CCGCGGGGCT GTGGCTCAGC CCCTCAGAAG GCGGTGGGCC CCATTTCCCC
31101 CATGGGGGCC AAACAGCTCA TTTGAGAGTG AGAGGTTTTA ACTTAGATCC
31151 AAGCCATTTT GTGTCTGAGC AAACCAGACA CCTGAAACAT GCATCAGAAA
31201 GGGCCACAAG TGGTCCGGAG CACTGGGTGG TTGTTAGTGA CAGTGTTCGT
31251 TTTTTTCTTT CTTTCTCTTT CTTTCTTTCT TTCATTCATT AATAGACTTT
31301 ATATTTTAAA GCAGTTTTAG ATTTACAGAA AATTGAGCAG ATAGTACAGA
31351 AAGTACCTAA AACCCCATGC CACCCTCCCC CACCCACCAT CTTTCCCTGC
31401 CTATTATCCA CATCTTGCAT TAGTGGGGGA TATTTGAGGT TTTGGGGGAT
31451 TTCTCTTTTT TTTGTTTTCG TTTTTGTTTT TGTTTTTTGA GACAAAGTCT
31501 CGCTCTGTCG CCCAGGCTGG AGCGCAGTGG CGTGATCTCG GCTTACTGCA
31551 AGCTCCTCCT CCCGGGTTCA TGCCATTCTC CTGCCTCAGC CTCCCGAGTA
31601 GCTGGGACTA CAGGCGCCCA CCACCACGCC TGGCTATTTT TTTTTTTTTT
31651 TTTTTTTTTG TATTTTTAGT AGAGACGGGG TTTCACTGTG TTAGCCAGGA
31701 TGGTGTTGAT CTCCTGACCT CGTGATCCAC CCGCCTCGGC CTCCCAAAGT
31751 GCTGGGATTA CAGGCGTGAG CCACCGCGCC TGGCCTTTGT TTTGGGGAAT
31801 TTCTTTAGAG ACAGGGTCTC CCTCTGTGAT CCAGGCTACA GTGCAGTGGT
31851 GCTGCATAGC TCACTGCAGC CTCAAACTCC TGGGCTCAAG CAATCTCCAG
31901 CCCCAGCCTC CCAAGTAGCT AGGACCACAG GTGTGCACCA CTATGCCCAG
```

FIGURE 3

```
31951 CTAATTTTTA AATTTTTTTT TGTAGATACA GGGTCTTGCT GTGTTGCCTA
32001 GGCTGATCTT GAACTCCTGG CCTCAAGTGA TCCTCCCACT TCTGCCCCCC
32051 GAGTAGCTGG GACTAGAGGT GCATGCCACC ACACCACACC CAGTTTATTT
32101 TTTATTGTTT GTAGAGACAG GGTCTCTCTG TGTTGCCTGG GCCGATCTCA
32151 AACTCCTGGG CTCAAGTGAC CCTCCTGCCT CGGCCTCCCG GAGTGTTGGA
32201 ATTCCAGGTG TGAGCCACCA TATTAGTCAT GTCTGATGAA CCATTGTTGA
32251 TACATTATTA TTAACTAAAG TCCATAATTT ACATGAGAGT TCACTTACTG
32301 TGTTGTACAG GTCTATGGGT TTTTTTGTTT GTTTTGTTTT GTTTTTGTTT
32351 GTTTTTGTTT TTGTTTGATA CGGAGTTTCA CTCTTGTTGC CCAGGCTGGA
32401 GTGCAATGGC ACGATCTCGG CTCACTGCAA CCTACGCCTC CAGGGTTCAA
32451 GTGATTCTCT TGCCTCAGCC TCCCCAGTAA CTCGGATTAC AGGCATGCAC
32501 CACCACGCCC GGGTAATTTT TTTGTATTTT TAGTAGAGAC GGGGTTTCTC
32551 CATGTTGGTC AGGCTTGTCT TGAACTCCTG ACCTCAGGTG ATCTGCCCGC
32601 CTCAGCCTCC CAAAGTGCTG GATTACAGGT GTGAGCCACC GCGCCCAGCC
32651 TGGGTCTATG GGTTTTGACT CATTCATAAT GGCATGCATC CACCACTGCA
32701 GTATTATATG GCATAGCCTC ACTGTCCTAA AAATGCCCTG TGCTCGGCCT
32751 AGTCATCCCT CTGACCCTGG GGACCACTAC ATTAGTGACA TTTTCAGATC
32801 CCTCTGGCCT GTTTCCATCA GTGACACTTC TCACACCAAA TGTATGGGTT
32851 TTCCACACCA ACAACCTAGT CTCCAGCTCT CTGGACACCA CTGGGTGTCC
32901 GATGATTCCG TTCAATTGTA ACACTGCCCA GAGCTGGCGT CAGAACCCTC
32951 AGGTGAAGGG CTCAGTCGCA CCACACTGCC CCCACCTCTG ACGCAATTGC
33001 AAGTTCCTGC CTCCCGTACT TCTGTCCAAT TGGCTGTAAA TTGGGGGTTC
33051 CCACGACCCC TCCCTCAGGT TTGATAGTTT GCTAGAACAG CTCACGGAAC
33101 TCAGAAAGGC ATTTTCCCTA CATTTACCTG TTTATCAAGT TCAGGAACAG
33151 CCAGATAGAA GAGACTCACA GAGTAAGGGA TCGGGGAGAG ACCCAGAGCT
33201 TCCAGGCCCT CTTTGAGCAC CTCAATGTGT TCACCAATCC TAAGGGTACA
33251 GGTAAAAAGG CTCAGGGGCC AGGCGCAGTG GCTCACGTCT GCAATCCCAG
33301 CACTTTGGGA GGGCAAGCCA GGTGGATCAA CGGAGGTCAG AGTTCGAGAC
33351 CAGCCTGGCC AACATGGCGA AACCCTGTCT CTACTAAAAA CACAAAAATT
33401 AGCCGGGCGT GGTGGTGCAT GCCTGTAATC CCAGCTACTT GGGAGGCTGA
33451 GGCAGGAGAA TCGCTGGAAC CTCGGAGGTG GAGGTTGCTG TGAGCCGAGA
33501 TCGCCCCACT GCACTCCATC CTGGGAAACA TAGCAAGACT CCATCTCAAT
33551 AAAAAAATAA AATTAAAATT AAAAAATAAA AGGCTCAGGA AGGTGACTCA
33601 GCTAAGGAGA TATTTAGAGA GTCGAGAATG TGAATGGGAA TTTTTGAGGC
33651 GGACAGAGTG GAAGGGCACA TGAAGGAGAC AGAATGCCAT CTGCAAAGGC
33701 TCAGAGGTAC AACCAGGCAG GGGCAGGAGG ACAGCTGTCC CGGGGAGGCC
33751 CAGGCACAGT GTCACCAGGT GTGGGAGGGC AGGTTGGGGT TGGTATATTA
33801 CAGAGTCAGG ACTCTGTATT CCAGGCAACG GGGAGCCATG GAGGGCTTCA
33851 AAGCAGGGGA CTAACACCCA GCCTTCTTGG TTGTCTGTAT TTAGAGCAAG
33901 GACAAGTATG CATTTTTTCT TTGGAAGCCA TGGGATTGCA CACCTCTGCA
33951 CGCATTGAGG AAATGGGCAC AAATCTCCCA GCTGTGTCTG GGCACCGGAG
34001 CACAGGCCCT GTAGCCACAG AGCAGGTTTC TCGGTTCGCA GGCAGGGGCT
34051 AGGCCGCGCA GAGGCTTCCC AGGCCATCCC AAGCATGCAG CACCCACCCT
34101 TCCCTTCTCC TCTCCCCGCT GCAGGAGAAT CCCTACCTAT GCAGCAACGA
34151 GTGTGACGCC TCCAACCCGG ACCTGGCCCA CCCGCCCAGG CTCATGTTCG
34201 ACAAGGAGGA GGAGGGCCTG GCCACCTACT GGCAGAGCAT CACCTGGAGC
34251 CGCTACCCCA GCCCGCTGGA AGCCAACATC ACCCTTTCGT GGAACAAGAC
34301 CGTGGAGCTG ACCGACGACG TGGTGATGAC CTTCGAGTAC GGCCGGCCCA
34351 CGGTCATGGT CCTGGAGAAG TCCCTGGACA ACGGGCGCAC CTGGCAGCCC
34401 TACCAGTTCT ACGCCGAGGA CTGCATGGAG GCCTTCGGTA TGTCCGCCCG
34451 CCGGGCCCGC GACATGTCAT CCTCCAGCGC GCACCGCGTG CTCTGCACCG
34501 AGGAGTACTC GCGCTGGGCA GGCTCCAAGA AGGAGAAGCA CGTGCGCTTC
34551 GAGGTGCGGG ACCGCTTCGC CATCTTTGCC GGCCCCGACC TGCGCAACAT
34601 GGACAACCTC TACACGCGGC TGGAGAGCGC CAAGGGCCTC AAGGAGTTTT
34651 TCACCCTCAC CGACCTGCGC ATGCGGCTGC TGCGCCCGGC GCTGGGCGGC
34701 ACCTATGTGC AGCGCGAGAA CCTCTACAAG TACTTCTACG CCATCTCCAA
34751 CATCGAGGTC ATCGGCAGGT AAGGCCGGGG GAAGCCCTGG ATGTCACCTG
34801 CAACCTGGGA TGCTATCTGT TACCTGGGAC GTTATTGGAT ACCTGGGATG
34851 TTACCTGGTA TGTGGAACAT GACCAGGTTC TTGGGATGTT ACCTGGTTCC
34901 CTGGGCGTTA CCTGGTATGT GATACATCGT TACCTGGTTC CCCGGGGGTT
34951 ACCTGGTATG TGATACATCG TTACCTGGTT CCCTGGGGGT TACCTGGTAT
35001 GTGACACATC CCCTGGTTCT TGGGATGTTA CCTGGTACCT GGGACATGAC
35051 CTGGTTCTTG GGATGTTATC TGGTACCTGG GACATGACCT GGTTCTTGGA
35101 ATGTTACCTG GTACCTGGGA CACTATCTAC TTCTTGAGAT GTTACTTGGT
35151 TTCCAGGAGG TTACCTAGCA CCTGGGACAT TACCTGGTTC TCAGGACATT
35201 ACCTGATATT ACATCATAGG GACTGCCTAG GATGGCTGGG GCTGCACAGA
35251 CCCAATGCCT GAGCTTTTCC ACGGAAGAAG GTGCACATGT CGGAGGAAGG
35301 AAGTGCATGT GTTGGAGGTG TGGAGGGCAG GAGTGGAGCC TGTCTGGGCA
35351 GAGCTGGGCT AACTCCTGGT GCTGACTGTG AGAAGTGCTC TGAGGCTGTC
35401 CACTGACTCA GTGAATGGGC TCTCATGTTC AGGGGTTTCC GAGACCGCCC
35451 TCAGGCTCCA TGATTCACTA GAAAATCTCA CAGAACTCAG CAAAGCTGCT
```

FIGURE 3

35501 ATATTCACAG TTATGGTTTA TTACAATGAA GGATACAGAT TAAAATCAGC
35551 AACAGGAAAA GGTGCACAGG GCTGGGTCCA GGAGATACCA GGTGCAAGCT
35601 TCTCATTGTC CTCTCCCAGT GGGGTTGTGC AGGCAGCACT TGATTCTCCC
35651 AGCAGTGACA TGGGACAATG CTTGTGAACT ATCACCAGGC AGGGATGCTT
35701 ACTGAAGCCT CAGTGTTCAG AGTTTTTACT GGGAGATGAT GGTATAGGCA
35751 TTGCTGATGA TCTGGCTGGC CTTTGGCTCC AATCCCTCCA GAGGTCGAGC
35801 TGATACTACA TGGCCCAAGG CATAGGCTCT CTGATGTGGC CCAAGACCCG
35851 CAGGCAAACA TGGACACTCT TCCTAGACAG GACACTCCAA CGGCTTATGG
35901 ATTAGAGGTC ACCTAGCAGA AACCAGGCCG GAGTCAGATC TTTCTTTGGA
35951 ATGTGCAGGG TGTGGACAAC CCAGGCCTGC TGTGGCAGCT CTTTCTTGCC
36001 CAGCTCTGGA GAGGAATGTG TTTTGAGGAA ACTTTGAGGA AGGCAAAAGA
36051 CTGTCATAGG AGATGAAAGC AGCCTTTACC TCCCAAAAAC TGTTTAGACT
36101 GTGGATGATT AAATGGTTCT GTGTTGTCCT GGGTCATTTG TACTTTTGGC
36151 AGATAATATG CCAGTCTGTG GGGGCAATGC CATCTACCCT GGGCCATCTG
36201 TAGAATGGGG ATGGTAACAA GCATAAGTCT CTTGCTGGAT TGTGGTGAGG
36251 GTTAAATGAG TTAAAGTTGA TGAAGCACAC CGAACAGTGC CTGGCACATA
36301 ATACGCACTA AAAAAAAGTG TTTAATAAAA AAATTGGACA AAATAATGGA
36351 GTGAGAGGGC TGGGGCGGGC TGAGGGAGGA GAGAGTTGAG ACCCAAGGGT
36401 GCTGGGGCAG TCCGGGCCAG GAGCCTGGCT TAGGCAGGTG CAGGCAGGGG
36451 CTGCACCTGG GGAGCCCAGC TGGGATGATC ACTGTGGGCA TCTCTCCACT
36501 GCAGGCTTTC AACGTCCACC CAAATCTGTA ATTCCTGACA AGTGGAAAAA
36551 GTGATTTGCC GTATGGCAGG TAGAGCATAA TCCTGTTTTT GTAAAATATA
36601 TAATATGCCT TGTATATCAT TTCTGCGTCT TGAACAGACA TTGGCTGTAA
36651 GTGTGTACAG GACCGGGAAA AACATCATGG GTTGGGGAGA AAGGCTGGAA
36701 GCCCCGTTTGC CAGCCATGTG CTGTGGTGAT GCCTGTGCCG TGAGGCTGGG
36751 CAGGCCGCTG TGGTGCTGCT GTTTCCAGCT TGTATTTGCA TTTTGTTTCA
36801 CTTTGACGCT GTGTGTGGAT TCCAGGTGGG CAAAGGCAAA GGCAGGTCGT
36851 TCATGAAGCT GGAGTCAGCT TCAAGTCTGG CAGATGGGAG TGAAGCAGGT
36901 GCAGACAGGG GCCCGTGGTT CTGACACCCG TGATCTCTAA CCCACCTCCA
36951 ATTGGGACAG GATATTCAAA ATCAGGAATG TTCTGGAAAC CACCCCATAT
37001 ATAGGCACTG TTGTGGAACT GGTGTCCAGA CTTTCAGCCC CTGGGCATCT
37051 CCGTGACCTC TGCCCTCGGG ATGCAATTGT CTGGTTGTTA CTAGACACCT
37101 GCCAATCCCT GCAGTCCCAC TGGGCTTAGC AGCTCAGGGC TGGGAACCAG
37151 CTCCCCAGGC CCTGCAGGCT CCTGGCTTGT CTTCTGACCT TGGACTGCTC
37201 AGGGAAACCC TCATCTTCAG GCAACTGTTC AAGGACCCTG CCTTACCTTT
37251 CATAAGGTGG CTCATGTGGG GCCCTCATGT CTGGGCCAGC TGTAGGATCC
37301 AGACTGGGAC TGTGATGCTC AGGGCCTGGT GGTGCCCCTT CAACCCCAGC
37351 AGGGTGACCT GGGGTTCAAA CCCTGCCTCT CTGCATCCTG CCTTGATTTC
37401 CCATCTGTGA CAAGAGGCCT GAGGCTGCCT GGAGCCCATG TCTCCCTGGA
37451 GGTGGTTCCG CTTGGCACCC AGCCTCTGCG GGCGGCGGGG AGGAGCCGAG
37501 GTGGTGGCGA GAGTGTACAA TCAATTCAGA GAGAGAGTAA ACACTCTTCA
37551 GGCGGCAGCT TGGCTTGTCG TTGAACTCGG TTGGGTTAGA CGGTTCAGCG
37601 AAGGACTGAG GTTACCTCCC CTTTTCCCTC TTCAGTTCCG CTGGGAAAGT
37651 CCAGCCATGC TGAGAAGCCA GAGACACAGG ACTGGTGGCA AAGGGGGATG
37701 TGGGACCCGA AGTTCCTGAG TGGGCACCAG GGACCCCTGG GGGTTCCCAG
37751 GCTCTGGGAT TTTACACCGA TCCTGTTCTG GGGATTCCCT CAGAATTTCG
37801 TGACTGCCAT GTTGCCCAGT GGCTCCGCTG TGGCCCACTA ACAAGAACAG
37851 CAAACGTTTA TCGAGCATTT GCTACGGGCT CATTTGCTAT TGAGCGCTGT
37901 TATTAACGGC TTCATGTGTA TTTTCTTGTT TAATCAATAG CCGTTTAATT
37951 CCCCAATGGT TATTGAATGG CTCAGATCCC ACAAGCCAAA GACACGGCAC
38001 TGAGCAAGCC AGACCTGATC CCCAACCCCT GGAGCGCTCC CCGCGGGGGC
38051 AGCGCTCCCC CAGCCTCATG TTTTCAGATG AAGAAGCCAA GGACTCTGAG
38101 GTCAGAGAAT AGAAGGAGAT TGCACAGCGC CACTACATGG CAGACAGCTA
38151 ACATCTAGAC ACAAGCATGT CCAGGCTCAA GACAGACTCA GCCACCGAGC
38201 TGTCCCTAGC CCTGGCCCGG TGGCTGCCTC ATTTCTGGTT CCAGCTCTGA
38251 CTTGAGCCCA CCTGGAGCTC AGGTAGCCTG TATGATGGAG CGATCCCACG
38301 GGACTCCCCT TTGAATTGAA GGTTCTTTGA TGGGTGGTGG CTCCCTGCTC
38351 TGCCACTCAA GTCAGTGATT TCCAGCTCCA AGGGGGACCC CAATGGCAGG
38401 CAGCAATGTA GGCAGGAGGG GGAGGACAGA GGCACCTGGC TGCAGAGGAG
38451 GAGCTGGAGC AGCTGCAGCA GCTGCAGCCA GCCCAGGGCC CTGGGAAGGA
38501 GGGACTGGTA CAAATGGTGA GGGGGACAAG AGGTGGGGAC AGACTGAGCT
38551 GCAGCCCCAC AGGGGGCAGG TGGGCTGCCG GTGGCTCTGA GAGAAGCACA
38601 GGCACCCGCA AGTCCCAGCC ACGCCCCCTC CCTGCCTCCC GCGGAGCCTG
38651 CTCCACGTGC CCATTCCTGG AAGGCAGGCA GGGGGTGGGA GCTATTTTCT
38701 CCTTCCACAC ATGGGGAAAC TCAGGTGCGG AGCCTGGAGA GTGGCTGGCT
38751 TCAGACTGTG CAGCAAATCT GGGGCAGAAT CTGGATTTGA GGAGGAATGA
38801 GATCATCCCT TTATGGTGCT TGGAGCGGGA GGGGCTTCCG GCCCCCACCC
38851 TCCACACAAC CATTAAGAGA AGCGGGGCCC AAACCTGACC CAGCATCTCA
38901 CAGGAGAACC ATTTAAGACA GGGTTCCCCA GCCCCACCAC TGAGAGCTTT
38951 GGGACCGGGA GGCTTGGGGT GAGGTCCAGA AGGCTGTGTC TTTGCAAAGC
39001 TCCTCAAGTG AGCTGGAAAG GTGGAAGCTG CTGCATCCAT GGAGATGGAT

FIGURE 3

```
39051 GACTGTAACC TGTTGGAACC TCAGTTTTCT CATCTGCAAA ATGGAGAAAT
39101 AGTAATGACT TCAGCTGCCT GTTGTGAGGA GTCGATGAAA TGCTATATTC
39151 AAAGAGCTTA TCCCAGTGTC TCATTCCTCT GGTTAGACGG TGGGAGAGGG
39201 AAGTGGGGGA GGCTTGGGGT CAGGGCTGCC TAGATATGAG GCCATTCCTG
39251 CACCCACCCC AGGTACTAGA AGGAGTGTGG AAGCATGCCT CTGGGTTCAC
39301 AGGGGTTGCA ATCCCAAATG TCTGCAGAGG CTGGAAACAC AACTTACGTG
39351 GAAGCCAGGC ATGGTGGTGT GCGCCTGTAG TCCTAGCTAC TTGGGAGGCT
39401 GAGGCAGAAG GATTATTTGA GCCTGGGAGG TTGAGGCTGC AGTGAGTTAT
39451 GATGGTACCA CAGCACTCCA GCCTGGGTGA CAGAGAGAGA TCCTGTCTCT
39501 AAAAAAATAA TAATAATAAA TTCAAATGAA TAATAAAAAT AATGAGAAGA
39551 AGATAATGTA CATGGGTACC TATGGTACCA GCGGAAGCGA TAGGGAAGGG
39601 TGTGGCCTGG GGAGAAATGG AGACAGCAGG CGCCTCACAA GACAGTGGAC
39651 CAAACATTCC AGATCTTCTG CGGTTACAAA GGAAGGTGCA CAGAAGTATC
39701 TATGTAAAGT CTCTTGCTTT CAAAAATATT GGCACCTACT TCAAAAAGTT
39751 GTTTAAACTC TGTGCAAGGA AAACAAAATA CCCCTGCAGC CCAGGGCAGC
39801 TCACGAGCTG CCAGGGAGCA CCCCTGCTCT GAAATGCTCA GTGTCCTCCT
39851 GCAGCCGTGG TAGCAGCTGC ATTCGATGGG CCCAAGGCAG GGCCAACACA
39901 CTGGAGCCAG GGCAGGGGCA AGATAGCAAA ACAGCCCAGC CAGCTCTGCC
39951 CACTCCAGAC CTGGGAGTCA GAGGGAAGGG GGCATCATCC CACGCAAGTT
40001 CCTCTGCCCA GAAAACCCAG GAAGCTCCAG GAAGGCAGCC CAGCGGGTCC
40051 TTTGCCAGCA GGAACAGGAA AGGAGGAATG AGTTCCCCAT TCCTGGAGGT
40101 ATTCAAGCCC AGTGGGTGGT GTTTGTGGGA GATGATGAAG GGAGGTCCTT
40151 TGCACCAAAT TAAGGTCTTG TGTGCCCATT CTGTTCAAAA GGAATAAATG
40201 TGGATTAAGG GAACAGGAAA GGCAGGTACC ATTTATTGGG AACCCATGGG
40251 TTACGTGCCA GGCATTGCAT CCTCACCCCC TTTCTGCCCA GGAGCAGGCT
40301 GAAGTCCTTG AGCGAGAGGA AACCCGTGAG TTCCAAGAGG CTGAGCTGGG
40351 CCTGGACGGG CAGATCTGAG CTCCAGACTG GCCAAGCCCG GCTCCTGCAG
40401 ACCCAGCTTC CCCAGACCCT GTGGGGGTCT AGCTGCTCTA TGTCCCTGGG
40451 CCTCCCCCAA CCTGTTTTTT CCTCAACACC CTCCCCTCTT TCTTGCCTTC
40501 TTGTTTTCTT CCATAAACGT CCTGTCCCTA TTGATCGCCC CGTAACTGGA
40551 GATGGACGGA CACCGGACTA GAATAGAAAC AGCAGAGGGA AGAAAACCAT
40601 GGAGGGTGGG CAGGTGGGCG CGGGGAGGGG AGTGCAGAGA CAGGCGATTT
40651 CAGCCACCCT ATTCCTGTAC CCACCCTCCC TGCATCTCTT TCTCAGCTGC
40701 TCTACAGGAA AAACAGGGGC AAGAAGGGAA TTCTGTGATG TTATTCTGCC
40751 TGCACTGACT CCTTGAAACT CTGGGAAGGC AGAGACAACT CCCTCAGCAA
40801 TATTAACTGT CATTCACTGA GAATTCCCAT AGCCCAGGTA GCATCCCGAG
40851 TCCTTTATGT GCATTTCCTC ATTAAGTCCT ATCCTGTGAG GATCATTGTC
40901 CCCATGACAC AGATGAGGAA ACTGAGGCCA GGAGGTTGAG ATATAGATGG
40951 AAGGGCTGGG TGTGGTGGCT CGTGCCTGTA ATCCTAGCAC TTTGGGAGGC
41001 TGACGCAGGA GGATCTCTTG AGTCCAGGAG TTCAAGACCA GCCTGGGCAA
41051 TATAGTGAGA CCCCCATCTC AATGAAAACA AAATAGATGG AGGTACCCAT
41101 GAGGATGGCT ATTATTGGAA AAGCAAAAAA TAACAAGTGT TGGCAAGGAT
41151 GTGGAAAAAT CGGAACCTTC ATGCATTGCT GGTGGGAACA AAAACGGTGC
41201 AGCCACAGGG GAGAACAGTA TGGTGGCTCC TGAACAAGTT ACATGTAGAT
41251 TTAGCAATTC CGCTCCTAGG TATAGACCCA ACAGAACTGA AAGAAGGGAC
41301 TTGAACAGAT ACTTGTACAC CCATGTTTAT AGCAGCTTTA TTCAGAGTAG
41351 TCAAAAGGTA GAAACAACAC AACTGTTCAT CAACAAATGA AGACATAAAA
41401 TGTGACCCAT CCATATAATG GAATATGATT CAGCCTTAAA AAGGAGGGAG
41451 AACCTGACAC ACGCTGCGAC ACAGATGAAC CCTGAGGACA TGATGCTAAG
41501 TGAAATAAGC CAGACACAAA AGGACAAGTA CCATATGATT CTGCTTACCC
41551 GACGTCCCCA GAGTCGTCAA ATTCATAGAG ACAGAAAGTA GAATGGTGGC
41601 TGCCAGGGGC TGGGGGAAGG CGGAATGGGG GTTAGTATTT AATGGAGACA
41651 GCTTCAGTTT GCGATGATTT AAAGTTCTGG AGAGGGATGG TAACGGTGCT
41701 TGCACAGCAC TGTGGATGTG CTAATGCCAC CAGACTGTAC ACTTAACAAT
41751 GGTAAGGTGG CTGGGCGCGG TGGCTCATGC CTGTAATCTC AGCACTTTGG
41801 GAGGCCGAGG CGGGCGGATC ACTTGAGGTC AGGAGCTCAA GACCAGCCTG
41851 GCCAACATGG CAAAACCCCA TCTCTACTAA AAATACAAAA ATTACCTAGG
41901 TGTGGGAGCG GGTGCCTGTA ATCCCACCTA CTTGGGAAGC CGAGACAGGA
41951 GAATCACTTG AACCAGGGAA GTGGAGGTTG CAGTGAACTG AGATCGCACC
42001 ACTGCACTCC AGCCTGGGCA ACAGAGAGAG ACTCCATCTC AAAAGAAAAA
42051 AACATGGTAA GGTGATAAAT ATATACGTAT ATTTTAGCAC ACTAAAAAAT
42101 GAAGGAAATA GATGGGTGGG TGGTCTCCAT GTCAGACTGT CCAGGTGCAA
42151 ATTCTGGATT AGGTGTGACC CTGGGCAAGT TGTCTAACCT GTCTTTGCCT
42201 CCTTCCCACC TCCACAAAAT GGAGACAATC ACCCTGGAGT TTAAGTGAGA
42251 CCATCCAGAG AGAGGAGCCA CAACTGTCAG GGTTACAGAG AAAGAACAGC
42301 AGATCACAGA GGGCCTGGAG GTTCTTCTAC TGGGGCCTAT GGGCTCTTAG
42351 GAGCGCAAAG GGTGGGTTTC AATCATCCAT GGACTCCCTG AAATATATGC
42401 TAAGTGTTGT GTGTGTGCAG TCAGTTGGTG AGAGCCCCTA GTTTTTGTTC
42451 CATTCTCAAA GGGGGTCCAT GATCCCAAAA TATTTGGAGA GTCCCTGGAA
42501 AGGCAGCTTT GTGTTCTGCT TTTCATTTTT TAAACCCAAC TCTTGGCCGG
42551 GCGTGGTGGC TCACGCCTGT AATCCCAACA CTTTGGGAGG CCGAGGCAGG
```

FIGURE 3

```
42601 TCCATCACCT GAGGTCAGGA GTTCAAGACC AGCCTGACCA ACATGGTAAA
42651 ACCCTGTCTC TACTAAAGAT ACAAAATTTA TCCAGGCGTG GTGGCGTGCA
42701 CCTGTAATCC CAGCTACTCG GGAGGCTGAG GCAGGAGAAT CACTTGAACC
42751 CAGAAGGCGG AGGTTGCAGT GAGCCGAGAT CGTGCCGTTG CACTCCAGCC
42801 TGGGTGACAG AGCAAGACTC CGTCTCAAAA AAATAAATAA ATAAATAAAA
42851 TAAACCCAAC TCTTTGGGGA CACCCAGCTC TTTGGGGACA GACATTTTGT
42901 TAAGTACAGT TCACACACCT TCAACTGCAT CTCCCAGGCC CCCTGAGCTA
42951 CTGCTTCCCA CCAAAAAGCG GGCATGCACG ATTCCAGACC ACATCAGCCT
43001 CCACTGAGAA GTGCCGGTGT TGGGAAGGGG CCAGACTATG CTGAAGAGTT
43051 GGGGGGTGGT GGGGTCCACT GCCAAGGCAG GTGGTTGACC CCTGGACCTG
43101 CGGTCTCCTC TGGTCTTTGG CTCTGCTCCA CTGCTGGGTG GGGCGGAGAG
43151 GGCAGGAGGG GCATCTCGGC CTTGGCCAGG GTCACAGGAG GCTTGGGGAG
43201 CCCTGCAGGT GCTTGGGAGA GAGTCACGCA CAGCAACGCC TTCCCAGCAG
43251 CACTAGGCAG AATCGGGGTT CCCTCTTCTC TTGCCTAAGA AGGCATTTCC
43301 TCTGCGAGCT TTCTCGCCAT CATATTAATT CACTCAACAG ATGTTTTTTC
43351 TCAGACCTTG CAGGGATGAA TTAACACATA TTTCTAATGG GCACTGCCTT
43401 GGTTTCTTGG AGCCTCCATA ACAAAGTACC ACTAACTGGG TGGAGGTCAG
43451 GAGTCTGAAA TCAGGGTGCA GCTGGCTTCG CTGTGGAACC TGCAGAGGGG
43501 AATCTGTCCC ATACCCTCCG CCCTCCGCTC GCTCCTGGTG GGGCCGGCAG
43551 CCTCTGGCGT TCCTGGGCTT CCAGCTGCGT CACTCCAGTC TCTGCCTTGT
43601 CTTGCGGCTG TCCGCTCCCT GAGCATCTGT CTTACACGCT GTCTCCCCGT
43651 GTCTCTTCCT ACAGGGACAC CAATCACATC GGATAAGGGC CCGCCCTCCT
43701 GCAGCATGAC CTCAGCTTAA CTTACATCCT CATCCACATC CACAAAACCC
43751 CATTTCCATG TAAGGTCACA TTCATAGGCA CTGGGGGTCA GGATGTCAAC
43801 ATATCTTTTC GGGAGACACG ATTTGGCCCA CAGCAGGCAC CTATGACATG
43851 CTGGACACAA GCAGGGTGCC AGGGTCCAAG AGCATGGAGG GGCTGACGCT
43901 CCCGTAGTGG GAGAGCAGAA ATCAACAAAC CCAGTACCCC CAAGTGCCCA
43951 CAGGTGCACG GTGAGAAAGA AACCAGGGCA AGGCCAGGCA CGGTGGCTCA
44001 CGCCTGTAAC CCCAGCACTT TGGGAGGCTG AGGCTGGCGG ATTGCTTGAG
44051 CGCAGGAGTT CAAGACCAGC TTGGGCCACA TGGTGAAACC CCATCACGAC
44101 AAAAAATACA AAAATTAGCC AGGCATGGTG GTGTGCGCCT GTGGTCTCAG
44151 ATACCTGGGA GGCTGAGGTG GGAGGATTGC TTGAGCCCAG GAAGTCAAGC
44201 CTGCAGTGAG TTATGACTGT GCTACTGCAC TCCAGCCTGG GCAACAGAGC
44251 GAGACCCTAT CTCAAAACCC AGGGCAATAT GGGAGTGGAG CTGGGAACCA
44301 ACCTGAGACC TGGACCCAGG GGTCAGGGAA GGCCCTCCAG GGAGGGGGGC
44351 TTTGAAGGAG AAACCTGAAG GGGGAACGGG GGCACGGCAT TTGTGGCAGG
44401 GGGACAGCAG GCAGGACAGG TGCAGTGCCT TCTAAGGACT CAATGTGATG
44451 TGGCTGTGGT GGGTGAGTAA GATTCCAAAG AAGAGCAGGC AGGGCCAGAT
44501 CACGCCAGCC CTGCAGCCCC GGGAGGAAGT TGGTTTTGGT TGAAGTTTGG
44551 TAGGAGGGCT ACAGGCAGGG GAGAGGTGTG ATCTGATTTC CGGTGTTTGG
44601 AAATCTGAGT TGGGGGTGGC GGTGCTGGAG GCCTGCAGGA GACCGGGCGG
44651 GCCCTACCAA GACTGACAGC AGCCTGGCCG GGTTCTGGCA GGGACAGGGG
44701 CTGTGATGTG GGCCGCCTGG AGGTGGTCTT TGGTGGCAGT GTTGCTGAGT
44751 AGCTGATGGG AGCAGGCGGT GGGAGGCATC ATGGAGGACT CCCAGGCATC
44801 TGGCTTTGAG ACACTGGGTG GATGGAAATT TTGCCTTGTC AAATGGGGAC
44851 AGAGGTGAGG GGATTTTCTG GAGGGGAATT TTTTTTAAT TTCTTAATTT
44901 TTTATTTTTA TATTTTCATA CAGCAAAGTG GACCTTTTGG GGGTACAATT
44951 CTAGGAATTT TAGCACGTGG ATAGAGTCAT GCAACCACCA CCACAGCCTG
45001 GGCCCACAAC AGCCCGTCTC TCCAAGAACT CCCGTCCCGA GACCCCTGGA
45051 AGCCCCCATC CTGCTCCCTG AAAATCTGCT GTAAGTGGAG TCGTGCGGCA
45101 GGTGGCTTTT TCAAGGGGTC TCCTTCGTTC TCACTGTGCC TCTGAGGCTT
45151 ATGGCGTGGC CGCCTGCGTG GCTCGGGCTC CTCACTCCTG AGTAGGACTG
45201 TGCCGTGTGC ACTGCACACC GTGGTCTGTT TACCCATTCT CCGTGGAAGA
45251 ACGTTTCGGT GGCTCCCAGT CTGGGGCGAT TATGAGTAGA GCTGCTATAA
45301 ACATTCGTGG AAAGGTTTTT GGGTGAACAT AAGTTTTCGC TTCTCTTGGG
45351 TAAACACCCA GGTGTGGACG GCTGGGTCGT CTGGTGAGCG TGTTTAACTC
45401 TATGAGAAAC TGCCAGGCCG TCTTCCAGCG CGGCTGTGCA CTGGAGGAAG
45451 CCTGACTTCT ATTTTGGACC TGTTGGTTGT GAGCGTTTTG GAGGCATCCG
45501 GAGCATTTGG GCCTCTGAGT GTGAAGCGCA GGCGAGACTG GGCGCGGGGA
45551 GTCCTGAGCT CCGCGGTGAC GATGACGCCC GCCGTGGGGG CGTGCGAGAC
45601 TTCCGAGAGG AGAGCCGGGG AGCCGGGCGG GCTGCAGCCC TGAGATCAGA
45651 GACAGGGGGA ATCTGGCAGG GCGGTGCCCC AAAGCCACGC GGGAGCGGTT
45701 CCTGCAGGGC AGCGGCCACC TGTGTCCATG CTGCGGCGGG TCCAAGTGGA
45751 GGAGGCCAGC GTGGGCCCAT GGACCTGGGA AGATGGAGGT CACAGTAGAC
45801 GCACCCAGAG CAGAGCCCGG GAGTCCGGGA TGAAGGCCGT CGTGGGGTGG
45851 TTTCAGGGGT GAATGCCAAC ACTTTTAAAA GAGTGAAATA GGGGTAAAAA
45901 AGAGGGGGTG TAGGCAAGCT TTTGAAAACG TTTTGCTCGG AAGGGGAGCA
45951 GAGAGATGGG GTGTGAGCTG CTGGGGGAAT GGGGAGGGGA GAGGGCTCTT
46001 CCTGAAAGCG GGAGGCGCCG CAGCGTTGGC GTTGGGATGG GAAGAATCCA
46051 CCAGGACTGA AGATAGCTGT GGGGAGGGCG GGAGGGACGC TGGGGACAGG
46101 AGCCGAGGAA GTGATCGGGG TGGGCGGAGA GCTGACAGGA CAGAGGGCCC
```

FIGURE 3

```
46151 ACCAGCATGT GTAAGGAGGA AAGGCTGACT GATTTCATAT TTAGGTCAGA
46201 AACCCTGCTT CTAACACCTT CTAGAGACTC CATGTCCATC TTGACAACCC
46251 AGGGTGGCCC ACCCTGGTGG CTCACAGAAG TCCCTTTGTC CTGAGTAAAG
46301 AGTGGCACAT CACGGTGCCA GCGAGCAGCC TGCCTGAGAT TCCAGCTCTT
46351 TCCATTTGCT CTTTTCTGGC AAGTCTCTTC CATTTCTCCA GGCCTCAGTT
46401 TCCCCACCTG TAAAATGGGA TCCCGCTCTT GTAAAGATTG TCTGAGACTT
46451 GCCCGCTCAA GTCCCAATAC TCAGGAGGTG CCTGAGGAAG GGGAGCCATT
46501 GATCCCTTGA AGGAGTCGGT ATTTTGTCAC CTGAACACAC AGACCATGTT
46551 CCTCCGTCAC TCGCCTGGGG CTGAGTGGAA TTCCTTGGCT TTATCTCCAG
46601 CTCTGTCCTC TGTCCCTTCA TCAGTCAACA GTCATGCAGG AATGAGAAGC
46651 TCAGTTCATA GTAGGTATAG CCAAAGGTAC TTTGGAGGAG AGTGCTGCTG
46701 CCAACAGCGG ACGGGGGACT GAGCCCCACC CAGTTCTGCC TGGGCCACTC
46751 CCTCTGGGTC TGTGGCCACG ACACCCGGCA GGCTCCTGTC CCCCCGACCC
46801 CAGCTCTGTG GCTTGCTGGG CCTCACCAGC TGCGATTGGG CAGGACAGAT
46851 GGGGTTGGAG AGGGGAAGTT GAGTGGGGAG AGGACAGGCC ACATCCAGGC
46901 TCATCCTCAG GCCTTGTTCT GCCCCTGGGC TGGGCTGCCT CCCTCCACAT
46951 CTCAGCAGAG GGAGCGTGGG GTGGGTCATT GTAGAGGATC CCTGGCCCTT
47001 CATCCCATGT GAAGCCTGTT TTAGACTCAC AGATGCTCCC TTAGATGTAC
47051 CCGCCAAGTC ACACACACAC GCAGCCATTC ACTCATCTGT AACATGCCCC
47101 TGACTCACAC ATAAACACAC ACCTGCCCCA CCCACACATG TGAATCCGCT
47151 TCCAGAAGCA CCCAGTCATA GACACATCAC ACTCCACATG GCTGCCTTGC
47201 ACCCCTGCTC ACAGACGCGG ACCCACAGGG TGAGCGGAGG CCTCCGCAGC
47251 AGCCCGGGGT TCACCAGAGG CCTTGGCTTG ACTGCCCTGG GCTCAGGGCT
47301 TCTCTGTGCT TCTCCCCCAT GTTGAGGCAC CCACTTCATG CTTGGGTTCC
47351 TGGAGGCCTC TTTCCACTCT GTGGGACAAA GCCCCCTCTG GGCTTCTAAC
47401 CATCCCTCCA TCCCACTCTC AGCCTGAACT AGGTCCCAAG GTGGACCCCC
47451 TATCAGCAGC CTTGCCCCCT AGAAGAGAGG TGCAGGCTGG CGTCAGGATT
47501 TGGTTTCGGG ACATGACTGG CCTCCCTCCC TCCCTTCAGT CTTGCCCTTT
47551 CTGTAGTTTT TTGTCCCTAA ACACAAACCC AGTCACTTCT CTGCACAATA
47601 TCTCTAGTGG TCTTCTACTG ACCTTAGACT CAAGTCTGTG CTCTGTGACC
47651 ACAGACTTTC CCTCTCACCT CCTTCTGTAG CCTGGGCTCT GAGACTCCCA
47701 AGATCAGCCA CCCCAAAGGC TGGGTCTCTC GGGCTGGCCT AGGGGCTAGT
47751 CCTTCTGCCT GGAACATGAC CTAGCCCTCC CCAGCCTTCC TGTCTGGCAA
47801 ACTCCTCACC CCTCAGGACT TGGCTCAGAG AGCACCTCTC TGGAGCTCCT
47851 GAAGGCTGGA GACCTGTGCA ACCAGCCCCT CGGTCCCCTT CTCCAGCCTT
47901 GGCGGCTTGC GACAATTCCA CAGCCTTGCC ATAGCTACCA TAGCTATGCA
47951 CGTACACGTC TGACACCCGC CCCCATAGGC TCTGAGCTCC AAGGGAGCAG
48001 GGATGGGGTC TGGCTTGTTC TCACCTTAGC ACCAGCACCC AGAACCAGGT
48051 AGGCGTGCTG TATATGTGTG GAATTCATTC ATTCATTCAG TCAGTCAGTC
48101 ATTCATTTAT TCATTCATGA ACAAATGAAC TTACATGAAC AAAGAAGTCT
48151 TACTATCTGG AACTCTTTCC AGAGAAGAAA GAGACTGGAA TATAGGCGCT
48201 AGGAGGCAGA GAGGCAGGAC CAGCTGGGCT TGCTAGCACC CACCCTTGTC
48251 GCCTCCCACC CTTCCACCAT CACACCCTGC TCCTTCTCCT GTTTTCCTTG
48301 TGTGAGAAAT GGGCCTCTCT GGGCTACCAC CTGGGCCTGG TCCTTCCCTC
48351 GGGGGTGTTG GGCCTTCCTG CAGGGGAAGA AGCAACCTTG CTGAAGTCAG
48401 CTAGGCTGCT TAGGCTGGGG GTGCCTTCGC AGGCCCTGGT CATACCCACC
48451 GTGTCACCCC TAAGCCGCAC ACCTCAAGCC TCCCAGCACC CGCCCGTCAG
48501 CCTCTGTGTT GTGTGCAAGG AAGCTGCCTC TGGCTTTGTA ATGGGTAATA
48551 GGATTTATCA ATAGACTGAG GAGGTGAGGT ATGTTAAAGC ACTTAATAGA
48601 AAAGGGCTTC GCACAGAAGC CCAAATTTGA TTTAGCCAAT GAACTCAATT
48651 GCCGGCCTGA TTGCATTCCA GGAGGCGCAG CAGCCAGCAT TTGTCCATGT
48701 TACCCTGGAA AAGCCAGGCT GCGCCAGGCG ACCAACCCAG ACCACCCAGA
48751 CCTCCCCTCT GCCCCCACGG CTCTGTTGTT GGTTCCTCGC TTTCTCAGGA
48801 TCTCAGGTTT GAATGGCAGT CCTTTGACCC AAACAGTCCC AAGTTCTCCA
48851 GCATCCAACA GCCTCCTTCC CTCTAGTGCC CAAGGCTTCC CCATCCCATG
48901 AATTAGCTAG AAGTGCAGTT TGCTACCATG TTCCTTTCAC CATGATCCTC
48951 AAATCCGTGT GTCCCCTGGT CACTTGTCAC CTCTTACTCA AGCTGGCCTC
49001 TGCAGGCTTC CCTCCTGACC ATTTGTTTCT TCTCTGGCTG CTGCCCATCT
49051 CCCCACACTG CTTGCCTGAG CCAAGGACAC ATCTCTCCCA ACCCCCCAAA
49101 GAAGACCAAG ATCCCCAGGC CACAGTGTTT CATGACGTCT TGACTCAGGC
49151 CTGGGCATAG CAGATGGGAA TTACCATGTA GAAAAGAAGG ACTTGAGGCC
49201 GGGCACAATG GCTCATGCCT GTAACCTCAA CAATTTGGGA AGGCAGATCA
49251 CCTGAGGTCG GGAGCTCGAG ACCAGCCTGG CCAACATAGT AAAACCCTCG
49301 TCTCTACTAA AAATACAAAA ATAAATTAAA AATTAGCCAA GCATAGTGGT
49351 GCATGCCTAT AATCCTGGAA ACTCAAGAGG TTGAGGCAGG AGAATCACTT
49401 GAACTTGGGA GGCGGAGGTT GCAGTGAGCT GAGATCGTGC CACTGCACTC
49451 CAGCCTCCAG CCTGGGCGAT GACAGAGTAA GACTCCATCT CAAAAAAAAA
49501 AAAAAAAGAA GGACTTGAGA GGATTTCCAT TAGCCATCAC CAACCTTGAG
49551 GCCGTCCATC GTCCATCCAT TGGCCACGGC CACCCACCGG CTGGGGAGGA
49601 GAAACCTCTT GGCCATGTAA GCCCGCAGCC CCCTCCCAGG GCAGCATCAT
49651 TCCTCACTGC GGTCCAGGTC CCCAACCCCA GCTCTGTGAG TCCTCAGAAA
```

FIGURE 3

```
49701 GCGAGGGCCC TTGTTGCTTC AGCCTCCTCC TTCATCCTGC AGAACACGCC
49751 AGCCTCCCTT GTGAGCATCT CCAAGCCACC GTAAGATCTG GGATTTGCTT
49801 TAGGTTATTT TGCCCATAAC CCAGAGGCTC ATGAAAAATG TTTTTCCCAA
49851 AACAAAAGAA CACTCTTCAC CCAAAGATAA ATGCGCTATC TGGCAAGAGA
49901 AATTGGAAAA CATTGCTGGC TGTGTTAAAC TAGTAGTCTA ACTTTAGCCC
49951 CCAAGGTACC AGAGCTCTGC GAGGCCAGTC TCAGTATACA CGACATGTAA
50001 TAATGTGGGG ATGGGTGGTA ACATACCAGG AAAGAAGGAC CAGGCAATGT
50051 GATTAATGAC CAGGAACCCC CATGGTCCTG CAAAGAGGTG CTTTGAGACA
50101 AGTTGGGGAC TAGCTCTCCC AGCCCAGCAC CTGCCCACCC CAGTTCAGAG
50151 CCATTGCCGT GAACACTCTG AAATCCCCGT GGGGCTTTGC CTTTGCAGAG
50201 AGCCAGCCCT GGGGCCTCCT CCTCCCTGCC CAGCTCCAGC ACCATCCCTG
50251 GCTCGCTCAC CCAAACTCAC CGCTTCCTCC GTCATTCCCC GCAAGGAGTG
50301 GATGACATCA CTCTTTCCGG CAGCCAGGAG TTGACCTACA TTCCACCCGC
50351 CCTCCTGTCT GGGAGACTCT CCAAACCCCC CGCTCCTCTC TCCCTGGGAG
50401 GAGGGAAGGG CCCCGCTCAC GATTTTTGTG GAGTGACGGT GCCAGGCCCG
50451 GGTCCAGATG GTGCCCCCGC AGCAGCTCCC AGGCACTGCC TGCCCCTCCC
50501 TGCAGCGCAG GGCACGCTCC TGCCTGGCAC GGGCCAGGGC CCCTGTCATC
50551 CTTCATGTGG CTGTCAGGCC CCCAGCTGGG CCTGCCACGT TTCCCAGAGA
50601 ACTGGTGTTA TCTGGGCTGG GCTCTCCCCT GGAGGTGAGG CCCGGTGCTG
50651 CCTACTAACA GTTGTGGTCT CCAGGGGCTT ACTAGGGACT CATCCATTCA
50701 AGAAAAAGGG AAACTTAGCT GAAAAGGTGG CTGTGGCTCT GTCCTTGCTG
50751 GCAGCAGGGC CAGCTTCCAG CAAGCAAGTG AGTGCGCCCC ATGTCAGGCC
50801 GTGAGAGAAG TCAGGGTCTG AGCAGAGGGG CCAGACAGCC ACACGGGTGA
50851 GCCGGGTAGC AGGTGGGCCT GCCAGATGAA ATACGGGATA CCCAGTTAAA
50901 CCTGAATTTC AGATAAACAA GGGAATCGTT GTTTTTTAGG GTAAGTATGT
50951 CCCAAATATT TCATGGAATA TACTTGTACT AACAAAATCA TTAGTTATTT
51001 ATCTAAAATT CAAAAAAAAA TTTTTTTTTT TTAGAGACAG GGTCTTGCTC
51051 TGTCACCCAG GCTGAAGGGC AGTGCAGTGG CACAATCACG GCTCACTGCA
51101 GCCTCAACCT CCTGGACTCA AGCGATCCTC CTGCCTCAGC CTCCGAAGTA
51151 GCTAAGATGA CAGGTGCTCA CCACCATGCC CTGATAAATT TTGTGTTTTT
51201 TTAAATTTTT TTGTAGAGAT TGGGGGGGGG GTCTCACTTT CTTGCCCAGG
51251 CTGGTCTCAA ACTCCTGTTC TCAGGTAATC CTCCTGCCTG GGCCTCCCAG
51301 AGTGCCCAGA TTACAAGCAT GAGCCACTGC ACCAGGCCTA AAATTCAAAT
51351 GTAACTCAGT GTCCTGTATT TTTATTTGGG AAATCTGGCA ACCATTGTGG
51401 CATGGGGCTA CTGTGGGGAA ATGACTCCAA GAGGCCAGTG GGGGCCAGGC
51451 ATGGTGGCTT ATACCTGTAA TCCCAGCATT TTGGGAGGCC GAGGTGGGCG
51501 GATCATTTGA GGTCAGGAGT TTGAGACCAG CCTGGCCAAC GTGGTGAAAC
51551 CCCATCTCTA CTAAAATACT AAAATTAGCC GGGTGTGGTG GCGGGCGCCT
51601 GTAATCCCAG CTACTCGGGA GGCTGAGGCA GAGAATTGCT TGAACCCAGG
51651 AGGTGGAGGT TGAAGTGAGC CAAGATCGAA CCACTGCACT CCGGTCTAGA
51701 TGACAAAGCG AGATTCCATC TCAAAATAAG TAAATAAATA AAGGATAAAT
51751 ATCTGGCAAT ATGTTACTGA ATCCTCCAAG AGGCCAGGGG GTGTGTCACC
51801 TCTTGGGGCG CCCCAGGGGC CTGGAGAATA GGTTGATTAT CTCTATTTTG
51851 GGGAGCTGCA GGCTCTGAGA GAGGCTCTGG GGTGGGTGTT GGACGGGACA
51901 GCATGAATAA GGGGCCCTGC CCTTGGGGTC AGGCACCTCA GGGCTGCCAG
51951 GTAAGGCCGA TCTTGGCACC TGGGAGCCCA TGTACTCTCA TCTCCGTCCC
52001 TCCCTCTATC ACCTGTTGGC TGTTGACAAT AGCAACAATA ATAATAGCTG
52051 ACGTTTCCGG AGGGCATGCT CTGTGCCAGA CCCAGTGCTG AGCCCTCAAG
52101 GGTACCAGCT CAGCTTGCCC TTCCCTGACC CTTGACAGTC CTGCCACATG
52151 GAAACAGCCA CTTCCTAGGG TTACAGGACA CCGACTGATT CCCCAAGGCT
52201 GGTCACCTTC CTGCCTTTGC CTAGCTTGAG GGGTCAGAGG TTTGAACCCG
52251 ACTGTCTGGC CACCGCGCTA CCTGGTGCTG CAGCCCAGCC GTGACACTCA
52301 CCGGCCTTTA ACCCAGCAGG TGCTCCATGC CGGTTGCTGC ACACAGAGCT
52351 CCCATGCATT ATCCCCTTCC TCCCTTGCTT CTTTCTGCAA ACATTGAGCA
52401 CCTACTGTGT GCCAGGCAGT GTGTTAGACA TTTACCAAAA ATTAAACCCA
52451 GAGAAGTGAC TTGCTCAAGG TCACCTAGCC AATCAGTAGC AAAACCTGGG
52501 TAAGAATCTG GCTGCTGACT CAGCCTCTCT CTCTCTCTCT CTCTCTCTCT
52551 CTCTCTCTCT CTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTCC
52601 CATGGAGCGT TCTTGCTGCC ATCTGAAGAG CTAAAGCCGT GGGATTCTGC
52651 AGTGGGGCAG GGGTGGAGAG GGAACAGCCC TGGTGGCCTG GGAGGGGTCC
52701 ACTGGCCTCA TGGGCTGGCC AGCATGCCCT ATCTCTATGT TCATTATTAA
52751 ATCATCCTTT CTGATGGATG AGCTGAAAAG TGCTGTGTCG GGGGATCAGG
52801 TGGCCCAGGT GATGGTTTTC TGGTGAGATT GATGGTTCCT GAGGGTCAAA
52851 TTCAGAGAGC GATCGCTTGG GAAAATTGAT CCCACAGAAG AAGGGGAAAT
52901 ATCTGGGCTG GAGTGCAGCA GGAGCCCAGA CTGCCCGCCC AGGTCCAGAT
52951 CCAGGCAGAG GCTGGTGCCA GGAGGGCACG TGGCAAAGGG GGCTTCTTTC
53001 TTCTTCTGTG GCTGCATTTC CTTGGCTGTG GCTGGGCCAC AGAATTGGTA
53051 AGCAGGGCCA TGGGCAGCAG GCATGGGGGC ATCTGGCTTA GTGGCCCCTT
53101 TCCTGGCCTT CTTTGTCCCC ATGGAATGGA CGCGAAGCCA GCACTGAGAG
53151 AGACGCAGGC AGCAGGTGGC TGTGAGATCC AGCTCTAGTG CCATGCAGCC
53201 CCGAGTGCCA AGCCTGCTTC TTCCACCCAC AGCCCTGGCC CTTGACCCAT
```

FIGURE 3

```
53251 CTGAGCCTCG CATTCTCACC TGTAAAGTGG GAAGGACAGT GCCTTCTGCG
53301 TTTTCACGAG GATTTGTTTA ATAATGCATG GCGTCTGTCC ACGTAGGAGC
53351 ACTCAGCAAA TTATTTCTAT GTGCTTGCTT ATTTATTTCA TTTACTTCAT
53401 CTCCCGAGGG CCTTCCAGAA TAGGGCAGAG AAATCACAGA AGGCAGGCAG
53451 CCTGAGGCAG GGAGGGGAAG GGTGTGCCTG TGTCTGGGGG AGCTGTGGGG
53501 TGGCTGCAGT TCCAGGAGCC TGAGTCTATC CAAGAGCCAG CACACTGAGA
53551 CTCTTCAGGC TCTGCCCACT GCCTGCCCAC TTCCTAAAGT GGGGTTCACC
53601 AGCTGTGTGG CCTTGGGCAA GTTGCTTAAC TTCTCTTGCC TCATTTCCTT
53651 GTCTGTGAAA TAGAGATAGT CATAGTATAC TCTGCATAAG GTTCCTGTGA
53701 AGAGGTGATA ACTTCACTTG TTGAAAGCTC TGAGGATCCC TGGCTCTAGG
53751 AGCTCCTGAG TATAAGAAAT GGAAAACTGG CATGTCCTAG GCACCTGTTT
53801 GTCTCATGCA AGAGCTTTAT AGACCGTCTC CCACTTGATC CTTCCAACAA
53851 CTCCATGAGG CAGGTGTTAC TGCCATCCTT ATTTTACAGA TTAGAAAACA
53901 GGCCAAGTGA AATTCAGCAA CTTTCCTGAG GTCACACAGT CGTTGGCAGA
53951 ACATCCCAGG CCAGTGGGGC CCAGGGGATG GGGTTACCAT CTCTCAGTCA
54001 CCCAGAAAGA CAAAAGGAAG GCACACAAAC CCCAGTGGGC ACCGCTCTGT
54051 GCCAGGGCTT TTGTTTCACC CCAACCGTGA TGAGTGTGGG TGATGTGGCC
54101 GGTGTGGTCC ATATCGCTCC TCAGGGCCAG TGTCACTGGT GTGCGTGGCA
54151 GCCTGTGGCA TACACAAGAT ATTGTTTCCC GCTCTTCAGG TGAGAAAAGC
54201 GAGACTCAGG TTGGCTAAGA ATTTGTGCAG GCCACACAGC TGTGTGCACC
54251 ATGAGCTGGG CAGAAGCCCT GCCCACCAGC ACCATGGGGG CTCCCGGCTC
54301 CTGCCACACA GGCCACCCGG GAGTGCAGAC ACTAAGTGTG GCTCTGAATC
54351 GAGCTGCTTG GATTCAAGTC CTGCGTCTGC CCCTTACCAG CTGTGGAACC
54401 TCAGGCAGCT CACTTCACCA CCCCAAGTCT CAGTTTCCCC ATCTGCACAA
54451 CGAAGATAAT AATGACATGT CATTCAGGGG ATTGCTGTGA AGGGCTAATG
54501 TTCTAATTCG GGTGGCACTC TCAGAACAGT TCCTCATGCA AAGTGAGCAC
54551 TCACGCATCA GCTGTCTACC CAGGAGACTT CCTTTGAGGC ACAATGTGAT
54601 AGCGTTTTGT TTTTTTGTTT TTTTTTGTTT TTCTGGTTTT TTTGAGACCA
54651 AGTGTTGCTC TTGTCCCCCA GGCTGGAGTG CAATGGCATG ACCTCGGCCT
54701 ACTGCAACCT CCACCTGTTG GGTTCAAGCG ATCCTCCTGC CTCAGCCTCC
54751 CGAGTAGCTG GGATTACAGG CGCATGCCAC CACACCCAGC TAATTTTTGT
54801 ATTTTTAGTA GAGACGGGAT TTCACCATGT TGGCCAGGCT GGTCTCGAAC
54851 TCCTGACCTC AGGTGATACA CCCGCCTCGG CTTCCCAAAG TGCTGGGATT
54901 ACAGGCGTGA GCCAACGCGC CCGGCCCCAA TGTGATAGGT TTATAAAAAT
54951 AGGAAAACAA ATCTCTGAAG ATTTTTAGGA GAGATACTGC AAGAAAAGCC
55001 TGGAAGCTGT AAGCCCTTCC CTGTTTCCAC ACATCCCGAT CATTCATTCA
55051 CTCACAGCTT TACCCAGATG TGATTCACAC GTTGTACAAT CCACACATGT
55101 AAAGCGTACA ATTCAGTCGC TTTTAGGTTA TTCACAGAGT TGTGCAACTA
55151 TCACCACAGC AGACTTTAGA ACATTTCCAT CACCCTAAAA AGACACTCTG
55201 GCCGGGTGTG GTGGCACACG CCTGTAATCC CAGCTACTTG GGAGGCTGAG
55251 GCAGGAGAAT CACTTGAACC CAGGAGGCAG ACATTGCAGT GAGCTGTGAT
55301 CGTGACACTA CACTCAGCCT GGGTGACAGA GCAAGACTCT GTCTCAAAAC
55351 AAAAAAAAAA AAGAAAAAAA AAGTCATCTA CCTATCCTAG ACATTTTGTG
55401 TAAATGAAAT TATACAAGAT GGCTGGGCAC GGTGCTTCAT GCCTATAATC
55451 CCAGCACTTT GGGAGGCCCA ACTGGGAGGA TCACTTGAGG TCAGGAGTTT
55501 AAGATTAGCA TGGCCAACAT GGCAAAACCC CATCTGTACT AAAAATACAA
55551 AAATTAGCTA GATGTGGTGG TGGGTGCCTG TAATCCCAGC TACTTAGCAG
55601 GCTGAGGCAG GAGAATCACT TGAACCCAGG AGGTGGAGGT TGCAGTGAGC
55651 CAAGATCATA CCACTGTACT CCAGCCTGGG TGACAGAGCA AGACTCTGTC
55701 TCAAAAAAAA AAAAAAAAAA GAAAGAAAGA AAGAAATTAT ACAATATGTA
55751 GTCTTTCATC ACTGACCGCT TTCACTTGGC ATAATGTTTT CAAGGTTCTT
55801 CCATGTCATA GTGTGGATCA GTATTTCATT ACTTTTTCTG GCTGTATAGT
55851 ATTCATTGCC TGCTTACACA TTTTGTTTTT CTGTTGATGG ATATTGGGTT
55901 GTTCCCACCT TTTGACTATT ATTAGCAATG CTGCTATCAA CATTCATGTA
55951 CACATTTTGT GTGGACATGT TTTCGTTTCT CTTGGTTATA TACCTAGGTA
56001 GAATTACTGG GCAACTCTAA GTTCACCTTT TGTAGAACTT CCAGACCACT
56051 TTCCAAAGCA GCTGCACAAT TTTACATTCC CACCAGCAGC GTATGAGGGT
56101 TCTGATTTCT CTGCATCCTC ACCAGCAGGT GATTATCCGT CTTTTTGACT
56151 CTGCCTATCC TGGTGAGTGT GAAGTGGTAT CTCATTATGG CTTTGGTGTG
56201 CATTTCCCTG ATGGCTGTAA TGTTGAACAT GTTTTCATCT GCTTCTTGGC
56251 CATTTGTATG CCTTCTTCGG AGAAATGTCT GTGCTGATCC TTTCCTCATT
56301 TTAAAAACTG GATTATTTGT CATTTTATTA TTGAGTTATA AGTGTTCATT
56351 AGATGTTCTA GATATAAGTT CCTTTTTAGC TATGTCATGA CTTGCAAATT
56401 TTTCTCCCAT TCTGTGGGTT GTCTTTTCAC TTTCCTGACA GTGTCCTCTG
56451 AAGTACAAAA ATATTTAGTT TTGATGAAGT CTAATTTATC TATTTTTTTT
56501 TCTTCTGTTG CTTTTGCTTT TGGCATCCTG TCTAAGAAAC CATTGCTTAA
56551 TTCAAGGTCA TGAAGATTTG CTGCTATGTT TCCTTCTTCT TTTTTTTTT
56601 TTTTTTGAGA CAGAGTCTCG CTCTGTGCTC AGGCCAGAGT GCAATCTCAG
56651 CTCACTGCAA CCTCTGCCTC CTGGGTTCAA GTGATTCTCA TGCCTCAGGA
56701 CCTCCCAAGT AGTTGGGATT ACAGGCACCC ACCACCACAC CTGGCTAATT
56751 TTTGTGATTT TAGTAGAGAC GGGGTTTCAC CATGTTGACC AACTTCAACT
```

FIGURE 3

```
56801 CCTGACCTCA GGTAACCTGC CTGCCTCAGC CTCCGAAAGT GCTCAGATTA
56851 CAGGCGTGAG CCACCGCACC TGGCTCCCTA CGTTTCCATC TGAGAGTTTT
56901 CTAGTTTGCA GCTCTTTGAT CCATTTTGAG TTAATTTTTA TATGTGATGT
56951 GAAGGGTCCA ACTTCATTCT TTTTACGTGG GTCTCCCATA TTATTTATTT
57001 ATTTGTTTAC CTAATAATTT CTTTAAATTA ATATACTTTA AAAATTTGGT
57051 TTCACTTAGG CAGTATTTTT GTGATCCCAG GTCTTATGAG TCTGACATTA
57101 AAAAAAACAT ACTGTGTATT CAAAAACAAA TAATAGCCTG TAGAGCTGTT
57151 TTACAGTCCA TGCGCCCATC TCCCCCAGTC ATCTCGTACA TCCCACTTTG
57201 GGAGCCAGTG CTGATCATGA TGAGTTTTAT GGGGAGGGAG AGTGTAGCAG
57251 GGGCTGATTT GAGGTAGGGG TTGTGAAAGG CCTCCCCAAA GAAGTGGCAT
57301 TTGGCCTAGA ATCTGACACC ACCTCTAAAT ATTAGCTCTA GGAGAAGGGG
57351 CAGAGCCTGT AGTACATTCC AGGCAGAGGG AACAGCATGT GCAAAGGCCC
57401 AGAGAAAAGC AAGATCAGAA CATCTGCTGG ACAGAGACAT GGAGAGAGAG
57451 ACAAAGAGAC AGGAGAAGAC GGAGACATGG AGAGAGGGGC ACTGAGAGAG
57501 CCAGAGAGGC CAAGAGGTCA TACACAAACA CCATGTCCAC ACACACCGAC
57551 ACAGAGGCAG TACCACCCAC AGTGACCTGC AAAAGTCAGA GAGACCAGAA
57601 GCACCTGCAT TGACCAGGTG CCATGTGTGT GCCAGCCTTG CCAGCCTTTC
57651 CCCACACAGT GCCAGCTCCT CCCTGCAGCC CTGCTAGACA GGGCAATTCT
57701 CTCCTTTGGA AGAGGAGGCA CTGTGCTTGT TCTGGGGCTC AGCCAGCGAG
57751 TTGAAGAGGC CGGGACATCT TCTCCTGTTC CCGAGGAAGA CAGGACTGAT
57801 AGGAGGGGCT GGCCAGGATG GAGCCTGGGC AGAGCCAGCT GGGGAGGGAC
57851 GTCTGGATTT GGGTCAGGAG AAACCAACTG TATGACTTTG GGCGACTTCT
57901 CCCTACCAGC ACCACCCAGC CTGTCCAGCT CTGACATCCA ATGGCTACCT
57951 ATTATCACTG TTACAATATG CTTCTTGGAT TCGGCCGGTC TTCTATGCAT
58001 CTTTACTAAG CCAAGGTCAA TGTTATCACC CAAAGTTGAT GGCAATCTTT
58051 TCTTTCGGTA GCTGATATCT AAAAATATCC CTTCCTCAGG CCAGCAGTGC
58101 GGAGCACAGC CCCGGTTCCG TACAGCTGAG GTGTTCCTCT CCAGCCAGCT
58151 CAGGGGGGTG TTAGATAGGA ATTCAGAAGT TGCGAATTAA AACCTCCCCA
58201 GGTCAGATTT TGCAGACAAA CTCTTGTTAG ACAAAGTCAG ACAGGTTTCT
58251 TCCCTGCAGG TCTTGTCAGA GCCTTTAATC TGCTGAGATG TATTATGACC
58301 CTCAGTGGAG TGAGGTAGGA TCAAGGAGCA GATGTCAGGG GATTTGGAGA
58351 GCATCATATG CTTGACCAAA AAGGTTTAAG CACAGAAACT TTCTGTGCCA
58401 AGGACCATCT CTGGGATCTG GCATCTCCAG AAAGAAGGGC CGCCCGTACC
58451 AATAGGTGCC TGCTGTCCCC CTCTGTGCCT TTCCCAGAGC TAAGCACATA
58501 GCAGTCCTGC AGCACGCATG TGGCAATGGA GGGATGAGAA TGCCAGGGAA
58551 AGAGTGAACA GGCTTTTTAA TGCTCTATTT ATTTCAGCTT GGTGTTCCAG
58601 TAAAACCTCG GCAGGGTTAG TGTCCAGGCT CTCAGGGGAA GATGGAGTTA
58651 AAATCACACA GACTTCAGCG GTGGCTAGTG CTGCTGGCCA GGGCTCTTGC
58701 TCTGCCTCAG AATTCTGTTC CAGAGAATTC CAGAGACGTC CTTCGTCCTG
58751 GGAACTGAGC ATCACGGCAC AGTGGATACA TGTAACCTAA AGCGGTGGGT
58801 TCCTGAGTGC CCCCTCTCCC ACGGCCCACC CAGGACCTGC AGGATTCTCT
58851 CCTCCTCCTG GGAGGAAAGG CGTCAGCTCA GAAACTCACT GCCTTCTCCA
58901 CTCTGGCCTG ACCTCATGTT TCTAAATAAC ATTTCTAAAA CATGGACTTT
58951 AGGCCAGTCG CAATGGCTCA CACCTATAAT CCCAGCACTT TGGGAGGCTG
59001 AGGTGGGCAG ATTGCTTGAA CTTCAACCAG GAGTTTGAGA CCAGCCTGGG
59051 CAACACAGGG AGACCCTATC TCCACAAAAA ATAAAAATGA AATTAGCCAG
59101 GCATGGTGGC ACGCGACTGT GGTCCCAGCT ACTCGGGAGG CTGAGGAGGG
59151 AGGATTGCTT GAGCCCGAGA GGCAGAAGTT GCAGTGAGCC GAGATTGTGC
59201 CACGGCACTC CAGCCTGGGT GACAGAGTGA GACTCGGTCT CAAAAAGCAA
59251 AAAGACTCCA TTGCCTGGAT TTACCATAAT TTCCCTAACC ATTCTGCAGG
59301 ATAATTCGAG GACAGAGAGA AGACCCCAAG GAGGAAGGCA GCCCTGGTTA
59351 CCAAAGCTGG CAGATGGGGG TTAGCTGGAA GCCCCAGGGC TGGTGGACGC
59401 AGGGGCCGCT GTTTCCCCAC CCAGATCCTT GCTCTGGAAG GCGGCCCCCA
59451 GGGGACCCTT CATTCCCACT CAGACAGGGA CAGAGGCGGG ACAGAGCCGA
59501 GGGAGGAGGG CTCAGATGAA GCACCTGGCA GGACTGAGAT ATGAGGGAGG
59551 CCGGATGCGA GGAGGGAGCT CTGCAGCCTG TGGTGTCCAG GAGGATTTGG
59601 GGAGTGTGAG GTGAGAAAAC AAAAAGCGTC ACCCCTCCCA GTGGAAGGGG
59651 AGCATGAAGA GAGAAGAAAA TGCCAGTTAC ACATCCTCAA AACAATCTCT
59701 GAATGGACGA GAGCCAGTCA GGGGGAAACG GAAGTTGCAG TGAGCCAAGA
59751 TTGTGCCACG GCACTCCAGC CTGGGTGACA GAGTGAGACT CTGTCTCAAA
59801 ACACAAAAAG ATGCCGCACA CCCCAGGCGG GACGGGTGGG CCCAGGTGCC
59851 CTGCCCTGGT CCAGGCCTCT ATTCGCAGGA TCTCACTGAT CTCTCCCTGC
59901 CCAGGCTTTT GTCTCCCAGA CAAACCAGTC TCCATGAAGC AGCCACAGCG
59951 TTATAAAATA TGGCCCAGAG CAGGACACGC CTAACCAGTC AAAGGCTTGC
60001 AGAGTAAAAT CCCGCCTCCG TCTCCTGGTC TGGCCTCCTC CACCCCCCAG
60051 CCTCATTGCA GACCTCACGG GTGGAACTTC TTTCAGCCCT AGGGCTTTCC
60101 TCATGCTCTT GCCTCTTTCT AGAAGGTTCT CCCCTTCTGG CCCGGCAACT
60151 CTTGATCCTC CTTCAGGTCT TAGTTTAAAA CTGTTCTTCC TGGAAACTGT
60201 GGACTCCCGT GATGCCTTCT TCTTTGCACA ACCCGTGGGC TCTTTTAACA
60251 GTTGAACACT GGAAGAGGTG TGTGATTCGT ATTGTTATAA TTAATAGACT
60301 TTATTTTTAG AGCAGTTTTA AGTTTACAGA AAATTGAGCA GATAGTACAG
```

FIGURE 3

```
60351 AAAGTTCCCA TTTCCCCACC CCCCTGCACA GTTTCCCCTA TTTTGGAGTA
60401 TATGTGTTAC AATGATCAGC TAACACTGAT ACAGTGTCAT TCACTAGGGC
60451 CTGTTATTTA CATTAGGACT CACTCTGTGT TGGACACTTC AGTGGGTTTT
60501 GCCATATGCA TAATGCCACA TATCTACCAT TGTGGCATCA TAAAGAATGG
60551 TTTCACCGCC CTAAAAATCC CTGTGTGCCA CCTATTTATC CTCCCTCCCT
60601 GTCCCCTCCA ATCACTGATT TTGTTCATTG TCTCCCTGCT AGACCCGAAG
60651 CACCATCAGG TCAGGGAGCC ATCTGCTTTG TTGACATTAT ACTCACTGTT
60701 CCCAGCTCAG AGCCCCTGCC ACATAACAGG TGCCTGATAA ATATTTTTCT
60751 TTGCATATTT TAATAAAAAT AGGATTCTAC TAAACCCATT GCTCTGCAAT
60801 TTGCTTTTTC CAAATATTGA CAATATATCT TGGAAATCTT CCCATATCAG
60851 AATATAGAGC ACACTCTGTC TCTCTTTCAC TTTTTATTAT GGAGAATTTC
60901 CAACACAAAA GAAGATGGTG TCTGCCATGA TGAACCTCTG TGTACCTCTC
60951 ACCCGACTGC AGTTACCAAC TAGTGAATCT TCTTGTTTCA TCTGTACCTC
61001 TGCCCATCAG ATAATTTCAT CTGTAAATAT TTCTATGTGT ATCTCTCTCT
61051 GCTTTTTTTT TTCCTTTTGA GATGTAGTCT CGCTCTGTCT CCCAGGCTGG
61101 AGTGCAGTGG TGCAATGGCT CACTGCAACC TCTGCCTCCT GGGTTCAAGC
61151 GATTCTCCTG CCTCAGCCTC CCCAGTAGCT GGGATTACAG GTGCACCATC
61201 ACACTTGGCT AATTTTGTAT TTTTAGTAGA GATGGGGTTT CACCATGTCG
61251 GACAGGCTGG TCTTGAACTC CTGACCTCAA GTGATCTGCC TGCCTTGGCC
61301 TCCCAAAGTG CTAGGAATAC AGGCATAAAC CACCATGGCT GGCCCTTTTG
61351 TTTTGTTTTG TTTTTGTTTT TGTTTTGTTT GTTTTTGTTT TGTTTTGTTT
61401 TGTTTTTTTT GGAGACAGAG TCTTGCTCTG TCTCCCAGGC TGGAGTGCAG
61451 TGGTGCAAAC ATGGCTTGCT ACAGCCTTGA ACTCCTGGGC TCAAGGGATC
61501 CTCCTGCCCT AGCCTCTTGA GTAGCTAGGA CTGCAGGCAT GTGCCACCAC
61551 ACTGGCTAAT TTTTTTATTT TCTGTAGAAA CAGGATCTTA CTATGTTACC
61601 CAGTCTGGCC TCAAATTCCT GGGCTCAATT GATCCTCCTG CGTCAGCCTC
61651 CTAAAGTGCT TGGATTATAG GTGTGCACCA CCACGCCTAG CCTGTGTGGT
61701 TCTCTAAAAG ATAAGGACTT TAAAAAACAT AACCACAATA TCATCATCAC
61751 ACCTAGTAAA CATTAGTATC TAATATAGGC AATGTCCAAA TTTCCAGTTA
61801 TCTCATAAAT ATACGTTCCA TTTGTTTGCG TGTTTGATTT TACAGGATGT
61851 TTGAACCAAG GACCAGAGGA GCTTCACTTC TGCCATTAAT TGGTATGCTT
61901 TTGAGGTTTC TCTTGCTCTA GGTCCATCCT TTTTCCGTAT CAATTTTTTC
61951 TCCACGCAAC TCATCTGTTG GAAGAAGTGG GTCATTACTG CCGGAGTTTC
62001 CTCGAGTCCA GATCTTGCGG ATGGCATCCC CTGGTGTCGT TCAGGGTGAT
62051 GCTCTGTCCT CTGTCTTTCC TGTGCGCTCA CAGCTGGGTC TAGAGGCTTG
62101 ATCTGATTCA GGTTCAGTTG CTTTGGCTTC GGGTGGGGGT GTGTTCTTCT
62151 AGAGACGCTC ACTGTCTGCT GGCCTCTCAG TGGGAAGTTC TGGGCTGGTG
62201 CGGCTCAGTG CCTGGAGCTA TGCATTCATC AGGGACTGCA AATTCTCTCG
62251 TGAAATGCTT CTATAAAAAG AAGCTTCCCC TCATCAACTT TCGGTTACCC
62301 CGAGGTACAG TTTATATAGG ACAGGCAGGA TAAATGCTTG ATTCTTTCCC
62351 TTTATTTACC AGTTTTCAGA ATAATGAGTC GCCTATAAAT ATTTGAATGA
62401 ATGGATGAAT AATACACCTG TGAATGAACT GACGGGGGTG TGGGAGTTGG
62451 GGGTTCTATC TGTTGCCCTC AAGGGCGTGG CTGTGGGCCA CAACCTGACA
62501 GCAGAGGTCC AGCCTGAAGC CAGGTGCCTT TCTACACAAT GAAGTGCAGC
62551 CCATGGCCAA GTCTCTGTCT ACACCAGGTG CTGGCAGCAG CCTCTGTCAC
62601 CCATCAACCC ACATCAGTCA AAGGCTAGGG TGATCAGAAG CTGCATTACT
62651 AAGAATCTAG CATCTGGGAC AAGGCAGTAT CATCACTCTC CCCAACTGAG
62701 ATGTGAGGAA TCCTTGAAAC CAGCATCAGA GCAGAGAGGA GAGCCGCGCA
62751 GTGACTGCAG GTGTGGCCTT TGGAACACGG CGTTGATCTC TCTGCAGGAA
62801 GGGGAATCAA GGAGTTTCTG GCCTAAAGGT TGGGCTGGTG GCCTCCAGGG
62851 TTTCTTCCTG GGCAGCCCAA CACCCTCCTG GGCCCCTCCT GGGAGGCGCT
62901 CCTTTCCCCA GAGGCCAGGC CAGGCTGCCC ACAAGCTCTC TGACATCTCT
62951 GCCCTCTCGG TGTCTCCCCA GGTGCAAGTG CAACCTGCAT GCCAACCTGT
63001 GCTCCATGCG CGAGGGCAGC CTGCAGTGCG AGTGCGAGCA CAACACCACC
63051 GGCCCCGACT GCGGCAAGTG CAAGAAGAAT TTCCGCACCC GGTCCTGGCG
63101 GGCCGGCTCC TACCTGCCGC TGCCCCATGG CTCTCCCAAC GCCTGTACGT
63151 GCCATGCCCC GGGGCCACGA GCCCACATGG CTATAATCTT CCCTGCCCGT
63201 CAATCCCAGG AGCTGTGGAT CATACACACG CACACAAACC TGCACACAGG
63251 CACATACGTG TGCACATGCA TGCAAACGTG CACACAGAAA CATACGAGCA
63301 TGCATGCACA GGCATGGGCA CACATATGAA TGCAAAAACA CGTGCATGCA
63351 CAGAAACACA CGTGCGTGCA TGCACCCCCA CACACACACC TTGTTCTACA
63401 GCTCCCAAAT GCCAGGTCTC ATAACAAGTC CCTCTAGCAT ACCTGCATCC
63451 TGCTGAATGC TAAGCTGCCC TTCCAGCCCT GGTCCACAGG GAAACCCGAG
63501 AGGAGCTGCT CAGCAGCATT CCTGCAACCC TTCGCCTTCT TGACCCTGAA
63551 GGCTGGCAGG TGGCCCCCAT GGGATGGCAG GGAGTGATGG GGTGGGACCC
63601 CCACTTCAGG TGAGGGGATC AGTAGCATCA TCCCTGATCA TAAGCATCTC
63651 TGGGGTGTTT GGAGCGAGCC ACCAAGGCCA AGTGCTGCGT GTTCAGCACT
63701 CACCTCCATC CTCAGAGCAA CCCCTGGGAA CTGAGCTTCC GTTATCCCTT
63751 TTTACAGATG TGTCCTCTAC GGCCTGCACC AGAGTCCCCC CGCACGTGCT
63801 CTCTGCCCAC CCCTCTCCTG TCCTCGCCCA GCCATGCCAG GGAGCCATCC
63851 TCAGGGCCTG CCCACTCACA TCTGCATCAT AGACCCCCTT CTGAGGACCA
```

FIGURE 3

63901 CCCACCCGGG CGCGCTAGCG TAGTCCACAC TGTCTCCATG AAATCAGAAT
63951 GCCAGGAAGC CCCTGGGCTG CAAGTGGAGA AGGCCACAGA CTGCCCCTGG
64001 GGTGGGCTCT CCCTGACCCC CACCACTGCC ACTTTACGGC CGTTCTCAAG
64051 TGTGTGGTTG GTGTCCGAGG CCCTCCTAGG CACTTTGTGC ATCTCGCTTC
64101 ACCCTCACAC AGCCCCGCAG GGTGGAGTCC TTATTACGGT CCCTTCTCCA
64151 GATGAGCAGA CTGAGGCCCA GCGAACCTCG GGGCCCTCCC AGGTTTCACA
64201 GTGAGTTGCA TGCCAACGCC AGGGCTTAGC GCAGCTACTG GATCTGATGC
64251 TCAGACCAGG CCTGGGCAGT GGCCTCCTGC ACCCCATCTG GGCATCCTCC
64301 CTCGACGGGT GAAAGATTTA TGTGACTTTA GGCTGATCAA AGCTGAGAGC
64351 CACGTGCAGA GGCCATGACA GGCATGTCAC GCGTGTGCTA CAGCAGGCGG
64401 TGAGACAGGC CTTTGAAGGG GTGACTTGGC AGGCACCCGA CAGCACTCTC
64451 GGGTCTTCAA AGGCACAAAG AGCACTCCGG GGGCTGACAG CCCACTCCAG
64501 CCGCTGCCTC CTCCCAGCTT CCTGTCCCCA TCTCTTTGTG GTCCAAACCC
64551 TTTCCTGAGG TCTCTCTTTA GTGGCCGTCA CTCTCTCCCT GACTCACAAA
64601 TTTGCTGCCC TTTCGTCTGT CCTGAGGGCC GCTCTCTCTT CCCTCCTCTC
64651 TGTCTCAGTT CTCTCCACCA TCACACCCTC CTTTTTTTTT TTTTCTTAAA
64701 GATAAGTCTC TCACTCTGTC GTCCAGGCTG GAGTGCAGTG GCGCAGTCAT
64751 GGCTCACTGT AACCTCAACC TCCTGGGCTT GAGCGATCCT CCCACTTCAG
64801 CCTCCCAAGT AGCTGGGACT GCAGGTGCAC ACCACCAGGC CCAGCTAATT
64851 TTCGTGTTTT TTTCAGAGAA AAGGTTTTGC CATGTTGCCC AGGCTGGTCT
64901 CGAGCTCCTG AGCTCTAGTA ACCCTCCTGC CTTGGCCTCC CAAAGTGCTG
64951 GGATTACAGG CATGAGCCAC CACGTCTGGC CATGTCTGGC TTTTTTCTTC
65001 CACTTGCCCC CTGCGCCTGG AGGGTTCTGC CCATCTCTGG CCACCTGAGA
65051 CTCTTACCTG CCAGTCACAC CTAGGAGGAT CCACTGTCCC CACCCTCACC
65101 CCCAGTCACC ACCTCCAGTC TAGCCAGCGA TGGATAGACA GGCCTGTAGG
65151 GGCTGGGCAG CTGGGTCTTC CTTACTTAGG GGAGACTCCT GAGCTGGCCC
65201 CACCTCCTGC TTCCCTGGGA GTCCCCGGGA ATTGCCTTTG AGGCCCCCAA
65251 GCCCCAGGAA GGAGGGGTTT CTTCTGGCAT TTGCAGGGTC ACAGGTGAGG
65301 CTGGAGGCGC TGCTGTCCTC AGCACCCAAG TCTCCTGCCT TCTCTGGGGT
65351 GCCAGCAGGA AACAGCCCAA AGAGACACAG AGGCTAATTT TGCCCTCACC
65401 CTGCCCCTAC CGTGACCCCT CCTCAGTAAG TGGCACCGAC ACCCACGGGG
65451 CCAAGTGCAA GTCTGGGGGT CACCTTGACT GGAGACCCTC CTCCCGCCAC
65501 GTTCTTTGAA CTTCCCTCCA TCCGCTCCAA GTCTCTCCCA ATGCCATCCT
65551 CAGCCCTGCA GCAGCCCTCA CTCCCGATGC CTTCCCACCT CCTCACCACT
65601 CTGCCCCCAC CTGGCCAGCC CATCACCCTC CAGGGCCCAA CTTGGAGCCC
65651 CCAGGACCTC CCCGTGCCCT GCCTGATGTC CCGCCTGTCC CCACAGAGCC
65701 TCACTTGGTC ACCACCCAGT CCTGGCCCTT GCTTACTGTG GCTGCACCCC
65751 GAGGTGTCCT CAGGGTCTAG CAGGTGGCTG CCCAGACATG GAGGTAGAGG
65801 AAGGAGTGGG TGGGGATGGG CTTGTCCTCC CAGGCCTCCC TGCCTGTCCT
65851 GCTGGCCACA GCCTTGGCTT GCCCAGGAGA AGCCCATGGG CCACACATCC
65901 CACTGCCAAT CCCACAGCGT CCTTTCTCGG GAACACCGTG GGGAAAGCTG
65951 TGGCACCAGC TCCTTCCTTT TGCAACTCTG ATGAATCTCA CCCAGGGATT
66001 TCAAGGCCCC TGGTCACACC AGGATCATAG GCCTCCCCCA TCCCCTGGAC
66051 ACACAGAGAC ACACCTGGAT TCAGGTCAGG CCTCGCCCAC TCTCGGCTAT
66101 ATTTCTCCCC AAGCCGTGTG TCCTCAGCTG TAGAATCAGG ACCATAAGGA
66151 AGTTCCCTCA TAGGGTTCTT GTCAGGACGG CACGATTTAC GTAGGGGATG
66201 CTGACACCGT GCCTGGCACG TGGGACGCAC TCCACCCGCG GCAGCCGCTC
66251 CCATGGCTTC TCAGTGAGTT TTCCAGCCAC ACTGCACTTC TTAGACAGGA
66301 ACACTCCATA CGATGTCCCT GTCCTGCACT GGATGGCCCA AAAATCTGAA
66351 ATAAGAGGAG GAGTGCGTGT GAAGCTCCCA GTGGAGCGTT TGGCACCTGT
66401 CCAGCATGTC CCCAAGGGCA AGTCACGGCT CTGAGATTCA GTGTCTCCTT
66451 CTGCAAAATG GGCCAATAGT GGTTCCTCCC TCCCAGGGCT GAAGTGAGGA
66501 TGAAATGGGA TAATCCACCC CCGTCCCCAC ACCCTGCAGG TCATCATCAT
66551 TGCTAGCAGT TGTGTGGTGG AGCAGGTGCT CTTGAGGGAG CGACACCTCC
66601 AGGTGCTCCC CTGCCCTGCT GGCCCCTCTG CAGGGAGGTG ACACCCAGGC
66651 CCCTTCCCCT GGGGCAGCCA GCTCACGCCC GTCTCTCTCC CACAGGTGCC
66701 GCTGCAGGTT CCTTTGGCAG TAAGTACACG CCTGGGGAGG GTGGCCAGGG
66751 CCCCCACTGC ACGAGCCTCT TTGCATGTCC TGGAAAAAGC TGGAGAGAAA
66801 AAAGGGGCTT CAGTGTCCCC TCTGGGACTT GGGCCTATTC ACTCCCTCCT
66851 CTAATTACAC CCCATCTGCT TCTCCACCTC TCCCCCCTCC ACCTCCCCCC
66901 CTCCACCCAT CCCCACTTCA CATCATATGC CATGTGTCAT GTGTCATTTT
66951 GCTGTGGCCT GTGGCCCAGC AACTCTCAGG CTCTCCCAGG AGCTCCATCA
67001 GTGCTGCTTT GGAAAACGGG ACAGGACTTT TTGCAGGTCT CTTGGCCCCT
67051 GGGTGGGCTC CCTGCTCCTC CTGCCACCCA CGCCACTTCT CTCACCTGGA
67101 TCTGGGAGAG CAGTCTCTCC TGCCAGTCAA GAGTGGGGTG ACCTTCCCCC
67151 ACCAGGGGCA GAATCCACCC CCTAGCCTAA CCATGGGGGC AGCCTCCCTC
67201 TGGGCAGCCT CTGCAGCCAG CTTGTCCCAG GGCTCTGCTC GTCCAGGTCA
67251 GCTCAGGTCC CAGGGGAGTC GGACCAGGGA GGGGCATCTG CAGGAGGTGG
67301 GGGTCCTGAG AGTTCCCCAG GAGGGCGAGG GCGACATGGC GCCCACAGGT
67351 TATCAGTAAA TGTCATCGAG ACTGTCCCCA GACACTCACA GGGTGCCAGG
67401 CACGGTCTCT CCTTTCAGCC TTGCAAACCC CTCCCCCTGG GAGGTCGCCA

FIGURE 3

```
67451 TCTGCTCTGC GAGGCAGCAG GAGAGGACTG GCCAATGTCA AAGAGCCAGC
67501 CGGGAGCAGA CCCCAAATCT CAGAGATGCT TCTGGGGTGC ACCGTCACCC
67551 TCCACCAGGG CTCTGTGGGG CCCCACATCC CACCCAAGTT GTCCCTCCCG
67601 GACCCAGGGG GCCCCTGCCT GGGAAGCCAG TGAGCCGAGA GGGCGCCAGA
67651 AAGAAGCTGG ACCCTGCAGG GACGCTGGTC TGCACAGCCG TCGTAAGTTG
67701 CTTCTCTGTG GTGTCCCCAC CCCGGCAACC CCCCAACCCT CTCTTGCTTT
67751 TCCCATCTCT CACCAGGCAT CAGCAGGTCC CAGAAAGACC CCGACCCCAA
67801 AGGCCCTGTG GCCACTGCGG CCACCACAGC CATGACAGGG GCCCCTACTA
67851 CTCCTGTCCC CTCCACGTCC ACTGCCTGGG CCCCCATGGC GCCCAGCACC
67901 CCACAGCCCA CAGGTGGGTG CCAGGGTACA GCGACCCCTG TCATCCCACC
67951 CTCTCCTGCT TCTAGCCTGG GTCCCTGCCT CTCTTGGGGT GGGAGGGTCG
68001 GCAGCCCTGG GCAGAGAGCA GGGGCTTGGC TCTTAGAATA GAGACGCTAG
68051 AACCCTAGAG CTGGGAGGCC ACAGGCCAAA GGGGCTTGAG GACACCTGGG
68101 TCAACCTGTT CCTGAGCCCA GCCAGGGGAT TCAGGGATCA GTTCAGCTTC
68151 CAAAGTCGTC TTCCTCCTGC CCTTCAAGCC ATTGCTTGGA AGGGCTCCCA
68201 GACCATTGTG GCCAGACGGC TGCAGGAACT GAGAGGAAAG GTGCTGGGGG
68251 CAGCGAGGCC ATCCTGACAT GCAGCCAAAG ACTGGCCTTA TCTCCCAATG
68301 GTGCTTCTGC CTCCGTGGTC CCTGGAGCCC CGCCCACACC CTGTCCCCAC
68351 CTGGCCCCCA GGGCCTCTCT GTCCTTAGCC CCTCAGCAGC ACACCGGTGG
68401 GATGGATGGA GCAGGGTTAG CCCAGAAAGC AAATGTCTCT GATCAGCAGG
68451 GCAAAGGGAG CCTCTGGAGC TGAGTTTGGA CACCGTGGGC TGCTGGGAAT
68501 GTGGAGGCTG TGTGTGTAGT GCAAGGCCAG GCCAGGGCCA GACGTCCTGC
68551 CCCCTCAGGG GTCTGCCACA GACAGGCATG GAAACCTGAT TCTCGCTCCC
68601 CTCCAACGGA GGGATTCACG TGTATTCAAG GCTGGGGGTG CTGGAGTGGG
68651 CCTCTGCTCT CACCTGGACT CACCTGGGGA GTATCCCTGC ACTCTGTGCA
68701 GTGCAGGTGC CAGGGGTCTG AAAGGATTTA TCCTTCCCAG AGGGCACCAG
68751 GAAGACGATG ACCAAGGGGA ATTCTTCCTG GTCCCAGCCA GGGAGGGGTG
68801 CTCCAATAGC CTGCCACACC CTGTCCCCCG CCACCCTGCA GGGAGGACCT
68851 GGTGGGGACT CCTGGCCCCT TGGGTAGTGC CCTGGCCCTC CATCTCTCTG
68901 ATCCAAGGAG ACCTGCCCCA CTGATCCTTC CCCCTTGGGG GGTGGCATTT
68951 CTAAAGGGCA GAGTCCCCTC CATCAGCTCC TGCCTCGGCC TGTTGCTGGG
69001 TGGACACTCA GGCTCCCCAG ACAGGGGCAA ATGCTGAGAG AAAGACCTCC
69051 TCCTTCCTAG GCCATCCAGA GCAGCTCCCC TGGGGGCAGC ACACCCCACC
69101 TCTTTCTACA TCCTTCCTTT TCTGCAGGAG GCATTTACAG GAGGCAGGGG
69151 CTAGCCAAAA GATTGGAGGA TTTCCGGGAA GCCTCCTGAC CCAGGAATCC
69201 TCTTTGGGGT GGAAGACATG GGTCACTCTG AGAATTCTGG ACTTCAGACA
69251 TAGGTTGGCC CAGCCACAAG GGACCTGTGC TTTGCTGATG AGCCTGTGGT
69301 GGGCAGACAG AAGCAAAAAC AGTGGTGGTG GGTGCTGTGC CTGTCTCCAA
69351 ACAGGGGTTT GGCTGGGAGG CCAGATACTC TCCATATCAC ATGTGCAAGT
69401 GCACACATGC ACACACACAC ATGCATGCAC ACACACAGGC ATGCACACGC
69451 ACATGTACAC ACACACACAC ACAGAGGAAT CCATTTGCAG AGCTGCTTCT
69501 GACTTGGTGC CAGGGCAGCC GTGGGAGGCT GGGCAGATTG TGCAAAGTTG
69551 GGAATTAAAG AGGAAAAGTC AGAGGCCAGA GTGGGAAATG CAGGGGAGTT
69601 GAGGGTCCCC AGGACCCCCT CAGTGAGCAG AAGGCACACC CTCCCTCTCG
69651 GCAAGACAGT GCTGCTCTGC ACCCTCAGCC CTGTATCAAG AAGCAGGACA
69701 TTAGGGGAGG AGGTGGCTCC AATGTGACAG CCAGTGGCCC CTACAGCCCA
69751 CATCTAGGGG CTCCTCCCTC CTCTTCAGCA ACTGAAGCCC CTGTCCAGAG
69801 CCCCCATTAA TGAAAACGAT CATTGCAGTA GCTGAGGGTG AGTTCTCCTG
69851 GGCTGTGCTC GTATCATTGT ATCATCATAT CATTGTATTC TGGGCTCACA
69901 GCTCCGTGAG ATGGAGGCTG TTATTTTCCT AGTCCCACAG GTGAGGGGAT
69951 CGAGGCTTAG GAAGAAGCAG CTGGATTTTA TGATATGTAA ATTACACCTC
70001 AATCAAGCTG TTTCAGAAGA AAAAAGGGGC AGCTGCTCAA GGTCTCAGAA
70051 TTATGGAGAG GCACGGGCAG GATTTGAACT CAGGGCTCGC CAACTCAGCC
70101 ACCCAAAGCT ATTGTCCTGA GGCCTCCAGG GGCTATGAGG TAGAGCTATC
70151 TTTTTTTTTT TTTTTTGAGA TGGAGTTTCG CTCTTGTCGC TGAGGCTGGA
70201 GTGCAATGGA GCAATCTCAG CTCACTGCAA CCTCCGCCCC CCCAGGTTCA
70251 AGCAATTCTC CTGCCTCAGC CTCCCGAGTA GCTGGGATTA CAGGCACCTG
70301 TCACCATGTT CAGCTACTTT TTGTCTTTTT AGAGAGACAG GGTTTCACCA
70351 TGTTGGTCAG GCTGGTGTTG AACTCCTGAC CTCAAGTGAT CCACCCGCCT
70401 CAGCCTCCCA AAGTGCTGGG ATTCCAGGCG TGAGCCACCG CACCCGGCCA
70451 AGTAGTGCTG TCTCCAAGGC CTGGCTTGCA GGGCTTCCCA GTTCCAAAGG
70501 AGCAGACCGG GCTTCCATGG GGCCTTGGCA CAGCACACAG GCCATGGCGA
70551 GAACTTGCTT CCCACACACC TGAGTGTGTC CCTGGGCAGC CAAGCCAGGA
70601 CTCCCTCCCT CCCCAAGACC CTGGTCCCTG AAAGATCCTG AATACCCCCG
70651 AGTGCCTCCC AACAGGTGCT TCGGGCTCTT TGAACAGAGT CCAGCTGGGC
70701 CTCTGAACTC CTGGGCCAGA TGTTTCTCCC GCCTGCCAAT GTCAAGCTGT
70751 CTGGAGGACA GCGCTGCGGG CGGAAAACGC CGCTGGAGAC ACTAATCCTT
70801 TCCTGGGCTG GGCCACGGAG GATGGAGGGA GACAGGCTCT GAAGCAAATG
70851 CCTTCAGGGC TGGCTTTCTC ATGGCTCTAA TTAAGCCCTT GCCAATTTGG
70901 GCCTGGCGGC CTCATCTTCC CACTGAACAT CATATTAAAG TCAATTCATG
70951 TCCAAAGCTC CCCGCTCCCA GCTGGAAAGT CTTCCGCACT TGTTAGCTGG
```

FIGURE 3

```
71001 TAGCTTTTCC TTTTCTTTCC CCACAGCCAC CGTTGTGTAT AATCCCTTCA
71051 AGAAGCGGAA AACAGCAGCG CTCCCCTGTC CCTCTGGGTT TGTCCTTTGA
71101 AATTTGGGCA CAGGGCAGTT CTTTGCCAGC CCTGCCTGCC TGCCTTGCTG
71151 GCTGTGTGTC CCGTTAGTCT ACGGGCTGAG CGTTGTGTCA TTGGTTCATG
71201 CTGGGGTCCC TGGTGAAAAT GGGCCAGGCC AGGGGTCAGG AAGGTAGAAG
71251 GGCAGTGATC AGGGAAGCAG GTCAGATGCT GGGGAAGGCT CCGGTCCCTG
71301 GATTGCGGCT GGACAGGAAG GACACCTTCC AGGACACTTC TGGACACATG
71351 TAAGATCTTG GCCGGAACAC ATGTCCCACT TCGCAGCCAT TAGCCAGAGA
71401 CATCAGCTCA GAGAGGTCTG GGCCCAGAGG CGGGACCTGG TCTAGCTCTG
71451 TCCTTCAGTC AGAACGGGGA CGGCACAGGG AGTGTAGAAG GGTCTCGCTG
71501 AAGAATATGC AGATTCTCAG GCATGGGTTC ACCTCTCATC TATCGGGCTT
71551 TAAGTCTGCA TGTGCCCTCC ACAGGCTGAA ATAGTGTAGA TGCTGCCTAT
71601 GTAGTAGATT TGGACCCAAT TCCTTTGGCC AGTGTAGACA GAGCCTCTCC
71651 TTATAGTGCT GCTGCTTCTA AGGGGCCTGT GGGGTGCGGG GCTGTGATGC
71701 CTCAGTATGT ACCCAGCTTC CCTCAGCACC ACCCCCTCGC ATAACTTGGT
71751 TTCTTCTCTT CTTCCCCCCA AGAGTGGACC AGGCCATCTA CGGCTGCCCC
71801 TCTCTCGAGC AGGTGGTCCC AGGTGGCCTC CCGTGCAGAA GGTATGGGGG
71851 GGCAAGGCCT GTGATGGGCC TGAGACCCCG GGGAAGCGCC CTCTTAGACT
71901 CGTAGGCCCC TCCCTCTGTA GTGGAAGTAG CAGGTGTGCA TGGTGGGGAC
71951 CTGAGGTTGG AGGGGGGCCG CAGGAACCAA CTGAGGGCAC GGGTGTAGAA
72001 TGTCGGTGCC TGGGGAGCTC TAGGGCACAG TGGTGAGGGA GCGGCCTGGT
72051 AGAGCAGGTC TACCAGCTCT GCCCCCAAGC TCACCTGCTT CAAGAGGTTC
72101 CATGTGGCAC CCCCACGCCA AGCCCTTCCA CCAGCACTCC CTCCGAGGGC
72151 TTCGGAGTCT GGTAGAGGCC CCGCCTCCCA CGACAGGAAC CCCCCTCTCC
72201 AGCTGCCCTT GCTCACAGGA CACCTGGGCA GTTGCTGGAT CAGAGAGTCA
72251 GAGGGGGCTT CCTGCAGGAG CGGGGGCCAT GAGACCTCGG AGGGTGGACT
72301 GTGGTGGGTG AAGGGAGAAG GCAGCACATT CCAGGCCGCA GGGCCAGCCG
72351 GGGCAAAGGC TTGGCAGTGG GATGGCAGGG AGCCTGACAA AGTGGAAAAT
72401 GTGTGGGTTA AAGGAGGGAG GGCGGGGTCC TGGAAGACAC TGACATCCTC
72451 CTGCTACGTG GGAGGAGACA CAGGGCTCAT CTGTAGCCAT AGACAGACAT
72501 GCCAAGGAAA CGCGCAGGCC TGCCCGACTC TCCAGAAGGG AAATTGTCCC
72551 TGGCCCCAGC TCACCAAGCC TGGGTGGGGA ATTAGGGCCT GAGGTCTAGG
72601 GAACAGGTGA GCTGTTCCTT CCAGCTCACA TGTTCAAATT TCCTCCAGCC
72651 CCAGCTCTGA GCAGCGAGCA GGGCTTTGAG CGCCCTCTAC TGGCAGGAAG
72701 CTCTGGCGCT GGAAGCATGT TTAGAGAGGG TCTGAGGCTC GGTTCCTAGA
72751 AACCTGGAGG ACCTGGGCCT GGTGTCCTCT GTGGTGATGG AGACAGAGCT
72801 GGCGGGAGCC ATCGCTTCCC TACCCTGGGC CAACCAGGGC ACCACAGACC
72851 CCCAGAGGGA AGCCAAGGTA GTGACGATCC CGGGACAGTG GCCTGCTCAC
72901 CCACAGATAG GGCGTTGGGG TCCCAGCGGG ATTCTGGGCA GTGGAAGGCA
72951 GGTGCCGTCC GTGTTCCTGG CTTGACAGCA CTTGCGAGTG GGACTCCAGG
73001 GACAGCGAAG GATTCACTTC GGCTGGAGCA GGAAGAGTGT TTCAGAAAGG
73051 AAGGGAGATG CCAAAGTCCT TAAATGCCAA GTTTAGTCTC TGGGTTTGAT
73101 GCTCCAGGAA GTTTGGAGAG GCGGTGGGGA GAGCAAGAGA CGGGCGTGGT
73151 GTGCAATGTG ATGTCAATCT ATCTAAAAAC AGTTTGGCTT CCAAGAAGGT
73201 CTTAGCAGGG CGCGGGGGTG TCAGGGGTTA CAGAAGTCAT TTGGAGGATT
73251 AATCCAGCCA GATGTGTCCA TGGTCTCAGA GAGGGGACCA AGGGCAGGGC
73301 TGATTTGCAA GCTTGGGATG TGCTGTGTTT CCTTCAGGAA GGGGCCCCAC
73351 CTCCCTGGGC TCTTCGAGGA GAGGGGCTGT GTGATTTGAG GCCAGAGGGG
73401 CCTCTCCCTC CCTCACATCT GAGCAGGCGA CAAGGCTGCC TGCCCTAGAG
73451 CTGGCCCAGG GCGGCTCGGA AGCCTTTGCT GGGCTCTTCC CTGGGCAGTG
73501 GGACCATGAC AGACGAAAGA ACCTGTTTCT CATCTCTCCA AGCTGTGGGC
73551 ACCCCTGCCG CTGCCCCTGC CCCTGCCAAG GGCTACAAAC TTTTCCAGCT
73601 CAAGCCCAAA TCTCCTCAAG TGATGCCTAT TGAAGAATTC CAAGGTAAGA
73651 GGATGGACCT GGGGCCCCAT CAGCCCTCCC TGACACCTGT TCCCCATCCG
73701 CCGCTGGAAA AAGACGGTGC AGGATAGAGG ACCGATGCCT GGCTCCGAAA
73751 ACCCTCCTGG AGTAGCTGGG TCAAGGTTAA ACTGAGTCTC TCTTCCCTAC
73801 AGGCCTCCCT CCCCAAGGGA GCTGGGAGCA GGTATGAGTC AGAAGCCAAC
73851 TTGGGCACAG TGGGCAGGCC ACACAGCAGG CAGAGCAGAT GCCAGAAATA
73901 GCCCATCCCG GCTCCCCTGG GAGGTGTGGC CCTGGGGCTC GTGTTGGTTG
73951 AAGCAGAATC TGGGACACAC GGGTCACCGA TGCTGCTCTT TGGGACACTT
74001 AGAGGATGCC TCATCTCCTC ATTATCTCTG GAGGGACAAA GTGAAGGGGG
74051 CAGGACTAGG TGGCCCACAG GTGGGAGTGC CCACCATCTC TCCTGGGCAC
74101 AGGCTGTTTC TCTAGTCTCC CATGCCCTTG ACCACTGGGT CAGTCCCTCA
74151 TCCCATCACA AAAGGGAAGC TGGGTCCTCT AGAGATACAC AGATGGTGTT
74201 TCAAGAGGGT GGCCGTTGTC CTTCCTTGTT CGGGGGCAGC CACATTGGCT
74251 TTCTTGCTGG AGGGTGGGTG GGTGGGTGAG TACTGTGTCC CTTCGTAGGA
74301 ACATCAAGGG ATGCCCCCCC ATTCTTAGGG ATGGTGACCT TCCTCACCAA
74351 ATCCTCCATT GACAATGTGG GATTCACCTC CAATCCCTGA GAGCCTTGCC
74401 CCAGGCAGTC ACGGGCTTGT CTGGTCCTTG GAGCGGAGCT GGTTAGGCAG
74451 GGGTCAGCCT GAGAACCACG TAGGGGTGGG GTGCAGGAGG CGGCAGGACA
74501 TGGTGGTGGT GGTCCTTGGT ATGAAACCAT GTGCTTCCAG GAGCAGCGAG
```

FIGURE 3

```
74551 TCAGAAGCCG GGCCAGGACC AGGGGGAGGC ATGCAGGTTC CCAGGGCTCC
74601 TGCTTTAAAG TGGCACTCAC TCTTAGCATC CTGCAAAACA ATCAAACTTG
74651 CAGAAAGCTC AGGCTAATAA GAAAGGGTCT GGCAGGTGGG CGTTTTCCTC
74701 CCAGCCATCT TCCAAAGCAG CATGGGCAGG AGCTCCTGGC CCATTGCATC
74751 TTGTCCAGCG TCCATCCATG CATTCATCTA CCCGAGGATA CCACGGCGAG
74801 CGCCGTGAAC CCAGGCGTCG CCCTCCCCCA GTGCACAGCC AGGTGGCATG
74851 ACCCTGCCCT CCTTGCATGA ATCACTTTCT AATCACCCCG GCATGTGGGC
74901 ATCCTTCAGC GAGCGCTTGG CCCTGGTGCC CAGCCAGGCA TTAGCAGGAG
74951 CTGCCCACTG GCCCTCCCTG GTTCCCTGCC CACAGGGCCA GGTGGGAATC
75001 CCTGGGCTCA GCCTACTCAG GTTCTCCTCT GGGCTCAAAG CAGGGAGGCC
75051 TCTCTCTTCC TGAATCCGAT GGAAGGGTGG GAGGCCTAGG GCACCTTCCG
75101 GTACCTTTTC CAAAGATGCC TTCCTCCGTC CCTGCATGAC CTGGGGTGAG
75151 TCCTTCCTCG CCCTGTCCCT CAGTTTCCCT GAATGCTCGC TGACCATTGG
75201 TATTTCTCCC ACTTGGCCGG CCCAGACTGC GAATGCTACG GTCACTCCAA
75251 CCGCTGCAGC TACATTGACT TCCTGAATGT GGTGACCTGC GTCAGCTGCA
75301 AGCACAACAC GCGAGGTCAG CACTGCCAGC ACTGCCGGCT GGGCTACTAC
75351 CGCAACGGCT CGGCAGAGCT GGATGATGAG AACGTCTGCA TTGGTGAGAG
75401 GGCACGGACA CGGCACAGGG AACTTGCTGG AATGCGTGCA GGGTGCACTG
75451 CCCTGCGAGG TGGCCTCTGG GGCCCCCTGC ATCAGAATCA CCTGGGGAGA
75501 CTGTGGGAAT TCTAACTCCA GGGCCCTCTC CAGTTGAGCA TCTCTAAGGA
75551 CAGAAAGCTC CAGAAACTGC TCTATTAGTA ACCTACCCTT GCGGTTCTCC
75601 GGTAAGTTTT GCACTGGAGT TGCAAAACTT ACCAGTGGCC CTTCCCTCTC
75651 TGGGCAACTG GAGGGGACAC TGACCCTTCC TGGCTCCAAA GAGCTGTGAC
75701 TCTGGCAGGT GGCAGGCACT CAGTGGCAGA GGCCACTGAG CATCTGTCTG
75751 GGGCTGGTGT GTGGGGCGGT CCCCCTCCAT AGCTCCTTTC CAGAAAGGTG
75801 GAGGAGCAGC CTATCCCTCC TCCTGCAGGG GCCCAGTTGG GGGCCAAAAG
75851 ATCGCCTTGC TGCGTGCATT TGTGCAAGTC CCTTCCCGTT CCTGGGCCTC
75901 AGCTTCCTCA TTCATCAAAT TGGGAGGCAG ATCAGATCAA AGGTTTTCAG
75951 CTCTTTTTTG TGGCTGAAGC TTTTCTTCAA ATGCTTTACC AGCCCAGGTC
76001 CAGCTATAAA GCTGCTCTTC ACCCCTGGTG GGCACCCAGT CTGCTTTCTT
76051 CCAAGTTCTA CTCAAGGACT GGCTTTTGGG TAGAGAAGGA AGTCCATCAG
76101 GGCCCTGGGC CTGGGCAAAG ACCAAAGCCA TGACCGCCAA CCAAAACGCA
76151 CCAGCCTGGA ATGGTTGCCC CTGTCGTCAG TAGAGGCCAG GTCTCGGCCT
76201 CAGGGGCTGT CCCCCAACCC TGCCCAGCCA GGCCCCTTGG GACACCATCA
76251 CCCATCCCCC ACCCAGCAGG AGGCTCTGGC TGCCCAGAGG AGGGGCTCCT
76301 GCAAAGCTGG AGCTGTCGGT CTGAATTCTG GCGGCAGCCT TCAGATAATT
76351 CCATCAACTC TAAGTGATCA AAGCCGCTGA CGTCACAGGG GGCCAGCTGC
76401 AGGGACAGGG CAGGGCCTTT GGATCCAATT AGAGGTGCCC ACACCCTGGC
76451 ACCCTCCTCC TCTCCCTGGC TCTCCCTGCC TCCACCCCGA GAGCCAGCAC
76501 TGAGCTGCAA GGTTTCTCAG GGTGGACGAT ATTCACCCTC TCCCACAGAG
76551 CCCCAAGGCA ACCAACTGGG CCCACCCCGG GAGCAGGAAT AGGCTGTTCC
76601 TCCACTCCCC TGCAAAGGAG CTATGGAGGG GGGCCACCCC ACAACACAGC
76651 AGCCCCAGAC ATGCTCAGTG GCCTCTGCTG AGTTTCTGCC ACCTGTCGGA
76701 GTCATAGCTC TTTGGAGATG GGAAGGACAG CGACCCCTCT AGTTGCCCAG
76751 AGAGGGGAAG GGGCTGACCC AGGCCACACC AGTGCCAGGG CGGGGAAGGT
76801 GGGGCTGGGA CGTGTTTGAT CCCAAGGAAG GAAGCCAGAG TCTTCTCTCC
76851 AGGCCTGGCC ACCCTGGGAA GTCCCCACCT GCCGTCCAGC CGCGGGCTCA
76901 CGTGGACCCA GTGTGGGGAG CATCCCCTGG GGAGTGTGGA GATGCTCCCT
76951 GCGAGGCCGG GAGAGTGGGG GTCCGAGAAG ACGGCGCCCA CACGTAGCCC
77001 TGACCGCGCG CCCGTGCCCG TGTCCGTCCA GAGTGTAACT GCAACCAGAT
77051 AGGCTCCGTG CACGACCGGT GCAACGAGAC CGGCTTCTGC GAGTGCCGCG
77101 AGGGCGCGGC GGGCCCCAAG TGCGACGACT GCCTCCCCAC GCACTACTGG
77151 CGCCAGGGCT GCTACCGTGA GTGCGCGCCG TCCCCCGTGG GCGGGGCCTG
77201 CGGAAAGGGG ACGGGGCAGG ACCGAGGCAG TGGGCGGGGC CTAGTGGGAC
77251 GGGGCAGGGG CGGTGGACTG GGCCTAGCAA GACGGGGCAG GGCCGGGGAA
77301 GTGGGTGGGG CCTAGTGGGA CGGGGAAGAG GCGGTGGGCG GGGCTCGCGA
77351 GACGGGGCAG GGCCGGGGCA GTGGGTGGGG CCTAGATGAG AGCGGGGCAG
77401 GGTTGGGATA GTTGGCAGGG GCCTGGTGAG ATGGGGCCGA CCCGGGGGCG
77451 GTGGACGGGG CCTAGCGAGA CGGAGCTGGC AGGTGGGCGG GGACAGGATG
77501 CTGCTGAGGT CCGGGGCCGG GCCGAGGGGC GGGTCCAAGA GCTCGGGGCG
77551 GGGCCTGATG CGACCTGAGG CACGGTGGTG CCTGGTGGGA ACTACGAGAA
77601 AGACCGAGCT GGGGTTGGTT GGAAAGGTAT TTGCGGGGAC AGAGGGAGGG
77651 AGGCTGTCCA AGTCGGCGTT AGCCGCGGGC ACAGGGTGAA AGGAGGCTCC
77701 AGGCGCGTGG AACAGCACGT GCACAGCTCT GGAGACTGCA GGCGCGTCTG
77751 AAGAACAGCA CCGAGGCCAG TGGGGCGGGG AGAGAGGGGC AGCGGTGGGA
77801 GGCAGCCGGG GGCCAGTATC TCGCCCGGGC GCCGTCACCC TCCGAGGGGG
77851 GACGTTTCGC ACCCAGCGCG CCTGGAGCCT CCTACATCCC CGGCCCAGAC
77901 GGCGCCCCCG GGATCTCGCA CACCCTGCTT CGCAGGAGCT CGGAGGTTGG
77951 CGGGGGGACC GGGCCACCCC CCGTGCTGAC CGCCCCCTCC GCCTGCAGCC
78001 AACGTGTGCG ACGACGACCA GCTGCTGTGC CAGAACGGAG GCACCTGCCT
78051 GCAGAACCAG CGCTGCGCCT GCCCGCGCGG CTACACCGGC GTGCGCTGCG
```

FIGURE 3

```
78101 AGCAGCCCCG CTGCGACCCC GCCGACGATG ACGGCGGTCT GGACTGCGAC
78151 CGCGCGCCCG GGGCCGCCCC GCGCCCCGCC ACCCTGCTCG GCTGCCTGCT
78201 GCTGCTGGGG CTGGCCGCCC GCCTGGGCCG CTGAGCCCCG CCCGGAGGAC
78251 GCTCCCCGCA CCCGGGAGCC GGGGGTCCCG GGGTCCCGGG GCGGGGCCGG
78301 CGTCCGAGGC CGGGCGGTGA GAAGGGTGCG GCCCGAGGTG CTCCCAGGTG
78351 CTACTCAGCA GGGCCCCCCG CCCGGCCCGC GCTCCCGCCC GCACTGCCCT
78401 CCCCCCGCAG CAGGGGCGCC TTGGGACTCC GGTCCCCGCG CCTGCGATTT
78451 GGTTTCGTTT TTCTTTTGTA TTATCCGCCG CCCAGTTCCT TTTTTGTCTT
78501 TCTCTCTCTC TCTTTTTTTT TTTTTTTTTC TGGCGGTGAG CCAGAGGGTC
78551 GGGAGAAACG CTGCTCGCCC CACACCCCGT CCTGCCTCCC ACCACACTTA
78601 CACACACGGG ACTGTGGCCG ACACCCCCTG GCCTGTGCCA GGCTCACGGG
78651 CGGCGGCGGA CCCCGACCTC CAGTTGCCTA CAATTCCAGT CGCTGACTTG
78701 GTCCTGTTTT CTATTCTTTA TTTTTCCTGC AACCCACCAG ACCCCAGGCC
78751 TCACCGGAGG CCCGGTGACC ACGGAACTCA CCGTCTGGGG GAGGAGGAGA
78801 GAAGGAAGGG GTGGGGGGCC TGGAAACTTC GTTCTGTAGA GAACTATTTT
78851 TGTTTGTATT CACTGTCCCC TGCAAGGGGG ACGGGGCGGG AGCACTGGTC
78901 ACCGCGGGGG CCGATGGTGG ACAATCCGAG GAGTAAAGAG TTTGCTCACT
78951 GCTGCCTCCA CGGCCTGTTT TCTTTCTGTG TTGGGGACGG TGGGCAGGTG
79001 TGGGGCTTAC AGAGGAATCC ACAACACAGC CTTAAAGAAA CGGTTTCCCT
79051 ACTGGGGCCA CCATTTCCCT GGGCCTTTCT GTGGATTCCA GCAGGCAGTG
79101 CCCCCTCCCC GCAGGCTTGG CTGGCAGAGT TTTCCACCCC GCGGCCAGGC
79151 TGCAGGTGCC CCACCTGTTA GGAGCCTCCC CACACTGAAA GGCTGCCTCC
79201 CTCCTTTCCC AAAAAAGAAA TCCGGAGTGT ATTGGCCCTT TTCTACAGAA
79251 GTCCAAGGGA AATGACTCAG GGAGAATCCT AGCAGAGGTT GAATCCAATG
79301 CTCTGATTTA TACTGTGTCT CGGTGGCCAC CTCCGATGGA TGTGTCATCT
79351 CAGACCTGTT GCAGCCGGAG CCTCAAGTCC AATATCAGAT GAAGCTGAAC
79401 CCACAATTCG GCCACCGCCT CCTTCCAGAG TTTCAGATGG CCAGGTGGGC
79451 AGAGGCGGGC AGTGCAGAGA CCCCAGACGT GCCGGCCCTG TCCTCCCTAC
79501 CTTCTCAAGA TTAGGAAGGG GTGCTGGAGG GGACAGGGCC AGCTTGGGAG
79551 TGGTGAGGAA GCTCCTAGAT TCGGGGCTCA TCCCCTGGGG CCTCTGATTC
79601 AGAGGATCCA GCAGTTCTCC CATCTCCGCT TGGTGTCTCC AGCCCTGGGG
79651 CCACACTTCC CCCTCGGTCC AGCCTCCTGT CCACCTATGT TTATTTCAGA
79701 GCAGTGCCGG GGGTCCGGTC CTGGTTGCTA ACTGCTGCCA CTGCTCCACC
79751 TGCAGGTGCT CCCAGCACTG GCTTCTGCCA CCACACCTGT TCTTTCCCAG
79801 CTGCGAGGTT TAGACCTGGG TCCTTCCCTT GAGTCCCCAA AGCTAAGCTA
79851 AGACCAAGTG GAACAAACTT GGCCTTGGGG ACAGCAGGAG ATTACAACAC
79901 AGAAAAGGAG GGGGAGGCAC AACGGGACAC TGCATAGGAC TCACAGTGTC
79951 CCGAGCCCAC ACACCAGCCC CTCCTGGCCT CCTCTCTCTG CTCCCACCCC
80001 CAGCACCCTG CTGACCCGGA AGTGCCTTCC GACAGGCCCT GCATCCTCCT
80051 GCAGCTGGCC CATCTCTACC CTCACATTCT TCCTTACGCA CAGAACCCCA
80101 CTGCCTGGGA CCCAAAGTGC CCAGATAAAA TAAAACACCT CTCTGGGGCT
80151 TCTTGTGGAT AGGAGTGGCC AGGGGACACA GCTCTGGGCA GTGAGATGTA
80201 AGCAGGACTG ATGGGTGGGC TTCCAGAAAG TTCTTAAAAG TCATCCCTTC
80251 CTCCACGCCC CACAAAGCCT CAGTTGTCCA AGTTTGTGGC TCTGGTTTTT
80301 CTCTCTCATC CCTCCCTCTG CTCTTCAGTC AGCAGGATGG GGAAGGAGCC
80351 TTCTTGAGAC ACTGAGCATA ACAGTGTCAC CCTCAGGATA GTGAAGGGTG
80401 GAGCCGGGGG TGGAAGTTGC AGAGCCTGGG ACACCTGCCT TGGCTGCAAC
80451 CTCTGCACTG CTTTTATTTA TCTACTTATT TTTATAGAGT TGAGGTCTCA
80501 CTATGTTGCC CAGGCTGGTC TCAAACTCCT AGGCTCAAGT GAGCCTCCGG
80551 CCTTGGCCTC CCGAAGTGCT GAGATTACAG GCATGAGCCA CCGCACCACC
80601 TGCCTGTCTT ACTGTGTGAG AAAAAAAATA AACGGTGATG TGATTAAAAC
80651 ACTAGAATTT CGGTTTTCTG TTCCATGCAC ACTTTCCAAG TTCTTGGCCC
80701 TGCCTGATTG GCATCCGGGC CCCCTGAGCC CTCTAAGGCC AGAGAGAGAC
80751 TGAAACCCAC CCTTCCTTGG CCGCAGCCCC CCATGGTGAC CCTCAAGTCA
80801 CCAGGAAGAG GAATGTCCAT GACTCAGGCC GAGGAGTCCA CCTCTGTCCA
80851 GAACAGGACT GCCCCTGCCT TCTAGTGAGC ACAGGGCAGC TGAGCAGAAC
80901 TCACCCCAGA GAAACAGCAG CGCTGGCCGC AGTCGGTAAA CAGCAGGCAT
80951 TTCCTCCTA (SEQ ID NO:3)
```

**FEATURES:**

| | |
|---|---|
| Start: | 3000 |
| Exon: | 3000-3212 |
| Intron: | 3213-34124 |
| Exon: | 34125-34768 |
| Intron: | 34769-62971 |
| Exon: | 62972-63144 |
| Intron: | 63145-66695 |
| Exon: | 66696-66719 |
| Intron: | 66720-75225 |
| Exon: | 75226-75393 |
| Intron: | 75394-77031 |

FIGURE 3

Exon: 77032-77166
Intron: 77167-77998
Exon: 77999-78231
Stop: 78232

**Allelic Variants (SNPs):**

| DNA Position | Major | Minor | Domain | Protein Position | Major | Minor |
|---|---|---|---|---|---|---|
| 337 | T | C | Beyond ORF(5') | | | |
| 389 | A | T | Beyond ORF(5') | | | |
| 817 | G | T | Beyond ORF(5') | | | |
| 874 | T | C | Beyond ORF(5') | | | |
| 3666 | A | G | Intron | | | |
| 4757 | G | A | Intron | | | |
| 4816 | G | A | Intron | | | |
| 5640 | - | A C | Intron | | | |
| 6535 | A | G | Intron | | | |
| 7503 | C | G | Intron | | | |
| 9001 | G | A | Intron | | | |
| 9617 | C | T | Intron | | | |
| 9672 | T | C | Intron | | | |
| 9761 | T | C | Intron | | | |
| 9811 | C | T | Intron | | | |
| 11617 | A | G | Intron | | | |
| 11835 | G | A | Intron | | | |
| 12837 | C | A | Intron | | | |
| 14651 | C | T | Intron | | | |
| 15701 | C | T | Intron | | | |
| 16587 | G | A | Intron | | | |
| 16710 | C | A | Intron | | | |
| 17888 | T | C G | Intron | | | |
| 19282 | G | A | Intron | | | |
| 19429 | A | G | Intron | | | |
| 19561 | G | C | Intron | | | |
| 19575 | G | A | Intron | | | |
| 19634 | G | A | Intron | | | |
| 20383 | C | G | Intron | | | |
| 22878 | T | A | Intron | | | |
| 22993 | - | T | Intron | | | |
| 23852 | - | A | Intron | | | |
| 23853 | - | G A | Intron | | | |
| 23888 | G | A | Intron | | | |
| 24151 | C | G | Intron | | | |
| 24984 | A | T | Intron | | | |
| 25681 | T | C G | Intron | | | |
| 25698 | - | G | Intron | | | |
| 25928 | A | T | Intron | | | |
| 31637 | T | - | Intron | | | |
| 36513 | C | T | Intron | | | |
| 38025 | A | C | Intron | | | |
| 38068 | A | G | Intron | | | |
| 42787 | A | G | Intron | | | |
| 43423 | A | T | Intron | | | |
| 43752 | A | G | Intron | | | |
| 44151 | - | T A | Intron | | | |
| 44443 | A | G | Intron | | | |
| 44644 | C | T | Intron | | | |
| 44888 | A | T | Intron | | | |
| 45250 | A | G | Intron | | | |
| 48665 | A | G | Intron | | | |
| 49661 | G | A | Intron | | | |
| 50012 | T | C | Intron | | | |
| 51203 | A | T | Intron | | | |
| 51222 | G | - | Intron | | | |
| 51230 | G | C | Intron | | | |
| 52864 | C | T | Intron | | | |
| 53973 | A | G | Intron | | | |
| 59408 | A | G | Intron | | | |
| 60143 | C | T | Intron | | | |

FIGURE 3

| | | | | | | |
|---|---|---|---|---|---|---|
| 60158 | C | G | Intron | | | |
| 60227 | T | C | Intron | | | |
| 62304 | G | A | Intron | | | |
| 63410 | T | C | Intron | | | |
| 64577 | G | C | Intron | | | |
| 65796 | A | G | Intron | | | |
| 65918 | C | T | Intron | | | |
| 66192 | T | C | Intron | | | |
| 66209 | G | A | Intron | | | |
| 66334 | T | G | Intron | | | |
| 66548 | C | G | Intron | | | |
| 68155 | G | A | Intron | | | |
| 69247 | G | A | Intron | | | |
| 69288 | A | C | Intron | | | |
| 70151 | - | T | Intron | | | |
| 70166 | - | T | Intron | | | |
| 70568 | A | G | Intron | | | |
| 70769 | G | A | Intron | | | |
| 71191 | T | C | Intron | | | |
| 71368 | G | C | Intron | | | |
| 71370 | G | C | Intron | | | |
| 71684 | T | G | Intron | | | |
| 73463 | G | C | Intron | | | |
| 73734 | G | A | Intron | | | |
| 75366 | G | C | Exon | 399 | E | Q |
| 75368 | G | A | Exon | 399 | E | E |
| 76076 | T | C | Intron | | | |
| 79643 | C | T | Beyond ORF(3') | | | |
| 80208 | C | G | Beyond ORF(3') | | | |

**Context:**

DNA
Position

337
CCTGAGGACTCTTAAAAGAAAATAAAGCACATTGATTCTATTTGTTTCTGGGAGCTGCAG
TTTCTTAATAATATCAGGTGAAGATAAATTTTCCACGGAGAAAACGATCCTCCGGGATGC
AGCTTCTTACTCTGAAAATTTCCCTGCCGACTCCTCACTCTCTGCGCTCCTCCTCGTTAT
CCGGGGACTCCTGCCTCTCTTCCCCCTTCTCTTTTTTCTTTTTGGCAGAACCCGCCTGCA
ATATTCGTGTGCTGAGCTCGTAATTCCCCCTGCGATGCCAGCAACGCCCAATTGATTGAC
[T,C]
AGTTGTAAACACATTTTTCCCCTGGCAGATTTTGTTGTTGTTAGGGTTTTTAAAATTTAT
TTATTTTCCAGGGAATGCGTGGCATTTAAACCAACAGGACTGCAATTAATAGATTTGCGA
GTTGCGCCGCGCGCGCCGCTCGCCCCAGCCTCCCGGCCTCCGGGCCTCGCTGCCTCCCCG
CGCCCGGCGGCGTCCAGCGCCCTGCAAGCCCCGAGCAGCCGCGGGTCCTGCAGCTGAAGG
AAGGTTGCAGCTGCGCCCTCCTTGCAAGCCGCAGCCCGGCGTCCTGGTTGTCCCAGCAGC

389
AGCTGCAGTTTCTTAATAATATCAGGTGAAGATAAATTTTCCACGGAGAAAACGATCCTC
CGGGATGCAGCTTCTTACTCTGAAAATTTCCCTGCCGACTCCTCACTCTCTGCGCTCCTC
CTCGTTATCCGGGGACTCCTGCCTCTCTTCCCCCTTCTCTTTTTTCTTTTTGGCAGAACC
CGCCTGCAATATTCGTGTGCTGAGCTCGTAATTCCCCCTGCGATGCCAGCAACGCCCAAT
TGATTGACTAGTTGTAAACACATTTTTCCCCTGGCAGATTTTGTTGTTGTTAGGGTTTTT
[A,T]
AAATTTATTTATTTTCCAGGGAATGCGTGGCATTTAAACCAACAGGACTGCAATTAATAG
ATTTGCGAGTTGCGCCGCGCGCGCCGCTCGCCCCAGCCTCCCGGCCTCCGGGCCTCGCTG
CCTCCCCGCGCCCGGCGGCGTCCAGCGCCCTGCAAGCCCCGAGCAGCCGCGGGTCCTGCA
GCTGAAGGAAGGTTGCAGCTGCGCCCTCCTTGCAAGCCGCAGCCCGGCGTCCTGGTTGTC
CCAGCAGCCAGGAGATCCCTACCTGTTAGTGAACAGTTAGGAGTCGACTGCTGGAAGAAT

817
GCGCCCGGCGGCGTCCAGCGCCCTGCAAGCCCCGAGCAGCCGCGGGTCCTGCAGCTGAAG
GAAGGTTGCAGCTGCGCCCTCCTTGCAAGCCGCAGCCCGGCGTCCTGGTTGTCCCAGCAG
CCAGGAGATCCCTACCTGTTAGTGAACAGTTAGGAGTCGACTGCTGGAAGAATTAATTAG
GAACGTGCTGTGCTCTGGGCAGCGCGAGCTCGGGTAGAGGCATCCAAACCTTTGCCGGCG
GCGCTATTTTATTTTTACTACATTTTCTCAGGTTGCAAAAATAGACACCGGGCACGTTCT
[G,T]
TCTTAGAGTTTTCTAGCAAGGAGCGCCTTCAAGGCCAGGCAGGCTCTGTAACAGGTTCCC
CTTTAAACAGCCAGAGGTGAGACGGGGAAAATGGTCCTGGCTGGGTTCTCGTTCATCTCC
ATCAGCAGTCCTTCACCCAGAGAGAGGGGCAGGGGTCGCCCTAACTCAGATGAATGAGTC
CCATGCCTGGAGCCCTGGGGCCCTGGCTGGGGGCTGCTCCGAGCCTCAGGTGCTCAGGGC
GCTCAGGGCAGCAAGTGTCCGCCACTTCGGTTTGTCATTTTTGGCAGGAGCGTTTTTCTG

FIGURE 3

```
874     AAGGAAGGTTGCAGCTGCGCCCTCCTTGCAAGCCGCAGCCCGGCGTCCTGGTTGTCCCAG
        CAGCCAGGAGATCCCTACCTGTTAGTGAACAGTTAGGAGTCGACTGCTGGAAGAATTAAT
        TAGGAACGTGCTGTGCTCTGGGCAGCGCGAGCTCGGGTAGAGGCATCCAAACCTTTGCCG
        GCGGCGCTATTTTATTTTTACTACATTTTCTCAGGTTGCAAAAATAGACACCGGGCACGT
        TCTTTCTTAGAGTTTTCTAGCAAGGAGCGCCTTCAAGGCCAGGCAGGCTCTGTAACAGGT
        [T,C]
        CCCCTTTAAACAGCCAGAGGTGAGACGGGGAAAATGGTCCTGGCTGGGTTCTCGTTCATC
        TCCATCAGCAGTCCTTCACCCAGAGAGAGGGGCAGGGGTCGCCCTAACTCAGATGAATGA
        GTCCCATGCCTGGAGCCCTGGGGCCCTGGCTGGGGGCTGCTCCGAGCCTGAGGTGCTCAG
        GGCGCTCAGGGCAGCAAGTGTCCGCCACTTCGGTTTGTCATTTTTGGCAGGAGCGTTTTT
        CTGTCTGGGTGGAGAATGGAGTTTCACGGAAACACAGTTAACTCTTCAGGGGCCTTGCAA

3666    TTGTCCCACCTTGGAAATGTGTCAAAACAGAGGACACCCTGAGGACACTGGGGTCATGTG
        ACATTGTTCTCTGGGGTAGGGGCATTCTCGGCCAGCTGGCCACTAGTTCAGTTCCCTCGG
        GAAGCCTATAGTATTCAGCTCCGCAGCCTTCAGGCTAAGCCCCACCTCCTGTGTAGGAAG
        TCAGCATTCTGGGCGAGTGAGCAAGATGCTACCTGCAACATGATACTGTAAGCTCCCTCT
        GTTCATCCTTTCTGTGGTGCAACCTCTTAGCCCACTCAATCCAATCCAGCAGACGGTACC
        [A,G]
        TGAAGCTAGAGCATGCCAAGCACTGAGCTGGCCCTTTGCATGGGGAGGTTTGACGGAGGC
        TCAGAGGGGATTCACGCAGGATGAAGTTGGGGGTTAAGCTGGGACAGACAGTGCAGCTTG
        GGTCCCCTTCCACTTCCATTCCTACATGCAATGATGGCAGCCCCTGGGTAAGTCGGGGGA
        AGGCAGACATTCAGGCGTTGTCCCTTGCCTCCCTAGCTAGAGAAGGAGGGGTCCTAGGGG
        CACGGAGTACTTCAGTACTCAAGTAATGTTAACAGCAACAACAACAGCAGAGATGGTTGC

4757    AGAAGCTCACAGGAGGGGCTTGATGCCATCCCCCAGGCAGCTGGTACCTCTGCGTGTCCT
        TGGAGGAGCCTCCAGGTCTCTGGGTTTGTGGTGGGGCTGACCCGGTGCCCCCACCTCAGA
        GTCCTGAGGACTTGATCTCATGGGCCGGCTCTGCTCACATCACAGGGCAGTCAGCCTGAG
        AAGGTAGCTTCTTACTCAGGCTTGTCAGTGGTGATGAGGCTGTCACTGTGGGTGGTGGCT
        GGGCCAGGGCCAGCTGGGGAGAGAGAGAGGGAGGGAGAGAGGGAAGGAGAAGGCGGCACG
        [G,A]
        AGCCAGGAGCTGGGGCTGGATAGTCTGTGGCCATAACTGCCCCGGGGACCGCAGGGCCGA
        GCAAGGGGGCTGGGCTCTGGAAGCCAGGAGGAAGGCCAGGATAGGGGCTGGTACTCAGTC
        CACATCTCAAAGCCGGTGGGAGGGTTCTCCACACGCTCTCGGGCACGGTCAACCTCTGTC
        TCTCGTATTAGAGTCTCGACTGTATTTTCTCTCTTAAATATTAATCACTGCCTTAACGTG
        CTTGGGGGAGCAGCTAAATCATAATTCTAGGACCAGCTTTGGGTCGAGGGCTTGAGGTGG

4816    TTGGAGGAGCCTCCAGGTCTCTGGGTTTGTGGTGGGGCTGACCCGGTGCCCCCACCTCAG
        AGTCCTGAGGACTTGATCTCATGGGCCGGCTCTGCTCACATCACAGGGCAGTCAGCCTGA
        GAAGGTAGCTTCTTACTCAGGCTTGTCAGTGGTGATGAGGCTGTCACTGTGGGTGGTGGC
        TGGGCCAGGGCCAGCTGGGGAGAGAGAGGGAGGGAGAGAGGGAAGGAGAAGGCGGCAC
        GGAGCCAGGAGCTGGGGCTGGATAGTCTGTGGCCATAACTGCCCCGGGGACCGCAGGGCC
        [G,A]
        AGCAAGGGGGCTGGGCTCTGGAAGCCAGGAGGAAGGCCAGGATAGGGGCTGGTACTCAGT
        CCACATCTCAAAGCCGGTGGGAGGGTTCTCCACACGCTCTCGGGCACGGTCAACCTCTGT
        CTCTCGTATTAGAGTCTCGACTGTATTTTCTCTCTTAAATATTAATCACTGCCTTAACGT
        GCTTGGGGGAGCAGCTAAATCATAATTCTAGGACCAGCTTTGGGTCGAGGGCTTGAGGTG
        GGGAGATGACCCTCAGAGTCAAGTCCGAGGCCCTCTCCTTGCCAAAGCTGTCTAGGGTGT

5640    AGTCCCTCCACAATCTCCCACATGGGCAAGTCCTCGGCGTCTCGCCACCTGTGTGATGGA
        CTTAAATATTTCATCATGGGCTGCCATCCAGCTCTGCTTTGATTACAAATGTGTGTCCGA
        TGAGGACGGGGAGGCCGCCGTGGCAGGTGGACGGCAGCCTTTTGCAGGGCTGGCTTTTGG
        AGGGCTGGCTTTGGAGGGCTGGCTTTTGGAGGGGTGGTTTTTGGAGGGGGTGGTTACTTC
        TAGATAGATCTGGGTTCAAACCCTGATGCCACAGTTTATTGTGGCTTCCCCTGGGAGGGT
        [-,A,C]
        CGGGCAGTAGGATGCTCTGAGGTAGGTCCTACTCTAATCTCATTTTTGGGAATAGAAAGT
        GTTGCCTAAATCCTGCCAACATCACTCCGTTGAGAAGGCAGGCTGGGAACTGCCTGTAGC
        CAAGGCCCATGCCCACCCACCAGGCGGCACAGCCTCCCCTCTCTCCACTTCTGACCTCTC
        TGAGTTTCACTGTCTTCCTCTGCCCAGTAGAAACCATAATGTAGGAGTGTCACGGTGCTG
        GCGGGACGGCACAGCCCGGCCGTGCTGCCGTGGTACGTGCGGACACATAGTAGGTACTCA

6535    CGGCAGGTGCTACGTGTGTCCCAACATCCAGCCCACGGGAGCGCAAGCCCGGCTGCACTC
        CGAGGGTACAGGAGGGACCCAGGGGCGGACGGCCTTTTCAGATGCGGGGTGCAGACCCTC
        TCTCTTTCCCAGGCTTCCTGTCCTTCAGCAAGGCCACCTGAGAACTGAATTGTAAATTCC
        AGCCTTGCAAGGAAGGGGAGGAGGCTGAAAGACAAGATAAATGAATAAATAAAAACCTAT
        TGGCTCTAAATGCACAATGAGAATTAATAGGGATGAGCTCTCAACAGAGGATTCTTGGGC
        [A,G]
        AGGACAAAATTGCAATGAGGGCTGTGGGAGAAGAGAGGCGGCCCCCACCCACATTCCCAG
        GGCCTGCCCCTGAGGTCAGCCCAGCAAGGCTGCAGGCCGAGAAGGAGGCAGGAAGGGCGA
        CGGGCAAAGGCATCTCAGCCCCTCACAGGGCAACTGCTGCTTGCAGGACCCTTGGAGATG
        GGGAGGGCGTGACTGGCATTGGGAGGTGCCCCTTAGCGCTCTGCAAGACACGTCTCCCGC
```

FIGURE 3

```
          CAGCCACCTGTGGTCCCCGTGAGAGGAAGGGGGAAACCCTTCCTTTGTGCCTTGGAGACA

7503      GGGGGTCCTGGGAATGGAGGGATGGAGCTGGGAGCCGAGATTCCCAGATGCGGGCCTTGC
          TTGCCAAGGCGGGGGGTCTGTGCTTCTGAGCAGGAGCCAGTGGAAGGTTCCAGAGAGAGG
          CAGCACTACCTCCTGACAAGAGCTACAGGCCTAGAATCTGAAGCCCTGAGCTCACATCCC
          AGTGTTGTCACTTGGGGGCTGGGTGGTCTGAGGTGAGTCATTCAAATTCTCCAAGCCTCA
          GTTTCCCTATCTGTAAAAGGGGTATGATCTTCACTGCCTTGCCCACCTCACCGACTTCCT
          [C,G]
          GGAGAGGAAACGCCTCTCACAACATAACAGAAGGTGCTTGGCAGAGACTGTAAGTTCTTT
          TCTAGTGTGAGGCTTGAATATTATTGTTTATCCAATCTTGATAAGCTTGACATTTCTACA
          TTGACATGGGTGAACATTCAAATTTCAGCCATGTTGGAGTCTCCAGCATACAGAGGCACC
          TTGGCTTTGGTGGACAGGAAAGCCAAGGGCAGGCTGTGAACGTGCATCTTTGAACACACA
          TCTTACTCAGTTAGAACCTGCTTTAGCTTCTCTGTGCCTCAGTTTACCCAGCTGCAGAGC

9001      TTTAGATGTCTGCTTTTCCCCTGGACTGTGAGCTCCTCAAAGGCAGGGCTGTCTCGCTAT
          CTTTGCATCCCCAGCACCTCCCTCTGGACTTGATCAGTGCCCAGTAAGCATTTCCTGAAT
          GAATGGATGCATGGGAGGAAGCATGGTGTGATCCAGGAACGCTTCTCAGAGGAGGCAGGA
          TCAGGGCTGTATTTTTTTTTTTTTTTTTTTTTTTTGGAGTCAAAGTCTCGCTCTGTCGC
          CCAGGCTGGAGTGCAGCGGCGCAATCTCGGCTCGCTGCAAGCTCTGCCTCCTGGGTTCAC
          [G,A]
          CCATTCTCCTGCCTCAGCCTCCCCAGTAGCTGGGACTACAGGCACCTGCCACCACGCCAG
          GCTAATTTTTTGTATTTTTAGTAGAGACGGGGTTTCACTGTGTTAGCCAGGATGGTCTCG
          ATCTCCTGACCTCGTGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGAGTG
          AGCCACTGCACTTGGCCAGGGCTGAGTCTTGAAGGAAAAACTGGGGTTTGGGTCAGGACA
          GAGGAGACCCTGGAAGCCCCTGCTTCTCTCCACTGCAGTCCCTGTTCTGTGGGATTTGCG

9617      GAGGTTTGCACAACTCTGTCCTTAAGTCAGTTGCAAGTGACTTCGGCACCTGAGCTGCAC
          CAGCCGTTAAAGCCACTCAGTCTCTTGAAATGCCCGAGGCAGGGCCCAGCCTAGGACAAG
          AATAGTTCTGTGAAATGACATCTTGTTGCACAGTGAAGTCTCCCTCCTGGGCAGTAGACA
          ATGAGAAGACCGAGGCCCGGGGCCCAGGGAGTGAGACCCTTGCTTCTGACTTCCCTTGAG
          GGAATGAGGTTGGGTCCAGACACCCCGTGGAAGGCAGGCAGCTGTGTGAAAGGGCCCAGA
          [C,T]
          GGGACATCTTTCCAAAGAATGTCAGAGACTTAGAGACCCCCAGACCTTTCCGGTTTGCGC
          ATCCCCACCTTCCCAGGCTGTCTTCCTCTATGCTTCCTAACTCTGATGTTTAATCCATTT
          CCCTTTTTCTCATTTACTGTGGGTATAATGACAAGCTGCCTCCAATCCCACCTGCGATGG
          GGCAGGCAGTGGACGGATGGACAGACGAACGGACAGACAGGCAGGCCGCACCATGCTGCG
          GATGAGACGGATGGATGGACAGACGGACAGACAGGCAGGAGCACCATGCTGCGGATGAGA

9672      TGCACCAGCCGTTAAAGCCACTCAGTCTCTTGAAATGCCCGAGGCAGGGCCCAGCCTAGG
          ACAAGAATAGTTCTGTGAAATGACATCTTGTTGCACAGTGAAGTCTCCCTCCTGGGCAGT
          AGACAATGAGAAGACCGAGGCCCGGGGCCCAGGGAGTGAGACCCTTGCTTCTGACTTCCC
          TTGAGGGAATGAGGTTGGGTCCAGACACCCCGTGGAAGGCAGGCAGCTGTGTGAAAGGGC
          CCAGACGGGACATCTTTCCAAAGAATGTCAGAGACTTAGAGACCCCCAGACCTTTCCGGT
          [T,C]
          TGCGGCATCCCCACCTTCCCAGGCTGTCTTCCTCTATGCTTCCTAACTCTGATGTTTAATC
          CATTTCCCTTTTTCTCATTTACTGTGGGTATAATGACAAGCTGCCTCCAATCCCACCTGC
          GATGGGGCAGGCAGTGGACGGATGGACAGACGAACGGACAGACAGGCAGGCCGCACCATG
          CTGCGGATGAGACGGATGGATGGACAGACGGACAGACAGGCAGGAGCACCATGCTGCGGA
          TGAGATGGATGGACGGACGAACGGAGAGGCAGGCAGGTCGAACCATACTGCGGATGAGAC

9761      GTTGCACAGTGAAGTCTCCCTCCTGGGCAGTAGACAATGAGAAGACCGAGGCCCGGGGCC
          CAGGGAGTGAGACCCTTGCTTCTGACTTCCCTTGAGGGAATGAGGTTGGGTCCAGACACC
          CCGTGGAAGGCAGGCAGCTGTGTGAAAGGGCCCAGACGGGACATCTTTCCAAAGAATGTC
          AGAGACTTAGAGACCCCCAGACCTTTCCGGTTTGCGCATCCCCACCTTCCCAGGCTGTCT
          TCCTCTATGCTTCCTAACTCTGATGTTTAATCCATTTCCCTTTTTCTCATTTACTGTGGG
          [T,C]
          ATAATGACAAGCTGCCTCCAATCCCACCTGCGATGGGGCAGGCAGTGGACGGATGGACAG
          ACGAACGGACAGACAGGCAGGCCGCACCATGCTGCGGATGAGACGGATGGATGGACAGAC
          GGACAGACAGGCAGGAGCACCATGCTGCGGATGAGATGGATGGACGGACGAACGGAGAGG
          CAGGCAGGTCGAACCATACTGCGGATGAGACGGACGGACGACAGACAGGCAGGCCACAC
          CATGCTGCGGATGAGATGGATGGATGGACAGACGGACAGACAGGCAGGCCGCACCATGCT

9811      GCCCGGGGCCCAGGGAGTGAGACCCTTGCTTCTGACTTCCCTTGAGGGAATGAGGTTGGG
          TCCAGACACCCCGTGGAAGGCAGGCAGCTGTGTGAAAGGGCCCAGACGGGACATCTTTCC
          AAAGAATGTCAGAGACTTAGAGACCCCCAGACCTTTCCGGTTTGCGCATCCCCACCTTCC
          CAGGCTGTCTTCCTCTATGCTTCCTAACTCTGATGTTTAATCCATTTCCCTTTTTCTCAT
          TTACTGTGGGTATAATGACAAGCTGCCTCCAATCCCACCTGCGATGGGGCAGGCAGTGGA
          [C,T]
          GGATGGACAGACGAACGGACAGACAGGCAGGCCGCACCATGCTGCGGATGAGACGGATGG
          ATGGACAGACCGACAGACAGGCAGGAGCACCATGCTGCGGATGAGATGGATGGACGGACG
          AACGGAGAGGCAGGCAGGTCGAACCATACTGCGGATGAGACGGACGGACGGACAGACAGG
```

FIGURE 3

```
            CAGGCCACACCATGCTGCGGATGAGATGGATGGATGGACAGACGGACAGACAGGCAGGCC
            GCACCATGCTGCGGATGAGATGGACGGACGGACGGACAGATGGACAGACAGGCAGGAGCA

11617       AGAAGTGGGTCTTGCAGGATGGGAAGGGTCAGGAGAGTGAGGATGTGGGCAGAAGCAATG
            GTTCGTGCAGAGACGCAGGGAAGGGGGCCAGGTGATTCGAGGGGAAAGTGCGTGGGTGAC
            AGAGAGGAGACAGTCCACTCCCCCTGCCCACCTCATCCAAGCCCCTGTAGGTCTGTTACT
            GTGCATCTGACCGGTGAATATTCTGAGACTTCTCAGAGCCCACTGAGTGTAGGAGCTGGG
            GTTCAGCCTTCCTGTGTCTGGCTCCTGACCGCTCGCTAGGGTTAGGAAGGATTAGGCCAC
            [A,G]
            GGCTCTGAAGGAGCAAGAGGGGCAGGAGGGCAATTGAGGGGCAATTGAGAGGAACCCAGA
            ACATGGAAGCCCTGTGCCGTGGGGCTGGTCCAGAGCTCACCAGGCTGGACCACGTGGTTG
            CTGAGCCATGGCCCCTGACCGGGGCTGACCTGGCCAGAGTCCCTGTGGCCAGCACTGATG
            CAGGGCTCCTTCCTAGAGGGGCCGGGCCATGAGGAACGGGAGAAACGGCAGATGATGCGG
            GAACCGGTCTGTTCGGCTTTGGTTTGCAGGATCCGATTTGTTTTTCATCAGCAGCAGATT

11835       CCCACTGAGTGTAGGAGCTGGGGTTCAGCCTTCCTGTGTCTGGCTCCTGACCGCTCGCTA
            GGGTTAGGAAGGATTAGGCCACGGGCTCTGAAGGAGCAAGAGGGGCAGGAGGGCAATTGA
            GGGGCAATTGAGAGGAACCCAGAACATGGAAGCCCTGTGCCGTGGGGCTGGTCCAGAGCT
            CACCAGGCTGGACCACGTGGTTGCTGAGCCATGGCCCCTGACCGGGGCTGACCTGGCCAG
            AGTCCCTGTGGCCAGCACTGATGCAGGGCTCCTTCCTAGAGGGGCCGGGCCATGAGGAAC
            [G,A]
            GGAGAAACGGCAGATGATGCGGGAACCGGTCTGTTCGGCTTTGGTTTGCAGGATCCGATT
            TGTTTTTCATCAGCAGCAGATTTGCTTAAGTATATGAAAATGTGTTTCTAATTCCCCGAG
            CACACACCAACTGCTGGCGGGGGAGGGAGCAGTGCATAGGAGCAGAGTGAATGCCACCGG
            GAGTCAGAGTGCTAGGCCCTGGCTGCTGAGAGAGCGAGAATACGCCCCCAGCCTCAGTTT
            CCCCAACTGAGCAGCCGGGGAAGATTTGGCTAGATTAACCAGTTCATTCAATGTTCCCTG

12837       TACATTCACAATGTTGTGCAACCAGCATCTCTACTTAGTCCCAACACGTTTCCATCGCCC
            CAATAGAAAACCCTGCACCCGTTAGTTACTCCCCATCTCCTTCCCCTGCCCCTGAAAACC
            ACGCGTCTACTTTTTGTCTCCATGAATTTAGCTATGCTAGACATTTCATACGAATGGAAT
            CAGACAATATGAGGCTCTTTGTGATGGCCTTCCTTCACTGGCAAAATGTTTCCAAGGTTT
            GTCCACATTGTCGCATGACTCAGTGCTTCATTCCTGTTTATGGCTGCATAATATGCCATC
            [C,A]
            TGTGGACACACCATATTTTGTGTATCCGTTTCCTAACTGATGGACATTTGAGCTGCTTCT
            GCTTTCTGGCTATTAGGAGTGATACTGCTGTGGACATTTGGGTCTCAGTTTTTGCATGTG
            TGTATGTCTTCATTTCTCTTGGCTGTCTACCTAAAAGTGGAGTTTCTGGGTCACAAGGTA
            ATTCTATGTGTAACTTTTTGGGGAGCCACCAAACTGTTTTCTACAGGTGCTGCACCTCTT
            ACGTTCCCACCAGCAATGTACGAGAATGCCAGTTTCTCCGAATCCTTGTCAACACTTGTT

14651       ATGAGGAAACTGAGACTCCAAAACTGAGTGAAGTCAAGGCTCAAACTCAGATCCCAGTCA
            TTTGATGACTAGGCCACAGTGAGGCCTGAGGAGGGGAAAAATCCCAATGGTTACCCTCCC
            CTTCCCCTCCCCACCCTCATTTTCTTCTCCCTCTTTCAGGCTGGGATGTGGACTTGGATT
            CTCAGAGCAGGGTCCTTGGAAGGAGATGCTGTGACTTCTCTCTGGCCTCCAAATACCTCC
            TCAGCCTCCAGTCCACCTCCGTCCCTCTCCCACGCAGCCAGGCACTGTTCTGTCCTCTTC
            [C,T]
            TTGTCCCACAGTCAGTGCTTGCATGTAGCAGGTACTTAATAAATGCTGAAGATAATTATC
            CATCATTTCAAATAGAGACACACAACTTAGAAGGCATGCTGGGATTGTCTAAGGCCAGAA
            AAACCCCAATGTCGATAAGCATGTTACAGTGAAATTGACTGCGCCCAGGAAAGGGGACCC
            CAGAAGCAGGTGGCTGGTGTCCCCCTACCCTGCCCCAGGCCCCGAGTTCCCCAATCCACC
            ACTAGGAAGTCCTGGGCTCCTGTGAAGACAATATAAAACCACTGATTAGGCCAAGTGTGG

15701       CCAGAGGTAGACTGGAGCCATGAGGACAGGGGCCCTCCCCAACCAGGTCTCTGTCCATCT
            ACACGTGCCCTGGATCTGACTTCACGTGATGGCATCTGGTGGGGGACACAGGATGCCTGC
            CCGGATGCCACCTGCAGCCAGTGGGGGCCGGAGCTGCCTCTTCAGGGTCAGTGAGGGTGA
            TACATCTACTTCCCAGCCTGCTTAGGTGAGCTCCCGCCTATGTGTCACTACTGGTGACTG
            GCATGGCTCAGAGCCAGATCTTGGGGGCCCTGAGGGGATCAAGAGCGTCCCCTAAGCCCA
            [C,T]
            CTGCCAGCTGCGGTCTTCTCTGTGGTGGCAGCATCACAGAAAGTGGACAGAAAGAGTGCT
            CTGTGCCAGGAGGGCAAGGCCGGGTAGGATGGTGGCTGGAATGCTGGCGATCGCAGCAAT
            GCCGGCGATCATGGTGCTGGGTTTTGGTGGTGTGCTGGACGCCTGGGAGCCTCATGAGTG
            AGAGACTGGGGCACACGTGCTTCCGTAGTGCCATGCACCGGTGGCAATTCAGAGAAAGAC
            GCTGTGCAAAGCACCCCATGTGTGCAGCTTTTTGCCCTCTCGTAACAGGACGGAGCCAGG

16587       CCAGAACTGCCCACCCTGTGGAAAGGGACTCAGGCCTGTCTTCAAGGACCTGGCATCCTT
            CTGTCCCAGGGCAGTTTGTCTTGGGTCTCTCAGGGACCGTTTGGGCCTCTTCAGCCCCTC
            ATTCCACTTCCCTCCTGCTGCCCAAGTCATTCGTCCACTTGACTCCAAGAGTCGGCTGGG
            GAAATAAAAGGAAATGAAACACGACCAGGCATTTTCCCTTGGCCGAAGCAGAAGTCTGCT
            GTCGGGCAAAAGGTGAAGAAGAGACCAATGAGAGATGAGCCCACGGTGCTCCTGCCCTCC
            [G,A]
            CCAAGGCAGGCCATCCTCTGCTGCCAGCCTGCAACAGGGCAGTGTCCTTCTGGGAGGTGT
            CCCTCCCTCTGGGGGATCAAGAGATGGCCAAAAGCAGGTGGCAGCAAGTGGAGAAGGCTG
```

FIGURE 3

```
                TTCATCCAGAACGCACCTTGTCTCTGCCCCTGTCCCCACCCAGGCAACATCCAAAACCTT
                TGCCCACAGTTCCGGGGCTGGCACCGTCCTGGGGCTCAGCTCCTAGGGACGGGGCTCCCC
                CAGGCACTGGCTGCCAGGAACTGGGTGGCCCCGGGCAAGTCTCTTCCCATTTCGGGGTAT

16710           CCACTTCCCTCCTGCTGCCCAAGTCATTCGTCCACTTGACTCCAAGAGTCGGCTGGGGAA
                ATAAAAGGAAATGAAACACGACCAGGCATTTTCCCTTGGCCGAAGCAGAAGTCTGCTGTC
                GGGCAAAAGGTGAAGAAGAGACCAATGAGAGATGAGCCCACGGTGCTCCTGCCCTCCGCC
                AAGGCAGGCCATCCTCTGCTGCCAGCCTGCAACAGGGCAGTGTCCTTCTGGGAGGTGTCC
                CTCCCTCTGGGGGATCAAGAGATGGCCAAAAGCAGGTGGCAGCAAGTGGAGAAGGCTGTT
                [C,A]
                ATCCAGAACGCACCTTGTCTCTGCCCCTGTCCCCACCCAGGCAACATCCAAAACCTTTGC
                CCACAGTTCCGGGGCTGGCACCGTCCTGGGGCTCAGCTCCTAGGGACGGGGCTCCCCCAG
                GCACTGGCTGCCAGGAACTGGGTGGCCCCGGGCAAGTCTCTTCCCATTTCGGGGTATAGA
                CTTCCTGCCTGTAAAATGAGGGGGTCTGCAGGTCAACCTCAGAGTCCCACTGTACCCCCA
                GATTCTGCTTCAGGGAGACGGAGAGAGAGAGAAAGAGAAAGAACGATAGAGAGATGCAAT

17888           ACTCCTGGGCTCACTTTCTTTTTTTTTTTTTTTGAGACGGAGTCTTGCTCTGTCACCCA
                GGCTGGAGGGCAGTGGCATGATCTTGGCTCACTGCAACCTCCACCTCCCAAGTTCAAGCA
                ATCCTCCTTCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCACCTGCCACCATGCCTGGC
                TAATTTTTGTATTTTTAGTAGAGACAAAGTTTCACCATGTTGGCCAGGCTGATCTCCTGA
                CCTTAAGTGATCCGCCCACCTCAGCCTCCCAAAGTGTCGGGATTACAGGCGTGAGCCACT
                [T,C,G]
                TGCCTGGCCATAGCCAGACTTTCTTGATTCTATATCCTTCTCCTCAGAGCAGAAACATCG
                AGCATTTGTTGAGTGCCTCATGTATACCAAGCCCTTAACCTAAGCTATAGCTCATTGAAC
                TCTCACAGAAGTCTTAAGGTAGAGCTTGTATTTAGATCCGTTTTGAATATGAGGAATCTC
                AGGTTCAGAGAATTTAAGCCACTTGCCGAAGGCCACACAGCTTCTAAGTAGAGGAGGCTG
                GCACCTCCAGCCTGGGCCGCCCGGCCCGGCATCCAGGTTCCTAAACAGCCTGCTCAGCTG

19282           TCAGGTTTTGTGGCTCTTTTCCATCTCCCAACACTTGGGGCAGTCTTCTCGAAGGCCTCA
                AGCCCAGCGGGCAGCTATGACCCCACCAGGAGCGGAGCGGGCAGGGACCAGGCTGCCCTC
                TAAGCCACTCGGCTGGCTCTCAGCCGGGGTGCACACTGGACTTGCCTGGGAGCTTTTCAT
                TCCCCCCCTGCCGCAGCTGCCCCCCAGACCAGCTTCAGCAGCCTCTCTGGGCGGCCCAGC
                AGGAGGAGGCATTAAACCTCCCCAGGTGGTCCCAGTGCACAGCCAAGTTTGAGAAGCACC
                [G,A]
                ATTGAAACCTCTCCCAGGCCTGCCCTGGAGCCCTTCCAGCCTGAAGCATCTTGTCGTCTT
                AAAACTGAAAGACCAGGAGGAAGAGAATTCCATGGCCTGCCTCGGCTCTCTCGGAGCCTC
                TCTCACATCTGAGCTGCAGGTGCTCCATCCTCCTTCTGGCTTCCTGGGTGCCGAGGGGTG
                CCAGCTCTCCAGGCTTGGGAGAGGGCCACTTAAGCCCTCACACTTTGTTCCCAGGCTCTT
                CACCTGTCTTCCTGGAAGGAGGGGGCCGGCCAGCATTAGGGCTGTCACGGGCGCTGCTTA

19429           GGTGCACACTGGACTTGCCTGGGAGCTTTTCATTCCCCCCCTGCCGCAGCTGCCCCCCAG
                ACCAGCTTCAGCAGCCTCTCTGGGCGGCCCAGCAGGAGGAGGCATTAAACCTCCCCAGGT
                GGTCCCAGTGCACAGCCAAGTTTGAGAAGCACCGATTGAAACCTCTCCCAGGCCTGCCCT
                GGAGCCCTTCCAGCCTGAAGCATCTTGTCGTCTTAAAACTGAAAGACCAGGAGGAAGAGA
                ATTCCATGGCCTGCCTCGGCTCTCTCGGAGCCTCTCTCACATCTGAGCTGCAGGTGCTCC
                [A,G]
                TCCTCCTTCTGGCTTCCTGGGTGCCGAGGGGTGCCAGCTCTCCAGGCTTGGGAGAGGGCC
                ACTTAAGCCCTCACACTTTGTTCCCAGGCTCTTCACCTGTCTTCCTGGAAGGAGGGGGCC
                GGCCAGCATTAGGGCTGTCACGGGCGCTGCTTAATGTCAAGCTGCCCATCTGGCTCCTGG
                CCTCCCTTTGGCCTTCTCTCCTGCGCTCCCCACCAAGCTCCTGGCTCAGCAGCGTGCATG
                CGTTAACCCATTGCCCCCCTGCAGTGTTTTGTGTGTCCAGCCTGGCCCTTTGCTCAGTCG

19561           CAGCCAAGTTTGAGAAGCACCGATTGAAACCTCTCCCAGGCCTGCCCTGGAGCCCTTCCA
                GCCTGAAGCATCTTGTCGTCTTAAAACTGAAAGACCAGGAGGAAGAGAATTCCATGGCCT
                GCCTCGGCTCTCTCGGAGCCTCTCTCACATCTGAGCTGCAGGTGCTCCATCCTCCTTCTG
                GCTTCCTGGGTGCCGAGGGGTGCCAGCTCTCCAGGCTTGGGAGAGGGCCACTTAAGCCCT
                CACACTTTGTTCCCAGGCTCTTCACCTGTCTTCCTGGAAGGAGGGGGCCGGCCAGCATTA
                [G,C]
                GGCTGTCACGGGCGCTGCTTAATGTCAAGCTGCCCATCTGGCTCCTGGCCTCCCTTTGGC
                CTTCTCTCCTGCGCTCCCCACCAAGCTCCTGGCTCAGCAGCGTGCATGCGTTAACCCATT
                GCCCCCCTGCAGTGTTTTGTGTGTCCAGCCTGGCCCTTTGCTCAGTCGACCTAGAGCACC
                ATCCTCCCAGACTAGTCGAGTGTCCCTCCACCTGTCCTGAGTCCAGATGAAATCCCACCT
                CCCCCAGGAAGCCTTCTGACTGCCCCAGCCCGTCACCTCCAGGGCTTGTCATCTGTGCCA

19575           AAGCACCGATTGAAACCTCTCCCAGGCCTGCCCTGGAGCCCTTCCAGCCTGAAGCATCTT
                GTCGTCTTAAAACTGAAAGACCAGGAGGAAGAGAATTCCATGGCCTGCCTCGGCTCTCTC
                GGAGCCTCTCTCACATCTGAGCTGCAGGTGCTCCATCCTCCTTCTGGCTTCCTGGGTGCC
                GAGGGGTGCCAGCTCTCCAGGCTTGGGAGAGGGCCACTTAAGCCCTCACACTTTGTTCCC
                AGGCTCTTCACCTGTCTTCCTGGAAGGAGGGGGCCGGCCAGCATTAGGGCTGTCACGGGC
                [G,A]
                CTGCTTAATGTCAAGCTGCCCATCTGGCTCCTGGCCTCCCTTTGGCCTTCTCTCCTGCGC
```

FIGURE 3

```
                TCCCCACCAAGCTCCTGGCTCAGCAGCGTGCATGCGTTAACCCATTGCCCCCCTGCAGTG
                TTTTGTGTGTCCAGCCTGGCCCTTTGCTCAGTCGACCTAGAGCACCATCCTCCCAGACTA
                GTCGAGTGTCCCTCCACCTGTCCTGAGTCCAGATGAAATCCCACCTCCCCCAGGAAGCCT
                TCTGACTGCCCCAGCCCGTCACCTCCAGGGCTTGTCATCTGTGCCACTCATGGGGACCAG

19634           TGTCGTCTTAAAACTGAAAGACCAGGAGGAAGAGAATTCCATGGCCTGCCTCGGCTCTCT
                CGGAGCCTCTCTCACATCTGAGCTGCAGGTGCTCCATCCTCCTTCTGGCTTCCTGGGTGC
                CGAGGGGTGCCAGCTCTCCAGGCTTGGGAGAGGGCCACTTAAGCCCTCACACTTTGTTCC
                CAGGCTCTTCACCTGTCTTCCTGGAAGGAGGGGGCCGGCCAGCATTAGGGCTGTCACGGG
                CGCTGCTTAATGTCAAGCTGCCCATCTGGCTCCTGGCCTCCCTTTGGCCTTCTCTCCTGC
                [G,A]
                CTCCCCACCAAGCTCCTGGCTCAGCAGCGTGCATGCGTTAACCCATTGCCCCCCTGCAGT
                GTTTTGTGTGTCCAGCCTGGCCCTTTGCTCAGTCGACCTAGAGCACCATCCTCCCAGACT
                AGTCGAGTGTCCCTCCACCTGTCCTGAGTCCAGATGAAATCCCACCTCCCCCAGGAAGCC
                TTCTGACTGCCCCAGCCCGTCACCTCCAGGGCTTGTCATCTGTGCCACTCATGGGGACCA
                GGACACAGGTGACTTCTCTGGTGGACACAGCAGAACGGTCAACATTCCCAAAAGGGAGCA

20383           CACACATCTGAGTCACCTGCTGAGGGTTTTGGTTAGTTGGTTGGTTAATTACTCGATTGG
                TTAGTTTGTTGGTTTGTTTGTTAGCTTTCAATTTGGAGTAGAATTTCTGACCTTAGGGCC
                CAGGAATGTTGCCACACCCCTCCCGTCCCCAGGTATTGTCATGTTTTCATGGGTGGCCAA
                ATCTAAGAGCTGCTTCTCTGGGGCACGAAGGATGTTCACAAATAGTTGATGAATAAGTGA
                ATGAATAAATCAATGAAACTTACCAGCCCAGCCTCACTACTCGCACCCACCCCCAACGAC
                [C,G]
                AGCCAGGGTTCATCCACAGAGGGGTGTACCTGTCCAGGTGTCCCCAGGTGTGGGCAGACC
                CAGTAACTTTACTCTTTCATCGGCCCCACCGCCTCTTAACTCCTCAGAGACCAGCAGGAA
                GAAACCCTCGGAGGTCGCAGCTTCTGGCTGTTCTCAGGGGCAGGCCCCGTCCATCGGGTG
                CTGTGTCTACTCCTAAGACCTGGTTCTGAGTATGGAACACCTGGAGAGGGAAGGGGCCGA
                GGAGGGGGAGTCACTCGGCTGTGTCAGGCTCCGCCCCTGCCTTCCTGAAGCACACAGTGG

22878           CAACTGGGATAGAATTTCCTGCCCCAAACATTCCTGGAAATCTGGCTGTGGGAAGAATCC
                ACATATGCCCAGGGCAAAGCAGAATGTGTCCTTTAAGAAAACAATAATACATTTTTAAGT
                TCCTGGAGAGATTAACCCTTGTCTAGCCAGAGCCATGGCAATGCCTCCCCGCCCACCACA
                CTCTGGTGGTTCGGCTGACGGAGGAGATCAGTCATTCAGGGGTCTGCGGTCCTGATGAGC
                AGTGGGTGCCCACACCAGGCCTGGCATTTCATCCTTGCTTTCTGACCTTGGCTTCCCAGT
                [T,A]
                GACCCTCTCCCGGGCAGCTCGTCCATCAGGGCAGCCCAATGCCCTCAGGTCCTCCGAAAG
                GATCTCAGGGTGTTCTGTGGGGGCAACCCGAATTGGTGTAAGAAGACTAAGCAGTCGATC
                TGCTGGAACAGCATCCCCAAAGCGGAGCGAAGCCCGCGGATGCCCACCGCCTCTCCCCCA
                GGCAGCGTCCTACCTGGATAGAACTGCCTGGAGCCACTGCAGAGGGTCCTCGCTCAGTTA
                GGGAATGTTTGTCATATACCGCTGTGTGCAAACAGCTGTTGGGAGTGTGGCGCAAAGGTG

22993           TAAGTTCCTGGAGAGATTAACCCTTGTCTAGCCAGAGCCATGGCAATGCCTCCCCGCCCA
                CCACACTCTGGTGGTTCGGCTGACGGAGGAGATCAGTCATTCAGGGGTCTGCGGTCCTGA
                TGAGCAGTGGGTGCCCACACCAGGCCTGGCATTTCATCCTTGCTTTCTGACCTTGGCTTC
                CCAGTTGACCCTCTCCCGGGCAGCTCGTCCATCAGGGCAGCCCAATGCCCTCAGGTCCTC
                CGAAAGGATCTCAGGGTGTTCTGTGGGGGCAACCCGAATTGGTGTAAGAAGACTAAGCAG
                [-,T]
                CGATCTGCTGGAACAGCATCCCCAAAGCGGAGCGAAGCCCGCGGATGCCCACCGCCTCTC
                CCCCAGGCAGCGTCCTACCTGGATAGAACTGCCTGGAGCCACTGCAGAGGGTCCTCGCTC
                AGTTAGGGAATGTTTGTCATATACCGCTGTGTGCAAACAGCTGTTGGGAGTGTGGCGCAA
                AGGTGGGTAAGGCCCCTGCTCTCCCAGAGTTCACACTCACAGAAGGTTCTGGAAGGAGGA
                ACACTGTGGGCAGGGTTGAAAGGCCTAAAGTGCTCCCTTTCCTCCCAAATAATGCGGGGT

23852           GTAATTGCTTGGAGCTGGGAGGAAAATTGCTCCAACCAGAAAACAAAACAGAAAAGCCGC
                CTTGGCCAGCTGCAGCTCCAGCCCTAAAATGCCAGGTTGGTTTACGCTGATTCACGAGCG
                GGGAGGGTGACCTTGCTGTCTGTTGTCCAGGGCCTGTGCACGAAGAGAATCTGGAAAGGG
                AAGGAGAGAGACACCTGCACGCTGGGGAAGGAATTAGCAGCACAGAGAGCAAGAGGGACA
                GCGATCAATGAAACCATAGAAGGAGAATGAGAAACACACACACAGAGAGCGAGAGGGAGC
                [-,A]
                AGAGAGAGAGAGAGAGAGAGAGACGGACACAAAACGAGAGGGAGGGAGGGAGGGAGAGCT
                CAGAGAGTTAGAGACCGTCAGGGCCGCTAGAATTAGAATCAGCTCTGAACAGAATCTCCG
                TTTCCGCTTTGTTAATAATTTATTCCCTCTGCAACTTTTCTTACCAATAAATAGGAAGTA
                ATCTGTTAAGGAGAATTCCCCTAGCACCCCGGCTTTCTCCCTGGAGTCAGGGGAGGAGGA
                TGTGTCTCTGTGCCCTTCCTCCCTAGCAGCATGGGGGCCTGAGGAACACGCAGAACTTCA

23853           TAATTGCTTGGAGCTGGGAGGAAAATTGCTCCAACCAGAAAACAAAACAGAAAAGCCGCC
                TTGGCCAGCTGCAGCTCCAGCCCTAAAATGCCAGGTTGGTTTACGCTGATTCACGAGCGG
                GGAGGGTGACCTTGCTGTCTGTTGTCCAGGGCCTGTGCACGAAGAGAATCTGGAAAGGGA
                AGGAGAGAGACACCTGCACGCTGGGGAAGGAATTAGCAGCACAGAGAGCAAGAGGGACAG
                CGATCAATGAAACCATAGAAGGAGAATGAGAAACACACACACAGAGAGCGAGAGGGAGCA
                [-,G,A]
```

FIGURE 3

```
GAGAGAGAGAGAGAGAGAGAGGGACAGAAAACGAGAGGGAGGGAGGGAGGGAGAGCTC
AGAGAGTTAGAGACCGTCAGGGCCGCTAGAATTAGAATCAGCTCTGAACAGAATCTCCGT
TTCCGCTTTGTTAATAATTTATTCCCTCTGCAACTTTTCTTACCAATAAATAGGAAGTAA
TCTGTTAAGGAGAATTCCCCTAGCACCCCGGCTTTCTCCCTGGAGTCAGGGGAGGAGGAT
GTGTCTCTGTGCCCTTCCTCCCTAGCAGCATGGGGGCCTGAGGAACACGCAGAACTTCAG

23888  CAGAAAACAAAACAGAAAAGCCGCCTTGGCCAGCTGCAGCTCCAGCCCTAAAATGCCAGG
TTGGTTTACGCTGATTCACGAGCGGGGAGGGTGACCTTGCTGTCTGTTGTCCAGGGCCTG
TGCACGAAGAGAATCTGGAAAGGGAAGGAGAGAGACACCTGCACGCTGGGGAAGGAATTA
GCAGCACAGAGAGCAAGAGGGACAGCGATCAATGAAACCATAGAAGGAGAATGAGAAACA
CACACACAGAGAGCGAGAGGGAGCAAGAGAGAGAGAGAGAGAGAGAGGGACAGAAAAC
[G,A]
AGAGGGAGGGAGGGAGGGAGAGCTCAGAGAGTTAGAGACCGTCAGGGCCGCTAGAATTAG
AATCAGCTCTGAACAGAATCTCCGTTTCCGCTTTGTTAATAATTTATTCCCTCTGCAACT
TTTCTTACCAATAAATAGGAAGTAATCTGTTAAGGAGAATTCCCCTAGCACCCCGGCTTT
CTCCCTGGAGTCAGGGGAGGAGGATGTGTCTCTGTGCCCTTCCTCCCTAGCAGCATGGGG
GCCTGAGGAACACGCAGAACTTCAGACTTTAGGATGTCAGGGTCAGAGGCGGACAGCCCA

24151  CAAGAGAGAGAGAGAGAGAGAGAGAGGGACAGAAAACGAGAGGGAGGGAGGGAGGGAGAG
CTCAGAGAGTTAGAGACCGTCAGGGCCGCTAGAATTAGAATCAGCTCTGAACAGAATCTC
CGTTTCCGCTTTGTTAATAATTTATTCCCTCTGCAACTTTTCTTACCAATAAATAGGAAG
TAATCTGTTAAGGAGAATTCCCCTAGCACCCCGGCTTTCTCCCTGGAGTCAGGGGAGGAG
GATGTGTCTCTGTGCCCTTCCTCCCTAGCAGCATGGGGGCCTGAGGAACACGCAGAACTT
[C,G]
AGACTTTAGGATGTCAGGGTCAGAGGCGGACAGCCCACTCCTGCCCGGTCATTTTGTGAA
CGGGGAAACCAAGGCACAGATAGGGCAAGGCCCTGGCCAAGGTCACACATGGTGTTAGGG
GCAGTCCCCTGAGTCCTAATTCCATGGCCCCACGGGTCAGGGCACCTATTGATTTATGCA
CCTGCCCAAGCCATAGGGTTTCCCCCGAAATGGCAGAGGCCACATCCAAGGAGGAGGGTG
GGGCTAGCTCGGCTGCCTTTCCTTGCCTTCCCCCACGATTGCTTCCCCCGTGCTCGAGTC

24984  GGTGAAGGGCACCATTTCCCAATTTGTCTCAGGGCCCAGCTCAGTGGGCCATCCCCTGGC
TTCTTATCCCACCTCAGCTGCTGCCGAGCCGCATGACCCTGCGACATTGCTCAGCCTCTC
TGAGTCTCGGTTTCCTGAGGATCGCACTCTCCAGGATCCCTGGGAGCGTGGGAGGTGGGG
TTGGGGCACACAGGGCGCCCCAGCACAGGGCCGAGGTGGAAGACATGCTCCCTAACGGCGG
GGCCTGCTGTTTGCTGAAGCACCAGGCCAGACAGTGGCCATGAATGTGCTCCCAGCATCC
[A,T]
TCACCCATGAGCTGGCACCACCGAGGCACTTGCCATGGTGCACCTGGCATCATTCCTATG
ACAACCCTGTGAAGCCAGTGCTAGTAACCTCATTGAGCGTTCATTCATTCTCCGAAGATT
TCCCGAGTCCCTGAGGAGGGCCGGGGGCTGGGGCTGGAGTGGGGACAGGATCAGATGTGG
TCGCTGCCCGCATGAAGCCTCCCCTCCAACAGAGAAGCTGAGGCTCTCGGGCAGGAGAAA
GATCTTTTCCTCACCCATTCTATGTTAGTGGCTGAGGGCCCATCATAACAGACAAATTAA

25681  GGTCTCACTCTGTCACTCAGGCTGGAGTGCAGTGGCACAGTTACGGCTCACTACAGCCTC
GACCTCCTGGGCCCAAGGGATCTTCCTATCTCAGCCACCCAAGTAGCTGGGACCACTGGT
GTGTGCCACCATGCCCGGCTAGTTTTCTTTTTTGTTTTTTGAGGTTTTTTTCTGTAGAGA
TGGCATCTCCCTGTGTTGCCTGGCCTCATGGGAGCTTTCATAAGGAATGAAGACCCAAAA
CATTGGTGAACATCTATTTTGTATGCTAGGTTTAATGGAGAAATAGTCATGGAGAAGTAC
[T,C,G]
ATTGGCTTAAAAAAAAGTATCATCTCCTGGTGATAAACTGGCGGGAATTTTGCAAGACCT
GTGTGTCCAGGTCCCTCTCTGTGACCCTGCATCTTTGGAGATGAGAATGTTCCTTCCTCC
GGGCATTGGGAGGGCACCTCTCGAATGAGCCTCATGTCCTGCTTCAGGGAAGAAGGGCAG
GGGAAGGTCAAAGAGTAACCTTCCGCTTCTGTGGTTTTCTCAAATCCCTTCAGCTTAAAA
AAAATTATTTTTTTGAGACGGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTGTG

25698  CAGGCTGGAGTGCAGTGGCACAGTTACGGCTCACTACAGCCTCGACCTCCTGGGCCCAAG
GGATCTTCCTATCTCAGCCACCCAAGTAGCTGGGACCACTGGTGTGTGCCACCATGCCCG
GCTAGTTTTCTTTTTTGTTTTTTGAGGTTTTTTTCTGTAGAGATGGCATCTCCCTGTGTT
GCCTGGCCTCATGGGAGCTTTCATAAGGAATGAAGACCCAAAACATTGGTGAACATCTAT
TTTGTATGCTAGGTTTAATGGAGAAATAGTCATGGAGAAGTACGATTGGCTTAAAAAAAA
[-,G]
TATCATCTCCTGGTGATAAACTGGCGGGAATTTTGCAAGACCTGTGTGTCCAGGTCCCTC
TCTGTGACCCTGCATCTTTGGAGATGAGAATGTTCCTTCCTCCGGGCATTGGGAGGGCAC
CTCTCGAATGAGCCTCATGTCCTGCTTCAGGGAAGAAGGGCAGGGGAAGGTCAAAGAGTA
ACCTTCCGCTTCTGTGGTTTTCTCAAATCCCTTCAGCTTAAAAAAAATTATTTTTTTGAG
ACGGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAA

25928  GAACATCTATTTTGTATGCTAGGTTTAATGGAGAAATAGTCATGGAGAAGTACGATTGGC
TTAAAAAAAAGTATCATCTCCTGGTGATAAACTGGCGGGAATTTTGCAAGACCTGTGTGT
CCAGGTCCCTCTCTGTGACCCTGCATCTTTGGAGATGAGAATGTTCCTTCCTCCGGGCAT
TGGGAGGGCACCTCTCGAATGAGCCTCATGTCCTGCTTCAGGGAAGAAGGGCAGGGGAAG
GTCAAAGAGTAACCTTCCGCTTCTGTGGTTTTCTCAAATCCCTTCAGCTTAAAAAAAATT
```

FIGURE 3

```
        [A,T]
        TTTTTTTGAGACGGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCTCAG
        CTCACTGCAACCTCTGCCTCCCAGGTTCAAATGNNNNNNNNNNNNNNNNNNNNNNNNNN
        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGGAGGCCGAGGCGGGC
        GGCTCATGAGGTCAGGAGATCGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTAC

31637   GCAGATAGTACAGAAAGTACCTAAAACCCCATGCCACCCTCCCCCACCCACCATCTTTCC
        CTGCCTATTATCCACATCTTGCATTAGTGGGGGATATTTGAGGTTTTGGGGGATTTCTCT
        TTTTTTTGTTTTCGTTTTTGTTTTTCTTTTTTGAGACAAAGTCTCGCTCTGTCGCCCAGG
        CTGGAGCGCAGTGGCGTGATCTCGGCTTACTGCAAGCTCCTCCTCCCGGGTTCATGCCAT
        TCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCCACCACCACGCCTGGCTA
        [T,-]
        TTTTTTTTTTTTTTTTTTTTTGTATTTTTAGTAGAGACGGGGTTTCACTGTGTTAGCCA
        GGATGGTGTTGATCTCCTGACCTCGTGATCCACCCGCCTCGGCCTCCCAAAGTGCTGGGA
        TTACAGGCGTGAGCCACCGCGCCTGGCCTTTGTTTTGGGGAATTTCTTTAGAGACAGGGT
        CTCCCTCTGTGATCCAGGCTACAGTGCAGTGGTGCTGCATAGCTCACTGCAGCCTCAAAC
        TCCTGGGCTCAAGCAATCTCCAGCCCCAGCCTCCCAAGTAGCTAGGACCACAGGTGTGCA

36513   GGTAACAAGCATAAGTCTCTTGCTGGATTGTGGTGAGGGTTAAATGAGTTAAAGTTGATG
        AAGCACACCGAACAGTGCCTGGCACATAATACGCACTAAAAAAAAAGTGTTTAATAAAAAA
        ATTGGACAAAATAATGGAGTGAGAGGGCTGGGGCGGGCTGAGGGAGGAGAGAGTTGAGAC
        CCAAGGGTGCTGGGGCAGTCCGGGCCAGGAGCCTGGCTTAGGCAGGTGCAGGCAGGGGCT
        GCACCTGGGGAGCCCAGCTGGGATGATCACTGTGGGCATCTCTCCACTGCAGGCTTTCAA
        [C,T]
        GTCCACCCAAATCTGTAATTCCTGACAAGTGGAAAAAGTGATTTGCCGTATGGCAGGTAG
        AGCATAATCCTGTTTTTGTAAAATATATAATATGCCTTGTATATCATTTCTGCGTCTTGA
        ACAGACATTGGCTGTAAGTGTGTACAGGACCGGGAAAAACATCATGGGTTGGGGAGAAAG
        GCTGGAAGCCCGTTTGCCAGCCATGTGCTGTGGTGATGCCTGTGCCGTGAGGCTGGGCAG
        GCCGCTGTGGTGCTGCTGTTTCCAGCTTGTATTTGCATTTTGTTTCACTTTGACGCTGTG

38025   CACCAGGCACCCCTGGGGGTTCCCAGGCTCTGGGATTTTACACCGATCCTGTTCTGGGGA
        TTCCCTCAGAATTTCGTGACTGCCATGTTGCCCAGTGGCTCCGCTGTGGCCCACTAACAA
        GAACAGCAAACGTTTATCGAGCATTTGCTACGGGCTCATTTGCTATTGAGCGCTGTTATT
        AACGGCTTCATGTGTATTTTCTTGTTTAATCAATAGCCGTTTAATTCCCCAATGGTTATT
        GAATGGCTCAGATCCCACAAGCCAAAGACACGGCACTGAGCAAGCCAGACCTGATCCCCA
        [A,C]
        CCCCTGGAGCGCTCCCCGCGGGGGCAGCGCTCCCCCAGCCTCATGTTTTCAGATGAAGAA
        GCCAAGGACTCTGAGGTCAGAGAATAGAAGGAGATTGCACAGCGCCACTACATGGCAGAC
        AGCTAACATCTAGACACAAGCATGTCCAGGCTCAAGACAGACTCAGCCACCGAGCTGTCC
        CTAGCCCTGGCCCGGTGGCTGCCTCATTTCTGGTTCCAGCTCTGACTTGAGCCCACCTGG
        AGCTCAGGTAGCCTGTATGATGGAGCGATCCCACGGGACTCCCCTTTGAATTGAAGGTTC

38068   CGATCCTGTTCTGGGGATTCCCTCAGAATTTCGTGACTGCCATGTTGCCCAGTGGCTCCG
        CTGTGGCCCACTAACAAGAACAGCAAACGTTTATCGAGCATTTGCTACGGGCTCATTTGC
        TATTGAGCGCTGTTATTAACGGCTTCATGTGTATTTTCTTGTTTAATCAATAGCCGTTTA
        ATTCCCCAATGGTTATTGAATGGCTCAGATCCCACAAGCCAAAGACACGGCACTGAGCAA
        GCCAGACCTGATCCCCAACCCCTGGAGCGCTCCCCGCGGGGGCAGCGCTCCCCCAGCCTC
        [A,G]
        TGTTTTCAGATGAAGAAGCCAAGGACTCTGAGGTCAGAGAATAGAAGGAGATTGCACAGC
        GCCACTACATGGCAGACAGCTAACATCTAGACACAAGCATGTCCAGGCTCAAGACAGACT
        CAGCCACCGAGCTGTCCCTAGCCCTGGCCCGGTGGCTGCCTCATTTCTGGTTCCAGCTCT
        GACTTGAGCCCACCTGGAGCTCAGGTAGCCTGTATGATGGAGCGATCCCACGGGACTCCC
        CTTTGAATTGAAGGTTCTTTGATGGGTGGTGGCTCCCTGCTCTGCCACTCAAGTCAGTGA

42787   GACAGTCCCTGGAAAGGCAGCTTTGTGTTCTGCTTTTCATTTTTTAAACCCAACTCTTGG
        CCGGGCGTGGTGGCTCACGCCTGTAATCCCAACACTTTGGGAGGCCGAGGCAGGTCCATC
        ACCTGAGGTCAGGAGTTCAAGACCAGCCTGACCAACATGGTAAAACCCTGTCTCTACTAA
        AGATACAAAATTTATCCAGGCGTGGTGGCGTGCACCTGTAATCCCAGCTACTCGGGAGGC
        TGAGGCAGGAGAATCACTTGAACCCAGAAGGCGGAGGTTGCAGTGAGCCGAGATCGTGCC
        [A,G]
        TTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCGTCTCAAAAAAATAAATAAATAAATA
        AAATAAACCCAACTCTTTGGGGACACCCAGCTCTTTGGGGACAGACATTTTGTTAAGTAC
        AGTTCACACACCTTCAACTGCATCTCCCAGGCCCCCTGAGCTACTGCTTCCCACCAAAAA
        GCGGGCATGCACGATTCCAGACCACATCAGCCTCCACTGAGAAGTGCCGGTGTTGGGAAG
        GGGCCAGACTATGCTGAAGAGTTGGGGGGTGGTGGGGTCCACTGCCAAGGCAGGTGGTTG

43423   CTGCTCCACTGCTGGGTGGGGCGGAGAGGGCAGGAGGGGCATCTCGGCCTTGGCCAGGGT
        CACAGGAGGCTTGGGGAGCCCTGCAGGTGCTTGGGAGAGAGTCACGCACAGCAACGCCTT
        CCCAGCAGCACTAGGCAGAATCGGGGTTCCCTCTTCTCTTGCCTAAGAAGGCATTTCCTC
        TGCGAGCTTTCTCGCCATCATATTAATTCACTCAACAGATGTTTTTTCTCAGACCTTGCA
```

FIGURE 3

```
         GGGATGAATTAACACATATTTCTAATGGGCACTGCCTTGGTTTCTTGGAGCCTCCATAAC
         [A,T]
         AAGTACCACTAACTGGGTGGAGGTCAGGAGTCTGAAATCAGGGTGCAGCTGGCTTCGCTG
         TGGAACCTGCAGAGGGGAATCTGTCCCATACCCTCCGCCCTCCGCTCGCTCCTGGTGGGG
         CCGGCAGCCTCTGGCGTTCCTGGGCTTGCAGCTGCGTCACTCCAGTCTCTGCCTTGTCTT
         GCGGCTGTCCGCTCCCTGAGCATCTGTCTTACACGCTGTCTCCCCGTGTCTCTTCCTACA
         GGGACACCAATCACATCGGATAAGGGCCCGCCCTCCTGCAGCATGACCTCAGCTTAACTT

43752    AGTCTGAAATCAGGGTGCAGCTGGCTTCGCTGTGGAACCTGCAGAGGGGAATCTGTCCCA
         TACCCTCCGCCCTCCGCTCGCTCCTGGTGGGGCCGGCAGCCTCTGGCGTTCCTGGGCTTG
         CAGCTGCGTCACTCCAGTCTCTGCCTTGTCTTGCGGCTGTCCGCTCCCTGAGCATCTGTC
         TTACACGCTGTCTCCCCGTGTCTCTTCCTACAGGGACACCAATCACATCGGATAAGGGCC
         CGCCCTCCTGCAGCATGACCTCAGCTTAACTTACATCCTCATCCACATCCACAAAACCCC
         [A,G]
         TTTCCATGTAAGGTCACATTCATAGGCACTGGGGGTCAGGATGTCAACATATCTTTTCGG
         GAGACACGATTTGGCCCACAGCAGGCACCTATGACATGCTGGACACAAGCAGGGTGCCAG
         GGTCCAAGAGCATGGAGGGGCTGACGCTCCCGTAGTGGGAGAGCAGAAATCAACAAACCC
         AGTACCCCCAAGTGCCCACAGGTGCACGGTGAGAAAGAAACCAGGGCAAGGCCAGGCACG
         GTGGCTCACGCCTGTAACCCCAGCACTTTGGGAGGCTGAGGCTGGCGGATTGCTTGAGCG

44151    CTGGACACAAGCAGGGTGCCAGGGTCCAAGAGCATGGAGGGGCTGACGCTCCCGTAGTGG
         GAGAGCAGAAATCAACAAACCCAGTACCCCCAAGTGCCCACAGGTGCACGGTGAGAAAGA
         AACCAGGGCAAGGCCAGGCACGGTGGCTCACGCCTGTAACCCCAGCACTTTGGGAGGCTG
         AGGCTGGCGGATTGCTTGAGCGCAGGAGTTCAAGACCAGCTTGGGCCACATGGTGAAACC
         CCATCACGACAAAAAATACAAAAATTAGCCAGGCATGGTGGTGTGCGCCTGTGGTCTCAG
         [-,T,A]
         TACCTGGGAGGCTGAGGTGGGAGGATTGCTTGAGCCCAGGAAGTCAAGCCTGCAGTGAGT
         TATGACTGTGCTACTGCACTCCAGCCTGGGCAACAGAGCGAGACCCTATCTCAAAACCCA
         GGGCAATATGGGAGTGGAGCTGGGAACCAACCTGAGACCTGGACCCAGGGGTCAGGGAAG
         GCCCTCCAGGGAGGGGGGCTTTGAAGGAGAAACCTGAAGGGGGAACGGGGGCACGGCATT
         TGTGGCAGGGGGACAGCAGGCAGGACAGGTGCAGTGCCTTCTAAGGACTCAATGTGATGT

44443    GGTCTCAGATACCTGGGAGGCTGAGGTGGGAGGATTGCTTGAGCCCAGGAAGTCAAGCCT
         GCAGTGAGTTATGACTGTGCTACTGCACTCCAGCCTGGGCAACAGAGCGAGACCCTATCT
         CAAAACCCAGGGCAATATGGGAGTGGAGCTGGGAACCAACCTGAGACCTGGACCCAGGGG
         TCAGGGAAGGCCCTCCAGGGAGGGGGGCTTTGAAGGAGAAACCTGAAGGGGGAACGGGGG
         CACGGCATTTGTGGCAGGGGGACAGCAGGCAGGACAGGTGCAGTGCCTTCTAAGGACTCA
         [A,G]
         TGTGATGTGGCTGTGGTGGGTGAGTAAGATTCCAAAGAAGAGCAGGCAGGGCCAGATCAC
         GCCAGCCCTGCAGCCCCGGGAGGAAGTTGGTTTTGGTTGAAGTTTGGTAGGAGGGCTACA
         GGCAGGGGAGAGGTGTGATCTGATTTCCGGTGTTTGGAAATCTGAGTTGGGGGTGGCGGT
         GCTGGAGGCCTGCAGGAGACCGGGCGGGCCCTACCAAGACTGACAGCAGCCTGGCCGGGT
         TCTGGCAGGGACAGGGGCTGTGATGTGGGCCGCCTGGAGGTGGTCTTTGGTGGCAGTGTT

44644    GGGGGGCTTTGAAGGAGAAACCTGAAGGGGGAACGGGGGCACGGCATTTGTGGCAGGGGG
         ACAGCAGGCAGGACAGGTGCAGTGCCTTCTAAGGACTCAATGTGATGTGGCTGTGGTGGG
         TGAGTAAGATTCCAAAGAAGAGCAGGCAGGGCCAGATCACGCCAGCCCTGCAGCCCCGGG
         AGGAAGTTGGTTTTGGTTGAAGTTTGGTAGGAGGGCTACAGGCAGGGGAGAGGTGTGATC
         TGATTTCCGGTGTTTGGAAATCTGAGTTGGGGGTGGCGGTGCTGGAGGCCTGCAGGAGAC
         [C,T]
         GGGCGGGCCCTACCAAGACTGACAGCAGCCTGGCCGGGTTCTGGCAGGGACAGGGGCTGT
         GATGTGGGCCGCCTGGAGGTGGTCTTTGGTGGCAGTGTTGCTGAGTAGCTGATGGGAGCA
         GGCGGTGGGAGGCATCATGGAGGACTCCCAGGCATCTGGCTTTGAGACACTGGGTGGATG
         GAAATTTTGCCTTGTCAAATGGGGACAGAGGTGAGGGGATTTTCTGGAGGGGAATTTTTTT
         TTTAATTTCTTAATTTTTTATTTTTATATTTTCATACAGCAAAGTGGACCTTTTGGGGGT

44888    TTCCGGTGTTTGGAAATCTGAGTTGGGGGTGGCGGTGCTGGAGGCCTGCAGGAGACCGGG
         CGGGCCCTACCAAGACTGACAGCAGCCTGGCCGGGTTCTGGCAGGGACAGGGGCTGTGAT
         GTGGGCCGCCTGGAGGTGGTCTTTGGTGGCAGTGTTGCTGAGTAGCTGATGGGAGCAGGC
         GGTGGGAGGCATCATGGAGGACTCCCAGGCATCTGGCTTTGAGACACTGGGTGGATGGAA
         ATTTTGCCTTGTCAAATGGGGACAGAGGTGAGGGGATTTTCTGGAGGGGAATTTTTTTTT
         [A,T]
         ATTTCTTAATTTTTTATTTTTATATTTTCATACAGCAAAGTGGACCTTTTGGGGGTACAA
         TTCTAGGAATTTTAGCACGTGGATAGAGTCATGCAACCACCACCACAGCCTGGGCCCACA
         ACAGCCCGTCTCTCCAAGAACTCCCGTCCCGAGACCCCTGGAAGCCCCCATCCTGCTCCC
         TGAAAATCTGCTGTAAGTGGAGTCGTGCGGCAGGTGGCTTTTTCAAGGGGTCTCCTTCGT
         TCTCACTGTGCCTCTGAGGCTTATGGCGTGGCCGCCTGCGTGGCTCGGGCTCCTCACTCC

45250    TCTAGGAATTTTAGCACGTGGATAGAGTCATGCAACCACCACCACAGCCTGGGCCCACAA
         CAGCCCGTCTCTCCAAGAACTCCCGTCCCGAGACCCCTGGAAGCCCCCATCCTGCTCCCT
         GAAAATCTGCTGTAAGTGGAGTCGTGCGGCAGGTGGCTTTTTCAAGGGGTCTCCTTCGTT
```

FIGURE 3

```
          CTCACTGTGCCTCTGAGGCTTATGGCGTGGCCGCCTGCGTGGCTCGGGCTCCTCACTCCT
          GAGTAGGACTGTGCCGTGTGCACTGCACACCGTGGTCTGTTTACCCATTCTCCGTGGAAG
          [A,G]
          ACGTTTCGGTGGCTCCCAGTCTGGGGCGATTATGAGTAGAGCTGCTATAAACATTCGTGG
          AAAGGTTTTTGGGTGAACATAAGTTTTCGCTTCTCTTGGGTAAACACCCAGGTGTGGACG
          GCTGGGTCGTCTGGTGAGCGTGTTTAACTCTATGAGAAACTGCCAGGCCGTCTTCCAGCG
          CGGCTGTGCACTGGAGGAAGCCTGACTTCTATTTTGGACCTGTTGGTTGTGAGCGTTTTG
          GAGGCATCCGGAGCATTTGGGCCTCTGAGTGTGAAGCGCAGGCGAGACTGGGCGCGGGGA

48665     TTCCTGCAGGGGAAGAAGCAACCTTGCTGAAGTCAGCTAGGCTGCTTAGGCTGGGGGTGC
          CTTCGCAGGCCCTGGTCATACCCACCGTGTCACCCCTAAGCCGCACACCTCAAGCCTCCC
          AGCACCCGCCCGTCAGCCTCTGTGTTGTGTGCAAGGAAGCTGCCTCTGGCTTTGTAATGG
          GTAATAGGATTTATCAATAGACTGAGGAGGTGAGGTATGTTAAAGCACTTAATAGAAAAG
          GGCTTCGCACAGAAGCCCAAATTTGATTTAGCCAATGAACTCAATTGCCGGCCTGATTGC
          [A,G]
          TTCCAGGAGGCGCAGCAGCCAGCATTTGTCCATGTTACCCTGGAAAAGCCAGGCTGCGCC
          AGGCGACCAACCCAGACCACCCAGACCTCCCCTCTGCCCCCACGGCTCTGTTGTTGGTTC
          CTCGCTTTCTCAGGATCTCAGGTTTGAATGGCAGTCCTTTGACCCAAACAGTCCCAAGTT
          CTCCAGCATCCAACAGCCTCCTTCCCTCTAGTGCCCAAGGCTTCCCCATCCCATGAATTA
          GCTAGAAGTGCAGTTTGCTACCATGTTCCTTTCACCATGATCCTCAAATCCGTGTGTCCC

49661     AATCCTGGAAACTCAAGAGGTTGAGGCAGGAGAATCACTTGAACTTGGGAGGCGGAGGTT
          GCAGTGAGCTGAGATCGTGCCACTGCACTCCAGCCTCCAGCCTGGGCGATGACAGAGTAA
          GACTCCATCTCAAAAAAAAAAAAAAAAGAAGGACTTGAGAGGATTTCCATTAGCCATCAC
          CAACCTTGAGGCCGTCCATCGTCCATCCATTGGCCACGGCCACCCACCGGCTGGGGAGGA
          GAAACCTCTTGGCCATGTAAGCCCGCAGCCCCCTCCCAGGGCAGCATCATTCCTCACTGC
          [G,A]
          GTCCAGGTCCCCAACCCCAGCTCTGTGAGTCCTCAGAAAGCGAGGGCCCTTGTTGCTTCA
          GCCTCCTCCTTCATCCTGCAGAACACGCCAGCCTCCCTTGTGAGCATCTCCAAGCCACCG
          TAAGATCTGGGATTTGCTTTAGGTTATTTTGCCCATAACCCAGAGGCTCATGAAAAATGT
          TTTTCCCAAAACAAAAGAACACTCTTCACCCAAAGATAAATGCGCTATCTGGCAAGAGAA
          ATTGGAAAACATTGCTGGCTGTGTTAAACTAGTAGTCTAACTTTAGCCCCCAAGGTACCA

50012     TGTTGCTTCAGCCTCCTCCTTCATCCTGCAGAACACGCCAGCCTCCCTTGTGAGCATCTC
          CAAGCCACCGTAAGATCTGGGATTTGCTTTAGGTTATTTTGCCCATAACCCAGAGGCTCA
          TGAAAAATGTTTTTCCCAAAACAAAAGAACACTCTTCACCCAAAGATAAATGCGCTATCT
          GGCAAGAGAAATTGGAAAACATTGCTGGCTGTGTTAAACTAGTAGTCTAACTTTAGCCCC
          CAAGGTACCAGAGCTCTGCGAGGCCAGTCTCAGTATACACGACATGTAATAATGTGGGGA
          [T,C]
          GGGTGGTAACATACCAGGAAAGAAGGACCAGGCAATGTGATTAATGACCAGGAACCCCCA
          TGGTCCTGCAAAGAGGTGCTTTGAGACAAGTTGGGGACTAGCTCTCCCAGCCCAGCACCT
          GCCCACCCCAGTTCAGAGCCATTGCCGTGAACACTCTGAAATCCCCGTGGGGCTTTGCCT
          TTGCAGAGAGCCAGCCCTGGGGCCTCCTCCTCCCTGCCCAGCTCCAGCACCATCCCTGGC
          TCGCTCACCCAAACTCACCGCTTCCTCCGTCATTCCCCGCAAGGAGTGGATGACATCACT

51203     TGAATTTCAGATAAACAAGGGAATCGTTGTTTTTTAGGGTAAGTATGTCCCAAATATTTC
          ATGGAATATACTTGTACTAACAAAATCATTAGTTATTTATCTAAAATTCAAAAAAAAATT
          TTTTTTTTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCTGAAGGGCAGTGCAGTGGCA
          CAATCACGGCTCACTGCAGCCTCAACCTCCTGGACTCAAGCGATCCTCCTGCCTCAGCCT
          CCGAAGTAGCTAAGATGACAGGTGCTCACCACCATGCCCTGATAAATTTTGTGTTTTTTT
          [A,T]
          AATTTTTTTGTAGAGATTGGGGGGGGGGTCTCACTTTCTTGCCCAGGCTGGTCTCAAACT
          CCTGTTCTCAGGTAATCCTCCTGCCTGGGCCTCCCAGAGTGCCCAGATTACAAGCATGAG
          CCACTGCACCAGGCCTAAAATTCAAATGTAACTCAGTGTCCTGTATTTTTATTTGGGAAA
          TCTGGCAACCATTGTGGCATGGGGCTACTGTGGGGAAATGACTCCAAGAGGCCAGTGGGG
          GCCAGGCATGGTGGCTTATACCTGTAATCCCAGCATTTTGGGAGGCCGAGGTGGGCGGAT

51222     GGAATCGTTGTTTTTTAGGGTAAGTATGTCCCAAATATTTCATGGAATATACTTGTACTA
          ACAAAATCATTAGTTATTTATCTAAAATTCAAAAAAAAATTTTTTTTTTTTAGAGACAGG
          GTCTTGCTCTGTCACCCAGGCTGAAGGGCAGTGCAGTGGCACAATCACGGCTCACTGCAG
          CCTCAACCTCCTGGACTCAAGCGATCCTCCTGCCTCAGCCTCCGAAGTAGCTAAGATGAC
          AGGTGCTCACCACCATGCCCTGATAAATTTTGTGTTTTTTTAAATTTTTTTGTAGAGATT
          [G,-]
          GGGGGGGGGTCTCACTTTCTTGCCCAGGCTGGTCTCAAACTCCTGTTCTCAGGTAATCCT
          CCTGCCTGGGCCTCCCAGAGTGCCCAGATTACAAGCATGAGCCACTGCACCAGGCCTAAA
          ATTCAAATGTAACTCAGTGTCCTGTATTTTTATTTGGGAAATCTGGCAACCATTGTGGCA
          TGGGGCTACTGTGGGGAAATGACTCCAAGAGGCCAGTGGGGGCCAGGCATGGTGGCTTAT
          ACCTGTAATCCCAGCATTTTGGGAGGCCGAGGTGGGCGGATCATTTGAGGTCAGGAGTTT

51230     TGTTTTTTAGGGTAAGTATGTCCCAAATATTTCATGGAATATACTTGTACTAACAAAATC
          ATTAGTTATTTATCTAAAATTCAAAAAAAAATTTTTTTTTTTTAGAGACAGGGTCTTGCT
```

FIGURE 3

```
                CTGTCACCCAGGCTGAAGGGCAGTGCAGTGGCACAATCACGGCTCACTGCAGCCTCAACC
                TCCTGGACTCAAGCGATCCTCCTGCCTCAGCCTCCGAAGTAGCTAAGATGACAGGTGCTC
                ACCACCATGCCCTGATAAATTTTGTGTTTTTTTAAATTTTTTTGTAGAGATTGGGGGGGG
                [G,C]
                GTCTCACTTTCTTGCCCAGGCTGGTCTCAAACTCCTGTTCTCAGGTAATCCTCCTGCCTG
                GGCCTCCCAGAGTGCCCAGATTACAAGCATGAGCCACTGCACCAGGCCTAAAATTCAAAT
                GTAACTCAGTGTCCTGTATTTTTATTTGGGAAATCTGGCAACCATTGTGGCATGGGGCTA
                CTGTGGGGAAATGACTCCAAGAGGCCAGTGGGGGCCAGGCATGGTGGCTTATACCTGTAA
                TCCCAGCATTTTGGGAGGCCGAGGTGGGCGGATCATTTGAGGTCAGGAGTTTGAGACCAG

52864           TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCCCATGGAGCGTTCTTGCTGCCATC
                TGAAGAGCTAAAGCCGTGGGATTCTGCAGTGGGGCAGGGGTGGAGAGGGAACAGCCCTGG
                TGGCCTGGGAGGGGTCCACTGGCCTCATGGGCTGGCCAGCATGCCCTATCTCTATGTTCA
                TTATTAAATCATCCTTTCTGATGGATGAGCTGAAAAGTGCTGTGTCGGGGGATCAGGTCG
                CCCAGGTGATGGTTTTCTGGTGAGATTGATGGTTCCTGAGGGTCAAATTCAGAGAGCGAT
                [C,T]
                GCTTGGGAAAATTGATCCCACAGAAGAAGGGGAAATATCTGGGCTGGAGTGCAGCAGGAG
                CCCAGACTGCCCGCCCAGGTCCAGATCCAGGCAGAGGCTGGTGCCAGGAGGGCACGTGGC
                AAAGGGGGCTTCTTTCTTCTTCTGTGGCTGCATTTCCTTGGCTGTGGCTGGGCCACAGAA
                TTGGTAAGCAGGGCCATGGGCAGCAGGCATGGGGGCATCTGGCTTAGTGGCCCCTTTCCT
                GGCCTTCTTTGTCCCCATGGAATGGACGCGAAGCCAGCACTGAGAGAGACGCAGGCAGCA

53973           TAGTATACTCTGCATAAGGTTCCTGTGAAGAGGTGATAACTTCACTTGTTGAAAGCTCTG
                AGGATCCCTGGCTCTAGGAGCTCCTGAGTATAAGAAATGGAAAACTGGCATGTCCTAGGC
                ACCTGTTTGTCTCATGCAAGAGCTTTATAGACCGTCTCCCACTTGATCCTTCCAACAACT
                CCATGAGGCAGGTGTTACTGCCATCCTTATTTTACAGATTAGAAAACAGGCCAAGTGAAA
                TTCAGCAACTTTCCTGAGGTCACACAGTCGTTGGCAGAACATCCCAGGCCAGTGGGGCCC
                [A,G]
                GGGGATGGGGTTACCATCTCTCAGTCACCCAGAAAGACAAAAGGAAGGCACACAAACCCC
                AGTGGGCACCGCTCTGTGCCAGGGCTTTTGTTTCACCCCAACCGTGATGAGTGTGGGTGA
                TGTGGCCGGTGTGGTCCATATCGCTCCTCAGGGCCAGTGTCACTGGTGTGCGTGGCAGCC
                TGTGGCATACACAAGATATTGTTTCCCGCTCTTCAGGTGAGAAAAGCGAGACTCAGGTTG
                GCTAAGAATTTGTGCAGGCCACACAGCTGTGTGCACCATGAGCTGGGCAGAAGCCCTGCC

59408           GGCACGCGACTGTGGTCCCAGCTACTCGGGAGGCTGAGGAGGGAGGATTGCTTGAGCCCG
                AGAGGCAGAAGTTGCAGTGAGCCGAGATTGTGCCACGGCACTCCAGCCTGGGTGACAGAG
                TGAGACTCGGTCTCAAAAAGCAAAAAGACTCCATTGCCTGGATTTACCATAATTTCCCTA
                ACCATTCTGCAGGATAATTCGAGGACAGAGAGAAGACCCCAAGGAGGAAGGCAGCCCTGG
                TTACCAAAGCTGGCAGATGGGGGTTAGCTGGAAGCCCCAGGGCTGGTGGACGCAGGGGCC
                [A,G]
                CTGTTTCCCCACCCAGATCCTTGCTCTGGAAGGCGGCCCCCAGGGGACCCTTCATTCCCA
                CTCAGACAGGGACAGAGGCGGGGACAGAGCCGAGGGAGGAGGGCTCAGATGAAGCACCTGG
                CAGGACTGAGATATGAGGGAGGCCGGATGCGAGGAGGGAGCTCTGCAGCCTGTGGTGTCC
                AGGAGGATTTGGGGAGTGTGAGGTGAGAAAACAAAAAGCGTCACCCCTCCCAGTGGAAGG
                GGAGCATGAAGAGAGAAGAAAATGCCAGTTACACATCCTCAAAACAATCTCTGAATGGAC

60143           CAGGTGCCCTGCCCTGGTCCAGGCCTCTATTCGCAGGATCTCACTGATCTCTCCCTGCCC
                AGGCTTTTGTCTCCCAGACAAACCAGTCTCCATGAAGCAGCCACAGCGTTATAAAATATG
                GCCCAGAGCAGGACACGCCTAACCAGTCAAAGGCTTGCAGAGTAAAATCCCGCCTCCGTC
                TCCTGGTCTGGCCTCCTCCACCCCCCAGCCTCATTGCAGACCTCACGGGTGGAACTTCTT
                TCAGCCCTAGGGCTTTCCTCATGCTCTTGCCTCTTTCTAGAAGGTTCTCCCCTTCTGGCC
                [C,T]
                GGCAACTCTTGATCCTCCTTCAGGTCTTAGTTTAAAACTGTTCTTCCTGGAAACTGTGGA
                CTCCCGTGATGCCTTCTTCTTTGCACAACCCGTGGGCTCTTTTAACAGTTGAACACTGGA
                AGAGGTGTGTGATTCGTATTGTTATAATTAATAGACTTTATTTTTAGAGCAGTTTTAAGT
                TTACAGAAAATTGAGCAGATAGTACAGAAAGTTCCCATTTCCCCACCCCCCTGCACAGTT
                TCCCCTATTTTGGAGTATATGTGTTACAATGATGAGCTAACACTGATACAGTGTCATTCA

60158           GGTCCAGGCCTCTATTCGCAGGATCTCACTGATCTCTCCCTGCCCAGGCTTTTGTCTCCC
                AGACAAACCAGTCTCCATGAAGCAGCCACAGCGTTATAAAATATGGCCCAGAGCAGGACA
                CGCCTAACCAGTCAAAGGCTTGCAGAGTAAAATCCCGCCTCCGTCTCCTGGTCTGGCCTC
                CTCCACCCCCCAGCCTCATTGCAGACCTCACGGGTGGAACTTCTTTCAGCCCTAGGGCTT
                TCCTCATGCTCTTGCCTCTTTCTAGAAGGTTCTCCCCTTCTGGCCCGGCAACTCTTGATC
                [C,G]
                TCCTTCAGGTCTTAGTTTAAAACTGTTCTTCCTGGAAACTGTGGACTCCCGTGATGCCTT
                CTTCTTTGCACAACCCGTGGGCTCTTTTAACAGTTGAACACTGGAAGAGGTGTGTGATTC
                GTATTGTTATAATTAATAGACTTTATTTTTAGAGCAGTTTTAAGTTTACAGAAAATTGAG
                CAGATAGTACAGAAAGTTCCCATTTCCCCACCCCCCTGCACAGTTTCCCCTATTTTGGAG
                TATATGTGTTACAATGATGAGCTAACACTGATACAGTGTCATTCACTAGGGCCTGTTATT

60227           AGTCTCCATGAAGCAGCCACAGCGTTATAAAATATGGCCCAGAGCAGGACACGCCTAACC
```

FIGURE 3

```
        AGTCAAAGGCTTGCAGAGTAAAATCCCGCCTCCGTCTCCTGGTCTGGCCTCCTCCACCCC
        CCAGCCTCATTGCAGACCTCACGGGTGGAACTTCTTTCAGCCCTAGGGCTTTCCTCATGC
        TCTTGCCTCTTTCTAGAAGGTTCTCCCCTTCTGGCCCGGCAACTCTTGATCCTCCTTCAG
        GTCTTAGTTTAAAACTGTTCTTCCTGGAAACTGTGGACTCCCGTGATGCCTTCTTCTTTG
        [T,C]
        ACAACCCGTGGGCTCTTTTAACAGTTGAACACTGGAAGAGGTGTGTGATTCGTATTGTTA
        TAATTAATAGACTTTATTTTTAGAGCAGTTTTAAGTTTACAGAAAATTGAGCAGATAGTA
        CAGAAAGTTCCCATTTCCCCACCCCCCTGCACAGTTTCCCCTATTTTGGAGTATATGTGT
        TACAATGATGAGCTAACACTGATACAGTGTCATTCACTAGGGCCTGTTATTTACATTAGG
        ACTCACTCTGTGTTGGACACTTCAGTGGGTTTTGCCATATGCATAATGCCACATATCTAC

62304   GAGTCCAGATCTTGCGGATGGCATCCCCTGGTGTCGTTCAGGGTGATGCTCTGTCCTCTG
        TCTTTCCTGTGCGCTCACAGCTGGGTCTAGACGGCTTGATCTGATTCAGGTTCAGTTGCTT
        TGGCTTCGGGTGGGGGTGTGTTCTTCTAGAGACGCTCACTGTCTGCTGGCCTCTCAGTGG
        GAAGTTCTGGGCTGGTGCGGCTCAGTGCCTGGAGCTATGCATTCATCAGGGACTGCAAAT
        TCTCTCGTGAAATGCTTCTATAAAAAGAAGCTTCCCCTCATCAACTTTCGGTTACCCCGA
        [G,A]
        GTACAGTTTATATAGGACAGGCAGGATAAATGCTTGATTCTTTCCCTTTATTTACCAGTT
        TTCAGAATAATGAGTCGCCTATAAATATTTGAATGAATGGATGAATAATACACCTGTGAA
        TGAACTGACGGGGGTGTGGGAGTTGGGGGTTCTATCTGTTGCCCTCAAGGGCGTGGCTGT
        GGGCCACAACCTGACAGCAGAGGTCCAGCCTGAAGCCAGGTGCCTTTCTACACAATGAAG
        TGCAGCCCATGGCCAAGTCTCTGTCTACACCAGGTGCTGGCAGCAGCCTCTGTCACCCAT

63410   CTACCTGCCGCTGCCCCATGGCTCTCCCAACGCCTGTACGTGCCATGCCCCGGGGCCACG
        AGCCCACATGGCTATAATCTTCCCTGCCCGTCAATCCCAGGAGCTGTGGATCATACACAC
        GCACACAAACCTGCACACAGGCACATACGTGTGCACATGCATGCAAACGTGCACACAGAA
        ACATACGAGCATGCATGCACAGGCATGGGCACACATATGAATGCAAAAACACGTGCATGC
        ACAGAAACACACGTGCGTGCATGCACCCCCACACACACACCTTGTTCTACAGCTCCCAAA
        [T,C]
        GCCAGGTCTCATAACAAGTCCCTCTAGCATACCTGCATCCTGCTGAATGCTAAGCTGCCC
        TTCCAGCCCTGGTCCACAGGGAAACCCGAGAGGAGCTGCTCAGCAGCATTCCTGCAACCC
        TTCGCCTTCTTGACCCTGAAGGCTGGCAGGTGGCCCCCATGGGATGGCAGGGAGTGATGG
        GGTGGGACCCCCACTTCAGGTGAGGGGATCAGTAGCATCATCCCTGATCATAAGCATCTC
        TGGGGTGTTTGGAGCGAGCCACCAAGGCCAAGTGCTGCGTGTTCAGCACTCACCTCCATC

64577   CTGCACCCCATCTGGGCATCCTCCCTCGACGGGTGAAAGATTTATGTGACTTTAGGCTGA
        TCAAAGCTGAGAGCCACGTGCAGAGGCCATGACAGGCATGTCACGCGTGTGCTACAGCAG
        GCGGTGAGACAGGCCTTTGAAGGGGTGACTTGGCAGGCACCCGACAGCACTCTCGGGTCT
        TCAAAGGCACAAAGAGCACTCCGGGGGCTGACAGCCCACTCCAGCCGCTGCCTCCTCCCA
        GCTTCCTGTCCCCATCTCTTTGTGGTCCAAACCCTTTCCTGAGGTCTCTCTTTAGTGGCC
        [G,C]
        TCACTCTCTCCCTGACTCACAAATTTGCTGCCCTTTCGTCTGTCCTGAGGGCCGCTCTCT
        CTTCCCTCCTCTCTGTCTCAGTTCTCTCCACCATCACACCCTCCTTTTTTTTTTTTTCTT
        AAAGATAAGTCTCTCACTCTGTCGTCCAGGCTGGAGTGCAGTGGCGCAGTCATGGCTCAC
        TGTAACCTCAACCTCCTGGGCTTGAGCGATCCTCCCACTTCAGCCTCCCAAGTAGCTGGG
        ACTGCAGGTGCACACCACCAGGCCCAGCTAATTTTCGTGTTTTTTTCAGAGAAAAGGTTT

65796   GCCACGTTCTTTGAACTTCCCTCCATCCGCTCCAAGTCTCTCCCAATGCCATCCTCAGCC
        CTGCAGCAGCCCTCACTCCCGATGCCTTCCCACCTCCTCACCACTCTGCCCCCACCTGGC
        CAGCCCATCACCCTCCAGGGCCCAACTTGGAGCCCCCAGGACCTCCCCGTGCCCTGCCTG
        ATGTCCCGCCTGTCCCCACAGAGCCTCACTTGGTCACCACCCAGTCCTGGCCCTTGCTTA
        CTGTGGCTGCACCCCGAGGTGTCCTCAGGGTCTAGCAGGTGGCTGCCCAGACATGGAGGT
        [A,G]
        GAGGAAGGAGTGGGTGGGGATGGGCTTGTCCTCCCAGGCCTCCCTGCCTGTCCTGCTGGC
        CACAGCCTTGGCTTGCCCAGGAGAAGCCCATGGGCCACACATCCCACTGCCAATCCCACA
        GCGTCCTTTCTCGGGAACACCGTGGGGAAAGCTGTGGCACCAGCTCCTTCCTTTTGCAAC
        TCTGATGAATCTCACCCAGGGATTTCAAGGCCCCTGGTCACACCAGGATCATAGGCCTCC
        CCCATCCCCTGGACACACAGAGACACACCTGGATTCAGGTCAGGCCTCGCCCACTCTCGG

65918   GCCCATCACCCTCCAGGGCCCAACTTGGAGCCCCCAGGACCTCCCCGTGCCCTGCCTGAT
        GTCCCGCCTGTCCCCACAGAGCCTCACTTGGTCACCACCCAGTCCTGGCCCTTGCTTACT
        GTGGCTGCACCCCGAGGTGTCCTCAGGGTCTAGCAGGTGGCTGCCCAGACATGGAGGTAG
        AGGAAGGAGTGGGTGGGGATGGGCTTGTCCTCCCAGGCCTCCCTGCCTGTCCTGCTGGCC
        ACAGCCTTGGCTTGCCCAGGAGAAGCCCATGGGCCACACATCCCACTGCCAATCCCACAG
        [C,T]
        GTCCTTTCTCGGGAACACCGTGGGGAAAGCTGTGGCACCAGCTCCTTCCTTTTGCAACTC
        TGATGAATCTCACCCAGGGATTTCAAGGCCCCTGGTCACACCAGGATCATAGGCCTCCCC
        CATCCCCTGGACACACAGAGACACACCTGGATTCAGGTCAGGCCTCGCCCACTCTCGGCT
        ATATTTCTCCCCAAGCCGTGTGTCCTCAGCTGTAGAATCAGGACCATAAGGAAGTTCCCT
        CATAGGGTTCTTGTGAGGACGGCACGATTTACGTAGGGGATGCTGACACCGTGCCTGGCA
```

FIGURE 3

66192
CACACATCCCACTGCCAATCCCACAGCGTCCTTTCTCGGGAACACCGTGGGGAAAGCTGT
GGCACCAGCTCCTTCCTTTTGCAACTCTGATGAATCTCACCCAGGGATTTCAAGGCCCCT
GGTCACACCAGGATCATAGGCCTCCCCCATCCCCTGGACACACAGAGACACACCTGGATT
CAGGTCAGGCCTCGCCCACTCTCGGCTATATTTCTCCCCAAGCCGTGTGTCCTCAGCTGT
AGAATCAGGACCATAAGGAAGTTCCCTCATAGGGTTCTTGTGAGGACGGCACGATTTACG
[T,C]
AGGGGATGCTGACACCGTGCCTGGCACGTGGGACGCACTCCACCCGCGGCAGCCGCTCCC
ATGGCTTCTCAGTGAGTTTTCCAGCCACACTGCACTTCTTAGACAGGAACACTCCATACG
ATGTCCCTGTCCTGCACTGGATGGCCCAAAAATCTGAAATAAGAGGAGGAGTGCGTGTGA
AGCTCCCAGTGGAGCGTTTGGCACCTGTCCAGCATGTCCCCAAGGGCAAGTCACGGCTCT
GAGATTCAGTGTCTCCTTCTGCAAAATGGGCCAATAGTGGTTCCTCCCTCCCAGGGCTGA

66209
ATCCCACAGCGTCCTTTCTCGGGAACACCGTGGGGAAAGCTGTGGCACCAGCTCCTTCCT
TTTGCAACTCTGATGAATCTCACCCAGGGATTTCAAGGCCCCTGGTCACACCAGGATCAT
AGGCCTCCCCCATCCCCTGGACACACAGAGACACACCTGGATTCAGGTCAGGCCTCGCCC
ACTCTCGGCTATATTTCTCCCCAAGCCGTGTGTCCTCAGCTGTAGAATCAGGACCATAAG
GAAGTTCCCTCATAGGGTTCTTGTGAGGACGGCACGATTTACGTAGGGGATGCTGACACC
[G,A]
TGCCTGGCACGTGGGACGCACTCCACCCGCGGCAGCCGCTCCCATGGCTTCTCAGTGAGT
TTTCCAGCCACACTGCACTTCTTAGACAGGAACACTCCATACGATGTCCCTGTCCTGCAC
TGGATGGCCCAAAAATCTGAAATAAGAGGAGGAGTGCGTGTGAAGCTCCCAGTGGAGCGT
TTGGCACCTGTCCAGCATGTCCCCAAGGGCAAGTCACGGCTCTGAGATTCAGTGTCTCCT
TCTGCAAAATGGGCCAATAGTGGTTCCTCCCTCCCAGGGCTGAAGTGAGGATGAAATGGG

66334
TCCCCCATCCCCTGGACACACAGAGACACACCTGGATTCAGGTCAGGCCTCGCCCACTCT
CGGCTATATTTCTCCCCAAGCCGTGTGTCCTCAGCTGTAGAATCAGGACCATAAGGAAGT
TCCCTCATAGGGTTCTTGTGAGGACGGCACGATTTACGTAGGGGATGCTGACACCGTGCC
TGGCACGTGGGACGCACTCCACCCGCGGCAGCCGCTCCCATGGCTTCTCAGTGAGTTTTC
CAGCCACACTGCACTTCTTAGACAGGAACACTCCATACGATGTCCCTGTCCTGCACTGGA
[T,G]
GGCCCAAAAATCTGAAATAAGAGGAGGAGTGCGTGTGAAGCTCCCAGTGGAGCGTTTGGC
ACCTGTCCAGCATGTCCCCAAGGGCAAGTCACGGCTCTGAGATTCAGTGTCTCCTTCTGC
AAAATGGGCCAATAGTGGTTCCTCCCTCCCAGGGCTGAAGTGAGGATGAAATGGGATAAT
CCACCCCCGTCCCCACACCCTGCAGGTCATCATCATTGCTAGCAGTTGTGTGGTGGAGCA
GGTGCTCTTGAGGGAGCGACACCTCCAGGTGCTCCCCTGCCCTGCTGGCCCCTCTGCAGG

66548
CTCCCATGGCTTCTCAGTGAGTTTTCCAGCCACACTGCACTTCTTAGACAGGAACACTCC
ATACGATGTCCCTGTCCTGCACTGGATGGCCCAAAAATCTGAAATAAGAGGAGGAGTGCG
TGTGAAGCTCCCAGTGGAGCGTTTGGCACCTGTCCAGCATGTCCCCAAGGGCAAGTCACG
GCTCTGAGATTCAGTGTCTCCTTCTGCAAAATGGGCCAATAGTGGTTCCTCCCTCCCAGG
GCTGAAGTGAGGATGAAATGGGATAATCCACCCCCGTCCCCACACCCTGCAGGTCATCAT
[C,G]
ATTGCTAGCAGTTGTGTGGTGGAGCAGGTGCTCTTGAGGGAGCGACACCTCCAGGTGCTC
CCCTGCCCTGCTGGCCCCTCTGCAGGGAGGTGACACCCAGGCCCCTTCCCCTGGGGCAGC
CAGCTCACGCCCGTCTCTCTCCCACAGGTGCCGCTGCAGGTTCCTTTGGCAGTAAGTACA
CGCCTGGGGAGGGTGGCCAGGGCCCCCACTGCACGAGCCTCTTTGCATGTCCTGGAAAAA
GCTGGAGAGAAAAAAGGGGCTTCAGTGTCCCCTCTGGGACTTGGGCCTATTCACTCCCTC

68155
TGTCCCCTCCACGTCCACTGCCTGGGCCCCCATGGCGCCCAGCACCCCACAGCCCACAGG
TGGGTGCCAGGGTACAGCGACCCCTGTCATCCCACCCTCTCCTGCTTCTAGCCTGGGTCC
CTGCCTCTCTTGGGGTGGGAGGGTCGGCAGCCCTGGGCAGAGAGCAGGGGCTTGGCTCTT
AGAATAGAGACGCTAGAACCCTAGAGCTGGGAGGCCACAGGCCAAAGGGGCTTGAGGACA
CCTGGGTCAACCTGTTCCTGAGCCCAGCCAGGGGATTCAGGGATCAGTTCAGCTTCCAAA
[G,A]
TCGTCTTCCTCCTGCCCTTCAAGCCATTGCTTGGAAGGGCTCCCAGACCATTGTGGCCAG
ACGGCTGCAGGAACTGAGAGGAAAGGTGCTGGGGGCAGCGAGCCCATCCTGACATGCAGC
CAAAGACTGGCCTTATCTCCCAATGGTGCTTCTGCCTCCGTGGTCCCTGGAGCCCCGCCC
ACACCCTGTCCCCACCTGGCCCCCAGGGCCTCTCTGTCCTTAGCCCCTCAGCAGCACACC
GGTGGGATGGATGGAGCAGGGTTAGCCCAGAAAGCAAATGTCTCTGATCAGCAGGGCAAA

69247
ATTTCTAAAGGGCAGAGTCCCCTCCATCAGCTCCTGCCTCGGCCTGTTGCTGGGTGGACA
CTCAGGCTCCCCAGACAGGGGCAAATGCTGAGAGAAAGACCTCCTCCTTCCTAGGCCATC
CAGAGCAGCTCCCCTGGGGGCAGCACACCCCACCTCTTTCTACATCCTTCCTTTTCTGCA
GGAGGCATTTACAGGAGGCAGGGGCTAGCCAAAAGATTGGAGGATTTCCGGGAAGCCTCC
TGACCCAGGAATCCTCTTTGGGGTGGAAGACATGGGTCACTCTGAGAATTCTGGACTTCA
[G,A]
ACATAGGTTGGCCCAGCCACAAGGGACCTGTGCTTTGCTGATGAGCCTGTGGTGGGCAGA
CAGAAGCAAAAACAGTGGTGGTGGGTGCTGTGCCTGTCTCCAAACAGGGGTTTGGCTGGG
AGGCCAGATACTCTCCATATCACATGTGCAAGTGCACACATGCACACACACACATGCATG
CACACACACAGGCATGCACACGCACATGTACACACACACACACACAGAGGAATCCATTTG
CAGAGCTGCTTCTGACTTGGTGCCAGGGCAGCCGTGGGAGGCTGGGCAGATTGTGCAAAG

FIGURE 3

```
69288   GCCTGTTGCTGGGTGGACACTCAGGCTCCCCAGACAGGGGCAAATGCTGAGAGAAAGACC
        TCCTCCTTCCTAGGCCATCCAGAGCAGCTCCCCTGGGGGCAGCACACCCCACCTCTTTCT
        ACATCCTTCCTTTTCTGCAGGAGGCATTTACAGGAGGCAGGGGCTAGCCAAAAGATTGGA
        GGATTTCCGGGAAGCCTCCTGACCCAGGAATCCTCTTTGGGGTGGAAGACATGGGTCACT
        CTGAGAATTCTGGACTTCAGACATAGGTTGGCCCAGCCACAAGGGACCTGTGCTTTGCTG
        [A,C]
        TGAGCCTGTGGTGGGCAGACAGAAGCAAAAACAGTGGTGGTGGGTGCTGTGCCTGTCTCC
        AAACAGGGGTTTGGCTGGGAGGCCAGATACTCTCCATATCACATGTGCAAGTGCACACAT
        GCACACACACACATGCATGCACACACACAGGCATGCACACGCACATGTACACACACACAC
        ACACAGAGGAATCCATTTGCAGAGCTGCTTCTGACTTGGTGCCAGGGCAGCCGTGGGAGG
        CTGGGCAGATTGTGCAAAGTTGGGAATTAAAGAGGAAAAGTCAGAGGCCAGAGTGGGAAA

70151   GGCTGTGCTCGTATCATTGTATCATCATATCATTGTATTCTGGGCTCACAGCTCCGTGAG
        ATGGAGGCTGTTATTTTCCTAGTCCCACAGGTGAGGGGATCGAGGCTTAGGAAGAAGCAG
        CTGGATTTTATGATATGTAAATTACACCTCAATCAAGCTGTTTCAGAAGAAAAAAGGGGC
        AGCTGCTCAAGGTCTCAGAATTATGGAGAGGCACGGGCAGGATTTGAACTCAGGGCTCGC
        CAACTCAGCCACCCAAAGCTATTGTCCTGAGGCCTCCAGGGGCTATGAGGTAGAGCTATC
        [-,T]
        TTTTTTTTTTTTTTTGAGATGGAGTTTCGCTCTTGTCGCTGAGGCTGGAGTGCAATGGAG
        CAATCTCAGCTCACTGCAACCTCCGCCCCCCCAGGTTCAAGCAATTCTCCTGCCTCAGCC
        TCCCGAGTAGCTGGGATTACAGGCACCTGTCACCATGTTCAGCTACTTTTTGTCTTTTTA
        GAGAGACAGGGTTTCACCATGTTGGTCAGGCTGGTGTTGAACTCCTGACCTCAAGTGATC
        CACCCGCCTCAGCCTCCCAAAGTGCTGGGATTCCAGGCGTGAGCCACCGCACCCGGCCAA

70166   ATTGTATCATCATATCATTGTATTCTGGGCTCACAGCTCCGTGAGATGGAGGCTGTTATT
        TTCCTAGTCCCACAGGTGAGGGGATCGAGGCTTAGGAAGAAGCAGCTGGATTTTATGATA
        TGTAAATTACACCTCAATCAAGCTGTTTCAGAAGAAAAAAGGGGCAGCTGCTCAAGGTCT
        CAGAATTATGGAGAGGCACGGGCAGGATTTGAACTCAGGGCTCGCCAACTCAGCCACCCA
        AAGCTATTGTCCTGAGGCCTCCAGGGGCTATGAGGTAGAGCTATCTTTTTTTTTTTTTTT
        [-,T]
        GAGATGGAGTTTCGCTCTTGTCGCTGAGGCTGGAGTGCAATGGAGCAATCTCAGCTCACT
        GCAACCTCCGCCCCCCCAGGTTCAAGCAATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGG
        ATTACAGGCACCTGTCACCATGTTCAGCTACTTTTTGTCTTTTTAGAGAGACAGGGTTTC
        ACCATGTTGGTCAGGCTGGTGTTGAACTCCTGACCTCAAGTGATCCACCCGCCTCAGCCT
        CCCAAAGTGCTGGGATTCCAGGCGTGAGCCACCGCACCCGGCCAAGTAGTGCTGTCTCCA

70568   AGCCTCCCGAGTAGCTGGGATTACAGGCACCTGTCACCATGTTCAGCTACTTTTTGTCTT
        TTTAGAGAGACAGGGTTTCACCATGTTGGTCAGGCTGGTGTTGAACTCCTGACCTCAAGT
        GATCCACCCGCCTCAGCCTCCCAAAGTGCTGGGATTCCAGGCGTGAGCCACCGCACCCGG
        CCAAGTAGTGCTGTCTCCAAGGCCTGGCTTGCAGGGCTTCCCAGTTCCAAAGGAGCAGAC
        CGGGCTTCCATGGGGCCTTGGCACAGCACACAGGCCATGGCGAGAACTTGCTTCCCACAC
        [A,G]
        CCTGAGTGTGTCCCTGGGCAGCCAAGCCAGGACTCCCTCCCTCCCCAAGACCCTGGTCCC
        TGAAAGATCCTGAATACCCCCGAGTGCCTCCCAACAGGTGCTTCGGGCTCTTTGAACAGA
        GTCCAGCTGGGCCTCTGAACTCCTGGGCCAGATGTTTCTCCCGCCTGCCAATGTCAAGCT
        GTCTGGAGGACAGCGCTGCGGGCGGAAAACGCCGCTGGAGACACTAATCCTTTCCTGGGC
        TGGGCCACGGAGGATGGAGGGAGACAGGCTCTGAAGCAAATGCCTTCAGGGCTGGCTTTC

70769   GCCTGGCTTGCAGGGCTTCCCAGTTCCAAAGGAGCAGACCGGGCTTCCATGGGGCCTTGG
        CACAGCACACAGGCCATGGCGAGAACTTGCTTCCCACACACCTGAGTGTGTCCCTGGGCA
        GCCAAGCCAGGACTCCCTCCCTCCCCAAGACCCTGGTCCCTGAAAGATCCTGAATACCCC
        CGAGTGCCTCCCAACAGGTGCTTCGGGCTCTTTGAACAGAGTCCAGCTGGGCCTCTGAAC
        TCCTGGGCCAGATGTTTCTCCCGCCTGCCAATGTCAAGCTGTCTGGAGGACAGCGCTGCG
        [G,A]
        GCGGAAAACGCCGCTGGAGACACTAATCCTTTCCTGGGCTGGGCCACGGAGGATGGAGGG
        AGACAGGCTCTGAAGCAAATGCCTTCAGGGCTGGCTTTCTCATGGCTCTAATTAAGCCCT
        TGCCAATTTGGGCCTGGCGGCCTCATCTTCCCACTGAACATCATATTAAAGTCAATTCAT
        GTCCAAAGCTCCCCGCTCCCAGCTGGAAAGTCTTCCGCACTTGTTAGCTGGTAGCTTTTC
        CTTTTCTTTCCCCACAGCCACCGTTGTGTATAATCCCTTCAAGAAGCGGAAAACAGCAGC

71191   GCCAATTTGGGCCTGGCGGCCTCATCTTCCCACTGAACATCATATTAAAGTCAATTCATG
        TCCAAAGCTCCCCGCTCCCAGCTGGAAAGTCTTCCGCACTTGTTAGCTGGTAGCTTTTCC
        TTTTCTTTCCCCACAGCCACCGTTGTGTATAATCCCTTCAAGAAGCGGAAAACAGCAGCG
        CTCCCCTGTCCCTCTGGGTTTGTCCTTTGAAATTTGGGCACAGGGCAGTTCTTTGCCAGC
        CCTGCCTGCCTGCCTTGCTGGCTGTGTGTCCCGTTAGTCTACGGGCTGAGCGTTGTGTCA
        [T,C]
        TGGTTCATGCTGGGGTCCCTGGTGAAAATGGGCCAGGCCAGGGGTCAGGAAGGTAGAAGG
        GCAGTGATCAGGGAAGCAGGTCAGATGCTGGGGAAGGCTCCGGTCCCTGGATTGCGGCTG
        GACAGGAAGGACACCTTCCAGGACACTTCTGGACACATGTAAGATCTTGGCCGGAACACA
        TGTCCCACTTCGCAGCCATTAGCCAGAGACATCAGCTCAGAGAGGTCTGGGCCCAGAGGC
```

FIGURE 3

```
          GGGACCTGGTCTAGCTCTGTCCTTCAGTCAGAACGGGGACGGCACAGGGAGTGTAGAAGG

71368     GCGCTCCCCTGTCCCTCTGGGTTTGTCCTTTGAAATTTGGGCACAGGGCAGTTCTTTGCC
          AGCCCTGCCTGCCTGCCTTGCTGGCTGTGTGTCCCGTTAGTCTACGGGCTGAGCGTTGTG
          TCATTGGTTCATGCTGGGGTCCCTGGTGAAAATGGGCCAGGCCAGGGGTCAGGAAGGTAG
          AAGGGCAGTGATCAGGGAAGCAGGTCAGATGCTGGGGAAGGCTCCGGTCCCTGGATTGCG
          GCTGGACAGGAAGGACACCTTCCAGGACACTTCTGGACACATGTAAGATCTTGGCCGGAA
          [G,C]
          ACATGTCCCACTTCGCAGCCATTAGCCAGAGACATCAGCTCAGAGAGGTCTGGGCCCAGA
          GGCGGGACCTGGTCTAGCTCTGTCCTTCAGTCAGAACGGGGACGGCACAGGGAGTGTAGA
          AGGGTCTCGCTGAAGAATATGCAGATTCTCAGGCATGGGTTCACCTCTCATCTATCGGGC
          TTTAAGTCTGCATGTGCCCTCCACAGGCTGAAATAGTGTAGATGCTGCCTATGTAGTAGA
          TTTGGACCCAATTCCTTTGGCCAGTGTAGACAGAGCCTCTCCTTATAGTGCTGCTGCTTC

71370     GCTCCCCTGTCCCTCTGGGTTTGTCCTTTGAAATTTGGGCACAGGGCAGTTCTTTGCCAG
          CCCTGCCTGCCTGCCTTGCTGGCTGTGTGTCCCGTTAGTCTACGGGCTGAGCGTTGTGTC
          ATTGGTTCATGCTGGGGTCCCTGGTGAAAATGGGCCAGGCCAGGGGTCAGGAAGGTAGAA
          GGGCAGTGATCAGGGAAGCAGGTCAGATGCTGGGGAAGGCTCCGGTCCCTGGATTGCGGC
          TGGACAGGAAGGACACCTTCCAGGACACTTCTGGACACATGTAAGATCTTGGCCGGAACA
          [G,C]
          ATGTCCCACTTCGCAGCCATTAGCCAGAGACATCAGCTCAGAGAGGTCTGGGCCCAGAGG
          CGGGACCTGGTCTAGCTCTGTCCTTCAGTCAGAACGGGGACGGCACAGGGAGTGTAGAAG
          GGTCTCGCTGAAGAATATGCAGATTCTCAGGCATGGGTTCACCTCTCATCTATCGGGCTT
          TAAGTCTGCATGTGCCCTCCACAGGCTGAAATAGTGTAGATGCTGCCTATGTAGTAGATT
          TGGACCCAATTCCTTTGGCCAGTGTAGACAGAGCCTCTCCTTATAGTGCTGCTGCTTCTA

71684     CAGCCATTAGCCAGAGACATCAGCTCAGAGAGGTCTGGGCCCAGAGGCGGGACCTGGTCT
          AGCTCTGTCCTTCAGTCAGAACGGGGACGGCACAGGGAGTCTAGAAGGGTCTCGCTGAAG
          AATATGCAGATTCTCAGGCATGGGTTCACCTCTCATCTATCGGGCTTTAAGTCTGCATGT
          GCCCTCCACAGGCTGAAATAGTGTAGATGCTGCCTATGTAGTAGATTTGGACCCAATTCC
          TTTGGCCAGTGTAGACAGAGCCTCTCCTTATAGTGCTGCTGCTTCTAAGGGGCCTGTGGG
          [T,G]
          TGCGGGGCTGTGATGCCTCAGTATGTACCCAGCTTCCCTCAGCACCACCCCCTCGCATAA
          CTTGGTTTCTTCTCTTCTTCCCCCCAAGAGTGGACCAGGCCATCTACGGCTGCCCCTCTC
          TCGAGCAGGTGGTCCCAGGTGGCCTCCCGTGCAGAAGGTATGGGGGGGCAAGGCCTGTGA
          TGGGCCTGAGACCCCGGGGAAGCGCCCTCTTAGACTCGTAGGCCCCTCCCTCTGTAGTGG
          AAGTAGCAGGTGTGCATGGTGGGGACCTGAGGTTGGAGGGGGGCCGCAGGAACCAACTGA

73463     GTCAATCTATCTAAAAACAGTTTGGCTTCCAAGAAGGTCTTAGCAGGGCGCGGGGGTGTC
          AGGGGTTACAGAAGTCATTTGGAGGATTAATCCAGCCAGATGTGTCCATGGTCTCAGAGA
          GGGGACCAAGGGCAGGGCTGATTTGCAAGCTTGGGATGTGCTGTGTTTCCTTCAGGAAGG
          GGCCCCACCTCCCTGGGCTCTTCGAGGAGAGGGGCTGTGTGATTTGAGGCCAGAGGGGCC
          TCTCCCTCCCTCACATCTGAGCAGGCGACAAGGCTGCCTGCCCTAGAGCTGGCCCAGGGC
          [G,C]
          GCTCGGAAGCCTTTGCTGGGCTCTTCCCTGGGCAGTGGGACCATGACAGACGAAAGAACC
          TGTTTCTCATCTCTCCAAGCTGTGGGCACCCCTGCCGCTGCCCCTGCCCCTGCCAAGGGC
          TACAAACTTTTCCAGCTCAAGCCCAAATCTCCTCAAGTGATGCCTATTGAAGAATTCCAA
          GGTAAGAGGATGGACCTGGGGCCCCATCAGCCCTCCCTGACACCTGTTCCCCATCCGCCG
          CTGGAAAAAGACGGTGCAGGATAGAGGACCGATGCCTGGCTCCGAAAACCCTCCTGGAGT

73734     GGCTGCCTGCCCTAGAGCTGGCCCAGGGCGGCTCGGAAGCCTTTGCTGGGCTCTTCCCTG
          GGCAGTGGGACCATGACAGACGAAAGAACCTGTTTCTCATCTCTCCAAGCTGTGGGCACC
          CCTGCCGCTGCCCCTGCCCCTGCCAAGGGCTACAAACTTTTCCAGCTCAAGCCCAAATCT
          CCTCAAGTGATGCCTATTGAAGAATTCCAAGGTAAGAGGATGGACCTGGGGCCCCATCAG
          CCCTCCCTGACACCTGTTCCCCATCCGCCGCTGGAAAAAGACGGTGCAGGATAGAGGACC
          [G,A]
          ATGCCTGGCTCCGAAAACCCTCCTGGAGTAGCTGGGTCAAGGTTAAACTGAGTCTCTCTT
          CCCTACAGGCCTCCCTCCCCAAGGGAGCTGGGAGCAGGTATGAGTCAGAAGCCAACTTGG
          GCACAGTGGGCAGGCCACACAGCAGGCAGAGCAGATGCCAGAAATAGCCCATCCCGGCTC
          CCCTGGGAGGTGTGGCCCTGGGGCTCGTGTTGGTTGAAGCAGAATCTGGGACACACGGGT
          CACCGATGCTGCTCTTTGGGACACTTAGAGGATGCCTCATCTCCTCATTATCTCTGGAGG

75366     CCGATGGAAGGGTGGGAGGCCTAGGGCACCTTCCGGTACCTTTTCCAAAGATGCCTTCCT
          CCGTCCCTGCATGACCTGGGGTGAGTCCTTCCTCGCCCTGTCCCTCAGTTTCCCTGAATG
          CTCGCTGACCATTGGTATTTCTCCCACTTGGCCGGCCCAGACTGCGAATGCTACGGTCAC
          TCCAACCGCTGCAGCTACATTGACTTCCTGAATGTGGTGACCTGCGTCAGCTGCAAGCAC
          AACACGCGAGGTCAGCACTGCCAGCACTGCCGGCTGGGCTACTACCGCAACGGCTCGGCA
          [G,C]
          AGCTGGATGATGAGAACGTCTGCATTGGTGAGAGGGCACGGACACGGCACAGGGAACTTG
          CTGGAATGCGTGCAGGGTGCACTGCCCTGCGAGGTGGCCTCTGGGGCCCCCTGCATCAGA
          ATCACCTGGGGAGACTGTGGGAATTCTAACTCCAGGGCCCTCTCCAGTTGAGCATCTCTA
```

FIGURE 3

AGGACAGAAAGCTCCAGAAACTGCTCTATTAGTAACCTACCCTTGCGGTTCTCCGGTAAG
TTTTGCACTGGAGTTGCAAAACTTACCAGTGGCCCTTCCCTCTCTGGGCAACTGGAGGGG

75368
GATGGAAGGGTGGGAGGCCTAGGGCACCTTCCGGTACCTTTTCCAAAGATGCCTTCCTCC
GTCCCTGCATGACCTGGGGTGAGTCCTTCCTCGCCCTGTCCCTCAGTTTCCCTGAATGCT
CGCTGACCATTGGTATTTCTCCCACTTGGCCGGCCCAGACTGCGAATGCTACGGTCACTC
CAACCGCTGCAGCTACATTGACTTCCTGAATGTGGTGACCTGCGTCAGCTGCAAGCACAA
CACGCGAGGTCAGCACTGCCAGCACTGCCGGCTGGGCTACTACCGCAACGGCTCGGCAGA
[G,A]
CTGGATGATGAGAACGTCTGCATTGGTGAGAGGGCACGGACACGGCACAGGGAACTTGCT
GGAATGCGTGCAGGGTGCACTGCCCTGCGAGGTGGCCTCTGGGGCCCCCTGCATCAGAAT
CACCTGGGGAGACTGTGGGAATTCTAACTCCAGGGCCCTCTCCAGTTGAGCATCTCTAAG
GACAGAAAGCTCCAGAAACTGCTCTATTAGTAACCTACCCTTGCGGTTCTCCGGTAAGTT
TTGCACTGGAGTTGCAAAACTTACCAGTGGCCCTTCCCTCTCTGGGCAACTGGAGGGGAC

76076
TCCATAGCTCCTTTCCAGAAAGGTGGAGGAGCAGCCTATCCCTCCTCCTGCAGGGGCCCA
GTTGGGGGCCAAAAGATCGCCTTGCTGCGTGCATTTGTGCAAGTCCCTTCCCGTTCCTGG
GCCTCAGCTTCCTCATTCATCAAATTGGGAGGCAGATCAGATCAAAGGTTTTCAGCTCTT
TTTTGTGGCTGAAGCTTTTCTTCAAATGCTTTACCAGCCCAGGTCCAGCTATAAAGCTGC
TCTTCACCCCTGGTGGGCACCCAGTCTGCTTTCTTCCAAGTTCTACTCAAGGACTGGCTT
[T,C]
TGGGTAGAGAAGGAAGTCCATCAGGGCCCTGGGCCTGGGCAAAGACCAAAGCCATGACCG
CCAACCAAAACGCACCAGCCTGGAATGGTTGCCCCTGTCGTCAGTAGAGGCCAGGTCTCG
GCCTCAGGGGCTGTCCCCCAACCCTGCCCAGCCAGGCCCCTTGGGACACCATCACCCATC
CCCCACCCAGCAGGAGGCTCTGGCTGCCCAGAGGAGGGGCTCCTGCAAAGCTGGAGCTGT
CGGTCTGAATTCTGGCGGCAGCCTTCAGATAATTCCATCAACTCTAAGTGATCAAAGCCG

79643
TGTCATCTCAGACCTGTTGCAGCCGGAGCCTCAAGTCCAATATCAGATGAAGCTGAACCC
ACAATTCGGCCACCGCCTCCTTCCAGAGTTTCAGATGGCCAGGTGGGCAGAGGCGGGCAG
TGCAGAGACCCCAGACGTGCCCGCCCTGTCCTCCCTACCTTCTCAAGATTAGGAAGGGGT
GCTGGAGGGGACAGGGGCAGCTTGGGAGTGGTGAGGAAGCTCCTAGATTCGGGGCTCATC
CCCTGGGGCCTCTGATTCAGAGGATCCAGCAGTTCTCCCATCTCCGCTTGGTGTCTCCAG
[C,T]
CCTGGGGCCACACTTCCCCCTCGGTCCAGCCTCCTGTCCACCTATGTTTATTTCAGAGCA
GTGCCGGGGGTCCGGTCCTGGTTGCTAACTGCTGCCACTGCTCCACCTGCAGGTGCTCCC
AGCACTGGCTTCTGCCACCACACCTGTTCTTTCCCAGCTGCGAGGTTTAGACCTGGGTCC
TTCCCTTGAGTCCCCAAAGCTAAGCTAAGACCAAGTGGAACAAACTTGGCCTTGGGGACA
GCAGGAGATTACAACACAGAAAAGGAGGGGGAGGCACAACGGGACACTGCATAGGACTCA

80208
GAGGGGGAGGCACAACGGGACACTGCATAGGACTCACAGTGTCCCGAGCCCACACACCAG
CCCCTCCTGGCCTCCTCTCTCTGCTCCCACCCCCAGCACCCTGCTGACCCGGAAGTGCCT
TCCGACAGGCCCTGCATCCTCCTGCAGCTGGCCCATCTCTACCCTCACATTCTTCCTTAC
GCACAGAACCCCACTGCCTGGGACCCAAAGTGCCCAGATAAAATAAAACACCTCTCTGGG
GCTTCTTGTGGATAGGAGTGGCCAGGGGACACAGCTCTGGGCAGTGAGATGTAAGCAGGA
[C,G]
TGATGGGTGGGCTTCCAGAAAGTTCTTAAAAGTCATCCCTTCCTCCACGCCCCACAAAGC
CTCAGTTGTCCAAGTTTGTGGCTCTGGTTTTTCTCTCTCATCCCTCCCTCTGCTCTTCAG
TCAGCAGGATGGGGAAGGAGCCTTCTTGAGACACTGAGCATAACAGTGTCACCCTCAGGA
TAGTGAAGGGTGGAGCCGGGGGTGGAAGTTGCAGAGCCTGGGACACCTGCCTTGGCTGCA
ACCTCTGCACTGCTTTTATTTATCTACTTATTTTTATAGAGTTGAGGTCTCACTATGTTG

FIGURE 3

# ISOLATED HUMAN SECRETED PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND USES THEREOF

## FIELD OF THE INVENTION

[0001] The present invention is in the field of secreted proteins that are related to the netrin-like secreted subfamily, recombinant DNA molecules, and protein production. The present invention specifically provides novel peptides and proteins that effect protein phosphorylation and nucleic acid molecules encoding such peptide and protein molecules, all of which are useful in the development of human therapeutics and diagnostic compositions and methods.

## BACKGROUND OF THE INVENTION

[0002] Secreted Proteins

[0003] Many human proteins serve as pharmaceutically active compounds. Several classes of human proteins that serve as such active compounds include hormones, cytokines, cell growth factors, and cell differentiation factors. Most proteins that can be used as a pharmaceutically active compound fall within the family of secreted proteins. It is, therefore, important in developing new pharmaceutical compounds to identify secreted proteins that can be tested for activity in a variety of animal models. The present invention advances the state of the art by providing many novel human secreted proteins.

[0004] Secreted proteins are generally produced within cells at rough endoplasmic reticulum, are then exported to the golgi complex, and then move to secretory vesicles or granules, where they are secreted to the exterior of the cell via exocytosis.

[0005] Secreted proteins are particularly useful as diagnostic markers. Many secreted proteins are found, and can easily be measured, in serum. For example, a 'signal sequence trap' technique can often be utilized because many secreted proteins, such as certain secretory breast cancer proteins, contain a molecular signal sequence for cellular export. Additionally, antibodies against particular secreted serum proteins can serve as potential diagnostic agents, such as for diagnosing cancer.

[0006] Secreted proteins play a critical role in a wide array of important biological processes in humans and have numerous utilities; several illustrative examples are discussed herein. For example, fibroblast secreted proteins participate in extracellular matrix formation. Extracellular matrix affects growth factor action, cell adhesion, and cell growth. Structural and quantitative characteristics of fibroblast secreted proteins are modified during the course of cellular aging and such aging related modifications may lead to increased inhibition of cell adhesion, inhibited cell stimulation by growth factors, and inhibited cell proliferative ability (Eleftheriou et al., *Mutat Res March-November* 1991;256(2-6):127-38).

[0007] The secreted form of amyloid beta/A4 protein precursor (APP) functions as a growth and/or differentiation factor. The secreted form of APP can stimulate neurite extension of cultured neuroblastoma cells, presumably through binding to a cell surface receptor and thereby triggering intracellular transduction mechanisms. (Roch et al., *Ann N Y Acad Sci Sep.* 24, 1993;695:149-57). Secreted APPs modulate neuronal excitability, counteract effects of glutamate on growth cone behaviors, and increase synaptic complexity. The prominent effects of secreted APPs on synaptogenesis and neuronal survival suggest that secreted APPs play a major role in the process of natural cell death and, furthermore, may play a role in the development of a wide variety of neurological disorders, such as stroke, epilepsy, and Alzheimer's disease (Mattson et al., *Perspect Dev Neurobiol* 1998;5(4):337-52).

[0008] Breast cancer cells secrete a 52K estrogen-regulated protein (see Rochefort et al., *Ann N Y Acad Sci* 1986;464:190-201). This secreted protein is therefore useful in breast cancer diagnosis.

[0009] Two secreted proteins released by platelets, platelet factor 4 (PF4) and beta-thromboglobulin (betaTG), are accurate indicators of platelet involvement in hemostasis and thrombosis and assays that measure these secreted proteins are useful for studying the pathogenesis and course of thromboembolic disorders (Kaplan, *Adv Exp Med Biol* 1978;102:105-19).

[0010] Vascular endothelial growth factor (VEGF) is another example of a naturally secreted protein. VEGF binds to cell-surface heparan sulfates, is generated by hypoxic endothelial cells, reduces apoptosis, and binds to high-affinity receptors that are up-regulated by hypoxia (Asahara et al., *Semin Interv Cardiol Sep.* 1, 1996;(3):225-32).

[0011] Many critical components of the immune system are secreted proteins, such as antibodies, and many important functions of the immune system are dependent upon the action of secreted proteins. For example, Saxon et al., *Biochem Soc Trans May* 5, 1997;(2):383-7, discusses secreted IgE proteins.

[0012] For a further review of secreted proteins, see Nilsen-Hamilton et al., *Cell Biol Int Rep September* 1982 6;(9):815-36.

[0013] Netrins

[0014] Experimental evidence has demonstrated that the netrin family of proteins are involved in embryonic nervous system development in both vertebrates and invertebrates. Specifically, they have been shown to provide guidance for developing axons. Tessier-Lavigne and Goodman, *Science* 274:1123-1133, (1996). For example, Netrin-1, a diffusable protein made by floor plate cells, has been shown to attract spinal commissural axons and repel trochlear axons in vitro, as well as play a vital role in mouse neuron development. Serafini, et al., *Cell* 87:1001-1014, (1996). Netrin has been shown to interact with a laminin protein to convert netrin-mediated attraction into repulsion. It has been suggested that repulsion from the region in which laminin and netrin are coexpressed may help to drive axons into the region where only netrin is present, providing a mechanism for their escape from the regions such as the retinal surface. Hopker, et al., *Nature* 401:69-73, (1999). Experimental evidence suggests that netrin provides guidance for axons by activating the neuronal DCC receptor, and that chemical inhibitors of metalloproteases increase netrin-mediated axon growth in vitro. Galko and Tessier-Lavigne, *Science* 289:1365-1367, (2000).

[0015] Netrin-G1 is a member of the netrin family, but unlike typical netrins, netrin-G1 consists of at least six

isoforms of which five were predominantly anchored to the plasma membrane via glycosyl phosphatidyl-inositol linkages. Netrin-G1 transcripts are expressed in mouse in midbrain and hindbrain regions by embryonic day 12, and reach their highest levels of expression at perinatal stages in various brain regions, including olfactory bulb mitral cells, thalamus, and deep cerebellar nuclei. Unlike typical netrin proteins, netrin-G1 proteins did not show appreciable affinity to any netrin receptor examined. Unlike netrin-1, secreted netrin-G1 does not attract circumferentially growing axons from the cerebellar plate. The expression pattern of netrin-G1 and its predicted neuronal membrane localization suggest it may also have novel signaling functions in nervous system development. For more information on Netrin-G1, see Nakashiba, et al., *J Neurosci September* 1;20(17):6540-50 (2000).

[0016] Secreted proteins, particularly members of the netrin-like secreted protein subfamily, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown members of this subfamily of secreted proteins. The present invention advances the state of the art by providing previously unidentified human secreted proteins that have homology to members of the netrin-like secreted protein subfamily.

## SUMMARY OF THE INVENTION

[0017] The present invention is based in part on the identification of amino acid sequences of human secreted peptides and proteins that are related to the netrin-like secreted protein subfamily, as well as allelic variants and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate secreted protein activity in cells and tissues that express the secreted protein. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma.

## DESCRIPTION OF THE FIGURE SHEETS

[0018] **FIG. 1** provides the nucleotide sequence of a cDNA molecule or transcript sequence that encodes the secreted protein of the present invention. (SEQ ID NO: 1) In addition, structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma.

[0019] **FIG. 2** provides the predicted amino acid sequence of the secreted protein of the present invention. (SEQ ID NO: 2) In addition structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence.

[0020] **FIG. 3** provides genomic sequences that span the gene encoding the secreted protein of the present invention. (SEQ ID NO: 3) In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence. As illustrated in **FIG. 3**, SNPs were identified at 90 different nucleotide positions.

## DETAILED DESCRIPTION OF THE INVENTION

[0021] General Description

[0022] The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a secreted protein or part of a secreted protein and are related to the netrin-like secreted protein subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human secreted peptides and proteins that are related to the netrin-like secreted protein subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these secreted peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the secreted protein of the present invention.

[0023] In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known secreted proteins of the netrin-like secreted protein subfamily and the expression pattern observed. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma. The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known netrin-like family or subfamily of secreted proteins.

## SPECIFIC EMBODIMENTS

[0024] Peptide Molecules

[0025] The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the secreted protein family of proteins and are related to the netrin-like secreted protein subfamily (protein sequences are provided in **FIG. 2**, transcript/cDNA sequences are provided in **FIG. 1** and genomic sequences are provided in **FIG. 3**). The peptide sequences provided in **FIG. 2**, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in **FIG. 3**, will be referred herein as the secreted peptides of the present invention, secreted peptides, or peptides/proteins of the present invention.

[0026] The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprise the amino acid sequences of the secreted peptides disclosed in the **FIG. 2**, (encoded by the nucleic acid molecule shown in **FIG. 1**, transcript/cDNA or **FIG. 3**, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

[0027] As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

[0028] In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

[0029] The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the secreted peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

[0030] The isolated secreted peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma. For example, a nucleic acid molecule encoding the secreted peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

[0031] Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in **FIG. 2** (SEQ ID NO: 2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in **FIG. 1** (SEQ ID NO: 1) and the genomic sequences provided in **FIG. 3** (SEQ ID NO: 3). The amino acid sequence of such a protein is provided in **FIG. 2**. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

[0032] The present invention further provides proteins that consist essentially of the amino acid sequences provided in **FIG. 2** (SEQ ID NO: 2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in **FIG. 1** (SEQ ID NO: 1) and the genomic sequences provided in **FIG. 3** (SEQ ID NO: 3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

[0033] The present invention further provides proteins that comprise the amino acid sequences provided in **FIG. 2** (SEQ ID NO: 2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in **FIG. 1** (SEQ ID NO: 1) and the genomic sequences provided in **FIG. 3** (SEQ ID NO: 3). A protein comprises an amino acid sequence when the amino acid sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the secreted peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

[0034] The secreted peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a secreted peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the secreted peptide. "Operatively linked" indicates that the secreted peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the secreted peptide.

[0035] In some uses, the fusion protein does not affect the activity of the secreted peptide per se. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant secreted peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

[0036] A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel et al., *Current Protocols in Molecular Biology,* 1992). Moreover, many expression vectors are

commercially available that already encode a fusion moiety (e.g., a GST protein). A secreted peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the secreted peptide.

[0037] As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

[0038] Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the secreted peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

[0039] To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[0040] The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology,* Lesk, A. M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects,* Smith, D. W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part* 1, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology,* von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer,* Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., et al., *Nucleic Acids Res.* 12(1):387 (1984)) (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

[0041] The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

[0042] Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the secreted peptides of the present invention as well as being encoded by the same genetic locus as the secreted peptide provided herein.

[0043] Allelic variants of a secreted peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the secreted peptide as well as being encoded by the same genetic locus as the secreted peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in **FIG. 3**, such as the genomic sequence mapped to the reference human. As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a secreted peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

[0044] **FIG. 3** provides information on SNPs that have been found in the gene encoding the receptor protein of the present invention. SNPs were identified at 90 different

nucleotide positions, including a non-synonymous coding SNP at positions 753666 and 75368. Changes in the amino acid sequence caused by these SNPs is indicated in **FIG. 3** and can readily be determined using the universal genetic code and the protein sequence provided in **FIG. 2** as a reference. Some of these SNPs that are located outside the ORF and in introns may affect gene expression. Positioning of each SNP in an exon, intron, or outside the ORF can readily be determined using the DNA position given for each SNP and the start/stop, exon, and intron genomic coordinates given in **FIG. 3**.

[**0045**] Paralogs of a secreted peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the secreted peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a secreted peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

[**0046**] Orthologs of a secreted peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the secreted peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a secreted peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

[**0047**] Non-naturally occurring variants of the secreted peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the secreted peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a secreted peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie et al., *Science* 247:1306-1310 (1990).

[**0048**] Variant secreted peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind substrate, ability to phosphorylate substrate, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. **FIG. 2** provides the result of protein analysis and can be used to identify critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

[**0049**] Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

[**0050**] Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham et al., *Science* 244:1081-1085 (1989)), particularly using the results provided in **FIG. 2**. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as secreted protein activity or in assays such as an in vitro proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith et al., *J. Mol. Biol.* 224:899-904 (1992); de Vos et al. *Science* 255:306-312 (1992)).

[**0051**] The present invention further provides fragments of the secreted peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in **FIG. 2**. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

[**0052**] As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a secreted peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the secreted peptide or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the secreted peptide, e.g., active site or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in **FIG. 2**.

[**0053**] Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the termninal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in secreted peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in **FIG. 2**).

[**0054**] Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme

moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

[0055] Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutarnic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins—Structure and Molecular Properties,* 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins, B. C. Johnson, Ed., Academic Press, New York* 1-12 (1983); Seifter et al. (*Meth. Enzymol.* 182:626-646 (1990)) and Rattan et al. (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

[0056] Accordingly, the secreted peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature secreted peptide is fused with another compound, such as a compound to increase the half-life of the secreted peptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature secreted peptide, such as a leader or secretory sequence or a sequence for purification of the mature secreted peptide or a pro-protein sequence.

[0057] Protein/Peptide Uses

[0058] The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein or ligand (such as, for example, in a secreted protein-effector protein interaction or secreted protein-ligand interaction), the protein can be used to identify the binding partner/ligand so as to develop a system to identify inhibitors of the binding interaction. Any or all of these uses are capable of being developed into reagent grade or kit format for commercialization as commercial products.

[0059] Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

[0060] The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, secreted proteins isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the secreted protein. Experimental data as provided in **FIG. 1** indicates that secreted proteins of the present invention are expressed in glioblastoma and adrenal cortex carcinoma. Specifically, a virtual northern blot shows expression in glioblastoma and adrenal cortex carcinoma. A large percentage of pharmaceutical agents are being developed that modulate the activity of secreted proteins, particularly members of the netrin-like subfamily (see Background of the Invention). The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in **FIG. 1**. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma. Such uses can readily be determined using the information provided herein, that which is known in the art, and routine experimentation.

[0061] The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to secreted proteins that are related to members of the netrin-like subfamily. Such assays involve any of the known secreted protein functions or activities or properties useful for diagnosis and treatment of secreted protein-related conditions that are specific for the subfamily of secreted proteins that the one of the present invention belongs to, particularly in cells and tissues that express the secreted protein. Experimental data as provided in FIG. **1** indicates that secreted proteins of the present invention are expressed in glioblastoma and adrenal cortex carcinoma. Specifically, a virtual northern blot shows expression in glioblastoma and adrenal cortex carcinoma.

[0062] The proteins of the present invention are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the secreted protein, as a biopsy or expanded in cell culture. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the secreted protein.

[0063] The polypeptides can be used to identify compounds that modulate secreted protein activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with the secreted protein. Both the secreted proteins of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the secreted protein. These compounds can be further screened against a functional secreted protein to determine the effect of the compound on the secreted protein

activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the secreted protein to a desired degree.

[0064] Further, the proteins of the present invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the secreted protein and a molecule that normally interacts with the secreted protein, e.g. a substrate or a component of the signal pathway that the secreted protein normally interacts (for example, another secreted protein). Such assays typically include the steps of combining the secreted protein with a candidate compound under conditions that allow the secreted protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the secreted protein and the target.

[0065] Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam et al., *Nature* 354:82-84 (1991); Houghten et al., *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang et al., *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, $F(ab')_2$, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

[0066] One candidate compound is a soluble fragment of the receptor that competes for substrate binding. Other candidate compounds include mutant secreted proteins or appropriate fragments containing mutations that affect secreted protein function and thus compete for substrate. Accordingly, a fragment that competes for substrate, for example with a higher affinity, or a fragment that binds substrate but does not allow release, is encompassed by the invention.

[0067] Any of the biological or biochemical functions mediated by the secreted protein can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly **FIG. 2**. Specifically, a biological function of a cell or tissues that expresses the secreted protein can be assayed. Experimental data as provided in **FIG. 1** indicates that secreted proteins of the present invention are expressed in glioblastoma and adrenal cortex carcinoma. Specifically, a virtual northern blot shows expression in glioblastoma and adrenal cortex carcinoma.

[0068] Binding and/or activating compounds can also be screened by using chimeric secreted proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a substrate-binding region can be used that interacts with a different substrate then that which is recognized by the native secreted protein. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which the secreted protein is derived.

[0069] The proteins of the present invention are also useful in competition binding assays in methods designed to discover compounds that interact with the secreted protein (e.g. binding partners and/or ligands). Thus, a compound is exposed to a secreted protein polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble secreted protein polypeptide is also added to the mixture. If the test compound interacts with the soluble secreted protein polypeptide, it decreases the amount of complex formed or activity from the secreted protein target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the secreted protein. Thus, the soluble polypeptide that competes with the target secreted protein region is designed to contain peptide sequences corresponding to the region of interest.

[0070] To perform cell free drug screening assays, it is sometimes desirable to immobilize either the secreted protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

[0071] Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., $^{35}S$-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of secreted protein-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a secreted protein-binding protein and a candidate compound are incubated in the secreted protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using

antibodies reactive with the secreted protein target molecule, or which are reactive with secreted protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

[0072] Agents that modulate one of the secreted proteins of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

[0073] Modulators of secreted protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the secreted protein pathway, by treating cells or tissues that express the secreted protein. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma. These methods of treatment include the steps of administering a modulator of secreted protein activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

[0074] In yet another aspect of the invention, the secreted proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Pat. No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with the secreted protein and are involved in secreted protein activity.

[0075] The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a secreted protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, in vivo, forming a secreted protein-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the secreted protein.

[0076] This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a secreted protein-modulating agent, an antisense secreted protein nucleic acid molecule, a secreted protein-specific antibody, or a secreted protein-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

[0077] The secreted proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma. The method involves contacting a biological sample with a compound capable of interacting with the secreted protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

[0078] One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

[0079] The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered secreted protein activity in cell-based or cell-free assay, alteration in substrate or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

[0080] In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can be detected in vivo in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

[0081] The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M. W. (*Clin. Chem.*

43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the secreted protein in which one or more of the secreted protein functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other substrate-binding regions that are more or less active in substrate binding, and secreted protein activation. Accordingly, substrate dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

[0082] The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma. Accordingly, methods for treatment include the use of the secreted protein or fragments.

[0083] Antibodies

[0084] The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

[0085] As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or $F(ab')_2$, and Fv fragments.

[0086] Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, Antibodies, Cold Spring Harbor Press, (1989).

[0087] In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in **FIG. 2**, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

[0088] Antibodies are preferably prepared from regions or discrete fragments of the secreted proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or secreted protein/binding partner interaction. **FIG. 2** can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

[0089] An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see **FIG. 2**).

[0090] Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidini-biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include $^{125}I$, $^{131}i$, $^{35}S$ or $^{3}H$.

[0091] Antibody Uses

[0092] The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Experimental data as provided in **FIG. 1** indicates that secreted proteins of the present invention are expressed in glioblastoma and adrenal cortex carcinoma. Specifically, a

virtual northern blot shows expression in glioblastoma and adrenal cortex carcinoma. Further, such antibodies can be used to detect protein in situ, in vitro, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

[0093] Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

[0094] The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

[0095] Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

[0096] The antibodies are also useful for tissue typing. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

[0097] The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the secreted peptide to a binding partner such as a substrate. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See **FIG. 2** for structural information relating to the proteins of the present invention.

[0098] The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nuleic acid arrays and similar methods have been developed for antibody arrays.

[0099] Nucleic Acid Molecules

[0100] The present invention further provides isolated nucleic acid molecules that encode a secreted peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the secreted peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

[0101] As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

[0102] Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

[0103] For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include in vivo or in vitro RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

[0104] Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in FIGS. **1** or **3** (SEQ ID NO: 1, transcript sequence and SEQ ID NO: 3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in **FIG. 2**, SEQ ID NO: 2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

[0105] The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in FIGS. **1** or **3** (SEQ ID NO: 1, transcript sequence and a SEQ ID NO: 3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in **FIG. 2**, SEQ ID NO: 2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

[0106] The present invention further provides nucleic acid molecules that comprise the nucleotide sequences shown in FIGS. **1** or **3** (SEQ ID NO: 1, transcript sequence and SEQ ID NO: 3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in **FIG. 2**, SEQ ID NO: 2. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

[0107] In **FIGS. 1 and 3**, both coding and non-coding sequences are provided. Because of the source of the present invention, humans genomic sequence (**FIG. 3**) and cDNA/transcript sequences (**FIG. 1**), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in **FIGS. 1 and 3** or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

[0108] The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case in situ, the additional amino acids may be processed away from the mature protein by cellular enzymes.

[0109] As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the secreted peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

[0110] Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

[0111] The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the secreted proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

[0112] The present invention further provides non-coding fragments of the nucleic acid molecules provided in **FIGS. 1 and 3**. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in developing screens to identify gene-modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in **FIG. 3**.

[0113] A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a CDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

[0114] A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

[0115] Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95%

or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene.

[0116] **FIG. 3** provides information on SNPs that have been found in the gene encoding the receptor protein of the present invention. SNPs were identified at 90 different nucleotide positions, including a non-synonymous coding SNP at positions 753666 and 75368. Changes in the amino acid sequence caused by these SNPs is indicated in **FIG. 3** and can readily be determined using the universal genetic code and the protein sequence provided in **FIG. 2** as a reference. Some of these SNPs that are located outside the ORF and in introns may affect gene expression. Positioning of each SNP in an exon, intron, or outside the ORF can readily be determined using the DNA position given for each SNP and the start/stop, exon, and intron genomic coordinates given in **FIG. 3**.

[0117] As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology,* John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6× sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2×SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

[0118] Nucleic Acid Molecule Uses

[0119] The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in **FIG. 2** and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in **FIG. 2**. As illustrated in **FIG. 3**, SNPs were identified at 90 different nucleotide positions.

[0120] The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

[0121] The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

[0122] The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter in situ expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

[0123] The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

[0124] The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of in situ hybridization methods.

[0125] The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

[0126] The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

[0127] The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

[0128] The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

[0129] The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

[0130] The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Experimental data as provided in **FIG. 1** indicates that secreted proteins of the present invention are expressed in glioblastoma and adrenal cortex carcinoma. Specifically, a virtual northern blot shows expression in glioblastoma and adrenal cortex carcinoma. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in secreted protein expression relative to normal results.

[0131] In vitro techniques for detection of mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detecting DNA include Southern hybridizations and in situ hybridization.

[0132] Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a secreted protein, such as by measuring a level of a secreted protein-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a secreted protein gene has been mutated. Experimental data as provided in **FIG. 1** indicates that secreted proteins of the present invention are expressed in glioblastoma and adrenal cortex carcinoma.

Specifically, a virtual northern blot shows expression in glioblastoma and adrenal cortex carcinoma.

[0133] Nucleic acid expression assays are useful for drug screening to identify compounds that modulate secreted protein nucleic acid expression.

[0134] The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the secreted protein gene, particularly biological and pathological processes that are mediated by the secreted protein in cells and tissues that express it. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma. The method typically includes assaying the ability of the compound to modulate the expression of the secreted protein nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired secreted protein nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the secreted protein nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

[0135] Thus, modulators of secreted protein gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of secreted protein mRNA in the presence of the candidate compound is compared to the level of expression of secreted protein mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

[0136] The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate secreted protein nucleic acid expression in cells and tissues that express the secreted protein. Experimental data as provided in **FIG. 1** indicates that secreted proteins of the present invention are expressed in glioblastoma and adrenal cortex carcinoma. Specifically, a virtual northern blot shows expression in glioblastoma and adrenal cortex carcinoma. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

[0137] Alternatively, a modulator for secreted protein nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the secreted protein nucleic acid expression in the cells and tissues that express the protein. Experimental data as provided in **FIG. 1** indicates expression in glioblastoma and adrenal cortex carcinoma.

[0138] The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the secreted protein gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

[0139] The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in secreted protein nucleic acid expression, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in secreted protein genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally occurring genetic mutations in the secreted protein gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the secreted protein gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a secreted protein.

[0140] Individuals carrying mutations in the secreted protein gene can be detected at the nucleic acid level by a variety of techniques. **FIG. 3** provides information on SNPs that have been found in the gene encoding the receptor protein of the present invention. SNPs were identified at 90 different nucleotide positions, including a non-synonymous coding SNP at positions 753666 and 75368. Changes in the amino acid sequence caused by these SNPs is indicated in **FIG. 3** and can readily be determined using the universal genetic code and the protein sequence provided in **FIG. 2** as a reference. Some of these SNPs that are located outside the ORF and in introns may affect gene expression. Positioning of each SNP in an exon, intron, or outside the ORF can readily be determined using the DNA position given for each SNP and the start/stop, exon, and intron genomic coordinates given in **FIG. 3**. Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Pat. Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al., *Science* 241:1077-1080 (1988); and Nakazawa et al., *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya et al., *Nucleic Acids Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, "mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under

conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

[0141] Alternatively, mutations in a secreted protein gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

[0142] Further, sequence-specific ribozymes (U.S. Pat. No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

[0143] Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and SI protection or the chemical cleavage method. Furthermore, sequence differences between a mutant secreted protein gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C. W., (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen et al., *Adv. Chromatogr.* 36:127-162 (1996); and Griffin et al., *Appl. Biochem. BiotechnoL* 38:147-159 (1993)).

[0144] Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers et al., *Science* 230:1242 (1985)); Cotton et al, *PNAS* 85:4397 (1988); Saleeba et al., *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita et al., *PNAS* 86:2766 (1989); Cotton et al., *Mutat. Res.* 285:125-144(1993); and Hayashi et al., *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers et al., *Nature* 313:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective amplification, and selective primer extension.

[0145] The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the secreted protein gene in an individual in order to select an appropriate compound or dosage regimen for treatment. **FIG. 3** provides information on SNPs that have been found in the gene encoding the receptor protein of the present invention. SNPs were identified at 90 different nucleotide positions, including a non-synonymous coding SNP at positions 753666 and 75368. Changes in the amino acid sequence caused by these SNPs is indicated in **FIG. 3** and can readily be determined using the universal genetic code and the protein sequence provided in **FIG. 2** as a reference. Some of these SNPs that are located outside the ORF and in introns may affect gene expression. Positioning of each SNP in an exon, intron, or outside the ORF can readily be determined using the DNA position given for each SNP and the start/stop, exon, and intron genomic coordinates given in **FIG. 3**.

[0146] Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

[0147] The nucleic acid molecules are thus useful as antisense constructs to control secreted protein gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of secreted protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into secreted protein.

[0148] Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of secreted protein nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired secreted protein nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the secreted protein, such as substrate binding.

[0149] The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in secreted protein gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered ex vivo and returned to the patient, are introduced into an individual where the cells produce the desired secreted protein to treat the individual.

[0150] The invention also encompasses kits for detecting the presence of a secreted protein nucleic acid in a biological sample. Experimental data as provided in **FIG. 1** indicates that secreted proteins of the present invention are expressed in glioblastoma and adrenal cortex carcinoma. Specifically, a virtual northern blot shows expression in glioblastoma and adrenal cortex carcinoma. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting secreted protein nucleic acid in a biological sample; means for determining the amount of secreted protein nucleic acid in the sample; and means for comparing the amount of secreted protein nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect secreted protein mRNA or DNA.

[0151] Nucleic Acid Arrays

[0152] The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid

molecules that are based on the sequence information provided in **FIGS. 1 and 3** (SEQ ID NOS:1 and 3).

[**0153**] As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in U.S. Pat. No. 5,837,832, Chee et al., PCT application W095/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93:10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown et al., U.S. Pat. No. 5,807,522.

[**0154**] The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

[**0155**] In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

[**0156**] In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/25 1116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

[**0157**] In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large-scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

[**0158**] Using such arrays, the present invention provides methods to identify the expression of the secreted proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the present invention and or alleles of the secreted protein gene of the present invention. **FIG. 3** provides information on SNPs that have been found in the gene encoding the receptor protein of the present invention. SNPs were identified at 90 different nucleotide positions, including a non-synonymous coding SNP at positions 753666 and 75368. Changes in the amino acid sequence caused by these SNPs is indicated in **FIG. 3** and can readily be determined using the universal genetic code and the protein sequence provided in **FIG. 2** as a reference. Some of these SNPs that are located outside the ORF and in introns may affect gene expression. Positioning of each SNP in an exon, intron, or outside the ORF can readily be determined using the DNA position given for each SNP and the start/stop, exon, and intron genomic coordinates given in **FIG. 3**.

[**0159**] Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be

found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques,* Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. et al., *Techniques in Immunocytochemistry,* Academic Press, Orlando, Fla. Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Lab oratory Techniques in Biochemistry and Molecular Biology,* Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

[0160] The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

[0161] In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

[0162] Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the Human genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

[0163] In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified secreted protein gene of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

[0164] Vectors/host cells

[0165] The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

[0166] A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

[0167] The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in prokarvotic or eukaryotic cells or in both (shuttle vectors).

[0168] Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

[0169] The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ, the lac, TRP, and TAC promoters from *E. coli,* the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

[0170] In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

[0171] In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual.* 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (1989).

[0172] A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors

for prokaryotic and eukaryotic hosts are described in Sambrook et al., *Molecular Cloning: A Laboratory Manual.* 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (1989).

[0173] The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

[0174] The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

[0175] The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli,* Streptomyces, and *Salmonella typhimurium.* Eukaryotic cells include, but are not limited to, yeast, insect cells such as Drosophila, animal cells such as COS and CHO cells, and plant cells.

[0176] As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith et al., *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., *Gene* 69:301-315 (1988)) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

[0177] Recombinant protein expression can be maximized in host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. (1990)119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli.* (Wada et al., *Nucleic Acids Res.* 20:2111-2118 (1992)).

[0178] The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, et al., *EMBO J.* 6:229-234 (1987)), pMFa (Kuijan et al., *Cell* 30:933-943(1982)), pJRY88 (Schultz et al., *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, Calif.).

[0179] The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al., *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow et al., *Virology* 170:31-39 (1989)).

[0180] In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman et al., *EMBO J.* 6:187-195 (1987)).

[0181] The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[0182] The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

[0183] The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

[0184] The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

[0185] Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid

molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

[0186] In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

[0187] Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

[0188] While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

[0189] Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as kinases, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

[0190] Where the peptide is not secreted into the medium, which is typically the case with kinases, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

[0191] It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

[0192] Uses of vectors and host cells

[0193] The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a secreted protein or peptide that can be further purified to produce desired amounts of secreted protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

[0194] Host cells are also useful for conducting cell-based assays involving the secreted protein or secreted protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native secreted protein is useful for assaying compounds that stimulate or inhibit secreted protein function.

[0195] Host cells are also useful for identifying secreted protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant secreted protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native secreted protein.

[0196] Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a secreted protein and identifying and evaluating modulators of secreted protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

[0197] A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the secreted protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

[0198] Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the secreted protein to particular cells.

[0199] Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Pat. Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Pat. No. 4,873,191 by Wagner et al. and in Hogan, B., *Manipulating the Mouse Embryo,* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

[0200] In another embodiment, transgenic non-human animals can be produced which contain selected systems that

allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, see, e.g., Lakso et al. *PNAS* 89:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman et al. *Science* 251:1351-1355 (1991). If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

[**0201**] Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. et al. *Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter $G_0$ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

[**0202**] Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an in vivo context. Accordingly, the various physiological factors that are present in vivo and that could effect substrate binding, secreted protein activation, and signal transduction, may not be evident from in vitro cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay in vivo secreted protein function, including substrate interaction, the effect of specific mutant secreted proteins on secreted protein function and substrate interaction, and the effect of chimeric secreted proteins. It is also possible to assess the effect of null mutations, that is, mutations that substantially or completely eliminate one or more secreted protein functions.

[**0203**] All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

---

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 5

<210> SEQ ID NO 1
<211> LENGTH: 3537
<212> TYPE: DNA
<213> ORGANISM: Human

<400> SEQUENCE: 1

ttggtgatcg tgagtgatgg caagaggatt tagcctcggc attaacttgg agcggagtgc 60

agggggcag tgaagcgccc gccatctggc ccgcgccgcg ccgggggat gccccggctc 120

cccgacgaga cgccgcgaag cccacccggg ccgggggctg cccggcgccc gagcgcgggt 180

cctccccggg ccgcccaggg gggccaaaaa gtttgcactt gttagcggcg acctcccgct 240

cagcccgggc gggcgatgcg ggcggcgcgg gcggcccccc cccccggccc gcgtctccgg 300

gacggctgcg ggcggccccc ccggcggccg gagggctccc tggccccgat ctgacggcgg 360

cggcggcggc ggccacagcg gcgggagcgg cgcggggaag gagcagcggc tcgcagccct 420

cggcccgcgc ccccacccag cgccagcccg agggggggagg cgcagcgccg gagggtggcg 480

gtcctcggcc ctcccaggtc tccgcgccgg gaagccgctc cgagccggga ttgtgcctat 540

gattgggggg ctgtttctca gtgcctggct cttcatggtt ggccagcacg tggagcactc 600

tgttcatctg tacatataga caagtcagac cccaggtacc tggatggatt gagagctgag 660

atctcagaaa cttcataata agttgccgat ggctaccagc caaggctgga gggttgtttt 720

gccgttgtgt tgagcacgtc acccattaag agccctttaa agacctggat tgattggaag 780

-continued

gacaaaaatt aaaagcaatc tgatccagcc tcatgcagga tccctgcgga ttttctcctt 840

atcccatttc catccactgt cacaatttga gaatctgcct gatttgatca gattcacctc 900

caggggaggt gtgataccag ggttaggagg acgtgaagtt atgggcaact ttctgatctg 960

tccatcagca gtctgagaaa cgctggctct gaattttccg tgtcggcctt ttggaaacaa 1020

caagttcctc gctgtttgca aagcttcagt gctcgggtcc ctgggacacc ccggccaccc 1080

tcgcctggta gatgtggcat ttccatgctg aggccgcgag tcccgcctga ccccgtcgct 1140

gcctctccag ggcttctctg ggccgcgcct ctgcagactg cgcagccatg ctgcatctgc 1200

tggcgctctt cctgcactgc ctccctctgg cctctgggga ctatgacatc tgcaaatcct 1260

gggtgaccac agatgagggc cccacctggg agttctacgc ctgccagccc aaggtgatgc 1320

gcctgaagga ctacgtcaag gtgaaggtgg agccctcagg catcacatgt ggagaccccc 1380

ctgagaggtt ctgctcccat gagaatccct acctatgcag caacgagtgt gacgcctcca 1440

acccggacct ggcccacccg cccaggctca tgttcgacaa ggaggaggag ggcctggcca 1500

cctactggca gagcatcacc tggagccgct accccagccc gctggaagcc aacatcaccc 1560

tttcgtggaa caagaccgtg gagctgaccg acgacgtggt gatgaccttc gagtacggcc 1620

ggcccacggt catggtcctg gagaagtccc tggacaacgg gcgcacctgg cagccctacc 1680

agttctacgc cgaggactgc atggaggcct tcggtatgtc cgcccgccgg gcccgcgaca 1740

tgtcatcctc cagcgcgcac cgcgtgctct gcaccgagga gtactcgcgc tgggcaggct 1800

ccaagaagga gaagcacgtg cgcttcgagg tgcgggaccg cttcgccatc tttgccggcc 1860

ccgacctgcg caacatggac aacctctaca cgcggctgga gagcgccaag ggcctcaagg 1920

agttcttcac cctcaccgac ctgcgcatgc ggctgctgcg cccggcgctg ggcggcacct 1980

atgtgcagcg ggagaacctc tacaagtact tctacgccat ctccaacatc gaggtcatcg 2040

gcaggtgcaa gtgcaacctg catgccaacc tgtgctccat gcgcgagggc agcctgcagt 2100

gcgagtgcga gcacaacacc accggccccg actgcggcaa gtgcaagaag aatttccgca 2160

cccggtcctg gcgggccggc tcctacctgc cgctgcccca tggctctccc aacgcctgtg 2220

ccgctgcagg ttcctttggc aactgcgaat gctacggtca ctccaaccgc tgcagctaca 2280

ttgacttcct gaatgtggtg acctgcgtca gctgcaagca caacacgcga ggtcagcact 2340

gccagcactg ccggctgggc tactaccgca acggctcggc agagctggat gatgagaacg 2400

tctgcattga gtgtaactgc aaccagatag gctccgtgca cgaccggtgc aacgagaccg 2460

gcttctgcga gtgccgcgag ggcgcggcgg gccccaagtg cgacgactgc ctccccacgc 2520

actactggcg ccagggctgc taccccaacg tgtgcgacga cgaccagctg ctgtgccaga 2580

acggaggcac ctgcctgcag aaccagcgct gcgcctgccc gcgcggctac accggcgtgc 2640

gctgcgagca gccccgctgc gaccccgccg acgatgacgg cggtctggac tgcgaccgcg 2700

cgcccggggc cgccccgcgc cccgccaccc tgctcggctg cctgctgctg ctggggctgg 2760

ccgcccgcct gggccgctga gccccgcccg gaggacgctc cccgcacccg gaggccgggg 2820

gtcccggggt cccggggcgg ggccggcgtc cgaggccggg cggtgagaag ggtgcggccc 2880

gaggtgctcc caggtgctac tcagcagggc cccccgcccg gcccgcgctc ccgcccgcac 2940

tgccctcccc ccgcagcagg ggcgccttgg gactccggtc cccgcgcctg cgatttggtt 3000

tcgtttttct tttgtattat ccgccgccca gttccttttt tgtctttctc tctctctctt 3060

-continued

```
ttttttttttt ttttctggc ggtgagccag agggtcggga gaaacgctgc tcgccccaca   3120

ccccgtcctg cctcccacca cacttacaca cacgggactg tggccgacac cccctggcct   3180

gtgccaggct cacgggcggc ggcggacccc gacctccagt tgcctacaat tccagtcgct   3240

gacttggtcc tgttttctat tctttatttt tcctgcaacc caccagaccc caggcctcac   3300

cggaggcccg gtgaccacgg aactcaccgt ctgggggagg aggagagaag gaaggggtgg   3360

ggggcctgga aacttcgttc tgtagagaac tatttttgtt tgtattcact gtcccctgca   3420

agggggacgg ggcgggagca ctggtcaccg cgggggccga tggtggagaa tccgaggagt   3480

aaagagtttg ctcactgctg cctccaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa      3537
```

<210> SEQ ID NO 2
<211> LENGTH: 530
<212> TYPE: PRT
<213> ORGANISM: Human

<400> SEQUENCE: 2

```
Met Leu His Leu Leu Ala Leu Phe Leu His Cys Leu Pro Leu Ala Ser
 1               5                  10                  15

Gly Asp Tyr Asp Ile Cys Lys Ser Trp Val Thr Thr Asp Glu Gly Pro
            20                  25                  30

Thr Trp Glu Phe Tyr Ala Cys Gln Pro Lys Val Met Arg Leu Lys Asp
        35                  40                  45

Tyr Val Lys Val Lys Val Glu Pro Ser Gly Ile Thr Cys Gly Asp Pro
    50                  55                  60

Pro Glu Arg Phe Cys Ser His Glu Asn Pro Tyr Leu Cys Ser Asn Glu
65                  70                  75                  80

Cys Asp Ala Ser Asn Pro Asp Leu Ala His Pro Pro Arg Leu Met Phe
                85                  90                  95

Asp Lys Glu Glu Glu Gly Leu Ala Thr Tyr Trp Gln Ser Ile Thr Trp
            100                 105                 110

Ser Arg Tyr Pro Ser Pro Leu Glu Ala Asn Ile Thr Leu Ser Trp Asn
        115                 120                 125

Lys Thr Val Glu Leu Thr Asp Asp Val Val Met Thr Phe Glu Tyr Gly
    130                 135                 140

Arg Pro Thr Val Met Val Leu Glu Lys Ser Leu Asp Asn Gly Arg Thr
145                 150                 155                 160

Trp Gln Pro Tyr Gln Phe Tyr Ala Glu Asp Cys Met Glu Ala Phe Gly
                165                 170                 175

Met Ser Ala Arg Arg Ala Arg Asp Met Ser Ser Ser Ser Ala His Arg
            180                 185                 190

Val Leu Cys Thr Glu Glu Tyr Ser Arg Trp Ala Gly Ser Lys Lys Glu
        195                 200                 205

Lys His Val Arg Phe Glu Val Arg Asp Arg Phe Ala Ile Phe Ala Gly
    210                 215                 220

Pro Asp Leu Arg Asn Met Asp Asn Leu Tyr Thr Arg Leu Glu Ser Ala
225                 230                 235                 240

Lys Gly Leu Lys Glu Phe Phe Thr Leu Thr Asp Leu Arg Met Arg Leu
                245                 250                 255

Leu Arg Pro Ala Leu Gly Gly Thr Tyr Val Gln Arg Glu Asn Leu Tyr
            260                 265                 270

Lys Tyr Phe Tyr Ala Ile Ser Asn Ile Glu Val Ile Gly Arg Cys Lys
```

-continued

```
        275                 280                 285

Cys Asn Leu His Ala Asn Leu Cys Ser Met Arg Glu Gly Ser Leu Gln
    290                 295                 300

Cys Glu Cys Glu His Asn Thr Thr Gly Pro Asp Cys Gly Lys Cys Lys
305                 310                 315                 320

Lys Asn Phe Arg Thr Arg Ser Trp Arg Ala Gly Ser Tyr Leu Pro Leu
                325                 330                 335

Pro His Gly Ser Pro Asn Ala Cys Ala Ala Ala Gly Ser Phe Gly Asn
            340                 345                 350

Cys Glu Cys Tyr Gly His Ser Asn Arg Cys Ser Tyr Ile Asp Phe Leu
        355                 360                 365

Asn Val Val Thr Cys Val Ser Cys Lys His Asn Thr Arg Gly Gln His
    370                 375                 380

Cys Gln His Cys Arg Leu Gly Tyr Tyr Arg Asn Gly Ser Ala Glu Leu
385                 390                 395                 400

Asp Asp Glu Asn Val Cys Ile Glu Cys Asn Cys Asn Gln Ile Gly Ser
                405                 410                 415

Val His Asp Arg Cys Asn Glu Thr Gly Phe Cys Glu Cys Arg Glu Gly
            420                 425                 430

Ala Ala Gly Pro Lys Cys Asp Asp Cys Leu Pro Thr His Tyr Trp Arg
        435                 440                 445

Gln Gly Cys Tyr Pro Asn Val Cys Asp Asp Asp Gln Leu Leu Cys Gln
    450                 455                 460

Asn Gly Gly Thr Cys Leu Gln Asn Gln Arg Cys Ala Cys Pro Arg Gly
465                 470                 475                 480

Tyr Thr Gly Val Arg Cys Glu Gln Pro Arg Cys Asp Pro Ala Asp Asp
                485                 490                 495

Asp Gly Gly Leu Asp Cys Asp Arg Ala Pro Gly Ala Ala Pro Arg Pro
            500                 505                 510

Ala Thr Leu Leu Gly Cys Leu Leu Leu Leu Gly Leu Ala Ala Arg Leu
        515                 520                 525

Gly Arg
    530
```

<210> SEQ ID NO 3
<211> LENGTH: 80959
<212> TYPE: DNA
<213> ORGANISM: Human
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)...(80959)
<223> OTHER INFORMATION: n = A,T,C or G

<400> SEQUENCE: 3

```
gagttttaat accctggcta tcagcccccc ttggttcctg aggactctta aaagaaaata     60

aagcacattg attctatttg tttctgggag ctgcagtttc ttaataatat caggtgaaga    120

taaattttcc acggagaaaa cgatcctccg ggatgcagct tcttactctg aaaatttccc    180

tgccgactcc tcactctctg cgctcctcct cgttatccgg ggactcctgc ctctcttccc    240

ccttctcttt tttctttttg gcagaacccg cctgcaatat tcgtgtgctg agctcgtaat    300

tccccctgcg atgccagcaa cgcccaattg attgactagt tgtaaacaca tttttcccct    360

ggcagatttt gttgttgtta gggtttttaa aatttattta ttttccaggg aatgcgtggc    420
```

-continued

atttaaacca acaggactgc aattaataga tttgcgagtt gcgccgcgcg cgccgctcgc 480

cccagcctcc cggcctccgg gcctcgctgc ctccccgcgc ccggcggcgt ccagcgccct 540

gcaagccccg agcagccgcg ggtcctgcag ctgaaggaag gttgcagctg cgccctcctt 600

gcaagccgca gcccggcgtc ctggttgtcc cagcagccag gagatcccta cctgttagtg 660

aacagttagg agtcgactgc tggaagaatt aattaggaac gtgctgtgct ctgggcagcg 720

cgagctcggg tagaggcatc caaacctttg ccggcggcgc tattttattt ttactacatt 780

ttctcaggtt gcaaaaatag acaccgggca cgttctttct tagagttttc tagcaaggag 840

cgccttcaag gccaggcagg ctctgtaaca ggttcccctt taaacagcca gaggtgagac 900

ggggaaaatg gtcctggctg ggttctcgtt catctccatc agcagtcctt cacccagaga 960

gaggggcagg ggtcgcccta actcagatga atgagtccca tgcctggagc cctggggccc 1020

tggctggggg ctgctccgag cctgaggtgc tcagggcgct cagggcagca agtgtccgcc 1080

acttcggttt gtcatttttg gcaggagcgt ttttctgtct gggtggagaa tggagtttca 1140

cggaaacaca gttaactctt caggggcctt gcaagtacag gaggtgaaga ggatgtcagg 1200

ggagagccag gtccagactg gacatttggg gtggtttggg aaatcaaatg caatcatcga 1260

agacatatta accagaataa ttaatcatgc aggcactttt ttactgcagt aacctttgcc 1320

ctattggcca atatttttgg ccagaatccc atgctggctg gacttggatt ctccgggtga 1380

cgtatccagt gtctggaaca caccacagta ctgcagtcgt gttatttccc aatgtaacat 1440

tcatgtaact ggtatctatt ttgatataat atattaatta tatctatttt gattttaaat 1500

aattaacaga agcttaaaat agcacagcaa taatcgtaat tgtaacccat ttatagacct 1560

accctctaag aatgatttgc aaaatgctgg ccttgatcag aaaaatctga actcacaaag 1620

cattgttacc tctttggcag tcttcaatat ccctagtttc ttacagttaa aaaaaattaa 1680

atttgccatt tcagattgtg cctatgattg gggggctgtt tctcagtgcc tggctcttca 1740

tggttggcca gcacgtggag cactctgttc atctgtacat ataggtatgg gtccatctgc 1800

acctatagtt acatgctcat ctttgccttt aacaactttg acattctgac ttgacagtca 1860

tggtatttta aagcaaccat taaatcttgg ctcctgggga tgcttttgaa agtctctgac 1920

ccccaagccg gggcacttct gctgaactaa cactcccata aatgagaaaa aaatgcatca 1980

ccttttaaat aacatgcccg atctccaaat gtgcaactct gatattaata aaaataaccc 2040

agttttctcg ggagactttg ctaatgcagc ctcatttttt gcacattttg ccggagagct 2100

tttgttctta tacgtctcta ttctcccccc ctttaatttg ttgcaggtgt tgttgctaat 2160

gagctctctc tctctcctcc cctcttacaa tgaaagacaa gtcagacccc aggtacctgg 2220

atggattgag agctgagatc tcagaaactt cataataagt tgccgatggc taccagccaa 2280

ggtcagctgg gccccattag tgccggcccc caccaaagca gaaccaaata gccttctccc 2340

cagtgaacac ctcagtagct tttcattcta gtccagtcac acagctgttg ccaccttaag 2400

ctcatatgaa agaaatctct tttattggtc tgagaaccca agtccagtcc caaagaggga 2460

acatgtttcc agctaacatg cccacctcct gattttattt tatctttaca acgcaggctg 2520

gagggttgtt ttgccgttgt gttgagcacg tcacccatta agagcccttt aaagacctgg 2580

attgattgga aggacaaaaa ttaaaagcaa tctgatccag cctcatgcag gatccctgcg 2640

gattttctcc ttatcccatt tccatccact gtcacaattt gagaatctgc ctgatttgat 2700

-continued

cagattcacc tccaggggag gtgtgatacc agggttagga ggacgtgaag ttatgggcaa 2760

ctttctgatc tgtccatcag cagtctgaga aacgctggct ctgaattttc cgtgtcggcc 2820

ttttggaaac aacaagttcc tcgctgtttg caaagcttca gtgctcgggt ccctgggaca 2880

ccccggccac cctcgcctgg tagatgtggc atttccatgc tgaggccgcg agtcccgcct 2940

gaccccgtcg ctgcctctcc agggcttctc tgggccgcgc ctctgcagac tgcgcagcca 3000

tgctgcatct gctggcgctc ttcctgcact gcctccctct ggcctctggg gactatgaca 3060

tctgcaaatc ctgggtgacc acagatgagg gccccacctg ggagttctac gcctgccagc 3120

ccaaggtgat gcgcctgaag gactacgtca aggtgaaggt ggagccctca ggcatcacat 3180

gtggagaccc ccctgagagg ttctgctccc atgtaagtcc acttactgct cttttgtttg 3240

cccaggccaa gtgggaggag gtctggagga gatccttggg gtggacgagc agagctggag 3300

gactgagatg caaggctgac tttcctgcct cttgaccagg tctggaccag acctggacca 3360

ggtctttgtc ccaccttgga aatgtgtcaa aacagaggac accctgagga cactggggtc 3420

atgtgacatt gttctctggg gtaggggcat tctcggccag ctggccacta gttcagttcc 3480

ctcgggaagc ctatagtatt cagctccgca gccttcaggc taagccccac ctcctgtgta 3540

ggaagtcagc attctgggcg agtgagcaag atgctacctg caacatgata ctgtaagctc 3600

cctctgttca tcctttctgt ggtgcaacct cttagcccac tcaatccaat ccagcagacg 3660

gtaccatgaa gctagagcat gccaagcact gagctggccc tttgcatggg gaggtttgac 3720

ggaggctcag aggggattca cgcaggatga agttgggggt taagctggga cagacagtgc 3780

agcttgggtc cccttccact tccattccta catgcaatga tggcagcccc tgggtaagtc 3840

gggggaaggc agacattcag gcgttgtccc ttgcctccct agctagagaa ggaggggtcc 3900

taggggcacg gagtacttca gtactcaagt aatgttaaca gcaacaacaa cagcagagat 3960

ggttgcgtgc gagctgtctg agtgtgtggg ccgagcctga tgccattgcg aggagtcctc 4020

ccctcagcta gccaggtgga ttctgtgttg tataaagcag gagcgtcagg ggagggctct 4080

ggcccagtga cgtctgtggg cttctgttct gtcatctgca aaatgggcac actaagagca 4140

cacactccta gggtcattgt gaggagtttg tggtttaatt aacgtaggta aagtgttcgg 4200

aataggacct ggcacagagt aagtgcacgc agatgttagc cgttgtcatt ctggtcatac 4260

aggtggggta actgaggcag cccaggggtg acgggtaaag gcatctggcc aaggtcacac 4320

tccaggaggt ggcgaagctg tgattccagc ttagagtggc tccaatgtct ctgagctaag 4380

ctgcttccca ctgggcagtg ctccggaggc cgtccctggc agggcagggc agagctgggt 4440

agggcagcca ggctgcagaa gctcacagga ggggcttgat gccatccccc aggcagctgg 4500

tacctctgcg tgtccttgga ggagcctcca ggtctctggg tttgtggtgg ggctgacccg 4560

gtgcccccac ctcagagtcc tgaggacttg atctcatggg ccggctctgc tcacatcaca 4620

gggcagtcag cctgagaagg tagcttctta ctcaggcttg tcagtggtga tgaggctgtc 4680

actgtgggtg gtggctgggc cagggccagc tggggagaga gagagggagg gagagaggga 4740

aggagaaggc ggcacggagc caggagctgg ggctggatag tctgtggcca taactgcccc 4800

ggggaccgca gggccgagca agggggctgg gctctggaag ccaggaggaa ggccaggata 4860

ggggctggta ctcagtccac atctcaaagc cggtgggagg gttctccaca cgctctcggg 4920

cacggtcaac ctctgtctct cgtattagag tctcgactgt attttctctc ttaaatatta 4980

-continued

atcactgcct taacgtgctt gggggagcag ctaaatcata attctaggac cagctttggg 5040

tcgagggctt gaggtgggga gatgaccctc agagtcaagt ccgaggccct ctccttgcca 5100

aagctgtcta gggtgtcatt ggccctggac ctctgccctg cccccaccct cagacagaga 5160

acccagtgca gtgggcggct gttctgggga ggtggtcacc cctcccagtc ccagtgccgg 5220

cagagcctca tcccaggcag ccagcctcaa gccctggggt ctagaagagt gctctctccc 5280

atccccatcc cctgctcctc tcctgggcag acaggtgggg aaggcagggg agaaagaaca 5340

gtccctccac aatctcccac atgggcaagt cctcggcgtc tcgccacctg tgtgatggac 5400

ttaaatattt catcatgggc tgccatccag ctctgctttg attacaaatg tgtgtccgat 5460

gaggacgggg aggccgccgt ggcaggtgga cggcagcctt ttgcagggct ggcttttgga 5520

gggctggctt tggagggctg gcttttggag gggtggtttt tggagggggt ggttacttct 5580

agatagatct gggttcaaac cctgatgcca cagtttattg tggcttcccc tgggagggtc 5640

cgggcagtag gatgctctga ggtaggtcct actctaatct catttttggg aatagaaagt 5700

gttgcctaaa tcctgccaac atcactccgt tgagaaggca ggctgggaac tgcctgtagc 5760

caaggcccat gcccacccac caggcggcac agcctcccct ctctccactt ctgacctctc 5820

tgagtttcac tgtcttcctc tgcccagtag aaaccataat gtaggagtgt cacggtgctg 5880

gcgggacggc acagcccggc cgtgctgccg tggtacgtgc ggacacatag taggtactca 5940

gtgtgcaata ctagtcacca gcttatcact gttgtgatta tggcctcggt gattttcctg 6000

gctttgcttc ttccccgggc tcctttgcct ctccccagcc ttggggacgc cacgccttgt 6060

attgctagtc tctggcagca tgtggaagat gcaggctggg ggagcttcct cgcctgcagc 6120

cccagcagct gttgttaacc agccaccagg gggcgcccta atggcggcca gatgcccact 6180

gcccccttag ccttgccatc gcccatgggt ccttgctccc cttcgccccc tcagcggcag 6240

gtgctacgtg tgtcccaaca tccagcccac gggagcgcaa gcccggctgc actccgaggg 6300

tacaggaggg acccaggggc ggacggcctt ttcagatgcg gggtgcagac cctctctctt 6360

tcccaggctt cctgtccttc agcaaggcca cctgagaact gaattgtaaa ttccagcctt 6420

gcaaggaagg ggaggaggct gaaagacaag ataaatgaat aaataaaaac ctattggctc 6480

taaatgcaca atgagaatta atagggatga gctctcaaca gaggattctt gggcaaggac 6540

aaaattgcaa tgagggctgt gggagaagag aggcggcccc cacccacatt cccagggcct 6600

gcccctgagg tcagcccagc aaggctgcag gccgagaagg aggcaggaag ggcgacgggc 6660

aaaggcatct cagcccctca cagggcaact gctgcttgca ggacccttgg agatggggag 6720

ggcgtgactg gcattgggag gtgcccctta gcgctctgca agacacgtct cccgccagcc 6780

acctgtggtc cccgtgagag gaaggggggaa acccttcctt tgtgccttgg agacaaaaac 6840

tagccacccg agactcaggg gtgagctgac gagcaggtga gagagagaga gagaactact 6900

gcccaggcag cagggtgcca agggaatcct gccactccca ctcagagcaa tcgcctcagt 6960

gcagcagcgc tgacctcacc ttgatagagg tgcagaatcc cagaccctgc cccacaactg 7020

ccgagccaga ctctgccttt taacaggaag tctctggagc gctggggtgg gggttggtcc 7080

aggatggtgg atgaagcctg ccgttgacga aggggccaga ggtctgtggg tcccatgggg 7140

tctgtgcagg ttaagaggag aggcattgct gaggggagac aggagccaac aggagagcac 7200

taggggggtcc tgggaatgga gggatggagc tgggagccga gattcccaga tgcgggcctt 7260

-continued

gcttgccaag gcgggggtc tgtgcttctg agcaggagcc agtggaaggt tccagagaga 7320

ggcagcacta cctcctgaca agagctacag gcctagaatc tgaagccctg agctcacatc 7380

ccagtgttgt cacttggggg ctgggtggtc tgaggtgagt cattcaaatt ctccaagcct 7440

cagtttccct atctgtaaaa ggggtatgat cttcactgcc ttgcccacct caccgacttc 7500

ctcggagagg aaacgcctct cacaacataa cagaaggtgc ttggcagaga ctgtaagttc 7560

ttttctagtg tgaggcttga atattattgt ttatccaatc ttgataagct tgacatttct 7620

acattgacat gggtgaacat tcaaatttca gccatgttgg agtctccagc atacagaggc 7680

accttggctt tggtggacag gaaagccaag ggcaggctgt gaacgtgcat ctttgaacac 7740

acatcttact cagttagaac ctgctttagc ttctctgtgc ctcagtttac ccagctgcag 7800

agcaaatggt tccattcctc ccccaggggg tctgctgggc tgttgaatgc agacagcagg 7860

caggtcctgg agtccttgag ccactcctca ctgtgacatt gaacaccaca tgcagaatca 7920

aggtcacaga cactctgcct cctgccagac cttgaggatc atgaaatgtt tccatccaac 7980

aatgcagagc tgaccggccc gatggctctg gcctccttcc ccctctgtgt ttctggcccc 8040

agcctccctc ctctcgccgt cacaaccctc cagtcggcag atgtttccag gtatgacctc 8100

atgccaggca cagttgtgat gggaaagtga gatgactaga gataaattaa gtgaaaaaca 8160

agagaaaaac actacagaca ttgaagacaa tacccagcta accacaaatg gtgacaccgg 8220

ggcctgatat gataggagct ggggaaggag gagggcgtgc ccctccccat gaactctccg 8280

ggcgctccag ccttggcccg tgctgtctcc tccacccgaa tgcctgcccc ctccttcctt 8340

tcttttcaag gcttccgctt ctacctggat aaccctggct ccttttttcaa ggttcaggtt 8400

ggaccctacc tcctcctcca ggaggccttc tctgactcct ttgcccagtg gggtgaggag 8460

gccctcattt ggtcatcatc ctgcttactg actggcctat cagtgaagag ttctggccct 8520

aggaatggag agtttagaat cccgtgctgt gtgatctctg atgtgagacc taacctctct 8580

gaaccttagt ttctcaatgg gagctattaa tagaacgtgc ataggagact gtcatgagca 8640

cggtgcttgg cacatagtag ttgctcactg cgtgctggct gttgggacca attctgtaac 8700

tttagatgtc tgcttttccc ctggactgtg agctcctcaa aggcagggct gtctcgctat 8760

ctttgcatcc ccagcacctc cctctggact tgatcagtgc ccagtaagca tttcctgaat 8820

gaatggatgc atgggaggaa gcatggtgtg atccaggaac gcttctcaga ggaggcagga 8880

tcagggctgt attttttttt ttttttttt ttttttggag tcaaagtctc gctctgtcgc 8940

ccaggctgga gtgcagcggc gcaatctcgg ctcgctgcaa gctctgcctc ctgggttcac 9000

gccattctcc tgcctcagcc tccccagtag ctgggactac aggcacctgc caccacgcca 9060

ggctaatttt ttgtattttt agtagagacg gggtttcact gtgttagcca ggatggtctc 9120

gatctcctga cctcgtgatc cacccacctc agcctcccaa agtgctggga ttacaggagt 9180

gagccactgc acttggccag ggctgagtct tgaaggaaaa actggggttt gggtcaggac 9240

agaggagacc ctggaagccc ctgcttctct ccactgcagt ccctgttctg tgggatttgc 9300

gattggatga agccgggagg tttgcacaac tctgtcctta agtcagttgc aagtgacttc 9360

ggcacctgag ctgcaccagc cgttaaagcc actcagtctc ttgaaatgcc cgaggcaggg 9420

cccagcctag gacaagaata gttctgtgaa atgacatctt gttgcacagt gaagtctccc 9480

tcctgggcag tagacaatga gaagaccgag gcccggggcc cagggagtga gaccttgct 9540

-continued

tctgacttcc cttgagggaa tgaggttggg tccagacacc ccgtggaagg caggcagctg 9600

tgtgaaaggg cccagacggg acatctttcc aaagaatgtc agagacttag agacccccag 9660

acctttccgg tttgcgcatc cccaccttcc caggctgtct tcctctatgc ttcctaactc 9720

tgatgtttaa tccatttccc tttttctcat ttactgtggg tataatgaca agctgcctcc 9780

aatcccacct gcgatggggc aggcagtgga cggatggaca gacgaacgga cagacaggca 9840

ggccgcacca tgctgcggat gagacggatg gatggacaga cggacagaca ggcaggagca 9900

ccatgctgcg gatgagatgg atggacggac gaacggagag gcaggcaggt cgaaccatac 9960

tgcggatgag acggacggac ggacagacag gcaggccaca ccatgctgcg gatgagatgg 10020

atggatggac agacggacag acaggcaggc cgcaccatgc tgcggatgag atggacggac 10080

ggacggacag atggacagac aggcaggagc accatgctgc ggatgagatg gatggacaga 10140

cgaacggaca ggcaggcagg tcgaaccatg ctacggatga gacggacgga tggacggaca 10200

ggcaggcagg ccacaccatg ctgcggatga gatggacaga cggacagaac ggacagatag 10260

gcaggccgca ccatgctgcg gatgagatgg acagacggac agacggacag acaggcaggc 10320

cacaccatgc tgcggatgag atggacagac ggacagacgg acagataggc aggccgcacc 10380

atgctgcgga tgagatggac agacggacag ataggcaggc cgcaccatgc tgcggatgag 10440

acggacggac ggacaggcag gtcgaaccat gctgcagatt agacggacag atggacggac 10500

ggacagacag gcaggccgca ccatgctgcg gatgagacgg acggagacag acggacagac 10560

agcaggtcga accatgctgc agattagacg gacagatgga cggacggaca gacaggcagg 10620

ccgcaccatg ctgtggatga gatggatgga tggacagacg gacagacagg caggccgcac 10680

catgctgcgg atgagaactt gggcttctgg agggaggaga tggggcccgg gggcatcccg 10740

cacttctggg atgtgggact tgggataagt cccttgtgac cctgagcctt ggttttctca 10800

tctgaaactg ggcatggcgc tggacacaac ctcgaaggac gtgtgtacaa ataagacgag 10860

accaggcgtg tgatgacctc agctgggagc cagcacaaaa ggaatgctca aaaaaagggc 10920

cgggtgcacg gtggctcaag cctgtaatcc cagctctgtg agaggctgag gtgggcggat 10980

cacctgaggt caggagttca agaacagcct ggccaacatg gggaaacccc gtctctacta 11040

aaaatacaaa aattagccgg gtgtggtggc gcgcgcctgt aatcccagct acttgggagt 11100

ctgaggcaag agaatcactt gaacctggga ggtggggttt gtagtgaact gagatcgtcc 11160

cactgcactc cagcctgggc gacagagcaa gactctgtct caaaaaaaaa aaagtgtatt 11220

tttatttta tccttttta ttctagaatt tagcttgagg gacagaagag gaccccatag 11280

gccaaaccca cagccagagg cacaggctgt gggctcagaa gtgggtcttg caggatggga 11340

agggtcagga gagtgaggat gtgggcagaa ggaatggttc gtgcagagac gcagggaagg 11400

gggccaggtg attcgagggg aaagtgcgtg ggtgacagag aggagacagt ccactccccc 11460

tgcccacctc atccaagccc ctgtaggtct gttactgtgc atctgaccgg tgaatattct 11520

gagacttctc agagcccact gagtgtagga gctggggttc agccttcctg tgtctggctc 11580

ctgaccgctc gctagggtta ggaaggatta ggccacgggc tctgaaggag caagaggggc 11640

aggagggcaa ttgaggggca attgagagga acccagaaca tggaagccct gtgccgtggg 11700

gctggtccag agctcaccag gctggaccac gtggttgctg agccatggcc cctgaccggg 11760

gctgacctgg ccagagtccc tgtggccagc actgatgcag ggctccttcc tagaggggcc 11820

-continued

gggccatgag gaacgggaga aacggcagat gatgcgggaa ccggtctgtt cggctttggt 11880

ttgcaggatc cgatttgttt ttcatcagca gcagatttgc ttaagtatat gaaaatgtgt 11940

ttctaattcc ccgagcacac accaactgct ggcgggggag ggagcagtgc ataggagcag 12000

agtgaatgcc accgggagtc agagtgctag gccctggctg ctgagagagc gagaatacgc 12060

ccccagcctc agtttcccca actgagcagc cggggaagat ttggctagat taaccagttc 12120

attcaatgtt ccctgctgat tgccaggtac attctgggag tttagggaaa tccagattgg 12180

tcagagacaa aaccacacaa aacagtggac tccagtgcag acagaggggt cctagatgta 12240

taccccgggc tcagcatagc aataatcatt ttaaaaaaga ttttaaaaca tttttaaaac 12300

tcaggtgaag ttcacataac ataaaattaa ccagttaaac aacgttttgg tgggtgcagt 12360

ggctcacacc tataacccca gcactttggg aggccgaggc aggaggatca cttgaggcca 12420

agagtttgag accagcctgg gcaatgtagt gagaccccat ctctcaaaaa aaaaaaagtc 12480

ttattgtgtc taacataaaa cttgcctttt aaactatttt acaatataca attcagtaca 12540

ttcacaatgt tgtgcaacca gcatctctac ttagtcccaa cacgtttcca tcgccccaat 12600

agaaaaccct gcacccgtta gttactcccc atctccttcc cctgcccctg aaaaccacgc 12660

gtctactttt tgtctccatg aatttagcta tgctagacat ttcatacgaa tggaatcaga 12720

caatatgagg ctctttgtga tggccttcct tcactggcaa aatgtttcca aggtttgtcc 12780

acattgtcgc atgactcagt gcttcattcc tgtttatggc tgcataatat gccatcctgt 12840

ggacacacca tattttgtgt atccgtttcc taactgatgg acatttgagc tgcttctgct 12900

ttctggctat taggagtgat actgctgtgg acatttgggt ctcagttttt gcatgtgtgt 12960

atgtcttcat ttctcttggc tgtctaccta aaagtggagt ttctgggtca caaggtaatt 13020

ctatgtgtaa ctttttgggg agccaccaaa ctgttttcta caggtgctgc acctcttacg 13080

ttcccaccag caatgtacga gaatgccagt ttctccgaat ccttgtcaac acttgttatt 13140

ttctggtttt gttttgtctt attaggatga gcctagtggg tgtggggcag tatcccatta 13200

tggtcttgat ttgcatttcc ccgatagcta atgatgtcag tgtgcttctt agtcattttt 13260

ttgtttttgt ttttgttgtg tgttgttttg agacagagtc tcattctgtc acctgggctg 13320

gagtgcagtg ttgcgatctt ggctcactgc aaccttcacc tcctgggttc aaatcattcc 13380

tgcctcagct cccaagtagc tgggattaca ggtacacacc accacaccca cctaattttt 13440

gtgtttttag tagagacagg gtttcaccat gttgcccagg ctggtctcga actcctggcc 13500

tcaagcgatc tgcccgcctg ggcctcccga agtgctggga ccacaggcgt gagccaccac 13560

gcccagccta ttttaaattt aatgaactcc aatgtgtgta tttttttctt ttgttgcttg 13620

tgcttttggt gtcatatcta agaaaccact gctaaatcca aggtcagcag tatttacccc 13680

catattttct tctaagactt ttatagtttt agctcttata tctaggtctt tgatccattt 13740

tgagttaatt tttgtatctg gtgtaaggga aaaggtctat ctttattctt ttgcatgtgg 13800

agatccagtt tccccaacac tatttgttga agagcctatt cttcccccac taaatgttct 13860

tggcaacctt gtcgaaaatc aattgagcat aatctatgca cttacttctg gactctcaaa 13920

tctctgggtt tttttgtttg tttgtttgtt tgttttggag tcagagtctt gctctgtcac 13980

ccaggctgga gtgcagtggt gcgatctcag ctcactgcaa cctccccctc ccgggttcaa 14040

gcgattctcc tgcctcagcc tcccaagcag ctgggattac aggcactagc cagcacgccc 14100

-continued

agctaatttt tgtattttta gtagagatgg ggtttcacta tttggccagg ctgatctcga 14160

actcctggac tcaagtgatc tgcccacctc ggcctcccaa agtgctggga ttacaggcgt 14220

gagccacagc gcctggcctc aaatctattc ctctgaagca tcaagcattc tatgtgcact 14280

acttcatgaa accctcctgg atattctgca ctgtagaaac gattactctc ctgttgtgcc 14340

cattttatag atgaggaaac tgagactcca aaactgagtg aagtcaaggc tcaaactcag 14400

atcccagtca tttgatgact aggccacagt gaggcctgag gaggggaaaa atcccaatgg 14460

ttaccctccc cttcccctcc ccaccctcat tttcttctcc ctctttcagg ctgggatgtg 14520

gacttggatt ctcagagcag ggtccttgga aggagatgct gtgacttctc tctggcctcc 14580

aaatacctcc tcagcctcca gtccacctcc gtccctctcc cacgcagcca ggcactgttc 14640

tgtcctcttc cttgtcccac agtcagtgct tgcatgtagc aggtacttaa taaatgctga 14700

agataattat ccatcatttc aaatagagac acacaactta gaaggcatgc tgggattgtc 14760

taaggccaga aaaaccccaa tgtcgataag catgttacag tgaaattgac tgcgcccagg 14820

aaaggggacc ccagaagcag gtggctggtg tccccctacc ctgccccagg ccccgagttc 14880

cccaatccac cactaggaag tcctgggctc ctgtgaagac aatataaaac cactgattag 14940

gccaagtgtg gtggctcaca cctgtaaatc ctagcactct gggaggctga ggcgggcgga 15000

ttgtctgagc tcaggagttt gaaaccagcc agggtgacat ggtgaaaccc catctctact 15060

aaaaatacaa aaaaaaaaaa aaaaaaaatt agctgggtgt ggtggtgcac acctgtagtc 15120

ccagctactc gggagactga ggcagaagaa tttcttgaac ctgggaggcg gaggcagagg 15180

ttgcagtgag ccgagactgt gccactgcac tccagcctgc acaacagagt gagactcggt 15240

ctcaaaaaaa aaaaaaaaaa aaaaactact gatgaggcac atccccccct ctcatttcct 15300

atgaaggaga aactgaggcc cagagggttg gagtgacttc cttgagcccc ccatgaggag 15360

cttcagaccc tggaggctcc accccaggcc aagggctctc ccagaggtag actggagcca 15420

tgaggacagg ggccctcccc aaccaggtct ctgtccatct acacgtgccc tggatctgac 15480

ttcacgtgat ggcatctggt gggggacaca ggatgcctgc ccggatgcca cctgcagcca 15540

gtgggggccg gagctgcctc ttcagggtca gtgagggtga tacatctact tcccagcctg 15600

cttaggtgag ctcccgccta tgtgtcacta ctggtgactg gcatggctca gagccagatc 15660

ttgggggccc tgaggggatc aagagcgtcc cctaagccca cctgccagct gcggtcttct 15720

ctgtggtggc agcatcacag aaagtggaca gaaagagtgc tctgtgccag gagggcaagg 15780

ccgggtagga tggtggctgg aatgctggcg atcgcagcaa tgccggcgat catggtgctg 15840

ggttttggtg gtgtgctgga cgcctgggag cctcatgagt gagagactgg ggcacacgtg 15900

cttccgtagt gccatgcacc ggtggcaatt cagagaaaga cgctgtgcaa agcaccccat 15960

gtgtgcagct ttttgccctc tcgtaacagg acggagccag gtcagagtgc agatgaggag 16020

aggaaggtgc agggaggtgg agatgctgac ccaagtttgc acagccaaaa ctaggatcgg 16080

tctccagggc ctccgtcact gtcctgtcct gccttctgtc acacaggagt tcgaatggtc 16140

gttctgaaat tgagagctag cggggctggg atctcactgg gcggccacag aggggtcccc 16200

tgacctcttg gggtctcgtt ggcaggaggg aattgtattg gaatatccag gtgtgtggat 16260

ccctgtgaat ctaaccctgg agtgttccag aactgcccac cctgtggaaa gggactcagg 16320

cctgtcttca aggacctggc atccttctgt cccagggcag tttgtcttgg gtctctcagg 16380

-continued

gaccgtttgg gcctcttcag cccctcattc cacttccctc ctgctgccca agtcattcgt 16440

ccacttgact ccaagagtcg gctggggaaa taaaaggaaa tgaaacacga ccaggcattt 16500

tcccttggcc gaagcagaag tctgctgtcg ggcaaaaggt gaagaagaga ccaatgagag 16560

atgagcccac ggtgctcctg ccctccgcca aggcaggcca tcctctgctg ccagcctgca 16620

acagggcagt gtccttctgg gaggtgtccc tccctctggg ggatcaagag atggccaaaa 16680

gcaggtggca gcaagtggag aaggctgttc atccagaacg caccttgtct ctgcccctgt 16740

ccccacccag gcaacatcca aaacctttgc ccacagttcc ggggctggca ccgtcctggg 16800

gctcagctcc tagggacggg gctcccccag gcactggctg ccaggaactg ggtggccccg 16860

ggcaagtctc ttcccatttc ggggtataga cttcctgcct gtaaaatgag ggggtctgca 16920

ggtcaacctc agagtcccac tgtaccccca gattctgctt cagggagacg gagagagaga 16980

gaaagagaaa gaacgataga gagatgcaat aacctcccag catccaggaa gacccagagg 17040

ggagaaatgc agggaaccta cccagaaaac cctggagcgg gagcttctca cttttaatgg 17100

tcatggcccc acttgagaat ccatggtgct cttcctagaa ccacgcatgt gcacacgtgt 17160

gtgcaaacac tgggctcatg cacaggcaca cacacacaca tataaggttg caaacacttt 17220

cagggacttc ccagatttct ctgagtccat ccgtggtcac tttcggtcaa tcatctggcc 17280

tgagcaggtt ctgccttctg ggggctcttc taccctcagg gaaatcaggg tttggttccc 17340

tgtaattgtc tggtccaatt gtctgaggac tttctctttt ttttgagaca gggtctcact 17400

ctgtcaccca ggctggagtg cagtgaagca gtcttggctc actgcagcct cgacctcctg 17460

ggctccagtg atcctcccac ctcagcctcc tgagcacctg ggatcacagg catgcaccac 17520

catgcctggc taatttttgt atttttgtag agatggggtt ttgccatgtt gcccaggctg 17580

gtcttaaact cctgggctca ctttcttttt ttttttttt tgagacggag tcttgctctg 17640

tcacccaggc tggagggcag tggcatgatc ttggctcact gcaacctcca cctcccaagt 17700

tcaagcaatc ctccttcctc agcctcccaa gtagctggga ttacaggcac ctgccaccat 17760

gcctggctaa tttttgtatt tttagtagag acaaagtttc accatgttgg ccaggctgat 17820

ctcctgacct taagtgatcc gcccacctca gcctcccaaa gtgtcgggat tacaggcgtg 17880

agccactgtg cctggccata gccagacttt cttgattcta tatccttctc ctcagagcag 17940

aaacatcgag catttgttga gtgcctcatg tataccaagc ccttaaccta agctatagct 18000

cattgaactc tcacagaagt cttaaggtag agcttgtatt tagatccgtt ttgaatatga 18060

ggaatctcag gttcagagaa tttaagccac ttgccgaagg ccacacagct tctaagtaga 18120

ggaggctggc acctccagcc tgggccgccc ggcccggcat ccaggttcct aaacaggctg 18180

ctcagctgac acgagtcgct ctggatctca gaaagcgcct gatagacgga ggcggctgtc 18240

atctgtgtgt gtgtgtgcgt gcgtgcagcc atgtgtgctg gtgagcatac tgtccctgtc 18300

agcctctcct cccccagcac accccggcag cccagagaag ggagggcccg gaggagtgac 18360

ggtgttcccc acccccctgcc ctttgagaca caatggagtc cgctaatcca gttacttgat 18420

aattcactta tttcatgtct atttggcagc gagcgtgctc ccacgcacca gctctgggga 18480

aggcgagatg gctttgcctg gaggaacctg attgtttttc tggggaagga gtgggggaaa 18540

aaattgcacc caacaatgga caataatggg cctaaaaata gagggtggag ggtgcagggg 18600

gtggaggagt gtgctgtctg ccaagggagg gctccaggcc tgtctgcttg gcacggggca 18660

-continued

gccctaccct cctgcccagt tcccctcccc tgcactgggt tggccgcctc tcagcctaga 18720

ggaggggcac tggaaggagg aggcccaagt gggtgggggg ctgggtggcc ttccttgctg 18780

tctctgcccg ctccaaaatg gaaaattgtc ccacccaggg ggtcctggag gcagcagcca 18840

cgaccttggg tggacgctgc gcctcatcag ccctgactag ccgtgatgcc caggaccttc 18900

ccccaagggg ctcaggcatc agctgagaac tgcagccttg ggtacagagt acgggttgtc 18960

tcccagcagg aaagggaggt ttcaggtttt gtggctcttt tccatctccc aacacttggg 19020

gcagtcttct cgaaggcctc aagcccagcg ggcagctatg accccaccag gagcggagcg 19080

ggcagggacc aggctgccct ctaagccact cggctggctc tcagccgggg tgcacactgg 19140

acttgcctgg gagcttttca ttccccccct gccgcagctg ccccccagac cagcttcagc 19200

agcctctctg ggcggcccag caggaggagg cattaaacct ccccaggtgg tcccagtgca 19260

cagccaagtt tgagaagcac cgattgaaac ctctcccagg cctgccctgg agcccttcca 19320

gcctgaagca tcttgtcgtc ttaaaactga aagaccagga ggaagagaat tccatggcct 19380

gcctcggctc tctcggagcc tctctcacat ctgagctgca ggtgctccat cctccttctg 19440

gcttcctggg tgccgagggg tgccagctct ccaggcttgg gagagggcca cttaagccct 19500

cacactttgt tcccaggctc ttcacctgtc ttcctggaag gaggggccg gccagcatta 19560

gggctgtcac gggcgctgct taatgtcaag ctgcccatct ggctcctggc ctccctttgg 19620

ccttctctcc tgcgctcccc accaagctcc tggctcagca gcgtgcatgc gttaacccat 19680

tgccccctg cagtgttttg tgtgtccagc ctggcccttt gctcagtcga cctagagcac 19740

catcctccca gactagtcga gtgtccctcc acctgtcctg agtccagatg aaatcccacc 19800

tcccccagga agccttctga ctgccccagc ccgtcacctc cagggcttgt catctgtgcc 19860

actcatgggg accaggacac aggtgacttc tctggtggac acagcagaac ggtcaacatt 19920

cccaaaaggg agcaaattgc ccgagtcacc agaagtgtga ccttgagcaa gcatctggct 19980

caggggctct tggcttcccc acctgtaggt aaaataacag gaacagtgtc atcgtgtggg 20040

ggcccttccc tggaccacct ggaccagcct ctcaaacctg gccacacatc tgagtcacct 20100

gctgagggtt ttggttagtt ggttggttaa ttactcgatt ggttagtttg ttggtttgtt 20160

tgttagcttt caatttggag tagaatttct gaccttaggg cccaggaatg ttgccacacc 20220

cctcccgtcc ccaggtattg tcatgttttc atgggtggcc aaatctaaga gctgcttctc 20280

tggggcacga aggatgttca caaatagttg atgaataagt gaatgaataa atcaatgaaa 20340

cttaccagcc cagcctcact actcgcaccc acccccaacg accagccagg gttcatccac 20400

agaggggtgt acctgtccag gtgtccccag gtgtgggcag acccagtaac tttactcttt 20460

catcggcccc accgcctctt aactcctcag agaccagcag gaagaaaccc tcggaggtcg 20520

cagcttctgg ctgttctcag gggcaggccc cgtccatcgg gtgctgtgtc tactcctaag 20580

acctggttct gagtatggaa cacctggaga gggaaggggc cgaggagggg gagtcactcg 20640

gctgtgtcag gctccgcccc tgccttcctg aagcacacag tggggagggg acacacccgt 20700

cattaaccga agaagccact ggggaaaact gtgacccagt gctcccttgg gactgggggg 20760

cagtggccag gggtgttttc cctgaggaaa agaaatttaa gcagacacct gccaaaggct 20820

ggagggagag ctgtagacag aagatggctc aacctgaaag ctccgcgggg tggagggggc 20880

atccaaaggg cgggagagac tggccagtag aaaacgaggc cagaagccgg acatggtggc 20940

-continued

tcacgcctgt aatcccagca ctttgggagg ccgaggcggg tggatcacct gaggtcagga 21000

gttcgagacc agcctgacca acatggggaa acccagtctc tactaaaaat acaaaattag 21060

ccaggtgtgg aggcacatcc ggtaatccca gctacttggg aggctgaggt aggagaatca 21120

cttgaacccg ggaggccgag gttgcagtga gccaagatcg taccagtgca ctccagcctg 21180

gacaacaaga gcgaaactct gtctcaaaga aaacaaggct ggaaagacaa gggagagagg 21240

cagagctggt ggcagagcca ggcccagggt ttggacttaa gaaggggaag gcgctggggc 21300

tctgaggagg agctggagac aaagggaagg gtgtctggct gcaggctgag aatgggctgg 21360

gggtggggag gctagagggg tcggtggagg ccccaaccaa ggcggggcag gggagaacag 21420

agtgagcctt taggagaaac cactgctggg gaccaggctt cactctcagc ccatccggaa 21480

acctctttca gctgatgctt cccccgaccc ccttccccac cctgggcctc tgttggaagc 21540

tggctgaggc cctgttgatc agcaagaaag cccagggcag ctctcagaga agaggagagg 21600

gggcccagaa aaggccccca ggatctgggg aggggatccg aggagaggca gctaccaagc 21660

gccccagcca ggggggcctg tccctcaccc caccccggca ctgaaaccct cacagccact 21720

tttttcctcc ctccgtgata aaatattcat ggcggcagag tgggccctct ttgggaaggc 21780

tgcctgggtc tagcttatgc tctgcacaag cttttaaaga gcagggcgct gttcctactc 21840

tctaagcatt ttctaagtcc tgaatcaata atgcactttc ctgggcttct ccggatgtag 21900

ccctcttcct ctcggtgcct ttcccccgcc cgccccctta tctttctttc ccttttcttc 21960

tctcccgttc tttcatcctt ccatctctcc tttcatttac tttttaaaaa agtatgaaag 22020

tgttgagctg tttgggtggc cagtgaagcc ctgagtaggg agtgggcagg aagggaggcg 22080

ccagactgag cccctgtgtg tgcagggagg aggaggagga gtgggaggag gaggaggagt 22140

gggaggagga ggaggagtgg gaggaggagg aggagggaaa ggaggaggag cagcagcggc 22200

tgagcgctcc cgctggccct gctagggaag tgtttgagga tcactgagct cctggtgtgg 22260

ggaaggagga gggcttagcc tcacccggcc tccctctctc ctttttctaa tcaattagaa 22320

agtgtttaca gcatagccag agaaaaatag ggaacctggg accaagaaaa aatgcaaagc 22380

accggccaaa tttcagcccc acactcgaag gagggagcag tggggtttca cctatctgcc 22440

ttctgtggta atgaaacccc tgtcgctaga ggtatgcaag agaaggggag ccttaccctg 22500

tctgagacag acgcccatgt gtctggtcca ttctgtcagt cgcctgtggg gtgccccagg 22560

gatgcacggg cactcttcaa ctgggataga atttcctgcc ccaaacattc ctggaaatct 22620

ggctgtggga agaatccaca tatgcccagg gcaaagcaga atgtgtcctt taagaaaaca 22680

ataatacatt tttaagttcc tggagagatt aacccttgtc tagccagagc catggcaatg 22740

cctccccgcc caccacactc tggtggttcg gctgacggag gagatcagtc attcaggggt 22800

ctgcggtcct gatgagcagt gggtgcccac accaggcctg gcatttcatc cttgctttct 22860

gaccttggct tcccagttga ccctctcccg ggcagctcgt ccatcagggc agcccaatgc 22920

cctcaggtcc tccgaaagga tctcagggtg ttctgtgggg gcaacccgaa ttggtgtaag 22980

aagactaagc agtcgatctg ctggaacagc atccccaaag cggagcgaag cccgcggatg 23040

cccaccgcct ctcccccagg cagcgtccta cctggataga actgcctgga gccactgcag 23100

agggtcctcg ctcagttagg gaatgtttgt catataccgc tgtgtgcaaa cagctgttgg 23160

gagtgtggcg caaaggtggg taaggcccct gctctcccag agttcacact cacagaaggt 23220

-continued

tctggaagga ggaacactgt gggcagggtt gaaaggccta aagtgctccc tttcctccca 23280

aataatgcgg ggtgaggggc ggtgaggaga gccgctctga gcaaccaagg aactgagatg 23340

cattttctgg gtctcctttt gagccgaggc aggttcgaga ggcagccaga gactctgggt 23400

tcaaggtgga cctgtgccca ggccatgccc actgtggccc ccctggggga ggagcagggg 23460

cggtcgccgt ggctttggga ggctcattgc tgggacaagc gagtccctgg ggaggcagcg 23520

cttggaggct cgcttgcctg cccctgcctg agtaattgct tggagctggg aggaaaattg 23580

ctccaaccag aaaacaaaac agaaaagccg ccttggccag ctgcagctcc agccctaaaa 23640

tgccaggttg gtttacgctg attcacgagc ggggagggtg accttgctgt ctgttgtcca 23700

gggcctgtgc acgaagagaa tctggaaagg gaaggagaga gacacctgca cgctggggaa 23760

ggaattagca gcacagagag caagagggac agcgatcaat gaaaccatag aaggagaatg 23820

agaaacacac acacagagag cgagagggag caagagagag agagagagag agagagggac 23880

agaaaacgag agggagggag ggagggagag ctcagagagt tagagaccgt cagggccgct 23940

agaattagaa tcagctctga acagaatctc cgtttccgct ttgttaataa tttattccct 24000

ctgcaacttt tcttaccaat aaataggaag taatctgtta aggagaattc ccctagcacc 24060

ccggctttct ccctggagtc aggggaggag gatgtgtctc tgtgcccttc ctccctagca 24120

gcatgggggc ctgaggaaca cgcagaactt cagactttag gatgtcaggg tcagaggcgg 24180

acagcccact cctgcccggt cattttgtga acggggaaac caaggcacag atagggcaag 24240

gccctggcca aggtcacaca tggtgttagg ggcagtcccc tgagtcctaa ttccatggcc 24300

ccacgggtca gggcacctat tgatttatgc acctgcccaa gccatagggt ttcccccgaa 24360

atggcagagg ccacatccaa ggaggagggt ggggctagct cggctgcctt tccttgcctt 24420

cccccacgat tgcttccccc gtgctcgagt cctggccctc tacctgggca cccacaccca 24480

gggcctctcc tgggcagcct ccagccttcc accttgtatg cggcagcagc ctcccgtcct 24540

ggtgaggctg aggggctgag gatgagaagg gttccgttgg caaatcagca acagcagtca 24600

agagacgtgc cgcctgcctc cccgtggaac ccgagtctgc gggagcacag tgcggcccag 24660

gcaacagcgt cctttccctt ttgggtgaag ggcaccattt cccaatttgt ctcagggccc 24720

agctcagtgg gccatcccct ggcttcttat cccacctcag ctgctgccga gccgcatgac 24780

cctgcgacat tgctcagcct ctctgagtct cggtttcctg aggatcgcac tctccaggat 24840

ccctgggagc gtgggaggtg gggttggggc acacagggcg cccagcacag ggccgaggtg 24900

gaagacatgc tccctaacgg cggggcctgc tgtttgctga agcaccaggc cagacagtgg 24960

ccatgaatgt gctcccagca tccatcaccc atgagctggc accaccgagg cacttgccat 25020

ggtgcacctg gcatcattcc tatgacaacc ctgtgaagcc agtgctagta acctcattga 25080

gcgttcattc attctccgaa gatttcccga gtccctgagg agggccgggg gctggggctg 25140

gagtggggac aggatcagat gtggtcgctg cccgcatgaa gcctcccctc caacagagaa 25200

gctgaggctc tcgggcagga gaaagatctt ttcctcaccc attctatgtt agtggctgag 25260

ggcccatcat aacagacaaa ttaataccac caaagcatac cagtggattt aatataagtt 25320

ttatgtgaaa aaggctttaa gcctttcttt ttttttttt ttttttttt ttttagacag 25380

ggtctcactc tgtcactcag gctggagtgc agtggcacag ttacggctca ctacagcctc 25440

gacctcctgg gcccaaggga tcttcctatc tcagccaccc aagtagctgg gaccactggt 25500

-continued

gtgtgccacc atgcccggct agttttcttt tttgtttttt gaggtttttt tctgtagaga 25560

tggcatctcc ctgtgttgcc tggcctcatg ggagctttca taaggaatga agacccaaaa 25620

cattggtgaa catctatttt gtatgctagg tttaatggag aaatagtcat ggagaagtac 25680

gattggctta aaaaaaagta tcatctcctg gtgataaact ggcgggaatt ttgcaagacc 25740

tgtgtgtcca ggtccctctc tgtgaccctg catctttgga gatgagaatg ttccttcctc 25800

cgggcattgg gagggcacct ctcgaatgag cctcatgtcc tgcttcaggg aagaagggca 25860

ggggaaggtc aaagagtaac cttccgcttc tgtggttttc tcaaatccct tcagcttaaa 25920

aaaaattatt tttttgagac ggagtctcac tctgtcaccc aggctggagt gcagtggtgt 25980

gatctcagct cactgcaacc tctgcctccc aggttcaaat gnnnnnnnnn nnnnnnnnnn 26040

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 26100

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ngggaggccg 26160

aggcgggcgg ctcatgaggt caggagatcg agaccatcct ggctaacacg gtgaaacccc 26220

gtctctacta aaaatacaaa aacttagctg ggcgtggtgg cgggcacctg taatcccagc 26280

tacttgggag gctgaggcag gagaatggtg tgaacccggg aggcagagct tgcagtgacc 26340

cgagatagcg ccactgcact ccagactggg cgaaagagag agactccatc tcaaaaaaaa 26400

aaaaaaaaaa aaaaaaaaaa agaaggtttt agggtagcat gtcctgggct tcattgggct 26460

caaggaaatt gggggactcg cccaagcctg tggagctggg ggagctgctg aggaagtgag 26520

ccagtgcatc cttccacaaa ccacagagcc ctgcccggga caccctggcg atttcatccc 26580

aactctgaga tttcatccca gcgcaggctt gggtgcaagg ccagctgcat gacgttgtct 26640

gttcttcttc ctgagacttg gtgaccattc cagtgaccgc ccctccacgg cctccttagt 26700

caggggcctg gaagttcaaa tggctgggct tcccacaggc acatttacct ccattcttgc 26760

taacagttca ctttcccatt caattatgtt ttttgcctgc agcttgccta tgatgtttac 26820

aacctggcct ctgactttga ctgtaccctt tgcacagaaa ataacataaa aggaaaagca 26880

tttaagtgcc catgacgagg cttgaaaagc atggcatgaa atggtttctc catagaatat 26940

ttcatgccag gagctcaggc ttggcatctg tgtaggaggc tcctcctcac ccggttctct 27000

ggtgctatgg gcctggctgg ctgccttccc ttccctctgc ccaccctcct ttgacttcta 27060

gccacactgc ctgacctttc aggtgacttg cctatgtgtg tccagagaga taggaaagct 27120

gtagacatga tgggcttggt ttccccaaga ttcctcaagg ttgggtcccg tggagtccag 27180

gggatgggta agtgatgtgc ggccatgggt acttagtgtc ttatctgaga cgtggagctg 27240

actgtagcac ctgcctcctg tgggttgctg agagggtgca gggaggagcg gctgtgaagt 27300

gcccaaccca gcccgggcac agggctggca ctcagtgaat gttacaatca tcacactctt 27360

ctgagtcagc cgttcccggg acagtccacg ccatgaagtt ccagggtttc tcatgcacaa 27420

gcctggtggt ctcagcctca tcccttcctc ctgtggaagg ttgctgggag tggagtgtcc 27480

ctgagttaat acgaagctgc tgtttcaaaa taatcgctcc cgtttggaga cacttcctag 27540

ggattctagg taatgtgaga cacagaccat ctcacatggc aatcaataag aaaatagaga 27600

ctcagagaag tcaagtgact ttgtcaaggt cacagagcct caaaggagag agctgagagg 27660

gaactcaggg ctttcagacg ccagggccca aaactgtatg gaaaatgggt atgttaactg 27720

catctttctt ttttttttc ttttagacag agtcttgctc tgtcgcccag gctggagtgc 27780

-continued

tggagtgcag tggcgcgatc tcggctcact gcaagctcca cctcctgggt tcacgccatt 27840

ctcctgcctc agcttcctga gtagctggga ctacaggcgc ccgccgctac acccggctaa 27900

ttttttgtat tttttagtag agacggggtt tcactgtgtt agccaggatg gtctcgatct 27960

ccttacctca tatccacccg cctcagcctc ccagagtgct tggattacgg gcgtgagcca 28020

ccgcgcccgg ccattaccgc atctttctag agaaaatccc aagactcttt ttaaaaatca 28080

gcggtatgat ttttgttgtt gttttaattt tcatgaaata tttaaagagg cagctactac 28140

ttctgatact atcaaagggc ggctttggag ccatgctgaa aggctagagg tgtgcctaac 28200

agtctctccc tatattaggc cactttgttc tgactgctgt ttttgtgatt agttgatctg 28260

ctgtcctgga gatgggtgga aacgcgcaga cccagaggag gctggcaaac cttcaccccc 28320

actcagagga gacactccca gtctcagccc caccccccatg acattatcag ctgtcagatg 28380

ctgactgggg actggggtgg ggggactggg gtaaactggt cttctaatac cccggtgggt 28440

ccccagaatt cccccagtga agaatagaag gtccaggtgc aatggctcac gcctgtaatc 28500

ccagcacttt gggaggccga ggtgggcgga tcacctgagg tcaggagttc gagaccagcc 28560

ttgccaacat ggtgaaaccc tgtctctact aaaaatacaa aaattagcca ggcgtggtgg 28620

cgcacgcctg tagtcccagc tacctgggag actgaggcag gagtatcgct tgaacccagg 28680

aagtggaggt tgcagtgagc cgagattgtg ccactgcatt ccagcctggg ccatggagca 28740

agattctgtc tcaaaaaaca aaacaaaaca aaacaagaat agaaggaaag tgagaggaag 28800

caatactgtg tttgtagagg aaaaggaagg tcaggtcagc gccaacagtg cctttgctgg 28860

gatgctccac actcctgacc agctcagtgg aattctgaga ccttggggag caaaatcaga 28920

gggggacaag gagaaagaca gaaagagaaa gagggagaga gaaaattgtg ttggccgcgg 28980

ctcgggattt atttattgta cttgctgttg actgagctcg ctccacagca ggcgaggggc 29040

ctgtaaacac acaggctgat tcattagttt ctgaccatct gcttcccggg ccggggcggg 29100

ggccggagtg ccccataggt tggagcagcc agcttagagc ccccatctcc agccagagcc 29160

agctgtgtca gtcttcccag tgacagacag gccgggtgca actgggggct ttgagggata 29220

ccctgtcccc cgccacaatg ccccccaccc atagaaccca tcaggggcct gggcagggca 29280

gggctaagta aggcaggaag atgaagagag accagggaaa aggtgaagtg tgggtggctc 29340

agcagttggg gatgacggga atactatcaa cactcaatag gagatatgtc aatatcactg 29400

tcattttatt aatagcttat taatatgatg ataataataa taagtataat tgagtgagtg 29460

gttactgagg tccaggccca atatgcaggc ctcactacac tgtccccaat aagcacatta 29520

acaactctat gatggatgat gagagaactg ggttcaaaga ttgattcaag tcaggagggc 29580

ctcctggagg aggcacatgt gggagggctc tgaaaaccag gaggacagac agtggccaaa 29640

ggagcacagg ctctgaggtg ggagggacg aagtggcctg ggaaatggca cagcttcagt 29700

gctgggttga taagagcttt ctgggtaggt gacgcagtgt gacgccactt ggtgccagcg 29760

tactgagaaa ccaggctggt ggaggagacc ccagagacat cctctgggga aataggatga 29820

aaaactccca attcgtctct acccctctga gccattcctc ccggcctgtt ccctctttgc 29880

tcacctggca tgaggcttgt cttagtcagc tccggccacc ataacaaagt agcacacgag 29940

acccaggtgg cttaaacaac aaaaatgtct ttccttgttc tggaggctga aagtccaaga 30000

tgaagttgtt ggcaggtttg gtttctggag aggcctctct ccctggcttg gagccagccg 30060

-continued

ccttctcact gtgccttcaa acggcctttt ctctgtgcac atgggcccca gcgtctcatc 30120

cggcagccaa acttcttctt cttccaaggc caccagtcag attggactag ggccaacctg 30180

agggtcccat ttaacctcaa tcacctcttg aaagagaatc tgtctccaaa tacagtccca 30240

ttctcagata ctagggttag ggcttcaaca cagaaatctg ggggacacca ttcagcccct 30300

aggaagtccc caggggcagc agaagcagca ctgggacttc aggaagggag tgatggctgg 30360

gctgggggaa ggggggcatg tgcctgggtg ccagagccca ggtgcaagtt cttggagcat 30420

atgcccatca ctcctcctcc cgagcggcac agacagagcc cattctgcat ccaggctcac 30480

agcacccttc ccggctgctg gctccctcac gggtcaagcc tgcctcctcc agcccagcac 30540

ctgctaccca ctcaggccat tgggccatcc ccacagcctc tcagtccctc accattccag 30600

gcaacagcac aggacagggt gtgtcttgcc cagagtcacc cggccagtga gtgaggaagc 30660

tgggaacccc cccgccccat gtccctctgc tgtcctggca tggtgggctt gccagtggtg 30720

tggatatctc tgatgataac agcagactcc tcctcagagg tgccatctgc tctcgagaag 30780

gctggggcca gtcctaccct ggcccgggcc tcagtccccc catctgctta aggagatatt 30840

ggactctatg gtgagctcct gccaccctga ggttgcatgg gggcctgagg acagagctat 30900

tttggtcaaa gagggatttg tggaacctgc ctcaatgaac cgagggtcag ccctgccacc 30960

gaggagcccc aggcagccag gggacctgtc cttcacccta gaggatagaa gcccaaaggc 31020

tgcggctgct gcagtggcac ctggtggggg ccgcggggct gtggctcagc ccctcagaag 31080

gcggtgggcc ccatttcccc catgggggcc aaacagctca tttgagagtg agaggtttta 31140

acttagatcc aagccatttt gtgtctgagc aaaccagaca cctgaaacat gcatcagaaa 31200

gggccacaag tggtccggag cactgggtgg ttgttagtga cagtgttcgt ttttttcttt 31260

ctttctcttt ctttctttct ttcattcatt aatagacttt atattttaaa gcagttttag 31320

atttacagaa aattgagcag atagtacaga aagtacctaa aaccccatgc caccctcccc 31380

cacccaccat ctttccctgc ctattatcca catcttgcat tagtgggggа tatttgaggt 31440

tttgggggat ttctcttttt tttgttttcg tttttgtttt tgtttttga gacaaagtct 31500

cgctctgtcg cccaggctgg agcgcagtgg cgtgatctcg gcttactgca agctcctcct 31560

cccgggttca tgccattctc ctgcctcagc ctcccgagta gctgggacta caggcgccca 31620

ccaccacgcc tggctatttt tttttttttt ttttttttg tatttttagt agagacgggg 31680

tttcactgtg ttagccagga tggtgttgat ctcctgacct cgtgatccac ccgcctcggc 31740

ctcccaaagt gctgggatta caggcgtgag ccaccgcgcc tggcctttgt tttggggaat 31800

ttctttagag acagggtctc cctctgtgat ccaggctaca gtgcagtggt gctgcatagc 31860

tcactgcagc ctcaaactcc tgggctcaag caatctccag ccccagcctc ccaagtagct 31920

aggaccacag gtgtgcacca ctatgcccag ctaattttta aatttttttt tgtagataca 31980

gggtcttgct gtgttgccta ggctgatctt gaactcctgg cctcaagtga tcctcccact 32040

tctgcccccc gagtagctgg gactagaggt gcatgccacc acaccacacc cagtttattt 32100

tttattgttt gtagagacag ggtctctctg tgttgcctgg gccgatctca aactcctggg 32160

ctcaagtgac cctcctgcct cggcctcccg gagtgttgga attccaggtg tgagccacca 32220

tattagtcat gtctgatgaa ccattgttga tacattatta ttaactaaag tccataattt 32280

acatgagagt tcacttactg tgttgtacag gtctatgggt ttttttgttt gttttgtttt 32340

-continued

gtttttgttt gtttttgttt ttgtttgata cggagtttca ctcttgttgc ccaggctgga 32400

gtgcaatggc acgatctcgg ctcactgcaa cctacgcctc cagggttcaa gtgattctct 32460

tgcctcagcc tccccagtaa ctgggattac aggcatgcac caccacgccc gggtaatttt 32520

tttgtatttt tagtagagac ggggtttctc catgttggtc aggcttgtct tgaactcctg 32580

acctcaggtg atctgcccgc ctcagcctcc caaagtgctg gattacaggt gtgagccacc 32640

gcgcccagcc tgggtctatg ggttttgact cattcataat ggcatgcatc caccactgca 32700

gtattatatg gcatagcctc actgtcctaa aaatgccctg tgctcggcct agtcatccct 32760

ctgaccctgg ggaccactac attagtgaca ttttcagatc cctctggcct gtttccatca 32820

gtgacacttc tcacaccaaa tgtatgggtt ttccacacca acaacctagt ctccagctct 32880

ctggacacca ctgggtgtcc gatgattccg ttcaattgta acactgccca gagctggcgt 32940

cagaaccctc aggtgaaggg ctcagtcgca ccacactgcc cccacctctg acgcaattgc 33000

aagttcctgc ctcccgtact tctgtccaat tggctgtaaa ttgggggttc ccacgacccc 33060

tccctcaggt ttgatagttt gctagaacag ctcacggaac tcagaaaggc attttcccta 33120

catttacctg tttatcaagt tcaggaacag ccagatagaa gagactcaca gagtaaggga 33180

tcggggagag acccagagct tccaggccct ctttgagcac ctcaatgtgt tcaccaatcc 33240

taagggtaca ggtaaaaagg ctcaggggcc aggcgcagtg gctcacgtct gcaatcccag 33300

cactttggga gggcaagcca ggtggatcaa cggaggtcag agttcgagac cagcctggcc 33360

aacatggcga aaccctgtct ctactaaaaa cacaaaaatt agccgggcgt ggtggtgcat 33420

gcctgtaatc ccagctactt gggaggctga ggcaggagaa tcgctggaac ctgggaggtg 33480

gaggttgctg tgagccgaga tcgccccact gcactccatc ctgggaaaca tagcaagact 33540

ccatctcaat aaaaaaataa aattaaaatt aaaaaataaa aggctcagga aggtgactca 33600

gctaaggaga tatttagaga gtcgagaatg tgaatgggaa tttttgaggc ggacagagtg 33660

gaagggcaca tgaaggagac agaatgccat ctgcaaaggc tcagaggtac aaccaggcag 33720

gggcaggagg acagctgtcc cggggaggcc caggcacagt gtcaccaggt gtgggagggc 33780

aggttggggt tggtatatta cagagtcagg actctgtatt ccaggcaacg gggagccatg 33840

gagggcttca aagcagggga ctaacaccca gccttcttgg ttgtctgtat ttagagcaag 33900

gacaagtatg catttttttct ttggaagcca tgggattgca cacctctgca cgcattgagg 33960

aaatgggcac aaatctccca gctgtgtctg ggcaccggag cacaggccct gtagccacag 34020

agcaggtttc tcggttcgca ggcaggggct aggccgcgca gaggcttccc aggccatccc 34080

aagcatgcag cacccaccct tcccttctcc tctccccgct gcaggagaat ccctacctat 34140

gcagcaacga gtgtgacgcc tccaacccgg acctggccca cccgcccagg ctcatgttcg 34200

acaaggagga ggagggcctg gccacctact ggcagagcat cacctggagc cgctacccca 34260

gcccgctgga agccaacatc acccttccgt ggaacaagac cgtggagctg accgacgacg 34320

tggtgatgac cttcgagtac ggccggccca cggtcatggt cctggagaag tccctggaca 34380

acgggcgcac ctggcagccc taccagttct acgccgagga ctgcatggag gccttcggta 34440

tgtccgcccg ccgggcccgc gacatgtcat cctccagcgc gcaccgcgtg ctctgcaccg 34500

aggagtactc gcgctgggca ggctccaaga aggagaagca cgtgcgcttc gaggtgcggg 34560

accgcttcgc catctttgcc ggccccgacc tgcgcaacat ggacaacctc tacacgcggc 34620

-continued

tggagagcgc caagggcctc aaggagtttt tcaccctcac cgacctgcgc atgcggctgc 34680

tgcgcccggc gctgggcggc acctatgtgc agcgggagaa cctctacaag tacttctacg 34740

ccatctccaa catcgaggtc atcggcaggt aaggccgggg gaagccctgg atgtcacctg 34800

caacctggga tgctatctgt tacctgggac gttattggat acctgggatg ttacctggta 34860

tgtggaacat gaccaggttc ttgggatgtt acctggttcc ctgggcgtta cctggtatgt 34920

gatacatcgt tacctggttc cccgggggtt acctggtatg tgatacatcg ttacctggtt 34980

ccctgggggt tacctggtat gtgacacatc ccctggttct tgggatgtta cctggtacct 35040

gggacatgac ctggttcttg ggatgttatc tggtacctgg gacatgacct ggttcttgga 35100

atgttacctg gtacctggga cactatctac ttcttgagat gttacttggt ttccaggagg 35160

ttacctagca cctgggacat tacctggttc tcaggacatt acctgatatt acatcatagg 35220

gactgcctag gatggctggg gctgcacaga cccaatgcct gagcttttcc acggaagaag 35280

gtgcacatgt cggaggaagg aagtgcatgt gttggaggtg tggagggcag gagtggagcc 35340

tgtctgggca gagctgggct aactcctggt gctgactgtg agaagtgctc tgaggctgtc 35400

cactgactca gtgaatgggc tctcatgttc aggggtttcc gagaccgccc tcaggctcca 35460

tgattcacta gaaaatctca cagaactcag caaagctgct atattcacag ttatggttta 35520

ttacaatgaa ggatacagat taaaatcagc aacaggaaaa ggtgcacagg gctgggtcca 35580

ggagatacca ggtgcaagct tctcattgtc ctctcccagt ggggttgtgc aggcagcact 35640

tgattctccc agcagtgaca tgggacaatg cttgtgaact atcaccaggc agggatgctt 35700

actgaagcct cagtgttcag agtttttact gggagatgat ggtataggca ttgctgatga 35760

tctggctggc ctttggctcc aatccctcca gaggtcgagc tgatactaca tggcccaagg 35820

cataggctct ctgatgtggc ccaagacccg caggcaaaca tggacactct tcctagacag 35880

gacactccaa cggcttatgg attagaggtc acctagcaga aaccaggccg gagtcagatc 35940

tttctttgga atgtgcaggg tgtggacaac ccaggcctgc tgtggcagct ctttcttgcc 36000

cagctctgga gaggaatgtg ttttgaggaa actttgagga aggcaaaaga ctgtcatagg 36060

agatgaaagc agcctttacc tcccaaaaac tgtttagact gtggatgatt aaatggttct 36120

gtgttgtcct gggtcatttg tacttttggc agataatatg ccagtctgtg ggggcaatgc 36180

catctaccct gggccatctg tagaatgggg atggtaacaa gcataagtct cttgctggat 36240

tgtggtgagg gttaaatgag ttaaagttga tgaagcacac cgaacagtgc ctggcacata 36300

atacgcacta aaaaaaagtg tttaataaaa aaattggaca aaataatgga gtgagagggc 36360

tggggcgggc tgagggagga gagagttgag acccaagggt gctggggcag tccgggccag 36420

gagcctggct taggcaggtg caggcagggg ctgcacctgg ggagcccagc tgggatgatc 36480

actgtgggca tctctccact gcaggctttc aacgtccacc caaatctgta attcctgaca 36540

agtggaaaaa gtgatttgcc gtatggcagg tagagcataa tcctgttttt gtaaaatata 36600

taatatgcct tgtatatcat ttctgcgtct tgaacagaca ttggctgtaa gtgtgtacag 36660

gaccgggaaa aacatcatgg gttggggaga aaggctggaa gcccgtttgc cagccatgtg 36720

ctgtggtgat gcctgtgccg tgaggctggg caggccgctg tggtgctgct gtttccagct 36780

tgtatttgca ttttgtttca ctttgacgct gtgtgtggat tccaggtggg caaaggcaaa 36840

ggcaggtcgt tcatgaagct ggagtcagct tcaagtctgg cagatgggag tgaagcaggt 36900

-continued

gcagacaggg gcccgtggtt ctgacacccg tgatctctaa cccacctcca attgggacag 36960

gatattcaaa atcaggaatg ttctggaaac caccccatat ataggcactg ttgtggaact 37020

ggtgtccaga ctttcagccc ctgggcatct ccgtgacctc tgccctcggg atgcaattgt 37080

ctggttgtta ctagacacct gccaatccct gcagtcccac tgggcttagc agctcagggc 37140

tgggaaccag ctccccaggc cctgcaggct cctggcttgt cttctgacct tggactgctc 37200

agggaaaccc tcatcttcag gcaactgttc aaggaccctg ccttaccttt cataaggtgg 37260

ctcatgtggg gccctcatgt ctgggccagc tgtaggatcc agactgggac tgtgatgctc 37320

agggcctggt ggtgcccctt caaccccagc agggtgacct ggggttcaaa ccctgcctct 37380

ctgcatcctg ccttgatttc ccatctgtga caagaggcct gaggctgcct ggagcccatg 37440

tctccctgga ggtggttccg cttggcaccc agcctctgcg ggcggcgggg aggagccgag 37500

gtggtggcga gagtgtacaa tcaattcaga gagagagtaa acactcttca ggcggcagct 37560

tggcttgtcg ttgaactcgg ttgggttaga cggttcagcg aaggactgag gttacctccc 37620

cttttccctc ttcagttccg ctgggaaagt ccagccatgc tgagaagcca gagacacagg 37680

actggtggca aaggggggatg tgggacccga agttcctgag tgggcaccag ggacccctgg 37740

gggttcccag gctctgggat tttacaccga tcctgttctg gggattccct cagaatttcg 37800

tgactgccat gttgcccagt ggctccgctg tggcccacta acaagaacag caaacgttta 37860

tcgagcattt gctacgggct catttgctat tgagcgctgt tattaacggc ttcatgtgta 37920

ttttcttgtt taatcaatag ccgtttaatt ccccaatggt tattgaatgg ctcagatccc 37980

acaagccaaa gacacggcac tgagcaagcc agacctgatc cccaacccct ggagcgctcc 38040

ccgcgggggc agcgctcccc cagcctcatg ttttcagatg aagaagccaa ggactctgag 38100

gtcagagaat agaaggagat tgcacagcgc cactacatgg cagacagcta acatctagac 38160

acaagcatgt ccaggctcaa gacagactca gccaccgagc tgtccctagc cctggcccgg 38220

tggctgcctc atttctggtt ccagctctga cttgagccca cctggagctc aggtagcctg 38280

tatgatggag cgatcccacg ggactcccct ttgaattgaa ggttctttga tgggtggtgg 38340

ctccctgctc tgccactcaa gtcagtgatt tccagctcca agggggaccc caatggcagg 38400

cagcaatgta ggcaggaggg ggaggacaga ggcacctggc tgcagaggag gagctggagc 38460

agctgcagca gctgcagcca gcccagggcc ctgggaagga gggactggta caaatggtga 38520

gggggacaag aggtggggac agactgagct gcagccccac agggggcagg tgggctgccg 38580

gtggctctga gagaagcaca ggcacccgca agtcccagcc acgccccctc cctgcctccc 38640

gcggagcctg ctccacgtgc ccattcctgg aaggcaggca gggggtggga gctattttct 38700

ccttccacac atggggaaac tcaggtgcgg agcctggaga gtggctggct tcagactgtg 38760

cagcaaatct ggggcagaat ctggatttga ggaggaatga gatcatccct ttatggtgct 38820

tggagcggga ggggcttccg gcccccaccc tccacacaac cattaagaga agcggggccc 38880

aaacctgacc cagcatctca caggagaacc atttaagaca gggttcccca gccccaccac 38940

tgagagcttt gggaccggga ggcttggggt gaggtccaga aggctgtgtc tttgcaaagc 39000

tcctcaagtg agctggaaag gtggaagctg ctgcatccat ggagatggat gactgtaacc 39060

tgttggaacc tcagttttct catctgcaaa atggagaaat agtaatgact tcagctgcct 39120

gttgtgagga gtcgatgaaa tgctatattc aaagagctta tcccagtgtc tcattcctct 39180

-continued

ggttagacgg tgggagaggg aagtggggga ggcttggggt cagggctgcc tagatatgag 39240

gccattcctg cacccacccc aggtactaga aggagtgtgg aagcatgcct ctgggttcac 39300

aggggttgca atcccaaatg tctgcagagg ctggaaacac aacttacgtg gaagccaggc 39360

atggtggtgt gcgcctgtag tcctagctac ttgggaggct gaggcagaag gattatttga 39420

gcctgggagg ttgaggctgc agtgagttat gatggtacca cagcactcca gcctgggtga 39480

cagagagaga tcctgtctct aaaaaaataa taataataaa ttcaaatgaa taataaaaat 39540

aatgagaaga agataatgta catgggtacc tatggtacca gcggaagcga tagggaaggg 39600

tgtggcctgg ggagaaatgg agacagcagg cgcctcacaa gacagtggac caaacattcc 39660

agatcttctg cggttacaaa ggaaggtgca cagaagtatc tatgtaaagt ctcttgcttt 39720

caaaaatatt ggcacctact tcaaaaagtt gtttaaactc tgtgcaagga aaacaaaata 39780

cccctgcagc ccagggcagc tcacgagctg ccagggagca cccctgctct gaaatgctca 39840

gtgtcctcct gcagccgtgg tagcagctgc attcgatggg cccaaggcag ggccaacaca 39900

ctggagccag ggcaggggca agatagcaaa acagcccagc cagctctgcc cactccagac 39960

ctgggagtca gagggaaggg ggcatcatcc cacgcaagtt cctctgccca gaaaacccag 40020

gaagctccag gaaggcagcc cagcgggtcc tttgccagca ggaacaggaa aggaggaatg 40080

agttccccat tcctggaggt attcaagccc agtgggtggt gtttgtggga gatgatgaag 40140

ggaggtcctt tgcaccaaat taaggtcttg tgtgcccatt ctgttcaaaa ggaataaatg 40200

tggattaagg gaacaggaaa ggcaggtacc atttattggg aacccatggg ttacgtgcca 40260

ggcattgcat cctcaccccc tttctgccca ggagcaggct gaagtccttg agcgagagga 40320

aacccgtgag ttccaagagg ctgagctggg cctggacggg cagatctgag ctccagactg 40380

gccaagcccg gctcctgcag acccagcttc cccagaccct gtgggggtct agctgctcta 40440

tgtccctggg cctcccccaa cctgtttttt cctcaacacc ctcccctctt tcttgccttc 40500

ttgttttctt ccataaacgt cctgtcccta ttgatcgccc cgtaactgga gatggacgga 40560

caccggacta gaatagaaac agcagaggga agaaaaccat ggagggtggg caggtgggcg 40620

cggggagggg agtgcagaga caggcgattt cagccaccct attcctgtac ccaccctccc 40680

tgcatctctt tctcagctgc tctacaggaa aaacaggggc aagaagggaa ttctgtgatg 40740

ttattctgcc tgcactgact ccttgaaact ctgggaaggc agagacaact ccctcagcaa 40800

tattaactgt cattcactga gaattcccat agcccaggta gcatcccgag tcctttatgt 40860

gcatttcctc attaagtcct atcctgtgag gatcattgtc cccatgacac agatgaggaa 40920

actgaggcca ggaggttgag atatagatgg aagggctggg tgtggtggct cgtgcctgta 40980

atcctagcac tttgggaggc tgacgcagga ggatctcttg agtccaggag ttcaagacca 41040

gcctgggcaa tatagtgaga cccccatctc aatgaaaaca aaatagatgg aggtacccat 41100

gaggatggct attattggaa aagcaaaaaa taacaagtgt tggcaaggat gtggaaaaat 41160

cggaaccttc atgcattgct ggtgggaaca aaaacggtgc agccacaggg gagaacagta 41220

tggtggctcc tgaacaagtt acatgtagat ttagcaattc cgctcctagg tatagaccca 41280

acagaactga aagaagggac ttgaacagat acttgtacac ccatgtttat agcagcttta 41340

ttcagagtag tcaaaaggta gaaacaacac aactgttcat caacaaatga agacataaaa 41400

tgtgacccat ccatataatg gaatatgatt cagccttaaa aaggagggag aacctgacac 41460

-continued

acgctgcgac acagatgaac cctgaggaca tgatgctaag tgaaataagc cagacacaaa 41520

aggacaagta ccatatgatt ctgcttaccc gacgtcccca gagtcgtcaa attcatagag 41580

acagaaagta gaatggtggc tgccaggggc tgggggaagg cggaatgggg gttagtattt 41640

aatggagaca gcttcagttt gcgatgattt aaagttctgg agagggatgg taacggtgct 41700

tgcacagcac tgtggatgtg ctaatgccac cagactgtac acttaacaat ggtaaggtgg 41760

ctgggcgcgg tggctcatgc ctgtaatctc agcactttgg gaggccgagg cgggcggatc 41820

acttgaggtc aggagctcaa gaccagcctg gccaacatgg caaaacccca tctctactaa 41880

aaatacaaaa attacctagg tgtgggagcg ggtgcctgta atcccaccta cttgggaagc 41940

cgagacagga gaatcacttg aaccagggaa gtggaggttg cagtgaactg agatcgcacc 42000

actgcactcc agcctgggca acagagagag actccatctc aaaagaaaaa aacatggtaa 42060

ggtgataaat atatacgtat attttagcac actaaaaaat gaaggaaata gatgggtggg 42120

tggtctccat gtcagactgt ccaggtgcaa attctggatt aggtgtgacc ctgggcaagt 42180

tgtctaacct gtctttgcct ccttcccacc tccacaaaat ggagacaatc accctggagt 42240

ttaagtgaga ccatccagag agaggagcca caactgtcag ggttacagag aaagaacagc 42300

agatcacaga gggcctggag gttcttctac tggggcctat gggctcttag gagcgcaaag 42360

ggtgggtttc aatcatccat ggactccctg aaatatatgc taagtgttgt gtgtgtgcag 42420

tcagttggtg agagccccta gtttttgttc cattctcaaa gggggtccat gatcccaaaa 42480

tatttggaga gtccctggaa aggcagcttt gtgttctgct tttcattttt taaacccaac 42540

tcttggccgg gcgtggtggc tcacgcctgt aatcccaaca ctttgggagg ccgaggcagg 42600

tccatcacct gaggtcagga gttcaagacc agcctgacca acatggtaaa accctgtctc 42660

tactaaagat acaaaattta tccaggcgtg gtggcgtgca cctgtaatcc cagctactcg 42720

ggaggctgag gcaggagaat cacttgaacc cagaaggcgg aggttgcagt gagccgagat 42780

cgtgccgttg cactccagcc tgggtgacag agcaagactc cgtctcaaaa aaataaataa 42840

ataaataaaa taaacccaac tctttgggga cacccagctc tttggggaca gacattttgt 42900

taagtacagt tcacacacct tcaactgcat ctcccaggcc ccctgagcta ctgcttccca 42960

ccaaaaagcg ggcatgcacg attccagacc acatcagcct ccactgagaa gtgccggtgt 43020

tgggaagggg ccagactatg ctgaagagtt ggggggtggt ggggtccact gccaaggcag 43080

gtggttgacc cctggacctg cggtctcctc tggtctttgg ctctgctcca ctgctgggtg 43140

gggcggagag ggcaggaggg gcatctcggc cttggccagg gtcacaggag gcttggggag 43200

ccctgcaggt gcttgggaga gagtcacgca cagcaacgcc ttcccagcag cactaggcag 43260

aatcggggtt ccctcttctc ttgcctaaga aggcatttcc tctgcgagct ttctcgccat 43320

catattaatt cactcaacag atgtttttc tcagaccttg cagggatgaa ttaacacata 43380

tttctaatgg gcactgcctt ggtttcttgg agcctccata acaaagtacc actaactggg 43440

tggaggtcag gagtctgaaa tcagggtgca gctggcttcg ctgtggaacc tgcagagggg 43500

aatctgtccc ataccctccg ccctccgctc gctcctggtg gggccggcag cctctggcgt 43560

tcctgggctt gcagctgcgt cactccagtc tctgccttgt cttgcggctg tccgctccct 43620

gagcatctgt cttacacgct gtctccccgt gtctcttcct acagggacac caatcacatc 43680

ggataagggc ccgccctcct gcagcatgac ctcagcttaa cttacatcct catccacatc 43740

-continued

cacaaaaccc catttccatg taaggtcaca ttcataggca ctgggggtca ggatgtcaac 43800

atatcttttc gggagacacg atttggccca cagcaggcac ctatgacatg ctggacacaa 43860

gcagggtgcc agggtccaag agcatggagg ggctgacgct cccgtagtgg gagagcagaa 43920

atcaacaaac ccagtacccc caagtgccca caggtgcacg gtgagaaaga aaccagggca 43980

aggccaggca cggtggctca cgcctgtaac cccagcactt tgggaggctg aggctggcgg 44040

attgcttgag cgcaggagtt caagaccagc ttgggccaca tggtgaaacc ccatcacgac 44100

aaaaaatata aaaattagcc aggcatggtg gtgtgcgcct gtggtctcag atacctggga 44160

ggctgaggtg ggaggattgc ttgagcccag gaagtcaagc ctgcagtgag ttatgactgt 44220

gctactgcac tccagcctgg gcaacagagc gagaccctat ctcaaaaccc agggcaatat 44280

gggagtggag ctgggaacca acctgagacc tggacccagg ggtcagggaa ggccctccag 44340

ggaggggggc tttgaaggag aaacctgaag ggggaacggg ggcacggcat ttgtggcagg 44400

gggacagcag gcaggacagg tgcagtgcct tctaaggact caatgtgatg tggctgtggt 44460

gggtgagtaa gattccaaag aagagcaggc agggccagat cacgccagcc ctgcagcccc 44520

gggaggaagt tggttttggt tgaagtttgg taggagggct acaggcaggg gagaggtgtg 44580

atctgatttc cggtgtttgg aaatctgagt tgggggtggc ggtgctggag gcctgcagga 44640

gaccgggcgg gccctaccaa gactgacagc agcctggccg ggttctggca gggacagggg 44700

ctgtgatgtg ggccgcctgg aggtggtctt tggtggcagt gttgctgagt agctgatggg 44760

agcaggcggt gggaggcatc atggaggact cccaggcatc tggctttgag acactgggtg 44820

gatggaaatt ttgccttgtc aaatggggac agaggtgagg ggattttctg gaggggaatt 44880

ttttttaat ttcttaattt tttatttta tattttcata cagcaaagtg gacctttgg 44940

gggtacaatt ctaggaattt tagcacgtgg atagagtcat gcaaccacca ccacagcctg 45000

ggcccacaac agcccgtctc tccaagaact cccgtcccga gacccctgga agcccccatc 45060

ctgctccctg aaaatctgct gtaagtggag tcgtgcggca ggtggctttt tcaaggggtc 45120

tccttcgttc tcactgtgcc tctgaggctt atggcgtggc cgcctgcgtg gctcgggctc 45180

ctcactcctg agtaggactg tgccgtgtgc actgcacacc gtggtctgtt tacccattct 45240

ccgtggaaga acgtttcggt ggctcccagt ctggggcgat tatgagtaga gctgctataa 45300

acattcgtgg aaaggttttt gggtgaacat aagttttcgc ttctcttggg taaacaccca 45360

ggtgtggacg gctgggtcgt ctggtgagcg tgtttaactc tatgagaaac tgccaggccg 45420

tcttccagcg cggctgtgca ctggaggaag cctgacttct attttggacc tgttggttgt 45480

gagcgttttg gaggcatccg gagcatttgg gcctctgagt gtgaagcgca ggcgagactg 45540

ggcgcgggga gtcctgagct ccgcggtgac gatgacgccc gccgtggggg cgtgcgagac 45600

ttccgagagg agagccgggg agccgggcgg gctgcagccc tgagatcaga gacaggggga 45660

atctggcagg gcggtgcccc aaagccacgc gggagcggtt cctgcagggc agcggccacc 45720

tgtgtccatg ctgcggcggg tccaagtgga ggaggccagc gtgggcccat ggacctggga 45780

agatggaggt cacagtagac gcacccagag cagagcccgg gagtccggga tgaaggccgt 45840

cgtggggtgg tttcaggggt gaatgccaac acttttaaaa gagtgaaata ggggtaaaaa 45900

agaggggtg taggcaagct tttgaaaacg ttttgctcgg aaggggagca gagagatggg 45960

gtgtgagctg ctgggggaat ggggagggga gagggctctt cctgaaagcg ggaggcgccg 46020

-continued

cagcgttggc gttgggatgg gaagaatcca ccaggactga agatagctgt ggggagggcg 46080

ggagggacgc tggggacagg agccgaggaa gtgatcgggg tgggcggaga gctgacagga 46140

cagagggccc accagcatgt gtaaggagga aaggctgact gatttcatat ttaggtcaga 46200

aaccctgctt ctaacacctt ctagagactc catgtccatc ttgacaaccc agggtggccc 46260

accctggtgg ctcacagaag tccctttgtc ctgagtaaag agtggcacat cacggtgcca 46320

gcgagcagcc tgcctgagat tccagctctt tccatttgct cttttctggc aagtctcttc 46380

catttctcca ggcctcagtt tccccacctg taaaatggga tcccgctctt gtaaagattg 46440

tctgagactt gcccgctcaa gtcccaatac tcaggaggtg cctgaggaag gggagccatt 46500

gatcccttga aggagtcggt attttgtcac ctgaacacac agaccatgtt cctccgtcac 46560

tcgcctgggg ctgagtggaa ttccttggct ttatctccag ctctgtcctc tgtcccttca 46620

tcagtcaaca gtcatgcagg aatgagaagc tcagttcata gtaggtatag ccaaaggtac 46680

tttggaggag agtgctgctg ccaacagcgg acgggggact gagccccacc cagttctgcc 46740

tgggccactc cctctgggtc tgtggccacg acacccggca ggctcctgtc cccccgaccc 46800

cagctctgtg gcttgctggg cctcaccagc tgcgattggg caggacagat ggggttggag 46860

aggggaagtt gagtggggag aggacaggcc acatccaggc tcatcctcag gccttgttct 46920

gcccctgggc tgggctgcct ccctccacat ctcagcagag ggagcgtggg gtgggtcatt 46980

gtagaggatc cctggccctt catcccatgt gaagcctgtt ttagactcac agatgctccc 47040

ttagatgtac ccgccaagtc acacacacac gcagccattc actcatctgt aacatgcccc 47100

tgactcacac ataaacacac acctgcccca cccacacatg tgaatccgct tccagaagca 47160

cccagtcata gacacatcac actccacatg gctgccttgc acccctgctc acagacgcgg 47220

acccacaggg tgagcggagg cctccgcagc agcccggggt tcaccagagg ccttggcttg 47280

actgccctgg gctcagggct tctctgtgct tctcccccat gttgaggcac ccacttcatg 47340

cttgggttcc tggaggcctc tttccactct gtgggacaaa gccccctctg ggcttctaac 47400

catccctcca tcccactctc agcctgaact aggtcccaag gtggaccccc tatcagcagc 47460

cttgccccct agaagagagg tgcaggctgg cgtcaggatt tggtttcggg acatgactgg 47520

cctccctccc tcccttcagt cttgcccttt ctgtagtttt ttgtccctaa acacaaaccc 47580

agtcacttct ctgcacaata tctctagtgg tcttctactg accttagact caagtctgtg 47640

ctctgtgacc acagactttc cctctcacct ccttctgtag cctgggctct gagactccca 47700

agatcagcca ccccaaaggc tgggtctctc gggctggcct aggggctagt ccttctgcct 47760

ggaacatgac ctagccctcc ccagccttcc tgtctggcaa actcctcacc cctcaggact 47820

tggctcagag agcacctctc tggagctcct gaaggctgga gacctgtgca accagcccct 47880

cggtcccctt ctccagcctt ggcggcttgc gacaattcca cagccttgcc atagctacca 47940

tagctatgca cgtacacgtc tgacacccgc ccccataggc tctgagctcc aagggagcag 48000

ggatggggtc tggcttgttc tcaccttagc accagcaccc agaaccaggt aggcgtgctg 48060

tatatgtgtg gaattcattc attcattcag tcagtcagtc attcatttat tcattcatga 48120

acaaatgaac ttacatgaac aaagaagtct tactatctgg aactctttcc agagaagaaa 48180

gagactggaa tataggcgct aggaggcaga gaggcaggac cagctgggct tgctagcacc 48240

caccttgtc gcctcccacc cttccaccat cacaccctgc tccttctcct gttttccttg 48300

-continued

tgtgagaaat gggcctctct gggctaccac ctgggcctgg tccttccctc gggggtgttg 48360

ggccttcctg caggggaaga agcaaccttg ctgaagtcag ctaggctgct taggctgggg 48420

gtgccttcgc aggccctggt catacccacc gtgtcacccc taagccgcac acctcaagcc 48480

tcccagcacc cgcccgtcag cctctgtgtt gtgtgcaagg aagctgcctc tggctttgta 48540

atgggtaata ggatttatca atagactgag gaggtgaggt atgttaaagc acttaataga 48600

aaagggcttc gcacagaagc ccaaatttga tttagccaat gaactcaatt gccggcctga 48660

ttgcattcca ggaggcgcag cagccagcat ttgtccatgt taccctggaa aagccaggct 48720

gcgccaggcg accaacccag accacccaga cctcccctct gcccccacgg ctctgttgtt 48780

ggttcctcgc tttctcagga tctcaggttt gaatggcagt cctttgaccc aaacagtccc 48840

aagttctcca gcatccaaca gcctccttcc ctctagtgcc caaggcttcc ccatcccatg 48900

aattagctag aagtgcagtt tgctaccatg ttcctttcac catgatcctc aaatccgtgt 48960

gtcccctggt cacttgtcac ctcttactca agctggcctc tgcaggcttc cctcctgacc 49020

atttgtttct tctctggctg ctgcccatct ccccacactg cttgcctgag ccaaggacac 49080

atctctccca acccccaaa gaagaccaag atccccaggc cacagtgttt catgacgtct 49140

tgactcaggc ctgggcatag cagatgggaa ttaccatgta gaaaagaagg acttgaggcc 49200

gggcacaatg gctcatgcct gtaacctcaa caatttggga aggcagatca cctgaggtcg 49260

ggagctcgag accagcctgg ccaacatagt aaaaccctcg tctctactaa aaatacaaaa 49320

ataaattaaa aattagccaa gcatagtggt gcatgcctat aatcctggaa actcaagagg 49380

ttgaggcagg agaatcactt gaacttggga ggcggaggtt gcagtgagct gagatcgtgc 49440

cactgcactc cagcctccag cctgggcgat gacagagtaa gactccatct caaaaaaaaa 49500

aaaaaaagaa ggacttgaga ggatttccat tagccatcac caaccttgag gccgtccatc 49560

gtccatccat tggccacggc cacccaccgg ctggggagga gaaacctctt ggccatgtaa 49620

gcccgcagcc ccctcccagg gcagcatcat tcctcactgc ggtccaggtc cccaacccca 49680

gctctgtgag tcctcagaaa gcgagggccc ttgttgcttc agcctcctcc ttcatcctgc 49740

agaacacgcc agcctccctt gtgagcatct ccaagccacc gtaagatctg ggatttgctt 49800

taggttattt tgcccataac ccagaggctc atgaaaaatg tttttcccaa aacaaaagaa 49860

cactcttcac ccaaagataa atgcgctatc tggcaagaga aattggaaaa cattgctggc 49920

tgtgttaaac tagtagtcta actttagccc ccaaggtacc agagctctgc gaggccagtc 49980

tcagtataca cgacatgtaa taatgtgggg atgggtggta acataccagg aaagaaggac 50040

caggcaatgt gattaatgac caggaacccc catggtcctg caaagaggtg ctttgagaca 50100

agttggggac tagctctccc agcccagcac ctgcccaccc cagttcagag ccattgccgt 50160

gaacactctg aaatccccgt ggggctttgc ctttgcagag agccagccct ggggcctcct 50220

cctccctgcc cagctccagc accatccctg gctcgctcac ccaaactcac cgcttcctcc 50280

gtcattcccc gcaaggagtg gatgacatca ctctttccgg cagccaggag ttgacctaca 50340

ttccacccgc cctcctgtct gggagactct ccaaaccccc cgctcctctc tccctgggag 50400

gagggaaggg ccccgctcac gatttttgtg gagtgacggt gccaggcccg ggtccagatg 50460

gtgcccccgc agcagctccc aggcactgcc tgcccctccc tgcagcgcag ggcacgctcc 50520

tgcctggcac gggccagggc ccctgtcatc cttcatgtgg ctgtcaggcc cccagctggg 50580

-continued

cctgccacgt ttcccagaga actggtgtta tctgggctgg gctctcccct ggaggtgagg 50640

cccggtgctg cctactaaca gttgtggtct ccaggggctt actagggact catccattca 50700

agaaaaaggg aaacttagct gaaaaggtgg ctgtggctct gtccttgctg gcagcagggc 50760

cagcttccag caagcaagtg agtgcgcccc atgtcaggcc gtgagagaag tcagggtctg 50820

agcagagggg ccagacagcc acacgggtga gccgggtagc aggtgggcct gccagatgaa 50880

atacgggata cccagttaaa cctgaatttc agataaacaa gggaatcgtt gttttttagg 50940

gtaagtatgt cccaaatatt tcatggaata tacttgtact aacaaaatca ttagttattt 51000

atctaaaatt caaaaaaaaa ttttttttt ttagagacag ggtcttgctc tgtcacccag 51060

gctgaagggc agtgcagtgg cacaatcacg gctcactgca gcctcaacct cctggactca 51120

agcgatcctc ctgcctcagc ctccgaagta gctaagatga caggtgctca ccaccatgcc 51180

ctgataaatt ttgtgttttt ttaaattttt ttgtagagat tgggggggg gtctcacttt 51240

cttgcccagg ctggtctcaa actcctgttc tcaggtaatc ctcctgcctg ggcctcccag 51300

agtgcccaga ttacaagcat gagccactgc accaggccta aaattcaaat gtaactcagt 51360

gtcctgtatt tttatttggg aaatctggca accattgtgg catggggcta ctgtggggaa 51420

atgactccaa gaggccagtg ggggccaggc atggtggctt atacctgtaa tcccagcatt 51480

ttgggaggcc gaggtgggcg gatcatttga ggtcaggagt ttgagaccag cctggccaac 51540

gtggtgaaac cccatctcta ctaaaatact aaaattagcc gggtgtggtg gcgggcgcct 51600

gtaatcccag ctactcggga ggctgaggca gagaattgct tgaacccagg aggtggaggt 51660

tgaagtgagc caagatcgaa ccactgcact ccggtctaga tgacaaagcg agattccatc 51720

tcaaaataag taaataaata aaggataaat atctggcaat atgttactga atcctccaag 51780

aggccagggg gtgtgtcacc tcttggggcg ccccaggggc ctggagaata ggttgattat 51840

ctctattttg gggagctgca ggctctgaga gaggctctgg ggtgggtgtt ggacgggaca 51900

gcatgaataa ggggccctgc ccttggggtc aggcacctca gggctgccag gtaaggccga 51960

tcttggcacc tgggagccca tgtactctca tctccgtccc tccctctatc acctgttggc 52020

tgttgacaat agcaacaata ataatagctg acgtttccgg agggcatgct ctgtgccaga 52080

cccagtgctg agccctcaag ggtaccagct cagcttgccc ttccctgacc cttgacagtc 52140

ctgccacatg gaaacagcca cttcctaggg ttacaggaca ccgactgatt ccccaaggct 52200

ggtcaccttc ctgcctttgc ctagcttgag gggtcagagg tttgaacccg actgtctggc 52260

caccgcgcta cctggtgctg cagcccagcc gtgacactca ccggccttta acccagcagg 52320

tgctccatgc cggttgctgc acacagagct cccatgcatt atccccttcc tcccttgctt 52380

ctttctgcaa acattgagca cctactgtgt gccaggcagt gtgttagaca tttaccaaaa 52440

attaaaccca gagaagtgac ttgctcaagg tcacctagcc aatcagtagc aaaacctggg 52500

taagaatctg gctgctgact cagcctctct ctctctctct ctctctctct ctctctctct 52560

ctgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtcc catggagcgt tcttgctgcc 52620

atctgaagag ctaaagccgt gggattctgc agtggggcag gggtggagag ggaacagccc 52680

tggtggcctg ggaggggtcc actggcctca tgggctggcc agcatgccct atctctatgt 52740

tcattattaa atcatccttt ctgatggatg agctgaaaag tgctgtgtcg ggggatcagg 52800

tggcccaggt gatggttttc tggtgagatt gatggttcct gagggtcaaa ttcagagagc 52860

-continued

gatcgcttgg gaaaattgat cccacagaag aaggggaaat atctgggctg gagtgcagca 52920

ggagcccaga ctgcccgccc aggtccagat ccaggcagag gctggtgcca ggagggcacg 52980

tggcaaaggg ggcttctttc ttcttctgtg gctgcatttc cttggctgtg gctgggccac 53040

agaattggta agcagggcca tgggcagcag gcatgggggc atctggctta gtggcccctt 53100

tcctggcctt ctttgtcccc atggaatgga cgcgaagcca gcactgagag agacgcaggc 53160

agcaggtggc tgtgagatcc agctctagtg ccatgcagcc ccgagtgcca agcctgcttc 53220

ttccacccac agccctggcc cttgacccat ctgagcctcg cattctcacc tgtaaagtgg 53280

gaaggacagt gccttctgcg ttttcacgag gatttgttta ataatgcatg gcgtctgtcc 53340

acgtaggagc actcagcaaa ttatttctat gtgcttgctt atttatttca tttacttcat 53400

ctcccgaggg ccttccagaa tagggcagag aaatcacaga aggcaggcag cctgaggcag 53460

ggaggggaag ggtgtgcctg tgtctggggg agctgtgggg tggctgcagt tccaggagcc 53520

tgagtctatc caagagccag cacactgaga ctcttcaggc tctgcccact gcctgcccac 53580

ttcctaaagt ggggttcacc agctgtgtgg ccttgggcaa gttgcttaac ttctcttgcc 53640

tcatttcctt gtctgtgaaa tagagatagt catagtatac tctgcataag gttcctgtga 53700

agaggtgata acttcacttg ttgaaagctc tgaggatccc tggctctagg agctcctgag 53760

tataagaaat ggaaaactgg catgtcctag gcacctgttt gtctcatgca agagctttat 53820

agaccgtctc ccacttgatc cttccaacaa ctccatgagg caggtgttac tgccatcctt 53880

attttacaga ttagaaaaca ggccaagtga aattcagcaa ctttcctgag gtcacacagt 53940

cgttggcaga acatcccagg ccagtggggc ccaggggatg gggttaccat ctctcagtca 54000

cccagaaaga caaaaggaag gcacacaaac cccagtgggc accgctctgt gccagggctt 54060

ttgtttcacc ccaaccgtga tgagtgtggg tgatgtggcc ggtgtggtcc atatcgctcc 54120

tcagggccag tgtcactggt gtgcgtggca gcctgtggca tacacaagat attgtttccc 54180

gctcttcagg tgagaaaagc gagactcagg ttggctaaga atttgtgcag gccacacagc 54240

tgtgtgcacc atgagctggg cagaagccct gcccaccagc accatggggg ctcccggctc 54300

ctgccacaca ggccacccgg gagtgcagac actaagtgtg gctctgaatc gagctgcttg 54360

gattcaagtc ctgcgtctgc cccttaccag ctgtggaacc tcaggcagct cacttcacca 54420

ccccaagtct cagtttcccc atctgcacaa cgaagataat aatgacatgt cattcagggg 54480

attgctgtga agggctaatg ttctaattcg ggtggcactc tcagaacagt tcctcatgca 54540

aagtgagcac tcacgcatca gctgtctacc caggagactt cctttgaggc acaatgtgat 54600

agcgttttgt ttttttgttt ttttttgttt ttctggtttt tttgagacca agtgttgctc 54660

ttgtccccca ggctggagtg caatggcatg acctcggcct actgcaacct ccacctgttg 54720

ggttcaagcg atcctcctgc ctcagcctcc cgagtagctg ggattacagg cgcatgccac 54780

cacacccagc taatttttgt atttttagta gagacgggat ttcaccatgt tggccaggct 54840

ggtctcgaac tcctgacctc aggtgataca cccgcctcgg cttcccaaag tgctgggatt 54900

acaggcgtga gccaacgcgc ccggccccaa tgtgataggt ttataaaaat aggaaaacaa 54960

atctctgaag atttttagga gagatactgc aagaaaagcc tggaagctgt aagcccttcc 55020

ctgtttccac acatcccgat cattcattca ctcacagctt tacccagatg tgattcacac 55080

gttgtacaat ccacacatgt aaagcgtaca attcagtcgc ttttaggtta ttcacagagt 55140

-continued

```
tgtgcaacta tcaccacagc agactttaga acatttccat caccctaaaa agacactctg  55200

gccgggtgtg gtggcacacg cctgtaatcc cagctacttg ggaggctgag gcaggagaat  55260

cacttgaacc caggaggcag acattgcagt gagctgtgat cgtgacacta cactcagcct  55320

gggtgacaga gcaagactct gtctcaaaac aaaaaaaaaa aagaaaaaaa aagtcatcta  55380

cctatcctag acattttgtg taaatgaaat tatacaagat ggctgggcac ggtgcttcat  55440

gcctataatc ccagcacttt gggaggccca actgggagga tcacttgagg tcaggagttt  55500

aagattagca tggccaacat ggcaaaaccc catctgtact aaaaatacaa aaattagcta  55560

gatgtggtgg tgggtgcctg taatcccagc tacttagcag gctgaggcag gagaatcact  55620

tgaacccagg aggtggaggt tgcagtgagc caagatcata ccactgtact ccagcctggg  55680

tgacagagca agactctgtc tcaaaaaaaa aaaaaaaaaa gaaagaaaga aagaaattat  55740

acaatatgta gtctttcatc actgaccgct ttcacttggc ataatgtttt caaggttctt  55800

ccatgtcata gtgtggatca gtatttcatt actttttctg gctgtatagt attcattgcc  55860

tgcttacaca ttttgttttt ctgttgatgg atattgggtt gttcccacct tttgactatt  55920

attagcaatg ctgctatcaa cattcatgta cacattttgt gtggacatgt tttcgtttct  55980

cttggttata tacctaggta gaattactgg gcaactctaa gttcaccttt tgtagaactt  56040

ccagaccact ttccaaagca gctgcacaat tttacattcc caccagcagc gtatgagggt  56100

tctgatttct ctgcatcctc accagcaggt gattatccgt ctttttgact ctgcctatcc  56160

tggtgagtgt gaagtggtat ctcattatgg ctttggtgtg catttccctg atggctgtaa  56220

tgttgaacat gttttcatct gcttcttggc catttgtatg ccttcttcgg agaaatgtct  56280

gtgctgatcc tttcctcatt ttaaaaactg gattatttgt cattttatta ttgagttata  56340

agtgttcatt agatgttcta gatataagtt ccttttttagc tatgtcatga cttgcaaatt  56400

tttctcccat tctgtgggtt gtcttttcac tttcctgaca gtgtcctctg aagtacaaaa  56460

atatttagtt ttgatgaagt ctaatttatc tatttttttt tcttctgttg cttttgcttt  56520

tggcatcctg tctaagaaac cattgcttaa ttcaaggtca tgaagatttg ctgctatgtt  56580

tccttcttct ttttttttttt ttttttgaga cagagtctcg ctctgtgctc aggccagagt  56640

gcaatctcag ctcactgcaa cctctgcctc ctgggttcaa gtgattctca tgcctcagga  56700

cctcccaagt agttgggatt acaggcaccc accaccacac ctggctaatt tttgtgattt  56760

tagtagagac ggggtttcac catgttgacc aacttcaact cctgacctca ggtaacctgc  56820

ctgcctcagc ctccgaaagt gctcagatta caggcgtgag ccaccgcacc tggctcccta  56880

cgtttccatc tgagagtttt ctagtttgca gctctttgat ccattttgag ttaattttta  56940

tatgtgatgt gaagggtcca acttcattct ttttacgtgg gtctcccata ttatttattt  57000

atttgtttac ctaataattt ctttaaatta atatacttta aaaatttggt ttcacttagg  57060

cagtattttt gtgatcccag gtcttatgag tctgacatta aaaaaaacat actgtgtatt  57120

caaaaacaaa taatagcctg tagagctgtt ttacagtcca tgcgcccatc tcccccagtc  57180

atctcgtaca tcccactttg ggagccagtg ctgatcatga tgagttttat ggggagggag  57240

agtgtagcag gggctgattt gaggtagggg ttgtgaaagg cctccccaaa gaagtggcat  57300

ttggcctaga atctgacacc acctctaaat attagctcta ggagaagggg cagagcctgt  57360

agtacattcc aggcagaggg aacagcatgt gcaaaggccc agagaaaagc aagatcagaa  57420
```

-continued

catctgctgg acagagacat ggagagagag acaaagagac aggagaagac ggagacatgg 57480

agagaggggc actgagagag ccagagaggc caagaggtca tacacaaaca ccatgtccac 57540

acacaccgac acagaggcag taccacccac agtgacctgc aaaagtcaga gagaccagaa 57600

gcacctgcat tgaccaggtg ccatgtgtgt gccagccttg ccagcctttc cccacacagt 57660

gccagctcct ccctgcagcc ctgctagaca gggcaattct ctcctttgga agaggaggca 57720

ctgtgcttgt tctggggctc agccagcgag ttgaagaggc cgggacatct tctcctgttc 57780

ccgaggaaga caggactgat aggaggggct ggccaggatg gagcctgggc agagccagct 57840

ggggagggac gtctggattt gggtcaggag aaaccaactg tatgactttg ggcgacttct 57900

ccctaccagc accacccagc ctgtccagct ctgacatcca atggctacct attatcactg 57960

ttacaatatg cttcttggat tcggccggtc ttctatgcat ctttactaag ccaaggtcaa 58020

tgttatcacc caaagttgat ggcaatcttt tctttcggta gctgatatct aaaaatatcc 58080

cttcctcagg ccagcagtgc ggagcacagc cccggttccg tacagctgag gtgttcctct 58140

ccagccagct caggggggtg ttagatagga attcagaagt tgcgaattaa aacctcccca 58200

ggtcagattt tgcagacaaa ctcttgttag acaaagtcag acaggtttct tccctgcagg 58260

tcttgtcaga gcctttaatc tgctgagatg tattatgacc ctcagtggag tgaggtagga 58320

tcaaggagca gatgtcaggg gatttggaga gcatcatatg cttgaccaaa aaggtttaag 58380

cacagaaact ttctgtgcca aggaccatct ctgggatctg gcatctccag aaagaagggc 58440

cgcccgtacc aataggtgcc tgctgtcccc ctctgtgcct ttcccagagc taagcacata 58500

gcagtcctgc agcacgcatg tggcaatgga gggatgagaa tgccagggaa agagtgaaca 58560

ggctttttaa tgctctattt atttcagctt ggtgttccag taaaacctcg gcagggttag 58620

tgtccaggct ctcaggggaa gatggagtta aaatcacaca gacttcagcg gtggctagtg 58680

ctgctggcca gggctcttgc tctgcctcag aattctgttc cagagaattc cagagacgtc 58740

cttcgtcctg ggaactgagc atcacggcac agtggataca tgtaacctaa agcggtgggt 58800

tcctgagtgc cccctctccc acggcccacc caggacctgc aggattctct cctcctcctg 58860

ggaggaaagg cgtcagctca gaaactcact gccttctcca ctctggcctg acctcatgtt 58920

tctaaataac atttctaaaa catggacttt aggccagtcg caatggctca cacctataat 58980

cccagcactt tgggaggctg aggtgggcag attgcttgaa cttcaaccag gagtttgaga 59040

ccagcctggg caacacaggg agaccctatc tccacaaaaa ataaaaatga aattagccag 59100

gcatggtggc acgcgactgt ggtcccagct actcgggagg ctgaggaggg aggattgctt 59160

gagcccgaga ggcagaagtt gcagtgagcc gagattgtgc cacggcactc cagcctgggt 59220

gacagagtga gactcggtct caaaaagcaa aaagactcca ttgcctggat ttaccataat 59280

ttccctaacc attctgcagg ataattcgag gacagagaga agaccccaag gaggaaggca 59340

gccctggtta ccaaagctgg cagatggggg ttagctggaa gccccagggc tggtggacgc 59400

aggggccgct gtttccccac ccagatcctt gctctggaag gcggccccca ggggaccctt 59460

cattcccact cagacaggga cagaggcggg acagagccga gggaggaggg ctcagatgaa 59520

gcacctggca ggactgagat atgagggagg ccggatgcga ggagggagct ctgcagcctg 59580

tggtgtccag gaggatttgg ggagtgtgag gtgagaaaac aaaaagcgtc acccctccca 59640

gtggaagggg agcatgaaga gagaagaaaa tgccagttac acatcctcaa aacaatctct 59700

-continued

gaatggacga gagccagtca gggggaaacg gaagttgcag tgagccaaga ttgtgccacg 59760

gcactccagc ctgggtgaca gagtgagact ctgtctcaaa acacaaaaag atgccgcaca 59820

ccccaggcgg gacgggtggg cccaggtgcc ctgccctggt ccaggcctct attcgcagga 59880

tctcactgat ctctccctgc ccaggctttt gtctcccaga caaaccagtc tccatgaagc 59940

agccacagcg ttataaaata tggcccagag caggacacgc ctaaccagtc aaaggcttgc 60000

agagtaaaat cccgcctccg tctcctggtc tggcctcctc cacccccag cctcattgca 60060

gacctcacgg gtggaacttc tttcagccct agggctttcc tcatgctctt gcctctttct 60120

agaaggttct ccccttctgg cccggcaact cttgatcctc cttcaggtct tagtttaaaa 60180

ctgttcttcc tggaaactgt ggactcccgt gatgccttct tctttgcaca acccgtgggc 60240

tcttttaaca gttgaacact ggaagaggtg tgtgattcgt attgttataa ttaatagact 60300

ttatttttag agcagtttta agtttacaga aaattgagca gatagtacag aaagttccca 60360

tttccccacc cccctgcaca gtttccccta ttttggagta tatgtgttac aatgatgagc 60420

taacactgat acagtgtcat tcactagggc ctgttattta cattaggact cactctgtgt 60480

tggacacttc agtgggtttt gccatatgca taatgccaca tatctaccat tgtggcatca 60540

taaagaatgg tttcaccgcc ctaaaaatcc ctgtgtgcca cctatttatc ctccctccct 60600

gtcccctcca atcactgatt ttgttcattg tctccctgct agacccgaag caccatcagg 60660

tcagggagcc atctgctttg ttgacattat actcactgtt cccagctcag agcccctgcc 60720

acataacagg tgcctgataa atatttttct ttgcatattt taataaaaat aggattctac 60780

taaacccatt gctctgcaat ttgctttttc caaatattga caatatatct tggaaatctt 60840

cccatatcag aatatagagc acactctgtc tctctttcac tttttattat ggagaatttc 60900

caacacaaaa gaagatggtg tctgccatga tgaacctctg tgtacctctc acccgactgc 60960

agttaccaac tagtgaatct tcttgtttca tctgtacctc tgcccatcag ataatttcat 61020

ctgtaaatat ttctatgtgt atctctctct gctttttttt ttccttttga gatgtagtct 61080

cgctctgtct cccaggctgg agtgcagtgg tgcaatggct cactgcaacc tctgcctcct 61140

gggttcaagc gattctcctg cctcagcctc cccagtagct gggattacag gtgcaccatc 61200

acacttggct aattttgtat ttttagtaga gatggggttt caccatgtcg gacaggctgg 61260

tcttgaactc ctgacctcaa gtgatctgcc tgccttggcc tcccaaagtg ctaggaatac 61320

aggcataaac caccatggct ggccctttg ttttgttttg tttttgtttt tgttttgttt 61380

gtttttgttt tgttttgttt tgtttttttt ggagacagag tcttgctctg tctcccaggc 61440

tggagtgcag tggtgcaaac atggcttgct acagccttga actcctgggc tcaagggatc 61500

ctcctgccct agcctcttga gtagctagga ctgcaggcat gtgccaccac actggctaat 61560

ttttttattt tctgtagaaa caggatctta ctatgttacc cagtctggcc tcaaattcct 61620

gggctcaatt gatcctcctg cgtcagcctc ctaaagtgct tggattatag gtgtgcacca 61680

ccacgcctag cctgtgtggt tctctaaaag ataaggactt taaaaaacat aaccacaata 61740

tcatcatcac acctagtaaa cattagtatc taatataggc aatgtccaaa tttccagtta 61800

tctcataaat atacgttcca tttgtttgcg tgtttgattt tacaggatgt ttgaaccaag 61860

gaccagagga gcttcacttc tgccattaat tggtatgctt ttgaggtttc tcttgctcta 61920

ggtccatcct ttttccgtat caatttttc tccacgcaac tcatctgttg gaagaagtgg 61980

-continued

gtcattactg ccggagtttc ctcgagtcca gatcttgcgg atggcatccc ctggtgtcgt 62040

tcagggtgat gctctgtcct ctgtctttcc tgtgcgctca cagctgggtc tagaggcttg 62100

atctgattca ggttcagttg ctttggcttc gggtgggggt gtgttcttct agagacgctc 62160

actgtctgct ggcctctcag tgggaagttc tgggctggtg cggctcagtg cctggagcta 62220

tgcattcatc agggactgca aattctctcg tgaaatgctt ctataaaaag aagcttcccc 62280

tcatcaactt tcggttaccc cgaggtacag tttatatagg acaggcagga taaatgcttg 62340

attctttccc tttatttacc agttttcaga ataatgagtc gcctataaat atttgaatga 62400

atggatgaat aatacacctg tgaatgaact gacgggggtg tgggagttgg gggttctatc 62460

tgttgccctc aagggcgtgg ctgtgggcca caacctgaca gcagaggtcc agcctgaagc 62520

caggtgcctt tctacacaat gaagtgcagc ccatggccaa gtctctgtct acaccaggtg 62580

ctggcagcag cctctgtcac ccatcaaccc acatcagtca aaggctaggg tgatcagaag 62640

ctgcattact aagaatctag catctgggac aaggcagtat catcactctc cccaactgag 62700

atgtgaggaa tccttgaaac cagcatcaga gcagagagga gagccgcgca gtgactgcag 62760

gtgtggcctt tggaacacgg cgttgatctc tctgcaggaa ggggaatcaa ggagtttctg 62820

gcctaaaggt tgggctggtg gcctccaggg tttcttcctg ggcagcccaa caccctcctg 62880

ggcccctcct gggaggcgct cctttcccca gaggccaggc caggctgccc acaagctctc 62940

tgacatctct gccctctcgg tgtctcccca ggtgcaagtg caacctgcat gccaacctgt 63000

gctccatgcg cgagggcagc ctgcagtgcg agtgcgagca caacaccacc ggccccgact 63060

gcggcaagtg caagaagaat ttccgcaccc ggtcctggcg ggccggctcc tacctgccgc 63120

tgccccatgg ctctcccaac gcctgtacgt gccatgcccc ggggccacga gcccacatgg 63180

ctataatctt ccctgcccgt caatcccagg agctgtggat catacacacg cacacaaacc 63240

tgcacacagg cacatacgtg tgcacatgca tgcaaacgtg cacacagaaa catacgagca 63300

tgcatgcaca ggcatgggca cacatatgaa tgcaaaaaca cgtgcatgca cagaaacaca 63360

cgtgcgtgca tgcaccccca cacacacacc ttgttctaca gctcccaaat gccaggtctc 63420

ataacaagtc cctctagcat acctgcatcc tgctgaatgc taagctgccc ttccagccct 63480

ggtccacagg gaaacccgag aggagctgct cagcagcatt cctgcaaccc ttcgccttct 63540

tgaccctgaa ggctggcagg tggcccccat gggatggcag ggagtgatgg ggtgggaccc 63600

ccacttcagg tgaggggatc agtagcatca tccctgatca taagcatctc tggggtgttt 63660

ggagcgagcc accaaggcca agtgctgcgt gttcagcact cacctccatc ctcagagcaa 63720

cccctgggaa ctgagcttcc gttatccctt tttacagatg tgtcctctac ggcctgcacc 63780

agagtccccc cgcacgtgct ctctgcccac ccctctcctg tcctcgccca gccatgccag 63840

ggagccatcc tcagggcctg cccactcaca tctgcatcat agacccccct ctgaggacca 63900

cccacccggg cgcgctagcg tagtccacac tgtctccatg aaatcagaat gccaggaagc 63960

ccctgggctg caagtggaga aggccacaga ctgcccctgg ggtgggctct ccctgacccc 64020

caccactgcc actttacggc cgttctcaag tgtgtggttg gtgtccgagg ccctcctagg 64080

cactttgtgc atctcgcttc accctcacac agccccgcag ggtggagtcc ttattacggt 64140

cccttctcca gatgagcaga ctgaggccca gcgaacctcg gggccctccc aggtttcaca 64200

gtgagttgca tgccaacgcc agggcttagg gcagctactg gatctgatgc tcagaccagg 64260

-continued

cctgggcagt ggcctcctgc accccatctg ggcatcctcc ctcgacgggt gaaagattta 64320

tgtgacttta ggctgatcaa agctgagagc cacgtgcaga ggccatgaca ggcatgtcac 64380

gcgtgtgcta cagcaggcgg tgagacaggc ctttgaaggg gtgacttggc aggcacccga 64440

cagcactctc gggtcttcaa aggcacaaag agcactccgg gggctgacag cccactccag 64500

ccgctgcctc ctcccagctt cctgtcccca tctctttgtg gtccaaaccc tttcctgagg 64560

tctctcttta gtggccgtca ctctctccct gactcacaaa tttgctgccc tttcgtctgt 64620

cctgagggcc gctctctctt ccctcctctc tgtctcagtt ctctccacca tcacaccctc 64680

ctttttttt ttttcttaaa gataagtctc tcactctgtc gtccaggctg gagtgcagtg 64740

gcgcagtcat ggctcactgt aacctcaacc tcctgggctt gagcgatcct cccacttcag 64800

cctcccaagt agctgggact gcaggtgcac accaccaggc ccagctaatt ttcgtgtttt 64860

tttcagagaa aaggttttgc catgttgccc aggctggtct cgagctcctg agctctagta 64920

accctcctgc cttggcctcc caaagtgctg ggattacagg catgagccac cacgtctggc 64980

catgtctggc ttttttcttc cacttgcccc ctgcgcctgg agggttctgc ccatctctgg 65040

ccacctgaga ctcttacctg ccagtcacac ctaggaggat ccactgtccc caccctcacc 65100

cccagtcacc acctccagtc tagccagcga tggatagaca ggcctgtagg ggctgggcag 65160

ctgggtcttc cttacttagg ggagactcct gagctggccc cacctcctgc ttccctggga 65220

gtccccggga attgcctttg aggcccccaa gccccaggaa ggaggggttt cttctggcat 65280

ttgcagggtc acaggtgagg ctggaggcgc tgctgtcctc agcacccaag tctcctgcct 65340

tctctggggt gccagcagga aacagcccaa agagacacag aggctaattt tgccctcacc 65400

ctgcccctac cgtgacccct cctcagtaag tggcaccgac acccacgggg ccaagtgcaa 65460

gtctgggggt caccttgact ggagaccctc ctcccgccac gttctttgaa cttccctcca 65520

tccgctccaa gtctctccca atgccatcct cagccctgca gcagccctca ctcccgatgc 65580

cttcccacct cctcaccact ctgcccccac ctggccagcc catcaccctc cagggcccaa 65640

cttggagccc ccaggacctc cccgtgccct gcctgatgtc ccgcctgtcc ccacagagcc 65700

tcacttggtc accacccagt cctggccctt gcttactgtg gctgcacccc gaggtgtcct 65760

cagggtctag caggtggctg cccagacatg gaggtagagg aaggagtggg tggggatggg 65820

cttgtcctcc caggcctccc tgcctgtcct gctggccaca gccttggctt gcccaggaga 65880

agcccatggg ccacacatcc cactgccaat cccacagcgt cctttctcgg gaacaccgtg 65940

gggaaagctg tggcaccagc tccttccttt tgcaactctg atgaatctca cccagggatt 66000

tcaaggcccc tggtcacacc aggatcatag gcctccccca tcccctggac acacagagac 66060

acacctggat tcaggtcagg cctcgcccac tctcggctat atttctcccc aagccgtgtg 66120

tcctcagctg tagaatcagg accataagga agttccctca tagggttctt gtgaggacgg 66180

cacgatttac gtaggggatg ctgacaccgt gcctggcacg tgggacgcac tccacccgcg 66240

gcagccgctc ccatggcttc tcagtgagtt ttccagccac actgcacttc ttagacagga 66300

acactccata cgatgtccct gtcctgcact ggatggccca aaaatctgaa ataagaggag 66360

gagtgcgtgt gaagctccca gtggagcgtt tggcacctgt ccagcatgtc cccaagggca 66420

agtcacggct ctgagattca gtgtctcctt ctgcaaaatg ggccaatagt ggttcctccc 66480

tcccagggct gaagtgagga tgaaatggga taatccaccc ccgtccccac accctgcagg 66540

-continued

tcatcatcat tgctagcagt tgtgtggtgg agcaggtgct cttgagggag cgacacctcc 66600

aggtgctccc ctgccctgct ggcccctctg cagggaggtg acacccaggc cccttcccct 66660

ggggcagcca gctcacgccc gtctctctcc cacaggtgcc gctgcaggtt cctttggcag 66720

taagtacacg cctggggagg gtggccaggg cccccactgc acgagcctct ttgcatgtcc 66780

tggaaaaagc tggagagaaa aaaggggctt cagtgtcccc tctgggactt gggcctattc 66840

actccctcct ctaattacac cccatctgct tctccacctc tccccctcc acctcccccc 66900

ctccacccat ccccacttca catcatatgc catgtgtcat gtgtcatttt gctgtggcct 66960

gtggcccagc aactctcagg ctctcccagg agctccatca gtgctgcttt ggaaaacggg 67020

acaggacttt ttgcaggtct cttggcccct gggtgggctc cctgctcctc ctgccaccca 67080

cgccacttct ctcacctgga tctgggagag cagtctctcc tgccagtcaa gagtggggtg 67140

accttccccc accaggggca gaatccaccc cctagcctaa ccatgggggc agcctccctc 67200

tgggcagcct ctgcagccag cttgtcccag ggctctgctc gtccaggtca gctcaggtcc 67260

caggggagtc ggaccaggga ggggcatctg caggaggtgg gggtcctgag agttccccag 67320

gagggcgagg gcgacatggc gcccacaggt tatcagtaaa tgtcatcgag actgtcccca 67380

gacactcaca gggtgccagg cacggtctct cctttcagcc ttgcaaaccc ctccccctgg 67440

gaggtcgcca tctgctctgc gaggcagcag gagaggactg gccaatgtca aagagccagc 67500

cgggagcaga ccccaaatct cagagatgct tctggggtgc accgtcaccc tccaccaggg 67560

ctctgtgggg ccccacatcc cacccaagtt gtccctcccg gacccagggg gcccctggct 67620

gggaagccag tgagccgaga gggcgccaga aagaagctgg accctgcagg gacgctggtc 67680

tgcacagccg tcgtaagttg cttctctgtg gtgtccccac cccggcaacc ccccaaccct 67740

ctcttgcttt tcccatctct caccaggcat cagcaggtcc cagaaagacc ccgaccccaa 67800

aggccctgtg gccactgcgg ccaccacagc catgacaggg gcccctacta ctcctgtccc 67860

ctccacgtcc actgcctggg cccccatggc gcccagcacc ccacagccca caggtgggtg 67920

ccagggtaca gcgacccctg tcatcccacc ctctcctgct tctagcctgg gtccctgcct 67980

ctcttggggt gggagggtcg gcagccctgg gcagagagca ggggcttggc tcttagaata 68040

gagacgctag aaccctagag ctgggaggcc acaggccaaa ggggcttgag gacacctggg 68100

tcaacctgtt cctgagccca gccaggggat tcagggatca gttcagcttc caaagtcgtc 68160

ttcctcctgc ccttcaagcc attgcttgga agggctccca gaccattgtg gccagacggc 68220

tgcaggaact gagaggaaag gtgctgggggg cagcgaggcc atcctgacat gcagccaaag 68280

actggcctta tctcccaatg gtgcttctgc ctccgtggtc cctggagccc cgcccacacc 68340

ctgtccccac ctggccccca gggcctctct gtccttagcc cctcagcagc acaccggtgg 68400

gatggatgga gcagggttag cccagaaagc aaatgtctct gatcagcagg gcaaagggag 68460

cctctggagc tgagtttgga caccgtgggc tgctgggaat gtggaggctg tgtgtgtagt 68520

gcaaggccag gccagggcca gacgtcctgc cccctcaggg gtctgccaca gacaggcatg 68580

gaaacctgat tctcgctccc ctccaacgga gggattcacg tgtattcaag gctgggggtg 68640

ctggagtggg cctctgctct cacctggact cacctgggga gtatccctgc actctgtgca 68700

gtgcaggtgc caggggtctg aaaggattta tccttcccag agggcaccag gaagacgatg 68760

accaagggga attcttcctg gtcccagcca gggaggggtg ctccaatagc ctgccacacc 68820

-continued

ctgtcccccg ccaccctgca gggaggacct ggtggggact cctggcccct tgggtagtgc 68880

cctggccctc catctctctg atccaaggag acctgcccca ctgatccttc ccccttgggg 68940

ggtggcattt ctaaagggca gagtcccctc catcagctcc tgcctcggcc tgttgctggg 69000

tggacactca ggctccccag acaggggcaa atgctgagag aaagacctcc tccttcctag 69060

gccatccaga gcagctcccc tgggggcagc acaccccacc tctttctaca tccttccttt 69120

tctgcaggag gcatttacag gaggcagggg ctagccaaaa gattggagga tttccgggaa 69180

gcctcctgac ccaggaatcc tctttggggt ggaagacatg ggtcactctg agaattctgg 69240

acttcagaca taggttggcc cagccacaag ggacctgtgc tttgctgatg agcctgtggt 69300

gggcagacag aagcaaaaac agtggtggtg ggtgctgtgc ctgtctccaa acaggggttt 69360

ggctgggagg ccagatactc tccatatcac atgtgcaagt gcacacatgc acacacacac 69420

atgcatgcac acacacaggc atgcacacgc acatgtacac acacacacac acagaggaat 69480

ccatttgcag agctgcttct gacttggtgc cagggcagcc gtgggaggct gggcagattg 69540

tgcaaagttg ggaattaaag aggaaaagtc agaggccaga gtgggaaatg caggggagtt 69600

gagggtcccc aggacccccт cagtgagcag aaggcacacc ctccctctcg gcaagacagt 69660

gctgctctgc accctcagcc ctgtatcaag aagcaggaca ttaggggagg aggtggctcc 69720

aatgtgacag ccagtggccc ctacagccca catctagggg ctcctccctc ctcttcagca 69780

actgaagccc ctgtccagag cccccattaa tgaaaacgat cattgcagta gctgagggtg 69840

agttctcctg ggctgtgctc gtatcattgt atcatcatat cattgtattc tgggctcaca 69900

gctccgtgag atggaggctg ttattttcct agtcccacag gtgaggggat cgaggcttag 69960

gaagaagcag ctggatttta tgatatgtaa attacacctc aatcaagctg tttcagaaga 70020

aaaaaggggc agctgctcaa ggtctcagaa ttatggagag gcacgggcag gatttgaact 70080

cagggctcgc caactcagcc acccaaagct attgtcctga ggcctccagg ggctatgagg 70140

tagagctatc ttttttttтt ttttttgaga tggagtttcg ctcttgtcgc tgaggctgga 70200

gtgcaatgga gcaatctcag ctcactgcaa cctccgcccc cccaggttca agcaattctc 70260

ctgcctcagc ctcccgagta gctgggatta caggcacctg tcaccatgtt cagctacttt 70320

ttgtcttttt agagagacag ggtttcacca tgttggtcag gctggtgttg aactcctgac 70380

ctcaagtgat ccacccgcct cagcctccca aagtgctggg attccaggcg tgagccaccg 70440

cacccggcca agtagtgctg tctccaaggc ctggcttgca gggcttccca gttccaaagg 70500

agcagaccgg gcttccatgg ggccttggca cagcacacag gccatggcga gaacttgctt 70560

cccacacacc tgagtgtgtc cctgggcagc caagccagga ctccctccct ccccaagacc 70620

ctggtccctg aaagatcctg aatacccccg agtgcctccc aacaggtgct tcgggctctt 70680

tgaacagagt ccagctgggc ctctgaactc ctgggccaga tgtttctccc gcctgccaat 70740

gtcaagctgt ctggaggaca gcgctgcggg cggaaaacgc cgctggagac actaatcctt 70800

tcctgggctg ggccacggag gatggaggga gacaggctct gaagcaaatg ccttcagggc 70860

tggctttctc atggctctaa ttaagccctt gccaatttgg gcctggcggc ctcatcttcc 70920

cactgaacat catattaaag tcaattcatg tccaaagctc cccgctccca gctggaaagt 70980

cttccgcact tgttagctgg tagcttttcc ttttctttcc ccacagccac cgttgtgtat 71040

aatcccttca agaagcggaa aacagcagcg ctcccctgtc cctctgggtt tgtcctttga 71100

-continued

aatttgggca cagggcagtt ctttgccagc cctgcctgcc tgccttgctg gctgtgtgtc 71160

ccgttagtct acgggctgag cgttgtgtca ttggttcatg ctggggtccc tggtgaaaat 71220

gggccaggcc aggggtcagg aaggtagaag ggcagtgatc agggaagcag gtcagatgct 71280

ggggaaggct ccggtccctg gattgcggct ggacaggaag gacaccttcc aggacacttc 71340

tggacacatg taagatcttg gccggaacac atgtcccact tcgcagccat tagccagaga 71400

catcagctca gagaggtctg ggcccagagg cgggacctgg tctagctctg tccttcagtc 71460

agaacgggga cggcacaggg agtgtagaag ggtctcgctg aagaatatgc agattctcag 71520

gcatgggttc acctctcatc tatcgggctt taagtctgca tgtgccctcc acaggctgaa 71580

atagtgtaga tgctgcctat gtagtagatt tggacccaat tcctttggcc agtgtagaca 71640

gagcctctcc ttatagtgct gctgcttcta aggggcctgt ggggtgcggg gctgtgatgc 71700

ctcagtatgt acccagcttc cctcagcacc acccccctcgc ataacttggt ttcttctctt 71760

cttcccccca agagtggacc aggccatcta cggctgcccc tctctcgagc aggtggtccc 71820

aggtggcctc ccgtgcagaa ggtatggggg ggcaaggcct gtgatgggcc tgagaccccg 71880

gggaagcgcc ctcttagact cgtaggcccc tccctctgta gtggaagtag caggtgtgca 71940

tggtggggac ctgaggttgg agggggggccg caggaaccaa ctgagggcac gggtgtagaa 72000

tgtcggtgcc tggggagctc tagggcacag tggtgaggga gcggcctggt agagcaggtc 72060

taccagctct gcccccaagc tcacctgctt caagaggttc catgtggcac ccccacgcca 72120

agcccttcca ccagcactcc ctccgagggc ttcggagtct ggtagaggcc ccgcctccca 72180

cgacaggaac ccccctctcc agctgccctt gctcacagga cacctgggca gttgctggat 72240

cagagagtca gaggggggctt cctgcaggag cgggggccat gagacctcgg agggtggact 72300

gtggtgggtg aagggagaag gcagcacatt ccaggccgca gggccagccg gggcaaaggc 72360

ttggcagtgg gatggcaggg agcctgacaa agtggaaaat gtgtgggtta aaggagggag 72420

ggcggggtcc tggaagacac tgacatcctc ctgctacgtg ggaggagaca cagggctcat 72480

ctgtagccat agacagacat gccaaggaaa cgcgcaggcc tgcccgactc tccagaaggg 72540

aaattgtccc tggccccagc tcaccaagcc tgggtgggga attagggcct gaggtctagg 72600

gaacaggtga gctgttcctt ccagctcaca tgttcaaatt tcctccagcc ccagctctga 72660

gcagcgagca gggctttgag cgccctctac tggcaggaag ctctggcgct ggaagcatgt 72720

ttagagaggg tctgaggctc ggttcctaga aacctggagg acctgggcct ggtgtcctct 72780

gtggtgatgg agacagagct ggcgggagcc atcgcttccc taccctgggc caaccagggc 72840

accacagacc cccagaggga agccaaggta gtgacgatcc cgggacagtg gcctgctcac 72900

ccacagatag ggcgttgggg tcccagcggg attctgggca gtggaaggca ggtgccgtcc 72960

gtgttcctgg cttgacagca cttgcgagtg ggactccagg gacagcgaag gattcacttc 73020

ggctggagca ggaagagtgt ttcagaaagg aagggagatg ccaaagtcct taaatgccaa 73080

gtttagtctc tgggtttgat gctccaggaa gtttggagag gcggtgggga gagcaagaga 73140

cgggcgtggt gtgcaatgtg atgtcaatct atctaaaaac agtttggctt ccaagaaggt 73200

cttagcaggg cgcgggggtg tcagggggtta cagaagtcat ttggaggatt aatccagcca 73260

gatgtgtcca tggtctcaga gaggggacca agggcagggc tgatttgcaa gcttgggatg 73320

tgctgtgttt ccttcaggaa ggggccccac ctccctgggc tcttcgagga gaggggctgt 73380

-continued

gtgatttgag gccagagggg cctctccctc cctcacatct gagcaggcga caaggctgcc 73440

tgccctagag ctggcccagg gcggctcgga agcctttgct gggctcttcc ctgggcagtg 73500

ggaccatgac agacgaaaga acctgtttct catctctcca agctgtgggc acccctgccg 73560

ctgcccctgc ccctgccaag ggctacaaac ttttccagct caagcccaaa tctcctcaag 73620

tgatgcctat tgaagaattc caaggtaaga ggatggacct ggggccccat cagccctccc 73680

tgacacctgt tccccatccg ccgctggaaa aagacggtgc aggatagagg accgatgcct 73740

ggctccgaaa accctcctgg agtagctggg tcaaggttaa actgagtctc tcttccctac 73800

aggcctccct ccccaaggga gctgggagca ggtatgagtc agaagccaac ttgggcacag 73860

tgggcaggcc acacagcagg cagagcagat gccagaaata gcccatcccg gctcccctgg 73920

gaggtgtggc cctggggctc gtgttggttg aagcagaatc tgggacacac gggtcaccga 73980

tgctgctctt tgggacactt agaggatgcc tcatctcctc attatctctg gagggacaaa 74040

gtgaaggggg caggactagg tggcccacag gtgggagtgc ccaccatctc tcctgggcac 74100

aggctgtttc tctagtctcc catgcccttg accactgggt cagtccctca tcccatcaca 74160

aaagggaagc tgggtcctct agagatacac agatggtgtt tcaagagggt ggccgttgtc 74220

cttccttgtt cgggggcagc cacattggct ttcttgctgg agggtgggtg ggtgggtgag 74280

tactgtgtcc cttcgtagga acatcaaggg atgccccccc attcttaggg atggtgacct 74340

tcctcaccaa atcctccatt gacaatgtgg gattcacctc caatccctga gagccttgcc 74400

ccaggcagtc acgggcttgt ctggtccttg gagcggagct ggttaggcag gggtcagcct 74460

gagaaccacg taggggtggg gtgcaggagg cggcaggaca tggtggtggt ggtccttggt 74520

atgaaaccat gtgcttccag gagcagcgag tcagaagccg ggccaggacc agggggaggc 74580

atgcaggttc ccagggctcc tgctttaaag tggcactcac tcttagcatc ctgcaaaaca 74640

atcaaacttg cagaaagctc aggctaataa gaaagggtct ggcaggtggg cgttttcctc 74700

ccagccatct tccaaagcag catgggcagg agctcctggc ccattgcatc ttgtccagcg 74760

tccatccatg cattcatcta cccgaggata ccacggcgag cgccgtgaac ccaggcgtcg 74820

ccctccccca gtgcacagcc aggtggcatg accctgccct ccttgcatga atcactttct 74880

aatcaccccg gcatgtgggc atccttcagc gagcgcttgg ccctggtgcc cagccaggca 74940

ttagcaggag ctgcccactg gccctccctg gttccctgcc cacagggcca ggtgggaatc 75000

cctgggctca gcctactcag gttctcctct gggctcaaag cagggaggcc tctctcttcc 75060

tgaatccgat ggaagggtgg gaggcctagg gcaccttccg gtacctttтc caaagatgcc 75120

ttcctccgtc cctgcatgac ctggggtgag tccttcctcg ccctgtccct cagtttccct 75180

gaatgctcgc tgaccattgg tatttctccc acttggccgg cccagactgc gaatgctacg 75240

gtcactccaa ccgctgcagc tacattgact tcctgaatgt ggtgacctgc gtcagctgca 75300

agcacaacac gcgaggtcag cactgccagc actgccggct gggctactac cgcaacggct 75360

cggcagagct ggatgatgag aacgtctgca ttggtgagag ggcacggaca cggcacaggg 75420

aacttgctgg aatgcgtgca gggtgcactg ccctgcgagg tggcctctgg ggcccctgc 75480

atcagaatca cctggggaga ctgtgggaat tctaactcca gggccctctc cagttgagca 75540

tctctaagga cagaaagctc cagaaactgc tctattagta acctaccctt gcggttctcc 75600

ggtaagtttt gcactggagt tgcaaaactt accagtggcc cttccctctc tgggcaactg 75660

-continued

gaggggacac tgacccttcc tggctccaaa gagctgtgac tctggcaggt ggcaggcact 75720

cagtggcaga ggccactgag catctgtctg gggctggtgt gtgggggggt ccccctccat 75780

agctcctttc cagaaaggtg gaggagcagc ctatccctcc tcctgcaggg gcccagttgg 75840

gggccaaaag atcgccttgc tgcgtgcatt tgtgcaagtc ccttcccgtt cctgggcctc 75900

agcttcctca ttcatcaaat tgggaggcag atcagatcaa aggttttcag ctcttttttg 75960

tggctgaagc ttttcttcaa atgctttacc agcccaggtc cagctataaa gctgctcttc 76020

acccctggtg ggcacccagt ctgctttctt ccaagttcta ctcaaggact ggcttttggg 76080

tagagaagga agtccatcag ggccctgggc ctgggcaaag accaaagcca tgaccgccaa 76140

ccaaaacgca ccagcctgga atggttgccc ctgtcgtcag tagaggccag gtctcggcct 76200

caggggctgt cccccaaccc tgcccagcca ggccccttgg gacaccatca cccatccccc 76260

acccagcagg aggctctggc tgcccagagg aggggctcct gcaaagctgg agctgtcggt 76320

ctgaattctg gcggcagcct tcagataatt ccatcaactc taagtgatca aagccgctga 76380

cgtcacaggg ggccagctgc agggacaggg cagggccttt ggatccaatt agaggtgccc 76440

acaccctggc accctcctcc tctccctggc tctccctgcc tccaccccga gagccagcac 76500

tgagctgcaa ggtttctcag ggtggacgat attcaccctc tcccacagag ccccaaggca 76560

accaactggg cccaccccgg gagcaggaat aggctgttcc tccactcccc tgcaaaggag 76620

ctatggaggg gggccacccc acaacacagc agccccagac atgctcagtg gcctctgctg 76680

agtttctgcc acctgtcgga gtcatagctc tttggagatg ggaaggacag cgacccctct 76740

agttgcccag agaggggaag gggctgaccc aggccacacc agtgccaggg cggggaaggt 76800

ggggctggga cgtgtttgat cccaaggaag gaagccagag tcttctctcc aggcctggcc 76860

accctgggaa gtccccacct gccgtccagc cgcgggctca cgtggaccca gtgtggggag 76920

catcccctgg ggagtgtgga gatgctccct gcgaggccgg gagagtgggg gtccgagaag 76980

acggcgccca cacgtagccc tgaccgcgcg cccgtgcccg tgtccgtcca gagtgtaact 77040

gcaaccagat aggctccgtg cacgaccggt gcaacgagac cggcttctgc gagtgccgcg 77100

agggcgcggc gggccccaag tgcgacgact gcctccccac gcactactgg cgccagggct 77160

gctaccgtga gtgcgcgccg tcccccgtgg gcggggcctg cggaaagggg acggggcagg 77220

accgaggcag tgggcggggc ctagtgggac ggggcagggg cggtggactg ggcctagcaa 77280

gacggggcag ggccggggaa gtgggtgggg cctagtggga cggggaagag gcggtgggcg 77340

gggctcgcga gacggggcag ggccggggca gtgggtgggg cctagatgag agcggggcag 77400

ggttgggata gttggcaggg gcctggtgag atggggccga cccgggggcg gtggacgggg 77460

cctagcgaga cggagctggc aggtgggcgg ggacaggatg ctgctgaggt ccggggccgg 77520

gccgaggggc gggtccaaga gctcggggcg gggcctgatg cgacctgagg cacggtggtg 77580

cctggtggga actacgagaa agaccgagct ggggttggtt ggaaaggtat ttgcggggac 77640

agagggaggg aggctgtcca agtcggcgtt agccgcgggc acagggtgaa aggaggctcc 77700

aggcgcgtgg aacagcacgt gcacagctct ggagactgca ggcgcgtctg aagaacagca 77760

ccgaggccag tggggcgggg agagaggggc agcggtggga ggcagccggg ggccagtatc 77820

tcgcccgggc gccgtcaccc tccgaggggg gacgtttcgc acccagcgcg cctggagcct 77880

cctacatccc cggcccagac ggcgcccccg ggatctcgca caccctgctt cgcaggagct 77940

-continued

cggaggttgg cgggggggacc gggccacccc ccgtgctgac cgccccctcc gcctgcagcc 78000

aacgtgtgcg acgacgacca gctgctgtgc cagaacggag gcacctgcct gcagaaccag 78060

cgctgcgcct gcccgcgcgg ctacaccggc gtgcgctgcg agcagccccg ctgcgacccc 78120

gccgacgatg acggcggtct ggactgcgac cgcgcgcccg gggccgcccc gcgccccgcc 78180

accctgctcg gctgcctgct gctgctgggg ctggccgccc gcctgggccg ctgagccccg 78240

cccggaggac gctccccgca cccgggagcc gggggtcccg gggtcccggg gcggggccgg 78300

cgtccgaggc cgggcggtga gaagggtgcg gcccgaggtg ctcccaggtg ctactcagca 78360

gggccccccg cccggcccgc gctcccgccc gcactgccct ccccccgcag caggggcgcc 78420

ttgggactcc ggtccccgcg cctgcgattt ggtttcgttt ttcttttgta ttatccgccg 78480

cccagttcct tttttgtctt tctctctctc tctttttttt ttttttttc tggcggtgag 78540

ccagagggtc gggagaaacg ctgctcgccc cacaccccgt cctgcctccc accacactta 78600

cacacacggg actgtggccg acacccccctg gcctgtgcca ggctcacggg cggcggcgga 78660

ccccgacctc cagttgccta caattccagt cgctgacttg gtcctgtttt ctattcttta 78720

tttttcctgc aacccaccag accccaggcc tcaccggagg cccggtgacc acggaactca 78780

ccgtctgggg gaggaggaga gaaggaaggg gtgggggggcc tggaaacttc gttctgtaga 78840

gaactatttt tgtttgtatt cactgtcccc tgcaaggggg acggggcggg agcactggtc 78900

accgcggggg ccgatggtgg agaatccgag gagtaaagag tttgctcact gctgcctcca 78960

cggcctgttt tctttctgtg ttggggacgg tgggcaggtg tggggcttac agaggaatcc 79020

acaacacagc cttaaagaaa cggtttccct actggggcca ccatttccct gggcctttct 79080

gtggattcca gcaggcagtg ccccctcccc gcaggcttgg ctggcagagt tttccacccc 79140

gcggccaggc tgcaggtgcc ccacctgtta ggagcctccc cacactgaaa ggctgcctcc 79200

ctcctttccc aaaaaagaaa tccggagtgt attggccctt ttctacagaa gtccaaggga 79260

aatgactcag ggagaatcct agcagaggtt gaatccaatg ctctgattta tactgtgtct 79320

cggtggccac ctccgatgga tgtgtcatct cagacctgtt gcagccggag cctcaagtcc 79380

aatatcagat gaagctgaac ccacaattcg gccaccgcct ccttccagag tttcagatgg 79440

ccaggtgggc agaggcgggc agtgcagaga ccccagacgt gccggccctg tcctccctac 79500

cttctcaaga ttaggaaggg gtgctggagg ggacaggggc agcttgggag tggtgaggaa 79560

gctcctagat tcggggctca tcccctgggg cctctgattc agaggatcca gcagttctcc 79620

catctccgct tggtgtctcc agccctgggg ccacacttcc ccctcggtcc agcctcctgt 79680

ccacctatgt ttatttcaga gcagtgccgg gggtccggtc ctggttgcta actgctgcca 79740

ctgctccacc tgcaggtgct cccagcactg gcttctgcca ccacacctgt tctttcccag 79800

ctgcgaggtt tagacctggg tccttccctt gagtccccaa agctaagcta agaccaagtg 79860

gaacaaactt ggccttgggg acagcaggag attacaacac agaaaaggag ggggaggcac 79920

aacgggacac tgcataggac tcacagtgtc ccgagcccac acaccagccc ctcctggcct 79980

cctctctctg ctcccacccc cagcaccctg ctgacccgga agtgccttcc gacaggccct 80040

gcatcctcct gcagctggcc catctctacc ctcacattct tccttacgca cagaacccca 80100

ctgcctggga cccaaagtgc ccagataaaa taaaacacct ctctggggct tcttgtggat 80160

aggagtggcc aggggacaca gctctgggca gtgagatgta agcaggactg atgggtgggc 80220

-continued

```
ttccagaaag ttcttaaaag tcatcccttc ctccacgccc cacaaagcct cagttgtcca  80280

agtttgtggc tctggttttt ctctctcatc cctccctctg ctcttcagtc agcaggatgg  80340

ggaaggagcc ttcttgagac actgagcata acagtgtcac cctcaggata gtgaagggtg  80400

gagccggggg tggaagttgc agagcctggg acacctgcct tggctgcaac ctctgcactg  80460

cttttattta tctacttatt tttatagagt tgaggtctca ctatgttgcc caggctggtc  80520

tcaaactcct aggctcaagt gagcctccgg ccttggcctc ccgaagtgct gagattacag  80580

gcatgagcca ccgcaccacc tgcctgtctt actgtgtgag aaaaaaaata aacggtgatg  80640

tgattaaaac actagaattt gggttttctg ttccatgcac actttccaag ttcttggccc  80700

tgcctgattg gcatccgggc cccctgagcc ctctaaggcc agagagagac tgaaacccac  80760

ccttccttgg ccgcagcccc ccatggtgac cctcaagtca ccaggaagag gaatgtccat  80820

gactcaggcc gaggagtcca cctctgtcca gaacaggact gcccctgcct tctagtgagc  80880

acagggcagc tgagcagaac tcaccccaga gaaacagcag cgctggccgc agtcggtaaa  80940

cagcaggcat ttcctccta                                               80959
```

<210> SEQ ID NO 4
<211> LENGTH: 510
<212> TYPE: PRT
<213> ORGANISM: Mus musculus

<400> SEQUENCE: 4

```
Gly His Tyr Asp Val Cys Lys Ser Leu Ile Tyr Thr Glu Glu Gly Lys
 1               5                  10                  15

Val Trp Asp Tyr Thr Ala Cys Gln Pro Glu Ser Thr Asp Met Thr Lys
            20                  25                  30

Tyr Leu Lys Val Lys Leu Asp Pro Pro Asp Ile Thr Cys Gly Asp Pro
        35                  40                  45

Pro Glu Ser Phe Cys Ala Met Gly Asn Pro Tyr Met Cys Asn Asn Glu
    50                  55                  60

Cys Asp Ala Ser Thr Pro Glu Leu Ala His Pro Pro Glu Leu Met Phe
65                  70                  75                  80

Asp Phe Glu Gly Arg His Pro Ser Thr Phe Trp Gln Ser Ala Thr Trp
                85                  90                  95

Lys Glu Tyr Pro Lys Pro Leu Gln Val Asn Ile Thr Leu Ser Trp Ser
            100                 105                 110

Lys Thr Ile Glu Leu Thr Asp Asn Ile Val Ile Thr Phe Glu Ser Gly
        115                 120                 125

Arg Pro Asp Gln Met Ile Leu Glu Lys Ser Leu Asp Tyr Gly Arg Thr
    130                 135                 140

Trp Gln Pro Tyr Gln Tyr Tyr Ala Thr Asp Cys Leu His Ala Phe His
145                 150                 155                 160

Met Asp Pro Lys Ser Val Lys Asp Leu Ser Gln His Thr Val Leu Glu
                165                 170                 175

Ile Ile Cys Thr Glu Glu Tyr Ser Thr Gly Tyr Ser Thr Asn Ser Lys
            180                 185                 190

Ile Ile His Phe Glu Ile Lys Asp Arg Phe Ala Phe Phe Ala Gly Pro
        195                 200                 205

Arg Leu Arg Asn Met Ala Ser Leu Tyr Gly Gln Leu Asp Thr Thr Lys
    210                 215                 220
```

-continued

```
Lys Leu Arg Asp Phe Phe Thr Val Thr Asp Leu Arg Ile Arg Leu Leu
225                 230                 235                 240

Arg Pro Ala Val Gly Glu Ile Phe Val Asp Glu Leu His Leu Ala Arg
                245                 250                 255

Tyr Phe Tyr Ala Ile Ser Asp Ile Lys Val Arg Gly Arg Cys Lys Cys
            260                 265                 270

Asn Leu His Ala Thr Ser Cys Leu Tyr Asp Asn Ser Lys Leu Thr Cys
        275                 280                 285

Glu Cys Glu His Asn Thr Thr Gly Pro Asp Cys Gly Lys Cys Lys Lys
    290                 295                 300

Asn Tyr Gln Gly Arg Pro Trp Ser Pro Gly Ser Tyr Leu Pro Ile Pro
305                 310                 315                 320

Lys Gly Thr Ala Asn Thr Cys Ile Pro Ser Ile Ser Ser Ile Gly Asn
                325                 330                 335

Cys Glu Cys Phe Gly His Ser Asn Arg Cys Ser Tyr Ile Asp Leu Leu
            340                 345                 350

Asn Thr Val Ile Cys Val Ser Cys Lys His Asn Thr Arg Gly Gln His
        355                 360                 365

Cys Glu Leu Cys Arg Leu Gly Tyr Phe Arg Asn Ala Ser Ala Gln Leu
    370                 375                 380

Asp Asp Glu Asn Val Cys Ile Glu Cys Tyr Cys Asn Pro Leu Gly Ser
385                 390                 395                 400

Ile His Asp Arg Cys Asn Gly Ser Gly Phe Cys Glu Cys Lys Thr Gly
                405                 410                 415

Thr Thr Gly Pro Lys Cys Asp Glu Cys Leu Pro Gly Asn Ser Trp Tyr
            420                 425                 430

Tyr Gly Cys Gln Pro Asn Val Cys Asp Asn Glu Leu Leu His Cys Gln
        435                 440                 445

Asn Gly Gly Thr Cys Gln Asn Asn Val Arg Cys Ala Cys Pro Asp Ala
    450                 455                 460

Tyr Thr Gly Ile Leu Cys Glu Lys Leu Arg Cys Glu Glu Ala Gly Ser
465                 470                 475                 480

Cys Gly Ser Glu Ser Gly Gln Gly Ala Pro Pro Arg Gly Ser Pro Ala
                485                 490                 495

Leu Leu Leu Leu Thr Met Leu Leu Gly Thr Ala Gly Pro Leu
            500                 505                 510


<210> SEQ ID NO 5
<211> LENGTH: 260
<212> TYPE: PRT
<213> ORGANISM: Human

<400> SEQUENCE: 5

Leu Tyr Lys Tyr Phe Tyr Ala Ile Ser Asn Ile Glu Val Ile Gly Arg
 1               5                  10                  15

Cys Lys Cys Asn Leu His Ala Asn Leu Cys Ser Met Arg Glu Gly Ser
            20                  25                  30

Leu Gln Cys Glu Cys Glu His Asn Thr Thr Gly Pro Asp Cys Gly Lys
        35                  40                  45

Cys Lys Lys Asn Phe Arg Thr Arg Ser Trp Arg Ala Gly Ser Tyr Leu
    50                  55                  60

Pro Leu Pro His Gly Ser Pro Asn Ala Cys Ala Ala Ala Gly Ser Phe
65                  70                  75                  80
```

-continued

```
Gly Asn Cys Glu Cys Tyr Gly His Ser Asn Arg Cys Ser Tyr Ile Asp
                85                  90                  95

Phe Leu Asn Val Val Thr Cys Val Ser Cys Lys His Asn Thr Arg Gly
            100                 105                 110

Gln His Cys Gln His Cys Arg Leu Gly Tyr Tyr Arg Asn Gly Ser Ala
        115                 120                 125

Glu Leu Asp Asp Glu Asn Val Cys Ile Glu Cys Asn Cys Asn Gln Ile
    130                 135                 140

Gly Ser Val His Asp Arg Cys Asn Glu Thr Gly Phe Cys Glu Cys Arg
145                 150                 155                 160

Glu Gly Ala Ala Gly Pro Lys Cys Asp Asp Cys Leu Pro Thr His Tyr
                165                 170                 175

Trp Arg Gln Gly Cys Tyr Pro Asn Val Cys Asp Asp Asp Gln Leu Leu
            180                 185                 190

Cys Gln Asn Gly Gly Thr Cys Leu Gln Asn Gln Arg Cys Ala Cys Pro
        195                 200                 205

Arg Gly Tyr Thr Gly Val Arg Cys Glu Gln Pro Arg Cys Asp Pro Ala
    210                 215                 220

Asp Asp Asp Gly Gly Leu Asp Cys Asp Arg Ala Pro Gly Ala Ala Pro
225                 230                 235                 240

Arg Pro Ala Thr Leu Leu Gly Cys Leu Leu Leu Leu Gly Leu Ala Ala
                245                 250                 255

Arg Leu Gly Arg
            260
```

That which is claimed is:

**1**. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence shown in SEQ ID NO: 2;

(b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO: 2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS: 1 or 3;

(c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO: 2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS: 1 or 3; and

(d) a fragment of an amino acid sequence shown in SEQ ID NO: 2, wherein said fragment comprises at least 10 contiguous amino acids.

**2**. An isolated peptide comprising an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence shown in SEQ ID NO: 2;

(b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO: 2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS: 1 or 3;

(c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO: 2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS: 1 or 3; and

(d) a fragment of an amino acid sequence shown in SEQ ID NO: 2, wherein said fragment comprises at least 10 contiguous amino acids.

**3**. An isolated antibody that selectively binds to a peptide of claim 2.

**4**. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO: 2;

(b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO: 2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS: 1 or 3;

(c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO: 2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS: 1 or3;

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO: 2, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

**5**. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO: 2;

(b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO: 2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS: 1 or 3;

(c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO: 2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS: 1 or 3;

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO: 2, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

**6**. A gene chip comprising a nucleic acid molecule of claim 5.

**7**. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.

**8**. A nucleic acid vector comprising a nucleic acid molecule of claim 5.

**9**. A host cell containing the vector of claim 8.

**10**. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.

**11**. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.

**12**. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows detection of the presence of the peptide in the sample and then detecting the presence of the peptide.

**13**. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.

**14**. A method for identifying a modulator of a peptide of claim 2, said method comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.

**15**. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.

**16**. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.

**17**. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.

**18**. A method for treating a disease or condition mediated by a human secreted protein, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.

**19**. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.

**20**. An isolated human secreted peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence shown in SEQ ID NO: 2.

**21**. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence shown in SEQ ID NO: 2.

**22**. An isolated nucleic acid molecule encoding a human secreted peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule shown in SEQ ID NOS: 1 or 3.

**23**. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule shown in SEQ ID NOS: 1 or 3.

* * * * *